



**ACTAS DEL
XXIX
CONGRESO NACIONAL
DE ESTADÍSTICA E
INVESTIGACIÓN OPERATIVA
Y DE LAS
III JORNADAS DE
ESTADÍSTICA PÚBLICA**

Joaquín Sicilia Rodríguez

Carlos González Martín

Miguel Angel González Sierra

David Alcaide López de Pablo

Editores

Actas del XXIX Congreso Nacional de
Estadística e Investigación Operativa y
de las III Jornadas de Estadística Pública

Editores:

Joaquín Sicilia Rodríguez
Carlos González Martín
Miguel Angel González Sierra
David Alcaide López de Pablo

ISBN: XXX-XXXXXX-X

Depósito legal:

Edición Técnica: Manuel D. García Román

Diseño de portada y contraportada: Carlos González Alcón

Impreso en La Laguna, mayo de 2006

XXIX CONGRESO NACIONAL
DE
ESTADÍSTICA E INVESTIGACIÓN OPERATIVA
Y
III JORNADAS DE
ESTADÍSTICA PÚBLICA

Organizado por:



Departamento de Estadística, Investigación Operativa y Computación



Sociedad de Estadística e Investigación Operativa



Instituto Canario de Estadística

Tenerife, 15 – 19 de mayo de 2006

Comité Organizador

Presidente

Joaquín Sicilia Rodríguez

Álvaro Dávila González

Carlos González Martín

Miguel Ángel González Sierra

David Alcaide López de Pablo

Carlos González Alcón

Sergio Alonso Rodríguez

Carlos Duque Gómez

Antonio Sedeño Noda

Enrique González Dávila

Arturo Fernández Rodríguez

Marcos Colebrook Santamaría

José Miguel Gutiérrez Expósito

Mercedes Suárez Rancel

Natividad Jiménez Saavedra

Beatriz Abdul-Jalbar Betancor

Roberto Dorta Guerra

Andrés Rodríguez González

Carlos Pérez González

Felipe Manuel Rosa González

Miguel Ángel Domínguez Ríos

Manuel García Román

Comité Científico

Presidente

Joaquín Sicilia Rodríguez

Juan Luis Moreno Rebollo

Marco Antonio López Cerdá

Ignacio García Jurado

Leandro Pardo Llorente

María Ángeles Gil Álvarez

Jesús López Fidalgo

Susana Montes Rodríguez

Juan García Laguna

Concepción Maroto Álvarez

Carlos González Martín

Miguel Ángel González Sierra

David Alcaide López de Pablo

Blas Pelegrín Pelegrín

Mariano J. Valderrama Bonnet

Emilio Carrizosa Priego

Gustavo Bergantiños Cid

Francisco Hernández Jiménez

Presentación

Este volumen recoge las versiones resumidas de los trabajos presentados en el XXIX Congreso Nacional de Estadística e Investigación Operativa y en las III Jornadas de Estadística Pública. A lo largo de la semana del 15 al 19 de mayo del 2006 se han expuesto en las diversas sesiones del congreso más de trescientas cincuenta comunicaciones, que cubren una amplia representación de la investigación española en el área de Estadística e Investigación Operativa. Además, se puede observar que nuestro congreso también es apreciado y reconocido fuera de nuestras fronteras. Prueba de ello, es la asistencia al mismo de profesores e investigadores de varios países europeos como son Italia, Portugal, Finlandia, Chipre, Irlanda, Hungría y de países latinoamericanos como México y Venezuela.

La edición de estas actas ha sido posible gracias al esfuerzo personal de los miembros encargados del equipo organizador y a la contribución económica de las instituciones colaboradoras. Así, queremos en primer lugar agradecer al Instituto de Estadística de Canarias, a la Universidad de La Laguna y a la Sociedad Nacional de Estadística e Investigación Operativa el apoyo constante y la ayuda inestimable que hemos recibido para la organización de este evento. También expresamos nuestro agradecimiento al resto de las instituciones y entidades colaboradoras, cuyos logotipos figuran en la contraportada del libro, las cuales nos han ayudado para que el Congreso tuviera el éxito deseado.

También deseamos expresar nuestro más sincero reconocimiento al Profesor D. Manuel García Román, editor técnico de estas actas, sin cuya labor hubiera sido impensable abordar la edición de las mismas.

Por último, queremos agradecer profundamente a los casi cuatrocientos congresistas inscritos en el congreso, su asistencia y presentación de trabajos, porque ellos son los verdaderos protagonistas y los que han hecho posible, con su presencia en Tenerife, la celebración de este evento. Gracias sinceramente a todos.

Joaquín Sicilia Rodríguez
Carlos González Martín
Miguel Ángel González Sierra
David Alcaide López de Pablo

Departamento de Estadística,
Investigación Operativa y Computación
Universidad de La Laguna

Índice de Trabajos

Conferencias Plenarias 1

- Computational strategies for deriving the best-subset regression models,
Erricos J. Kontoghiorghes¹, Cristian Gatu², Marc Hofmann², Petko Yanev³ 3

- Mathematical Programming Problems with Linear Complementarity Constraints, *Joaquim J. Júdice* 5

- Partisan Drawing of Electoral Districts: How bad can it be?, *N. Apollonio¹, R.I. Becker², Isabella Lari¹, Federica Ricca³, Bruno Simeone¹* 7

- Small Area Statistics: Methods and Applications in the Finnish Public Statistics, *Risto Lehtonen* 9

Sección Estadística 11

- A Family of Symmetric Circular Distributions, *M.C. Jones¹, Arthur Pewsey²* 13

- A new Multiblock PLS based method to estimate causal models, *Francisco Arteaga¹, Martina G. Gallarza², Irene Gil³* 15

- A simulation study of a nested sequence of multinomial loglinear models, *Nirian Martín¹, Leandro Pardo²* 17

- ABC-PLOT 1.1 Aplicación informática para la representación gráfica de datos multidimensionales, *C.A. Ávila Zarza¹, S. Rodríguez González², R.M. Martín Moreno³* 19

- Adaptive Combination of Forecast, with Application to Wind Energy, *Ismael Sánchez¹* 21

Adaptive Goodness-of-Fit Tests II: ARE-aided Basis Choice, <i>Aurea Grane</i> ¹ , <i>Josep Fortiana</i> ²	23
Ajuste de modelos lineales mixtos con efectos aleatorios anidados, <i>Domingo Morales</i> ¹ , <i>Agustín Pérez</i> ¹	25
Algorithm based on the residual variance of a linear model to find gene expression profiles of de-regulation and identify families of disease-altered genes, <i>M. J. Rivas-López</i> ¹ , <i>J. M. Sánchez-Santos</i> ¹ , <i>J. López-Fidalgo</i> ² , <i>C. Prieto-Sánchez</i> ³ , <i>J. de Las Rivas-Sanz</i> ³	27
Algoritmo de Balas Adaptado a la Detección e Imputación Automática de datos cualitativos (ABADIA), <i>Pedro Martínez Talaván</i>	29
Análisis bayesiano de un sistema de colas $G/M/1$ con proceso de llegadas de cola pesada, <i>Pepa R. Cobo</i> ¹ , <i>Rosa E. Lillo</i> ² , <i>Michael P. Wiper</i> ³	31
Análisis Canónico de Correspondencias de tres vías: Metodología y Aplicación, <i>José L. Vicente Villardón</i> ¹ , <i>Ángela Cortés</i> ² , <i>Mª José Fernández</i> ¹ , <i>Purificación Galindo</i> ¹	33
Análisis comparativo de distintas técnicas de apoyo a la interpretación de clusters sobre un caso real, <i>Karina Gibert</i> , <i>Gustavo Rodríguez</i> , <i>Gastón García</i>	35
Análisis comparativo provincial de los movimientos migratorios de la Comunidad de Castilla y León: Proyección de las inmigraciones, <i>P. Galindo</i> , <i>C. Ávila</i> , <i>I. Barrera</i> , <i>A. Blázquez-Zaballos</i> , <i>M.J. Fernández-Gómez</i> , <i>J. Martín-Vallejo</i> , <i>P. Vicente</i> , <i>J.L. Vicente-Villardón</i>	37
Análisis de datos composicionales con el CoDa-dendrograma, <i>V. Pawlowsky-Glahn</i> ¹ , <i>J.J. Egozcue</i> ²	39
Análisis de influencia conjunta en los subespacios generados por componentes principales, <i>A. Enguix-González</i> ¹ , <i>J.L. Moreno-Rebollo</i> ² , <i>J.M. Muñoz-Pichardo</i> ³	41
Análisis de la eficiencia en el muestreo estratificado. Aplicación al muestreo de la base de datos del impuesto sobre la renta de las personas físicas de la agencia tributaria, <i>César Pérez López</i> ¹ , <i>Santiago Díaz de Sarralde</i> ²	43

Análisis de la persistencia del IBEX-35, <i>Beatriz Rodríguez Flores</i>	45
Análisis del Comportamiento del RV (STATIS) y el Lg (AFM), ante cambios Alpha, Beta y Gamma, <i>P. Vicente¹, N. Bacala², P. Galindo³, M. J. Fernández-Gómez⁴</i>	47
Análisis espacio-temporal de la Fecundidad en la Comunidad de Castilla y León, <i>M.P. Galindo, C.A. Ávila-Zarza, I. Barrera, A. Bázquez-Zaballos, M.J. Fernández-Gómez, J. Martín-Vallejo, M.P. Vicente-Galindo, J.L. Vicente-Villardón</i>	49
Análisis Estadístico de Habilidades Motoras para la Natación en Niños., <i>R. Barea-Barrera, J.M. Fernández-Ponce, M.T. Gómez-Gómez, A. Ortiz-Bustos, J.L. Pino-Mejías</i>	51
Análisis estadístico sobre la inmigración ilegal en Melilla, <i>Juan Antonio Marmolejo Martín¹, Francisco Abad Montes², María Dolores Huete Morales³</i>	53
Análisis Factorial Múltiple y STATIS: una Aplicación a Datos de Afluencia Turística, <i>Sara Morgado Nunes¹, Juan Moncada Herrera², M. Purificación Galindo³</i>	55
Análisis Sensorial en el control y la mejora de procesos textiles. Caso de la Manta, <i>Pau Miró i Martínez, José Miguel Carot Sierra, José Manuel Jabaloyes Vivas, Mónica Martínez Gómez</i>	57
Analysis of Principal Alarms in Statistical Quality Control, <i>Isabel González¹, Ismael Sánchez²</i>	59
Aplicación de wavelets para el modelado de la serie de precios de la energía eléctrica: un enfoque bayesiano, <i>Elena Domínguez¹, José Mira², Carolina García-Martos³, María Jesús Sánchez⁴</i>	61
Aplicación de Modelos de Frontera Estocástica a Pequeñas Empresas, <i>Felipe M. Rosa-González¹, Antonio Arbelo-Álvarez², Enrique González-Dávila¹</i>	63
Aplicación de las redes neuronales y del análisis de la sensibilidad como método para la mejora de la calidad en el proceso del moldeo por inyección de termoplásticos, <i>Elena Pérez Bernabeu¹, Miguel Ángel Sellés Canto²,</i>	

Vicente Jesús Seguí Llinares ³ , Miguel Jorge Reig Pérez ⁴	65
Aplicación de técnicas composicionales al estudio del comportamiento de cerdas gestantes, J. Daunis-i-Estadella ¹ , G. Mateu-Figueras ² , N. Chapinal ³ , X. Manteca ⁴ , J.L. Ruiz de la Torre ⁵	67
Aplicación de técnicas estadísticas multivariantes a la identificación de vinos mediante descriptores aromáticos, Antonio Vaamonde Liste ¹ , Patricio Sánchez Bello ² , Patricio Sánchez Fernández ³	69
Aproximación de funciones de clasificación, Beatriz Lacruz ¹ , Rosa Eva Pruneda ²	71
Aproximaciones en tests χ^2 y F con modelos cercanos a la normal. Aplica- ciones en Robustez, A. García-Pérez	73
Bayesian estimation of the Gaussian mixture GARCH model, María Con- cepción Ausín ¹ , Pedro Galeano ²	75
Bayesian latent multinomial models: A case study, Javier Portela ¹	77
Biplot Interactivo como herramienta para el diagnóstico de la interacción en tablas de tres vías, A. Blázquez-Zaballos ¹ , J. L. Vicente-Villardón ¹ , M. Varela ² , P. Galindo ¹	79
Biplot: Una alternativa a la regresión lineal multivariante, Olesia Cárdenas ¹ , Purificación Galindo ² , José L. Vicente-Villardón ³	81
Bootstrap para la media bajo datos con sesgo por selección, Jorge Luis Ojeda Cabrera ¹ , José Tomás Alcalá Nalváiz ² , José Antonio Cristóbal Cristóbal ³	83
Bootstrap semiparamétrico para Modelos Lineales Mixtos Generalizados, José A. Sánchez-Espigares, Jordi Ocaña	85
Caracterización Multivariante de los perfiles de las mujeres en situación laboral irregular: el caso de Salamanca, P. Vicente ¹ , C. Patino ² , P. Galindo ³	87
Characterization and spatial distribution of heavy metals in soils irrigated	

with untreated wastewater, <i>J. Carlos García-Díaz</i>	89
Clasificación de series temporales mediante profundidad para datos funcionales, <i>Andrés M. Alonso¹, David Casado², Sara López-Pintado³, Juan Romo⁴</i>	91
Cluster-MDS con restricciones de contigüidad espacial, <i>Rodrigo Macías, José Fernando Vera, José Miguel Angulo, Miguel Angel Montero</i>	93
Cómo surge la Estadística a partir de un conjunto de datos, <i>A. Lara Porras, Y. Román Montoya, J. García Leal</i>	95
Comparación de modelos para el espectro poblacional, <i>P. Saavedra, C.N. Hernández, J. Artiles, A. Santana, I. Luengo</i>	97
Competencias y estrategias en la inserción laboral de los titulados universitarios, <i>R.M. Crujeiras Casais¹, P. Faraldo Roca¹, B.M. Fernández de Castro¹, C.L. Iglesias Patiño¹, M.J. Lombardía Cortiña¹, J.M. Matías Fernández², S. Naya Fernández³</i>	99
Complejidad Computacional de Clasificadores Estadísticos y Neuronales, <i>Andrés Jiménez Jiménez¹, José María Gutiérrez Pérez², Angel Berihuete Macias²</i>	101
Comportamiento variado del consumidor: un análisis bayesiano dinámico del escenario de compra, <i>Carmen Berne¹, Pilar Gargallo², Noemí Martínez³, Manuel Salvador⁴</i>	103
Construcción de un observador en un modelo depredador-presa, <i>I. López¹, M. Gámez², J. Garay³, Z. Varga⁴</i>	105
Contenidos de Probabilidad en el entorno CDPYE, <i>M. J. García-Ligero Ramírez, A. Hermoso Carazo, P. López García, J. A. Maldonado Jurado, P. Román Román, F. Torres Ruiz</i>	107
Contraste robusto de cointegración frente a no linealidades, <i>Alvaro Escrivano¹, Ana E. García² y M^a Teresa Santos³</i>	109
Indicadores asociados a porcentajes de incidencias con series cortas en periodo base, <i>Juan José Barreiro¹, Miquel Salicrú², Carlos González³</i>	111

Covariance Changes Detection in Multivariate Time Series, <i>Pedro Galeano¹, Daniel Peña²</i>	113
D- y c-optimalidad para un modelo binomial negativo de regresión, <i>Carmelo Rodríguez Torreblanca¹, Juan M. Rodríguez Díaz²</i>	115
Design and optimization of EWMA control charts with asymmetric control limits for regions of maximum and minimum ARL, <i>J. Carlos García-Díaz, Francisco Aparisi</i>	117
Determinación de Estados Estacionarios e Identificación de Parámetros de Procesos Multivariantes: Aplicación a un Proceso de Tratamiento Anaeróbico de Aguas, <i>Maria Castellano Méndez¹, Gonzalo Ruiz², Wenceslao González Manteiga¹, Enrique Roca Bordello², Juan Lema²</i>	119
Determinación de Niveles de Ozono y NOx en Galicia, <i>Mª Castellano Méndez¹, Amaya Franco², David Cartelle³, Manuel Febrero¹, Enrique Roca²</i>	121
Determinación del tamaño muestral mediante técnicas bayesianas, <i>Lourdes Rodero¹, Josep Ginebra¹, Xavi Puig¹</i>	123
Determinación óptima en reacciones enzimáticas con inhibición por exceso de sustrato, <i>M.J. Rivas-López¹, M.T. Santos¹, I. Martínez²</i>	125
Diseño de planes de aceptación óptimos para la distribución exponencial biparamétrica, <i>Arturo J. Fernández¹, Carlos J. Pérez-González²</i>	127
Diseño óptimo de experimentos con variables explicativas censuradas, <i>Sandra Garcet Rodríguez¹, Jesús López Fidalgo²</i>	129
Diseños experimentales para modelos de crecimiento del tamaño de una población, <i>J. López Fidalgo¹, I.M. Ortiz Rodríguez², Y. del Águila del Águila³</i>	131
Diseños factoriales D-óptimos a dos niveles para modelos de Poisson, <i>Roberio Dorta-Guerra¹, Enrique González-Dávila¹, Josep Ginebra²</i>	133
Diseños óptimos para discriminar entre observaciones log-normales, <i>Jesus López-Fidalgo¹, Chiara Tommasi², Paula Camelia Trandafir³</i>	135

Distribución geogràfica y temporal de la mortalidad por cáncer de estómago en Catalunya, 1986-2000, <i>Xavi Puig¹, Rosa Gispert², Lourdes Rodero¹, Josep Ginebra¹</i>	137
Distribuciones ε -Contaminadas en Tablas de Contingencia, <i>Miguel A. Gómez Villegas¹, Beatriz González Pérez²</i>	139
Economic design of EWMA control chart using regions of maximum and minimum ARL, <i>J. Carlos García-Díaz, Francisco Aparisi</i>	141
El determinante de la matriz de información de Fisher y su aditividad, <i>Enrique González-Dávila¹, Roberto Dorta-Guerra¹, Josep Ginebra²</i>	143
El modelo de regresión parcialmente lineal semi-funcional para series de tiempo, <i>Germán Aneiros Pérez¹, Philippe Vieu²</i>	145
El papel del Boxplot based discretization en la Interpretación conceptual de una clasificación jerárquica, <i>Alejandra Pérez-Bonilla¹, Karina Gibert²</i>	147
El tratamiento de la variabilidad en datos procedentes de escalas de diferenciación semántica, <i>Lluís Marco Almagro¹, Xavier Tort-Martorell Llabrés²</i>	149
Ensayos en motores diesel mediante técnicas de diseño de experimentos para desarrollar nuevos conceptos de combustión, <i>Jose Miguel Carot¹, José Jabaloyes¹, Ricardo Novella², Jesús Benajes², Mónica Martínez¹, Andrea Conchado¹</i>	151
Entorno virtual de soporte a la docencia en Estadística, <i>J. Daunis-i-Estadella¹, J. Poch García², S. Thió-Henestrosa³, F. Prados Carrasco⁴</i>	153
Estadísticos basados en divergencias para el diagnóstico de modelos, <i>M.D. Esteban¹, T. Hobza², Y. Marhuenda¹, D. Morales¹</i>	155
Estimación de la dependencia condicional de dos tests diagnósticos binarios en presencia del sesgo de verificación, <i>J. A. Roldán Nofuentes¹, J. D. Luna del Castillo²</i>	157
Estimación de la exactitud corregida por azar de un test diagnóstico binario en presencia de verificación parcial, <i>J. A. Roldán Nofuentes¹, J. D. Luna</i>	

<i>del Castillo</i> ²	159
Estimación de parámetros en modelos de mantenimiento imperfecto, <i>R. Mullor</i> ¹ , <i>A. Sánchez</i> ² , <i>N. Martínez</i> ³	161
Estimación de Theil para el modelo logit de regresión logística, <i>José Trujillo Carmona</i>	163
Estimación Máximo verosímil en la Distribución Lognormal triparamétrica mediante simulación, <i>Juan Antonio Muñoz Orellana</i> ¹ , <i>José Fernando Vera Vera</i> ² , <i>Miguel Ángel Montero Alonso</i> ³	165
Estimación No Paramétrica de la Varianza condicional en un modelo de regresión con respuesta faltante, <i>A. Pérez González</i> ¹ , <i>J.M. Vilar Fernández</i> ² , <i>W. González Manteiga</i> ³	167
Estimación simultánea de parámetros en regresión lineal con heterocedascidad, <i>Beatriz Lacruz</i> ¹ , <i>Ana Pérez-Palomares</i> ²	169
Estimación y contrastes en modelos de retención de partículas radiactivas, <i>M. Amo Salas</i> , <i>J.M. Rodríguez Díaz</i> ¹ , <i>G. Sánchez León</i> ²	171
Estimaciones para un conjunto de señales: Un estudio de simulación, <i>I. Luengo</i> ¹ , <i>C.N. Hernández</i> ² , <i>P. Saavedra</i> ²	173
Estimadores de Dependencia Fuerte con Anisotropía Espacio-Temporal, <i>María P. Frías</i> ¹ , <i>María D. Ruiz-Medina</i> ² , <i>José M. Angulo</i> ² , <i>Francisco J. Alonso</i> ²	175
Estimadores robustos en la fase de ajuste de un gráfico de control, <i>José Luis Alfaro Navarro</i> ¹ , <i>Juan Fco. Ortega Dato</i> ²	177
Estimadores <i>shrinkage</i> basados en la descomposición atómica, <i>María D. Ruiz-Medina</i> ¹ , <i>José M. Angulo</i> ¹	179
Estudio de la varianza poblacional en el muestreo en ocasiones sucesivas, <i>Inmaculada Oña Casado</i> ¹ , <i>Amelia V. García Luengo</i> ¹ , <i>Eva M. Artés Rodríguez</i> ¹	181
Estudio de supervivencia de pacientes con SIDA, <i>E. Corpas Nogales</i> ¹ , <i>J.</i>	

<i>García Leal², A. Lara Porras², J. M. Quesada Rubio²</i>	183
<i>Estudio del Comportamiento Extremal de Sucesiones Periódicas, Manuel G. Scotto¹, Andreia Hall²</i>	185
<i>Evaluación de selectores del parámetro de suavizado mediante SiZer Map, M.D. Martínez Miranda¹, R. Raya Miranda¹, W. González Manteiga², A. González Carmona¹</i>	187
<i>Exactitud de las Soluciones de MDS mediante técnicas de Remuestreo, Miguel Ángel Montero Alonso¹, José Fernando Vera Vera, Juan Antonio Muñoz Orellana, Rodrigo Macías Páez, Andrés González Carmona</i>	189
<i>Explotación por horario: un modelo de gestión para Transports de Barcelona, Miquel Salicrú¹, Josep Maria Armengol²</i>	191
<i>Extensión del espacio paramétrico de la distribución de Sichel, Xavi Puig¹, Marta Pérez-Casany², Josep Ginebra¹, Alex Riba¹</i>	193
<i>Extremes of Periodic Integer-valued Moving Average Sequences, Manuel G. Scotto¹, Andreia Hall², João Pedro Cruz³</i>	195
<i>Factores determinantes en la supervivencia de la banca rusa al inicio de la transición, J. Orbe¹, V. Núñez-Antón¹</i>	197
<i>Fiabilidad de una medición en presencia de varios factores, Juan L. González Caballero¹, Juan L. Peralta Sáez², José Almenara Barrios³, Carolina Lagares Franco⁴</i>	199
<i>Filtrado e inter/extrapolación funcional de procesos espacio-temporales fractales, Rosaura Fernández-Pascual¹, María D. Ruiz-Medina², José M. Angulo²</i>	201
<i>Fitting grouped Cox model using phi-divergences, M^a del Carmen Pardo</i>	203
<i>Funcionamiento Diferencial de los Ítems y del Test (DFIT): Su aplicación en el QUALEFFO, M. Sánchez Barba¹, P. Vicente Galindo²</i>	205
<i>Guía interactiva de autoaprendizaje de SPSS, J. García Leal, E. Hidalgo Ruiz, A. Lara Porras, J. M. Quesada Rubio, E. Ramos Ábalos, R. Raya</i>	

Miranda, J. E. Ruiz Castro, I. R. Sánchez Borrego, J. Tarifa Blanco 207

- Identificación y caracterización de tipologías docentes. Aplicación a la UPV,
Mónica Martínez¹, Jose Miguel Carot², José Jabaloyes Vivas³, Pau Mi-
ró i Martínez⁴ 209

- Indicador local de riqueza/pobreza: caracterización y relación con el indica-
 dor de renta municipal en Galicia, *Patricia Sánchez Fernández¹, Esther*
López Vizcaíno², Antonio Vaamonde List³ 211

- Influencia Local: estimadores de razón y de Horvitz-Thompson, *I. Barranco*
Chamorro¹, A. Enguix González², J.L. Moreno Rebollo^{3,4}, J.M. Muñoz
Pichardo⁵ 213

- Influencia macroeconómica en el mercado de la vivienda, *Cristina Oyáguez,*
Patricia Rodríguez, Andrés Toledo 215

- Intervalos de predicción en series temporales utilizando el método de Holt-
 Winters, *José D. Bermúdez¹, José V. Segura², Enriqueta Vercher³* 217

- Introducción al análisis de la covarianza con intervalos aleatorios, *M. Montenegro¹,*
G. González-Rodríguez¹, A. Colubi¹, R. Coppi² 219

- La envolvente convexa esperada, *I. Cascos Fernández* 221

- La Inferencia Estadística en la PAU, *M. Candelaria Espinel¹, Carmen*
Ramos², Rosa Ramos³ 223

- La *k*-agrupabilidad: Un nuevo método para distinguir conjuntos agrupados
 y dispersos, *Agustí Solanas* 225

- Las regiones europeas: una perspectiva demográfica, *A. Arroyo¹, S. Bermúdez¹,*
R. Blanquero¹, A. Pozo¹ 227

- Mapa euclídeo de la Red Internet basado en medidas de retardo, *Francisco*
Javier Delgado¹, Purificación Galindo², José Luis Vicente-Villardón³ 229

- Matriz de Dispersión Inducida, *Andrés Jiménez Jiménez¹, José María Gu-*
tiérrez Pérez², Francisco Alvarez Gonzalez³ 231

Maximal correlation between minima, <i>A. Castaño Martínez</i> ¹ , <i>F. López Blázquez</i> ² , <i>B. Salamanca Miño</i> ²	233
ML solutions for a finite mixture discrete model, <i>Javier Portela</i> ¹	235
Modelización de datos de conteo multivariantes, <i>J.M. Muñoz-Pichardo</i> ¹ , <i>J. García-Heras</i> ² , <i>M.D. Cubiles-Vega</i> ³ , <i>R. Pino-Mejías</i> ⁴	237
Modelización temporal de los contaminantes registrados por la red automática de monitoreo en la ciudad de Valencia, <i>A. Montesinos Guillot</i> ¹ , <i>C. Capilla Romá</i> ² , <i>S. Balasch Parisi</i> ³	239
Modelización y optimización multiobjetivo de las condiciones de trabajo de un motor diesel, <i>Andrea Conchado</i> ¹ , <i>José Jabaloyes</i> ¹ , <i>Jose Miguel Carot</i> ¹ , <i>Leonor Hernandez</i> ² , <i>Jose María Desantes</i> ² , <i>Mónica Martínez</i> ¹	241
Modelo basado en factores inobservados para la predicción de precios de la energía eléctrica a medio y largo plazo, <i>Andrés M. Alonso</i> ¹ , <i>Carolina García-Martos</i> ² , <i>Julio Rodríguez</i> ³ , <i>María Jesús Sánchez</i> ⁴	243
Modelo bayesiano espacio-temporal para detectar actividad cerebral en imágenes de Resonancia Magnética funcional, <i>Raquel Montes</i> ¹ , <i>Alicia Quirós</i> ²	245
Modelo de Zwinderman para evaluación del cambio en Calidad de Vida Relacionada con la Salud, <i>Sara Morgado Nunes</i> ¹ , <i>M. Purificación Galindo</i> ²	247
Modelo para la predicción de accidentes con participación de vehículos industriales en carreteras interurbanas, <i>B. Arenas</i> ¹ , <i>E. Bernardos</i> ² , <i>F. Aparicio</i> ³ , <i>C. González</i> ⁴ , <i>M. J. Sánchez</i> ⁵ , <i>F. J. Páez</i> ⁶ , <i>A. Gómez</i> ⁷	249
Modelos mixtos para la predicción de precios de la energía eléctrica a corto plazo, <i>Carolina García-Martos</i> ¹ , <i>Julio Rodríguez</i> ² , <i>María Jesús Sánchez</i> ³	251
Monitorización on-line de procesos por lotes con el método de regresión sobre los scores truncados, <i>Francisco Arteaga</i> ¹ , <i>Alberto Ferrer</i> ²	253
Non-linear association on R^2 along a curve, <i>Marcelo Smrekar</i> ¹ , <i>Pedro Delicado</i> ²	

Nonparametric forecasting in time series. A comparative study, <i>Juan M. Vilar Fernández¹, Ricardo Cao²</i>	257
On a variant of integral trimmed regions, <i>Pedro Terán</i>	259
On ANOVA tests for a weighted sample of density functions, <i>Pedro Delicado Useros¹</i>	261
Planes de muestreo de aceptación aproximados con censura progresiva, <i>Carlos J. Pérez-González¹, Arturo J. Fernández²</i>	263
Presmoothed estimation with left truncated and right censored data, <i>M^a del Carmen Iglesias Pérez¹, M^a Amalia Jácome Pumar²</i>	265
Programa de simulación de superficies de respuesta, <i>Josep Maria Mateo Sanz¹</i>	267
Proyecciones de la población española, <i>Andrés M. Alonso¹, Daniel Peña², Julio Rodríguez³</i>	269
Quantile Curves and Dependence Structure for Bivariate Distributions, <i>F. Belzunce¹, A. Castaño-Martínez², A. Olvera-Cervantes², A. Suárez-Llorens²</i>	271
Rational bounds for the median of the $\Gamma(n+1,1)$ distribution, <i>José A. Adell, P. Jodrá</i>	273
Redes neuronales para la predicción del fracaso empresarial, <i>Toni Moreno-García</i>	275
Reglas de clasificación para poblaciones normales ordenadas, <i>Cristina Rueda¹, Miguel Fernández, Bonifacio Salvador</i>	277
Regresión basada en distancias en presencia de heteroscedasticidad, <i>Eva Boj¹, Josep Fortiana², M. Mercè Claramunt¹</i>	279
Restricted Minimum Phi-Divergence Estimators in Logistic Regression Mod-	

els, <i>Julio Angel Pardo</i>	281
Robustez bayesiana de la tasa de falsos positivos (positive false discovery rate), <i>Julián de la Horra</i> ¹	283
Selección bayesiana de modelos lineales espacio-temporales autorregresivos, <i>M^a Asunción Beamonte</i> ¹ , <i>Pilar Gargallo</i> ² , <i>Manuel Salvador</i> ³	285
Selección de transformaciones en la predicción de series temporales con el modelo de Holt, <i>Jose D. Bermúdez</i> ¹ , <i>Ana Corberán</i> ² , <i>Enriqueta Vercher</i> ³	287
Selection of Variables for Multivariate Control Charts, <i>Isabel González</i> ¹ , <i>Ismael Sánchez</i> ²	289
Sensibilidad en Redes Bayesianas Gausianas mediante una Medida de Divergencia, <i>Miguel A. Gómez-Villegas</i> ¹ , <i>Paloma Maín</i> ² , <i>Rosario Susi</i> ³	291
Sistema integrado de control estadístico y regulación automática de un proceso continuo de polimerización. Análisis de robustez del controlador predictivo DMC., <i>S. Barcelo</i> ¹ , <i>J. Sanchis</i> ² , <i>A. Ferrer</i> ³	293
Sobre los estimadores intrínsecos, <i>G. García</i> ¹ , <i>J.M. Oller</i> ²	295
Software JavaScript para comparación de r curvas de supervivencia, <i>Emilio Letón</i> ¹ , <i>Pilar Zuluaga</i> ²	297
Some Recent Contributions to Spatio-Temporal Modelling, <i>Emilio Porcu</i> ¹	299
Stability under products of sufficient and minimal sufficient σ -fields, <i>J.E. Chacón, J. Montanero, A.G. Nogales, P. Pérez</i>	301
Statistical Inference in Poisson Loglinear Models: An approach using Phi-divergences, <i>Nirian Martín</i> ¹ , <i>Leandro Pardo</i> ²	303
STATMEDIA II: Sistema de elaboración y gestión de prácticas multimedia individualizadas de estadística, <i>M. Calvo, A. Villarroya, A. Miñarro, A. Arcas, S. Vives, E. Vegas, F. Reverter,</i>	305

Tablas de mortalidad dinámicas. Modelos edad-periodo-cohorte., <i>A. Debón¹, F. Montes², F. Martínez-Ruiz³</i>	307
TAID: Algoritmo de segmentación basado en árboles ternarios, <i>P. Galindo¹, C. R. Castro²</i>	309
Técnicas paramétricas de clasificación automática para datos composicionales: resultados preliminares, <i>J.A. Martín Fernández¹, J. Daunis i Estadella², G. Mateu Figueiras³</i>	311
Test de bondad de ajuste multivariante basado en profundidad, <i>Ángel López¹, Juan Romo²</i>	313
Time-varying joint distribution through copulas, <i>María Concepción Ausín¹, Hedibert Freitas Lopes²</i>	315
TIRPE, un tutor informatizado para la resolución de problemas de Bioestadística, <i>Pedro Femia Marzo, Juan de Dios Luna del Castillo</i>	317
Un enfoque bootstrap para la estimación no paramétrica de una función de regresión a partir de tiempos de recurrencia, <i>José A. Cristóbal¹, Pilar Olave², Tomás Alcalá³</i>	319
Un Estudio Comparativo de Técnicas de Clasificación para la Diagnosis del Alzheimer Mediante SPECT y CAMCOG, <i>J.M. Fernández-Ponce, M.T. Gómez-Gómez, J.L. Pino-Mejías, A. Ruiz-Gándara</i>	321
Un experimento computacional para la simulación estocástica de la dinámica de burbujas, <i>Javier Jiménez-Fernández¹, Julio Rodríguez², José Mira³, Camino González⁴</i>	323
Un modelo híbrido para la detección precoz del abandono de clientes, <i>Susana San Matías¹, Ángeles Caldúch², Vicent Giner³, Miguel Ángel Edo⁴</i>	325
Un nuevo estadístico para detectar el desequilibrio de ligamiento global entre loci multialélicos, <i>A. Cancela¹, C. Carollo², C. Zapata³</i>	327
Un nuevo procedimiento de tratamiento de la autocorrelación en control estadístico de la calidad, <i>José Luis Alfaro Navarro¹, José Mondéjar</i>	

<i>Jiménez, Manuel Vargas Vargas²</i>	329
Un refinamiento del algoritmo de k-medias mediante profundidad y boots-trap, <i>Aurora Torrente¹ y Juan Romo²</i>	331
Una alternativa al precontrol clásico basada en redes neuronales artificiales, <i>Vicent Giner¹, Susana San Matías², Ángeles Caldúch³</i>	333
Una aplicación de la regresión logística a la recuperación de imágenes, <i>G. Ayala¹, E. De Ves², J. Domingo², T. León¹, P. Zucarello²</i>	335
Una aproximación logística para modelizar la fecundidad, <i>Pedro César Alvarez Esteban, Cristina Rueda Sabater</i>	337
Una comparativa entre el wild bootstrap y el bootstrap de los residuos suavizados, <i>Tomás R. Cotos-Yáñez¹</i>	339
Una mejora del gráfico de control <i>U</i> mediante la aplicación de la metodología de Daudin, <i>Elena Pérez Bernabeu¹, José Jabaloyes Vivas², Andrés Carrión García³</i>	341
Una metodología para la aplicación del análisis conjunto al marketing bancario, <i>Ángeles Caldúch¹, Vicent Giner², Susana San Matías³</i>	343
Una métrica entre distribuciones basada en una representación ‘fuzzy’, <i>Gil González-Rodríguez, Ana Colubi, Manuel Montenegro, María Ángeles Gil</i>	345
Una propuesta para Análisis Cluster Robusto con grupos heterogéneos, <i>L. A. García-Escudero¹, A. Gordaliza²</i>	347
Una variante del análisis multicanónico para tablas múltiples heterogéneas con grupos de variables cuantitativas y grupos de datos textuales. Aplicación en marketing, <i>Mónica Bécue-Bertaut¹, Jérôme Pagès², Ramón Álvarez Esteban³, José Luis Vázquez Burguete⁴</i>	349
Valores composicionales por debajo del límite de detección: un reemplazamiento paramétrico, <i>Javier Palarea Albaladejo¹, Josep Antoni Martín-Fernández²</i>	351

Variables auxiliares óptimas para la estimación de cuantiles, <i>M.D. Martínez Miranda¹, M. Rueda¹, A. Arcos¹</i>	353
Escala y evaluación de modelo Poisson-GPD , <i>M.I. Ortego¹, J.J. Egozcue²</i>	355
¿Cómo obtener razones de fallo con forma de bañera?, <i>Jorge Navarro Camacho¹, Pedro J. Hernández Martinez², Eshetu T. Wondmagegnehu³</i>	357
Sección Estadística Pública	359
Comparación de soluciones a la multicolinealidad en regresión logística funcional, <i>M. Escabias¹, A.M. Aguilera², M.J. Valderrama³</i>	361
Datos faltantes en estudios longitudinales: un nuevo enfoque de análisis, <i>Wilmer Fermín¹, Purificación Galindo², Javier Martín²</i>	363
DELFOS: explotación de datos estadísticos vía web, <i>Sergio Alonso¹, Carlos Duque-Gómez², Manuel García-Román³, Carlos Pérez-González¹</i>	365
Estimación de errores cuadráticos medios de estimadores de áreas pequeñas en la encuesta de población activa de Canarias, <i>Jorge Saralegui¹, Montserrat Herrador¹, Domingo Morales², Mª Dolores Esteban², Ángel Sánchez², Laureano Santamaría², Yolanda Marhuenda², Agustín Pérez²</i>	367
Estimación en áreas pequeñas en la Encuesta Industrial de la C.A. de Euskadi, <i>Josu Iradi¹, Haritz Olaeta², Cristina Prado³</i>	369
Estimadores en Áreas Pequeñas aplicados a la Estadística Pública Canaria, <i>Alberto González-Yanes¹, Miguel A. González-Sierra², Enrique González-Dávila², Roberto Dorts-Guerra², Sergio Alonso², Domingo Morales³</i>	371
Estimation of Unemployment Rates in Small Areas by a Multinomial Logit Mixed Model, <i>Isabel Molina¹, Ayoub Saei², María J. Lombardía³</i>	373
Evaluación empírica de los estimadores de áreas pequeñas en la estimación del desempleo, <i>Antonio Vaamonde Liste¹, Esther López Vizcaíno²</i>	375
Fay-Herriot Models with Nonparametric Estimation of Heteroscedasticity,	

M.J. Lombardía¹, W. González-Manteiga¹, I. Molina², D. Morales³, L. Santamaría³ 377

La Encuesta Continua de Presupuestos Familiares 2006 (ECPF 2006), *Carmen Ureña Ureña¹, Encarnación Martín Bernia²* 379

Posprocesado sobre microagregación multivariante: un enfoque con grupos de cardinalidad variable, *Josep Maria Mateo Sanz¹, Agusti Solanas², Josep Domingo Ferrer²* 381

Solución de sistemas de las ecuaciones simultáneas por métodos de altas prestaciones, *José Juan López Espín¹, Domingo Jiménez Cánovas²* 383

TEIDE: Una nueva herramienta para depurar encuestas, *Sergio Delgado Quintero¹, Juan José Salazar González², Fayna Álamo Santana³* 385

Sección Investigación Operativa 387

A conic IPM decomposition approach for large scale SDP, *Kartik K. Sivaramakrishnan¹, Gema Plaza², Tamás Terlaky³* 389

A family of cooperative games obtained by delegation, *Rafael Amer¹, José Miguel Giménez²* 391

A GRASP algorithm for VRPTW, *F. Cano Sevilla¹, C. Simón de Blas², S. Cano Alsúa³, A. Felipe Ortega⁴* 393

A Greedy approach to solve the Multidimensional Assignment Problem, *F. Perea¹, H. W. de Waard²* 395

A value for games with asymmetric cooperation among the players, *Rafael Amer¹, José Miguel Giménez², Antonio Magaña³* 397

Agregación de preferencias y DEA: Un análisis de los métodos propuestos para discriminar candidatos eficientes, *Bonifacio Llamazares¹, Teresa Peña²* 399

Algoritmo Genético (GA) para una línea de producción con posibilidad de resecuenciar considerando almacenes limitados, *Gerrit Färber, Anna M.*

<i>Coves Moreno</i>	401
Algoritmos heurísticos para secuenciación de proyectos con costes sobre los recursos, <i>R. Alvarez-Valdés¹, G. Martín², J. M. Tamarit¹</i>	403
Algoritmos para la construcción del conjunto de árboles generadores bajo un sistema de preferencia: condiciones y diseño, <i>Miguel Ángel Domínguez-Ríos, Sergio Alonso, Marcos Colebrook, Antonio Sedeño-Noda</i>	405
An evolutionary algorithm for the resource-constrained project scheduling problem subject to temporal constraints, <i>Vicente Valls¹, Francisco Ballestín², Agustín Barrios³</i>	407
Análisis del efecto de las incertidumbres en la optimización de los intervalos de pruebas y mantenimiento bajo criterios de coste y riesgo, <i>Sofía Carlos¹, Ana Sánchez², Sebastián Martorell³</i>	409
Aplicación de ecuaciones estructurales para la valoración de sistemas de armas y materiales en un simulador de combate, <i>M. Molero, A. Planells, O. Muñoz, J. Padura, A. Castellote, F. Bargueño</i>	411
Aplicación de Inteligencia Artificial en un Simulador Táctico de Combate para Adiestramiento (STCA), <i>F. Bargueño¹, A. Castellote¹, M. Molero¹, O. Muñoz¹, J. Padura¹, A. Planells¹</i>	413
Aplicaciones de la Teoría de Juegos a sistemas de comunicaciones multislot, <i>Javier Gozámez¹, Natividad Llorca², M^a Carmen Lucas³, Manuel A. Pulido⁴, Joaquín Sánchez-Soriano⁵</i>	415
Asignación de Turnos de Riego. Parte II: Modelo cuadrático separable con variables 0-1 para la asignación de turnos de riego, <i>M. Almiñana, L.F. Escudero, M. Landete, J.F. Monge, A. Rabasa, J. Sánchez-Soriano</i>	417
Asignación de Turnos de Riego. Parte I: Descripción de la aplicación y Generación de Prioridades, <i>M. Almiñana, L.F. Escudero, M. Landete, J.F. Monge, A. Rabasa, J. Sánchez Soriano</i>	419
Assignment situations with multiple ownership and their games, <i>Silvia Miquel¹, Bas van Velzen, Herbert Hamers, Henk Norde</i>	421

Búsqueda de óptimos globales en problemas de localización discreta mediante un algoritmo heurístico, <i>P. Fernández¹, B. Pelegrín¹, J.L. Redondo², I. García², P.M. Ortigosa²</i>	423
Cálculo de valores extremos en teoría de colas, <i>J.R. Artalejo</i>	425
Coalition formation and partnerships in the Catalonia Parliament, <i>Francesc Carreras¹, Maria Dolors Llongueras², Maria Albina Puente³</i>	427
Coloreando grafos con redes neuronales, <i>Javier Yáñez¹, Pedro Martínez Talaván²</i>	429
Competencia espacial duopolística con demanda elástica y política de precios en destino, <i>P. Dorta González¹, D.R. Santos Peñate¹, B. Pelegrín Pelegrín², R. Suárez Vega¹</i>	431
Condiciones de optimalidad de orden superior para mínimos estrictos, <i>César Gutiérrez¹, Bienvenido Jiménez², Vicente Novo³</i>	433
Construcción de paneles de datos con información de fuentes administrativas a partir de muestreo. aplicación al panel de datos del impuesto sobre la renta de las personas físicas del instituto de estudios fiscales, <i>César Pérez López¹, Santiago Díaz de Sarralde²</i>	435
Coordination in discrete review inventory situations, <i>Luis A. Guardiola</i>	437
Cualificación de restricciones en programación multiobjetivo, <i>Luis Coladas Uría¹</i>	439
Definición y uso de un problema Core para el problema de transporte con costos fijos, <i>Jesús Sáez Aguado</i>	441
Determinación del precio y del tamaño del lote para una curva demanda-precio tipo racional, <i>L.A. San-José¹, J. Sicilia², J. García-Laguna³</i>	443
Diseño de rutas de recogida de residuos sólidos urbanos en el área metropolitana de Barcelona, <i>Joaquín Bautista¹, Jordi Pereira², Elena Fernández³</i>	445
Distribución del número de reintentos efectuados por un cliente, <i>M.J. López-</i>	

Divergencia no probabilística entre conjuntos borrosos, <i>Jorge Jiménez¹, Susana Montes²</i>	449
EAGH (Empirically Adjusted Greedy Algorithms) para la determinación de secuencias en una máquina multiproducto, <i>Albert Corominas¹, Rafael Pastor¹, Ana Sánchez¹</i>	451
Eficiencia de diferentes modelos de PLE para resolver SALBP, <i>Rafael Pastor¹, Laia Ferrer²</i>	453
El comportamiento emprendedor en la cultura empresarial: una aplicación del modelo de Rasch, <i>J. Febles Acosta¹, J.R. Oreja Rodríguez²</i>	455
El concepto de KT-invexidad en un problema de control, <i>M. Arana-Jiménez¹, R. Osuna-Gómez², A. Rufián-Lizana³, G. Ruiz-Garzón⁴</i>	457
El plan de extensión óptimo de la red ferroviaria de alta velocidad, <i>V. Blanco¹, J. Puerto²</i>	459
El problema de la mediana ordenada con normas l_p , <i>I. Espejo Miranda, A.M. Rodríguez-Chía, C. Valero Franco</i>	461
El problema de localización con distancias esperadas, <i>Concepción Valero Franco¹, Antonio M. Rodríguez Chía, Inmaculada Espejo Miranda</i>	463
Equilibrado de la línea de montaje del motor del modelo Pathfinder en la planta de NISSAN de Barcelona, <i>Joaquín Bautista Valhondo¹, Jordi Pereira Gude²</i>	465
Equilibrado de Líneas con Alternativas de Montaje, <i>Liliana Capacho Betancourt¹, Rafael Pastor Moreno²</i>	467
Equilibrado de líneas con setup. Un modelo matemático, <i>Rafael Pastor¹, Carlos Andrés², Cristóbal Miralles³</i>	469
Esqueletos Orientados a Objeto para la técnica de Ramificación y Acotación, <i>Isabel Dorta¹, Coromoto León¹, Carlos Segura¹</i>	471

Estimación de Índices de Eficiencia de Bibliotecas Universitarias mediante Análisis Envolvente de Datos, <i>J. García-Laguna¹, B. Salvador², J. A. Tapia³</i>	473
Estudio Experimental sobre el Comportamiento de Reglas de Despacho en un Taller de Máquinas Paralelas no Relacionadas con Llegada Dinámica de Trabajos y Setup Dependiente de la Secuencia, <i>Larissa Stania¹, Pedro Gómez¹, Carlos Andrés¹</i>	475
Evaluación y optimización de recursos sanitarios en el Hospital San Juan de Dios de Pamplona, Colombia, <i>Aurora Inés Gáfaro Rojas¹, Cristina Azcárate Camio², M^a Luisa Eraso Goicoechea³</i>	477
Exceso de Información en Optimización Lineal Paramétrica, <i>Miguel Ángel Goberna, Valentín Jornet, Mariola Molina¹</i>	479
Extended Classification for Flowshops with Resequencing, <i>Gerrit Färber, Anna M. Coves Moreno</i>	481
Familia de medidas de divergencia entre conjuntos borrosos, <i>Inés Couso¹, Susana Montes¹</i>	483
Fiabilidad de un sistema bajo reparaciones imperfectas gobernado por distribuciones tipo-fase, <i>D. Montoro-Cazorla¹, R. Pérez-Ocón², M. Carmen Segovia³</i>	485
Flexible Discrete Location Problems with Capacity Constraints, ¹ <i>J. Kalcsics, ²S. Nickel, ³J. Puerto, ⁴A.M. Rodríguez-Chía</i>	487
Funciones e-convexas: Propiedades, <i>V. Jornet¹, M. Rodríguez²</i>	489
Gestión de tablas de decisiones óptimas, <i>Juan A. Fernández del Pozo, Concha Bielza</i>	491
GRASP para el Diseño de Territorios de Ventas, <i>Elena Fernández¹, Roger Z. Ríos Mercado²</i>	493
Heurística para la resolución de un problema de líneas de montaje, <i>Alberto Gómez, David de la Fuente, José Parreño, Isabel Fernández</i>	495

Heurísticas constructivas y de mejora local para la secuenciación de modelos mixtos de vehículos en líneas de montaje, <i>Joaquín Bautista Valhondo¹, Jaime Cano Belmán²</i>	497
Heurísticas para determinar secuencias en una máquina multiproducto sujeta a fallos y con costes cuadráticos, <i>Albert Corominas¹, Rafael Pastor¹, Ana Sánchez¹</i>	499
Heurísticas para el Problema de Secuenciación de Flujo Regular de Permutación con Máquinas No Disponibles, <i>Paz Pérez González, José Manuel Framiñán Torres</i>	501
Heurísticas para talleres de flujo sin esperas y tiempos de cambio, <i>Rubén Ruiz¹, Ali Allahverdi²</i>	503
Heurísticos para la Secuenciación de Proyectos con Recursos Parcialmente Renovables, <i>R. Álvarez-Valdés¹, E. Crespo², J. M. Tamarit³, M. F. Villa⁴</i>	505
Información de segundo orden en optimización no lineal con cotas simples, <i>Javier M. Moguerza¹, Alberto Olivares¹</i>	507
Interval values for strategic games in which players cooperate, <i>Luisa Carpente¹, Balbina Casas-Méndez², Ignacio García-Jurado³, Anne van den Nouweland⁴</i>	509
KT-invexidad en optimización escalar no regular, <i>Beatriz Hernández Jiménez¹, Rafaela Osuna Gómez², Marko A. Rojas-Medar³, Antonio Beato Moreno⁴</i>	511
La evolución histórica de la “robustez”: entre la Decisión Multicriterio y los Métodos Bayesianos, <i>María del Carmen Escrivano Ródenas¹, Gabriela Mónica Fernández Barberis²</i>	513
Location and size selection for new competing centres on a network, <i>R. Suárez Vega, D.R. Santos Peñate, P. Dorta González</i>	515
Mal planteamiento primal-dual en optimización lineal semi-infinita: Una aplicación a los problemas cónicos, <i>M.J. Cánovas¹, M.A. López², J. Parra¹, F.J. Toledo¹</i>	517

Métodos de Optimización Indirecta en Aplicaciones Biotecnológicas, <i>A. Marín-Sanguino</i> ¹ , <i>C. González-Alcón</i> ² , <i>N.V. Torres Darias</i> ¹	519
Minimización del tiempo total de retraso en el taller de flujo de permutación, <i>Eva Vallada, Rubén Ruiz, Gerardo Minella</i>	521
Modelado de redes inalámbricas ad hoc: aspectos energéticos, <i>David Soler</i> ¹ , <i>José Albiach</i> ¹ , <i>Eulalia Martínez</i> ¹ , <i>Pietro Manzoni</i> ²	523
Modelización estadística para la predicción de fallos en los conductos de agua potable, <i>Andrés Carrión García</i> ¹ , <i>Ana Debón Aucejo</i> ² , <i>Liney A. Manjarres</i> ³	525
Modelo de PLEM para la asignación de tareas al personal y su aplicación a un caso real, <i>Ericka Zulema Rodríguez Calvo</i> ¹ , <i>Anna Maria Coves Moreno</i> ²	527
Modelos de localización líder-seguidor con demanda elástica, <i>Dolores R. Santos Peñate, Rafael Suárez Vega, Pablo Dorta González</i>	529
Modelos de subasta en un duopolio eléctrico con costes correlados, <i>Estrella Alonso Pérez</i> ¹ , <i>Juan Tejada Cazorla</i> ²	531
Modelos matemáticos para el problema de la ubicación de productos en estanterías, <i>M. Palmer</i> ¹ , <i>J. Albarracín</i> ¹ , <i>J. P. García</i> ¹ , <i>M. Babiloni</i> ¹ , <i>A. Esparcia</i> ²	533
MOM: Un Sistema Software para la optimización del tráfico ferroviario, <i>P. Tormos</i> ¹ , <i>A. Lova</i> ¹ , <i>F. Barber</i> ² , <i>L. Ingolotti</i> ² , <i>M. Abril</i> ² , <i>M.A. Salido</i> ²	535
Multidimensional Assignment Games, <i>F. Perea</i> ¹	537
New Results on the Public Good Index, <i>J. M. Alonso-Mejide</i> ¹ , <i>B. Casas-Méndez</i> ² , <i>M. J. Holler</i> ³ , <i>S. Lorenzo-Freire</i> ⁴	539
Non Advantageous Merging in cost spanning tree problems. A new characterization of the Bird rule., <i>María Gómez - Rúa</i> ¹ , <i>Juan J. Vidal - Puga</i> ²	541
Non-cooperative cost spanning tree games with budget restrictions, <i>Gustavo</i>	

<i>Bergantiños¹, Leticia Lorenzo²</i>	543
Nuevos criterios generalizados para modelar las preferencias del decisor en los métodos de relaciones de superación, <i>Gabriela Fernández Barberis, María del Carmen Escribano Ródenas</i>	545
Obtención de referentes más cercanos en la frontera eficiente DEA, <i>Inmaculada Sirvent¹, Juan Aparicio², José Luis Ruiz³</i>	547
Obtención de targets globales e individuales mediante un nuevo modelo de asignación de recursos centralizado bajo retorno de escala constante, <i>Jesús T. Pastor¹, Juan Aparicio²</i>	549
On the convenience of keeping components at a level greater than one half, <i>Francesc Carreras¹, Josep Freixas², María Albina Puente²</i>	551
On the election of optimal persuadable voters, <i>Josep Freixas, Montserrat Pons</i>	553
On the Shapley value of a minimum cost spanning tree problem, <i>Gustavo Bergantiños¹, Juan Vidal-Puga²</i>	555
On-line search algorithms for fault diagnostics in large-scale computer communication networks, <i>David Alcaide¹, Eugene Levner², Ilya Levin³, Sergey Frenkel⁴, Viktor Zakharov⁴</i>	557
Operadores de recombinación para Algoritmos Evolutivos Multiobjetivo basados en ranking Pareto, <i>P. Mateo Collazos¹, I. Alberto Moralejo²</i>	559
Optimal control of fed-batch processes with particle swarm optimization, <i>A. Ismael F. Vaz¹, Eugénio C. Ferreira²</i>	561
Optimistic weighted Shapley rules in minimum cost spanning tree problems, <i>Gustavo Bergantiños¹, Silvia Lorenzo-Freire²</i>	563
Optimización conjunta del mantenimiento y los recursos considerando criterios RAMS+C, <i>A. Sánchez¹, S. Martorell², J. Sanz³, G. Clemente⁴</i>	565
Optimización y estimación robusta de carteras financieras, <i>Francisco J.</i>	

<i>Nogales¹, Victor deMiguel²</i>	567
<i>Partnership formation and binomial semivalues, Francesc Carreras¹, Maria Dolors Llongueras², Maria Albina Puente³</i>	569
<i>Planificación de la producción en una acería. Análisis de un caso real, Fermín Mallor Giménez¹, Pedro M. Mateo Collazos²</i>	571
<i>Planificación de la producción ganadera mediante programación lineal, Sara Verónica Rodríguez Sánchez¹, Luis M. Plà Aragones²</i>	573
<i>Política óptima de un modelo EOQ con demanda parcialmente acumulable, L.A. San-José¹, J. Sicilia², J. García-Laguna³</i>	575
<i>Procedimiento constructivo por combinación de reglas para secuenciar modelos mixtos de vehículos en líneas de montaje, Joaquín Bautista Valhondo¹, Jaime Cano Belmán²</i>	577
<i>Procedimientos de Flujo en Redes para Problemas de Planificación Open-Shop Bicriterio, Antonio Sedeño-Noda, David Alcaide, Carlos González-Martín</i>	579
<i>Programación múltiple con funciones convexas generalizadas, A. Beato Moreno¹, M.B. Hernández Jiménez², Rafaela Osuna Gómez³</i>	581
<i>Programación por metas con criterios fraccionales no lineales, Carmen Castrodeza¹, Teresa Peña²</i>	583
<i>Propiedades de los sistemas completos borrosos, Susana Díaz¹, Pedro Gil¹, Susana Montes¹</i>	585
<i>Reglas hidráulicas para problemas de bancarrota con referencias, Mercedes Alemán Jiménez¹, Carlos González-Alcón²</i>	587
<i>Regularidad métrica del conjunto de soluciones óptimas en Programación Semi-Infinita, M. J. Cánovas¹, D. Klatte², M. A. López³, J. Parra¹,</i>	589
<i>Reparto de holguras en una red PERT, Javier Castro¹, Daniel Gómez², Juan Tejada³</i>	591

Resolución del Response Time Variability Problem mediante programación matemática, <i>Albert Corominas</i> ¹ , <i>Wieslaw Kubiak</i> ² , <i>Rafael Pastor</i> ¹	593
Resolución óptima del Multi-Min-Max Assembly Line Balancing Problem (MMM-ALBP) mediante programación matemática, <i>Rafael Pastor</i> ¹ , <i>Jordi Ojeda</i> ^{1,2} , <i>Carles Sentinella</i> ³ , <i>Jordi Solsona</i> ³	595
Retrospectiva y desafíos del problema de planificación de la producción, <i>J. Gutiérrez</i> ¹ , <i>A. Sedeño-Noda</i> ² , <i>M. Colebrook</i> ³ , <i>B. Abdul-Jalbar</i> ⁴ , <i>J. Sicilia</i> ⁵	597
Risk measures in MPSSP, <i>A. Alonso-Ayuso</i> ¹ , <i>L.F. Escudero</i> ² , <i>C. Pizarro</i> ³	599
Secuenciación mediante algoritmos evolutivos en complejos talleres flexibles, <i>Rubén Ruiz</i> ¹ , <i>Thijs Urlings</i> ¹ , <i>Funda Sivrikaya Serifoğlu</i> ²	601
SEDD, un sistema experto para el diagnóstico de desastres, <i>Antonio Omaña Prieto</i> ¹ , <i>Begoña Vitoriano Villanueva</i> ² , <i>M. Teresa Ortúñoz Sánchez</i> ³	603
Segmentación de imágenes digitales basada en problemas de coloracion para grafos valorados, <i>Daniel Gómez</i> ¹ , <i>Victoria López</i> ² , <i>Javier Montero</i> ² , <i>Javier Yáñez</i> ²	605
Selección de carteras utilizando funciones borrosas de riesgo lateral, <i>Jose D. Bermúdez</i> ¹ , <i>José V. Segura</i> ² , <i>Enrique Vercher</i> ³	607
Selección de variables y obtención de coeficientes de la función de discriminación lineal: Uso de Búsqueda Tabú, <i>Silvia Casado</i> ¹ , <i>Joaquín Pacheco</i> ¹ , <i>Laura Nuñez</i> ²	609
Sharing costs in a proportionate flow shop situatio., <i>M. A. Mosquera</i> ¹ , <i>P. Borm</i> ² , <i>H. Hamers</i> ² , <i>A. Estévez-Fernández</i> ²	611
Simulación de una subasta en el sector de la construcción, <i>Juan Momparler</i> ¹ , <i>Mario Hidalgo</i> ²	613
Sistema Geo ^a /Geo ^b /1, <i>María Cruz Valsero Blanco</i> , <i>Ana Lorente Marín</i>	615
Sistemas de ecuaciones e inecuaciones lineales: Estabilidad y distancia al mal planteamiento, <i>M. J. Cánovas</i> , <i>F.J. Gómez-Senent</i> , <i>J. Parra</i>	617

Sistemas de Inventario/Distribución con demanda potencial y rotura, <i>Beatriz Abdul-Jalbar¹, Joaquín Sicilia², Roberto Dorta³, José M. Gutiérrez⁴</i>	619
Sistemas Estratégicos de Rutas de Vehículos para Operaciones de Paz, <i>¹Eduardo Ramos Méndez, ²José A. Carrillo Ruiz</i>	621
Sobre una clase de problemas inversos en control óptimo estocástico, <i>Ricardo Josa Fombellida¹, Juan Pablo Rincón Zapatero²</i>	623
Solving a multiple instance problem via separating concentric hyperspheres, <i>José Gordillo¹, Emilio Carrizosa¹, Frank Plastria²</i>	625
Solving a School Bus Routing Problem, <i>Jorge Riera-Ledesma, Juan-José Salazar-González</i>	627
Solving the Ordered Median Spanning Tree Problem with a greedy algorithm, <i>Elena Fernández¹, Justo Puerto², Antonio M. Rodríguez Chía³</i>	629
Técnicas Bayesianas aplicadas a la fusión de datos en un Robot Modular, <i>A. Moreno Díaz¹, M.I. Rodríguez Galiano¹, M.A. Virtu García²</i>	631
Técnicas para la programación de tráfico ferroviario heterogéneo, <i>A. Lova¹, P. Tormos¹, F. Barber², L. Ingolotti², M. Abril², M.A. Salido²</i>	633
Teoría de Juegos y Cohesividad en Redes de Comunicación Probabilísticas, <i>Daniel Gómez¹, Enrique González Arangüena¹, Conrado Manuel¹, Guillermo Owen², Mónica del Pozo¹</i>	635
The Harsanyi paradox and the 'right to talk' in bargaining among coalitions, <i>Juan J. Vidal-Puga¹</i>	637
The use of relative priorities in optimizing the performance of a queueing system, <i>Justo Puerto¹, Francisco R. Fernández², Rafael Hassin³</i>	639
The virtually subgame perfect equilibrium for extensive games, <i>Ignacio García-Jurado¹, Julio González-Díaz²</i>	641
Ubicación de productos en estanterías de grandes superficies: una revisión	

sión bibliográfica, <i>J. M. Albarracín¹, M. Palmer¹, J. P. García¹, M. Babiloni¹, A. Esparcia²</i>	643
Un algoritmo basado en la entropía para la construcción de redes Bayesanas, <i>Ana María Martínez Rodríguez¹, Luis G. Vargas², Luis Parras Guijosa³</i>	645
Un algoritmo de Generación de Columnas para Máquinas de Vector de Apoyo, <i>Emilio Carrizosa¹, Belén Martín-Barragán², Dolores Romero Morales³</i>	647
Un algoritmo descendente local para generar la región eficiente de un problema de programación lineal multiobjetivo, <i>Jesús M. Jorge Santiso</i>	649
Un algoritmo genético híbrido basado en una codificación innovadora para el RCPSP, <i>Javier Alcaraz Soria¹, Concepción Maroto Álvarez²</i>	651
Un algoritmo heurístico para Clasificación Semisupervisada, <i>Rafael Blanquero¹, Emilio Carrizosa¹, Norberto Díaz²</i>	653
Un algoritmo para optimizar una función lineal sobre la región eficiente de un problema multiobjetivo lineal entero, <i>Jesús M. Jorge Santiso¹, Marianela Carrillo Fernández²</i>	655
Un estudio sobre la coherencia de preferencias lingüísticas, <i>Susana Díaz¹, José Luis García Lapresta², Susana Montes¹</i>	657
Un método simplex para problemas de caminos mínimos con pivotaje múltiple, <i>C. González Martín, A. Sedeño Noda, S. Alonso Rodríguez</i>	659
Un modelo de asignación en el mercado inmobiliario con intervención del estado, <i>Antonio Armada Estévez¹, Joaquín Sicilia Rodríguez²</i>	661
Un modelo de simulación para planificación de tráfico ferroviario, <i>Iosu Mirem Martínez¹, Begoña Vitoriano², Antonio Fernández-Cardador³, Paloma Cucala⁴</i>	663
Un modelo EOQ con demanda exponencialmente dependiente del precio, <i>L.A. San-José¹, J. Sicilia², J. García-Laguna³</i>	665

Un nuevo algoritmo GRASP para el problema del empaquetamiento bidimensional, <i>R. Álvarez-Valdés¹, F. Parreño², J. M. Tamarit³</i>	667
Un nuevo método simplex eficiente para resolver problemas de caminos mínimos, <i>A. Sedeño Noda, C. González Martín</i>	669
Un sistema multicomponentes reparable en tiempo discreto, <i>Juan Eloy Ruiz Castro¹, Rafael Pérez Ocón², Gemma Fernández-Calvillo Villodre</i>	671
Una aproximación de programación matemática para la depuración de datos categóricos, <i>Sergio Delgado Quintero¹, Juan José Salazar González²</i>	673
Una generalización del modelo con retropedidos parciales de Lee-Nahmias, <i>L.A. San-José¹, J. Sicilia², J. García-Laguna³</i>	675
Una nueva herramienta software para resolver problemas de localización multicriterio sobre redes, <i>Marcos Colebrook, Sergio Alonso, Joaquín Sicilia</i>	677
Unequal Probability Selection in multivariate finite population sampling. A Vector-Optimization approach, <i>Emilio Carrizosa¹</i>	679
Who gains and how much by cooperation in decentralized supply chains, <i>Ana Meca</i>	681

Sección Probabilidad 683

A closed-form expression for the quantile function of the $\Gamma(2, \lambda)$ distribution, <i>José A. Adell, P. Jodrá</i>	685
Berry-Esseen Bounds for Standardized Lévy Processes via Moduli of Smoothness, <i>José A. Adell¹, Alberto Lekuona²</i>	687
Campos de difusión biparamétricos no constantes en los ejes. El caso log-normal, <i>R. Gutiérrez¹, C. Roldán²</i>	689

Caracterización de distribuciones de vida mediante la ordenación estocástica de variables doblemente truncadas, <i>Miguel A. Sordo¹</i>	691
--	-----

Cercanía de la distribución de un vector aleatorio a la de un conjunto aleatorio, <i>I. Cascos Fernández¹, M. López-Díaz²</i>	693
Characterization of Distributions through Martingales, <i>R. Gouet¹, F.J. López², G. Sanz³</i>	695
Comparación Dispersiva de Distribuciones Wishart, <i>J.M. Fernández Ponce¹, M.T. Gómez Gómez¹, J.L. Pino Mejías¹, Rosario Rodríguez Griñolo¹</i>	697
Contrastes para órdenes convexos basados en L-estadísticos, <i>Miguel A. Sordo¹, Héctor M. Ramos², Jorge Ollero³</i>	699
Cuasi probabilidad en el anillo \mathbf{B}^* , <i>Joaquín Ferrer Sánchez¹, Juan Cuadra Carreño¹, Francisco Herrera Cuadra¹</i>	701
De la geometría composicional al espacio de Hilbert de densidades, <i>J.J. Egozcue¹, V. Pawlowsky-Glahn², J.L. Díaz-Barrero¹</i>	703
Densidades de cónulas considerando la estructura de su espacio soporte, <i>M.I. Ortego¹, G. Mateu-Figueras²</i>	705
Duality for Multi-State Probabilistic Cellular Automata, <i>F.J. López¹, G. Sanz², M. Sobottka³</i>	707
Estimación cuadrática a partir de observaciones con retraso aleatorio corregido, <i>R. Caballero¹, A. Hermoso², J. Jiménez¹, J. Linares², S. Nakamori³</i>	709
Estimación de señales mediante aproximación de mixturas, <i>R. Caballero¹, A. Hermoso², J. Jiménez¹, J. Linares², S. Nakamori³</i>	711
Estudio de variables temporales asociadas a procesos estocásticos, <i>R. Gutiérrez Jáimez, P. Román Román, D. Romero Molina, J.J. Serrano Pérez, F. Torres Ruiz</i>	713
Interpretación geométrica de λ para la familia de distribuciones gaussianas discretas, <i>Miguel Ángel Tíscar Soria¹, José Rodríguez Avi², Antonio J. Sáez Castillo³, Antonio Conde Sánchez⁴</i>	715
Observaciones imprecisas: variables aleatorias difusas, <i>Laura Garrido Blanco¹,</i>	

Teófilo Brezmes Brezmes², Pedro Gil Álvarez³ 717

Ordenación stop-loss multivariante de vectores de sumas aleatorias en ambientes dependientes, *Félix Belzunce², Eva-María Ortega¹, Franco Pellerey³, José M. Ruiz²* 719

Paralelización embarazosa distribuida sin modificación del código fuente, *C. Carleos¹, J. Á. Baro², N. Corral¹* 721

Probability Inequalities for the Ruin Probability in a Markov Controlled Risk Model, *Maikol Diasparra¹, Rosario Romera²* 723

Proceso estocástico lognormal triparamétrico: Simulación y aplicación, *R. Gutiérrez¹, R. Gutiérrez-Sánchez, A. Nafidi, E. Ramos* 725

Procesos Puntuales de Incrementos Condicionalmente Independientes, *Miguel Martín Díaz¹, Tomás Prieto Rumeau², Ricardo Vélez Ibarrola³* 727

Redundancia activa y en espera para sistemas con componentes dependientes, *Félix Belzunce, Helena Martínez, José María Ruiz* 729

Relaciones entre las esperanzas de Aumann y Herer de un conjunto aleatorio, *Pedro Terán¹* 731

Restauración de imágenes ruidosas usando información de covarianza, *M. J. García-Ligero¹, A. Hermoso¹, J. Linares¹, S. Nakamori²* 733

Una generalización de la distribución Beta-Binomial. Aplicaciones y propiedades, *Antonio Conde Sánchez¹, Mª José Olmo Jiménez², José Rodríguez Avil³, Antonio José Sáez Castillo⁴* 735

Una solución al problema de estimación no lineal basada en los desarrollos en serie aproximados, *R. M. Fernández Alcalá¹, J. Navarro Moreno², J. C. Ruiz Molina³* 737

Uso del Método de Rayleigh-Ritz para la Simulación de Campos Aleatorios, *A. Oya Lechuga¹, J. Navarro Moreno², J.C. Ruiz Molina³* 739

Índice de autores

740

1

Conferencias Plenarias

Computational strategies for deriving the best-subset regression models

Erricos J. Kontoghiorghes¹, Cristian Gatu², Marc Hofmann²,
Petko Yanev³

¹University of Cyprus and Birkbeck College University of London, UK

²University of Neuchâtel, Switzerland

³CNR, Pisa, Italy

Abstract

An efficient branch-and-bound algorithm for computing the best-subset regression models is proposed. The algorithm avoids the computation of the whole regression tree that generates all possible subset models. It is formally shown that if the branch-and-bound test holds, then the current subtree together with its right-hand side subtrees are cut. This reduces significantly the computational burden of the proposed algorithm when compared to an existing leaps-and-bounds method which generates two trees. Specifically, the proposed algorithm, which is based on orthogonal transformations, outperforms by $O(n^3)$ the leaps-and-bounds strategy. The criteria used in identifying the best subsets are based on monotone functions of the residual sum of squares (RSS) such as R^2 , adjusted R^2 , mean square error of prediction and C_p . Strategies and heuristics which improve the computational performance of the proposed algorithm are investigated. A computationally efficient heuristic version of the branch-and-bound strategy which decides to cut subtrees using a tolerance parameter is proposed. The heuristic algorithm derives models close to the best ones. However, it is shown analytically that the relative error of the RSS, and consequently the corresponding statistic, of the computed subsets is smaller than the value of the tolerance parameter which lies between zero and one. Extensions of the heuristic strategies are investigated. The tolerance is either a function of the submodel size or of the depth of the node in the regression tree. Several tolerance configurations are inspected and shown to perform very well. A weighted directed graph approach to compute the best subset model is proposed. Computational results and experiments on random and real data are presented and analyzed.

1. Bibliography

- [1] C. Gatu and E.J. Kontoghiorghes. *Parallel algorithms for computing all possible subset regression models using the QR decomposition*. Parallel Computing, **29**(4), 505-521, 2003.
- [2] C. Gatu and E.J. Kontoghiorghes. *Computing all possible subset models*

- derived from a VAR process.* Computational Management Science, **4**(4), 253-278 2005.
- [3] C. Gatu and E.J. Kontoghiorghes. *A branch and bound algorithm for computing the best subset regression models.* Journal of Computational and Graphical Statistics, **15**(1), 1-19, 2006.
 - [4] C. Gatu, and E.J. Kontoghiorghes. *Estimating all possible SUR models with permuted exogenous data matrices derived from a VAR process.* Journal of Economic Dynamics & Control, 2006 (in press).
 - [5] M. Hofmann, C. Gatu and E.J. Kontoghiorghes. *Algorithms for computing the best-subset regression models.* Computational Statistics & Data Analysis, 2006 (submitted).
 - [6] P. Foschi and E.J. Kontoghiorghes. *Estimation of VAR models: computational aspects.* Computational Economics, **21**(1-2): 3-22, 2003.
 - [7] E.J. Kontoghiorghes. *Computational methods for modifying seemingly unrelated regressions models.* Journal of Computational and Applied Mathematics, **62**(1): 247-261, 2004.
 - [8] P. Yanev, C. Gatu, and E.J. Kontoghiorghes. *Regression graphs and trees for regression subset selection* (under review).
 - [9] P. Yanev, P. Foschi and E.J. Kontoghiorghes. *Algorithms for computing the QR decomposition of a set of matrices with common columns.* Algorithmica, **39**: 83-93, 2004.

Mathematical Programming Problems with Linear Complementarity Constraints

Joaquim J. Júdice

Universidade de Coimbra and Instituto de Telecomunicações, Portugal

A Mathematical Programming Problem with Linear Complementarity Constraints (MPLCP) is a nonlinear program where a linear or a nonlinear continuously differentiable function is minimized over a nonconvex set defined by linear and complementarity constraints. The problem can be mathematically stated as

$$\begin{aligned} & \text{Minimize} && f(x, y) \\ & \text{subject to} && Ew = q + Mx + Ny \\ & && x \geq 0, w \geq 0, y \in K_y \\ & && x_i w_i = 0, i = 1, \dots, n \end{aligned}$$

where $E, M \in \mathbb{R}^{m \times n}$, $N \in \mathbb{R}^{m \times p}$, f is a continuously differentiable function on $\mathbb{R}^{n \times p}$, $w, x \in \mathbb{R}^n$, $y \in \mathbb{R}^p$ and K_y is the polyhedron

$$K_y = \{y \in \mathbb{R}^p : Ay = b, y \geq 0\}$$

This optimization problem has received much attention during the past several years, mainly because of their number of applications in different areas of science, engineering and economics. Furthermore there are many nonconvex NP-hard optimization problems that can be stated as MPLCPs. In this talk, some of the most important of these formulations are first discussed. These include Bilevel, Bilinear, Concave and Nonconvex Quadratic Programming Problems, Zero–One Integer Programs, the Linear Complementarity Problem and an Eigenvalue Complementarity Problem.

Algorithms for the solution of the MPLCP are then analysed. A Complementarity Active–Set algorithm is first discussed for computing a stationary point of the MPLCP. The algorithm relies on the active–set methodology and maintains complementarity during the whole procedure. An important issue for this type of approach is the computation of a first complementarity solution. This topic is also addressed in this talk.

Tree–Search algorithms have been proposed for finding a global minimum for the MPLCP. These include a complementarity branch–and–bound method, a sequential complementarity algorithm and an enumerative method. These procedures are based on a binary tree that is constructed by exploiting the dichotomy of the complementarity constraints.

Computational experience is reported in the last part of the talk and highlights the efficacy of the algorithms for solving MPLCPs associated with the NP–hard optimization problems mentioned above.

Partisan Drawing of Electoral Districts: How bad can it be?

*N. Apollonio¹, R.I. Becker², Isabella Lari¹, Federica Ricca³,
Bruno Simeone¹*

¹Dept. of Statistics, La Sapienza University, Roma. Italy

²Dept. of Mathematics and Applied Mathematics, University of Cape Town,
South Africa

³Dept. of Systems and Institutions for Economics, University of L'Aquila,
Italy

Abstract

Gerrymandering - the partisan and crafty manipulation of electoral districts - is a well-known ailment of electoral systems, especially first-past-the-post ones. In this talk, we provide some theoretical and experimental answers to the following questions: 1) How bad can the bias of seat apportionment be under the effect of gerrymandering? 2) How effective is compactness as a shield against gerrymandering? Accordingly, the talk is divided into two parts. In the first one, an idealized combinatorial model of gerrymandering is presented; in the second one, a more realistic multi-objective graph-partitioning model is adopted and the powerful Old Bachelor Acceptance local search heuristic is used in order to find good district designs. In a nutshell, our results for the theoretical model indicate that gerrymandering may have an extreme effect in reversing the election outcome and that compactness is indeed a very good remedy. These conclusions are confirmed to a large extent by the experimental results obtained with the latter model on some medium-large real-life test problems, especially when compactness is used in conjunction with the other two common criteria of population equality and administrative conformity.

Keywords: Electoral Systems, Graph Theory, Local Search

AMS: 91B12, 90C59, 05C15

Small Area Statistics: Methods and Applications in the Finnish Public Statistics

Risto Lehtonen

risto.lehtonen@helsinki.fi,

Department of Mathematics and Statistics, University of Helsinki

Abstract

A register-based infrastructure for the production of statistics for population subgroups and small areas is described in the paper. In Finland, the infrastructure relies on administrative registers and combinations of register data and sample survey data. Design-based and model-dependent estimation procedures are discussed and illustrated for the production of small area statistics by using synthetic data and real data bases of Statistics Finland.

Keywords: Administrative registers, EBLUP, GREG estimator.

1. Statistical infrastructure

There are different options for the production of reliable statistics for regions and other types of population subgroups or domains, depending on the statistical infrastructure in a country. In Finland, most statistics for domains and small areas are produced using administrative registers and micro-merged register data and sample survey data. These data can be linked at the unit level by using unique identification keys, which are available for example for persons, enterprises, buildings and real estate. Combined unit-level data bases amended with geographical coordinate data and unit-level register panels provide a basis for spatial and GIS techniques and flow statistics for regional statistics production [1], [2], [3]. Similar features also are becoming visible in many other European countries.

2. Methods and applications

The methods for small area statistics production discussed in the paper include selected model-assisted methods and model-dependent methods. Generalized regression (GREG) estimators represent examples of model-assisted estimation procedures [4], [5], [6], [7]. Synthetic estimators and EBLUP (Empirical Best Linear Unbiased Predictor) type estimators feature examples of model-dependent estimation procedures [8]. The models used in the construction of domain and small area estimators are members of the family of generalized linear mixed models. The estimators are compared by Monte Carlo simulation experiments and are illustrated by empirical examples. Selected computational tools are introduced. Examples are taken from Statistics Finland's Labour Force Survey and Income Distribution Survey. Empirical examples also are taken from the EURAREA project [9].

3. Contact

Address: Risto Lehtonen

University of Helsinki, Department of Mathematics and Statistics
P.O.Box 68, FI-00014 University of Helsinki, Finland

Telephone: +358-9-191 51511

Fax: +358-9-191 51400

e-mail: risto.lehtonen@helsinki.fi

4. Acknowledgement

The paper is based on joint work with Mr Kari Djerf, Mr Mikko Myrskylä, Ms Marja Tammilehto-Luode and Dr Ari Veijanen of Statistics Finland and Prof. Carl-Erik Särndal of University of Montreal.

5. Bibliography

- [1] Lehtonen, R., Djerf, K. and Tammilehto-Luode, M. (2003). *Production of statistics for domains and small areas by using micro-merged data from registers and sample surveys: benefits and challenges*. Invited paper, Bulletin of the International Statistical Institute, 54th Session, Berlin, August 2003, Vol. LX, Book 1, 432–435.
- [2] Djerf, K. and Lehtonen, R. (2005). *Small area estimation with borrowing strength in spatial and temporal dimensions: some Finnish experiences*. Contributed paper, 55th Session of the International Statistical Institute, Sydney, April 2005 (CD rom).
- [3] Statistics Finland (2001). *Population Census 2000*. Helsinki: Statistics Finland, Handbooks 35c.
- [4] Särndal, C.E., Swensson, B. and Wretman, J.H. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.
- [5] Lehtonen, R. and Veijanen, A. (1998). *Logistic generalized regression estimators*. *Survey Methodology*, 24, 51–55.
- [6] Lehtonen, R., Särndal, C.-E. and Veijanen, A. (2003). *The effect of model choice in estimation for domains, including small domains*. *Survey Methodology*, 29, 33–44.
- [7] Lehtonen, R., Särndal, C.-E. and Veijanen, A. (2005). *Does the model matter? Comparing model-assisted and model-dependent estimators of class frequencies for domains*. *Statistics in Transition*, 7, 649–673.
- [8] Rao, J.N.K. (2003). *Small Area Estimation*. Hoboken: Wiley.
- [9] Heady, P. and Ralphs, M. (2005). *EURAREA: an overview of the project and its findings*. *Statistics in Transition*, 7, 557–570.

Sección 2

Estadística

A Family of Symmetric Circular Distributions

M.C. Jones¹, Arthur Pewsey²

¹m.c.jones@open.ac.uk, Department of Statistics, The Open University, U.K.

²apewsey@guadiana.unex.es, Departamento de Matemáticas, Universidad de Extremadura.

Abstract

Consideramos una familia de distribuciones circulares unimodales simétricas que contiene como casos especiales las distribuciones uniforme, de von Mises, cardioide, y wrapped Cauchy, entre otras. La forma básica de las densidades de la familia es muy simple, aunque su constante de normalización requiere una función de Legendre asociada.

Keywords: Associated Legendre function; von Mises distribution; wrapped Cauchy distribution.

AMS: 60E05

1. Introduction

We consider a simple general family of symmetric unimodal distributions on the circle that incorporates, amongst others, the von Mises (VM), cardioid, wrapped Cauchy (WC) and circular uniform distributions as special cases. Constructions leading to the family result on conditioning and projecting certain bivariate spherically and elliptically symmetric distributions on to the circle.

The family has density

$$f_\psi(\theta) = \frac{(\cosh(\kappa\psi) + \sinh(\kappa\psi)\cos(\theta - \mu))^{1/\psi}}{2\pi P_{1/\psi}(\cosh(\kappa\psi))}, \quad \mu - \pi < \theta \leq \mu + \pi, \quad (1)$$

where $\mu \in [0, 2\pi]$ is a location parameter, $\kappa \geq 0$ is a concentration parameter equivalent to the usual parameter of the von Mises distribution, $\psi \in \mathbb{R}$ is an index controlling shape, and $P_{1/\psi}(z)$ is the associated Legendre function of the first kind of degree $1/\psi$ and order 0 (Gradshteyn & Ryzhik, 1994, Section 8.7–8.8).

The wrapped normal (WN) distribution is not a member of the family but, as is well known, the VM and WN distributions are very closely matched. Our new family provides an alternative to the wrapped symmetric stable (WSS) distributions. The special case common to both families is the WC distribution. Our family includes the VM distribution while the WSS family includes the WN distribution.

2. Basic properties

It is readily shown that the density (1) is unimodal with mode at $\theta = \mu$. The values of $\psi = 0, 1$ and -1 give, respectively, the VM, cardioid and WC distributions. $0 < \psi$ correspond to Cartwright's power-of-cosine distribution. A Dirac delta function is obtained for $-2 \leq \psi \leq 0$, $\kappa \rightarrow \infty$, whereas a new distribution related to that of Cartwright results for $-\infty < \psi < -2$, $\kappa \rightarrow \infty$. $\kappa = 0$ and $\psi \rightarrow \pm\infty$, κ finite, both yield the circular uniform distribution.

For a symmetric circular density like f_ψ , $\phi_p = E(e^{ip\Theta}) = \alpha_p$, where $\{\phi_p : p = 0, \pm 1, \dots\}$ is the characteristic function of Θ and $\alpha_p = E(\cos(p\Theta))$ is the p th cosine moment. Using Gradshteyn and Ryzhik (1994, (8.711.2)), we obtain

$$\phi_p = \begin{cases} \frac{\Gamma((1/\psi)+1)P_{1/\psi}^p(\cosh(\kappa\psi))}{\Gamma((1/\psi)+p+1)P_{1/\psi}^p(\cosh(\kappa\psi))}, & \psi > 0, \\ I_p(\kappa)/I_0(\kappa), & \psi = 0, \\ \frac{\Gamma((1/|\psi|)-p)P_{1/\psi}^p(\cosh(\kappa\psi))}{\Gamma(1/|\psi|)P_{1/\psi}(\cosh(\kappa\psi))}, & \psi < 0, \end{cases}$$

where $P_{1/\psi}^p(z)$ is the associated Legendre function of the first kind of degree $1/\psi$ and order p and $I_p(\kappa)$ is the modified Bessel function of the first kind of order p .

3. Maximum likelihood estimation

Standard asymptotic likelihood theory applies: the score equations and the observed and expected information matrices are summarized in Jones & Pewsey (2004). The maximum likelihood (ML) estimate of μ is asymptotically independent of the ML estimates of the other two parameters. The asymptotic correlation between $\hat{\kappa}$ and $\hat{\psi}$ is small to moderate for most κ and ψ , with the exception of higher correlations for small $|\psi|$, especially $\psi < 0$, and large κ . It is positive throughout.

Applications of the family in the analysis of real data are illustrated in Jones & Pewsey (2004, 2005).

4. Bibliography

- [1] Jones, M.C. and Pewsey, A. (2004). *A Family of Symmetric Distributions on the Circle*, Technical report 04/07, The Open University, Department of Statistics, available at <http://statistics.open.ac.uk/TechnicalReports/TechnicalReportsIntro.htm>.
- [2] Jones, M.C. and Pewsey, A. (2005). A Family of Symmetric Distributions on the Circle. *Journal of the American Statistical Association* 100, 1422-1428.
- [3] Gradshteyn, I.S. and Ryzhik, I.M. (1994). *Table of Integrals, Series, and Products*, fifth edition edited by A. Jeffrey. Academic Press.

A new Multiblock PLS based method to estimate causal models

Francisco Arteaga¹, Martina G. Galarza², Irene Gil³

¹francisco.arteaga@ucv.es, ²mggalarza@ucv.es, Facultad de Estudios de la Empresa, Universidad Católica de Valencia San Vicente Mártir

³irene.gil@uv.es, Facultad de Ciencias Económicas, Universitat de Valencia

Abstract

En este trabajo se presenta un nuevo algoritmo para estimar modelos causales basado en multiblock PLS. El algoritmo se prueba con un conjunto de datos reales y varios simulados, comparando los resultados con el método LVPLS. Ambos métodos confirman las relaciones estructurales propuestas, pero el Nuevo método explica una mayor proporción de la varianza.

Palabras Clave: MBPLS, LVPLS, Modelos causales

1. Introducción

Partial Least Squares (PLS) es un método de regresión desarrollado por Herman Wold y sus colaboradores [1, 2, 3]. Stone and Brooks [4] demuestran que PLS puede verse como un proceso en dos etapas en el que un conjunto de k variables predictoras se transforman primero en un nuevo conjunto de A ($A < K$) factores maximizando su covariance con la variable respuesta, con la restricción de ortogonalidad entre los factores. Por comparar con otros métodos de regresión, en Principal Components Regression (PCR) los factores se determinan maximizando su varianza con la restricción de ortogonalidad y en Ordinary Least Squares (OLS) se elige el factor de manera que se maximiza la correlación con la variable respuesta.

PLS Path Modelling (PLSPM) [5] es una extensión de PLS útil para estimar los coeficientes de modelos causales, en los que puede haber más de un bloque de variables predictoras y una estructura de relaciones lineales más compleja: varios bloques de variables preditoras, varios bloques de variables predichas y varios bloques intermedios, que son simultáneamente predotores y predichos.

En PLSPM se asume que cada bloque de variables se puede resumir por una única variable latente (LV) y que se dispone de un conocimiento teórico que fundamenta la existencia de un conjunto de relaciones lineales entre las diferentes variables latentes.

En el presente trabajo se presenta un nuevo método para la estimación de los coeficientes de las citadas relaciones lineales basado en Multiblock PLS

(MBPLS) [6] y lo comparamos con el método comúnmente empleado en la literatura denominado LVPLS.

Para la comparación se emplea un estudio del comportamiento post-compra. Ambos métodos respaldan las relaciones estructurales propuestas, aunque el nuevo algoritmo lo hace consiguiendo explicar una ligeramente mayor proporción de la varianza. La ligera superioridad del nuevo método se comprueba a partir de datos simulados.

2. Bibliography

- [1] Wold, H., in Research Papers in Statistics; Festschrift for J. Neyman, ed. by F.N. David, pp. 411-444, Wiley, New York (1966).
- [2] Wold, H., in Multivariate Analysis III, ed. By P. Krishnaiah, pp. 383-407, Academic, New York (1973).
- [3] Jöreskog K.G., Wold H. (1982): Systems under indirect observation, North Holland, Amsterdam. Parts I and II.
- [4] Stone, M. and Brooks, J. R. Stat. Soc. B, 52, 237-269 (1990).
- [5] Lohmöller J.B. (1989), *Latent variables path modeling with partial least squares*, Physica Verlag, Heidelberg.
- [6] Wangen L.E., Kowalski B.R. (1988), A multiblock partial least squares algorithm for investigating complex chemical systems. Journal of Chemometrics, 3, 3-20.

A simulation study of a nested sequence of multinomial loglinear models

Nirian Martín¹, Leandro Pardo²

¹nirian@estad.ucm.es, Statistics and O.R. III, Complutense U. of Madrid

²lpardo@mat.ucm.es, Statistics and O.R. I, Complutense U. of Madrid

Abstract

En este trabajo se presenta un estudio de simulación para analizar el comportamiento de los estadísticos Phi-divergencia en la elección del modelo loglineal con restricciones lineales de entre una secuencia jerarquizada de modelos en el supuesto de muestreo multinomial.

Keywords: Multinomial loglinear models, Minimum ϕ -divergence estimators, ϕ -divergence test statistics.

1. Introduction

We consider a discrete model, i.e. statistical sample Y_1, Y_2, \dots, Y_n of size $n \in \mathbb{N}$ with components taking on values in $\mathcal{Y} = \{C_1, \dots, C_k\}$ and i.i.d. as a probability distribution $\mathbf{p}(\boldsymbol{\theta}_0) = (p_1(\boldsymbol{\theta}_0), \dots, p_k(\boldsymbol{\theta}_0))^T$, $\boldsymbol{\theta}_0 \in \Theta \subset \mathbb{R}^{t+1}$. We denote $\widehat{\mathbf{p}} = (\widehat{p}_1, \dots, \widehat{p}_k)^T$ with $\widehat{p}_j = N_j/N$ and $N_j = \sum_{i=1}^n I_{\{C_j\}}(Y_i)$. Then (N_1, \dots, N_k) is multinomially distributed with parameters $(n; p_1(\boldsymbol{\theta}_0), \dots, p_k(\boldsymbol{\theta}_0))$ and its vector mean will be denoted by $\mathbf{m}(\boldsymbol{\theta}_0) = (m_1(\boldsymbol{\theta}_0), \dots, m_k(\boldsymbol{\theta}_0))^T$. Given a $k \times (t+1)$ matrix \mathbf{X} and $\text{rank}(\mathbf{X}) = t+1$, the set

$$C(\mathbf{X}) = \{\log \mathbf{m}(\boldsymbol{\theta}) \in \mathbb{R}^k : \log \mathbf{m}(\boldsymbol{\theta}) = \mathbf{X}\boldsymbol{\theta}, \boldsymbol{\theta} \in \mathbb{R}^{t+1}\} \quad (2.1)$$

represents the class of the loglinear models associated with \mathbf{X} . We suppose, in the following, that $\mathbf{J} = (1, \dots, 1)^T \in C(\mathbf{X})$. Taking into account $m_j(\boldsymbol{\theta}_0) = np_j(\boldsymbol{\theta}_0)$, $j = 1, \dots, k$, the parameter space is defined by $\Theta' = \{\boldsymbol{\theta} \in \mathbb{R}^{t+1} : \log \mathbf{m}(\boldsymbol{\theta}) = \mathbf{X}\boldsymbol{\theta} \text{ and } \mathbf{J}^T \mathbf{m}(\boldsymbol{\theta}) = n\}$. Now in addition to the previous model we shall assume that we have $r-1 < t$ linear constraints defined by $\mathbf{C}^T \mathbf{m}(\boldsymbol{\theta}) = \mathbf{d}^*$, where \mathbf{C} and \mathbf{d}^* are $k \times (r-1)$ and $(r-1) \times 1$ matrices, respectively. If we consider the linear constraint $\mathbf{J}^T \mathbf{m}(\boldsymbol{\theta}) = n$, associated with the multinomial sampling, we can write the parameter space for this new model by

$$\Theta^* = \{\boldsymbol{\theta} \in \mathbb{R}^{t+1} : \log \mathbf{m}(\boldsymbol{\theta}) = \mathbf{X}\boldsymbol{\theta} \text{ and } \mathbf{L}^T \mathbf{m}(\boldsymbol{\theta}) = \mathbf{d}\}, \quad (2.2)$$

where $\mathbf{L} = (\mathbf{J}, \mathbf{C})$, $\mathbf{d} = (n, (\mathbf{d}^*)^T)$ and $\text{rank}(\mathbf{L}) = \text{rank}(\mathbf{L}^T, \mathbf{d}) = r$.

2. Hierarchical Hypotheses and ϕ -divergences

Now we consider the problem of testing a hierarchical sequence of log-linear models. Consider a nested sequence of hypotheses

$$H_l : \mathbf{p} = \mathbf{p}(\boldsymbol{\theta}) \text{ and } \boldsymbol{\theta} \in \Theta_0^{(l)}; \quad l = 1, \dots, m, \quad m \leq t < k-1, \quad (2.3)$$

where $\Theta_0^{(1)} \supset \Theta_0^{(2)} \supset \dots \supset \Theta_0^{(m)}$ with $\dim(\Theta_0^{(l)}) = t_l + 1$; $\text{rank}(\mathbf{L}_l) = r_l$, $l = 1, \dots, m$, such that $t_{l+1} \leq t_l$ and $r_{l+1} \geq r_l$; $l = 1, \dots, m - 1$, where at least one of both inequalities is strict. In this framework, there is an integer m^* ($1 \leq m^* \leq m$) for which H_{m^*} is true but H_{m^*+1} is not true. The suggested hypothesis testing strategy is to test successively,

$$H_{\text{Null}} : H_{l+1} \text{ against } H_{\text{Alt}} : H_l; \quad l = 1, \dots, m - 1, \quad (2.4)$$

where we continue to test as long as the null hypothesis is accepted, and we infer m^* to be the first l for which H_{l+1} is rejected as a null hypothesis. Using the result obtained in Martín and Pardo (2006b) we must reject the null hypothesis iff

$$T_{\phi_1, \phi_2}^{(l)} = \frac{2n}{\phi_1''(1)} D_{\phi_1}(\mathbf{p}(\widehat{\boldsymbol{\theta}}_l^{(r), \phi_2}), \mathbf{p}(\widehat{\boldsymbol{\theta}}_{l+1}^{(r), \phi_2})) > \chi_{t_l - t_{l+1} - r_l + r_{l+1}}^2 (1 - \alpha), \quad (2.5)$$

where $\widehat{\boldsymbol{\theta}}_l^{(r), \phi_2}$ and $\widehat{\boldsymbol{\theta}}_{l+1}^{(r), \phi_2}$ are the minimum ϕ -divergence estimators under the alternative and null hypothesis respectively, and $\chi_{t_l - t_{l+1} - r_l + r_{l+1}}^2 (1 - \alpha)$ the $(1 - \alpha)$ -th quantile of a $\chi_{t_l - t_{l+1} - r_l + r_{l+1}}^2$ distribution. Minimum ϕ -divergence estimators in this type of log-linear models were introduced in Martín and Pardo (2006a). In this paper we study the behavior of the family of statistics in (2.5) on the basis of a simulation study.

3. Acknowledgement

This work was supported by Grants BMF2003-00892 and UCM2005-910707.

4. Bibliography

- [1] Martín, N. and Pardo, L. (2006a): Minimum phi-divergence estimators for loglinear models with linear constraints and multinomial sampling. To appear in *Statistical Papers*.
- [2] Martín, N. and Pardo, L. (2006b): Phi-divergence test statistics in multinomial sampling for hierarchical sequences of Loglinear Models with Linear Constraints. *Ninth International Conference Zaragoza-Pau on Applied Mathematics and Statistics*.
- [3] Pardo, L. (2006). *Statistical Inference Based on Divergence Measures*. Statistics: Textbooks and Monographs. Chapman & Hall/CRC, New York.

ABC-PLOT 1.1 Aplicación informática para la representación gráfica de datos multidimensionales

C.A. Ávila Zarza¹, S. Rodríguez González², R.M. Martín Moreno³

¹caaz@usal.es, ²srg@usal.es

Departamento de Estadística

³rosa.mmoreno@gmail.com, Departamento de Informática y Automática
Universidad de Salamanca

Abstract

Se presenta en este trabajo la aplicación informática ABC-PLOT en su versión 1.1.; aplicación que permite el desarrollo y utilización del método gráfico multivariante ABC-PLOT, y por tanto representar gráficamente información k-dimensional, lo que facilita su visualización y análisis estadístico exploratorio y confirmatorio.

Palabras Clave: ABC-PLOT 1.1., Software de Análisis Multivariante

AMS: 62-04, 62-09, 62H99

1. Introducción

El método ABC-PLOT [1] es un Método Gráfico Multivariante que constituye en si mismo un procedimiento descriptivo, ya que utiliza el alfabeto como base en la construcción de gráficos figurativos o *configuraciones* [2] para la representación y análisis de datos multidimensionales, al modo de otros métodos (*Curvas de Andrews*, *Caras de Chernoff*, etc), pero solventando claramente algunos de los inconvenientes que presentan éstos. Al igual que la generalidad de los MGM, uno de los inconvenientes que presenta es la necesidad de un software para realizar los gráficos y llevar a cabo el análisis. Con este trabajo pretendemos soslayar este inconveniente al presentar la aplicación informática ABC-PLOT en su versión 1.1.

2. Descripción y Aspectos relevantes de la aplicación informática ABC-PLOT 1.1

Esta aplicación ha sido desarrollada siguiendo el patrón modelo-vista-controlador (MVC), patrón de diseño software que separa los datos de una aplicación, la interfaz de usuario y la lógica de control en tres componentes distintos de forma que las modificaciones al componente de visualización pueden ser hechas con un mínimo impacto en el componente del modelo de datos.

Las características del lenguaje *Java* [3] hicieron que éste fuese el lenguaje elegido para el desarrollo y la implementación de la aplicación en diversos

entornos tales como Unix, Pc, Mac, etc. La creación de gráficos que lleva a cabo ABC-PLOT, se ha desarrollado utilizando las bibliotecas de clases proporcionadas por Java para la generación de gráficos 2D. Como documentación adicional se ha utilizado la herramienta *Javadoc* generando una ayuda en formato html enfocada a programadores o bien a gente interesada en el código fuente.

Los requisitos mínimos de esta aplicación, son:

Entorno hardware: Procesador PC/AT AMD K6 II a 500Mhz ó superior /Pentium II a 500Mhz o superior, 128 Mb de RAM, 200Mb de espacio libre en disco duro si no se dispone de maquina virtual Java.

Entorno software: Cualquier sistema operativo actual con interprete de Java o entorno capaz de ejecutar aplicaciones Java como JDK 1.5.0_01 ó superior.

3. Contactar

Carmelo A. Ávila Zarza

Dirección:

Departamento de Estadística
C/ Espejo 2
37007 – Salamanca
Spain

Teléfono: (+34) 923294500 Ext 1921 **Fax:** (+34) 923294619

e-mail: caaz@usal.es

4. Bibliografía

- [1] Ávila Zarza, C.A. (1993): 'Métodos Gráficos Multivariantes y su aplicación en las ciencias de la vida'. Trabajo de Grado. Universidad de Salamanca.
- [2] Wilkinson, L. (2005): 'The grammar of Graphics'. Springer, Chicago.
- [3] Gosling, J. (1990): Sun Microsystems, Inc. 4150 Network Circle. Santa Clara, CA 95054.

Adaptive Combination of Forecast, with Application to Wind Energy

Ismael Sánchez¹

¹Departamento de Estadística, Universidad Carlos III de Madrid

Abstract

En este trabajo se presenta un método de combinación de predicciones adaptativo, de manera que los pesos de la combinación van cambiando con el tiempo. El método es aplicado a la predicción de energía eólica.

Keywords: Combination of Forecast; Wind Energy.

1. Introduction

It is not infrequent in wind energy to have access to more than one prediction of the production for the next hours. In those cases, the adaptive combination of forecasts might be a useful methodology to generate an efficient single forecast. This work proposes an adaptive combination method called AEC that is similar to always using the best individual predictor. Then, a two steps combination methodology is proposed to combine a set of alternative predictions. This two steps procedure aims to take the advantage of the different approaches of forecast combination. In the first step, several combinations methods are used, being the AEC one of them. In the second step, the AEC method is used to combine the alternative combinations of the first step. The application to a real data set illustrates the usefulness of the proposed methods to obtain the best output from a set of alternative predictions.

2. Contact

Address:

Ismael Sánchez
Department of Statistics, University Carlos III of Madrid
Av. Universidad 30
28911 - Leganés (Madrid)
Spain

Telephone: (+34) 916249179

Fax: (+34) 916249430

e-mail: ismael@est.econ.uc3m.es

Adaptive Goodness-of-Fit Tests II: ARE-aided Basis Choice

Aurea Grané¹, Josep Fortiana²

¹aigrane@est-econ.uc3m.es, Dep. de Estadística, Univ. Carlos III, Spain.

²fortiana@ub.edu, Dep. de Probabilitat, Lògica i Estadística,
Fac. de Matemàtiques, Univ. de Barcelona, Spain.

Abstract

Proponemos un procedimiento, basado en la eficiencia relativa asintótica de Bahadur, para seleccionar una base ortonormal con la que construir estadísticos de bondad de ajuste a una ley uniforme para muestras univariantes, adaptados al objetivo de detectar una familia de alternativas.

Keywords: Goodness of fit; Bahadur asymptotic relative efficiency; L-statistics.

AMS: 62G30, 62G20

1. Introduction

In [1] we studied a test of goodness-of-fit to the $(0, 1)$ -Uniform distribution for univariate n -samples. Its test statistic, Q_n , evolved from the concept of maximum Hoeffding correlation between two distributions, namely the sample empirical distribution and the null-hypothesized. The outstanding performance of Q_n in detecting a wide variety of alternative distributions makes it a choice omnibus test, often improving on standard ones such as Kolmogorov's or Cramér-von Mises'. For some specific families of alternative distributions, however, the power of the maximum correlation statistic is remarkably low. Analysis of this phenomenon suggested an adaptive test, as a matter of fact a perturbation of the decomposition of Q_n along an orthonormal sequence naturally associated with it by means of a Karhunen-Loève expansion (see [2]).

The answer in the negative ([3]) to the question whether this KL orthonormal basis has a specific good power performance leads to the problem of comparing different bases which we address in the present work.

2. Approximate Bahadur slope in comparison of tests

Given a family $\{F_\theta\}$ of distributions containing the $(0, 1)$ -Uniform for $\theta = \theta_0$ (the null hypothesis), we derive an expression for the approximate Bahadur slope $c(\theta)$ for the test statistic $T = \sum_{j=0}^p \lambda_j \Phi_j$, where $\Phi_{nj} = \int_0^1 F_n^{-}(t) \phi_j(t) dt$, F_n^{-} is the g -inverse of the sample empirical distribution and $\phi = \{\phi_j\}_{j \geq 0}$ is an orthonormal sequence in $L^2[0, 1]$. The result is a closed formula, expressing $c(\theta)$ in terms of quantities easily calculable from $\{F_\theta\}$ and ϕ .

Assuming that two test statistics $T_1 = \sum_{j=0}^p \lambda_{1,j} \Phi_{nj}$ and $T_2 = \sum_{j=0}^p \lambda_{2,j} \Psi_{nj}$ are constructed as described above from two orthonormal sequences, $\phi = \{\phi_j\}_{j \geq 0}$ and $\psi = \{\psi_j\}_{j \geq 0}$, T_1 is asymptotically more efficient than T_2 if $c_1(\theta)/c_2(\theta) > 1$. We show some examples of how, keeping a fixed alternative family of distributions $\{F_\theta\}$, we can select the best orthonormal basis from two given ones, ϕ and ψ , by comparing the functions $c_1(\theta)$ and $c_2(\theta)$.

A question remains open: whether, given $\{F_\theta\}$, there exists an optimal orthonormal basis in the sense of maximum power and, hopefully, some easy path towards locating it.

3. Contact

Address: Aurea Grané

Departamento de Estadística,
Universidad Carlos III,
Calle Madrid, 126, 28903 Getafe (Madrid),
Spain.

Telephone: (+34) 91 624 8917

Fax: (+34) 91 624 9849

e-mail: agrane@est-econ.uc3m.es

www: <http://www.est.uc3m.es/agrane>

4. Acknowledgement

Work supported in part by grants: TEC2004-07853-C02-01 (Ministerio de Ciencia y Tecnología), MTM2005-02370 (Ministerio de Educación y Ciencia - FEDER).

5. Bibliography

- [1] Fortiana, J. and Grané, A. (2003). Goodness-of-fit Tests Based on Maximum Correlations and their Orthogonal Decompositions. *Journal of the Royal Statistical Society B* **37**, 216–237.
- [2] Grané, A. and Fortiana, J. (2006). An Adaptive Goodness-of-fit Test. *Communications in Statistics A. Theory and Methods* **35** (6).
- [3] Grané, A. and Fortiana, J. (2006). Karhunen-Loève Basis in Goodness-of-fit Tests Decomposition: An Evaluation. (*Working Paper*)

Ajuste de modelos lineales mixtos con efectos aleatorios anidados

Domingo Morales¹, Agustín Pérez¹

¹d.morales@umh.es y agustin.perez@umh.es, Centro de Investigación Operativa, Universidad Miguel Hernández de Elche.

Abstract

En el presente trabajo se proponen tres procedimientos para el ajuste de modelos lineales mixtos con efectos aleatorios anidados. Para evaluar la idoneidad de estos procedimientos se realiza un experimento de simulación Monte Carlo que permite la comparación de determinadas propiedades entre ellos.

Claves: Modelos lineales mixtos, efectos aleatorios anidados, algoritmos de ajuste.

AMS: 62J05, 65C05

1. Introducción

La particularidad de los modelos lineales mixtos es que son idóneos para la estimación de efectos de factores con muchos niveles, que son simplificados a efectos aleatorios en los cuales solo se estima un único parámetro de varianza. Los modelos lineales mixtos adoptan la expresión general

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}.$$

Habitualmente se supone normalidad e independencia en los errores, en los efectos aleatorios y entre ellos.

En este trabajo, se utilizan tres procedimientos para resolver el problema de estimación de los parámetros de regresión y de las componentes de la varianza en modelos lineales mixtos con un factor aleatorio y dos factores aleatorios anidados. Tales procedimientos son:

1. El método de la máxima verosimilitud (véase [2]), que obtiene estimaciones de los parámetros maximizando la log-verosimilitud del modelo. En este caso se ha de recurrir a algoritmos de cálculo numérico, tales como Fisher-Scoring, para la obtención de los estimadores debido a la imposibilidad de obtener fórmulas explícitas.
2. El método de la máxima verosimilitud residual (véase [2]), que transforma el modelo original para estimar las componentes de la varianza en primer lugar y los parámetros de regresión en segundo lugar y por separado. Al

igual que el método anterior, los estimadores se obtienen maximizando una función de verosimilitud (la del modelo transformado). De este modo se reduce el coste computacional de la obtención de los valores numéricos de los estimadores.

3. El método 3 de Henderson (véase [1]), que evita el coste computacional de los anteriores métodos pues aporta fórmulas explícitas. En una primera fase se tratan todos los efectos como si fuesen efectos fijos. Se ajusta el modelo de efectos fijos y se obtienen sus sumas de cuadrados. Se calculan los valores esperados de las mismas y se aplica el método de los momentos para obtener un sistema triangular de ecuaciones cuyas incógnitas son las componentes de la varianza. El estimador del vector de parámetros de regresión se obtiene reemplazando las componentes de la varianza por sus estimadores en la fórmula explícita obtenida por el método de los mínimos cuadrados.

Se elaboran algoritmos para la utilización de cada uno de los tres métodos y se implementan en C++. Con el fin de evaluar los tres métodos, se realizan experimentos de simulación Monte Carlo. Para ello se simulan 1.000 muestras y se estiman los parámetros del modelo en cada una de ellas y para cada uno de los tres procedimientos de estimación. Este proceso se repite para diferentes tamaños muestrales con el fin de estudiar el comportamiento de los procedimientos de estimación. Se analizan con detalle aquellos casos en los que los tamaños muestrales son pequeños dentro de los niveles de los factores aleatorios. Se calcula el error cuadrático medio y el sesgo para cada parámetro estimado y para cada conjunto de tamaños muestrales. El trabajo finaliza con la presentación de conclusiones.

2. Bibliografía

- [1] Henderson, C.R. (1953). Estimation of variance and covariance components. *Biometrics*, **9**, 226–252.
- [2] Searle, S.R., Casella, G. and McCulloch, C.E. (1992). *Variance Components*. John Wiley. New York.

Algorithm based on the residual variance of a linear model to find gene expression profiles of de-regulation and identify families of disease-altered genes

M. J. Rivas-López¹, J. M. Sánchez-Santos¹, J. López-Fidalgo², C. Prieto-Sánchez³, J. de Las Rivas-Sanz³

¹chusrl@usal.es, jose@usal.es, Department of Statistics, Faculty of Science,
University of Salamanca

²jesus.lopezfidalgo@uclm.es, Department of Mathematics, E.T.S.I.
Industriales, University of Castilla la Mancha

³cprietas@usal.es, jrivas@usal.es, Bioinformatics and Functional Genomics
Research, Cancer Research Center (CIC, USAL-CSIC)

Abstract

Construimos un algoritmo para buscar genes “desregulados”, es decir, que presentan una alteración significativa en la variabilidad de su nivel de expresión genómica entre un grupo de pacientes sanos y otro de pacientes enfermos. El algoritmo calcula grupos de genes junto con sus niveles de significación, usando la varianza residual de un modelo lineal con 2 factores.

Keywords: Analysis of the Variance, Residual Variance, F-scores, Correlated Genes.

1. Introduction

We use RMA method [1] for background correction and normalization of the expression signals for all the microarrays analysed. In our work [2] we find groups of genes that present a similar expression pattern in control samples but suffer a distinct alteration of their expression profiles in the disease samples (as we show in Figure 1).

2. Score function for altered expression and statistical adjustment

Each experimental set of microarrays, with G genes from S samples, can be assigned to a $|G| \times |S|$ -dimensional matrix $\mathbf{G} \times \mathbf{S} = (e_{ij})$, where e_{ij} is the normalized intensity of the i -th gene for the j -th sample. Due to the fact that the aim is to compare only two states, the samples set are always divided in two independent subsets: \mathbf{S}_a (*altered or disease samples*) and \mathbf{S}_c (*control samples*); so $|S| = |S_a| + |S_c|$. For each expression value e_{ij} , we can consider a linear model: $e_{ij} = \alpha_i + \beta_j + \gamma + \epsilon_{ij}$ with the gene effects (α_i), the sample effects (β_j), a constant value common to all the data in a set (γ), and the errors (ϵ_{ij}). To find the variability of expression values, we use the *residual variance* \mathbf{RV} [3]: $RV(G,S) = (e_{ij} - \bar{e}_{i\cdot} - \bar{e}_{\cdot j} - \bar{e}\cdot)^2$

The *residual variance* \mathbf{RV} is the ratio between a *Pearson's chi-square* distribution and its degrees of freedom. To compare \mathbf{RV} for the two independent subsets of samples, altered and control (\mathbf{S}_a and \mathbf{S}_c), in a given subset of genes \mathbf{G} , we calculate each *residual variance* $\mathbf{RV}_c = \mathbf{RV}(\mathbf{G}, \mathbf{S}_c)$ and $\mathbf{RV}_a = \mathbf{RV}(\mathbf{G}, \mathbf{S}_a)$, and the comparison can be done by the *relative residual variance*: $RRV =$

The division of these two *Pearson's chi-square* distributions follows a *Fisher-Snedecor's* distribution \mathbf{F}_{n_1, n_2} with $n_1 = (|\mathbf{G}| - 1) \cdot (|\mathbf{S}_c| - 1)$ and $n_2 = (|\mathbf{G}| - 1) \cdot (|\mathbf{S}_a| - 1)$ degrees of freedom. This allows constructing for each subset of genes \mathbf{G} , a statistical hypotheses test with its own significance levels:

$$H_0: \sigma_a^2 \leq \sigma_c^2 \text{ (no difference in variability between control and disease)}$$

$$H_1: \sigma_a^2 > \sigma_c^2 \text{ (difference in variability between control and disease)}$$

The algorithm looks for the genes of H_1 , i.e.: genes with a great variability in altered samples but a small one in control samples. See Figure 1. Our method is implemented in an R package called AlteredExpression available in <http://bioinfow.dep.usal.es/Altered-Expression/>

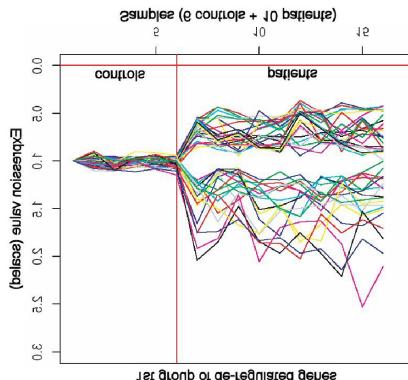


Figure 1: 1st group of de-regulated genes

3. Bibliography

- [1] Irizarry R.A., Bolstad B.M., Collin F., Cope L.M. et al (2003a). Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.*, **31**, e15.
- [2] Prieto C., Rivas M.J., Sánchez J.M., López-Fidalgo J. and De las Rivas J. (2006). Algorithm to find gene expression profiles of de-regulation and identify families of disease-altered genes. *Bioinformatics*. In press.
- [3] Kostka D. and Spang R. (2004). Finding disease specific alterations in the coexpression of genes. *Bioinformatics*, **20 Suppl 1**, I194-I199.

Algoritmo de Balas Adaptado a la Detección e Imputación Automática de datos cualitativos (ABADIA)

Pedro Martínez Talaván

¹ptalavan@ine.es, Subdirección General de Informática Estadística, Instituto Nacional de Estadística

Abstract

En este artículo se propone, a partir del esquema de Balas, un nuevo algoritmo para la depuración de datos cualitativos. Con él se evitan algunos de los problemas de la metodología Fellegi-Holt, como la necesidad de partir las depuraciones en trozos o asignar imputaciones con valores de muy baja frecuencia.

Palabras Clave: Depuración de cuestionarios, Imputación Automática.

AMS: Clasificación AMS (MSC 2000): 62-02, 90C09, 90C27

1. Introducción

La principal metodología utilizada hasta la fecha para la depuración de datos cualitativos de encuestas es la desarrollada por Fellegi-Holt (1976) que se basa en los siguientes principios:

- Se deben especificar reglas de inconsistencias entre datos, pero no se debe especificar como corregirlo.
- Se debe corregir el cuestionario imputando los menos campos posibles.
- La asignación de códigos a los campos erróneos se debe hacer respetando unas distribuciones marginales y sobre todo conjuntas de los campos en los ficheros originales señaladas previamente.

El principal problema de esta metodología es que son necesarias, en la práctica totalidad de los casos, particiones de la depuración en partes más pequeñas, ya que una de las etapas previas a la depuración del fichero se hace inviable con tamaños medios. Esto, además de provocar más imputaciones de las necesarias, genera una gran cantidad de trabajo.

Por otro lado, cuando se imputan valores a los campos, se dan con relativa facilidad asignaciones de datos 'extraños', es decir, combinaciones de datos que no aparecen o son raros en las frecuencias originales.

2. ABADÍA

Es fácil comprobar que, para datos cualitativos, el problema de minimizar el número de imputaciones que superen las reglas de inconsistencias, es un problema de optimización lineal binario con una función objetivo muy sencilla. A partir de esta idea, se ha planteado utilizar, como base, el algoritmo de Balas para el diseño de un nuevo procedimiento, que superase las dificultades de la metodología de Fellegi-Holt.

Con este planteamiento, se ha desarrollado un nuevo algoritmo, cuyo núcleo principal es una generalización del algoritmo de Balas para variables discretas, al cual se le han añadido pasos para incluir imputaciones determinísticas. Además, una vez obtenida una solución del problema de optimización combinatoria, se realiza un sorteo entre un conjunto de posibles soluciones similares; asignando probabilidades mediante los ficheros de frecuencias guías fijadas previamente por el usuario. Las diferencias de este nuevo algoritmo respecto a la metodología Fellegi-Holt son las siguientes:

- No necesita partir la depuración, ya que no son necesarias las etapas previas de la metodología de Fellegi-Holt.
- En la metodología de Fellegi-Holt, en primer lugar se buscan los campos a imputar y después se asignan valores. Sin embargo, en este algoritmo se tratan simultáneamente campos y valores. Con esto se dan menos casos de imputaciones 'extrañas'.
- Cuando se utilizan frecuencias guías en la imputación secuencial de la metodología Fellegi-Holt, estas sólo se utilizan en un sentido, es decir, si tenemos una frecuencia bidimensional con los campos (A, B), esta sólo se utiliza si se va a imputar el campo B , no para imputar el campo A . Sin embargo, con este modelo siempre utiliza la conjunta aunque se imputen los dos campos. Esto también evita casos de imputaciones 'extrañas'.

En la actualidad está programada una versión Beta que se está probando con la nueva EPA; esta depuración está dividida en 5 partes para la metodología Fellegi-Holt y se ha conseguido resolver de forma conjunta con el aBadía. Estos ficheros están siendo analizados por el servicio promotor.

3. Bibliografía

- [1] Fellegi IP, Holt D. (1976) *LATEX A Systematic Approach to Automatic Edit and Imputation*. Journal if the American Statistical Association, 71, 17-35.
- [2] Balas E. (1965) *LATEX An additive algorithm for solving the linear programs with zero-one variables*. The Journal of the Operations Research Society of America, 13, 4 (July- Aug. 1965): 517-546.

Análisis bayesiano de un sistema de colas $G/M/1$ con proceso de llegadas de cola pesada

Pepa R. Cobo¹, Rosa E. Lillo², Michael P. Wiper³

¹jrcobo@est-econ.uc3m.es, ²lillo@est-econ.uc3m.es, ³mwiper@est-econ.uc3m.es

Departamento de Estadística, Universidad Carlos III de Madrid

Abstract

Las distribuciones de cola pesada han adquirido gran relevancia en la modelización del tráfico de datos en Internet. En este trabajo, se estudia desde una perspectiva bayesiana, el sistema de colas $G/M/1$ con la particularidad de que se utiliza una mixtura de distribuciones Pareto como distribución que caracteriza el proceso de llegadas. El análisis se realiza mediante un método que aproxima la transformada de Laplace de la mixtura de Pareto.

Claves: cola pesada, inferencia bayesiana, sistemas de colas.

AMS: 62F15, 68M20

1. Introducción

Desde que Internet se introdujo de forma tan determinante en nuestra sociedad, existen muchos trabajos dedicados a estudiar cuáles son las propiedades que caracterizan el tráfico de datos en la red con fines de modelización, predicción y solución de problemas habituales como la congestión. Se ha observado que ciertas variables como el tiempo de transmisión, los tamaños de archivos descargados o las longitudes de los servicios tienen distribuciones de cola pesada ([1]). Estas variables se caracterizan por carecer total o parcialmente de momentos y, además, no poseen una fórmula cerrada o explícita de la transformada de Laplace, dificultando así el análisis clásico de algunos sistemas de colas, que depende tanto del tiempo esperado entre las llegadas, como de la transformada de Laplace de esa distribución. En particular, para obtener la distribución estacionaria de la ocupación del sistema en el modelo $G/M/1$ es necesario resolver la ecuación:

$$r_0 = A^*[\mu(1 - r_0)],$$

donde $A^*(\cdot)$ es la transformada de Laplace de la distribución de los tiempos entre llegadas y μ es el parámetro de la exponencial que regula el servicio y $r_0 \in (0, 1)$ es el parámetro de interés que hay que encontrar. Cuando la distribución del tiempo entre llegadas es de cola pesada no existe una expresión analítica de $A^*(\cdot)$, siendo necesario utilizar alguna alternativa para el estudio del sistema, como puede ser un método de aproximación de la transformada de Laplace, que es lo que se describe en esta presentación.

2. Modelo de colas y procedimiento de estimación

En este trabajo se considera la mixtura de distribuciones Pareto como ejemplo de distribución de cola pesada que gobierna los tiempos entre llegadas en el sistema de colas $G/M/1$. Mediante inferencia bayesiana se estiman las distribuciones típicas y las cantidades de interés asociadas al sistema, habiendo previamente aproximado la transformada de Laplace. Finalmente, aplicamos nuestros resultados tanto a datos simulados como reales. En concreto, la función de densidad para los tiempos entre llegadas es la siguiente,

$$f(x|\mathbf{w}, \boldsymbol{\beta}, \mathbf{b}) = \sum_{r=1}^k w_r \frac{\beta b^\beta}{(x+b)^{\beta+1}}, \quad \text{para } \beta > 0, b > 0 \text{ y } x > 0$$

donde k es el número de componentes de la mixtura, $w_r > 0$ para $r = 1, \dots, k$ y $w_1 + \dots + w_k = 1$.

Para estimar los parámetros correspondientes al modelo, las medidas del sistema y las distribuciones implicadas en un estudio habitual de un modelo de colas, se realiza inferencia bayesiana mediante un algoritmo de Monte Carlo para cadenas de Markov con procesos de nacimiento y muerte (BDMCMC, ver [2]). Como la distribución del tiempo entre llegadas es de cola pesada, es necesario aproximar la transformada de Laplace. Para ello, se generaliza el método propuesto en [3], conocido como *Transform Approximation method* (o *TAM*) adaptándolo al modelo de mixturas Pareto propuesto. Este método aproxima la transformada de Laplace $A^*(\cdot)$ de la distribución del tiempo entre llegadas por

$$A_N^*(z) = \frac{1}{N} \sum_{i=1}^N e^{-zx(i)},$$

donde N es suficientemente grande y los $x(i), i = 1, 2, \dots, N$ se eligen de manera que, $A(x(i)) = i/(N+1)$, siendo $A(\cdot)$ la función de distribución del tiempo entre llegadas.

3. Bibliografía

- [1] Crovella, M., Taqqu, M., Bestavros, A. (1998). Heavy-tailed probability distributions in the world wide web. *A Practical Guide To Heavy Tails*. Capítulo 1, Chapman & Hall, New York. 3-26
- [2] Stephens, M. (2000) Bayesian analysis of mixtures with an unknown number of components -an alternative to reversible jump methods. *Annals of Statistics*, 28:40-74.
- [3] Harris, C., Brill, P., Fisher, M. (2000) Internet-type queues with power-tailed interarrival times and computational methods for their analysis. *INFORMS Journal on Computing*. 12(4): 261-271.

Análisis Canónico de Correspondencias de tres vías: Metodología y Aplicación

José L. Vicente Villardon¹, Ángela Cortés², M^a José Fernández¹, Purificación Galindo¹

¹villardon@usal.es, Departamento de Estadística, Universidad de Salamanca,
España

² Ministerio de Educación, Chile

Abstract

Un problema que se presenta habitualmente en el análisis de datos ecológicos es el de relacionar un conjunto de especies con una serie de características ambientales y comprobar la estabilidad de las relaciones en el tiempo. En este trabajo presentamos una generalización del Análisis Canónico de Correspondencias para datos de tres vías.

Palabras Clave: Biplot, Análisis Canónico de Correspondencias, datos de tres vías

1. Introducción

En el análisis de tablas de tres vías a menudo se dispone de información complementaria o adicional que es necesario incorporar al estudio de la estructura de este tipo de datos multivariantes. Por ejemplo, es habitual en el estudio de datos ecológicos disponer de matrices de abundancias de especies, replicadas en el tiempo, que han de relacionarse con un conjunto de características ambientales medias sobre los mismos lugares. Ninguna de las técnicas presentes en la bibliografía permite recoger en el análisis las características de este tipo de datos.

2. Análisis Canónico de Correspondencias de dos vías

Cuando se dispone de una matriz de abundancias en una única ocasión el análisis de la misma suele llevarse a cabo utilizando el Análisis de Correspondencias. Si además se dispone de información ambiental adicional y se desea relacionar la composición natural con las características ambientales, la técnica utilizada es el Análisis Canónico de Correspondencias (ACC) [5]. Desde un punto de vista teórico más general, el ACC puede considerarse como un Análisis de Componentes Principales Restringido [4], en el que los ejes de ordenación de las especies se restringen para que sean combinaciones lineales de las características ambientales.

3. Análisis Canónico de Correspondencias de tres vías: Metodología

Cuando la matriz de abundancias está replicada en el tiempo, el análisis de la misma puede realizarse mediante el Análisis de Componentes Principales (Correspondencias) de tres vías [2]; sin embargo, cuando se dispone de información

ambiental adicional para explicar el comportamiento de la composición natural, sólo se ha encontrado en la literatura el denominado STATIS Canónico [6], que es útil para una única variable ambiental cualitativa.

En este trabajo se propone una generalización del ACC a datos de tres vías basada en la generalización, a tres vías, del Análisis de Componentes Principales Restringido [1].

4. Análisis Canónico de Correspondencias de tres vías: Aplicación

Los resultados teóricos se aplicarán a datos procedentes del proyecto "Shallow Wetland Lake Function and Restoration in a Changing European Environment (SWALE)" financiado por la Unión Europea , en el que se estudia la evolución natural de un lago mediterráneo cuando se simulan, mediante experimentación, las condiciones de intervención humana sobre el mismo. Los detalles del experimento pueden consultarse en [3].

5. Bibliografía

- [1] CORTES-SAUD, A. (2005). *Contribuciones al análisis de tables de tres vías restringido*. Tesis Doctoral. Universidad de Salamanca.
- [2] KROONENBERG, P.M. and De LEEUW, J. (1980). 'Principal component analysis of three-mode data by means of alternating least squares algorithms'. *Psychometrika* **45**, 69-97.
- [3] STEPHEN, D., BALAYLA, D., BÉCARES, et al. (2004). Continental-scale patterns of nutrient and fish effects on shallow wetland lakes: introduction to a pan-European mesocosm experiment. *Freshwater Biology* **49**: 1517-1524.
- [4] TAKANE, Y. and HUNTER, M.A. (2001). 'Constrained Principal Component Analysis: A Comprehensive Theory'. *Applicable Algebra in Engineering, Communication and Computing*, AAECC **12**, 391-419.
- [5] Ter BRAAK, C.J.F. (1986). 'Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis'. *Ecology* **67**, 1167-1179.
- [6] VALLEJO, A., VICENTE-VILLARDÓN, J. L. & GALINDO-VILLARDÓN, M. P. (2006) Canonical STATIS: Biplot analysis of multi-table group structured data based on STATIS-ACT methodology. *Computacional Statistics and Data Análisis*. (en prensa).

Análisis comparativo de distintas técnicas de apoyo a la interpretación de clusters sobre un caso real

Karina Gibert, Gustavo Rodríguez, Gastón García

karina.gibert@upc.edu, Departamento de Estadística e Investigación

Operativa, Universitat Politècnica de Catalunya

Campus Nord, Ed. C5. C. Jordi Girona 1-3, Barcelona 08034

Una de las técnicas más empleadas dentro del ámbito no supervisado es el Cluster Analysis. El objetivo es agrupar elementos en clusters o grupos de manera que los elementos dentro de un mismo grupo sean lo más parecidos posible, mientras que elementos de distintos grupos sean lo más diferentes posible.

Al ser el Cluster Analysis una técnica no supervisada, los grupos (clases) no se conocen de antemano, y por lo tanto una de las cuestiones más delicadas es decidir el número de conglomerados a considerar. Sin embargo, muchas veces la problemática no termina una vez decidido el número final de grupos a retener, que en el caso del cluster jerárquico se puede decidir a posteriori, ya que es entonces cuando hace falta saber si las clases que se proponen se corresponden o no con la realidad así como obtener una descripción y/o interpretación de las clases formadas que permita comprender el significado de dichos clusters y su utilización posterior, ya sea con fines meramente descriptivos o predictivos.

El problema de la validación en el contexto del cluster es que nos enfrentamos a un dominio no supervisado, donde generalmente no se conoce la estructura real de los datos (puesto que de conocerse sería mucho más adecuado plantearse otro tipo de técnicas para su reconocimiento, como el análisis discriminante u otras). En este contexto, aparte de probar las diferencias significativas entre las clases, a menudo se toma la "interpretabilidad" de las clases como un elemento de validación fundamental para determinar si esa clasificación va a ser utilizable o no[2]. Sin embargo, en materia de descripción y/o interpretación de las clases formadas poco se ha hecho, dejándose en general en manos del analista que la hace de forma más o menos artesanal[4].

El nivel de detalle en la descripción de los clusters depende principalmente de la información aportada sobre los objetos. Según [4] cada clase puede ser descrita por k (≥ 1) objetos representativos (*centroídes* o *medoides* para los cuales la suma de sus disimilitardades con los objetos de su cluster es mínima) y por una medida de dispersión. Los gráficos de disimilitud entre cada objeto y cada objeto representativo del cluster y otras medidas de pertenencia pueden ayudar a distinguir los miembros de un cluster. Este tipo de información constituye una herramienta para que el analista proceda personalmente a interpretar los clusters. Pero cuando el número de clases y variables es grande, conviene apoyarse en herramientas más potentes. Así, en la literatura se

encuentran distintas posibilidades[7]:

1. Derivar reglas para asignar objetos a aquella clase con la que tienen más semejanza.
2. Establecer para cada clase, un conjunto de propiedades que sean satisfechas por al menos un $\alpha\%$ de los miembros de la clase.
3. Establecer para cada clase, un conjunto de propiedades que sean satisfechas por al menos un $\alpha\%$ de sus miembros pero no más de un $\beta\%$ para los miembros de otras clases.

Dentro de la categoría 1, el conjunto de métodos conocidos como árboles de decisión (Breiman et al., 1984) aporta una descripción (mezquina) de las propiedades de cada clase. A su vez, las técnicas dentro de las categorías 2 y 3 provienen del campo del Aprendizaje Automático, Clustering Conceptual y KDD-Minería de Datos (e.g., Michalski et al., 1983; Fisher, 1987; Ho et al., 1988; Fayyad et al., 1996).

En este trabajo se presenta la aplicación de algunas de estas técnicas (*la Regresión Logística, los Árboles de Decisión y el Análisis Discriminante*) a un conjunto de datos real previamente clasificado[6] con el objetivo de identificar las situaciones características que se pueden dar en una planta de depuración de aguas residuales. Ello pone de manifiesto, por un lado, cuáles resultan herramientas auxiliares más o menos útiles para proceder a la interpretación de las clases y, por otro, el nivel de concordancia entre los resultados de unas y otras.

1. Bibliografía

- [1] S. Backer: *Unsupervised Pattern Recognition: Dimensionality Reduction and Classification*, University of Antwerp, (2002)
- [2] K. Gibert. *Técnicas Híbridas de Inteligencia Artificial y Estadística para el descubrimiento de conocimiento y la Minería de Datos*, Eds. R. Giráldez, J.C. Riquelme, J.S. Aguilar, 2002.
- [3] Gibert, K. and Roda, I. (2000.) Identifying characteristic situations in wastewater treatment plants. *Workshop BEASAI (ECAI2000)*, V1, 1–9.
- [4] A.D. Gordon: *Cluster Description*, University of St. Andrews Scotland,

Análisis comparativo provincial de los movimientos migratorios de la Comunidad de Castilla y León: Proyección de las inmigraciones

P. Galindo, C. Ávila, I. Barrera, A. Blázquez-Zaballos, M.J. Fernández-Gómez, J. Martín-Vallejo, P. Vicente, J.L. Vicente-Villardón

Departamento de Estadística, Universidad de Salamanca

Abstract

Se ha realizado un estudio de los movimientos migratorios interiores y exteriores que se producen en las nueve provincias de Castilla y León. Además, y dada la importancia que presentan los flujos de inmigrantes en los últimos años, se han calculado las proyecciones de la población inmigrante a partir de las poblaciones de la Renovación Patronal de 2003 con un horizonte temporal de 25 años.

Palabras Clave: Migraciones, proyecciones, Modelo Logístico

1. Introducción

La migración es uno de los principales componentes en el crecimiento o decrecimiento de una población y se considera que la inmigración puede resolver el debilitamiento del ritmo de crecimiento demográfico de los países desarrollados [1]. Generalmente, estos movimiento están asociados a cambios sociales y económicos [2]. En España, la incorporación de la población inmigrante viene incrementándose de forma paulatina durante los últimos años. Este considerable aumento de las cifras de inmigrantes en España y la diversidad de factores que producen dichos movimientos, suponen un grado de incertidumbre en los flujos de entradas y de salidas lo que hace necesario un análisis pormenorizado de estos movimientos. El crecimiento de inmigrantes es la única solución posible para resolver el preocupante problema del descenso en la natalidad en la comunidad autónoma de Castilla y León, la cual aunque presentó en 2002 un índice de natalidad de 7.12 nacimientos por cada 1000 habitantes, muy por debajo de la media nacional (10.03 nacimientos), ha alcanzado el nivel más alto desde 1993. Según el INE este aumento se debe en exclusiva a los partos de madres extranjeras, mientras que en las españolas sigue descendiendo. Desde 1996 se ha triplicado el número de nacimientos del colectivo inmigrante en España.

El objetivo de este trabajo es analizar los movimientos migratorios que se producen en la Comunidad de Castilla y León con una especial atención a la inmigraciones exteriores que sin duda son las mas influyentes en el cambio de la estructura poblacional de esta comunidad.

2. Metodología

Los datos de migración utilizados se han obtenido y/o elaborados a partir de la información proporcionada por el Instituto Nacional de Estadística (INE).

El periodo del estudio está comprendido entre 1999 y 2003. La población de inmigrantes exteriores se han obtenido por grupos quinquenales hasta los 75 años y por sexo. Se calculan la Tasas de migración brutas. Las proyección de la componente de la inmigración se ha llevado a cabo a partir de la composición de la población por edad y sexo y separadamente para cada una de las provincias. Se establece como año base el 2003 con un horizonte temporal de 25 años. La proyección de las migraciones se ha hecho según un modelo logístico.

3. Resultados

El estudio evolutivo comparativo provincial de las inmigraciones internacionales nos indica que Segovia, Soria y Burgos son las provincias que han experimentado el incremento mayor en los últimos años del estudio. León presenta un patrón más homogéneo mientras que Palencia y Zamora no experimentan un ascenso tan pronunciado. La proyecciones de las inmigraciones en el escenario más optimista se observa un crecimiento importante en el número de inmigrantes hasta el año 2010 para las provincias de Ávila, Burgos, León y Salamanca. El resto de las provincias presentan un incremento menos acentuado.

4. Contactar

Javier Martín-Vallejo

Dirección:

Departamento de Estadística
C/Espejo nº2
37007 -Salamanca (Salamanca)
Spain

Teléfono: (+34) 923288118 ext:1852 **Fax:** (+34) 923294619

e-mail: jmv@usal.es

5. Agradecimientos

Este trabajo ha sido financiado por el proyecto de investigación OTP/05/04 Subvencionado por la Consejería de Hacienda de la Junta de Castilla y León.

6. Bibliografía

- [1] León, B. (2004). *La inmigración: ¿una opción ante el envejecimiento de la población?*. *Estudios de Economía Española*. .182. Fedea. Madrid
- [2] Rowland, D.J. (2003). *Demographic methods and concepts*. Oxford University Press.

Análisis de datos composicionales con el CoDa-dendrograma

V. Pawlowsky-Glahn¹, J.J. Egozcue²

¹vera.pawlowsky@udg.es, Dep. de Informática y Matemática Aplicada,
Universidad de Girona, España.

²juan.jose.egozcue@upc.edu, Dep. de Matemática Aplicada III, Universidad
Politécnica de Cataluña. España.

Abstract

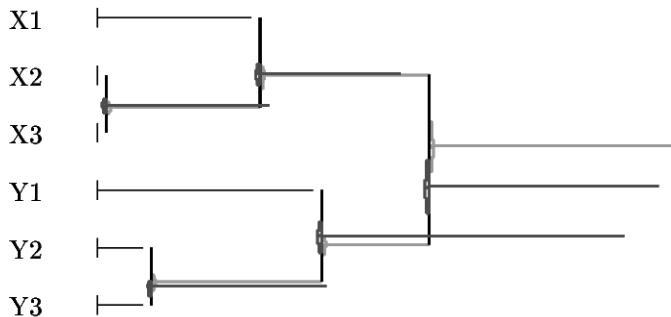
El espacio muestral de los datos composicionales es un espacio Euclídeo. La construcción de bases ortonormales a partir de particiones secuenciales binarias del vector composicional sugiere la representación conjunta de la partición y de medidas características de las coordenadas en un gráfico que resulta ser sumamente útil para el análisis exploratorio.

Claves: balances, geometría de Aitchison, simplex.

1. Composición, partición secuencial binaria y balances

Una composición es un vector aleatorio, cuyo espacio muestral es el simplex, $\mathcal{S}^D = \{\mathbf{x} = [x_1, x_2, \dots, x_D] \in \Re^D; x_i > 0, 1 \leq i \leq D; \sum_{i=1}^D x_i = \kappa\}$. La constante κ refleja las unidades de medida (tantos por uno, porcentajes, ppm, ...). \mathcal{S}^D admite estructura de espacio Euclídeo [2], lo que permite trabajar en coordenadas respecto a una base ortonormal. De las diversas estrategias para construir tales bases, elegimos una basada en una partición secuencial binaria del vector composicional \mathbf{x} [1], pues sugiere una representación en forma de dendrograma sumamente útil para el análisis exploratorio de datos composicionales. Para ilustrar la estrategia, consideremos el siguiente problema: un médico precisa distinguir entre dos tipos de cierta enfermedad, y ha determinado la composición de un conjunto de extractos sanguíneos, divisible en dos grupos, $[x_1, x_2, x_3]$ e $[y_1, y_2, y_3]$, como posible fuente discriminante. Dispone de 25 muestras para cada tipo de enfermedad, y se pregunta si los extractos tienen algún valor diagnóstico y, en caso afirmativo, como debiera utilizarse la información composicional para realizar un diagnóstico de nuevos pacientes. Una posible partición secuencial binaria del vector que separe primero los dos conjuntos de extractos es $x_1 ||| x_2 || x_3 ||| y_1 || y_2 | y_3$, donde el número decreciente de barras indica el orden creciente de la partición. Las coordenadas, o balances, correspondientes a la base ortonormal asociada son $b_1 = \sqrt{3/2} \ln [g(x_1 x_2 x_3)/g(y_1 y_2 y_3)]$, $b_2 = \sqrt{2/3} \ln [x_1/g(x_2 x_3)]$, $b_3 = \sqrt{2/3} \ln [y_1/g(y_2 y_3)]$, $b_4 = \sqrt{1/2} \ln [x_2/x_3]$, y $b_5 = \sqrt{1/2} \ln [y_2/y_3]$, siendo $g(\cdot)$ la media geométrica [1]. Para una muestra, la varianza total se descompone en suma de varianzas de los b_i , y estos pueden describirse mediante elementos estándar (media, quantiles).

2. CoDa-dendrograma



La figura muestra conjuntamente la partición, y los valores y medidas características de los balances. Por analogía con un dendrograma habitual es fácil reconocer: (a) la agrupación de partes debida a la partición; (b) la descomposición de la varianza por balances que determina la longitud de los segmentos horizontales; (d) la media de cada balance y sus quantiles reflejados por las uniones y las cajas respectivamente. Vemos que para los datos del médico los dos grupos de extractos discriminan muy bien (medias en b_1 muy distintas, varianzas similares, poca superposición); b_3 es poco informativo a pesar de la separación de medias por la alta variabilidad en uno de los dos tipos de enfermedad; y los demás balances apenas aportan información discriminante.

3. Agradecimientos

Esta investigación está financiada por la Dirección General de Enseñanza Superior del Ministerio de Educación y Cultura (BFM2003-05640/MATE).

4. Bibliografía

- [1] Egozcue, J. J. and V. Pawlowsky-Glahn (2005). Groups of parts and their balances in compositional data analysis. *Mathematical Geology* 37(7), 795–828.
- [2] Pawlowsky-Glahn, V. and J. J. Egozcue (2001). Geometric approach to statistical analysis on the simplex. *Stochastic Environmental Research and Risk Assessment (SERRA)* 15(5), 384–398.

Análisis de influencia conjunta en los subespacios generados por componentes principales

A. Enguix-González¹, J.L. Moreno-Rebollo², J.M. Muñoz-Pichardo³

¹aenguix@us.es, ²jlmoreno@us.es, ³juanm@us.es,

Departamento de Estadística e Investigación Operativa, Universidad de Sevilla

Abstract

En este trabajo se estudia la influencia conjunta ejercida por un conjunto de observaciones sobre el subespacio generado por subconjuntos de autovectores de la matriz de covarianzas. Para ello se utilizarán las versiones muestrales de la función de influencia y se propondrá el uso de aproximaciones para evitar distintos problemas prácticos.

Claves: Análisis de influencia conjunta, subespacios generados, Análisis de Componentes Principales.

AMS: Clasificación AMS: 62H25

1. Descripción del trabajo

Antes de utilizar una técnica estadística para estudiar un conjunto de datos, es conveniente realizar un análisis previo de la influencia que las observaciones obtenidas pueden ejercer en los estadísticos involucrados en tal técnica.

En la literatura existen distintos trabajos que abordan el Análisis de Influencia en el Análisis de Componentes Principales. En ellos, los estadísticos sobre los que se ha estudiado la influencia han sido, principalmente, los autovalores y los autovectores de la matriz de covarianzas muestrales. La mayoría de estos trabajos se han basado en las funciones de influencia y se han desarrollado desde el punto de vista de la influencia individual. Existen también algunos artículos publicados en los que se aborda la influencia sobre subespacios generados por las componentes principales, como Bénasseni (1990) o Tanaka (1988), de nuevo, desde el punto de vista de la influencia individual.

El Análisis de Influencia en autovalores y autovectores a través de las versiones muestrales de la función de influencia, presenta una serie de dificultades prácticas que conviene afrontar. Estas dificultades surgen al tener que comparar o relacionar autovalores y autovectores antes y después de eliminar observaciones: posible cambio en el orden de los autovalores y la determinación del sentido de los autovectores para la comparación. También es de señalar el elevado número de problemas de autovalores que es necesario resolver, suponiendo un grave problema computacional. Para evitar las dificultades prácticas señaladas, Enguix-González *et al.* (2005) propusieron el uso de aproximaciones de

los autovalores y autovectores muestrales calculados tras la omisión de observaciones mediante desarrollos en serie que se pueden obtener resolviendo un único problema de autovalores, el que se tiene con la muestra completa. Esto se abordó desde el punto de vista de la influencia conjunta.

En este trabajo se proponen medidas de la influencia que ejerce un conjunto de observaciones en el subespacio generado por un subconjunto de autovectores. Las dificultades prácticas que se tienen para los autovectores se extienden también a estos subespacios. Por ello, en este trabajo proponemos, de nuevo, el uso de aproximaciones basadas en desarrollos en serie para su cálculo. Para ello, se generalizarán las aproximaciones de la función de influencia muestral del subespacio generado por un subconjunto de autovectores obtenidas por Tanaka (1988), para en Análisis de Influencia Conjunto. Además, como las funciones de influencia (y sus versiones muestrales) tienen la dimensión del parámetro (o estadístico) estudiado, para obtener diagnósticos de influencia se calculará su norma de forma que las observaciones se podrán ordenar de forma lógica.

Finalmente, se darán ejemplos ilustrativos para mostrar el uso y la eficacia de las medidas de influencia propuestas.

2. Agradecimientos

Este trabajo está subvencionado por el Ministerio de Educación y Ciencia de España (MTM2004-01433).

3. Bibliografía

- [1] Bénasseni, J. (1990). Sensitivity Coefficients for the Subspaces Spanned by Principal Components. *Commun. Statist.-Theory Meth.*, 19, 2021-2034.
- [2] Enguix-González, A; Muñoz-Pichardo, J.M.; Moreno-Rebollo, J.L.; Pino-Mejías, R. (2005). Influence Analysis in principal components through power-series expansions. *Communications in Statistics. Theory and Methods*, 34, 2025-2046.
- [3] Pack, P.; Jolliffe, I.T; Morgan, B.J.T. (1988). Influential Observations in Principal Component Analysis: a Case Study. *J. Appl. Statist.*, 15, 39-52.
- [4] Tanaka, Y. (1988). Sensitivity Analysis in Principal Component Analysis: Influence on the Subspace Spanned by Principal Components. *Commun. Statist.-Theory Meth.*, 17, 3157-3175.

Análisis de la eficiencia en el muestreo estratificado. Aplicación al muestreo de la base de datos del impuesto sobre la renta de las personas físicas de la agencia tributaria

César Pérez López¹, Santiago Díaz de Sarralde²

¹cesar.perez@ief.minhac.es, Departamento de Estadística e Investigación Operativa III, Universidad Complutense de Madrid e Instituto de Estudios Fiscales

²santiago.diaz@ief.minhac.es, Instituto de Estudios Fiscales y Universidad Rey Juan Carlos de Madrid

Abstract

El presente trabajo tiene por objeto realizar un análisis exhaustivo de la eficiencia de los distintos tipos de afijación en el muestreo estratificado. Como aplicación práctica se presenta el diseño muestral realizado para muestrear la base de datos del Impuesto sobre la Renta de las Personas Físicas del Departamento de Informática Tributaria de la Agencia Tributaria.

Palabras Clave: Muestreo, Afijación, Simulación

1. Contenido

El trabajo comienza analizando la ganancia en precisión obtenida al considerar las distintas alternativas de afijación muestral en el muestreo estratificado y la ganancia en precisión de todas ellas respecto de la alternativa del muestreo aleatorio simple.

Se tendrá en cuenta la selección sin reposición y la selección con reposición contrastando también las eficiencias de las afijaciones en ambos casos. Asimismo se cuantificarán los errores de las estimaciones habituales estrictamente para cada tipo de afijación y tipo de selección y se analizará su incidencia en la determinación del tamaño muestral necesario para cometer un error de muestreo dado.

Como aplicación práctica se presenta el diseño muestral realizado para llevar a cabo el muestreo de la base de datos del Impuesto sobre la Renta de las Personas Físicas de la Agencia Tributaria.

Se analizan las variables de estratificación y se justifica la utilización de un muestreo estratificado con afijación de mínima varianza de acuerdo al marco teórico anterior.

A continuación se analiza la calidad de esta muestra presentándose los principales resultados de la misma.

Adicionalmente se describe cómo se utiliza esta muestra para la construcción de un simulador fiable del Impuesto sobre la Renta de las Personas Físicas

Finalmente se hace hincapié en la importancia práctica de este diseño muestral derivada del uso de la muestra y del simulador basado en ella para tomar decisiones de política fiscal en el marco de la Administración del Estado y las Comunidades Autónomas.

Análisis de la persistencia del IBEX-35

Beatriz Rodríguez Flores

brodriguez003@ehu.es, Departamento de Economía Aplicada III,
Universidad del País Vasco (UPV-EHU)

Abstract

El propósito de este trabajo es estudiar la naturaleza de la dependencia temporal en la volatilidad de los activos del IBEX-35. Asimismo, se analiza la naturaleza de la dependencia temporal en el volumen de negocio y la posible relación en el comportamiento de ambas variables.

Claves: Memoria larga; volatilidad estocástica.

1. Motivación y resumen de las aportaciones

El estudio de la dependencia temporal en distintas series financieras ha acaparado la atención de muchos investigadores puesto que tiene implicaciones sobre la valoración de un activo y permite además confirmar o refutar la hipótesis de eficiencia de los mercados.

Es un hecho característico de muchas series financieras que, si bien la rentabilidad de un activo presenta una escasa estructura dinámica, las autocorrelaciones muestrales para el logaritmo del cuadrado de dicha serie, u otras potencias del valor absoluto son significativamente distintas de cero. Este hecho confirma la presencia de una estructura dinámica en la volatilidad. En este contexto, el propósito de este trabajo es estudiar (contrastar y estimar) la naturaleza de la dependencia temporal en la volatilidad de los activos que componen el IBEX-35. Un análisis similar se aplica al volumen de negocio con el propósito de analizar la posible relación en el comportamiento de ambas variables.

Al objeto de modelizar la heterocedasticidad condicional tradicionalmente utilizada para recoger la estructura dinámica de la volatilidad, en este trabajo se emplea el denominado Modelo de Volatilidad Estocástica de Memoria Larga propuesto por Breidt et al. (1998) y Harvey (1998), que supone que la volatilidad depende de una variable estocástica no observable.

Los contrastes de dependencia temporal aplicados en este trabajo no requieren sin embargo la estimación del modelo LMSV. En este trabajo se propone el uso combinado de los contrastes de raíz unitaria de Perron y Ng (1996) con un contraste semiparamétrico de memoria corta propuesto por Lobato y Robinson (1998). La combinación de los resultados de ambos contrastes indica que tanto la volatilidad como el volumen no quedan bien modelizados ni a través de procesos integrados de orden uno ni con procesos con dependencia débil, lo cual, puede ser un signo de presencia de memoria larga.

Una vez detectada la memoria larga, es preciso su estimación a efectos inferenciales y predictivos. En este trabajo estimamos el parámetro de memoria a través de dos métodos semiparamétricos; el método de la regresión GPH (Geweke y Porter-Hudak, 1983), y el método local de Whittle (Robinson, 1995). Ambos métodos dan lugar a estimaciones similares del parámetro de memoria. En todo caso, los contrastes realizados indican que el parámetro de dependencia temporal es significativamente distinto de cero y menor que uno. Además, el parámetro de memoria estimado es siempre menor que 1 y, en casi todos los casos, menor que 1/2 (y mayor que 0), por lo que las series de volatilidades y volúmenes del IBEX-35 manifiestan una dependencia temporal de memoria larga pero estacionaria, es decir; el efecto de un *shock* se transmite a lo largo de un período de tiempo elevado pero no es permanente, sino que tiende a diluirse con el transcurso del tiempo.

Por último, en este trabajo se investiga de forma empírica la interrelación entre los movimientos de precio y volumen en el mercado bursátil, examinando si ambos procesos están conducidos por el mismo componente de memoria larga, esto es, si el volumen y la volatilidad están (fraccionalmente) cointegrados. De los resultados obtenidos se deriva que, a pesar de que la volatilidad y el volumen de algunos activos del IBEX-35 parecen compartir parámetro de memoria, la hipótesis de cointegración se rechaza en todos los casos analizados.

2. Bibliografía

- [1] Breidt, F.J., Crato, N. y Lima, P. (1998). The detection and estimation of long memory in stochastic volatility, *Journal of Econometrics* 83, 325-348.
- [2] Geweke, J. y Porter-Hudak, S. (1983). The estimation and application of long memory time series models, *Journal of Time Series Analysis* 4, 221-237.
- [3] Harvey, A.C. (1998). Long Memory in stochastic volatility. In: Oxford, Butterworth-Haineman (eds.), *Forecasting Volatility in Financial Markets*, 307-320.
- [4] Lobato, I.N. y Robinson, P.M. (1998). A nonparametric test for I(0), *Review of Economic Studies* 65, 475-495.
- [5] Perron, P. y Ng, S. (1996). Useful modifications to some unit root tests with dependent errors and their local asymptotic properties, *Review of Economic Studies* 63, 435-463.
- [6] Robinson, P.M. (1995). Gaussian semiparametric estimation of long range dependence, *The Annals of Statistics* 23, 1630-1661.

Análisis del Comportamiento del RV (STATIS) y el Lg (AFM), ante cambios Alpha, Beta y Gamma

P. Vicente¹, N. Bacala², P. Galindo³, M. J. Fernández-Gómez⁴

¹purivic@yahoo.com, ³pgalindo@usal.es, ⁴mjfg@usal.es

Departamento de Estadística, Universidad de Salamanca, España

²norab@crub.uncoma.edu.ar, Centro Regional Universitario Bariloche,
Universidad Nacional del Comahue, Argentina

Abstract

En este trabajo se analiza el comportamiento de los métodos STATIS y AFM ante la presencia de cambios Alfa, Beta y Gamma y se pone de manifiesto que ni el RV utilizado en el método STATIS, ni el Lg utilizado en el AFM, son útiles a la hora de definir qué tipo de cambio se produce entre una estructura y otra.

Palabras Clave: AFM, STATIS. Cambio

1. Introducción

Las medidas de relación entre dos tablas o grupos de variables t y t' que utilizan ambos métodos es el producto escalar entre las configuraciones normadas t y t' .

En el *Statis*, si se trabaja con configuraciones normadas, dicho producto escalar es el coeficiente RV; en el *AFM* es el coeficiente L_g . Ambos coeficientes tratan de medir el grado de semejanza entre dos estructuras, o dos grupos de variables, observadas sobre el mismo conjunto de individuos, es decir, evalúan si se han producido *cambios* entre las estructuras en ambas tablas.

2. Tipos de Cambio

El concepto de cambio es muy complejo y es entendido de forma muy diferente por distintos autores. Considerando los conceptos de MORTIMER et al. en 1982 y GOLEMBIEWSKI et al., en 1976, diremos que dos grupos de variables t y t' presentan **Invarianza estructural** si poseen: el mismo **número de dimensiones** y el mismo **patrón** en los factores de carga. En caso contrario diremos que se ha producido un *cambio estructural*, o *cambio gamma* según la tipología del cambio introducida por Golembiewski et al.

Si existe Invarianza estructural, es posible evaluar si los resultados son estables y/o consistentes, o por el contrario se ha producido algún otro tipo de cambio, en el estudio. La *Estabilidad* puede ser *Normativa* si se refiere a la estabilidad de los rangos de orden de las preferencias sobre un atributo de interés, a lo largo del tiempo, o *Cuantitativa* si se refiere a la estabilidad en la magnitud o cantidad media de preferencia de un atributo de interés a lo largo del tiempo. El

cambio convencional, es justo el concepto opuesto a la estabilidad cuantitativa, y que fue llamado *Cambio Alpha* por Golembieswki.

GOLEMBIESWKI et al (op. ct.) consideran otro tipo de cambio, *Cambio Beta*, el que se puede producir cuando hay cambios en la escala de medida y/o en el punto de referencia (recalibración).

3. Análisis del Comportamiento del RV y el Lg ante la presencia de Invarianza estructural

La ocurrencia de un *cambio Gamma* presupone un cambio en las principales direcciones de inercia de ambos grupos, por tanto, *el RV y el Lg decrecen a medida que crece el número y la importancia respecto a la principal dirección de inercia, de las dimensiones no comunes*. Cuando se presenta un *cambio Beta* alguna de las variables contributivas, relativas al mismo aspecto del constructo, tendrán distinta importancia en los dos grupos; es decir los valores de los factores de carga pueden cambiar y/o la variabilidad de las dimensiones latentes. Si este tipo de cambio ocurre en una estructura unidimensional, puede suceder que la/s variable/s contributiva/s en la dimensión, modifiquen el valor de su carga o importancia, de un grupo a otro, aunque la dimensión mantenga su importancia común a ambos grupos; el RV y el Lg *decrecen* porque no consideran esa dimensión como común.

En este trabajo se demuestra que: el RV y el Lg no son útiles a la hora de definir qué tipo de cambio se produce entre una estructura y otra. Ahora bien, cuanto más próximo a 1 sea el RV mayor evidencia de que no se han producido cambios en la dirección de máxima inercia en las dos estructuras. El valor del Lg estará indicando el número de dimensiones que comparten, o son comunes, a ambos grupos, pero valores bajos del Lg no son condición suficiente para caracterizar ninguno de los tipos de cambio.

4. Bibliografía

- [1] Escofier, B. et Pagès, J. (1984): 'Analyse factorielle multiple'. *Cahiers du BURG*, **2**. ISUP. Paris.
- [2] Golembiewski, R. T.; Billingsley, K. & Yeager, S. (1976): 'Measuring change and persistence in human affairs: types of changes generated by OD designs'. *Journal of Applied Behavioural Science*, **12**: 133-157
- [3] L'Hermier des Plantes, H.. (1976). *Structuration des tableaux à trois indices de la Statistique*. Thèse de Doctorat. Université de Montpellier.
- [4] Mortimer, J. D.; Finch, M. D. & Kumka, D. (1982): 'Persistence and change in development: the multidimensional self-concept'. In: P. B. Baltes & O. G. Brim (eds.). *Life-Span Development and Behavior*, **4**: 263-309. Academic Press.

Análisis espacio-temporal de la Fecundidad en la Comunidad de Castilla y León

*M.P. Galindo, C.A. Ávila-Zarza, I. Barrera, A. Báñez-Zaballos,
M.J. Fernández-Gómez, J. Martín-Vallejo, M.P. Vicente-Galindo,
J.L. Vicente-Villardón*

pgalindo@usal.es, Departamento de Estadística, Universidad de Salamanca

Abstract

La fecundidad en España está a la cola de los países de nuestro entorno, lejos del 1.53 de la media de la UE y más aún de los valores de los países líderes en Europa (Irlanda y Francia). La Comunidad de Castilla y León presenta los valores más bajos; la tasa de natalidad ha descendido a la mitad en los últimos 25 años. En este trabajo se analiza la situación actual de las provincias de la Comunidad de Castilla y León utilizando métodos Biplot y el método STATIS.

Palabras Clave: Tasa específica de fecundidad, BIPLOT, STATIS

1. Introducción

El estudio y conocimiento de nuestra población ha de ser un requisito imprescindible si se quiere planificar, de manera racional, en los diversos campos de la vida social: consumo, educación, sanidad, etc.

El envejecimiento de nuestras poblaciones es un hecho conocido, achacable en parte, al descenso de las tasas de fecundidad. El número y variedad de factores determinantes de este descenso hacen que el cálculo de nacimientos futuros se convierta en una tarea difícil que requiere de dos tipos de análisis del fenómeno: Por un lado conviene tener una caracterización actual de la fecundidad (en el área de interés: Comunidad, provincia, municipio, etc...), y por otra parte, analizar retrospectivamente este fenómeno.

Para llevar a cabo el estudio nos hemos centrado en la comunidad Castellano-Leonesa, una de las más deprimidas de la geografía española. El estudio retrospectivo se ha llevado a cabo desde 1975.

2. Material y Métodos

Se ha considerado, para cada año del estudio, la matriz que contiene las tasas específicas de fecundidad, en cada una de las 9 provincias de la Comunidad de Castilla y León (en filas), y en columnas las edades, en grupos quinquenales, de las mujeres que pueden tener hijos (desde los 15 hasta los 49 años de edad).

Para abordar el primer objetivo, consistente en caracterizar el estado actual de la fecundidad en la Comunidad de Castilla y León, hemos elegido los Métodos Biplot, ya que se trata de una técnica factorial que permite reducir la

dimensionalidad del problema para facilitar su interpretación, y representa simultáneamente las filas y las columnas de la matriz de datos. [1],[2],[3].

Una vez caracterizado el estado actual de la problemática que nos ocupa, resulta interesante realizar un estudio retrospectivo para ver la evolución que este fenómeno ha tenido en las últimas décadas, información fundamental con fines predictivos.

Para conseguir este segundo objetivo del estudio, se ha utilizado el método STATIS [4]. Este método trata de estudiar la inter-estructura entre los distintos estudios y si ésta evidencia un comportamiento similar se define una estructura compromiso y se representan los diferentes estudios sobre dicha estructura compromiso para poder analizar las intra-estructuras a través de las trayectorias.

3. Contactar

Dirección:

M. P. GALINDO-VILLARDÓN
 Departamento de Estadística
 C/ Espejo, 2
 337007 – Salamanca
 Spain

Teléfono: (+34) 923294400 Ext. 1852 **Fax:** (+34) 923294619

e-mail: pgalindo@usal.es

4. Agradecimientos

Un primer avance de este trabajo fue financiado por un proyecto de investigación subvencionado por la Consejería de Hacienda de la Junta de Castilla y León en 2003. (Referencia: OTP/03/12)

5. Bibliografía

- [1] Gabriel, K.R (1971). 'The Biplot-graphic display of matrices with application to principal component analysis'. *Biometrika*, 58, 453-467.
- [2] Galindo, M.P: (1986). 'Una alternativa de representación simultánea: HJ-Biplot'. *Questio* 10(1),13-23
- [3] Galindo, M.P. y Cuadras, C.M. (1986). 'Una extensión del método Biplot y su relación con otras técnicas'. *Publicaciones de Bioestadística y Biomatemática*. Universidad de Barcelona. Nº 17.
- [4] L'Hermier des Plantes, H.. (1976). *Structuration des tableaux à trois indices de la Statistique*. Thèse de Doctorat. Université de Montpellier.

Análisis Estadístico de Habilidades Motoras para la Natación en Niños.

*R. Barea-Barrera, J.M. Fernández-Ponce, M.T. Gómez-Gómez,
A. Ortiz-Bustos, J.L. Pino-Mejías*

Departamento de Estadística e I.O.

Universidad de Sevilla

*rbarea@us.es, ferpon@us.es, terego@us.es,
jalberto.ortiz@gmail.com, jlpino@us.es*

Abstract

El objetivo de este estudio descriptivo es el de localizar las posibles variables, observables sin necesidad de instrumentos complicados, que determinan o influyen la propulsión en natación. Para ello se ha procedido a recoger una serie de medidas anatómicas determinadas genéticamente y se las relaciona con la velocidad propulsiva de los pies.

1. Planteamiento del Problema

Es muy común la aplicación de técnicas estadísticas descriptivas en el estudio de cualidades para alguna modalidad deportiva en niños pertenecientes a escuelas deportivas. Es bien conocido que uno de los deportes más completos es la natación. Por ello se requieren de test físicos motores para la evaluación de dichas cualidades en jóvenes promesas. Aunque existen en la actualidad una batería extensa de dichos tests, la gran mayoría de ellos requieren de unos recursos técnicos prácticamente inexistentes en los complejos deportivos municipales o públicos. De ahí la necesidad de proponer un test físico de fácil y cómoda aplicación que además esté avalado por técnicas estadísticas.

La prueba consiste en una evaluación de la propulsión-deslizamiento. Para ello se estableció una distancia de cinco metros (medida con cinta métrica de construcción) delimitada por material de piscina (pull-boys). Esta delimitación se realizó en el borde lateral de la piscina a 4 metros del principio de la misma. Estos cuatro metros de separación están previstos para eliminar de algún modo la velocidad adquirida por el impulso en la pared. La toma de tiempo se hizo de forma manual con cronómetro, el cual se ponía en marcha justo cuando la tabla rebasaba la primera señalización (principio de los cinco metros) y se paraba cuando ésta se llegaba a la última.

La muestra ha sido formada por los niños y niñas de tres grupos de competición de tres pueblos distintos de la provincia de Sevilla que comparten la misma piscina para entrenar.

El objetivo del estudio es relacionar cada una de las variables con velocidades de propulsión y de deslizamiento y caracterizando el fenotipo más relacionado

con tales variables. También se persigue la posibilidad de demostrar la bondad de la prueba de propulsión (en el sentido de detección parcial de posible talento natatorio) así como la posible influencia de las medidas anatómicas en dichas velocidades.

Claves: Análisis Discriminante; MANOVA; Índice de Masa Corporal.

AMS: 62H30, 62P99

Dirección:

Departamento de Estadística e Investigación Operativa
Facultad de Matemáticas - Av. Reina Mercedes s/n
41012 - Sevilla
Spain

Teléfono: (+34) 954557927 **Fax:** (+34) 954622800

web <http://destio.us.es/>

2. Bibliografía

- [1] Gil-Flores, J; García-Jiménez, E. and Rodríguez-Gómez, G. (2001). Análisis Discriminante. *Editorial La Muralla*.
- [2] Geoffrey J. MacLachlan (1992). Discriminant Analysis and Statistical Pattern Recognition *John Wiley and Sons, INC*

Análisis estadístico sobre la inmigración ilegal en Melilla

*Juan Antonio Marmolejo Martín¹, Francisco Abad Montes²,
María Dolores Huete Morales³*

Departamento de Estadística e Investigación Operativa, Universidad de
Granada

¹jamarmo@ugr.es, ²fabad@ugr.es, ³mdhuete@ugr.es

Abstract

Tras los últimos acontecimientos en la Ciudad Autónoma de Melilla, efectuamos un análisis estadístico sobre la inmigración en dicha Ciudad con datos reales. En este análisis utilizamos técnicas demográficas y técnicas multivariantes, reflejando en esta comunicación los resultados obtenidos.

Palabras Clave: Técnicas multivariantes. Demografía

1. Introducción

A finales del año 2005 la Ciudad Autónoma de Melilla fue objeto de numerosos intentos de entrada ilegal por parte de inmigrantes de muy diversas nacionalidades. Tenemos contrastadas hasta 51 nacionalidades distintas que se pueden clasificar en tres grandes grupos inmigrantes subsaharianos, inmigrantes argelinos e inmigrantes de otras nacionalidades, generalmente asiáticos.

La inmigración ilegal utiliza diversas formas para acceder a la Ciudad, entre otras, la entrada por el perímetro fronterizo, el mar o la costa, ocultarse en vehículos e incluso utilizar documentación falsa.

El trabajo que hemos realizado comenzó, independientemente de los últimos acontecimientos, hace más de un año con la recogida de datos. Éstos se obtuvieron a través del Centro de Estancia Temporal de Inmigrantes de Melilla (C.E.T.I.), Cruz Roja Española, la Consejería de Bienestar Social y la Comandancia de la Guardia Civil de dicha ciudad. Los datos están divididos de la siguiente manera: Personas que han causado alta en el C.E.T.I., personas que han causado baja en el C.E.T.I. y personas que ocupan plaza al final de cada año en el C.E.T.I. En cada caso hemos considerado las variables sexo y grupo de edad. La variable “edad del inmigrante” está agrupada en intervalos: menor de 18 años, entre 18 y 34 años, entre 35 y 49 años y mayor de 50 años. Los datos de los que disponemos datan desde enero de 2002 hasta diciembre de 2005.

Este volumen de entradas de inmigrantes extranjeros cuyo acceso se produce de forma ilegal, ha generado significativos cambios demográficos en la Ciudad Autónoma de Melilla en los últimos años.

En el trabajo que aquí presentamos hemos realizado un cálculo resumen del número de entradas ilegales por sexo y por grupos de edad, que permita analizar de forma clara este fenómeno tan de actualidad en este momento. Para

visualizar la influencia de la variable edad en el número total de inmigraciones, se han construido las pirámides de edad para cada año de observación, sin distinción de la nacionalidad del inmigrante.

Por otra parte, se han obtenido las tasas de inmigración ilegal, tomando como población de referencia, la registrada en la Ciudad de Melilla. Esto nos ha permitido realizar una relativización del fenómeno, pudiendo incluso comparar estos indicadores con los registrados en la Comunidad de Andalucía para inmigración “legal”. Utilizando esta información también hemos obtenido un patrón de la inmigración ilegal, mediante la utilización de spline cúbicos.

En relación a los países de procedencia, se ha realizado un análisis de las entradas de ilegales, según la relevancia de dicho país de origen en lo que se refiere al volumen de personas que entran del mismo.

Por último, realizamos un análisis cluster para ver el grado de asociación entre las distintas variables.

2. Bibliografía

- [1] Bonard,N. (1989). Mouvements et modèles de population. Yaoundé, IFORD.
- [2] Boque, Arriage. (1993). Readings on population research methodology. Vol. 2 . U.N.F.P.A., SDC Chicago
- [3] Keyfitz, N. (1985). Applied mathematical demography. Springer Verlag, New York
- [4] Land, Kenneth C. Y Rogers, A. (1982). Multidimensional Mathematical Demography. Academic Press. New York.
- [5] McNeil, DR., Trusell, TJ, Turner, JC. (1977). Spline interpolation of demographics data. Demography, 14:245-252
- [6] Newel, C. (1994). Methods and models in demography. John Wiley and sons, Chicago.
- [7] United Nations. (1983). Indirect thecniques for demographic estimation. Manual X. Naciones Unidas. New York.
- [8] Vallin, J. (1995). La demografía. Alianza Universidad, Madrid.
- [9] Wunsch, G. (1984). Mèthodes d'analyse demnographique pour les pays en développement. Ed. Ordene Liège.

Análisis Factorial Múltiple y STATIS: una Aplicación a Datos de Afluencia Turística

Sara Morgado Nunes¹, Juan Moncada Herrera², M. Purificación Galindo³

¹sara@esg.ipcb.pt, Instituto Politécnico de Castelo Branco, Portugal

²jmoncada@uct.cl, Departamento de Cs. Matemáticas y Físicas, Universidad Católica de Temuco, Chile

³pgalindo@usal.es, Departamento de Estadística, Universidad de Salamanca, España

Abstract

El AFM y el STATIS son técnicas que permiten la exploración simultánea de varias tablas de datos, teniendo como base la búsqueda de una estructura común a los conjuntos de datos en análisis. En este trabajo se aplican las dos metodologías a datos de afluencia turística con el objetivo de compararlas, poniendo de manifiesto sus ventajas e inconvenientes.

Palabras Clave: Datos Múltiples, Análisis Factorial Múltiple y STATIS

AMS: 62H20, 62H25

1. Introducción

En un contexto de Turismo, surge muchas veces la necesidad de tratar e integrar información proveniente de diversas matrices de datos. Para tal efecto es necesario disponer de herramientas que posibiliten el análisis de conjuntos de datos múltiples. En general, en este tipo de datos, la información está organizada en un conjunto de matrices que se refieren a un conjunto de individuos sobre los cuales se observan variables en diversas ocasiones o situaciones experimentales.

La necesidad de explorar matrices de datos múltiples ha motivado el desarrollo de diversas técnicas entre las cuales destaca el Análisis Factorial Múltiple (AFM) y el STATIS. Ambos métodos posibilitan un análisis simultáneo de las diversas tablas en estudio, así como la obtención de una estructura compromiso capaz de sintetizar toda la información disponible.

En este trabajo se presenta una breve revisión de las metodologías AFM y STATIS, se presentan los resultados de la aplicación de las dos técnicas a datos de afluencia turística y se comparan los resultados obtenidos, evidenciando sus ventajas e inconvenientes.

2. El AFM y el STATIS

El AFM ha sido propuesto por Escofier & Pages (1985) con el objetivo de analizar la información contenida en K subtablas resultantes de la observación de K grupos de variables cuantitativas y/o cualitativas. En los grupos se pueden tener variables medidas en distintos instantes o variables que hayan sido

agrupadas según algún criterio. El objetivo es la obtención de un sistema de ejes común que permita representar simultáneamente las subtablas en estudio. La representación euclídea proporcionada por el AFM posibilita el análisis de la evolución de las posiciones a lo largo de las K subtablas, como también la comparación global de las K subtablas.

El método STATIS (“Structuration des Tableaux A Trois Índices de la Statistique”) ha sido propuesto por L’Hermier des Plantes (1976) y desarrollado por Lavit (1988), teniendo como objetivo principal la extracción de información relevante contenida en conjuntos de datos múltiples. El STATIS posibilita la exploración simultánea de T matrices de datos cuantitativos obtenidos en diferentes ocasiones sobre los mismos individuos (las variables pueden ser diferentes).

3. Aplicación a Datos de Afluencia Turística

Se dispone de un conjunto de datos relativos al número de turistas que han visitado diversas regiones portuguesas entre los años 1993 y 2001, según su proveniencia geográfica.

Se pretende estudiar la evolución registrada en términos de proveniencia geográfica de los visitantes y comparar globalmente las distintas regiones, durante el periodo en estudio. Se pretende de esta manera averiguar la existencia de una estructura común a las matrices de datos en análisis que posibilite la integración de la información disponible en una única estructura.

Los datos en estudio son sometidos al AFM y al STATIS y se comparan los resultados obtenidos.

4. Bibliografía

- [1] Escofier, B. & Pagès, J. (1985). *Mise en oeuvre de L'Analyse Factorielle Multiple pour des Tableaux Numériques, Qualitatif, ou Mixtes*. Rapports de Recherche N° 429. Centre de Rennes, IRISA. France.
- [2] Lavit, Ch. (1988). *Analyse conjointe de tableaux quantitatifs*. Masson, Paris.
- [3] L’Hermier des Plantes (1976). *Structuration des tableaux a trois indices de la statistique*. Thèse de 3 ème cycle, Université de Montpellier.
- [4] Pagès J. (1996). “Eléments de Comparaison entre L’Analyse Factorielle Multiple et la Méthode STATIS”. *Rev. Statistique Appliquée*, XLIV(4), 81-95.

Análisis Sensorial en el control y la mejora de procesos textiles. Caso de la Manta

Pau Miró i Martínez, José Miguel Carot Sierra, José Manuel Jabaloyes Vivas, Mónica Martínez Gómez

pamimar@eio.upv.es, Departamento de Estadística, Investigación Operativa Aplicadas y Calidad, Universidad Politécnica de Valencia

Abstract

El uso del Análisis Sensorial en procesos textiles es una práctica poco habitual, existiendo una gran falta de sistematización y de homogenización. En este trabajo se propone una metodología para la implantación de un sistema de control sensorial a un proceso textil y se expone el caso estudio realizado para el control de la manta en una empresa del sector.

Claves: Análisis Sensorial, procesos textiles.

1. Introducción

Se tiende a creer que la única manera de obtener información fiable sobre las características de un producto, para obtener resultados reproducibles y analizables estadísticamente, es aplicando métodos de ensayo científicamente exactos. En el sector textil el uso de los sentidos como método de evaluación de la calidad está muy extendido. En cambio en los procesos de producción aunque son habituales este tipo de evaluaciones no se están realizando siguiendo las recomendaciones habituales de las técnicas de Análisis Sensorial. Esta situación de no control estandarizado tiene el consiguiente peligro de la perdida de calidad por el aumento de la variabilidad del producto. Este trabajo quiere servir de herramienta para poder superar la problemática comentada, proponiendo una metodología para el seguimiento sensorial de productos textiles en procesos de producción de manera que se garantice la calidad sensorial del producto. La metodología propuesta se aplicó al caso de la manta para cama.

2. Objetivos

Los objetivos que se persiguen con la aplicación de la metodología de seguimiento sensorial para procesos de producción textil son: mejorar el proceso creando un sistema estandarizado de control basado en el análisis sensorial, facilitar el aprendizaje de los operarios noveles y optimizar las características sensoriales del producto. En consecuencia se quiere aumentar la fidelidad de proveedores y consumidores.

3. Metodología

El análisis sensorial en el proceso de producción difiere bastante del análisis

clásico puesto que hay determinados aspectos que no se pueden considerar tal y como se conciben normalmente, de ahí el carácter innovador de la propuesta.

Las fases que forman la metodología propuesta son:

1. Fase de Iniciación: se realizan las reuniones/entrevistas con los expertos.
2. Fase de Definición: se deben reconocer los puntos de inspección del proceso y escoger las muestras referencia de cada característica para consensuar las características, las referencias y los puntos de inspección a controlar.
3. Fase de Entrenamiento: se entrena a los operarios con la metodología y se realiza un periodo de pruebas.
4. Fase de Implementación: se establece el procedimiento definitivo y luego se incorpora al sistema de producción.
5. Fase de Control Continuo: se deben analizar los datos que genera el sistema para conseguir monitorizar el proceso.

4. Conclusiones

Las conclusiones a las que se llega mediante la aplicación de la metodología propuesta al caso de la manta son: se ha reducido la variabilidad del producto, se ha sistematizado el control sensorial y se homogeneizado la respuesta, se domina la producción sabiendo reaccionar en el corto plazo a desviaciones del proceso, se acortan los períodos de aprendizaje de los operarios y se deja de depender de la experiencia de unos pocos. Por lo tanto se ha conseguido perfeccionar el sistema de control y la calidad del producto.

5. Bibliografía

- [1] Philippe, F. et al. (2001). Développement d'une méthodologie d'analyse sensorielle tactile des textiles. *International Journal of Clothing* 15, 268-275.
- [2] Stone, H. and Sidel, J. L. (1992). *Sensory Evaluation Practices*. Academic Press Inc.
- [3] XP V 09-502 (2002). *Analyse sensorielle. Directives générales pour un suivi, par approche sensorielle, de la qualité d'un produit au cours de sa fabrication*. Association Française de Normalisation AFNOR.

Analysis of Principal Alarms in Statistical Quality Control

Isabel González¹, Ismael Sánchez²

¹Departamento de Ing. Mecánica, Universidad Carlos III de Madrid

²Departamento de Estadística, Universidad Carlos III de Madrid

Abstract

En este trabajo se propone una metodología para la simulación de situaciones fuera de control multivariantes utilizando datos bajo control. El método se basa en encontrar los factores independientes que producen la variabilidad del proceso. Estos factores se alteran uno a uno, y se calcula el cambio en términos de las variables observadas. Estos cambios se denominan por alarmas principales.

Keywords: Multivariate control charts; Principal alarms; Statistical process control

1. Introduction

The quality assessment of a production process is often attained by the joint monitoring of several correlated characteristics. It is of course possible to monitor the individual characteristics using univariate control charts. However, to the extent that the characteristics are correlated, a multivariate monitoring of the vector of characteristics can be more effective.

The use of multivariate control charts is, however, subject to practical drawbacks. The main problem is the interpretation, in terms of the original quality variables, of an out of control signal. The identification of the cause of the alarm can be a complex issue, since it can be related to a single variable or to a subset of variables. This identification problem has been addressed by a number of authors.

It is important to note that the available procedures are based on the analysis when the process is already out of control. Before the alarm, these procedures do not supply information about potential alarms. In this article, we propose a methodology that helps to predict the main shifts (both in mean or scale), denoted as principal alarms, that can be expected in the process using the available in-control data. The outcome of the procedure can be used to improve the quality of the process in many ways. First, it helps to identify the weakest elements of the process, which will be responsible of most of the variability. This result can lead to quality engineers to implement some modifications in the process. This capacity to improve the process is especially appealing in the initial stages of the production process, where the engineers need to adjust the machinery.

Second, it allows to improve the SPC activity because: (i) it allows to choose the best type of control chart for the process, and (ii) it eases the search for responsible variables when an alarm already takes place.

The method is based on finding the independent factors that explain the variability of the process. The principal component analysis is used to find these factors. Once these factors are estimated, a mean shift is introduced in them, one by one. The translation of these shifts, in terms of the original observed variables, is finally obtained. These shifts can be interpreted as a simulation of out-of-control situations provoked by each independent factor. The shifts provoked by the most important independent factors are called principal alarms.

These principal alarms are represented graphically, allowing the practitioner to visualize the main expected deviation of the process. The principal alarms are also used to select the best multivariate control chart. This comparison is made by estimating the ARL, of each competing chart, for shifts of different size. A resampling procedure for ARL estimation using principal alarms is proposed.

2. Contact

Address:

Ismael Sánchez
Department of Statistics, University Carlos III of Madrid
Av. Universidad 30
28911 - Leganés (Madrid)
Spain

Telephone: (+34) 916249179

Fax: (+34) 916249430

e-mail: ismael@est.econ.uc3m.es

Aplicación de wavelets para el modelado de la serie de precios de la energía eléctrica: un enfoque bayesiano

Elena Domínguez¹, José Mira², Carolina García-Martos³, María Jesús Sánchez⁴

¹edominguez@etsii.upm.es, Departamento de Matemática Aplicada a la Ingeniería Industrial

²jmira@etsii.upm.es, ³cgarciam@etsii.upm.es, ⁴mjsan@etsii.upm.es
Laboratorio de Estadística

Escuela Técnica Superior de Ingenieros Industriales, Universidad Politécnica de Madrid

Abstract

Este trabajo es una aplicación de las wavelets a la predicción de los precios del mercado eléctrico español. Se utiliza el enfoque bayesiano y las técnicas computacionales MCMC.

Palabras Clave: Wavelets, mercado eléctrico, Enfoque bayesiano, MCMC

1. Introducción

El mercado eléctrico funciona actualmente en España de forma competitiva. Actualmente existen dos posibles formas de operación en el mercado eléctrico: los contratos bilaterales y el pool.

Para los contratos bilaterales la reducción del riesgo presente en los mismos se puede hacer prediciendo los precios con un horizonte que cubra la duración del contrato, habitualmente un año. En el pool tanto las compañías eléctricas como los consumidores envían al operador del mercado (OMEL) sus respectivas ofertas de generación y consumo para cada hora del día siguiente.

El disponer de buenas predicciones a corto plazo permite a los productores organizar su producción para maximizar sus beneficios. El productor asume cierto riesgo, pero este es tanto más bajo cuanto más precisa sea la predicción de que dispone. Cualquier estrategia de oferta se apoya en predicciones de precios. Por ello, el modelado de los precios es de gran interés.

2. Metodología

En este trabajo se aplica un modelo bayesiano para la predicción de los precios horarios de la energía utilizando wavelets. Se aplica un enfoque bayesiano, con un modelo jerárquico completo. La estimación del modelo se realiza a través de técnicas MCMC.

La idea de las wavelets es conseguir una parametrización que capte la mayor parte de la variabilidad con pocos parámetros, de acuerdo con el principio de parquedad (parsimony). Se trabaja con la transformada wavelet de la señal original, después se suprime un subconjunto de los coeficientes (wavelet shrinkage o "encogido") y se deshace la transformación con los restantes. Dos aplicaciones de los wavelets a la estimación de precios de la energía eléctrica son Conejo et al (2005a) y Conejo, et al (2005b). En la primera se combinan las wavelets con los modelos ARIMA utilizando solamente cuatro coeficientes de la transformación wavelet y en la segunda se comparan las predicciones de los wavelets con las de redes neuronales, regresión dinámica, modelos ARIMA y funciones de transferencia.

El artículo seminal de wavelet shrinkage es Donoho y Johnstone (1995). El enfoque bayesiano está desarrollado en Abramovich, Sapatinas y Silverman (1998). La extensión a errores correlacionados es Johnstone y Silverman (1997).

3. Agradecimientos

José Mira, Carolina García-Martos y María Jesús Sánchez agradecen la financiación del proyecto MTM2005-08897 del Ministerio de Educación y Ciencia.

4. Bibliografía

- [1] Conejo, A., Plazas, M., Espínola, R., (2005a), *"Day-Ahead Electricity Price Forecasting Using the Wavelet Transform and ARIMA Models"*. IEEE Transactions on Power Systems, vol. 20, num 2, 135-1042.
- [2] Conejo, A., Conteras, J., Espínola, R. y Plazas, M., (2005b), *"Forecasting electricity prices for a Day-Ahead Pool-based electric energy market"*. International Journal of Forecasting, vol 21, 435-462.
- [3] Abramovich, F., Sapatinas, T. y Silverman , B.W. (1998), *"Wavelet Thresholding via a Bayesian Approach"*, Journal of the Royal Statistical society B, vol. 60, num 4, 725-749.
- [4] Donoho, D., y Johnstone, I., (1995) *"Ideal Spatial Adaptation via Wavelet Shrinkage"*, Biometrika, 81,425-455.
- [5] Johnstone, I. y Silverman, B.W, (1997), *"Wavelet Threshold Estimators for Data with Correlated Noise"*, Journal of The Royal Statistical Society, B, vol 59, num 2, 319-351.

Aplicación de Modelos de Frontera Estocástica a Pequeñas Empresas

Felipe M. Rosa-González¹, Antonio Arbelo-Álvarez², Enrique González-Dávila¹

¹ frosag@ull.es, egonzale@ull.es, Departamento de Estadística, Investigación Operativa y Computación, Universidad de La Laguna

²aarbelo@ull.es, Departamento de Economía y Dirección de Empresas, Universidad de La Laguna

Abstract

En este trabajo se aplican diferentes modelos de frontera estocástica a un conjunto de datos de panel con información sobre pequeñas empresas del ámbito de la Comunidad Autónoma de Canarias. Se intenta hacer una comparación de las distintas estimaciones de la eficiencia técnica.

Claves: Modelos de frontera estocástica; Eficiencia; LIMDEP.

AMS: 91B70; 62P20

1. Introducción

El uso de los modelos de frontera estocástica han sido ampliamente utilizados en el estudio de la medida de la eficiencia técnica en las empresas (Kumbhakar y Lovell (2000)). Basándonos en la literatura existente aplicaremos algunos modelos dados por diferentes autores a un conjunto de datos referidos a pequeñas empresas del ámbito de la Comunidad Autónoma de Canarias, con el fin de comparar las diferentes estimaciones de la eficiencia técnica dados por estos modelos.

Los modelos de frontera estocástica responden de forma general a la ecuación:

$$y_{it} = \beta' x_{it} + \nu' z_i + v_i t - u_{it} \quad (2.6)$$

con $i = 1, \dots, N$ y $t = 1, \dots, T$, y además con las condiciones distribucionales $v_{it} \sim N(0, \sigma_v^2)$, y $u_{it} = |U_{it}|$ donde $U_{it} \sim N(0, \sigma_u^2)$. Siendo el interés de estos modelos su robustez frente a posibles cambios externos de variables no controladas por el investigador, tales como temperatura, capacidad de trabajo, etc.

2. Descripción de los datos

Usaremos un panel de datos compuesto por información recogida en 284 pequeñas empresas. En este panel se han medido 21 variables o factores económicos sobre ellas durante tres períodos anuales, 2001, 2002 y 2003.

3. Modelos planteados

Si bien nos centramos en los modelos resueltos por el software LIMDEP (Greene (2002, 2005)), tanto en su versión de efectos fijos como aleatorios, probaremos otras posibilidades como las planteadas por Schmidt y Sickles (1984), Cornwell *et. al* (1990) y Pitt y Lee (1981). Las estimaciones de las eficiencias técnicas dadas por los distintos modelos serán comparadas entre sí.

4. Bibliografía

- [1] Cornwell, C.; Sickles, R. y Schmidt, P. (1990). Production Frontiers with Cross-Sectional and Time Series Variation in Efficiency Levels. *Journal of Econometrics* 46 (1), 185-200.
- [2] Greene, W. (2002). *LIMDEP*. Econometrics Software Inc.
- [3] Greene, W. (2005). Fixed and Random Effects in Stochastic Frontier Models. *Journal of Productivity Analysis* 23, 7-32.
- [4] Kumbhakar,S. y Lovell,K. (2000). *Stochastics Frontier Analysis*. Cambridge University Press.
- [5] Pitt, M. y Lee, L. (1981). The Measurement and Sources of Technical Inefficiency in Indonesian Weaving Industry. *Journal of Development Economics* 9, 43-64.
- [6] Schmidt, P. y Sickles, R. (1984). Production Frontier with Panel Data. *Journal of Business and Economic Statistics* 2 (4), 367-374.

Aplicación de las redes neuronales y del análisis de la sensibilidad como método para la mejora de la calidad en el proceso del moldeo por inyección de termoplásticos

Elena Pérez Bernabeu¹, Miguel Ángel Sellés Canto², Vicente Jesús Seguí Llinares³, Miguel Jorge Reig Pérez⁴

¹elenapb@eio.upv.es, Departamento de Estadística e Investigación Operativa Aplicadas y de Calidad, Universidad Politécnica de Valencia

^{2,3,4}maselles@dimm.upv.es, Departamento de Ingeniería Mecánica y de Materiales, Universidad Politécnica de Valencia

Abstract

El proceso del moldeo por inyección de termoplásticos es el proceso de fabricación de plásticos más utilizado en todo el mundo. Para poder fabricar piezas con la máxima calidad, se desarrolla una metodología que modeliza mediante una red neuronal dicho proceso. El error medio obtenido es despreciable, siendo comprobado posteriormente mediante ensayos reales.

Palabras Clave: red neuronal, calidad, inyección termoplásticos

1. Introducción

El estudio del proceso de moldeo por inyección de forma pormenorizada y extensa coincide con el aumento de la capacidad de producción de los equipos de inyección y del simultáneo incremento de la demanda de productos. Es de hecho el método de fabricación de plásticos más utilizado en todo el mundo. Por sus amplias posibilidades de generación de formas complejas y sus bajos costes, éste se convierte en un proceso de elevado volumen de producción y cuyos productos, que, sin embargo, presentan calidades finales fuertemente influenciadas por las condiciones de proceso. No obstante, la tendencia actual del mercado en la introducción de productos plásticos en sectores como automoción, electrónica o medicina, está exigiendo niveles de calidad muy elevados. De hecho, la calidad de las piezas obtenidas por este método de fabricación es el resultado de una combinación compleja entre el material, la pieza y el diseño del molde, y las condiciones de proceso, pudiendo ser cuantificada en función de su apariencia, dimensiones y características mecánicas [1].

Hasta hace poco tiempo, la calidad se controlaba en las empresas realizando ajustes en las máquinas de inyección mediante métodos de prueba-error durante la fase de puesta en marcha, en los que una serie de piezas se fabrican variando ligeramente los parámetros de proceso hasta que se alcanzan los requerimientos mínimos de calidad. Estos métodos, aparte de significar un esfuerzo adicional, incurren en un mayor gasto tanto de tiempo como de dinero. El propósito de este trabajo es el de desarrollar una metodología que pueda ser utilizada como

ayuda para alcanzar con mayor brevedad las condiciones de proceso óptimas, que son las que permitirán obtener la mayor calidad en las piezas.

2. Experimental

En primer lugar, se ha elaborado un diseño de experimentos, en el cual se han tomado en consideración cinco variables de proceso, controlables directamente por el operador de máquina: tiempo de inyección, temperatura de inyección, temperatura del refrigerante, nivel de presión de compactación y transición llenado-compactación. Éstas serán las entradas del proceso, mientras que las salidas estarán formadas por 8 variables resultado: presión máxima alcanzada, temperatura promedio ponderada, tensión de cizalla, velocidad de cizalla, fuerza de cierre, rango de contracciones, presión mínima alcanzada y contracción media al final del proceso..

Tras aplicar el método de Taguchi y de utilizar sus gráficas lineales, se ha confeccionado una tabla formada por 81 experimentos, a realizar en la máquina de inyección de termoplásticos del Departamento de Ingeniería Mecánica y de Materiales de la EPSA (UPV). Una vez obtenidos los resultados, los datos han sido introducidos en un software de redes neuronales, para que después de ser entrenada la red y haber utilizado el algoritmo BP (BackPropagation), genere una configuración que proporcione el menor error posible. El diseño óptimo se ha conseguido después de 17363 ciclos de entrenamiento, con 429 conexiones. La configuración es $5 \times 33 \times 8$. El error medio alcanzado es de $5 \cdot 10^{-5}$.

Posteriormente, mediante el análisis de sensibilidad, se ha conseguido simplificar la red neuronal generada, pasando a una de configuración $3 \times 22 \times 8$, con un error de $4,9 \cdot 10^{-3}$.

3. Conclusiones

Como se puede apreciar, el error del modelo simplificado continua siendo muy bajo, con lo que la aproximación al comportamiento real del proceso es totalmente válida.

Esta aproximación ha sido contrastada posteriormente con ensayos reales realizados directamente sobre máquina, corroborando los resultados proporcionados por la red neuronal reducida.

4. Bibliografía

- [1] S.A. PATEL, P.K. MALLICK. Development of a methodology of defect reduction in injection molding using process simulations. Part 1: Effects of inlet melt temperature and coolant temperature. *Journal of Injection Molding Technology*, 2(4), 1998.

Aplicación de técnicas composicionales al estudio del comportamiento de cerdas gestantes

J. Daunis-i-Estadella¹, G. Mateu-Figueras², N. Chapinal³, X. Manteca⁴, J.L. Ruiz de la Torre⁵

¹josep.daunis@udg.es, ²gloria.mateu@udg.es, ³nuria.chapinal@uab.es,
⁴xavier.manteca@uab.es, ⁵joseluis.ruiz@uab.es,

^{1,2}Departament d'Informàtica i Matemàtica Aplicada, Universitat de Girona,

^{3,4,5}Departament de Biologia Cel·lular, Fisiologia i Immunologia, Universitat Autònoma de Barcelona

Abstract

La metodología del biplot adaptada a datos composicionales, junto con las técnicas de análisis composicional son aplicadas al estudio multivariante del comportamiento de cerdas gestantes, en los aspectos referentes a su actividad oronasofacial. Estos análisis pueden ayudar a la toma de decisiones en referencia al sistema de estabulación y el bienestar de los animales.

Palabras Clave: biplot, logcocientes, comportamiento animal.

AMS: 62P10, 62H99, 92D50

1. Introducción

Los datos composicionales [1] pueden ser tomados como las realizaciones de un vector aleatorio de componentes estrictamente positivas y de suma constante. Dicha tipología de datos es muy frecuente en campos muy diversos: geología, química, etología, sociología, demografía,...

Cuando el número de componentes o partes de la composición es elevado se requieren técnicas de visualización multivariantes como el biplot [3], que nos ayudan a describir las filas (individuos) y columnas (componentes) de nuestros datos.

La adaptación de las técnicas biplot a la particularidad de los datos composicionales [2][3] conlleva una interpretación que no es la habitual en este tipo de gráficos. En este trabajo se expone una aplicación de técnicas biplot y el análisis composicional, para detectar diferencias entre el comportamientos de cerdas gestantes, por lo que hace referencia a distintos tipos de estabulación.

2. Datos

El conjunto de datos se corresponde a composiciones de 7 partes correspondientes a conductas oronasofaciales [4] obtenidas mediante muestreo de barrido [5] (bebida, masticación en vacío, interacción con el suelo, manipulación de barras, interacción con el comedero, interacción con el dispensador de pienso y

residuo) de 180 cerdas en tres sistemas de estabulación diferente: jaulas individuales, sistema de alimentación en caída lenta (trickle feeding) y sistema de alimentación electrónico (electronic sow feeding).

3. Aproximación metodológica

Desde [1] se ha aceptado que un correcto estudio de un conjunto de datos composicionales debe basarse en las transformaciones logcociente y en el uso de técnicas y métodos adaptados a la naturaleza de los datos.

En una primera fase del análisis que se presenta se tratan los datos obtenidos (diferentes para los tres sistemas de estabulación) y la presencia de ceros. La segunda fase consiste en detectar las configuraciones de las componentes representadas en el biplot composicional y el uso de éstas para el análisis más detallado mediante técnicas composicionales.

Finalmente se abordará una fase de análisis subcomposicional, mediante el uso de técnicas ANOVA para logcocientes, para estudiar la diferencia de comportamiento según la estabulación.

4. Agradecimientos

Este trabajo ha recibido financiación mediante los proyectos BFM2003-05640/MATE y PETRI 95-0639.OP (Dir. General de Investigación; Minis. de Ciencia y Tecnología).

5. Bibliografía

- [1] Aitchison, J. (1986) *The statistical analysis of compositional data*. London: Chapman and Hall. Reprinted in 2003 by Blackburn Press.
- [2] Daunis-i-Estadella, J., Tolosana-Delgado, R. y Martín-Fernández, J.A. (2004), Estrategias de análisis de datos composicionales mediante técnicas biplot y representaciones ternarias. In: Dept. EIO, UCA (Eds.), XXVIII Congreso Nacional de Estad. e Invest. Oper., Cádiz (E), CD-ROM, ISBN: 84-689-0438-4, 11 p.
- [3] Aitchison, J. y Greenacre, M. (2002). Biplots of compositional data. *Applied Statistics*, vol. 51, Part. 4, pp. 375-392.
- [4] McGlone, J. J., von Borell, E., Deen, J., Johnson, A. K., Levis, D. G., Meunier-Salaun, M. C., Morrow, J., Reeves, D., Salak-Johnson, J. L. and Sundberg, P. L. (2004). Compilation of the scientific literature comparing housing systems for gestating sows and gilts using measures of physiology, behavior, performance and health. *The Professional Animal Scientist* 20:105-117.
- [5] Martin, P. R. y Bateson, P. (1993) *Measuring behaviour: an introductory guide*. Cambridge University Press, 2nd ed. 222p. ISBN 0-521-44614-7

Aplicación de técnicas estadísticas multivariantes a la identificación de vinos mediante descriptores aromáticos

Antonio Vaamonde Liste¹, Patricio Sánchez Bello², Patricio Sánchez Fernández³

¹vaamonde@uvigo.es, Depto de Estadística e I.O. Universidad de Vigo

²psanchez@uvigo.es, Depto de Estadística e I.O. Universidad de Vigo

³patricio@uvigo.es, Depto de Ec. Financ. y Cont. Universidad de Vigo

Abstract

La identificación de vinos se realiza habitualmente mediante análisis organoléptico. En este trabajo se utilizan muestras de vinos procedentes de dos conocidas zonas vinícolas de Galicia: Rias Baixas (Albariño) y Ribeiro. Las puntuaciones de los catadores permiten elaborar perfiles aromáticos. El Análisis Discriminante contribuye a una identificación razonablemente fiable.

Palabras Clave: wine, identification, discriminant analysis

AMS: 62P30

1. Introducción

El vino es un producto de gran interés comercial que plantea numerosos problemas de identificación. El aseguramiento de la calidad ha llevado a establecer controles de denominación de origen, mediante consejos reguladores que establecen normas para tratar de mantener la fiabilidad y autenticidad de los productos vendidos con la etiqueta de la denominación.

La complejidad del vino, con varios cientos de componentes cuyos efectos sobre la calidad no son todavía muy bien conocidos, ha supuesto tradicionalmente una dificultad insalvable: tan solo el recurso a catadores especializados con gran experiencia y habilidad ha permitido establecer razonablemente –aunque con altos niveles de incertidumbre- el origen, las variedades de uva utilizadas, el año de la cosecha, o la calidad del vino.

2. Perfil varietal

Una herramienta útil para la identificación de los vinos es el perfil sensorial de aromas. Para su construcción se utiliza la norma ISO 11035 y complementarias, mediante comités de cata especialmente entrenados para la detección de aromas específicos. El procedimiento estadístico comienza con la elaboración de una lista de descriptores relacionados con los vinos de la denominación, inicialmente extensa (un centenar o más), que se reduce –mediante la opinión de los expertos y utilizando un procedimiento normalizado- a un conjunto limitado (en torno a una docena), en el que se incluyen los descriptores más importantes con los cuales se construirá el perfil. En una segunda fase los catadores son entrenados

con los productos de referencia de cada descriptor hasta conseguir que detecten pequeñas concentraciones mediante pruebas ciegas. Por último los catadores evalúan en una escala de 0 a 5 la intensidad percibida de cada descriptor en una muestra de varios vinos (al menos 10).

Mediante este procedimiento se han elaborado los perfiles aromáticos medios de los vinos de dos zonas de Galicia: Rias Baixas (albariño) y Ribeiro. Los comités de cata son distintos en cada zona, ya que los catadores son especializados y poseen una larga experiencia dentro de la Denominación de Origen. También son distintos los descriptores seleccionados en cada zona, aunque muchos de ellos son comunes.

3. Método Estadístico

El objetivo principal de este trabajo es la identificación de la zona de procedencia (Rias Baixas o Ribeiro) de cada vino a partir de las puntuaciones obtenidas en los diferentes descriptores aromáticos. Se utilizan únicamente los descriptores que son comunes a ambas denominaciones de origen, y los catadores para los que se dispone de datos completos. En total se trata de 435 pruebas de identificación de aromas, 26 vinos (16+10), 20 catadores (11+9) y 16 descriptores (14 de ellos comunes).

	DENOMINACION	GRUPO PRONOSTICADO		TOTAL
		ALBARIÑO	RIBEIRO	
ORIGINAL	ALBARIÑO	231 90,6 %	24 9,4%	255 100,0%
	RIBEIRO	2 1,1%	178 98,9%	180 100,0%
VALIDACIÓN CRUZADA	ALBARIÑO	228 89,4%	27 10,6%	255 100,0%
	RIBEIRO	5 2,8%	175 97,2%	180 100,0%

Tabla 1: Resultados de la clasificación

4. Conclusiones

Los resultados muestran con claridad que el Análisis Discriminante hace posible identificar fiablemente (94,0 % de los casos originales, 92,6 % en la validación cruzada) la zona de origen a partir de las puntuaciones de los catadores. De este modo, se comprueba que el perfil varietal puede ser utilizado como una referencia válida para asegurar la autenticidad de los vinos de cada denominación.

Aproximación de funciones de clasificación

Beatriz Lacruz¹, Rosa Eva Pruneda²

¹lacruz@umizar.es, Departamento de Métodos Estadísticos, Universidad de Zaragoza

²Rosa.Pruneda@uclm.es, Departamento de Matemáticas, Universidad de Castilla-La Mancha

Abstract

En este trabajo presentamos un método de clasificación consistente en aproximar la función clasificadora mediante una combinación lineal de funciones linealmente independientes. Discutimos las ventajas de esta técnica y la ilustramos aplicándola a varios conjuntos de datos simulados y reales. Finalmente, comparamos su eficacia con otros métodos como son las redes neuronales y la regresión logística.

Palabras Clave: Clasificación, ecuaciones funcionales

AMS: 6207, 39B22

Aproximaciones en tests χ^2 y F con modelos cercanos a la normal. Aplicaciones en Robustez

A. García-Pérez

agar-per@ccia.uned.es, Departamento de Estadística, Investigación Operativa y Cálculo Numérico, Universidad Nacional de Educación a Distancia (UNED), Paseo Senda del Rey 9, 28040-Madrid

Abstract

En el trabajo se obtienen buenas aproximaciones analíticas del p-valor y de los puntos críticos de tests χ^2 y F de Snedecor, cuando la distribución modelo es cercana a la habitual distribución normal. Con estas aproximaciones es posible analizar la robustez de este tipo de contrastes.

Claves: Aproximaciones von Mises y saddlepoint. Robustez en contraste de hipótesis.

AMS: 62F35 ; 62E17 ; 62F03

1. Introducción

En García-Pérez (2003) se obtuvieron aproximaciones von Mises (VOM) y saddlepoint (SAD) del punto crítico de un test de hipótesis con estadístico de contraste T_n . Aquí se proponen, de nuevo, aproximaciones VOM+SAD, tanto para el punto crítico como para el p-valor, de los habituales tests χ^2 , t de Student y F de Snedecor, pero cuando la distribución modelo F ya no es la habitual distribución normal $\Phi_{\mu,\sigma}$ sino otra cercana a ella.

La idea de las aproximaciones VOM+SAD es usar los dos primeros términos del desarrollo von Mises del funcional correspondiente (p-valor $p_n^F = P_F\{T_n > t\}$, o punto crítico k_n^F), el cual depende de la TAIF (*Tail Area Influence Function*, Field and Ronchetti, 1985) calculada en otro modelo G ,

$$p_n^F \simeq p_n^G + \int \text{TAIF}(x; t; T_n, G) dF(x)$$

$$k_n^F \simeq k_n^G + \frac{1}{g_n(k_n^G)} \int \text{TAIF}(x; k_n^G; T_n, G) dF(x)$$

en donde g_n es la densidad de T_n bajo el modelo G . Si F y G son suficientemente próximas, este tipo de aproximaciones son buenas.

Por esta razón, son utilizadas habitualmente en Estadística sustituyendo el modelo F por la distribución empírica F_n^* y tomando un tamaño muestral suficientemente grande.

Aquí mantenemos ambos modelos F y G , cuya cercanía valoramos mediante la ordenación de distribuciones $<_t$ introducida por Loh (1984). Si consideramos como modelo G la distribución normal $\Phi_{\mu,\sigma}$, el test para el que obtenemos la aproximación (VOM) es uno de los habituales tests χ^2 , t de Student, o F de Snedecor, pero ahora, cuando el modelo ya no es la normal.

En muchas situaciones, la función de influencia TAIF es descocida y debe de ser aproximada. Nosotros la aproximamos mediante aproximaciones saddlepoint (SAD), de extraordinaria precisión incluso para muestras de tamaño $n = 1$ ó 2, que incorporadas a la aproximación VOM dan como resultado las aproximaciones VOM+SAD propuestas en el trabajo.

Estas aproximaciones analíticas VOM+SAD permiten estudiar, entre otras cosas, la robustez de este tipo de contrastes.

Las aproximaciones VOM y VOM+SAD han dado muy buenos resultados para tests χ^2 en García-Pérez (2006a) y para tests t de Student en García-Pérez (2006b). Actualmente se han obtenido aproximaciones VOM+SAD para el caso general de tests F de Snedecor en García-Pérez (2006c).

2. Bibliografía

- [1] Field C.A. and Ronchetti E. (1985). A tail area influence function and its application to testing. *Communications in Statistics*, 4, 19-41.
- [2] García-Pérez A. (2003). von Mises approximation of the critical value of a test. *Test*, 12, 385-411.
- [3] García-Pérez A. (2006a). Chi-square tests under models close to the normal distribution. *Metrika*, 63.
- [4] García-Pérez A. (2006b). t -tests with models close to the normal distribution. In: Balakrishnan N., Castillo E y Sarabia J.M. (eds.), *Advances in Distributions, Order Statistics, and Inference*. Springer-Verlag.
- [5] García-Pérez A. (2006c). Approximations for F -tests with models close to the normal distribution. Enviado.
- [6] Loh W.Y. (1984). Bounds on AREs for restricted classes of distributions defined via tail-orderings. *The Annals of Statistics*, 12, 685-701.

Bayesian estimation of the Gaussian mixture GARCH model

Maria Concepción Ausín¹, Pedro Galeano²

¹mausin@udc.es, Departamento de Matemáticas, Universidade da Coruña

²pgaleano@usc.es, Departamento de Estadística e Investigación Operativa,
Universidade de Santiago de Compostela

Abstract

Bayesian inference and prediction for a generalized autoregressive conditional heteroskedastic (GARCH) model where the innovations are assumed to follow a mixture of two Gaussian distributions is performed. The mixture GARCH model can capture the patterns usually exhibited by many financial time series such as volatility clustering, large kurtosis and extreme observations. A Griddy Gibbs sampler implementation is proposed for parameter estimation and volatility prediction. Bayesian prediction of the Value at Risk is also addressed providing point estimates and predictive intervals. The method is illustrated using the Swiss Market Index.

Claves: Bayesian mixtures, GARCH models, Value at Risk

1. Introducción

A large amount of theoretical and empirical research has been carried out on analysis of financial time series in the last two decades. The main features exhibited by many of these series are time-varying volatility, heavy-tailed distributions, large kurtosis and extreme events.

Many models have been proposed for modeling the time-varying volatility in financial time series, including the autoregressive conditional heteroskedastic (ARCH) model, its generalization, the GARCH model, and the stochastic volatility model. These models assume that the conditional variance of the series is a function of the current information and have been successful in modeling large periods of tranquility followed by small periods of high volatility.

However, less attention has been paid to explaining heavy-tailed distributions, large kurtosis and extreme events. The usual assumption in fitting models to financial data is that the returns, i.e., the first difference of the logarithm of the series, are conditionally normally distributed. However, the normal GARCH model is known to be inconsistent with high kurtosis, heavy tails and extreme events. The normal stochastic volatility model can capture some leptokurtosis but not large enough to explain the high sample kurtosis found in real data series. To explain these facts, different approaches have been considered in

the literature, as modeling the innovations of the GARCH model with a t-distribution, considering the use of the generalized error distribution and non-parametric approaches. Alternatively, Bai et al. (2003) proposed modeling the innovations distribution with a mixture of two zero mean normal distributions with different variances. This is a model distribution which postulates that a large number of innovations are generated by a normal density with a small variance, while a small number of innovations are generated by a normal density with a large variance. This specification can capture volatility clustering, high kurtosis, heavy tails and the presence of extreme events.

Inference on ARCH and GARCH models has been traditionally carried out using maximum likelihood, quasi- maximum likelihood or the generalized method of moments. There has been much less progress in the analysis of these models from the Bayesian perspective. The Bayesian methodology offers a natural way to introduce parameter uncertainty in the estimation of volatilities. Also, predictive distributions of the future volatilities can be obtained which are more informative than simple point forecasts. Furthermore, predictive distributions of Value at Risk (VaR) can also be obtained. This is one of the most important issues in Risk Management and the Bayesian approach provides a measure of precision for VaR estimates via predictive intervals.

In this paper, we present a procedure for Bayesian inference and prediction of the GARCH model with Gaussian mixture innovations based on MCMC methods. Although mixture models are intrinsically difficult to analyze, the Bayesian approach allows for data augmentation techniques where indicator variables can be introduced to simplify the likelihood and the derivation of the posterior distributions. Our Bayesian approach combines these ideas with the Griddy Gibbs sampling algorithm which is a modified Gibbs sampler where the conditional posterior distributions are approximated by numerical integration methods. The Griddy Gibbs sampler has also been used by Bauwens and Lubrano (1998) for a GARCH model with t-distributed errors. This approach is easier to implement than Metropolis Hastings algorithms or importance sampling and, although it is more costly in computer time, the problems in finding an appropriate proposal distribution or importance function in these other methods are avoided.

2. References

- Bai, X., Rusell, J.R., Tiao, G.C., 2003. Kurtosis of GARCH and stochastic volatility models with non-normal innovations. *J. Econometrics*, 114, 349-360.
- Bauwens, L., Lubrano, M., 1998. Bayesian inference on GARCH models using Gibbs sampler. *Econometrics J.*, 1, 23-46.

Bayesian latent multinomial models: A case study

Javier Portela¹

¹jportela@estad.ucm.es, Dpto de Estadística e I. Operativa III, Universidad Complutense de Madrid

Abstract

In remembering events, we can visualize ourselves in the recalled image (Observer perspective), visualize the event image as in normal perception (Field perspective), or alternate between those points of view. Latent multinomial Bayesian Models are applied to a survey where subjects reported their remembering points of view for a number of events.

Keywords: bayesian models, multinomial, latent models

1. Comparison of three basic models

Three different Bayesian models are compared in modelling the three-way categorical variable **POV** (1=Observer, 2=Field, 3=Changing). Gibbs sampling is used to draw posterior inferences.

I. Simple Multinomial Model. The likelihood is set as $\text{POV}_i \sim \text{Multi}(p_1, p_2, p_3, n_i)$, with prior distributions $n_i \sim \text{Bin}(q, 102)$ and $(p_1, p_2, p_3) \sim \text{Dirichlet}(\alpha_1, \alpha_2, \alpha_3)$. Gibbs sampling is used to draw samples from the Dirichlet posterior distribution.

II. Multinomial 'By Subject' Model . The POV Variable is strongly subject-dependent. These strong differences between subjects in Points of View justify to use a different multinomial model for each subject. Settings are similar to the latter model, except for the variable $\text{POV}_i \sim \text{Multi}(p_{1i}, p_{2i}, p_{3i}, n_i)$.

III. Multinomial Latent Model . It looks like subjects could be clustered in large groups depending on their usual Point of View in recalling . We study this possibility through the latent categorical variable Z . For simplicity in the notation, we present the formulae for the case where Z is a two-category variable: $\text{POV}_i | Z_i = 1 \sim \text{Multi}(p_{11}, p_{21}, p_{31}, n_i)$, $\text{POV}_i | Z_i = 2 \sim \text{Multi}(p_{12}, p_{22}, p_{32}, n_i)$, $(p_{11}, p_{21}, p_{31}) \sim \text{Dirch}(\alpha_{11}, \alpha_{21}, \alpha_{31})$, $(p_{12}, p_{22}, p_{32}) \sim \text{Dirch}(\alpha_{12}, \alpha_{22}, \alpha_{32})$. $Z_i \sim \text{Bin}(r, 1)$, with prior set for r is a $\text{Beta}(1, 1)$. Priors for $(\alpha_{11}, \alpha_{21}, \alpha_{31})$ and $(\alpha_{12}, \alpha_{22}, \alpha_{32})$ are set like in the first model.

Comparisons of the three models are made through the "Deviance Information Criterion", DIC . It is interesting to see that subjects can be roughly grouped in three clusters using the latent model: 14 "Observers ($Z=3$)", 11 "Field ($Z=2$)" and 8 "Changing POV ($Z=1$)" subjects.

2. Conditional Multinomial latent model

For each event, subjects fill the questionnaire for a second time, at least one month after the series of 102 days has finished. Changes of point of view for each event are studied. The latent model is used , clustering the individuals in three types of subjects. POV_1 is modelled as in the 'By Z' Multinomial Model presented in previous Section. POV_2 is modelled via a Conditional Multinomial Model, conditioning on values taken in POV_1 : $(POV_2 | POV_1 = j, Z) \sim Multi(p_{1z}^j, p_{2z}^j, p_{3z}^j, n_j)$, $j = 1, 2, 3$

As we expect that subjects conserve the original Point Of View , that is, that $POV_2 = POV_1$, we adopt priors for p_{iz}^j that take into account this fact: $(p_{1z}^j, p_{2z}^j, p_{3z}^j) \sim Dirichlet(\alpha_{1z}^j, \alpha_{2z}^j, \alpha_{3z}^j)$, where $\alpha_{iz}^j = 1$ if $i = j$ and $\alpha_{iz}^j = 0$ if $i \neq j$

This modelling approach leads directly to Dirichlet conditional distributions in the Gibbs settings. The results obtained show that subjects tend to 'keep' or 'change to' their original Type-POV for the events.

3. A Bayesian logit model for changes over time

When the series of 102 daily events is finished, subjects wait approximately one month and refill the questionnaire for each event , with one month as maximum delay. Variables POV_1 and POV_2 are coded in two categories: Observer and Non-Observer . Let p_{ij} = the probability of remembering an event as Observer in the second remembering time ($p_{ij} = P(POV_2 = 1)$). t_{ij} is the time elapsed since the first remembering, and we seek to study if the β parameter is significant. A logit random effects model is set . Gibbs sampling is used to the model

$$\text{logit } (p_{ij}) = \alpha_i + \mu_i(POV_1 - \bar{POV}_1) + \beta(t_{ij} - \bar{t})$$

Results do not show a clear relationship of the portion of time elapsed with changes in perspective.

4. Bibliography

- [1] Congdon, P. (2003) *Applied Bayesian Modelling* .Wiley.
- [2] Chen, M. , Shao, Q. and Ibrahim, J. G. (2000). *Monte Carlo methods in Bayesian Computation*. Springer Verlag. New York.
- [3] Nigro, G., Neisser, U. (1983) . Point of View in personal memories. *Cognitive Psychology*, 15, 467-482

Biplot Interactivo como herramienta para el diagnóstico de la interacción en tablas de tres vías

A. Blázquez-Zaballos¹, J. L. Vicente-Villardón¹, M. Varela², P. Galindo¹

¹abz@usal.es, Dpto. de Estadística, Universidad de Salamanca

²Instituto Nacional de Ciencias Agrícolas. Habana (Cuba)

Abstract

En este trabajo se estudia la interacción genotipo-ambiente-tiempo mediante una extensión de los modelos AMMI al caso de tres factores de variación. Se considera la diagnosis de modelos para datos continuos de 3 vías a partir del Biplot Interactivo, visto como una generalización del Biplot clásico.

Palabras Clave: AMMI, Modelo de Tucker, Biplot

AMS: 62H99

1. Introducción

En experimentos agrícolas en los que el rendimiento de varios genotipos se prueba en distintos ambientes, es importante el análisis de la interacción genotipo-ambiente. Los modelos AMMI [4], mantienen la descomposición de la interacción en términos multiplicativos, sin forzarla a tener una forma específica, ni un número concreto de términos. Estos modelos posibilitan la representación conjunta genotipos y ambientes, lo que nos permite estudiar la estabilidad de cada uno de ellos a lo largo de las categorías del otro y, por tanto, seleccionar que genotipos son más estables y aptos para el cultivo.

Estos modelos se conocen también como bilineales [3] o biaditivos [1], entre otros.

Si en lugar de dos vías, se introduce un factor más como el tiempo o distintas condiciones de cultivo, se incorpora al modelo un término de interacción triple. En este caso el modelo de Tucker [6] permite reducir la dimensionalidad simultáneamente en los tres factores y además, presenta una estructura multiplicativa en los parámetros, similar al caso de dos vías. En definitiva, preserva la esencia de los modelos AMMI (efectos principales aditivos y efecto interacción multiplicativo) para datos con tres factores.

2. Diagnóstico Biplot de la interacción de orden superior

Partiendo de que es posible explicar los residuales de interacción de orden dos o superior en tablas multivía, a partir del Biplot (modelos AMMI) o una

generalización del Biplot (modelo de Tucker con biplots interactivos), en este trabajo veremos que con una simple inspección de la posición geométrica de los marcadores, podemos detectar la presencia, o no, de interacción de tercer orden y diagnosticar el modelo lineal que mejor se ajusta a los datos, en los casos de ausencia de interacción triple.

Descomponer la tabla de tres vías a partir del modelo de Tucker permite generalizar los modelos a un número reducido de situaciones simples (ocho en el caso de tres vías) a partir de las cuales son estudiadas el resto [7].

Para estimar los parámetros se siguen las ideas de Kroonenberg y De Leeuw [5] los cuales proponen el algoritmo Tuckals3, que es un proceso de mínimos cuadrados alternados que minimiza la suma de cuadrados residual, manteniendo la estructura de tres vías.

El método propuesto permite el análisis de cualquier tipo de interacción triple y no solo de la interacción genotipo-ambiente-tiempo.

3. Bibliografía

- [1] Blazquez-Zaballos, A. (1998). *Análisis Biplot basado en modelos lineales generalizados*. Tesis Doctoral. Universidad de Salamanca
- [2] Carlier, A. y Kroonenberg, P. M. (1996). Decompositions and Biplots in three-way Correspondence Analysis. *Biometrika* 58, 453-467
- [3] Gabriel, K. R. (1971). The biplot-graphic display of matrices with application to principal component analysis. *Biometrika* 58, 453-467
- [4] Gauch, H. G. (1988). Model selection and validation for yields trials with interaction. *Biometrics* 44, 705-715
- [5] Kroonenberg, P. M. y De Leeuw, J. (1980). The biplot-graphic display of matrices with application to principal component analysis. *Psychometrika* 61(2), 355-373.
- [6] Tucker, L. R. (1966). The biplot-graphic display of matrices with application to principal component analysis. *Biometrika* 58, 453-467.
- [7] Varela, M. (2002). *Los Métodos BIPLOT como herramienta de análisis de interacción de orden superior en un modelo lineal/bilineal*. Tesis Doctoral. Universidad de Salamanca.

Biplot: Una alternativa a la regresión lineal multivariante

Olesia Cárdenas¹, Purificación Galindo², José L. Vicente-Villardón³

¹olesiacardenas@cantv.net, Postgrado en Estadística, Universidad Central de Venezuela

²pgalindo@aida.usal.es, ³villardon@aida.usal.es

Departamento de Estadística, Universidad de Salamanca

Abstract

En este trabajo se presenta la comparación entre la aproximación de una matriz obtenida a través del ajuste de un Biplot como un Modelo Bilineal y una Regresión con Componentes Principales, cuando las variables se tratan en forma simétrica, su distribución es normal y se introduce convenientemente una métrica en el espacio de las mismas.

Palabras Clave: Biplot, Modelos Bilineales, Ajustes no lineales

1. Introducción

Es posible enfocar los Biplots Clásicos (Gabriel, 1971) como ajustes a través de Modelos Bilineales (Cárdenas y Galindo, 2004), en cuyo caso al igual que los Modelos de Regresión Lineal Multivariante (RLM) su objetivo es la aproximación en baja dimensión de una matriz de datos (individuos por variables) a través de la búsqueda de subespacios de ajuste óptimo en el sentido de los mínimos cuadrados. La diferencia estriba en que mientras en Regresión se asignan papeles asimétricos a dos conjuntos de variables (dependientes e independientes) entre los cuales se establece una relación funcional con el fin de realizar predicciones de las variables dependientes, en un Biplot no se establece diferenciación alguna por lo que se realiza la aproximación sin hacer supuestos sobre distribuciones subyacentes ni utilizar modelo alguno.

En el trabajo se presenta un resumen que compara ambos análisis, un método alternativo de estimación para ajustar un Biplot a través de Modelos Bilineales (BMB) y una aplicación práctica donde se resaltan las bondades de los BMB, principalmente en el caso de ajustes no lineales.

2. Materiales y Métodos

Si consideramos la matriz de datos centrada por columnas, introducimos la métrica identidad en el espacio de las variables y se obtienen las componentes principales de la matriz de variables independientes, resulta que los resultados obtenidos a través de una Regresión con Componentes Principales (RCP) son similares a los de un BMB (Cárdenas *et al.*, 2006).

Para el ajuste de un BMB se pueden estimar las matrices de parámetros a través de la realización iterativa de regresiones alternadas, partiendo de una matriz

inicial de estimaciones para las filas de la matriz tal que contemple el problema de la multicolinealidad. Es posible utilizar los métodos de Mínimos Cuadrados y de Máxima Verosimilitud, dependiendo de la distribución probabilística de las variables y la forma en que se realizan las regresiones alternadas, columna a columna, fila a fila, o todas las columnas y/o filas simultáneamente (Cárdenas y Galindo, 2004).

Se ilustra la aproximación de una matriz comparando la bondad en los resultados obtenidos con los dos métodos, poniendo de manifiesto la potencialidad de los BMB como una alternativa confiable para el caso de ajustes lineales y no lineales.

3. Comentarios Finales

La aproximación BMB tiene las siguientes ventajas:

- Permite la realización de ajustes no lineales cuando las variables tienen cualquier distribución de la familia exponencial, permitiendo visualizar confiablemente sobre un plano las asociaciones entre individuos y/o las correlaciones entre variables.
- Permite realizar predicciones para todas las variables en la matriz de datos, sin considerar relaciones de dependencia entre ellas.
- Brinda la posibilidad de introducir asimetría en el análisis si se conocen relaciones de dependencia entre las variables, tal que el método nos permita predecir las variables dependientes a partir de las independientes como en Regresión (Cárdenas *et al*, 2005).

4. Bibliografía

- [1] Cárdenas, O., Noguera C., Galindo P. y Vicente-V., J.L. (2006). "Alternativa a la regresión con componentes principales basada en biplots de regresión". *Interciencia*, 31 (6). En imprenta.
- [2] Cárdenas, O., Álvarez W. y Noguera C. (2005). "Biplot de regresión asimétricos". LV Convención Asociación Venezolana para el Avance de la Ciencia, Universidad Central de Venezuela.
- [3] Cárdenas, O. y Galindo P. (2004). *Biplot con información externa basado en modelos bilineales generalizados*. Ediciones Consejo de Desarrollo Científico y Humanístico, Universidad Central de Venezuela.
- [4] Gabriel, K. (1971). The Biplot-graphic display of matrices with applications to principal component analysis. *Biometrika* **58**, 453-467.

Bootstrap para la media bajo datos con sesgo por selección

Jorge Luis Ojeda Cabrera¹, José Tomás Alcalá Nalváiz², José Antonio Cristóbal Cristóbal³

¹jojeda@unizar.es, ²jtalcal@unizar.es, ³cristo@unizar.es

Dept. de Métodos Estadísticos. Fac. Ciencias. Universidad de Zaragoza. Pedro Cerbuna, num. 12, 50009. Zaragoza.

Abstract

En este trabajo se presentan y comparan tres esquemas bootstrap para realizar contrastes para la media en el contexto de datos sesgados por selección. Además de un estudio de simulación en el que se aborda el comportamiento de dichos contrastes, se presentan algunos ejemplos y aplicaciones.

Claves: bootstrap, sesgo por selección.

AMS: 62F40, 62F03, 62G09

1. Introducción

Cada día con más frecuencia los investigadores aplicados se enfrentan con problemas en los que por diversos motivos no se dispone de observaciones directas del fenómeno bajo estudio. En estos casos, los investigadores deben basar sus estudios y conclusiones en observaciones que, con respecto al fenómeno de interés, contienen información *parcial* y/o *distorsionada*. Se habla entonces de *observaciones con sesgo* o *muestras sesgadas*, véase [5], [1], [3] y las referencias que citan.

En este trabajo se presenta una metodología basada en la compensación del sesgo que permite realizar estimación e inferencia de forma relativamente sencilla para algunos casos de datos u observaciones sesgadas. Además de analizar la forma en que influyen los datos sesgados en el proceso de estimación, se propone un estimador para la media basado en la compensación, estudiando tanto su consistencia como su distribución asintótica. Estos desarrollos asintóticos están relacionados con los obtenidos en [2] en el contexto de la regresión. A partir de dichos desarrollos asintóticos se propone una metodología bootstrap para desarrollar contrastes de bondad de ajuste para la media en este contexto.

El primero de los esquemas bootstrap que se proponen y estudian en el presente trabajo está basado en el remuestreo de las observaciones, y en este sentido es una extensión del bootstrap *naif* a este contexto, en que los datos están sesgados. Los otros dos métodos utilizan tanto las desviaciones a la media, como las características de los estimadores compensados para obtener el comportamiento distribucional de los estimadores.

Si bien los tres esquemas bootstrap dan lugar a contrastes consistentes, las diferencias en cuanto a la concepción y fundamento de los mismos (es decir: los diferentes mecanismos que emplean para evitar el sesgo presente en los datos), hacen pensar que bajo diferentes condiciones poblacionales, y bajo diferentes sesgos unos esquemas sean preferibles a otros. Para mostrar si existen o no diferencias entre los esquemas que se presentan, se propone un estudio de simulación bajo diferentes condiciones poblacionales, así como una aplicación de estos resultados a datos reales (véase [4]).

2. Agradecimientos

Este trabajo ha sido subvencionado parcialmente por el proyecto MAT2005-01464, así como por el proyecto 511 (Grupo consolidado DGA2005). La investigación de Jorge Luis Ojeda Cabrera en este proyecto ha sido también subvencionada en parte por el "Programa Europa" (DGA (CONSI+D) y CAI).

3. Bibliografía

- [1] Cristóbal, J. A. and Alcalá, J. T. (2001). An overview of nonparametric contributions to the problem of functional estimation from biased data. *Test*, vol. 10, num. 2, pág. 309–332
- [2] Jacobo de Uña–Alvarez (2003). Large Sample results under biased sampling when covariables are present *Statist. Probab. Lett.* vol. 63, num. 3, pág. 287–293
- [3] Mahfoud, M. and Patil, G. P. (1982). On Weighted Distributions. *Statistics and Probability. Essays in Honor of C. R. Rao*, pág. 479–492.
- [4] Olave, P. and Salvador, M. and Miguel, J. A. and Muñoz L. (1998). La fecundidad y el periodo de desempleo en la mujer española. Un estudio no-paramétrico. Minist. Trabajo y Asuntos sociales. España.
- [5] Patil, G. P. and Rao, C. R. (1978). Weighted distributions and size-biased sampling with applications to wildlife populations and human families. *Biometrics*, vol. 34, num. 2, pág. 179–189.

Bootstrap semiparamétrico para Modelos Lineales Mixtos Generalizados

José A. Sánchez-Espigares, Jordi Ocaña

Departamento de Estadística, Universidad de Barcelona

Abstract

En este trabajo se analizan diferentes métodos de remuestreo en el entorno de los modelos mixtos generalizados. Se considera el método de estimación PQL y las diferentes definiciones de residuos a que da lugar. Se proponen estrategias de reconstrucción de la variable respuesta en base al efecto que provocan en el error cuadrático medio de los estimadores.

Palabras Clave: Bootstrap semiparamétrico, GLMM

1. Introducción

Los modelos mixtos lineales generalizados [1] extienden el modelo lineal introduciendo la posibilidad de tratar datos agrupados (medidas repetidas, datos longitudinales), así como variables respuesta con distribución de probabilidad de la familia exponencial (binomial, Poisson, normal, etc...)

Dado el vector de respuestas $Y_i = (Y_{i1}, \dots, Y_{in_i})'$ para el grupo i-ésimo, la correlación inducida por la agrupación se modeliza mediante la inclusión de efectos aleatorios en el predictor lineal: $g(\mu_{ij}) = \theta_{ij} = X_{ij}\beta + Z_{ij}b_i$, donde g corresponde a la función link considerada, b_i representa el factor aleatorio y $E(Y_{ij}|b_i) = \mu_{ij}$

El método de estimación PQL (Penalized Quasi-Likelihood) corresponde a una aproximación de la estimación Máximo-verosímil. En este contexto, el concepto de “residuos” admite diferentes definiciones:

Residuos basados en respuesta ($Y_{ij} - \hat{\mu}_{ij}$), residuos de Pearson ($e_{ij} = \frac{Y_{ij} - \hat{\mu}_{ij}}{\sqrt{a_{ij}V(\hat{\mu}_{ij})}}$) de Pearson estandarizados ($\frac{e_{ij}}{\sqrt{1 - h(\hat{\mu}_{ij})}}$) residuos en la escala del predictor lineal ($\frac{g(Y_{ij}) - g(\hat{\mu}_{ij})}{\sqrt{a_{ij}g'(\hat{\mu}_{ij})^2V(\hat{\mu}_{ij})(1 - h_{ij})}}$), residuos de deviancia ($sign(Y_{ij} - \hat{\mu}_{ij})\sqrt{d_{ij}}$).

En el caso en que la respuesta sea Normal, los residuos de Pearson y en la escala del predictor lineal, coinciden con la definición de residuos estandarizados del modelo lineal.

2. Objetivo

En este trabajo se analizan los métodos bootstrap semiparamétricos, basados en la reconstrucción de la variable respuesta [2], aplicando el remuestreo de

las fuentes de variación en base a la distribución empírica, tanto en los efectos aleatorios como en los diferentes tipos de residuos.

Se ha implementado en R la generación de remuestras a partir de un modelo ajustado que incorpora los parámetros y las matrices de diseño. Para los efectos aleatorios, es posible indicar si se deben ajustar los momentos muestrales a los parámetros de modelo (opción Adjusted) y/o la selección de los residuos se debe restringir al grupo seleccionado (opción Nested). Para los residuos se puede indicar que definición de residuos se utiliza en la generación de la respuesta.

La librería permite la estimación del modelo sobre el conjunto remuestreado y la extracción de los estimadores obtenidos. Se han realizado dos ejercicios de simulación: un modelo con respuesta Poisson con observaciones atípicas, y otro con respuesta binomial. En ambos casos, se ha analizado el error cuadrático medio para cada método bootstrap considerado.

3. Contactar

Dirección:

José A. Sánchez-Espigares
 Departamento de Estadística
 Facultad de Biología – Av. Diagonal 645
 08028 – Barcelona
 Spain

Teléfono: (+34) 934035896

e-mail: jossanchez@ub.edu

4. Bibliografía

- [1] Verbeke G, Molenberghs G (2000) *Linear Mixed Models for longitudinal data*. New York: Springer-Verlag.
- [2] Davison A, Hinkley D. (1997). *Bootstrap Methods and their application*. Cambridge: Cambridge University Press.

Caracterización Multivariante de los perfiles de las mujeres en situación laboral irregular: el caso de Salamanca

P. Vicente¹, C. Patino², P. Galindo³

¹purivic@yahoo.com, ²mcarmen_patino@yahoo.es, ³pgalindo@usal.es

Departamento de Estadística, Universidad de Salamanca

Abstract

En este trabajo se presentan los resultados de un estudio, el primero de estas características a nivel de la Comunidad de Castilla y León, cuyo objetivo es identificar los perfiles socioeconómicos de las mujeres salmantinas, en situación de irregularidad laboral, en comparación con las características de las mujeres que desarrollan su actividad en un marco laboral regular.

Palabras Clave: Trabajo irregular, mujer, HOMALS

1. Introducción

En este trabajo entenderemos por “**empleo irregular**”, aquellas actividades económicas que eluden la normativa fiscal. El presente estudio aborda el análisis de la presencia femenina en el empleo irregular, sea cual sea la actividad económica desarrollada por las trabajadoras. Se presta especial atención a las empleadas de hogar por su relevancia en este contexto.

La estimación del empleo irregular femenino en España se realiza comparando, dos fuentes estadísticas que miden empleo: la EPA del INE. Considerando que la EPA detecta a las mujeres ocupadas en actividades por las que obtienen una remuneración, la mayoría de las cuales generan obligación de cotizar a la Seguridad Social, la diferencia entre ambas fuentes permite estimar situaciones de irregularidad laboral. La EPA no proporciona datos desagregados a nivel municipal por tanto resulta imposible utilizar este método para realizar una estimación del número de mujeres en empleo irregular que hay en los diferentes municipios. Eso implica que el único procedimiento viable para aproximarnos al conocimiento del problema requiere un procedimiento de estimación “directo”; es decir, un procedimiento basado en entrevistas realizadas a las personas implicadas.

2. Objetivos

- Identificar los sectores esencialmente femeninos y estructuralmente irregulares, en la provincia de Salamanca.
- Conocer las características personales, familiares, profesionales y económicas de las mujeres salmantinas, en situación de irregularidad laboral, en comparación con las características de las mujeres salmantinas que desarrollan su actividad en un marco laboral regular.
- Identificar los motivos que conducen a las mujeres salmantinas a recurrir, y/o

permanecer, en situación de irregularidad laboral, identificar las consecuencias, y las principales demandas que las mujeres salmantinas afectadas, hacen a la Sociedad y a las Instituciones.

3. Material y Métodos

Se ha realizado un muestreo por conglomerados (los cuales se han obtenido del análisis multivariante, HJ-Biplot de las características socioeconómicas de los diferentes municipios salmantinos, tomadas del Anuario Social de España (<http://www.anuarisoc.lacaixa.comunicaciones.com>)). En una segunda etapa se realiza, en los conglomerados elegidos en la primera etapa, un muestreo en bola de nieve, en el que se garantizarán las cuotas por edad y por sector económico. El tamaño de muestra en cada estrato se elige según afijación proporcional. Para el muestreo en bola de nieve el origen estará en los datos suministrados por colectivos relacionados con las personas en riesgo de exclusión social. Se han analizado 1000 encuestas.

El cuestionario consta de 50 preguntas que recogen Características Sociodemográficas, Formación y Cualificación, Situación laboral y trabajo, Consecuencias de la situación de irregularidad sobre la vida laboral y personal, Demandas y Medidas.

Se utiliza como herramienta estadística para buscar los perfiles sociodemográficos de las mujeres en situación irregular, el análisis multivariante no lineal HOMALS (ACP para variables medidas en escala nominal u ordinal), realizado con el SPSS, el cual permite detectar perfiles de irregularidad y describir las variables responsables de los clusters encontrados.

4. Resultados más relevantes

Se han encontrado tres clusters de mujeres en situación laboral irregular perfectamente diferenciados, los cuales, a su vez, se diferencian del colectivo de mujeres en situación regular. Se han descrito sus perfiles, y se han identificado sus demandas prioritarias.

5. Bibliografía

- [1] Cabo, G. ; González, A.; Roces, P. y Muñoz, S. (2002). *La presencia de las mujeres en el empleo irregular* Centro de Estudios Económicos Tomillo.
- [2] Ruesga, S. (1982). *La economía oculta: aproximación al caso español*. Dirección Gen. de Estudios y Documentación del Ministerio de Presidencia.
- [3] Galindo, M.P. (1986): Una alternativa de representación simultánea: HJ-biplot. *Questiío*, 1, 13-23.

Characterization and spatial distribution of heavy metals in soils irrigated with untreated wastewater

J. Carlos García-Díaz

Dpto. de Estadística e Investigación Operativa Aplicadas y Calidad

Universidad Politécnica de Valencia, 46022 Valencia, España

juagardi@eio.upv.es

Abstract

Heavy metals are one of the serious pollutants in our natural environment due to their toxicity, persistence and bioaccumulation problems. The Huerta from Valencia is a complex environment where both natural and anthropogenic processes influence the characteristics and distributions of soils and contaminants, which in turn is of importance for maintenance dredging and pollution control. This study investigated the characteristics and spatial distribution of heavy metals from soils irrigated with untreated urban wastewater. Kriging was then used to estimate the values at unsampled places of properties that vary in space.

Keywords: Soil pollution, Heavy metals, spatial distribution

AMS: 62P12

1. Introduction

Traditionally, the soil has been considered as a depository of unlimited quantities of human and animal wastes because of its regenerative and buffer capacity. Soils used in the irrigation ditch of Favara in Valencia's fertile region belong to the Holocene period. Rice, citric fruits and vegetables are cultivated, although rice crops dominate most of the area. Non-treated residual waters from the Turia River are used for irrigation, which are supposed to carry heavy metals and organic composites. Once the heavy metals are deposited in the soil, they tend to stay, at least initially, on superficial slices, getting submitted to several retention processes, similar to those related to natural soils, but incorporated in greater concentrations and under variable complex forms, depending on the polluting source. The persistence of these heavy metals in soils is much greater than in any other environment of the biosphere, and the pollution produced by them is virtually permanent. Furthermore, their toxic character causes significant damage to plants, animals and people. The high and sustained amount of pollutants applied to the soils can overwhelm their retention and adsorption capability, contributing to the presence of polluters and their transition to plants or subterranean water. The heavy metal cadmium, Cd^{2+} , is naturally present in soils at trace concentrations. However, some industrial activities and agricultural practices increase its levels in the soil.

2. Study area and sampling

An extensive investigation of soil was conducted in the Favara Channel area. The agricultural site, located at south-east from Valencia City (Fig. 1), is regular-shaped with two sub-areas of different soils cultivated with rice and citrus crop. Each sub-area is divided into four predefined parcels. For the soil types, a systematic sampling was carried out at different depths according to the horizon distribution.

3. Geostatistical techniques

Geostatistics was introduced into soil science some 20 years ago as an alternative to classification for representing spatial variation. Fundamental to geostatistics is the concept of a variable that is both random and spatially auto-correlated. Kriging was then used to estimate the values at unsampled places of properties that vary in space. In order to describe mathematically the way the variance of a property changes as the distance and direction separating any two points vary, a semivariogram is used. Experimental semivariograms were used as a measure of spatial dependence.

4. Bibliography

- [1] Deutsch, C.V. y A.G. Journel (1997) GSLIB: Geostatistical Software Library and User's Manual Guide. Oxford University Press, New York.
- [2] Goovaerts, P. (1997). Geostatistics for Natural Resources Evaluation. Oxford University Press. New York.
- [3] Méndez, R. F. (2003) Estudio de la salinidad, sodicidad y metales pesados presentes en suelos regados por la Acequia de Favara bajo un enfoque espacio temporal. Tesis Doctoral, Universidad Politécnica de Valencia. 513 p.

Clasificación de series temporales mediante profundidad para datos funcionales

Andrés M. Alonso¹, David Casado², Sara López-Pintado³, Juan Romo⁴

¹andres.alonso@uc3m.es, Departamento de Estadística,
Universidad Carlos III de Madrid

²david.casado@uc3m.es, Departamento de Estadística,
Universidad Carlos III de Madrid

³saral@stat.rutgers.edu, Departamento de Estadística, Rutgers University

⁴juan.romo@uc3m.es, Departamento de Estadística,
Universidad Carlos III de Madrid

Abstract

En este trabajo se propone un método de clasificación de series temporales mediante técnicas discriminantes para datos funcionales. En particular, utilizamos los procedimientos basados en profundidad propuestos por López y Romo (2005). Se ilustra el procedimiento propuesto con un ejemplo con datos reales.

Claves: series temporales; clasificación; profundidad.

AMS: 62H30; 62M10.

1. Introducción

La clasificación de series temporales es un problema estadístico de interés y aparece en campos de aplicación muy diversos. Por ejemplo, (i) en Medicina interesa hacer diagnósticos basados en señales biomédicas tomadas a los pacientes; (ii) en Geología para distinguir entre las ondas producidas por terremotos y las producidas por explosiones; (iii) en Ingeniería, para identificar estados de un proceso de producción. El problema de la clasificación de series se aborda desde dos posibles aproximaciones: los métodos en el dominio del tiempo y los métodos en el dominio de la frecuencia (ver, por ejemplo, Kakizawa, Shumway y Tanigushi, 1998).

En este trabajo, proponemos un procedimiento para clasificar series temporales utilizando estimaciones de la función de densidad espectral. Se basa en técnicas discriminantes de datos funcionales basadas en profundidad propuestas por López-Pintado y Romo (2005). La ventaja de los métodos basados en la idea de profundidad sobre otras formas de clasificación para datos funcionales es su estabilidad frente a datos atípicos. Los resultados, con datos simulados y reales, igualan o mejoran las tasas de error de clasificación obtenidas por métodos alternativos.

2. Procedimiento de clasificación

Sea $\{X_1, \dots, X_n\}$ una muestra de series temporales de la población P_1 , y sea $\{Y_1, \dots, Y_m\}$ una muestra de series temporales de la población P_2 . El procedimiento de clasificación tiene los siguientes pasos:

1. Se obtienen estimaciones de las funciones de densidad espectral $\{f_{X_1}, \dots, f_{X_n}\}$ y $\{f_{Y_1}, \dots, f_{Y_m}\}$ de las series $\{X_1, \dots, X_n\}$ e $\{Y_1, \dots, Y_m\}$, respectivamente.
2. Se calcula la profundidad por bandas generalizada $GS(f)$ para cada función obtenida en el paso anterior (ver detalles en López-Pintado y Romo, 2005).
3. Se obtiene la media funcional α -truncada de cada muestra: \bar{f}_X^α y \bar{f}_Y^α .
4. Dada una nueva serie Z cuya función espectral estimada es f_Z , se clasifica en P_1 si $\|f_Z - \bar{f}_X^\alpha\|_1 < \|f_Z - \bar{f}_Y^\alpha\|_1$, y en P_2 en caso contrario.

3. Ejemplo con datos reales

Los datos analizados corresponden a las mediciones, mediante un sismógrafo, de las vibraciones producidas por ocho terremotos y ocho explosiones. Cada una de estas dieciséis series consta de 2048 datos, los 1024 primeros datos corresponden a las ondas primarias y los siguientes a las ondas secundarias. La aplicación del procedimiento a ambas ondas da lugar a una tasa de error por validación cruzada similar a la que aparece en Kakizawa et al (1998) y Chinipardaz y Cox (2004).

4. Agradecimientos

A.M.A. agradece la financiación del programa “Juan de La Cierva”. J.R. agradece la financiación del proyecto SEJ2005-06454.

5. Bibliografía

- [1] Chinipardaz, R. y Cox, T. (2004). Nonparametric discrimination of time series data. *Metrika* 59, 13-20.
- [2] López-Pintado, S. y Romo, J. (2005). Depth-based classification for functional data. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, American Mathematical Society.
- [3] Kakizawa, Y., Shumway, R.H. y Taniguchi, M. (1998). Discrimination and clustering for multivariate time series. *Journal of the American Statistical Association* 93, 328-340.

Cluster-MDS con restricciones de contigüidad espacial

Rodrigo Macías, José Fernando Vera, José Miguel Angulo, Miguel Angel Montero

Departamento de Estadística e I.O., Facultad de Ciencias
Universidad de Granada

Abstract

En el análisis de los procesos espacio-temporales, la estimación de la estructura de covarianza espacial no estacionaria es un problema conocido en el que recientemente, Multidimensional Scaling no métrico ha jugado un papel importante (véase Sampson and Guttorp, 1992). En este trabajo se propone un modelo métrico simultáneo de MDS-cluster que permita la representación, no de las estaciones sino de los centros de las clases estimadas en base al variograma espacial en un plano.

Palabras Clave: Multidimensional Scaling, Análisis Cluster, procesos espacio-temporales.

En el análisis de los procesos espacio-temporales relacionados con estudios medioambientales, la suposición de estacionalidad en las estructuras de covarianzas espaciales usualmente no solo no es real sino que no resulta una hipótesis adecuada (véase Sampson and Guttorp, 1992), siendo la estimación de la estructura de covarianza espacial no estacionaria un problema conocido en el que recientemente, Multidimensional Scaling ha jugado un papel importante.

Basándose en una versión ponderada del algoritmo no métrico de MDS de Kruskal (1964a, 1964b), Sampson and Guttorp (1992) propusieron una aproximación no paramétrica a la estimación global de la estructura de covarianzas espaciales de una función aleatoria, $Z(x, t)$, observada repetidamente durante un periodo t_i , $i = 1, \dots, T$ sobre un número finito (y reducido, debido a las restricciones de dimensionalidad del MDS) de estaciones $x_i, i = 1, \dots, N$ sobre un plano, asumiendo estacionalidad temporal pero no espacial ni isotropía. Así, mediante el variograma como una métrica natural para la estructura de covarianza espacial, la dispersión es modelizada como una función suave de las coordenadas geográficas de las estaciones.

La consideración del MDS en esta metodología es una estrategia recomendable para resolver el problema de extrapolación en la situación no estacionaria, habiendo sido adoptado, entre otros, por Arbia y Lafratta (2002) o por Löland y Höst (2003), entre otros. No obstante, es conocido que una representación exacta en MDS es obtenida en dimensión $N - 1$, siendo la restricción de una representación bidimensional menos condicional para un procedimiento no métrico al precio de obtener una configuración que solo preserva el orden respecto

a los datos originales. Incluso en ese caso, cuando el número de estaciones es elevado, resulta difícil obtener una representación adecuada en un plano.

En este trabajo se propone la consideración de un procedimiento de Cluster-MDS simultáneo basado en el propuesto por Heiser y Groenen (1997) mediante el cual, no las estaciones sino los centros de los cluster, sean representados en un plano. Esta metodología puede ser aplicada cuando el número de estaciones sea demasiado elevado para una representación bidimensional de MDS o cuando una estructura de cluster esté subyacente en el espacio de dispersión.

Bibliografía

- Arbia G, Lafratta G. 1997. Evaluating and updating the sample design in repeated environmental surveys: monitoring air quality in Padua. *Journal of Agricultural, Biological, and Environmental Statistics* **2**: 251-466.
- Arbia G, Lafratta G. 2002. Anisotropic spatial sampling designs for urban pollution. *Applied Statistics* **51**: 223-234.
- Heiser WJ, Groenen PJF. 1997. Cluster differences scaling with a within-clusters loss component and a fuzzy successive approximation strategy to avoid local minima. *Psychometrika* **62(1)**; 63-83.
- Kruskal JB. 1964. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* **29**;1-27.
- Kruskal JB. 1964. Nonmetric multidimensional scaling: A numerical method. *Psychometrika* **29**; 28-42.
- Löland A, Höst G. 2003. Spatial covariance modelling in a complex coastal domain by multidimensional scaling. *Environmetrics* **10**; 307-321.
- Sampson PD, Guttorp P. 1992. Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association* **87**; 108-119.

Cómo surge la Estadística a partir de un conjunto de datos

A. Lara Porras, Y. Román Montoya, J. García Leal

Departamento de Estadística e Investigación Operativa, Facultad de Ciencias,
Universidad de Granada, Campus de Fuentenueva, s/n, 18071, Granada
e-mail: alara@ugr.es, yroman@ugr.es, juliagl@ugr.es

Abstract

Este trabajo surge como respuesta a nuestras inquietudes sobre cómo impartir Estadística. Pretendemos que los alumnos se sientan más motivados y participativos. Planteamos la introducción de conceptos y métodos estadísticos mediante un proyecto estadístico basado en datos reales, que permita que los alumnos encuentren relación entre lo explicado en clase y la realidad.

Palabras Clave: Proyecto estadístico, didáctica de la estadística.

AMS: 97U60; 97U70

1. Introducción

El proyecto que proponemos está orientado a propiciar una mejor comprensión y asimilación de los conceptos y métodos estadísticos. Para ello planteamos la utilización de datos reales, haciendo hincapié en el diseño del estudio, en la realización del análisis y la interpretación de los resultados. Desde el comienzo del curso, los alumnos tienen acceso a la información sobre los datos, las variables de trabajo y los objetivos que se persiguen alcanzar.

Nuestro interés se centra en que sean los alumnos los que deduzcan qué herramientas les proporciona la Estadística. A la vista de un conjunto de datos, ¿cómo manejarlos?, ¿cómo obtener conclusiones? Nuestra experiencia docente ha evidenciado la necesidad de unir los contenidos explicados en clase a la realidad y de establecer de este modo una estrecha relación entre teoría y práctica. Los conocimientos teóricos que han de adquirir los alumnos no tienen sentido si no pueden aplicarlos a datos concretos y las aplicaciones prácticas requieren de un importante fundamento teórico. La clásica distinción entre horas de teoría y horas de prácticas se convierte así en análisis de supuestos reales, pretendiendo de este modo dar solución a la continua incertidumbre que tenemos los profesores sobre la efectividad de nuestra labor docente. Ante la “tradicional” pregunta de los alumnos: *cuando estoy ante una situación real, ¿cómo la resuelvo?*, ante la apatía de los alumnos que no encuentran ninguna conexión entre lo explicado en clase y la realidad, ante toda la problemática que presenta una clase cuando el alumno no encuentra sentido a lo que intentamos que aprenda, nos planteamos la realización de este proyecto.

2. Descripción de la experiencia

Nuestra propuesta de metodología se ha centrado en la Licenciatura de Ciencias Ambientales. Nos planteamos introducir los conceptos y técnicas estadísticas de la asignatura “Estadística” de 3º Curso mediante un proyecto estadístico basado en datos cedidos por el Dpto. de Ecología de la Universidad de Granada. Estos datos proporcionan diferentes medidas sobre las hojas de encinas en cinco zonas diferentes de la provincia de Granada. El objetivo es estudiar la simetría de las hojas y ver si esta se modifica ante situaciones de stress tales como herbivoría, escasez de agua, falta de nutrientes,...

Desde el comienzo de las clases, el alumno se va a enfrentar con un conjunto de datos reales. El profesor describe los datos y las variables con que se trabaja. Se induce a los alumnos a plantearse qué hacer con los datos. El siguiente paso, es dejar que la Estadística “nazca” como fruto de las preguntas que se cuestionan en clase por parte de alumnos y profesores. Nuestro objetivo es motivar a los alumnos de forma que sean ellos los que demanden los conceptos teóricos con los que resolver una situación práctica. Son varios los objetivos que nos planteamos:

- Intentamos promover el trabajo autónomo y autoaprendizaje del alumno
- Intentamos que el aprendizaje de los alumnos sea eficaz y estimulante, éstos han de querer *aprender a aprender* estadística
- Pretendemos que los alumnos adviertan la relación entre distintas materias de su “currículo”, fomentando la interdisciplinariedad

3. Bibliografía

- [1] Batanero, C. (2001). *Didáctica de la estadística. Grupo de Investigación en Educación Estadística.* Granada (Spain) (available in <http://www.ugr.es/local/batanero>).
- [2] García Leal, J; Lara Porras A.; Navarrete Álvarez, E.; Rico Castro, N.; Román Montoya, Y.; Ruiz Castro, J.E. and Sánchez Borrego, I. B. (2004). *Estadística para la Biología.* Granada (Spain)
- [3] García Leal, J; Lara Porras A.; Navarrete Álvarez, E.; Rico Castro, N.; Román Montoya, Y.; Ruiz Castro, J.E. and Sánchez Borrego, I. B. (2003b). *Técnicas Estadísticas para Investigación en Biología.* Granada (Spain).
- [4] Lara Porras, A.M.. (2002). “Estadística para Ciencias Biológicas y Ciencias Ambientales. Problemas y Exámenes resueltos”. Proyecto Sur. Granada (España).
- [5] Lara Porras, A; Román Montoya, Y.; García Leal, J. (2005). “*¿Qué respuestas me da la Estadística?*”. IX Congreso de Metodología de las Ciencias Sociales y de la Salud.

Comparación de modelos para el espectro poblacional

P. Saavedra, C.N. Hernández, J. Artiles, A. Santana, I. Luengo

Departamento de Matemáticas. Universidad de Las Palmas de Gran Canaria,
e-mail: jartiles@dma.ulpgc.es

Abstract

Para estimar la función de densidad espectral de procesos lineales con coeficientes aleatorios pueden utilizarse varios modelos de estimación alternativos. Por medio de un estudio de simulación comparamos el estimador paramétrico propuesto en [2] con dos estimadores no paramétricos basados, respectivamente, en suavizado de núcleo y en splines.

Claves: Procesos lineales de coeficientes aleatorios, espectro poblacional, estimación de núcleo.

AMS: Clasificación AMS (Opcional)

1. Introducción

El análisis de diferentes conjuntos de series temporales es bastante común en el campo de la biomedicina. Cada conjunto de series se observa sobre un sujeto perteneciente a una muestra aleatoria que se elige de una población determinada. Estas series temporales son generalmente no homogéneas en el sentido de que no son generadas por el mismo patrón. Diggle y Al Wasel en [1] analizan los niveles de la hormona LH en la sangre de una muestra de sujetos de una población, observando que estas series temporales no pueden considerarse realizaciones sólo de procesos estocásticos lineales estacionarios, por lo que proponen un modelo de efectos aleatorios basado en la representación asintótica del periodograma para procesos lineales generales. El modelo involucra un parámetro llamado espectro poblacional, una componente aleatoria específica de cada sujeto y un término relativo a los residuos de cada periodograma. Hernández et al. [3] usando un modelo más general estiman el espectro poblacional por medio del bootstrap y analizan la consistencia de este método usando la métrica de Mallows. Por su parte, Saavedra et al. [5] presentan algunos resultados teóricos sobre procesos estacionarios doblemente estocásticos que permiten analizar un conjunto de series temporales replicadas en el dominio de las frecuencias.

En este trabajo consideramos en particular la clase de procesos lineales que se caracterizan por tener coeficientes aleatorios. Denominamos *procesos lineales doblemente estocásticos* a esta clase de procesos. Cuando se selecciona al azar un conjunto de coeficientes, queda seleccionada también la densidad espectral correspondiente al proceso lineal resultante. De esta forma, cada función de densidad espectral puede considerarse como una realización de un proceso

estocástico cuya función de medias es el llamado espectro poblacional. Para estimar esta función a partir de un conjunto de series temporales generadas por tal proceso pueden utilizarse varios modelos de estimación alternativos. Por medio de un estudio de simulación comparamos el estimador paramétrico propuesto en [2] con dos estimadores no paramétricos basados, respectivamente, en suavizado de núcleo y en splines.

2. Discusión

Para comparar la eficiencia de los distintos estimadores hemos utilizado el valor del MISE (error cuadrático medio integrado), calculado a partir de diversas simulaciones realizadas con el paquete estadístico R [4]. A partir del estudio realizado observamos que en general los estimadores no paramétricos tienen un buen comportamiento en todas las condiciones. En el caso particular de que se conozca la verosimilitud asociada al modelo que genera los datos, la estimación paramétrica puede resultar ventajosa, pero en la práctica esta circunstancia es poco habitual, y la verosimilitud debe inferirse a partir de los datos disponibles. En ese caso el estimador paramétrico resulta en general inferior a los estimadores no paramétricos propuestos.

3. Bibliografía

- [1] Diggle, P. J. Al-Wasel, I. 1993, On periodogram-based spectral estimation for replicated time series, in S. Rao, ed., 'Developments in Time Series Analysis', Chapman and Hall, Great Britain, pp. 341–354.
- [2] Diggle, P. J. and Al-Wasel, I. 1997, 'Spectral analysis of replicated biomedical time series', *Appl. Statist.* **46**, 31–71.
- [3] Hernández-Flores, C., Artiles-Romero, J. and Saavedra-Santana, P. 1999, 'Estimation of the population spectrum with replicated time series', *Comp. Stat. and Data Anal* **30**, 271–280.
- [4] R Development Core Team 2003, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria. url: <http://www.R-project.org>
- [5] Saavedra, P., Hernández, C. and Artiles, J. 2000, 'Spectral analysis with replicated time series', *Communications in Statistics, Theory and Methods* **29**, 2343–2362.

Competencias y estrategias en la inserción laboral de los titulados universitarios

R.M. Crujeiras Casais¹, P. Faraldo Roca¹, B.M. Fernández de Castro¹, C.L. Iglesias Patiño¹, M.J. Lombardía Cortiña¹, J.M. Matías Fernández², S. Naya Fernández³.

¹rcrujeir@usc.es, Departamento de Estadística e Investigación Operativa,
Universidade de Santiago de Compostela

²Departamento de Estatística e Investigación Operativa, Universidade de Vigo

³Departamento de Matemáticas, Universidade da Coruña

Abstract

En este trabajo se presenta un análisis de los aspectos valorados en la contratación y las competencias demandadas por las empresas/instituciones vinculadas al Programa de Prácticas del Consello Social de la Universidad de Santiago de Compostela.

Claves: Competencias, inserción laboral.

1. Introducción

La evaluación de las nuevas competencias, capacidades y habilidades que necesitan los titulados universitarios como trabajadores cualificados para que su inserción laboral sea satisfactoria, es uno de los elementos claves para contrastar la eficacia de la formación universitaria para la exigencias del mercado de trabajo.

En este trabajo se ha obtenido, a través de encuestas a empresas/instituciones del Programa de Prácticas del Consello Social de la Universidad de Santiago de Compostela, la visión de los empleadores, agentes externos a la universidad. Se pretenden conocer las competencias que los titulados universitarios deben poseer para su incorporación al mercado laboral, y que deben ser adquiridas en el sistema educativo así como qué factores son determinantes en el momento de la contratación.

Se presenta un análisis de la valoración de las competencias y los aspectos valorados en la contratación, por grupos de empresas/instituciones atendiendo a características relevantes de las mismas. La información derivada del análisis de estos datos puede utilizarse para mantener o mejorar la capacidad de los títulos universitarios en cuanto a la preparación que proporcionan para la inserción en el mercado de trabajo.

2. Agradecimientos

Este trabajo ha sido promovido por el Consello Social da Universidad de Santiago de Compostela.

3. Bibliografía

- [1] *Incorporación do titulado universitario ao mercado de traballo. Visión den de a empresa.* Consello Social. Universidade de Santiago de Compostela. (2006).
- [2] *Inserción Laboral dos Titulados no Sistema Universitario de Galicia 2001-2003.* ACSUG. (2005).
- [3] *La universidad y la empresa española.* Fundación CYD. (2005).
- [4] *Estudio da Inserción Laboral dos Titulados do Sistema Universitario de Galicia. 1996-2001.* ACSUG.(2004).
- [5] *Educació superior i treball a Catalunya.* AQU. (2003).

Complejidad Computacional de Clasificadores Estadísticos y Neuronales

Andrés Jiménez Jiménez¹, José María Gutiérrez Pérez², Angel Berihuete Macias²

¹ andres.jimenez@uca.es, CITI., Univ. Cádiz

²josema.gutierrez@uca.es, Dpto. de Estadística e IO, Univ.

³Doctorando del Dpto. de Estadística e IO, Univ. Cádiz

Abstract

Se aborda experimentalmente la complejidad de los clasificadores estadísticos y neuronales, considerándolos como si en ambos casos fuesen algoritmos probabilísticos. La experiencia se orienta a la búsqueda de modelos de regresión que ajusten suficientemente bien los tiempos de ejecución de distintos clasificadores, en función de la dimensión del problema, el numero de datos y los parámetros de cada uno de ellos.

Palabras Clave: Complejidad Computacional, Clasificación

1. Modelado de los tiempos de computo

Al igual que ocurre con los algoritmos determinísticos, se supondrá que el orden de complejidad de un clasificador ya sea neuronal o estadístico, depende del tamaño del problema: la dimensión de la muestra y el número de datos. Por tanto, es posible obtener una idea del comportamiento general de un clasificador repitiendo el proceso de clasificación para los distintos valores de los citados parámetros, además de los propios de cada clasificador y ajustando adecuadamente la muestra obtenida.

Los modelos de tiempo de cómputo de los clasificadores estadísticos en función de la dimensiones muestrales se obtuvieron mediante ajustes con regresión polinómica del tipo

$$t = a_0 + a_1 x + a_2 y + a_3 x^2 + a_4 y^2 + a_5 xy \quad \text{donde } x = \dim(\mathbf{X}), y = \text{card}(\mathbf{m}).$$

Algunos ejemplos de los resultados obtenidos se exponen en la tabla y figuras que siguen. La tabla muestra el ajuste del tiempo de cómputo de cinco clasificadores conocidos, el valor del estadístico F asociado con análisis de la varianza para contrastar la idoneidad de los modelos, y el valor del coeficiente de determinación ajustado, R_{aj}^2 para establecer la bondad de cada ajuste. Todos los coeficientes que aparecen junto con F son significativos para el nivel $\alpha = 0,05$.

Modelo	F (Anova)	R_{aj}^2	$\hat{t} = f(x, y)$
ADL	1012	0.85	$\hat{t} = -1,89 - 3,33 \cdot 10^{-2}x + 5,27 \cdot 10^{-3}x^2 + 3,54 \cdot 10^{-3}y$
3-NN	291575	0.99	$\hat{t} = -5,06 + 4,34 \cdot 10^{-3}xy$
5-NN	249701	0.99	$\hat{t} = -4,31 + 4,10 \cdot 10^{-3}xy$
Bayes	5800	0.94	$\hat{t} = -0,57 + 1,96 \cdot 10^{-3}y + 1,10 \cdot 10^{-3}x^2$
ADR	6325	0.94	$\hat{t} = 3,19 \cdot 10^{-3}x^2 + 9,01 \cdot 10^{-8}xy$

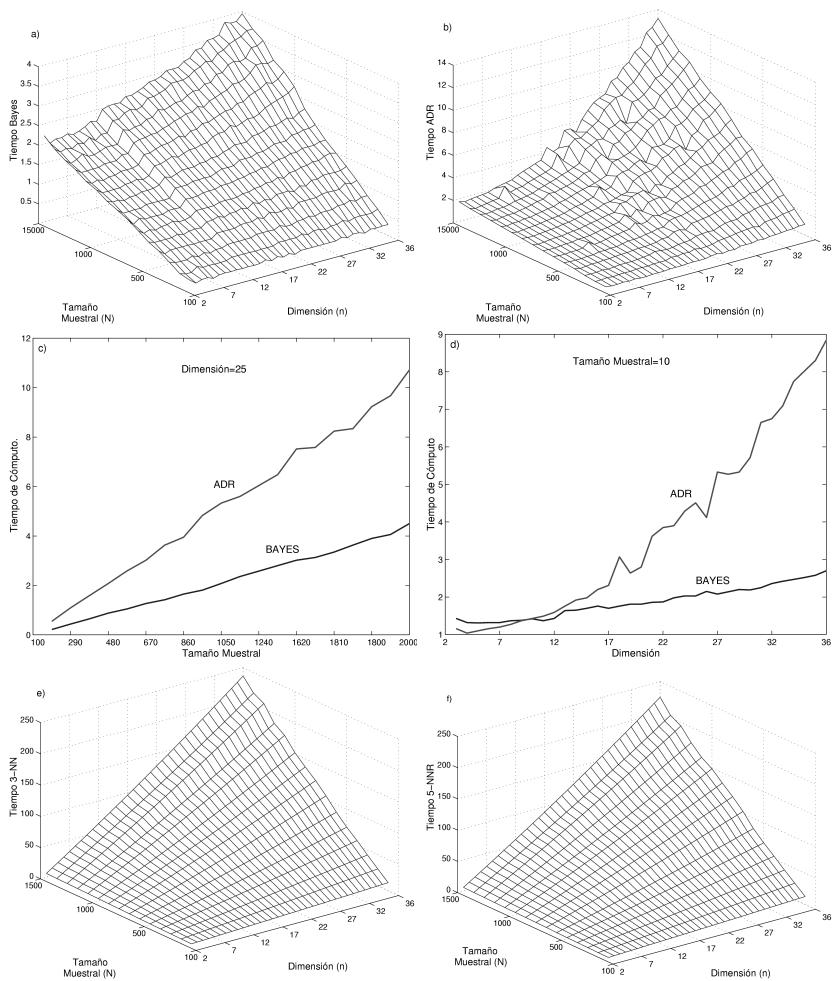


Figura 1: Comparativa de los tiempos de cómputo de distintos clasificadores.

Comportamiento variado del consumidor: un análisis bayesiano dinámico del escenario de compra

Carmen Berne¹, Pilar Gargallo², Noemí Martínez³, Manuel Salvador⁴

¹ cberne@unizar.es, ³noemar@unizar.es

Departamento de Economía y Dirección de Empresas, Universidad de Zaragoza

²pigarga@unizar.es, ⁴salvador@unizar.es

Departamento de Estructura e Historia Económica, Universidad de Zaragoza

Abstract

En este trabajo se plantea un modelo estadístico bayesiano dinámico con el fin de explorar las relaciones entre el comportamiento de compra de los hogares españoles y las características socio-demográficas que lo determinan. El modelo analiza la variación mensual del escenario de compra distinguiendo entre compradores leales a un único establecimiento de aquellos que muestran un comportamiento variado.

Palabras Clave: Comportamiento Variado, Escenario de Compra, Modelos Bayesianos Dinámicos

AMS: 62F15, 62M10, 91B42

1. Introducción

Durante las tres últimas décadas, el comportamiento variado de los consumidores ha atraído la atención de los investigadores de marketing y, en concreto, de aquellos interesados en el estudio del comportamiento del consumidor.

En este sentido, existen diversos trabajos que postulan que, cuando existen alternativas disponibles, las familias alternan y/o complementan la compra en su establecimiento habitual con compras en otros establecimientos (i.e. [1],[2]). Se sigue así un modelo de compra que incluye un componente denominado comportamiento variado del consumidor, el cual debe ser analizado y gestionado en la medida en que cuanto más presente está en un consumidor, más relevante es la parte del presupuesto total de la unidad de consumo no destinada al gasto en el establecimiento habitual o principal de compra. Según [2], aunque los consumidores resuelven la compra de alimentación y productos para el hogar a menudo en varias tiendas, generalmente realizan la mayor parte de sus compras en el denominado ‘establecimiento principal’. A los distribuidores minoristas les interesa especialmente que su establecimiento sea el preferido por los consumidores ya que, según [3], a pesar de que los compradores fieles suelen gastar, en términos generales, una parte más pequeña del presupuesto

mensual que los cambiadores (*switchers*), los clientes leales gastan el doble en su establecimiento principal de compra. Es una cuestión de reparto de gasto.

El objetivo de este trabajo se concreta en estudiar los determinantes de la variación en el tiempo del escenario de compra habitual de la unidad familiar (construido teniendo en cuenta los diferentes establecimientos que son utilizados para resolver la compra de productos de uso frecuente) y de la lealtad a los establecimientos que forman parte de dicho escenario. A tal fin se construye un modelo estadístico bayesiano dinámico que analiza la variación mensual del escenario de compra distinguiendo entre compradores leales a un único establecimiento de aquéllos que muestran un comportamiento variado. Para ello se plantea un modelo en dos fases: la primera analiza los determinantes de la decisión de compra mientras que la segunda permite discriminar entre el tipo de comprador. Debido a la complejidad del análisis planteado, la estimación de los parámetros del modelo se realiza mediante métodos MCMC utilizando técnicas de aumento de datos ([4]).

Tomando como punto de partida la revisión de la literatura, el estudio empírico partirá de la formulación de una serie de hipótesis de trabajo que serán contrastadas, mediante la utilización de la adecuada metodología, sobre una base de datos del panel de consumidores de AC Nielsen.

2. Bibliografía

- [1] Kahn, B. E. and McAlister, L. (1997). *Grocery Revolution: The New Focus on the Consumer*, Reading, MA: Addison Wesley, Longman
- [2] Rhee, H. and Bell, D. R. (2002). The inter-store mobility of supermarket shoppers, *Journal of Retailing* 78 (4), 225-237.
- [3] Knox S. D. and Denison, T. J. (2000). Store Loyalty: Its Impact on Retail Revenue - An Empirical Study of Purchasing Behaviour in the UK, *Journal of Retailing and Consumer Services* 7 (January), 33-45.
- [4] Rossi, P.; Allenby, G.M. and McCulloch, R. (2005). *Bayesian Statistics and Marketing*. Wiley Series in Probability and Statistics.

Construcción de un observador en un modelo depredador-presa

I. López¹, M. Gámez², J. Garay³, Z. Varga⁴

¹milopez@ual.es, ²mgomez@ual.es

Departamento de Estadística y Matemática Aplicada, Universidad de Almería

³ garayj@elte.ludens.hu, Department of Plant Taxonomy and Ecology,
Ecological and Theoretical Research Group of Hungarian Academy of Science,

L. Eötvös University, Budapest, Hungría

⁴ Varga.Zoltan@gek.szie.hu, Institute of Mathematics and Informatics, Szent
István University, Godollo, Hungría

Abstract

Para un sistema de poblaciones dado, la necesidad de construir un sistema auxiliar (observador), cuya solución tiende a la del sistema original, surge cuando a partir de la observación del estado de alguna/s de estas poblaciones, queremos reconstruir el estado de todo el sistema en función del tiempo, que no es tarea fácil. Construiremos un observador para un modelo depredador-presa.

Palabras Clave: Sistema de poblaciones, observabilidad, sistema observador

1. Análisis del problema

El problema del monitoraje en un ecosistema con varias poblaciones surge cuando observando únicamente el estado de una o varias de ellas queremos reconstruir el estado de todo el sistema en función del tiempo. Ello puede resultar conveniente, ya que en ocasiones la observación directa del estado de todas las poblaciones que se consideran en el ecosistema puede resultar complicada o costosa. Para ello buscaremos un sistema auxiliar llamado *observador* cuya solución tiende a la del sistema original.

Un concepto importante para resolver este problema es el de observabilidad: “*a partir de la observación de una o varias variables se puede recuperar el proceso de estado de todas las variables de la población, de manera única, sin determinar un método constructivo para obtener dicho proceso.*”

En este trabajo usaremos una condición general suficiente para la observabilidad de sistemas dinámicos no lineales que fue probada por Lee y Markus (1971). Esta condición también se aplicó a varios modelos en poblaciones genéticas y a modelos evolucionarios dependientes de las frecuencias en López, Gámez y Carreño (2004). Además en Varga, Scarelli y Shamandy (2003) se hizo un estudio sobre observabilidad en modelos Lotka-Volterra. La observabilidad en ecosistemas simples también ha sido objeto de estudio en Shamandy (2005).

Finalmente, para la construcción del observador de un modelo depredador-presa consideraremos el estudio geométrico del diseño de un observador local para sistemas no lineales que llevó a cabo Sundarapandian (2002).

2. Agradecimientos

Esta publicación ha sido realizada en el marco de la colaboración científica y tecnológica húngaro-española intergubernamental, con el soporte del Fondo de Innovación Tecnológica y de Investigación (de Hungría) y del MEC (de España), y también ha sido soportada por el Fondo Nacional Húngaro de Investigación (OTKA T 037271) y por parte de József Garay también por el Konrad Lorenz Institute for Evolution and Cognition (Austria).

3. Bibliografía

- [1] Kalman, R. E., Falb, P. L., Arbib, M. A., (1969). *Topics in Mathematical System Theory*. McGraw-Hill, New York.
- [2] Lee, E.B., Markus, L., (1971). *Foundations of Optimal Control Theory*. Wiley, New York.
- [3] López, I., Gámez, M. y Carreño, R. (2004). Observability in dynamic evolutionary models. *Biosystems*, 73, 99-109.
- [4] Markus, L. (1965). Controllability of Nonlinear Processes. *J. SIAM Control* 3, 78-90.
- [5] Shamandy, A. (2005). Monitoring of trophic chains. *Biosystems*, 81, (1), 43-48.
- [6] Sundarapandian, V. (2002). Local observer design for nonlinear systems. *Mathematical and computer modelling* 35, 25-36.
- [7] Varga, Z., Scarelli, A., Shamandy, A. (2003). State monitoring of a population system in changing environment. *Community Ecology*, 4, (1) 73-78.

Contenidos de Probabilidad en el entorno CDPYE

M. J. García-Ligero Ramírez, A. Hermoso Carazo, P. López García, J. A. Maldonado Jurado, P. Román Román, F. Torres Ruiz

{mjgarcia, ahermoso, jamaldo, proman, fdeasis}@ugr.es

Departamento de Estadística e Investigación Operativa

Universidad de Granada

Abstract

Se presenta la mejora e inclusión de contenidos de probabilidad en el entorno CDPYE (entorno interactivo de apoyo docente en el estudio del Cálculo de Probabilidades y Estadística Matemática para alumnos de primer ciclo).

Claves: Didáctica, Probabilidad, Entorno interactivo.

1. Introducción

En el anterior Congreso de la SEIO presentamos un entorno interactivo de apoyo docente en el estudio del Cálculo de Probabilidades y Estadística Matemática para alumnos de primer ciclo [1]. Dicho entorno interactivo hipertextual permite la navegación entre sus módulos con interrelaciones entre temas y asignaturas e incluye además presentaciones multimedia (simulaciones de explicaciones en pizarra), programas interactivos que permiten realizar aplicaciones a problemas concretos, así como la posibilidad de que el alumno pueda comprobar de forma interactiva los conocimientos adquiridos, bien mediante problemas desarrollados paso a paso en los que se le incentiva por medio de preguntas hacia el siguiente paso que debe seguir en la resolución de un problema, o bien a través de autoevaluaciones de cada tema o globales de cada asignatura.

En esta comunicación presentamos la evolución de nuestro trabajo en dos sentidos: por una parte, la mejora del propio entorno y, por otra, la inclusión de contenidos de Cálculo de Probabilidades en él.

En cuanto a la primera cuestión, se ha prestado especial atención a la funcionalidad del entorno, y a su independencia del dispositivo, restringiendo las necesidades de adecuación a aquellas consideradas de habitual uso o, en cualquier caso, abiertas a software libre disponible.

En cuanto a la inclusión de contenidos en el entorno, se han abordado los temas referentes a probabilidad, probabilidad condicionada, variable aleatoria y distribuciones de probabilidad unidimensionales. Destacamos, entre otros, aplicaciones que permiten al alumno definir una variable aleatoria y solicitar el

cálculo detallado de características asociadas a ella, el desarrollo de problemas de urnas en distintas etapas mediante diagramas de árbol; presentaciones multimedia de problemas de cambio de variable, manejo de tablas de distribuciones y demostraciones con largos desarrollos como, por ejemplo, la referente a que la mediana minimiza la media de las desviaciones absolutas respecto de un valor.

2. Agradecimientos

Este trabajo ha sido financiado por el Vicerrectorado de Planificación, Calidad y Evaluación Docente de la Universidad de Granada, a través del Proyecto de Innovación "Elaboración de material docente sobre Probabilidad para su inclusión en el entorno CDPYE".

3. Bibliografía

- [1] Cerdá, M. V., García-Ligero, M. J., Hermoso, A., Maldonado, J. A., Román, P. y Torres, F. (2004). CDPYE: un entorno interactivo de apoyo a la docencia del Cálculo de Probabilidades y la Estadística Matemática. En: Actas del XXVIII Congreso Nacional de Estadística e Investigación Operativa. Ref. 296. ISBN 84-609-0438-5.

Contraste robusto de cointegración frente a no linealidades

Alvaro Escribano¹, Ana E. García² y M^a Teresa Santos³

¹Departamento de Economía, Universidad Carlos III de Madrid.

²Departamento de Estadística e Investigación Operativa , Universidad Rey Juan Carlos de Madrid. anaelizabeth.garcia@urjc.es

³ Departamento de Estadística, Universidad de Salamanca.

Abstract

En este artículo proponemos un contraste no paramétrico para detectar la presencia de cointegración cuando la relación entre las series es no lineal. En dicho contraste el estadístico es construido a partir de funciones de estadísticos de orden, haciendo un estudio comparativo con los tradicionales contrastes.

Claves: Cointegración, Estadísticos de orden inducido y Contrastos de Engle y Granger.

1. Introducción

A partir de que Granger (1981) introduce el concepto de cointegración se han podido explicar de forma más detallada la mayoría de las relaciones existentes entre series temporales. Muchos han sido los autores que han estudiado este concepto proponiendo diversos contrastes como son Engle y Granger (1987) (EG), Johansen (1988,1991) entre otros. Sin embargo, cabe la posibilidad de que existan en la práctica relaciones de largo plazo tanto no lineales como variantes en el tiempo para las cuales dichos contrastes no son robustos, de tal manera que una inadecuada especificación del modelo puede dar lugar a que estas aparezcan como no cointegradas. Diversos estudios ofrecen evidencia empírica en favor de este tipo de relaciones. Nos centraremos en los métodos no paramétricos para contrastar relaciones de cointegración bajo no linealidad. La principal ventaja es que no imponen restricciones a la forma de la relación, ni la especificación de modelo alguno para las series de partida. Con esta metodología en lugar de trabajar con las variables actuales, lo haremos con los estadísticos de orden inducido.

2. Contraste de cointegración inducido

Para una serie temporal de tamaño n , es decir, x_1, \dots, x_n , los estadísticos de orden de x_t vienen dados por $x_{1,n} \leq \dots \leq x_{n,n}$ obtenidos después de una permutación de los índices $1, \dots, n$ tales que $x_{i,n} \leq x_{i+j,n}, \forall j > 0$. Dadas dos series temporales x_t, y_t , se define para la serie Y el estadístico de orden inducido basado en el orden de x_t como $\hat{y}_{i,n} = y_j$ si $x_{i,n} = x_j$. Definimos el siguiente

estadístico

$$C1 = \sup_{j=1,n} |\hat{F}_Y^{(n)}(\hat{y}_{l_s(j),n}) - \hat{F}_Y^{(n)}(y_{j,n})|,$$

donde $\hat{F}_Y^{(n)}(y)$ es la función de distribución empírica obtenida de una muestra de tamaño n de la serie temporal y_t , es decir, $\hat{F}_Y^{(n)}(y) = n^{-1} \sum_{t=1}^n 1(y_t \leq y)$,

donde $1(\cdot)$ es la función indicatriz; s es el signo del parámetro de cointegración, y $l_s(j)$ es igual a j o $n-j$, dependiendo de si $s = 1$ o $s = -1$, respectivamente. Consideraremos como hipótesis nula dos paseos aleatorios independientes y dentro de la alternativa diversos casos de cointegración tanto lineal como no lineal, donde el proceso generador de datos (PGD) es un modelo de corrección de error (MCE). Cuando el estadístico tome valores pequeños rechazamos la hipótesis nula por tanto estamos ante un contraste unilateral por la izquierda para discriminar entre cointegración y no cointegración.

3. Potencia del contraste

Consideramos el siguiente PGD donde el modelo de corrección de error MCE es lineal y la relación de cointegración es no lineal. $PGD : H_1 \Delta y_t = c + a \Delta x_t + b(y_{t-1} - g(x_{t-1}, \alpha)) + w_{1t}$, $\Delta x_t = w_{2t}$ donde w_{1t} y w_{2t} son Normales estandar mutuamente independientes, con $\alpha = 1$ y $a = 0$. En este ejemplo $g(z_{t-1}, \alpha) = z_{t-1}^j$ con 10,000 simulaciones de Monte Carlo, analizamos la potencia del contraste de orden inducido COI con un nivel de significación del 5% para $b = -0,05$, comparando los resultados con los de EG test, cuyos resultados se muestran en la siguiente tabla 1.

$b = -0,05$

$j \setminus n$	100		250		500	
	COI	EG	COI	EG	COI	EG
2	0.83	0.08	0.94	0.02	0.98	0.01
3	0.94	0.1	1	0.03	1	0.008
4	0.92	0.08	1	0.01	1	0.006

Cuadro 1:

4. Bibliografía

- [1] Granger C. (1981). *Some properties of Time Series Data and their use in Econometric Model Specification*. Journal of Econometrics, 16, 121-130.

Indicadores asociados a porcentajes de incidencias con series cortas en periodo base

Juan José Barreiro¹, Miquel Salicrú², Carlos González³

¹jjbi@uvigo.es, Dep. de Economía Aplicada, Universidad de Vigo

²msalicru@ub.edu, Dep. de Estadística, Universidad de Barcelona

³cgzlez@uvigo.es, Dep. Estadística e Inv. Operativa, Univ. de Vigo

Abstract

Obtenemos una corrección del estadístico estandarizado que mejora la eficiencia en el control de procesos caracterizados por el porcentaje de incidencias.

1. Introducción

En organizaciones dedicadas a servicios, la satisfacción y/o el coste operativo viene relacionado –en muchas ocasiones– con el comportamiento de indicadores asociados a porcentajes de incidencias. En este contexto, y debido a la importancia estratégica del indicador, son frecuentes acciones periódicas que provocan cambios en los parámetros de proceso. En este contexto, para procesos estables en el tiempo con porcentaje de incidencias p y muestras escogidas en el periodo base:

Número de muestra	1	2	s
Tamaño de muestra	m_1	m_2	m_s
Porcentaje incidencias	\hat{p}_1	\hat{p}_2	\hat{p}_s

la teoría clásica considera el estadístico estandarizado,

$$Z_i = \frac{n_i^{1/2}[\hat{p}_i - \hat{p}]}{[\hat{p}(1 - \hat{p})]^{1/2}} \quad \text{siendo} \quad \hat{p} = \sum_{j=1}^s \frac{m_j \hat{p}_j}{\sum_{k=1}^s m_k}$$

al que asocia a una ley $N(0,1)$ y establece los límites de control: $LCS(I) = \pm n_{\epsilon/2}$.

Cuando el número de muestras considerado en el periodo base es reducido y se mantiene el tamaño de muestra en el seguimiento del proceso, la estimación de p es de varianza apreciable. Por este motivo, ensayamos una corrección del estadístico estandarizado y realizamos un estudio de simulación para poner de manifiesto el interés de la corrección propuesta.

2. Distribución asintótica del estadístico estandarizado

De considerar el desarrollo de Taylor del estadístico $Z_i = Z_i(\hat{p}_i, \hat{p})$ en el punto (p, p) , la convergencia a la ley normal del término:

$$\left(\frac{n_i n^*}{n_i + n^*} \right)^{1/2} [\hat{p}_i - p, \hat{p} - p] \approx N((0, 0), \Sigma)$$

siendo

- Σ la matriz: $\sigma_{11}^2 = \lambda p(1-p)$, $\sigma_{12}^2 = \sigma_{21}^2 = 0$ y $\sigma_{22}^2 = (1-\lambda)p(1-p)$
- $n^* = (\sum_{k=1}^s m_k)^{-1}$ y $\lambda = \lim \frac{n_i^+}{n_i+n^*}$
y la convergencia a cero del resto de orden 2, se tiene el siguiente resultado:

Teorema 2.1. El estadístico corregido

$$Z_i^c = \frac{n_i^{-1/2}(\hat{p}_i - \hat{p})}{[\hat{p}(1 - \hat{p})]^{1/2}} \cdot \left[1 + \frac{n_i}{\sum_{k=1}^s m_k} \right]^{-1/2}$$

se distribuye asintótica de acuerdo con una ley normal $N(0, 1)$.

3. Estudio de simulación

Para contrastar la eficiencia del modelo estándar con la corrección propuesta, hemos realizado un estudio de simulación con:

- *Variable respuesta:* ARL.
- *Condiciones de experimentación*

1. p recorre el conjunto 0.01; 0.005; 0,001; 0.0005; 0.0001.
2. s recorre el conjunto 6; 12; 24
3. Los tamaños de muestras en periodo base y en proceso de control:
 - a) $m_1=m_2=\dots=m_s=n_i=\alpha^{-1}$
 - b) $m_i=n_i=0.8\cdot\alpha^{-1}$ si $i = \dot{2} + 1$; $m_i=n_i=1.2\cdot\alpha^{-1}$ si $i = \dot{2}$
 - c) $m_i=n_i=0.8\cdot\alpha^{-1}$ si $i = \dot{3} + 1$; $m_i=n_i=\alpha^{-1}$ si $i = \dot{3} + 2$; $m_i=n_i=1.2\cdot\alpha^{-1}$ si $i = \dot{3}$
4. α recorre el conjunto p^{-1} ; $5p^{-1}$; $10p^{-1}$; $20p^{-1}$; $30p^{-1}$; $40p^{-1}$; $50p^{-1}$; $60p^{-1}$
5. Estadísticos evaluados: Z_i y Z_i^c con límites de control $LCS(I) = \pm n_{\epsilon/2}$

- *Repeticiones:* 100 series de s valores y 100.000 valores en proceso.

4. Algunos resultados

	$m_i=n_i=\alpha$	$m_i,n_j=0.8\cdot\alpha^{-1}, 1.2\cdot\alpha^{-1}$	$m_i,n_j=0.8\cdot\alpha^{-1}, \alpha^{-1}, 1.2\cdot\alpha^{-1}$		
	Z_i	Z_i^c	Z_i	Z_i^c	Z_i
$\alpha=p^{-1}$	47.9	53.4	32.4	40.6	41.4
$\alpha=10p^{-1}$	117.9	190.3	135.5	223.8	120.2
$\alpha=30p^{-1}$	176.8	345,3	163,2	295,1	165,7
$\alpha=60p^{-1}$	136.3	261.3	194.7	394.4	168.0
					305.3

Tabla 1. Valores de ARL para $s=6$ y $p=0.005$

Covariance Changes Detection in Multivariate Time Series

Pedro Galeano¹, Daniel Peña²

¹pgaleano@usc.es, Departamento de Estadística e Investigación Operativa,
Universidad de Santiago de Compostela

²daniel.pena@uc3m.es, Departamento de Estadística, Universidad Carlos III
de Madrid

Abstract

This paper studies the detection of step changes in the variances and in the correlation structure of the components of a vector of time series. Two procedures based on the likelihood ratio test (LRT) statistic and on a cumulative sums (cusum) statistic are considered and compared in a simulation study. We conclude that for a single covariance change the cusum procedure is more powerful in small and medium samples whereas the likelihood ratio test is more powerful in large samples. However, for several covariance changes the cusum procedure works clearly better. The procedures are illustrated in a real data examples.

Keywords: Cusum test; Likelihood ratio test; Step Changes.

AMS: AMS Classification (62M10)

1. Introduction

The problem of detection of a sudden change in the marginal variance of a univariate time series has been extensively studied. For instance, Tsay (1988) studied outliers, level shifts and variance changes in ARIMA models. Inclán and Tiao (1994) proposed an iterative procedure based on a cumulative sum of squares statistic for the detection of several variance changes in Gaussian independent observations. The cusum statistic they proposed for variance changes was related to similar statistics studied for level shifts by Andrews (1993) and Bai and Perron (1998). Chen and Gupta (1997) considered an information theoretic approach based on the Bayesian Information Criteria (BIC) for detection of step changes. Lee and Park (2001) extended the Inclán and Tiao approach to moving average models.

The case of multivariate sequences has, to the best of our knowledge, not been considered yet. The multivariate case is more challenging than the univariate case because then we can consider changes in both the variances of the components and in the correlation structure among them. The effect of a change in the covariance structure depends on the magnitude of the change and on the parameters of the model. In fact, if we look for variance changes in each of the components of the series, we can find that several of these variance changes

can be explained by a single covariance change. In this article we study the detection of step changes in the variance and in the correlation structure of the components of a vector autoregressive moving average (VARMA) model. Two approaches are introduced and compared. The first one is a likelihood ratio approach and the second is a cusum approach.

The organization of the talk is as follows. First, we present the model for covariance changes and two statistics that can be used for testing for the change. Then we consider the particular case of changes only in the variances of the components, and present two statistics for testing for such a change. Next we study two different procedures for detection and estimation of these changes, which are compared in a Monte Carlo experiment for different models, sample sizes, number of changes and location of the changepoints. Finally, we illustrate the procedures by means of two real data examples. We conclude that the procedure based on the cusum statistic has an overall better performance than the one based on the likelihood ratio test, whereas for a single change and large sample size the likelihood ratio test appears to be more powerful.

2. Acknowledgement

We acknowledge the financial support by project SEJ2004-03303 and by Fundación BBVA, Spain. The first author also acknowledge financial support by Xunta de Galicia under the Isidro Parga Pondal Program

3. Bibliography

- [1] Andrews, D. W. K. (1993) Tests For Parameter Instability and Structural Change With Unknown Change Point. *Econometrica*, 61, 821-856.
- [2] Bai, J. and Perron, P. (1998) Estimating and Testing Linear Models With Multiple Structural Changes. *Econometrica*, 66, 47-78.
- [3] Chen, J. and Gupta, A. K. (1997). Testing and Locating Variance Changepoints with Application to Stock Prices. *Journal of the American Statistical Association*, 92, 739-747.
- [4] Inclán, C. and Tiao, G. C. (1994) Use of Cumulative Sums of Squares for Retrospective Detection of Changes of Variance. *Journal of the American Statistical Association*, 89, 913-923.
- [5] Lee, S. and Park, S. (2001) The Cusum of Squares Test for Scale Changes in Infinite Order Moving Average Models. *Scandinavian Journal of Statistics*, 28, 625-644.
- [6] Tsay, R. S. (1988) Outliers, Level Shifts and Variance Changes in Time Series. *Journal of Forecasting*, 7, 1-20.

D- y c-optimalidad para un modelo binomial negativo de regresión

Carmelo Rodríguez Torreblanca¹, Juan M. Rodríguez Díaz²

¹ Dpto de Estadística y Mat. Aplicada, Universidad de Almería *crt@ual.es*

² Dpto de Estadística, Universidad de Salamanca *juanmrod@usal.es*

Abstract

Este trabajo se ocupará del problema de estimación del parámetro que mide la probabilidad de ocurrencia del suceso para una variable respuesta Binomial Negativa, y cuya forma responde a la de un modelo de regresión exponencial biparamétrico. Se obtendrán diseños D- y c-óptimos para estos casos.

Claves: D y c-optimización, Distribución binomial negativa, Modelo de regresión exponencial

1. Introducción

Los modelos de regresión exponencial son modelos no lineales ampliamente utilizados en diversos campos experimentales como la agricultura, biología y química. En el caso unidimensional, la forma más general de estos modelos es

$$Y = ae^{-\beta x} + \epsilon,$$

donde a , β son constantes reales, siendo β positiva, y la variable independiente x es elegida en un intervalo de diseño de la recta real $\mathcal{X} = [x_{\min}, x_{\max}]$. El caso en que el error ϵ se supone distribuido normalmente con media cero y varianza constante σ^2 ya ha sido tratado anteriormente desde el punto de vista del diseño óptimo de experimentos. Por ejemplo, el diseño localmente D-óptimo no depende del valor del parámetro a y pone masa $\frac{1}{2}$ en cada uno de los puntos $x_1 = x_{\min}$ y $x_2 = \min\{x_{\max}, x_{\min} + \frac{1}{\beta}\}$ (Han y Chaloner, 2003).

2. Otros Modelos

Cuando se supone otra estructura distinta para la varianza de los errores nos estamos fijando en otros modelos. Por ejemplo, el modelo

$$Y = ae^{-\beta x} + \epsilon, \quad \text{con} \quad \text{Var}(Y|x) = \sigma^2 ae^{-\beta x}$$

es el que se adecúa perfectamente para la estimación del parámetro $\lambda(x) = ae^{-\beta x_i}$ de una variable respuesta de Poisson. Y el modelo que verifica

$$Y = ae^{-\beta x} + \epsilon, \quad \text{con} \quad \text{Var}(Y|x) = \sigma^2 ae^{-\beta x} (1 + \frac{1}{\theta} ae^{-\beta x}) \quad (2.7)$$

es justamente el adecuado para la estimación del parámetro $\lambda(x)$ de una variable respuesta Binomial Negativa: $y_i \sim BN(\theta, \lambda(x_i))$, donde y_i suele representar el número de fracasos hasta obtener θ éxitos individuos y la probabilidad de fracaso es $q(x_i) = \frac{\lambda(x_i)}{\lambda(x_i) + \theta}$.

Tanto para el caso de Poisson, parcialmente estudiado (Wang y otros, 2005), como para el Binomial Negativo, se han obtenido diversos resultados. Los más destacables por novedosos son los que se refieren a este último modelo, que se pueden resumir en el siguiente

Teorema

- Para el modelo (2.7) el diseño localmente D-óptimo pone masa $\frac{1}{2}$ en cada uno de los puntos $x_1 = x_{\min}$ y $x_2 = \min\{x_{\max}, x_s\}$, donde x_s es la solución de la ecuación no lineal $x = x_{\min} + \frac{2}{\beta}(1 + \frac{a}{\theta}e^{-\beta x})$.
- Los diseños localmente c-óptimos para los vectores canónicos $c_1 = (1, 0)^t$ y $c_2 = (0, 1)^t$ están soportados en los puntos $x_1^* = x_{\min}$ y $x_2^* = \min\{x_{\max}, x_s\}$, donde x_s es la solución de la ecuación no lineal

$$x = x_{\min} + \frac{2}{\beta} \left[1 + \frac{a}{\theta} e^{-\beta x} + \left(\frac{a}{\theta} + 1 \right)^{\frac{1}{2}} \left(\frac{a}{\theta} + e^{-\beta(x-x_{\min})} \right)^{\frac{1}{2}} \right].$$

El peso óptimo del punto x_1^* será

$$p_{c_1}^* = \frac{x_2^*(a + \theta e^{-\beta x_2^*})^{\frac{1}{2}}}{x_2^*(a + \theta e^{-\beta x_2^*})^{\frac{1}{2}} + x_{\min}(a + \theta e^{-\beta x_{\min}})^{\frac{1}{2}}}$$

para c_1 -optimización, y

$$p_{c_2}^* = \frac{(a + \theta e^{-\beta x_2^*})^{\frac{1}{2}}}{(a + \theta e^{-\beta x_2^*})^{\frac{1}{2}} + (a + \theta e^{-\beta x_{\min}})^{\frac{1}{2}}}$$

para c_2 -optimización. Cuando $x_{\min} = 0$ el diseño c_1 -óptimo es el diseño singular concentrado en el 0.

3. Bibliografía

- [1] Han, C., Chaloner, K., D- and c-optimal designs for exponential regression models used in viral dynamics and other applications. *Journal of Statistical Planning and inference* **115**, (2003) 585–601.
- [2] Wang, Y., Myers, R.H., Smith, E.P., Ye, k. D-optimal designs for Poisson regression models . *Journal of Statistical Planning and inference* En prensa, (2005).

Design and optimization of EWMA control charts with asymmetric control limits for regions of maximum and minimum ARL

J. Carlos García-Díaz, Francisco Aparisi

Dpto. de Estadística e Investigación Operativa Aplicadas y Calidad

Universidad Politécnica de Valencia, 46022 Valencia, España

juagardi@eio.upv.es

Abstract

The EWMA control chart is one of the best available options to use when good performance is needed to detect small process shifts. This paper presents a method for design of EWMA charts for control processes, in which the detection of small shifts is not necessary, and at the same time is effective in detecting important shifts. In such cases the EWMA control chart can also be designed successfully to cope with these requirements. A Markov chain approach is also applied to determine the *ARL* of the modified EWMA control chart. Genetic algorithms have been used to carry out this design.

Keywords: Statistical Process Control, In-control, Indifference, and Out-of-control regions, Genetic Algorithms.

AMS: 62P30

1. Introduction

Exponentially weighted moving average (EWMA) schemes were introduced by Roberts as an alternative to the Shewhart control chart. The EWMA chart accumulates data from past samples and is therefore more sensitive to average small-sustained process shifts.

However, situations may arise where it is not relevant to detect small-magnitude shifts. For example, attempting to adjust a process when the mean shift is very small may lead to the phenomenon of over-adjustment, and to the appearance of extra variability in the process. For the reasons described above, Woodall first recommends defining which shift sizes are important to be detected quickly, as a criterion for the design of control charts.

2. Regions of maximum and minimum ARL

Woodall studied the statistical design of control charts and recommended taking a selection of shift sizes important to be detected as a criterion for the design of control charts. For this, he proposed the definition of three regions: an in-control region (region of maximum *ARL*), an indifferent region and a region out of control (region of minimum *ARL*).

One of the possible methods for determining the in-control region, and the out-of-control region, will be developed. This method is based on the Taguchi loss function.

Since the values plotted on the chart are not independent, Markov chains have been employed to calculate the *ARL* of the EWMA control charts. In addition, in this EWMA chart design the control limits are commonly asymmetric ($L_1 \neq L_2 \neq L$), therefore, the computation of the *ARL* requires a specific approach, and the previous references cannot be used directly. The Markov chain method proposed by Fu et al. was adapted for the calculations needed in this work.

3. Optimization

In order to find the optimal EWMA chart that satisfies the previous requirements, it is necessary to search for the optimal values of control limits, smoothing parameter and sample size. Genetic algorithms have been used to carry out this design.

4. Conclusions

The contribution is the development of a procedure used to compute the *ARL* of the EWMA control chart with asymmetric control limits, based on the Markov chain approach. On the other hand, the search for the parameters of the EWMA chart has been posed as an optimization problem solved by using genetic algorithms (GA). The user, that may not have prior knowledge of GAs, can easily find the best EWMA control chart for each case by using the developed software. The solutions provided by this optimization software are fully compatible with the specified requirements.

5. Bibliography

- [1] Woodall, W.H. The statistical design of quality control charts. *The Statistician* 1985; 34, pp. 155-160
- [2] Aparisi, F. and Garcia-Diaz, J.C. Optimization of Univariate and Multivariate Exponentially Weighted Moving-average Control Charts using Genetic Algorithm. *Computers and Operations Research* 2004; 31, pp. 1437-1454.
- [3] Fu, C. J.; Spiring, A.F.; Xie, H. On the average run lengths of quality control schemes using a Markov chain approach. *Statistics & Probability Letters* 2002; 56, pp. 369-380.

Determinación de Estados Estacionarios e Identificación de Parámetros de Procesos Multivariantes: Aplicación a un Proceso de Tratamiento Anaeróbico de Aguas

Maria Castellano Méndez¹, Gonzalo Ruiz², Wenceslao González Manteiga¹, Enrique Roca Bordello², Juan Lema²

¹mcaste@usc.es, Departamento de Estadística e I.O., Universidad de Santiago de Compostela

²Departamento de Ingeniería Química, Universidad de Santiago de Compostela

Abstract

La determinación de estados estacionarios (EE) es clave para el análisis de cualquier proceso. En este trabajo se ha aplicado una metodología on-line para la detección de EE, y se han empleado estos datos para la estimación de los parámetros de un modelo cinético. La distribución de los estimadores de los parámetros se ha aproximado con un Wild Bootstrap Paramétrico.

Palabras Clave: Componentes Principales, Estado Estacionario, Bootstrap

1. Introducción

La detección de estados estacionarios es un punto clave para el análisis de proceso: el mantenimiento, la optimización, control,... Existen diversos trabajos sobre la detección de estados estacionario, pero pocos sobre algoritmos de detección on-line. Estos trabajos además suelen referirse a procesos univariantes. En este trabajo se emplea Análisis de Componentes Principales (PCA) para la reducción del número de variables, y para obtener variables incorreladas. Este proceso permitió pasar a un problema bidimensional [1], de modo que las nuevas variables se utilizaron para la determinación de los estados estacionarios, empleando la metodología propuesta por Cao y Rhinehart [2],

Una vez detectados los estados estacionarios, los parámetros de los modelos cinéticos pueden ser estimados. Tras la estimación de los parámetros se empleó Bootstrap paramétrico para aproximar la distribución de los estimadores de los parámetros con el fin de obtener regiones de confianza para estos, y para el modelo estimado.

2. Material y Métodos

Se dispone de datos de una planta de tratamiento anaeróbico de aguas residuales, compuesta por un reactor híbrido UASB-UAF de 1m³, que se monitoriza a través de 17 variables. Se aplicó PCA para reducir la dimensión del problema, y la la metodología de Cao y Rhinehart para detectar los estados estacionarios. Se estimaron los parámetros del modelo de Haldane empleando mínimos cuadrados ponderados, de modo que los diferentes estados estacionarios tuviesen

el mismo peso. Se aplicó Wild Bootstrap paramétrico, para la estimación de la distribución de los estimadores de los parámetros.

3. Resultados

La Figura 1 presenta el ejemplo en el que se trabajó, un modelo cinético de Haldane, que relaciona la tasa de reacción con la concentración de un reactivo. Las estimaciones bootstrap muestran que los parámetros estimados presentan una clara relación de dependencia entre sí, lo que reduce el tamaño de la región de confianza. Asimismo se determinó para cada valor del DOC un intervalo de confianza bootstrap para el valor del caudal de metano, al 95 %.

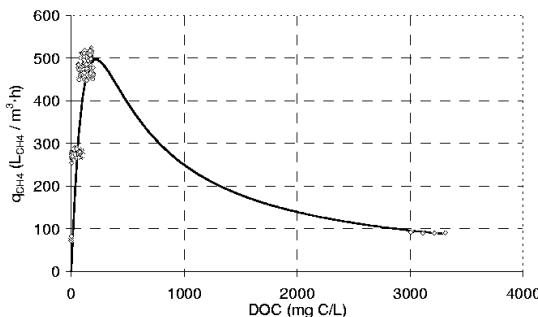


Figura 1: Datos estacionarios detectados y modelo Haldane estimado, que relaciona el carbono orgánico disuelto (DOC) y el caudal de metano (q_{CH_4})

4. Agradecimientos

Este trabajo ha sido financiado por el MEyC (apoyo Europeo FEDER incluido) a través de los proyectos ANACOM(CTQ2004-07811-C02-01) y MTM2005-00820.

5. Bibliografía

- [1] G. Ruiz, M. Castellano, W. González, E. Roca and J.M. Lema. (2004). Algorithm for steady state detection of multivariate process: Application to anaerobic wastewater digestion process. In: Proceedings of the 2nd International IWA Conference on Automation in Water Quality Monitoring (AutMoNet). G. Langergraber, S. Winkler, N. Fleischmann, A. Pressl and R. Haberl (eds). BOKU University. Abril 2004. Viena – Austria. Pp 181-188.
- [2] Cao,S.L. and R.R. Rhinehart. (1995). *Journal of Process Control* 5 (6), 363-374.

Determinación de Niveles de Ozono y NO_x en Galicia

M^a Castellano Méndez¹, Amaya Franco², David Cartelle³, Manuel Febrero¹, Enrique Roca²

¹mcaste@usc.es, Departamento de Estadística e Investigación Operativa,
Universidad de Santiago de Compostela

²Departamento de Ingeniería Química, Universidad de Santiago de
Compostela

³Laboratorio Medioambiental de Galicia, Xunta de Galicia, A Coruña

Abstract

El estudio de la calidad del aire es cada vez más importante en la sociedad actual. Se han recogido datos de calidad ambiental de tres zonas diferentes. Se han establecido umbrales de concentraciones de ozono y óxidos de nitrógeno para reconocer situaciones anormales (episodios), y se han establecido empleando series de tiempo las relaciones con algunas variables meteorológicas.

Palabras Clave: Ozono, Series de Tiempo, Niveles de Confianza

1. Introducción

La evaluación de calidad del aire ha sido un importante asunto medioambiental durante las últimas décadas debido al incremento de la emisión de contaminantes a la atmósfera por parte del hombre. La Directiva 96/62/CE del Consejo, de 27 de septiembre de 1996, sobre evaluación y gestión de la calidad del aire ambiente (Directiva Marco), proporciona un marco comunitario para las medidas nacionales, regionales y locales destinadas a mejorar la calidad del aire, estableciendo umbrales para los óxidos de nitrógeno (NO_x) y para el ozono; estos valores límite son, para las medias horarias de NO_x y ozono de 200 and 180 $\mu\text{m}/\text{m}^3$, respectivamente y para valores puntuales de 240 $\mu\text{m}/\text{m}^3$ para el ozono. En este marco se ha realizado un estudio de la calidad del aire en distintas regiones de Galicia, A Guardia, Corribedo y Verín (dos costeras y una interior), con el objetivo de promover el desarrollo sostenible, y de proteger estos entornos naturales. Los objetivos de este trabajo son: 1) Analizar el comportamiento del ozono y NO_x en las tres zonas de estudio; 2) Determinar los niveles básicos y medios de ambos contaminantes e identificar las posibles fuentes de emisión; 3) Establecer umbrales definiendo niveles de confianza, para poder identificar situaciones anormales; 4) Establecer una relación entre los datos de inmisión de ozono y NO_x y las variables meteorológicas de cada zona y; 5) Predecir con series de tiempo los valores del ozono y NO_x.

2. Metodología y Resultados

Se han definido dos niveles puntuales, uno basal (percentil 5 %) y otro de localización (mediana) con el fin de determinar el perfil característico del área y facilitar la determinación de episodios. Así mismo se han considerado modelos ARMAX de series de tiempo [1] para el estudio temporal de las concentraciones, empleando la información meteorológica (radiación, lluvia, viento,...) de las estaciones de muestreo como variables exógenas.

3. Resultados y Discusión

Los resultados respecto al Ozono son razonables, en tanto que A guardia, zona más industrial presenta mayores valores que Xurés y este a su vez mayores que Corrubedo, que es un parque natural. Para los valores de NOx se detectó que Corrubedo, presentaba valores anormalmente altos e irregulares, lo que hace suponer que este comportamiento es debido a un fenómeno de transporte a media distancia [2]. Los episodios detectados a partir del análisis funcional han sido analizados de modo que es posible determinar las causas que los provocaron, en general industrias circundantes o vertederos.

Los resultados del estudio de series de tiempo muestran que los niveles de ozono no dependen de variables meteorológicas, salvo en Verín donde la radiación es relevante.

4. Agradecimientos

Este trabajo ha sido financiado por el MEyC (apoyo Europeo FEDER incluido) a través del proyecto MTM2005-00820.

5. Bibliografía

- [1] Dueñas, C., Fernández, M.C., Cañete, S., Carretero, J., Liger, E. (2004). Analyses of ozone in urban and rural sites in Málaga (Spain). *Chemosphere*, 56, 631-639.
- [2] Wei, W.W.S. (1993). *Time Series Analysis: Univariate and Multivariate Methods*. Addison-Wesley.

Determinación del tamaño muestral mediante técnicas bayesianas

Lourdes Rodero¹, Josep Ginebra¹, Xavi Puig¹

¹lourdes.rodero@upc.edu, josep.ginebra@upc.edu, xavier.puig@upc.edu

Departament d'Estadística i Investigació Operativa,
Universitat Politècnica de Catalunya

Abstract

El problema de la determinación del tamaño muestral es uno de los más comunes y controvertidos en la estadística práctica, así como una de las cuestiones principales en el diseño de experimentos. En esta comunicación exploramos este problema en el marco de la teoría de la decisión Bayesiana.

Palabras Clave: Tamaño muestral, Análisis bayesiano, medidas de información

1. Introducción

Determinar el tamaño muestral es una de los problemas esenciales en el diseño de un experimento. Una mala elección del tamaño de la muestra al planificar un experimento puede ser crítico, ya que muestras demasiado grandes malgastan tiempo y recursos mientras que muestras demasiado pequeñas no permiten obtener la información necesaria sobre los parámetros del modelo.

La determinación del tamaño de una muestra es un problema ampliamente tratado. En el enfoque frecuentista se tiende a escoger aquel tamaño de muestra que permite realizar cierta prueba de significación con la garantía de que se tendrá un determinado nivel de confianza.

En esta comunicación se presentará un método de selección del tamaño muestral dentro del contexto de la teoría de la decisión bayesiana, para el caso en el que se observe una variable aleatoria con una distribución de la familia exponencial.

El método encontrado pasa por relacionar la información contenida en la muestra medida a través del *valor esperado de la información* o la *información mutua*, con el tamaño de muestra utilizado.

Se presentarán ejemplos de elección del tamaño muestral para las distribuciones Normal, Exponencial, Poisson, Binomial y Negativa Binomial.

2. Bibliografía

- [1] Raiffa, h. y Schlaifer, R.O. (1961) *Applied Statistical Decision Theory*. M.I.T. Press.

- [2] Chaloner, K. y Verdinelli, I. (1995) *Bayesian Experimental Design: A review*. Statistical Science, Vol. 10, N°3.
- [3] Parmigiani, G. y Berry, D.A. *Applications of Lindley information measure to the Design of Clinical Experiments*. Aspects of Uncertainty, chapter 20.

Determinación óptima en reacciones enzimáticas con inhibición por exceso de sustrato

M.J. Rivas-López¹, M.T. Santos¹, I. Martínez²

¹chusrl@usal.es, maysam@usal.es, Departamento de Estadística, Universidad de Salamanca

² ijmartin@ual.es, Departamento de Estadística y Matemática Aplicada, Universidad de Almería

Abstract

El diseño óptimo de experimentos, que permite la mejor estimación de parámetros presentes en la modelización de cualquier fenómeno, es aplicado en este trabajo, en el estudio de la velocidad de las reacciones enzimáticas que presentan inhibición por exceso de sustrato.

Claves: D-optimalidad, c-optimalidad, modelos racionales.

AMS: 62K05

1. Introducción

La velocidad de las reacciones enzimáticas se mide en unidades de cantidad de sustrato transformado por unidad de tiempo y la unidad standard es el μmol de sustrato transformado por minuto. Las medidas de actividad enzimática deben hacerse en condiciones prefijadas en los parámetros que influyen en la velocidad de reacción, tales como la concentración de enzima, de sustrato, de producto y de inhibidores, el pH y la temperatura.

Como la cantidad de enzima es limitada, a altas concentraciones de sustrato toda la enzima estará complejada con el sustrato y no podrá mejorarse la reacción poniendo más sustrato. A bajas concentraciones de sustrato la adición de sustrato aumentará la cantidad del complejo enzima-sustrato, y por tanto, la velocidad de reacción. La ecuación de Michaelis-Menten permite explicar el fenómeno de saturación, la velocidad de la reacción enzimática aumenta al incrementar la concentración de sustrato, pero a partir de una determinada concentración, la velocidad de reacción no se ve afectada. Esta ecuación de Michaelis-Menten es:

$$V = \frac{V_{max}[S]}{K_m + [S]} + \epsilon \quad (2.8)$$

correspondiente a la expresión de una hipérbola rectangular con asíntotas $V = V_{max}$ y $[S] = -K_m$, donde la constante de Michaelis, K_m , representa la concentración de sustrato para la cual la velocidad es la mitad de la máxima, V_{max} .

Los inhibidores enzimáticos son sustancias que al combinarse con alguna forma enzimática producen una disminución de la velocidad de reacción. Un caso particular de inhibición es el causado por el producto y por el exceso de producto. El exceso de sustrato puede inhibir porque dos moléculas de sustrato pueden unirse al mismo centro activo de forma no productiva. La ecuación que regula la velocidad de reacción ante este fenómeno, es dada por el modelo lineal inverso cuadrático,

$$V = \frac{V_{max}[S]}{K_m + [S] + [S]^2/K_i} + \epsilon \quad (2.9)$$

donde K_m es la constante de Michaelis-Menten y K_i es una nueva constante debida a la inhibición.

En 1969, Westley sugirió, que es posible, que esta orientación incorrecta del sustrato, también se presente a bajas concentraciones donde aparentemente no tenemos inhibición por sustrato, modificando la ecuación que regiría el sistema y proponiendo el siguiente modelo:

$$V = \frac{V_{max}[S]}{K_m + [S](1 + K_m/K_i)} + \epsilon \quad (2.10)$$

2. Aplicación del diseño óptimo de experimentos.

Las ecuaciones cinéticas de las reacciones enzimáticas que presentan inhibición por exceso de sustrato dependen de tres parámetros (V_{max} , K_m , K_i) que pueden estimarse, aunque presentando bastante variabilidad, aplicando la metodología de los gráficos doble-recíproco (V_{max} y K_m) y recíproco (K_i).

En este trabajo se obtienen diseños D-óptimos locales, que permiten la mejor estimación conjunta de la velocidad máxima de la reacción y de las constantes de inhibición involucradas en las reacciones, así como diseños c-óptimos locales, para los vectores canónicos, que proporcionan la mejor estimación de cada uno de los parámetros de forma individual, todos ellos a partir de los trabajos de Haines (1992) y Song & Wong (1998).

3. Bibliografía

- [1] Dixon, M & Webb, E.C. (1979). *Enzymes* (3^a Ed.). Longman. New York.
- [2] Haines, L.M. (1992). Optimal design for inverse quadratic polynomials. *South african Statistical Journal* 26, 25-41.
- [3] Song, D & Wong, W.K. (1998). Optimal two-points designs for the Michaelis-Menten model with heteroscedastic errors. *Communications in Statistics* 27, 1503-1516.

Diseño de planes de aceptación óptimos para la distribución exponencial biparamétrica

Arturo J. Fernández¹, Carlos J. Pérez-González²

¹ajfern@ull.es,²cpgonzal@ull.es, Departamento de Estadística e
Investigación Operativa, Universidad de La Laguna

Abstract

Este artículo presenta un procedimiento exacto para el diseño de planes óptimos de aceptación de la fiabilidad cuando la duración de los dispositivos en estudio tiene una distribución exponencial biparamétrica. El procedimiento determina el mínimo tamaño muestral, además del criterio de decisión, y admite la existencia de censura de tipo II.

Claves: Censura de tipo II, función operativa característica, riesgos del productor y del consumidor

1. Preliminares

Los planes de aceptación proporcionan reglas de decisión para aceptar o rechazar la efectividad de un producto. Cuando los productos están agrupados en lotes, cada lote se acepta o rechaza en función de los resultados obtenidos en una muestra. En análisis de fiabilidad, estos planes deciden la aceptación o no de un lote de dispositivos utilizando las duraciones de una muestra de unidades elegidas al azar. A menudo por motivos económicos, el test finaliza antes del fallo de todos los dispositivos.

Supongamos que la duración de cierto dispositivo, X , tiene una distribución exponencial biparamétrica, $Exp(\mu, \sigma)$, con función de densidad

$$f(x; \mu, \sigma) = \sigma^{-1} \exp\{-(x - \mu)/\sigma\}, \quad x > \mu, \quad (\mu \geq 0, \sigma > 0),$$

y que n unidades elegidas al azar de un lote de gran tamaño de estos dispositivos se ponen a prueba al mismo tiempo. El experimento finaliza cuando se observe el m -ésimo fallo ($m \leq n$), i.e., el experimento está censurado por el número de fallos (censura de tipo II) y $q = 1 - m/n$ es la proporción de datos censurados.

2. Criterio de decisión

Si $X_{1:n}, X_{2:n}, \dots, X_{m:n}$ representan los tiempos de fallo observados en el experimento anterior y $m \geq 2$, los estimadores de máxima verosimilitud de μ y σ son $\hat{\mu} = X_{1:n}$ y $\hat{\sigma} = (T - nX_{1:n})/m$, donde $T = \sum_{i=1}^m X_{i:n} + (n - m)X_{m:n}$ es el tiempo total del test. Además, $2n(\hat{\mu} - \mu)/\sigma \sim \chi_2^2$ y $2m\hat{\sigma}/\sigma \sim \chi_{2m-2}^2$ son independientes. Asumimos que C es la duración mínima aceptable, i.e., cualquier unidad con duración inferior que C se considera defectuosa. La proporción de

unidades defectuosas sería pues $p = \Pr(X < C)$, con lo cual C coincidiría con $\xi_p = \mu - \sigma \ln(1-p)$, el cuantil p de X . Obviamente, utilizando la cantidad pivotal $Y_p = (\hat{\mu} - \xi_p)/\hat{\sigma}$ es posible contrastar la aceptabilidad de un lote. En tal caso, el lote se aceptaría cuando Y_p es suficientemente grande, digamos $Y_p > k$, o, equivalentemente, cuando $\hat{\mu} - k\hat{\sigma} > \xi_p$.

3. Diseño de planes óptimos

En nuestro caso, un plan de aceptación se caracteriza por el tamaño muestral n , el grado de censura q (o el numero de fallos observados m) y la constante de aceptación k . La función operativa característica (OC) asociada al plan de muestreo (n, m, k) se define como la probabilidad de aceptación del lote cuando p es la proporción de unidades defectuosas. Luego, la expresión de la función OC sería

$$L(p) \equiv L(p; n, m, k) = \Pr(\hat{\mu} - k\hat{\sigma} > \xi_p), \quad 0 < p < 1,$$

o de forma más sencilla

$$L(p) = \Pr\left(\frac{1}{2n}\chi_2^2 - \frac{k}{2m}\chi_{2m-2}^2 > -\log(1-p)\right), \quad 0 < p < 1.$$

En la práctica, n y k se eligen mediante un acuerdo entre el productor y el consumidor de forma que el plan acepte lotes con $p = p_\alpha$ con una alta probabilidad de al menos $1 - \alpha$, y rechaze lotes con $p = p_\beta$ con una baja probabilidad de a lo sumo β ; p_α y p_β son los niveles de calidad aceptables y rechazables, mientras que α y β son los riesgos del productor y del consumidor, respectivamente. En resumen, el plan debe verificar que

$$L(p_\alpha; n, m, k) \geq 1 - \alpha \quad \text{y} \quad L(p_\beta; n, m, k) \leq \beta.$$

Una vez prefijado (a menudo, de acuerdo con el coste y duración del experimento) el nivel de censura $q \in [0, 1]$, además de α , β , p_α y p_β , se presenta en este artículo un procedimiento exacto para determinar el plan de muestreo con menor tamaño muestral n (y los correspondientes valores de m y k) que satisface las desigualdades anteriores. El procedimiento es exacto pues es posible obtener la función OC de forma analítica. A modo de ejemplo, supongamos que la duración de un fusible es $Exp(\mu, \sigma)$, y que se desea determinar un plan de aceptación tal que, en cualquier lote, un 3.19 % de unidades defectuosas es aceptable, y un 9.42 % es rechazable. El riesgo del productor es $\alpha = 0.05$ y el del consumidor es $\beta = 0.10$. Los planes de aceptación óptimos serían entonces:

- $q = 0.0 \implies n = 36, m = 36$ y $k \in [-0.03534, -0.03501]$.
- $q = 0.5 \implies n = 39, m = 19$ y $k \in [-0.04034, -0.03926]$.
- $q = 0.9 \implies n = 78, m = 8$ y $k \in [-0.06175, -0.06061]$.

Diseño óptimo de experimentos con variables explicativas censuradas

Sandra Garcet Rodríguez¹, Jesús López Fidalgo²

¹sandra_garcet@usal.es, Departamento de Estadística, Universidad de Salamanca

²jesus.lopezfidalgo@uclm.es, Departamento de Matemáticas (Estadística e I.O.), E.T.S.I. Industriales, Universidad de Castilla La Mancha

Abstract

El objetivo de este trabajo es proporcionar un algoritmo que permita el cálculo aproximado de diseños D-óptimos para modelos lineales donde los valores de una de las variables explicativas (o conjunto de variables) vienen censurados de acuerdo a una distribución de probabilidad conocida..

Palabras Clave: Matriz de información, D-optimización

1. Introducción

En la teoría clásica de diseño óptimo de experimentos para modelos lineales se supone que todas las variables explicativas pueden tomar valores en todo el espacio del diseño. Sin embargo en muchas áreas de aplicación es posible encontrar variables explicativas cuyos valores están censurados. Por ejemplo, Varela, Cordovilla, Jiménez, Novoa (2001) utilizan un test de ejercicio para predecir las complicaciones del tratamiento quirúrgico del cáncer de pulmón en el cual la variable Tiempo de Ejercicio tiene una cierta distribución dentro de la población, por lo que el experimento diseñado vendrá censurado en dicha variable generándose otro diseño. Nuestro objetivo es determinar, para un modelo lineal $E[y] = \alpha\eta(t)$, donde t denota una variable aleatoria con función de distribución de probabilidad $F: \mathfrak{S} \rightarrow [0,1]$, cómo debe ser el diseño asociado al modelo $\hat{\xi}$ si queremos que el diseño ξ que se genera sea un diseño lo mas cercano posible al óptimo.

2. Fecha

Dado que F es conocida es posible calcular una estimación de ξ . Sean $\mathfrak{S}=[0,T]$ y $F(t) = \int_0^t f(x) dx$, entonces $\xi(t) = f(t) \int_t^T \frac{\xi(x)}{F(x)} dx$.

La inversa $\hat{\xi}(t) = \left[\frac{\xi(t)f'(t) - \xi'(t)f(t)}{f^2(t)} \right] F(t)$ de la expresión anterior podría ser utilizada para determinar de manera directa el diseño $\hat{\xi}$ a aplicar para obtener en la práctica un diseño ξ óptimo. Esta solución solamente es viable cuando ξ es un diseño.

Consideremos la matriz de información $M(\xi) = \int_{\mathfrak{S}} \xi(t)\eta(t)\eta^T(t) dt$. Dado que $\xi(t) = f(t) \int_t^T \frac{\hat{\xi}(x)}{F(x)} dx$, entonces

$$\begin{aligned} M(\xi) &= \left[\int_0^T \frac{\hat{\xi}(t)}{F(t)} dt \right] \left[\int_0^T f(t)\eta(t)\eta^T(t) dt \right] \\ &\quad - \int_0^T \frac{\hat{\xi}(t)}{F(t)} \left(\int_0^t f(x)\eta(x)\eta^T(x) dx \right) dt \end{aligned}$$

Si suponemos que $\hat{\xi}$ es un diseño exacto y $g(t) = f(t)\eta(t)\eta^T(t)$ entonces $M(\xi) = \sum_{j=1}^n \hat{e}_j \left(\frac{G(t_j) - G(0)}{F(t_j)} \right)$.

Proponemos un algoritmo para calcular $\hat{\xi}$ de manera que ξ sea óptimo, utilizando la expresión anterior de la matriz de información de ξ

1. Sea $\hat{\xi}_0$ un diseño inicial no singular.
2. Se define $\hat{\xi}_n = (1 - \alpha_n)\hat{\xi}_{n-1} + \alpha_n \hat{\xi}_t$ donde

$$\alpha_n = \frac{1}{1+n} ; \quad t = \max_t |M(\xi_n)| ; \quad \xi_n = f(t) \sum_{t_j \geq t}^n \frac{\hat{\xi}_n(t_j)}{F(t_j)} ;$$

3. El algoritmo se detiene cuando $\frac{|M(\xi_{n-1})|}{|M(\xi_n)|} \geq \delta$.

3. Bibliografía

- [1] Varela G., Cordovilla, R. Jiménez, M. F., Novoa, N. (2001). *Utility of Standardized Exercise Oximetry to Predict Cardiopulmonary Morbidity After Lung Resection*. European Journal of Cardiothoracic Surgery, 19, 351-354.

Diseños experimentales para modelos de crecimiento del tamaño de una población

J. López Fidalgo¹, I.M. Ortiz Rodríguez², Y. del Águila del Águila³

¹jesus.lopezfidalgo@uclm.es, Departamento de Matemáticas. Universidad de Castilla la Mancha

²iortiz@ual.es, Departamento de Estadística y Matemática Aplicada. Universidad de Almería

³yaguila@ual.es, Departamento de Estadística y Matemática Aplicada. Universidad de Almería

Abstract

En este trabajo se considera un modelo de regresión no lineal que es una generalización de la curva logística con cuatro parámetros. Para este modelo se buscan diseños óptimos locales bajo diferentes criterios y se hace un estudio de eficiencia para comprobar la robustez de los diseños cuando se cambia el criterio o los valores iniciales de los parámetros.

Claves: Modelo logístico, diseño óptimo, diseño aproximado.

AMS: 62K05

1. Introducción

Los modelos de regresión no lineal se usan para caracterizar muchas relaciones entre variables en campos como la agricultura, biología o química. En este trabajo se estudia el modelo usado en (Jolicoeur y Pontier, 1989) para representar el crecimiento y decrecimiento de una población

$$E(Y_t) = \frac{1}{C_1 e^{t/D_1} + C_2 e^{-t/D_2}}, \quad t \in [0, \infty), \quad (2.11)$$

donde $C_1, C_2, D_1, D_2 \geq 0$, t es el tiempo e Y_t es el tamaño o proporción de población en el tiempo t . Los parámetros D_1 y D_2 están relacionados con el decrecimiento y crecimiento de la población, respectivamente y C_1 y C_2 controlan simultáneamente las razones relativas de crecimiento y decrecimiento así como la altura de la curva (tamaño de la población).

El modelo (2.11) es una generalización de la curva logística y como casos particulares se obtienen los modelos exponenciales de crecimiento ($C_1 = 0$) y decrecimiento ($C_2 = 0$) y los logísticos de crecimiento ($D_1 \rightarrow \infty$) y decrecimiento ($D_2 \rightarrow \infty$)

Los 4 parámetros desconocidos (C_1, C_2, D_1, D_2) del modelo general (2.11) son no lineales. Con la reparametrización $A = 1/C_1$, $B = D_1$, $D = D_2$ y $C = C_2/C_1$ podemos dejar solo 3 parámetros con un comportamiento no lineal.

Dos características importantes del modelo (2.11) son la respuesta máxima y el punto donde ésta se alcanza, t_{max} :

$$t_{max} = \frac{D_1 D_2}{D_1 + D_2} \ln \frac{C_2 D_1}{C_1 D_2}. \quad (2.12)$$

Para planificar un nuevo experimento con el fin de estimar el modelo (2.11) habrá que partir de un diseño adecuado. Según los objetivos prioritarios en la estimación del modelo se fijará un criterio de optimización para la búsqueda del diseño óptimo. Si el objetivo es estimar con máxima precisión los parámetros del modelo, un criterio muy usado es D-optimización y si sólo interesa estimar con precisión alguna característica como el punto de máxima respuesta, se usa el criterio c-optimización. Estos criterios son funciones de la matriz de información y para modelos no lineales dependen de parámetros desconocidos, por lo que se precisa algún tipo de información sobre ellos para obtener el diseño óptimo. Si se dispone inicialmente de un determinado valor para los parámetros se habla de diseño óptimo local.

En este trabajo se caracterizan los diseños D-óptimos locales y se buscan diseños c-óptimos locales para los parámetros individuales del modelo, así como para las características t_{max} e $Y(t_{max})$. Puesto que los diseños obtenidos dependen de los valores iniciales de los parámetros, se hace un estudio de eficiencia para cuantificar esta dependencia y además se estudia el comportamiento de los diseños óptimos frente a otros criterios de optimización.

2. Bibliografía

- [1] Dette H., Martínez I., Ortiz I. y Pepelyshev A. (2005). Maximin efficient design of experiment for exponential regression models. *Journal of Statistical Planning and Inference*, in press.
- [2] Fedorov V.V. (1972). *Theory of Optimal Experiments*. Academic Press, New York.
- [3] Jolicoeur P. and Pontier J. (1989). Population growth and decline: a four-parameter generalization of the logistic curve. *J. theor. Biol.* 141, 563-571.
- [4] López-Fidalgo J. and Rodríguez-Díaz J.M. (2004). Elfving's method for m -dimensional models. *Metrika* 59, 235-244.
- [5] López-Fidalgo J., Rodríguez-Díaz J.M., Sánchez G. and Santos-Martín M.T. (2005). Optimal designs for compartmental models with correlated observations. *Journal of applied statistics* 32, 1075-1088.

Diseños factoriales D-óptimos a dos niveles para modelos de Poisson

Roberto Dorta-Guerra¹, Enrique González-Dávila¹, Josep Ginebra²

¹*rodorta@ull.es, egonzale@ull.es, Departamento de Estadística, Investigación Operativa y Computación, Universidad de La Laguna*

²*josep.ginebra@upc.edu, Departament d'Estadística i Investigació Operativa, Universitat Politècnica de Catalunya*

Abstract

Se estudian los diseños factoriales a dos niveles que maximizan el determinante de la matriz de información para un modelo de Poisson con una y dos variables regresoras. El valor del determinante es no acotado y aumenta conforme los puntos del diseño están más dispersos, haciendo el estudio en regiones acotadas.

Claves: Modelos de Poisson; diseños factoriales; D-optimalidad

AMS: 62K05, 62K15, 62J12

1. Introducción

En modelos lineales normales, la cantidad de información que aportan los diseños factoriales a dos niveles, medida a través del determinante de la matriz de información, no depende de dónde se encuentre centrado el diseño, ni de cómo esté orientado. Además, las asignaciones balanceadas son más informativas que las no balanceadas cuando se trabaja con el mismo número de ensayos. En general, los diseños factoriales a dos niveles son óptimos, o se aproximan a los diseños óptimos, para la mayoría de los criterios de optimalidad y para una amplia gama de regiones experimentales de interés. Sin embargo, en algunos casos, las variables respuestas no pueden ser modeladas a través de modelos lineales normales. Por ejemplo, cuando se miden los efectos de crecimiento o la tasa de mortalidad de determinados organismos a los que previamente se le han aplicado concentraciones o dosis de un agente tóxico. En estos casos las respuestas siguen una distribución de Poisson, y el modelo asociado a dichas variables pertenece a la clase de los modelos lineales generalizados.

En la literatura podemos encontrar varios trabajos que analizan los diseños D-óptimos para modelos de Poisson. Así, Chiachirini (1996) calculó los diseños localmente D-óptimos para modelos de Poisson con un factor. Mullekom y Myers (2001) analizan el caso de dos factores y dedujeron que el diseño localmente D-óptimo era un diseño de tres puntos. A partir de estas referencias se puede comprobar que los diseños D-óptimos para los modelos de Poisson no pertenecen a la familia de los diseños factoriales a dos niveles. En este trabajo

se calculan los diseños factoriales a dos niveles localmente D-óptimos para modelos de Poisson con uno y dos factores, ya que, aún cuando no son los diseños D-óptimos generales, son los más utilizados en la práctica habitual.

2. Modelos de Poisson

Supongamos un experimento donde observamos un conjunto de respuestas, y_i , bajo un conjunto de q condiciones experimentales distintas, cada una caracterizada por los valores de $x_i = (x_{1i}, \dots, x_{ki})'$ para $i = 1, \dots, q$ y $k = 1$ ó 2. En modelos de Poisson, las distribuciones de $y_i|x_i$ se asumen independientes tal que $y_i \sim \text{Poisson}(\lambda(x_i, \beta))$, con $\beta = (\beta_0, \dots, \beta_k)$. El modelo de Poisson viene dado según la expresión

$$\lambda(x_i, \beta) = \exp(x_i' \beta)$$

y la matriz de información de Fisher para dicho modelo es de la forma

$$I(\beta) = \sum_{i=1}^q h(z_i) (1, x_{1i}, \dots, x_{ki})' (1, x_{1i}, \dots, x_{ki}) \quad (2.13)$$

con $h(z_i) = (e^{z_i}/(1 - e^{z_i}))$ y $z_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki}$.

3. Diseños factoriales D-óptimos

El criterio de la D-optimalidad maximiza el determinante de la matriz de información de Fisher. En este trabajo hemos calculado los diseños factoriales 2^k D-óptimos centrados en $x_0 = (x_{10}, \dots, x_{k0})$, y cuyos puntos son de la forma $x_i = (x_{10} + a_{i1}R_1, \dots, x_{k0} + a_{ik}R_k)$ con $i = 2^k$, $k = 1$ ó 2, y a_{ik} igual a 1 ó -1. Sustituyendo dichos puntos en la expresión (2.13) y calculando el determinante, se demuestra que la expresión obtenida es no acotada. Esto es, cuanto más dispersos estén los puntos mayor será el valor del determinante. Para evitar esto se llevan a cabo acotaciones sobre la región experimental. Se calcula tanto el diseño D-óptimo factorial completo para un x_0 dado, como el diseño D-óptimo factorial completo en la región acotada, obteniendo diseños que no necesariamente son los más dispersos en dichas regiones.

4. Bibliografía

- [1] Chiachierrini L.M. (1996). Experimental Design Issues in Impaired Reproduction Studies. *Ph.D. Thesis, Virginia Tech.*
- [2] Van Mullekom J. y Myers R.H. (2001). Optimal experimental designs for Poisson impaired reproduction. *Technical Report, Department of Statistics, Virginia Tech.* vol. 01-1.

Diseños óptimos para discriminar entre observaciones log-normales

Jesus López-Fidalgo¹, Chiara Tommasi², Paula Camelia Trandafir³

¹jesus.lopezfidalgo@uclm.es, Departamento de Matemáticas, Universidad Castilla La Mancha

²chiara.tommasi@unimi.it, Dipartimento di Economia Politica e Aziendale, Università degli Studi di Milano

³a110619@usal.es, Departamento de Matemáticas, Universidad de Salamanca

Abstract

En este trabajo se presenta un nuevo criterio, basado en la distancia de Kullback–Leibler que se puede utilizar para discriminar entre dos modelos con una distribución no-normal. Se presenta una aplicación de este criterio para discriminar entre el modelo de Michaelis–Menten y una extensión del mismo bajo la condición de log–normalidad.

Claves: T–optimalidad, distancia de Kullback–Liebler, modelo no lineales, KL–optimalidad

1. Introducción

El criterio definido para discriminar entre dos modelos homocedasticos bajo la condición de normalidad ha sido introducido por Atkinson y Fedorov ([1]). Sea el modelo general de regresión no lineal:

$$y = \eta(x) + \epsilon, \quad x \in \mathcal{X}, \quad \epsilon \sim \mathcal{N}(0, \sigma^2). \quad (2.14)$$

donde y es la variable respuesta y x es la variable explicativa.

Sea $\eta(x) = \eta_1(x, \theta_1)$ el modelo verdadero y $\eta_2(x, \theta_2)$ el otro modelo. La función criterio T es:

$$T_{21}(\xi) = \min_{\theta_2 \in \Omega_2} \int_{\mathcal{X}} [\eta(x) - \eta_2(x, \theta_2)]^2 \xi(dx). \quad (2.15)$$

Por definición, un diseño ξ^* que maximiza $T_{21}(\xi)$ se denomina *T–óptimo*.

Para el caso de heteroscedasticidad para los modelos con múltiple respuesta Ucinski y Bogacka [2] introducen una generalización al criterio T.

2. Discriminación entre los modelos no–normales

Sea $f_1(y, x, \theta_1, \tau)$ y $f_2(y, x, \theta_1, \tau)$ dos funciones de densidad, donde θ_i , $i = 1, 2$, y, x son definidas antes y τ es el parámetro de la varianza. Se considera también que $f(y, x, \tau) = f_1(y, x, \theta_1, \tau)$ es el modelo "verdadero".

Usando esta notación, la distancia Kullback–Liebler definida entre el modelo "verdadero" $f(y, x, \tau)$ y el otro modelo $f_2(y, x, \theta_1, \tau)$ es

$$\mathcal{I}(f, f_2, x, \theta_2, \tau) = \int f(y, x, \tau) \log \left(\frac{f(y, x, \tau)}{f_2(y, x, \theta_2, \tau)} \right) dy, \quad \forall x \in \mathcal{X}$$

donde \log es el logaritmo neperiano. El KL–criterio se define como

$$I_{21}(\xi) = \min_{\theta_2 \in \Omega_2} \int_{\mathcal{X}} \mathcal{I}(f, f_2, x, \theta_2) \xi(dx). \quad (2.16)$$

Un diseño que maximiza $I_{21}(\xi)$ se llama *KL–óptimo*.

Se proporciona un teorema de equivalencia que permite formular un algoritmo de construcción de diseños KL–óptimos.

3. Aplicación a las observaciones log–normales

Sea el modelo de Michaelis–Menten (MM)

$$y = \frac{Vx}{K + x} + \varepsilon \quad x \in \mathcal{X}, \quad (2.17)$$

y una modificación del modelo de Michaelis–Menten (MMM)

$$y = \frac{Vx}{K + x} + Fx + \varepsilon. \quad (2.18)$$

Se supone que las observaciones son distribuidas con distribución log–normal. En este caso el KL–criterio buscará encontrar un diseño que maximice la función:

$$\min_{K,V} \sum_i \left[\log \left(\frac{V_0 x_i}{K_0 + x_i} + F_0 x_i \right) - \log \left(\frac{V x_i}{K + x_i} \right) \right]^2 \xi(x_i).$$

Usando un algoritmo iterativo se obtiene numéricamente el KL–óptimo

$$\left\{ \begin{array}{ccc} 0 & 1,4515 & 5 \\ 0,2938 & 0,4996 & 0,2066 \end{array} \right\}$$

con una eficiencia de 0,999.

4. Bibliografía

- [1] Atkinson A. C., Fedorov V.V. (1975). The designs of experiments for discriminating between two rival models. *Biometrika*, 62, 57–70.
- [2] Uciński D., Bogacka B. (2004). T-optimum designs for multiresponse heteroscedastic models. In Di Bucchianico, A., Läuter, H. and Wynn, P.H. (eds.) *mODa 7-Advances in Model-Oriented Design and Analysis*. Physica-Verlag, Heidelberg, New-York, 191–199.

Distribución geográfica y temporal de la mortalidad por cáncer de estómago en Catalunya, 1986-2000

Xavi Puig¹, Rosa Gispert², Lourdes Rodero¹, Josep Ginebra¹

¹xavier.puig@upc.edu, Departament d'Estadística i Investigació Operativa,
Universitat Politècnica de Catalunya

² Servei d'Informació i Estudis. Departament de Salut. Generalitat de
Catalunya

Abstract

Este estudio analiza los cambios en la distribución espacial, de la mortalidad por cáncer de estómago en Cataluña, durante el período 1986-2000. Para el análisis se ha utilizado un modelo jerárquico bayesiano. El patrón observado se caracteriza por concentrar las zonas con menor riesgo de mortalidad en el litoral y las de mayor riesgo en el interior.

Palabras Clave: Mortalidad. Distribución espacial. Análisis bayesiano

1. Introducción

El estudio de los patrones de la mortalidad por cáncer es una herramienta de síntesis de la información útil para monitorizar la salud de la población, y ayudar en la evaluación y planificación de los recursos necesarios para su control. A pesar de la disminución de la mortalidad por cáncer de estómago en el marco europeo, éste sigue siendo una causa de muerte muy frecuente. Además es uno de los tumores con peor pronóstico.

2. Objetivo del trabajo

El objetivo de este estudio es analizar la distribución espacial y temporal de la mortalidad por cáncer de estómago por sexo en Catalunya, durante el período 1986-2000.

3. Material y Métodos

Las defunciones por cáncer de estómago de los años 1986 al 2000 proceden del Registro de Mortalidad de Catalunya, y la población del Institut d'Estadística de Catalunya. Para el análisis espacial como unidades geográficas se han seleccionado las 41 comarcas catalanas. Para cada quinquenio (1986-1990, 1991-1995 y 1996-2000) y sexo se ha calculado la Razón de Mortalidad estandarizada suavizada utilizando el modelo jerárquico espacial propuesto por Besag, York y Mollié, y se han comparado dichos modelos al omitir la componente espacial mediante el *Deviance Information Criterion*. Finalmente, para integrar la componente temporal se ha utilizado la generalización del modelo propuesta por Bernardinelli.

4. Resultados

El análisis espacial muestra que, para los dos sexos, las zonas con más bajo riesgo se encuentran en el litoral, y el riesgo aumenta de forma gradual hacia las zonas del interior. No obstante en el último quinquenio este patrón ya no se evidencia de forma tan clara, y los valores de la *Deviance Information Criterion* disminuyen al omitir la componente espacial del modelo. Respecto a la tendencia temporal observamos una relación inversa con el riesgo debido a una homogenización de éstos.

5. Bibliografía

- [1] Bernardinelli L, Clayton D, Pascutto C, Montomoli C, Ghislandi M, Sonnini M. Bayesian analysis of space-time variation in disease risk. *Stat Med* 1995;14:2433-2443.
- [2] Besag J, York JC, Mollié A. Bayesian image restoration, with two applications in spatial statistics. *Ann Inst of Stat Math*, 1991; 43:1-20.
- [3] Levi F, Lucchini F, Gonzalez JR, Fernandez E, Negri E, La Vecchia C. Monitoring falls in gastric cancer mortality in Europe. *Ann Oncol*. 2004;15:338-45.
- [4] Spiegelhalter KJ, Thomas A, Best N, Lunn D. WinBugs: Bayesian inference Using Gibbs Sampling. Manual Version 1.4, Medical Research Council Biostatistics Unit. Cambridge, 2003 [consultado 01-10-2005]: Disponible en <http://www.mrc-bsu.cam.ac.uk/bugs/>.

Distribuciones ε -Contaminadas en Tablas de Contingencia

Miguel A. Gómez Villegas¹, Beatriz González Pérez²

¹ma.gv@mat.ucm.es, Departamento de Estadística e Investigación Operativa,
Universidad Complutense de Madrid

²beatrizg@mat.ucm.es, Departamento de Estadística e Investigación
Operativa, Universidad Complutense de Madrid

Abstract

Se desarrolla un procedimiento bayesiano para contrastar la hipótesis nula de homogeneidad en tablas de contingencia utilizando distribuciones ε -contaminadas. Se calcula una cota inferior de la probabilidad a posteriori de la hipótesis nula y se muestra que la cota se puede hacer próxima al p-valor con un valor adecuado de la masa inicial de dicha hipótesis.

Claves: tablas de contingencia, distribuciones ε -contaminadas, p-valor.

AMS: 62F15

1. Introducción

Se pretende investigar el comportamiento de una característica Y común a r poblaciones. Para ello, se dispone de r muestras aleatorias independientes de tamaños n_i , $\sum_{i=1}^r n_i = n$ y se clasifican sus individuos en s clases disjuntas, C_j , $j = 1, \dots, s$. La frecuencia observada en C_j para la muestra i es n_{ij} , $m_j = \sum_{j=1}^s n_{ij}$.

Sea $\Theta = \left\{ \mathbf{p} = (p_1, \dots, p_s) \in (0, 1)^s, \sum_{j=1}^s p_j = 1 \right\}$ y $\mathbf{p}_i = (p_{i1}, \dots, p_{is}) \in \Theta$, con $p_{ij} = P(Y_{ij} \in C_j)$, $j = 1, \dots, s$, $i = 1, \dots, r$. El objetivo es contrastar

$$H_0 : \mathbf{p}_1 = \dots = \mathbf{p}_r = \mathbf{p}_0 \text{ versus } H_1 : \exists i \neq j, \mathbf{p}_i \neq \mathbf{p}_j, \quad (2.19)$$

donde $\mathbf{p}_0 = (p_{01}, \dots, p_{0s}) \in \Theta$ denota el valor común en la hipótesis nula. Por conveniencia y sin pérdida de generalidad se va a suponer que \mathbf{p}_0 es conocido.

2. Distribución a Priori

Sea la clase de distribuciones ε -contaminadas,

$$\Gamma = \left\{ \pi = (1 - \varepsilon) q_0 + \varepsilon q, q \in Q \right\}. \quad (2.20)$$

q_0 es una distribución a priori particular, Q es la clase de todas las posibles distribuciones y ε fijo, $0 < \varepsilon < 1$, es la contaminación introducida en q_0 .

Se propone utilizar (Gómez-Villegas y Sanz, 2000)

$$\pi^*(\theta) = \pi_0 I_{H_0}(\theta) + (1 - \pi_0) \pi(\theta) I_{H_1}(\theta), \quad (2.21)$$

donde $\theta = (\mathbf{p}_1, \dots, \mathbf{p}_r) \in \Theta^r$, $\pi_0 = \int_{B(\theta_0, \delta)} \pi(\theta) d\theta$, $\theta_0 = (\mathbf{p}_0, \dots, \mathbf{p}_0) \in \Theta^r$, $\pi(\theta) \in \Gamma$ y $\delta > 0$ suficientemente pequeño.

Con esta elección, $\pi_0 = (1 - \varepsilon) \pi_{q_0}^0 + \varepsilon \pi_q^0$, $\pi_{q_0}^0 = \int_{B(\theta_0, \delta)} q_0(\theta) d\theta$ y $\pi_q^0 = \int_{B(\theta_0, \delta)} q(\theta) d\theta$.

3. Acotación de la Probabilidad a Posteriori

En la misma línea de Berger y Sellke (1987), se pretende minimizar la probabilidad a posteriori de H_0 en la clase Γ dada en (2.20).

Teorema 1. Sea el contraste (2.20) y la distribución a priori dada en (2.21). Entonces, cuando $\pi \in \Gamma$,

$$P(H_0 | n_{11}, \dots, n_{rs}) \geq \left[1 + \frac{1 - (1 - \varepsilon) \pi_{q_0}^0}{(1 - \varepsilon) \pi_{q_0}^0} \eta_\varepsilon(n_{11}, \dots, n_{rs}) \right]^{-1}, \quad (2.22)$$

donde $\eta_\varepsilon(n_{11}, \dots, n_{rs}) = (1 - \varepsilon) \eta(n_{11}, \dots, n_{rs}) + \varepsilon \frac{\sup_{\theta \notin H_0} f(n_{11}, \dots, n_{rs} | \theta)}{\prod_{j=1}^s p_{0j}^{\sum_{i=1}^r n_{ij}}}$, con
 $\eta(n_{11}, \dots, n_{rs}) = \frac{\int_{\Theta^r} \prod_{i=1}^r \prod_{j=1}^s p_{ij}^{n_{ij}} q_0(\theta) d\theta}{\prod_{j=1}^s p_{0j}^{\sum_{i=1}^r n_{ij}}}$ y $f(n_{11}, \dots, n_{rs} | \theta)$ la verosimilitud.

4. Comparaciones

Para las comparaciones numéricas entre la cota inferior calculada en el Teorema 1 y el p-valor del test χ^2 de Pearson se han considerado tablas 2×2 , $p_0 = 1/2$, $\varepsilon = 0$ y distribuciones a priori uniformes independientes. Mediante un análisis de datos se obtiene que si se elige $\pi_{q_0}^0 \in (0,09, 0,11)$, entonces para $\delta \in (0,17, 0,19)$ las cotas inferiores son aproximadamente iguales a los p-valores usuales. Además la discrepancia es más acusada cuando se utiliza $\pi_{q_0}^0 = 1/2$.

5. Bibliografía

- [1] Berger, J. O. and Sellke, T. (1987). Testing a Point Null Hypothesis: The Irreconcilability of p-Values and Evidence, (with Discussion). *J. Amer. Statist. Assoc.*, **82**, 112-139.
- [2] Gómez-Villegas, M. A. and Sanz, L. (2000). ε -contaminated priors in testing point null hypothesis: a procedure to determine the prior probability. *Statistics & Probability Letters*, **47**, 53-60.

Economic design of EWMA control chart using regions of maximum and minimum ARL

J. Carlos García-Díaz, Francisco Aparisi

Dpto. de Estadística e Investigación Operativa Aplicadas y Calidad

Universidad Politécnica de Valencia, 46022 Valencia, España

jugardi@eio.upv.es

Abstract

The exponentially weighted moving average control chart (EWMA), although originally developed to detect effectively small shifts, it can be designed to cope with these requirements. This work presents a method for the economic-statistical design of EWMA charts for control processes in which no small shifts need to be detected, efficient at the same time in detecting important shifts. Genetic algorithms have been used to optimize this design. A sensitivity analysis of the optimal solution was carried out to determine the influence of some factors on the economic model.

Keywords: Statistical Process Control, In-control, Indifference, and Out-of-control regions, Genetic Algorithms.

AMS: 62P30

1. Introduction

Nowadays it is frequent to find industries with very capable productive processes or process that are very difficult to adjust. In these cases, to detect very small shifts may be not interesting due to the possible extra variation introduced in the process. It would be more interesting in these situations to decide what shift size is important to detect, and to design the chart to detect quickly this shift, having low probability of false alarms for the shifts that we do not want to detect [Woodall (1985)].

Exponentially weighted moving average (EWMA) schemes were introduced by Roberts as an alternative to the Shewhart control chart. The EWMA chart accumulates data from past samples and is therefore more sensitive to average small-sustained process shifts.

However, situations may arise where it is not relevant to detect small-magnitude shifts. For example, attempting to adjust a process when the mean shift is very small may lead to the phenomenon of over-adjustment, and to the appearance of extra variability in the process. For the reasons described above, Woodall first recommends defining which shift sizes are important to be detected quickly, as a criterion for the design of control charts.

2. Cost model for regions of maximum and minimum ARL

Montgomery et al. present the economic design of the EWMA chart, extending the Lorenzen-Vance flexible cost model. Woodall studied the statistical design of control charts and recommended taking a selection of shift sizes important to be detected as a criterion for the design of control charts. For this, he proposed the definition of three regions: an in-control region (region of maximum *ARL*), an indifferent region and a region out of control (region of minimum *ARL*).

The aim is to find the value of the design parameters of the EWMA control chart, using regions of maximum and minimum *ARL* that minimise the mean cost function per hour of production, for a shift of magnitude δ in the mean.

3. Design

Genetic Algorithms have been used in this paper as an optimisation tool to find the parameters of the sampling plan that minimise the cost function.

4. Conclusions

An economic model has been developed for EWMA control charts for regions of maximum and minimum *ARL* in order to obtain optimal sampling plans, from the perspective of obtaining minimum mean costs of operation per time unit to satisfy statistical constraints, imposed as in-control and out-of-control regions.

5. Bibliografía

- [1] Woodall, W.H. The statistical design of quality control charts. *The Statistician* 1985; 34. pp. 155-160
- [2] Lorenzen, T. J. and Vance, L. C. (1986). The economic design of control charts: A unified Approach. *Technometrics*, 28, 3-10.
- [3] Montgomery, D.C.; Torng, J. C.; Cochran, J. K. and Lawrence, F. P. (1995). Statistically constrained economic design of the EWMA control chart. *Journal of Quality Technology*, 27, 250-256.
- [4] Aparisi, F. and Garcia-Diaz, J.C. Optimization of Univariate and Multivariate Exponentially Weighted Moving-average Control Charts using Genetic Algorithm. *Computers and Operations Research* 2004; 31, pp. 1437-1454.

El determinante de la matriz de información de Fisher y su aditividad

Enrique González-Dávila¹, Roberto Dorta-Guerra¹, Josep Ginebra²

¹egonzale@ull.es, rodorta@ull.es, Departamento de Estadística, Investigación Operativa y Computación, Universidad de La Laguna

²josep.ginebra@upc.edu, Departamento d'Estadística i Investigació Operativa, Universitat Politècnica de Catalunya

Abstract

La matriz de información de Fisher de una secuencia de experimentos independientes es la suma de sus matrices de información. Si bien el criterio más usado para medir la información es el determinante de ellas, la aditividad no se tramite directamente, aunque sí se verifica un tipo especial de aditividad.

Claves: D-Optimalidad; Experimentos a dos niveles; Modelos lineales generalizados.

AMS: 62K05; 62K15; 62J12

1. Introducción

La matriz de información de Fisher de una secuencia de experimentos independientes es la suma de sus matrices de información. La información de un experimento dado es habitual medirla a través de valores reales obtenidos de dichas matrices, generalmente a partir de sus determinantes o trazas (D- y A-optimalidad). Bajo estas circunstancias no se verifica que tal información sea la suma de los valores que ellos toman sobre los experimentos individuales. No obstante, cuando la información es medida a través del determinante de la matriz de Información de Fisher se verifica una aditividad especial.

Consideremos que observamos un conjunto de respuestas, y_{ij} , con $j = 1, \dots, n_i$, bajo un conjunto de q condiciones experimentales distintas, cada una caracterizada por los valores de $x_i = (x_{1i}, \dots, x_{ki})$, para $i = 1, \dots, q$. En experimentos lineales generalizados, las distribuciones de $y_{ij}|x_i$ se asumen independientes y de la familia exponencial, con

$$E[y_{ij}|x_i] = F(\beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki}) = F(z_i), \quad (2.23)$$

donde $F(\cdot)$ es una función monótona conocida, y donde $z_i = F^{-1}(E[y_{ij}|x_i])$ podría ser un polinomio de p variables regresoras. La familia de experimentos lineales generalizados incluyen situaciones de observaciones de modelos lineales normales, logísticos y log-lineales (ver McCullagh and Nelder (1989)).

La matriz de información de Fisher para este tipo de experimentos es una matriz $(k+1) \times (k+1)$ y puede ser escrita como

$$I(\beta) = \sum_i^q \begin{pmatrix} n_i h(z_i) & n_i h(z_i)x_{1i} & \dots & n_i h(z_i)x_{ki} \\ n_i h(z_i)x_{1i} & n_i h(z_i)x_{1i}^2 & \dots & n_i h(z_i)x_{1i}x_{ki} \\ \dots & \dots & \dots & \dots \\ n_i h(z_i)x_{ki} & n_i h(z_i)x_{1i}x_{ki} & \dots & n_i h(z_i)x_{ki}^2 \end{pmatrix}, \quad (2.24)$$

con $h(z_i) = F'(z_i)^2 / (\psi V[F(z_i)])$ y ψ parámetro de escala.

2. El determinante de $I(\beta)$ y su aditividad

La elección de un experimento requiere de la definición de los valores de q, x_i y n_i , típicamente tomando el experimento que maximiza el determinante de la matriz de información de Fisher (minimizando el volumen del elipsoide de confianza usual para β).

Sea X la matriz $q \times (k+1)$ del diseño, cuya i -ésima fila es $(1, x_{1i}, \dots, x_{ki})$, y sea $X_{\{i_1, \dots, i_{k+1}\}}$ la submatriz $(k+1) \times (k+1)$ de X formada por las filas i_1, i_2, \dots, i_{k+1} . El siguiente resultado, presentado en Fedorov (1972, p.84) para modelos lineales normales, expone un tipo diferente de aditividad, descomponiendo la información de cualquier experimento lineal generalizado de cualquier orden en la suma de las informaciones de todos los $\binom{q}{k+1}$ sub-experimentos superpuestos formados al considerar las observaciones asociadas a $k+1$ de las q filas de X . Esto es,

$$\det(I(\beta)) = \sum_{\{i_1, \dots, i_{k+1}\}} \det(I_{\{i_1, \dots, i_{k+1}\}}(\beta)), \quad (2.25)$$

donde $I_{\{i_1, \dots, i_{k+1}\}}(\beta)$ representa la información del experimento restringido a las observaciones asociadas a las filas i_1, \dots, i_{k+1} de la matriz de diseño X . Además se verifica que,

$$\det(I_{\{i_1, \dots, i_{k+1}\}}(\beta)) = (\det(X_{\{i_1, \dots, i_{k+1}\}}))^2 n_{i_1} h(z_{i_1}) \dots n_{i_{k+1}} h(z_{i_{k+1}}). \quad (2.26)$$

Estos resultados son aplicados a experimentos a dos niveles soportados sobre los 2^p vértices de un hiper-rectángulo en \mathbb{R}^p incluyendo diseños factoriales.

3. Bibliografía

- [1] Box, G.E.P., Hunter, W.G. and Hunter, J.S. (2005). *Statistics for Experiments, 2nd ed.* Wiley, New York.
- [2] Fedorov, V.V. (1972). *Theory of Optimal Experiments*. Academic Press, New York.
- [3] McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models, 2nd ed.* Chapman Hall, London.

El modelo de regresión parcialmente lineal semi-funcional para series de tiempo

Germán Aneiros Pérez¹, Philippe Vieu²

¹ganeiros@udc.es, Departamento de Matemáticas, Universidade da Coruña

²vieu@cict.fr, Laboratoire de Statistique et Probabilités, Université Paul Sabatier, Toulouse³

Abstract

Presentamos un modelo de regresión cuya función de regresión se expresa como la suma de una componente lineal y otra no paramétrica, siendo de naturaleza funcional la variable explicativa de la componente no paramétrica. Bajo condiciones de dependencia tipo mixing, se obtienen algunos resultados asintóticos. Mostramos una aplicación a datos reales.

Claves: Datos funcionales, Inferencia no paramétrica, Series de tiempo.

1. Introducción

El modelo parcialmente lineal semi-funcional puede ser expresado como

$$Y = \sum_{j=1}^p X_j \beta_j + m(T) + \varepsilon, \quad (2.27)$$

donde las variables explicativas X_j son reales, T es otra variable explicativa pero de naturaleza funcional (véase [3] para una monografía referente a la estimación no paramétrica con datos funcionales) y ε es un error aleatorio verificando

$$E(\varepsilon | X_1, \dots, X_p, T) = 0.$$

Suponemos que $\{(Y_i, X_{i1}, \dots, X_{ip}, T_i)\}_{i=1}^n$ es una muestra de vectores idénticamente distribuidos según la distribución de (Y, X_1, \dots, X_p, T) , y que proviene de algún proceso estacionario. Además, T toma valores en algún espacio semi-métrico abstracto \mathcal{H} . En el caso particular en que $\mathcal{H} = R$ ($\dim(\mathcal{H}) = 1 < \infty$), puede verse en [1] un estudio reciente del modelo (2.27) bajo dependencia. En [2] se presenta una investigación del modelo (2.27) bajo independencia y $\dim(\mathcal{H}) = \infty$.

Estimamos el vector de parámetros β y la función m en (2.27) por medio de

$$\widehat{\beta}_h = \left(\tilde{\mathbf{X}}_h^T \tilde{\mathbf{X}}_h \right)^{-1} \tilde{\mathbf{X}}_h^T \tilde{\mathbf{Y}}_h$$

y

$$\widehat{m}_h(t) = \sum_{i=1}^n w_{n,h}(t, T_i) (Y_i - \mathbf{X}_i^T \widehat{\beta}_h),$$

respectivamente. En estos estimadores h es un parámetro de suavizado que típicamente aparece en cualquier escenario de estimación no paramétrica. Además, hemos denotado $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)^T$ con $\mathbf{X}_i = (X_{i1}, \dots, X_{ip})^T$, $\mathbf{Y} = (Y_1, \dots, Y_n)^T$ y, para cualquier $(n \times q)$ -matrix \mathbf{A} , $\tilde{\mathbf{A}}_h = (\mathbf{I} - \mathbf{W}_h)\mathbf{A}$, donde $\mathbf{W}_h = (w_{n,h}(T_i, T_j))_{i,j}$ con $w_{n,h}(\cdot, \cdot)$ siendo una función de pesos que puede tomar diferentes expresiones. En este trabajo nos centramos en los pesos

$$w_{n,h}(t, T_i) = \frac{K(d(t, T_i)/h)}{\sum_{j=1}^n K(d(t, T_j)/h)},$$

donde $d(\cdot, \cdot)$ es la semi-métrica asociada a \mathcal{H} y K una función núcleo.

2. Resultados asintóticos

Asumiendo que la muestra $\{(Y_i, X_{i1}, \dots, X_{ip}, T_i)\}_{i=1}^n$ procede de un proceso fuertemente mixing cuyos coeficientes mixing decaen algebraicamente, obtenemos la normalidad asintótica de $\hat{\beta}_h$, así como las razones de convergencia casi segura de $\hat{\beta}_h$ y \hat{m}_h . Por tanto, generalizamos los resultados de [2] (obtenidos bajo independencia) al caso de dependencia fuertemente mixing.

3. Aplicación a datos reales

Una aplicación a datos reales ilustrará la utilidad del modelo.

4. Agradecimientos

La investigación de G. Aneiros ha estado parcialmente financiada por el MEC (ERDF incluido) bajo el proyecto MTM2005-00429. P. Vieu desea agradecer a todos los participantes en el grupo de trabajo STAPH sobre Estadística Funcional de la Universidad Paul Sabatier de Toulouse sus estimulantes y continuos comentarios. Las actividades de este grupo están disponibles en <http://www.lsp.ups-tlse.fr/Fp/Ferraty/staph.html>.

5. Bibliografía

- [1] Aneiros-Pérez, G., González-Manteiga, W. and Vieu, P. (2004). Estimation and testing in a partial regression model under long-memory dependence. *Bernoulli* 10, 49-78.
- [2] Aneiros-Pérez, G. and Vieu, P. (2006). Semi-functional partial linear regression. *Statist. Probab. Lett.*, to appear.
- [3] Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis*. Springer, New York, in print.

El papel del Boxplot based discretization en la Interpretación conceptual de una clasificación jerárquica

Alejandra Pérez-Bonilla¹, Karina Gibert²

¹alejandra.perez@upc.edu, Departamento de Estadística e Investigación Operativa, Universitat Politècnica de Catalunya.

²karina.gibert@upc.edu, EIO-UPC.

Abstract

Se presenta la metodología *Caracterización Conceptual por Condicionamientos Sucesivos (CCCS)*, para la generación automática de descripciones conceptuales de clasificaciones y su aplicación a la interpretación de clases identificadas previamente en una planta depuradora de aguas residuales. El método permite interpretar una partición obtenida sobre un dominio poco estructurado a partir de una clasificación jerárquica.

Claves: Boxplot múltiple, Distribución condicionada, inducción de reglas.

En un proceso de clasificación automática en que se descubren las clases que componen un determinado dominio, quizás uno de los problemas más importantes y menos sistematizados que hay que enfrentar es el proceso de *interpretación* de las clases, íntimamente ligado a la *validación* de las mismas y decisivo en la posterior *utilidad* del conocimiento adquirido [7]. La interpretación de las clases, tan fundamental para entender el significado de la clasificación obtenida y, en consecuencia, la estructura del dominio, se realiza habitualmente de forma muy artesanal. Este proceso se complica enormemente conforme el número de clases y las variables que intervienen en la descripción del dominio crece. En este trabajo, se pretende abordar el problema de la generación automática de interpretaciones de una clasificación, con el objetivo de consolidar, a largo plazo, una herramienta de apoyo a ésta tarea y contribuir a la sistematización de la misma en contextos de Data Mining, donde la interpretación de los patrones generados a partir de los datos es tan o más importante que su hallazgo [3].

La metodología *CCCS* [5], parte de una clasificación jerárquica para generar un sistema de reglas, mediante el que se obtiene una interpretación automática de las clases que puede dar soporte a la construcción de *sistemas inteligentes de soporte a la toma de decisiones* en plantas depuradoras de aguas residuales. La propuesta se apoya en una serie de trabajos previos [8], integrando distintos elementos (como el *boxplot múltiple* [10], y que en nuestro contexto supone una herramienta ágil y potente para tratar variables numéricas) en una única herramienta metodológica que aprovecha la estructura jerárquica de la clasificación para superar algunas de las limitaciones de trabajos anteriores. La metodología *CCCS* [5] es un método rápido y eficaz que genera conceptos que serán de

gran apoyo al proceso de decisión posterior. A efectos interpretativos el método propuesto aporta información más rica que los métodos clásicos de inducción de árboles de decisión [9] donde en cada nodo se selecciona una única variable y, a pesar de producir interpretaciones multivariales, es más fácil de analizar que los resultados de métodos de reglas de asociación [1] que producen mayor número de reglas y más complejas [2]. Existe una primera propuesta que se ha aplicado con éxito a datos reales que vienen de una planta de tratamiento de aguas residuales. Las reglas de mayor probabilidad asociadas a cada clase se han analizado globalmente para entender la naturaleza de las clases y la interpretación obtenida está en concordancia con la que en su momento dieron los expertos a partir del análisis directo de los resultados de la clasificación [6], lo que hace pensar que la metodología *CCCS* es muy prometedora. El punto clave consiste en haber hallado una forma rápida de calcular los puntos de corte de las variables numéricas donde cambian las intersecciones entre clases [4]. En éste trabajo se compara nuestra propuesta con la anterior [11] donde la forma de generar los intervalos presentaba deficiencias que distorsionaban la interpretación y la propuesta aquí presentada resuelve esta situación.

1. Bibliografía

- [1] Agrawal, R. and Sikrant, R. (1994) *Fast algorithms for mining association rules*. In: VLDB94. Procs. OF 20th Trtl Conf. on Very large Data Bases. 487–493.
- [2] Comas, J. et alt. (2001). *Knowledge discovery by means of inductive methods in wastewater treatment plant data*. AI Communications, 14(1):45–62.
- [3] Fayyad, U. et alt. (1996). From Data Mining to Knowledge Discovery: An overview. *Advances in Knowledge Discovery and Data Mining*. AAAI/MIT Press.
- [4] Gibert, K. (2004). Técnicas híbridas de Inteligencia Artificial y Estadística para el descubrimiento de conocimiento y la minería de datos. *Tendencias de la minería de datos en España*. Red Nacional de MiDA. Digital3. 119–130.
- [5] Gibert, K. and Pérez-Bonilla, A. (2005) *Ventajas de la estructura jerárquica del clustering en la interpretación automática de clasificaciones*. III TAMIDA Workshop, Granada, THOMPSON 67-76.
- [6] Gibert, K. and Roda, I. (2000.) Identifying characteristic situations in wastewater treatment plants. *Workshop BESAI (ECAI2000)*, V1, 1–9.
- [7] Gordon, A. D. (1994). Identifying genuine clusters in a classification. *Computational Statistics and Data Analysis* 18, 561–581.
- [8] Pérez-Bonilla, A. and Gibert, K. (2005) *Análisis y propiedades de la metodología Caracterización Conceptual por Condicionamientos Sucesivos (CCCS)*. DR 2005/14. EIO-UPC, Barcelona.
- [9] Quinlan, J.R. (1993). *C4.5, programs for machine learning*. Machine Learning.
- [10] Tukey, J.W. (1977). *Exploratory Data Analysis*. Addison-Wesley.
- [11] Vázquez, F and Gibert, K. (2001). *Generación Automática de Reglas Difusas en Dominios Poco Estructurados con Variables Numéricas*. In Proc. CAEPIA01 , v. 1, pp 143–152.

El tratamiento de la variabilidad en datos procedentes de escalas de diferenciación semántica

Lluís Marco Almagro¹, Xavier Tort-Martorell Llabrés²

¹lluis.marco@upc.edu, ²xavier.tort@upc.edu

Departamento de Estadística e Investigación Operativa, Universitat
Politècnica de Catalunya

Abstract

En la ingeniería kansei se utiliza a menudo el método de escalas de diferenciación semántica para capturar las impresiones de los usuarios ante prototipos de producto. En el análisis posterior suele trabajarse con el promedio de las puntuaciones. Plantearemos los riesgos de no pensar en la variabilidad presente y estudiaremos alternativas para considerarla.

Palabras Clave: Kansei Engineering, Semantic Differential Method

AMS: 62P30, 62H99

1. La ingeniería kansei

La ingeniería kansei es un método para convertir las impresiones y sensaciones de los usuarios en parámetros de diseño de los productos. El método fue inventado por Mitsuo Nagamachi en la década de 1970. En la ingeniería kansei se pretende capturar las impresiones de los usuarios respecto a determinados prototipos con parámetros de diseño diferentes, de forma que sea posible después determinar las emociones que el producto transmitirá en función de su diseño.

2. El método de las escalas de diferenciación semántica

A menudo se utiliza el método de escalas de diferenciación semántica (Osgood, 1957) para recoger los datos sobre las impresiones de los usuarios. Se tiene entonces una matriz de tres dimensiones: los distintos prototipos, las distintas escalas de valoración (las palabras kansei) y los distintos usuarios que han participado en el estudio. Cada celda contiene una puntuación (por ejemplo, de 1 a 7).

La matriz de tres dimensiones suele colapsarse a una matriz de 2 dimensiones tomando la media de todos los datos de uno de los ejes. Sobre esta nueva matriz colapsada pueden aplicarse técnicas de análisis factorial para intentar reducir el número de palabras kansei de los que se dispone.

En este trabajo analizaremos, por un lado, las dificultades asociadas al hecho que la variabilidad en las valoraciones que dan los usuarios puede ser elevada y que esta información se pierde al colapsar la matriz usando la media de

todos los participantes en el estudio. Por otro lado, estudiaremos sistemas para intentar centrarnos no solo en la media, sino también en la varianza, sin perder simplicidad.

3. Bibliografía

- [1] Schüte, S., Eklund, J., Axelsson, J., Nagamachi, M. (2004). *Concepts, methods and tools in Kansei Engineering*. Theoretical issues in Ergonomics Science 5 (3), 214-231.
- [2] Schüte, S. (2005). *Engineering Emotional Values in Product Design*. Tesis doctoral, Linköpings universitet.
- [3] Osgood, C.E., Suci, G.J., Tannenbaum, P.H. (1957). *The Measurement of Meaning*. University of Illinois Press.

Ensayos en motores diesel mediante técnicas de diseño de experimentos para desarrollar nuevos conceptos de combustión

Jose Miguel Carot¹, José Jabaloyes¹, Ricardo Novella², Jesús Benajes², Mónica Martínez¹, Andrea Conchado¹

¹Departamento de Estadística, Investigación Operativa Aplicadas y Calidad,
Universidad Politécnica De Valencia

^{2,3}CMT, Máquinas y Motores, Universidad Politécnica de Valencia
jcarot@eio.upv.es

Abstract

La complejidad en el funcionamiento de los motores Diesel ha crecido de forma considerable en los últimos años debido a las nuevas soluciones técnicas propuestas para reducir las emisiones contaminantes y aumentar la eficiencia energética (disminuyendo por tanto el consumo de combustible). Esto ha provocado que en el desarrollo de motores, la modelización de estas variables dependa cada vez más de un número elevado de parámetros operativos de motor. De este modo, el número de ensayos experimentales necesarios para poder predecir y optimizar conjuntamente emisiones y consumo aumenta de forma importante, incrementando de esta forma el tiempo y coste de la fase experimental. Además, la realización de estos ensayos experimentales condiciona la aplicabilidad y validez de los modelos de predicción. La aplicación de las técnicas de diseño estadístico de experimentos a los ensayos con motores puede aportar a este campo las siguientes ventajas: reducir el número de ensayos a realizar, estimar los efectos individuales de los factores controlados y las interacciones entre ellos, proporcionar un mejor conocimiento de cómo afectan los parámetros de operación del motor al proceso de combustión y a la formación de contaminantes, proporcionar modelos matemáticos válidos para la predicción de estas variables, manejar todas las variables respuesta (emisiones contaminantes y consumo) simultáneamente y finalmente determinar las condiciones óptimas de funcionamiento del motor de acuerdo con determinados objetivos como cumplir la futura normativa anticontaminante europea para los motores Diesel (EURO V).

Palabras Clave: Diseño de experimentos, Motores diesel, modelización, optimización

Entorno virtual de soporte a la docencia en Estadística

*J. Daunis-i-Estadella¹, J.Poch García², S. Thió-Henestrosa³,
F.Prados Carrasco⁴*

¹josep.daunis@udg.es, ²jordi.poch@udg.es, ³santiago.thio@udg.es,
⁴ferran.prados@udg.es,

Departament d'Informàtica i Matemàtica Aplicada, Universitat de Girona

Abstract

La plataforma de e-learning ACME es un entorno virtual de soporte a la docencia presencial que se encuentra en funcionamiento y que permite la asignación personalizada de problemas, corrección automática y su posterior evaluación. Presentamos el proyecto de ampliación que se realiza sobre la plataforma ACME para ser usada como soporte a la docencia de la Estadística.

Palabras Clave: docencia virtual, evaluación continuada, innovación docente

AMS: 97U50

1. Introducción

Es evidente que las nuevas tecnologías abren en el campo de la enseñanza una serie de nuevas posibilidades que nos obligan a replantear la metodología que estamos aplicando en nuestro quehacer diario. Uno de los aspectos que se pueden replantear es el control y la incentivación del trabajo continuado de los alumnos con el objetivo de incrementar su rendimiento académico final, sin que se haga necesaria la presencia ni la acción directa del profesor. Esta incentivación del trabajo regular es posible mediante las facilidades de uso y la gran implantación de las tecnologías de la información y la comunicación y, en particular, de la plataforma ACME.

El Proyecto ACME (Evaluación Continuada y Mejora de la Enseñanza [1]), se inició en el año 1998 en el Departament d'Informàtica i Matemàtica Aplicada (IMA) de la Universitat de Girona (UdG) y se enmarcó dentro de un replanteamiento de la docencia en las asignaturas de matemáticas de las ingenierías de la Escola Politècnica Superior (EPS) de la UdG.

Los objetivos principales de la plataforma ACME consisten en implementar un sistema eficiente de evaluación y trabajo continuados, mediante la asignación de problemas de forma personalizada para cada alumno, ofreciendo un sistema de ayuda a la resolución de problemas, facilitando la comunicación recíproca profesor-alumno y el seguimiento y evaluación de éstos por parte del profesor.

La plataforma se ha continuado desarrollando, ampliando su planteamiento inicial restringido a las matemáticas, a asignaturas tan diversas como programación, física o economía. A su vez se han visto ampliadas sus funcionalidades

con nuevos módulos de carácter pedagógico, capacidad de corrección de distintos tipos de problemas ([2]), creación y gestión de grupos, calculadora y un largo etcétera.

Paralelamente, en el año 2000, los profesores de EIO del Departamento IMA de la UdG iniciaron el Proyecto ESTIC (La EStadística y la Tecnologías de la Información y la Comunicación [3]) con la creación de una base de problemas tipo test para el autoaprendizaje y la ejercitación no evaluable por parte de los alumnos y para la evaluación. Se disponía también de unas herramientas de seguimiento de usuarios y administración de preguntas basadas en las funcionalidades de la Unidad de soporte a la docencia virtual (USDV) de la UdG. La extinción de la USDV truncó el desarrollo del proyecto.

En ese trabajo mostramos la integración del proyecto ESTIC dentro del proyecto ACME, las dificultades de implementación y las ventajas que presenta la plataforma ACME.

Asimismo se presentan los resultados preliminares que se han obtenido en el primer cuatrimestre de este curso 2005-2006, donde se ha llevado a cabo una primera fase piloto en la que se ha aplicado a los estudios de ingenierías técnicas y a la ingeniería técnica en informática de gestión de la EPS, así como el planteamiento de nuevos recursos estadísticos como pueden ser los problemas parametrizados con corrección automática, la generación de problemas basados en datos aleatorios y la creación de un repositorio de teoría enlazado a los temas.

2. Bibliografía

- [1] Poch García, J.; Barrabés Vera, J.; Juher Barrot, D.; Ripoll Misse, J.; Soler Villanueva, J. y Calsina Ballesta, A. (2003) ACME 2.0 un sistema de evaluación continuada y ayuda a la resolución de problemas. Actas XI Congreso Universitario de Innovación Educativa en las Enseñanzas Técnicas. Vilanova i la Geltrú. ISBN 84-688-2216-7.
- [2] Prados Carrasco, F.; Boada Oliveras, I; Soler Masó, J. and Poch García, J. (2005) Automatic generation and correction of technical exercises. International Conference on Engineering and Computer Education: ICECE'05, Madrid. ISBN 84-6098149-5
- [3] Barceló-Vidal, C.; Daunis-Estadella, J.; Martín-Fernández, J. A.; Martín-Verdejo, F.; Mateu-Figueras, G. y Thió-Henestrosa, S. (2003). La Estadística y las Tecnologías de la Información y la Comunicación: Proyecto de Innovación Docente. Actas del XXVII Congreso Nacional de Estadística e Investigación Operativa, Lleida (E). CD-ROM. ISBN 84-8409-955-5, p. 1021-1031.

Estadísticos basados en divergencias para el diagnóstico de modelos

M.D. Esteban¹, T. Hobza², Y. Marhuenda¹, D. Morales¹

¹md.esteban@umh.es, y.marhuenda@umh.es, d.morales@umh.es, Centro de Investigación Operativa, Universidad Miguel Hernández de Elche

²hobza@km1.fjfi.cvut.cz, Departamento de Matemáticas, Universidad Técnica Checa, Praga

Abstract

En este trabajo, se introduce una familia de estadísticos basados en divergencias para el caso de observaciones no idénticamente distribuidas utilizando la idea propuesta por Jiang (2001). Se obtiene su distribución asintótica y su versión bootstrap. Finalmente, se realiza un análisis de su comportamiento para el diagnóstico de modelos lineales generalizados.

Claves: Divergencias, modelos lineales generalizados, bootstrap.

1. Introducción

Los problemas de bondad de ajuste a una distribución en la recta real, $H_0 : F = F_0$, generalmente se tratan particionando el rango de los datos en intervalos disjuntos y realizando el contraste $H_0 : \mathbf{p} = \mathbf{p}_0$ de una distribución multinomial. Sean Y_1, \dots, Y_n variables aleatorias i.i.d. con función de distribución F , E_1, \dots, E_m una partición de $R = (-\infty, \infty)$ en m intervalos, $\mathbf{p} = (p_1, \dots, p_m)$ y $\mathbf{p}_0 = (p_{01}, \dots, p_{0m})$ las probabilidades reales e hipotéticas de los intervalos E_k , i.e. $p_{0k} = \int_{E_k} dF_0$, $p_k = \int_{E_k} dF$, $k = 1, \dots, m$. Definimos $N_k = \sum_{j=1}^n \mathbb{1}_{(Y_j \in E_k)} = \#\{1 \leq j \leq n : Y_j \in E_k\}$ y la probabilidad estimada $\widehat{\mathbf{p}} = (\widehat{p}_1, \dots, \widehat{p}_m)$ con $\widehat{p}_k = N_k/n$, $k = 1, \dots, m$. Para realizar el contraste H_0 uno de los estadísticos más utilizados es el de Pearson, que es un caso particular de la familia de estadísticos divergencia

$$T_n^\phi(\widehat{\mathbf{p}}, \mathbf{p}_0) = \frac{2n}{\phi''(1)} D_\phi(\widehat{\mathbf{p}}, \mathbf{p}_0) = \frac{2n}{\phi''(1)} \sum_{k=1}^m p_{0k} \phi\left(\frac{\widehat{p}_k}{p_{0k}}\right), \quad (2.28)$$

donde $D_\phi(\cdot, \cdot)$ denota la ϕ -divergencia de dos distribuciones de probabilidad introducidas en [1] para ϕ en el conjunto Φ de las funciones reales convexas definidas en $[0, \infty)$, continuamente diferenciables entorno al 1 y que satisfacen $\phi(1) = \phi'(1) = 0$, $\phi''(1) > 0$.

Los contrastes de hipótesis compuestas se suelen resolver asumiendo que la función de distribución F es miembro de una familia paramétrica $\{F_\theta\}_{\theta \in \Theta}$ para un subconjunto abierto $\Theta \subset R^d$. En estos casos la probabilidad de las celdas depende de un parámetro desconocido θ , i.e. $p_k(\theta) = \int_{E_k} dF_\theta$, de forma que

pueden ser estimados con estimadores de mínima ϕ -divergencia que verifican $\widehat{\theta}_\phi = \operatorname{arginf}_{\theta \in \Theta} D_\phi(\widehat{\mathbf{p}}, \mathbf{p}(\theta))$, y que contienen como caso particular al estimador de máxima verosimilitud (EMV) basado en datos cuantificados. Bajo $H_0 : F = F_\theta$ y para cualquier $\phi_1, \phi_2 \in \Phi$, en [2] se dan condiciones de regularidad para que $T_n^{\phi_1}(\widehat{\mathbf{p}}, \mathbf{p}(\widehat{\theta}_{\phi_2})) \xrightarrow{L} \chi_{m-d-1}^2$ cuando $n \rightarrow \infty$. Sin embargo, si el EMV $\widehat{\theta}$ se basa en los datos originales, la distribución asintótica de $T_n^\phi(\widehat{\mathbf{p}}, \mathbf{p}(\widehat{\theta}))$ bajo H_0 es una combinación lineal de variables independientes χ_1^2 .

Si las variables originales son independientes con funciones de distribución F_1, \dots, F_n , dependientes de un parámetro desconocido $\theta \in \Theta \subset R^d$ abierto, la hipótesis de interés es $H_0 : Y_1 \sim F_1, \dots, Y_n \sim F_n$. Definimos $p_k(\theta) = \mathbb{E}_\theta[N_k]/n$, con $\mathbb{E}_\theta[N_k] = \sum_{j=1}^n P_\theta(Y_j \in E_k)$. Para contrastar H_0 , Jiang [3] propone el estadístico $\chi_J^2(\widehat{\mathbf{p}}, \mathbf{p}(\widehat{\theta})) = n \sum_{k=1}^m (\widehat{p}_k - p_k(\widehat{\theta}))^2$, donde $\widehat{\theta}$ es un estimador consistente de θ y da condiciones de regularidad bajo las cuales la distribución asintótica de $\chi_J^2(\widehat{\mathbf{p}}, \mathbf{p}(\widehat{\theta}))$ es una combinación lineal de variables independientes χ_1^2 .

En este trabajo se amplian los resultados de Jiang [3] a la clase de estadísticos $T_n^\phi(\widehat{\mathbf{p}}, \mathbf{p}(\widehat{\theta}))$ y se obtienen sus distribuciones asintóticas. Se introducen las versiones bootstrap y se aplican estos estadísticos al diagnóstico de modelos lineales generalizados. Finalmente se dan recomendaciones sobre la elección de ϕ utilizando simulación Monte Carlo.

2. Bibliografía

- [1] Csiszár I. (1963). Eine Informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. *Publications of the Mathematical Institute of Hungarian Academy of Sciences, Series A*, 8, 85-108.
- [2] Morales D., Pardo L. and I. Vajda (1996). Divergence between various estimates of quantized information sources. *Kybernetika*, 32, 4, 395-407.
- [3] Jiang J. (2001). A nonstandard χ^2 -test with application to generalized linear model diagnostics. *Statistics and Probability Letters*, 53, 101-109.

Estimación de la dependencia condicional de dos tests diagnósticos binarios en presencia del sesgo de verificación

J. A. Roldán Nofuentes¹, J. D. Luna del Castillo²

¹jaroldan@ugr.es, ²jdluna@ugr.es

Departamento de Estadística e Investigación Operativa. Bioestadística.

Facultad de Medicina. Universidad de Granada

Abstract

En este trabajo nosotros proponemos un algoritmo EM para estimar la dependencia condicional de dos tests diagnósticos binarios en presencia de verificación parcial de la enfermedad. Los errores estándares de los estimadores de la dependencia condicional se han estimado utilizando el algoritmo SEM y aplicando el método delta.

Palabras Clave: Dependencia condicional, Sesgo de verificación, Verificación parcial.

1. Introducción

En la práctica clínica es frecuente el uso de dos o más tests de screening para detectar la presencia de un determinada enfermedad, por ejemplo en el diagnóstico del cáncer, enfermedades mentales, etc. Si los tests diagnósticos son condicionalmente dependientes a la enfermedad, el procedimiento diagnóstico es más eficiente utilizando solamente el test de screening con más sensibilidad. Si los tests de screening son condicionalmente independientes a la enfermedad, la aplicación de los tests diagnósticos en serie incrementa la sensibilidad global del procedimiento diagnóstico. Por tanto, si los tests de screening son condicionalmente independientes la sensibilidad global del procedimiento diagnóstico es mayor que si los tests son condicionalmente dependientes. Así pues, el conocimiento de la estructura de dependencia condicional de múltiples tests diagnósticos es un elemento indispensable en programas de detección de enfermedades.

El objetivo de nuestro trabajo es estimar y hacer inferencia sobre la dependencia condicional de dos tests diagnósticos binarios cuando no todos los pacientes son verificados con el gold estándar, es decir, en presencia del sesgo de verificación.

2. El modelo

Sean π_k y ν_k la sensibilidad y especificidad del k-ésimo test diagnóstico y la especificidad, con $k=1,2$. La dependencia condicional entre dos tests diagnósticos [1] cuando el paciente está enfermo es $\varepsilon_1 = \pi_1\pi_2 - \alpha_1 - 1$, y $\varepsilon_0 = 1 - \nu_1 - 1 - \nu_2 - \alpha_0 - 1$ cuando el paciente está sano, donde α_1

α_0 es la covarianza entre los dos tests diagnósticos cuando el paciente está enfermo (sano).

Obtener las expresiones explícitas de los estimadores máximo verosímiles (EMV) de α_k es complicado, sin embargo sus valores se pueden estimar aplicando el algoritmo EM [2]. Una vez obtenidos los EMV de α_k , sus varianzas se han estimado mediante el algoritmo SEM [3], y aplicando el método delta se han estimado las varianzas de los EMV de ε_k . Se han realizado unos experimentos de simulación para investigar la cobertura asintótica de los intervalos de confianza de ε_k , obteniéndose que, en términos generales, se necesitan muestras de entre 100 y 500 pacientes, dependiendo de la prevalencia de la enfermedad, para que los intervalos de confianza al 95 % de ε_k tengan una cobertura satisfactoria. Los resultados obtenidos se han aplicado al diagnóstico de la enfermedad de Alzheimer.

3. Contactar

Dirección:

José Antonio Roldán Nofuentes
 Departamento de Estadística e Investigación Operativa
 Bioestadística. Facultad de Medicina – Av. De Madrid s/n
 18071 – Granada. Spain
Teléfono: (+34) 958244080 **Fax:** (+34) 958246117
e-mail: jaroldan@ugr.es

4. Agradecimientos

Este trabajo ha sido subvencionado por el Ministerio de Ciencia y Tecnología, proyecto BFM2003-08950.

5. Bibliografía

- [1] Vacek, P. M. (1985). The effect of conditional dependence on the evaluation of diagnostic tests. *Biometrics* 41, 959-968.
- [2] Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B* 39, 1-38.
- [3] Meng, X. L. and Rubin, D. B. (1991). Using EM to obtain asymptotic variance-covariance matrices: the SEM algorithm. *Journal of the American Statistical Association* 86, 899-909.

Estimación de la exactitud corregida por azar de un test diagnóstico binario en presencia de verificación parcial

J. A. Roldán Nofuentes¹, J. D. Luna del Castillo²

¹jaroldan@ugr.es, ²jdluna@ugr.es

Departamento de Estadística e Investigación Operativa. Bioestadística.

Facultad de Medicina. Universidad de Granada

Abstract

En este trabajo se deducen, a partir del modelo de corrección de azar de Brenner y Gefeller (1994), los estimadores máximo verosímiles y varianzas de la sensibilidad y especificidad corregidas por azar de un test diagnóstico binario en presencia de verificación parcial de la enfermedad.

Palabras Clave: Corrección por azar, Sesgo de verificación, Verificación parcial.

1. Introducción

Cuando un test diagnóstico se aplica a una muestra de n pacientes, algunos de los pacientes pueden ser clasificados correctamente por el test diagnóstico debido al azar. En este trabajo, se obtienen los estimadores máximo verosímiles y varianzas de la sensibilidad y especificidad corregidas por azar cuando algunos pacientes no son verificados con el gold estándar.

2. El modelo

Sean π y γ la sensibilidad y especificidad del test diagnóstico. En presencia de verificación parcial de la enfermedad la aplicación del test diagnóstico a una muestra de n pacientes da lugar a la Tabla 1. Suponiendo que el mecanismo de datos faltantes es MAR [2], utilizando el modelo de corrección de azar propuesto por Brenner y Gefeller (1994), los estimadores máximo verosímiles de la sensibilidad y especificidad corregidas por azar son

$$\hat{\pi}_c = \frac{n_1 s_1 r_0 - s_0 r_1}{r_0 m_1} \quad \text{y} \quad \hat{\gamma}_c = \frac{n_0 s_1 r_0 - s_0 r_1}{s_1 m_0}$$

y, aplicando el método delta, los estimadores de sus varianzas son

$$\widehat{Var}(\hat{\pi}_c) \approx \frac{\{\hat{\pi}_c(1-\hat{\pi}_c)\}^2 m_1 r_0}{\{m_0 n_1 r_1 + m_1 n_0 s_0\}^2} \times \left[\frac{n^2 m_0^2 s_1 r_1 r_0}{(s_1 r_0 - s_0 r_1)^2} + \frac{\{m_0^2 n_1 s_1 r_1 + m_1 n_0 (s_0^2 r_1 + s_1 r_0^2)\}^2}{m_1 m_0 s_0 (s_1 r_0 - s_0 r_1)^2} + \frac{n m_1 n_0 r_0}{n_1} \right]$$

y

	Test positivo	Test negativo
Verificados		
Enfermos	r_1	s_0
Sanos	r_0	s_1
No verificados	n_1	n_0
Total	n_1	n_0

Cuadro 2: Tabla de frecuencias observadas.

$$\widehat{Var}(\hat{\nu}_c) \approx \frac{\{\hat{\nu}_c(1-\hat{\nu}_c)\}^2 m_0 s_1}{\{m_0 n_1 s_1 + m_1 n_0 s_0\}^2} \times \left[\frac{\{nm_1^2 s_0 r_0 + n_1(s_1 r_0 - s_0 r_1)^2\}^2}{m_1 m_0 r_1 (s_1 r_0 - s_0 r_1)^2} + \frac{n^2 m_1^2 s_1 r_0 s_0}{(s_1 r_0 - s_0 r_1)^2} + \frac{nm_0 n_1 s_1}{n_0} \right],$$

siendo $m_i = s_i + r_i$ y $n = n_1 + n_0$. Se han realizado unos experimentos de simulación para investigar la cobertura de los intervalos de confianza de los parámetros corregidos por azar, investigándose el efecto que sobre ella tienen las probabilidades de verificación y la prevalencia. Los estimadores corregidos por azar también se han deducido en presencia de covariables discretas.

3. Contactar

Dirección: José Antonio Roldán Nofuentes

Departamento de Estadística e Investigación Operativa
Bioestadística. Facultad de Medicina – Av. De Madrid s/n
18071 – Granada. Spain

Teléfono: (+34) 958244080 **Fax:** (+34) 958246117

e-mail: jaroldan@ugr.es

4. Agradecimientos

Este trabajo ha sido subvencionado por el Ministerio de Ciencia y Tecnología, proyecto BFM2003-08950.

5. Bibliografía

- [1] Brenner, H., and Gefeller, O. (1994). Chance-corrected measures of the validity of a binary diagnostic test. *Journal of Clinical Epidemiology*, 47, 627-633.
- [2] Rubin, D.B. (1976). Inference and missing data. *Biometrika* 4, 73-89.
- [3] Zhou, X.H. (1993). Maximum likelihood estimators of sensitivity and specificity corrected for verification bias. *Communication in Statistics – Theory and Methods* 22, 3177-3198.

Estimación de parámetros en modelos de mantenimiento imperfecto

R. Mullor¹, A. Sánchez², N. Martínez³

¹ruben.mullor@ua.es, Departamento de Estadística, Investigación Operativa,
Universidad de Alicante

²aisanche@eio.upv.es, ³nalzamor@eio.upv.es

Departamento de Estadística, Investigación Operativa Aplicadas y Calidad,
Universidad Politécnica de Valencia

Abstract

Los modelos de fiabilidad dinámicos permiten simular el comportamiento de los equipos considerando aspectos tales como sus condiciones de servicio o el efecto que las actividades de mantenimiento tienen sobre los mismos. Un problema importante en estos modelos es la estimación de los diferentes parámetros involucrados. Esta ponencia se centra en la estimación de dichos parámetros utilizando la función de máxima verosimilitud.

Palabras Clave: estimación, máxima verosimilitud, mantenimiento imperfecto

1. Introducción

El proceso de envejecimiento de los equipos se inicia desde el momento de su instalación. Este proceso depende de las diferentes condiciones operacionales y ambientales a las que se encuentran sometidos los equipos a lo largo de su vida operativa. Asimismo, y con el objetivo de controlar este proceso de envejecimiento, se realizan diferentes actividades de mantenimiento preventivo.

Lógicamente, el grado de envejecimiento de un equipo tiene una gran influencia en las características de fiabilidad del mismo y, por tanto, en su tasa de fallos. Así, resulta necesario disponer de un modelo de fiabilidad dinámico que considere explícitamente tanto las condiciones de servicio (ambientales y operacionales) de los equipos como el efecto de las diferentes actividades de mantenimiento preventivo realizadas.

En dichos modelos de fiabilidad las condiciones de servicio pueden ser incorporadas mediante un modelo de vida acelerada mientras que el efecto de las actividades de mantenimiento sobre el estado del equipo puede ser incluido utilizando un modelo de mantenimiento imperfecto. En la literatura existen diferentes modelos de mantenimiento imperfecto, tales como el modelo PAR (Proporcional Age Reduction) [1], el cual considera que cada actividad de mantenimiento reduce la edad ganada desde el último mantenimiento dependiendo de un parámetro de efectividad, o el modelo PAS (Proporcional Age Set-back)

[1], que asume que cada actividad de mantenimiento reduce la edad que el componente tenía inmediatamente antes de entrar en mantenimiento en función de su efectividad.

Los dos modelos anteriores, así como otros propuestos en la literatura, que permiten calcular la edad y fiabilidad de los equipos considerando mantenimiento imperfecto son función de un conjunto de parámetros que pueden ser desconocidos o inciertos tales como la tasa de fallo, la probabilidad de fallo a la demanda, la efectividad del mantenimiento, etc. Por tanto, una tarea importante cuando se utilizan dichos modelos para simular el comportamiento de un equipo consiste en la estimación de dichos parámetros.

2. Objetivos

La presente ponencia se centra en la estimación de diferentes parámetros incluidos en modelos de fiabilidad dinámicos, tales como el PAR o PAS. El proceso de estimación de dichos parámetros se realiza a partir de la función de máxima verosimilitud y utilizando un método directo de búsqueda para maximizar dicha función. Asimismo, además de obtener las estimaciones de los parámetros, se obtiene la matriz de información de Fisher y la matriz de varianzas-covarianzas.

Finalmente, se presenta un caso de aplicación sobre unas válvulas de seguridad pertenecientes a una central nuclear a partir de datos específicos de planta. En esta aplicación se asume una tasa de fallos lineal y se estiman los parámetros correspondientes a la tasa de envejecimiento y la efectividad de mantenimiento asumiendo un comportamiento PAS o PAR del equipo.

3. Bibliografía

- [1] S. Martorell, A. Sánchez, V. Serradell. (1999). Age-dependent reliability model considering effects of maintenance and working conditions. *Reliability engineering and system safety*, 64, 19-31.

Estimación de Theil para el modelo logit de regresión logística

José Trujillo Carmona

trujillo@unex.es, Departamento de Matemáticas, Universidad de Extremadura

Abstract

Ser propone una modificación de la estimación no paramétrica de Theil de los coeficientes de regresión, para la estimación del modelo de regresión logística binaria "logit".

Palabras Clave: Regresión no paramétrica

AMS: 62G08

1. Introducción

La estimación de los parámetros de la regresión logística del modelo "logit" suele hacerse por el método de máxima verosimilitud. Cuando se dispone de observaciones repetidas para la variable independiente también es posible utilizar de la estimación mínimo cuadrática.

Un procedimiento alternativo de estimación para la regresión lineal y polinómica fue propuesto por Theil en 1950 [1]. Theil propone como estimadores a los que hacen independientes, en el sentido de la correlación no paramétrica de la tau de Kendall, a las variables independientes respecto de los residuos de la regresión [2].

Este procedimiento tienen las ventajas y los inconvenientes usuales en la estadística no paramétrica de rangos: máxima robustez a cambio de una menor potencia.

En el presente trabajo se presenta una modificación del procedimiento de estimación propuesto por Theil para adaptarlo a la estimación del modelo "logit" de regresión logística.

2. Modelo de regresión logit

Denominaremos ventaja (.°dd") al cociente p/q de una variable binomial. El modelo de regresión "logit" propone una función de regresión lineal entre el logaritmo neperiano de la ventaja y la variable independiente cuantitativa:

$$\ln(p/q) = \alpha + \beta X \implies p = \frac{1}{1 + e^{-\alpha - \beta X}} = \frac{1}{1 + ke^{-\beta X}}$$

El modelo propone que el parámetro p de la distribución Binomial de las observaciones cambia en función de la variable independiente, de acuerdo con una

función logística de Verhulst.

3. Estimaciones

Las observaciones de la variable dependiente se denotan como valores unidad cuando se produce el "éxito" binomial y ceros cuando no se produce el éxito. La estimación usual es la de máxima verosimilitud.

Cuando se dispone de observaciones repetidas de la variable independiente es posible la estimación de las frecuencias relativas condicionadas a la variable independiente \hat{p}_x y \hat{q}_x y por tanto del logaritmo de la ventaja para estas estimaciones de frecuencias condicionadas $\ln[(\hat{p}/\hat{q})|x]$. En estas circunstancias es posible la estimación mínimo cuadrática de la función.

Theil propone como estimadores de los parámetros de una regresión lineal: $\hat{\beta} = \text{Mediana} \left\{ b_{ij} | b_{ij} = \frac{y_i - y_j}{x_i - x_j} \right\}$: $i < j$ y $\hat{\alpha} = \tilde{y} - \hat{\beta}\tilde{x}$. La recta que pasa por cada par de puntos es una estimación de la recta del modelo; la estimación final es la mediana del conjunto de pendientes que se pueden calcular con todos los pares de datos.

La ecuación de Verhulst es la solución de una ecuación diferencial que en el caso del modelo logit es $p' = \beta p(1 - p)$. Una estimación análoga a la de Theil se puede obtener directamente a partir de la ecuación anterior estimando las probabilidades a partir de las frecuencias relativas cuando se dispone de ellas:

$$\hat{p} = f \quad f'_{ij} = \frac{f_j - f_i}{x_j - x_i} \quad f_{ij} = \frac{f_i + f_j}{2}$$

$$b_{ij} = \frac{f'_{ij}}{f_{ij}(1 - f_{ij})} \quad \hat{\beta} = \tilde{b} = \text{Mediana}|b_{ij}|$$

Este procedimiento de estimación de la función logística, al igual que el procedimiento de Máxima Verosimilitud no precisa disponer de valores repetidos para los valores de la variable independiente ya que f_i pueden ser valores binomiales, 0 y 1.

4. Bibliografía

- [1] Theil, H. (1950). A Rank-Invariant Method of Linear and Polynomial Regression Analysis. I. *Nederl. Akad. Wetensch. Proc. Ser. A* 53, 386-392.
- [2] Sen, P.K. (1969). Estimates of the regression coefficient based on Kendall's tau. *J. Amer. Statist. Assoc.* 63, 1379-1389.

Estimación Máximo verosímil en la Distribución Lognormal triparamétrica mediante simulación

Juan Antonio Muñoz Orellana¹, José Fernando Vera Vera²,
Miguel Ángel Montero Alonso³

¹ orebert@yahoo.com. OTRI-Universidad de Granada. Complejo Administrativo Triunfo, pab 1. Cuesta del Hospicio s/n. 18071- Granada.
²jfvera@ugr.es. Dpto. de Estadística e I.O. Facultad de Ciencias. Universidad de Granada. 18071 - Granada.

³mmontero@ugr.es. Dpto. de Estadística e I.O.- E.U. Ciencias Sociales, Campus de Melilla. Universidad de Granada.

Abstract

Se propone un algoritmo basado en técnicas de simulación que permite maximizar las verosimilitudes empleadas por Munro y Wixley, Wingo y Komori e Hirose asociadas a la distribución lognormal triparamétrica.

Claves: Estimación Máximo Verosímil, lognormal, optimización

AMS: Clasificación AMS (Opcional)

1. Introducción

La distribución lognormal triparamétrica juega un papel muy importante en un gran número de campos en los que se aplica la estadística y resulta de gran interés para modelizar variables que presentan una clara asimetría positiva y no tienen su origen determinado en 0, como por ejemplo, niveles de contaminación, volúmenes fluviales, longitud de determinadas especies de plantas o reclamaciones a las compañías de seguros.

Los estimadores máximo verosímiles de esta distribución son valores infactibles para los cuales la función de verosimilitud asociada toma el valor $+\infty$.

$$L(\mu = +\infty, \sigma = +\infty, \theta = x_{(1)}) = +\infty$$

Este hecho impulsó la búsqueda de los denominados ESTIMADORES MÁXIMO VERO SIMILES LOCALES (EMVLs) que corresponden al máximo valor finito de la función de Verosimilitud.

En este trabajo se propone un algoritmo basado en técnicas de simulación que maximiza las verosimilitudes empleadas por Munro y Wixley [5], Wingo [6] y Komori e Hirose [4] asociadas a la distribución lognormal triparamétrica.

Con este método se obtienen los mismos resultados que con los propuestos por dichos autores presentando la ventaja de su sencilla implementación y escalabilidad a casos multivariantes.

2. Bibliografía

- [1] Aitchison, J. y Brown, J. A. C. (1969). *"The Lognormal Distribution with special reference to its use in economics"*. Cambridge at the University Press.
- [2] Hill, B.M. (1963). *"The three-parameter lognormal distribution and Bayesian analysis of a point-source-epidemic"*. J.A.S.A., 58. págs. 72-84.
- [3] Hirose, H. (1997). *"Maximum likelihood parameter estimation in the three-parameter log-normal distribution using the continuation method"* Computational Statistics & Data Analysis, 24. págs. 139-152.
- [4] Komori, Y. y Hirose, H. (2004). *"Easy estimation by a new parametrization for the three-parameter lognormal distribution"*. Journal of Statistical Computation & Simulation, Vol. 74, Nº 1. págs. 63-74.
- [5] Munro, A.H. y Wixley, A.J. (1970) *"Estimators based on order statistics of small samples from a three-parameter lognormal distribution"*. Journal of the American Statistical Association, Vol. 65. Issue 329. págs. 212-225.
- [6] Wingo, D. R. (1984). *"Fitting Three-Parameter Lognormal models by Numerical global optimization"*. Computational Statistics & Data Analysis, 2. págs. 13-15.
- [7] Zilinskas, A. (1978). *"Optimization of one-dimensional multimodal functions. Algorithm AS133"*. Applied Statistics 27, págs. 367 375.

Estimación No Paramétrica de la Varianza condicional en un modelo de regresión con respuesta faltante

A. Pérez González¹, J.M. Vilar Fernández², W. González Manteiga³

¹ Dep. Estadística e Investigación Operativa,(Univ. de Vigo)

Fac. de Ciencias Empresariales y Turismo, Orense. e-mail: anapg@uvigo.es

²Dep. Matematicas (Univ. de A Coruña),

Fac. de Informática, A Coruña. e-mail: ejvilar@udc.es

³Dep. Estadística e Investigación Operativa,(Univ. de Santiago)

Fac. de Matemáticas, Santiago de Compostela, e-mail: wenceslao@usc.es

Abstract

El objetivo de este trabajo consiste en estudiar la estimación no paramétrica de la varianza condicional en un modelo de regresión heterocedástico donde la variable respuesta puede tener observaciones faltantes.

1. Introducción

En los modelos de regresión heterocedásticos uno de los principales objetivos es la estimación de la varianza condicional. Nuestro objetivo es realizar un estudio comparativo de dos procedimientos de estimación no paramétricos de la función de varianza condicional en un contexto de diseño fijo y observaciones faltantes en la variable respuesta. Consideremos el siguiente modelo de regresión

$$Y_{t,n} = m(x_{t,n}) + s(x_{t,n})\varepsilon_{t,n}, \quad 1 \leq t \leq n,$$

donde $\varepsilon_{t,n}$, $1 \leq t \leq n$, son v. a. independientes, no observables, con media cero y varianza 1. Los puntos del diseño $x_{t,n}$, $1 \leq t \leq n$, siguen un diseño regular generado por una función de densidad f . Puesto que $Y_{t,n}$ puede no ser observado para algún índice t , introducimos una variable δ en el modelo tal que $\delta_{t,n} = 1$ si $Y_{t,n}$ es observada y cero si $Y_{t,n}$ es faltante. Vamos a suponer que los datos son "missing at random", es decir:

$$P(\delta_{t,n} = 1 | Y_{t,n}, x_{t,n}) = P(\delta_{t,n} = 1 / x_{t,n}) = p(x_{t,n}).$$

Es necesario un estimador no paramétrico de la función de regresión. Para ello utilizaremos los estimadores estudiados por Pérez-González, Vilar-Fernández y González-Manteiga (2006) para el caso de errores correlados y respuesta faltante. Uno de ellos consiste en usar sólo las observaciones completas (SIMPLIFICADO), mientras que el otro imputa las observaciones faltantes y luego estima con la muestra completada (IMPUTADO). El caso de datos completos y errores correlados ha sido estudiado por Vilar-Fernández y Francisco-Fernández(2006). Como estimadores no paramétricos de la varianza nos basaremos en los estimadores del tipo de Härdle y Tsybakov(1997) y de Fan y Yao(1998).

2. Estimadores No Paramétricos de la varianza

Estimadores tipo Härdle y Tsybakov (1997)

Sea $v(x) = s^2(x) = E[Y^2/x] - E[Y/x]^2$ y sea $g(x) = E[Y^2/x]$. Entonces podemos estimar $v(x)$ como

$$\hat{v}_n(x) = \hat{g}_n(x) - (\hat{m}_n(x))^2$$

En nuestro caso, como hay observaciones faltantes en Y , aplicaremos los estimadores polinómico locales (SIMPLIFICADO E IMPUTADO) para estimar no paramétricamente g y m y así obtener estimaciones de la varianza.

Estimadores tipo Fan y Yao (1998)

En primer lugar se obtienen los residuos de una estimación piloto no paramétrica de la función de regresión $\hat{m}_{h_1}(x)$:

$$\hat{r}_t = (Y_t - \hat{m}_{h_1}(x_t))^2$$

Luego se aplica el estimador polinómico local considerando como variable respuesta esos residuos al cuadrado. En nuestro caso, puede que no observemos alguno de los residuos. La estimación simplificada consiste en utilizar sólo los residuos observados, mientras que la otra imputa sobre los residuos y así completa la muestra de los residuos y estima la varianza con la muestra completada.

Nuestro objetivo es observar mediante un estudio de simulación el comportamiento de las versiones simplificada e imputada de los estimadores anteriores.

3. Bibliografía

- [1] Fan, J. y Yao, Q.(1998) Efficient estimation of conditional variance functions in stochastic regression. Biometrika 85,no. 3, 645–660.
- [2] Härdle, W. y Tsybakov, A.(1997) Local polynomial estimators of the volatility function in nonparametric autoregression. J. Econometrics 81,1, 223-242.
- [3] Pérez-González, A. Vilar Fernández, J. M. y González-Manteiga, W. (2006) Local polynomial estimator in a regression model with correlated errors and missing data. Submitted.
- [4] Vilar-Fernández y Francisco-Fernández (2006) Nonparametric estimation of the conditional variance function with correlated errors. Submitted.

Estimación simultánea de parámetros en regresión lineal con heterocedasticidad

Beatriz Lacruz¹, Ana Pérez-Palomares²

¹lacruz@unizar.es, Departamento de Métodos Estadísticos, Universidad de Zaragoza

²anapp@unizar.es, Departamento de Métodos Estadísticos, Universidad de Zaragoza

Abstract

Presentamos un método para modelar la varianza en problemas con heterocedasticidad que permite obtener la estimación de todos los parámetros del modelo simultáneamente. Además, el método se ilustra con su aplicación tanto a datos simulados como a datos reales y se compara con otras técnicas clásicas.

Claves: Regresión lineal, heterocedasticidad, ecuaciones funcionales.

AMS: 65J02, 39B22

Estimación y contrastes en modelos de retención de partículas radiactivas

M. Amo Salas, J.M. Rodríguez Díaz¹, G. Sánchez León²

¹ Dpto de Estadística, Universidad de Salamanca *juanmrod@usal.es*

² Dpto de Economía e Historia Económica, Universidad de Salamanca
guillermo@usal.es

Abstract

A partir de nuevos modelos de retención de partículas radiactivas en el cuerpo humano se calcularán diseños óptimos para la estimación de los parámetros principales, así como contrastes sobre los mismos. Varios ejemplos ilustrarán el trabajo.

Claves: D-optimización, Modelos compartimentales, Retención de radiactividad

1. Introducción

El objetivo principal del presente trabajo era la obtención de información sobre los parámetros de modelos descritos en ICRP (1997) referidos a la retención de partículas radiactivas en el cuerpo humano. Se buscarán los mejores diseños para su estimación y se procurarán diversos contrastes sobre los mismos. Se supone una situación de accidente en instalaciones que manejan materiales radiactivos.

En el caso objeto de estudio sólo habría un trabajador en el lugar del accidente, lo que implicaría que las posibles muestras tomadas sobre él estarían en principio relacionadas. Queda imposibilitada además la realización de pruebas simultáneas, ya que es necesario un periodo mínimo de descanso entre muestras consecutivas. Debido al procedimiento de detección y al tiempo necesario para la preparación del trabajador y las instalaciones en que tienen lugar las muestras, la primera de ellas podría ser tomada como mínimo 12 horas después del accidente. Estas y otras consideraciones pueden ser revisadas en un trabajo previo, López-Fidalgo y otros (2005), en el que se justificaba además la realización de la primera prueba lo antes posible, es decir, en el instante $t = 0,5$ (t en días). En dicho artículo llegaban a calcularse diseños óptimos con soporte en sólo dos puntos debido a la dificultad del modelo, obtenido a partir de los modelos compartimentales que describen el fenómeno, y formado por un largo sumatorio de cocientes de términos exponenciales.

En el presente trabajo se proponen otros diseños con soporte en distinto número de puntos, que reducirán la incertidumbre en la estimación de los parámetros del modelo. La elección del diseño adecuado dependerá de factores temporales,

n	t_0	t_1	t_2	t_3	t_4	t_5	t_6
2	0,5	69,0073					
3	0,5	64,8227	73,4266				
4	0,5	61,2305	69,3202	77,4498			
5	0,5	57,6342	65,5647	73,1975	81,1296		
6	0,5	4,91997	69,0085	76,6553	84,0896	91,8645	
7	0,5	4,83376	66,0458	73,5112	80,6306	87,7905	95,4029

Cuadro 1: Diseños localmente D-óptimos para $p = 5$ y $\sigma^2 = \rho = 1$

fisiológicos e incluso económicos. El cálculo de los nuevos diseños ha sido posible gracias a la obtención de modelos de retención que pueden ser expresados simplemente como suma (grande, eso sí) de términos exponenciales, unido a la realización de un software eficiente que ha permitido alcanzar los resultados en un tiempo finito.

2. Cálculo de diseños y realización de contrastes

El modelo que nos ocupa refleja la retención pulmonar de partículas de Urano (cuyas características están descritas por el parámetro p) como función de t (en días) tras una ingestión inicial I .

$$\eta(I, t, p) = I \sum_i \gamma_i e^{\alpha_i p + \beta_i t}.$$

La relación entre muestras vendrá dada por la matriz $\Sigma = \sigma^2 (s_{i,j})$, donde $s_{i,j} = \exp(-\rho|t_i - t_j|)$, indicando que dicha relación disminuirá con la distancia en tiempo entre ellas. El parámetro ρ es característico del trabajador. Se ha considerado el criterio más habitual, D -optimización. Los diseños óptimos calculados para $\sigma^2 = \rho = 1$ y tomando como valor inicial del parámetro no lineal $p_0 = 5$ se muestran en la Tabla 1.

Se partirá del método de la razón de verosimilitudes para diseñar contrastes sobre los parámetros utilizando la información proporcionada por la estructura de los diseños calculados.

3. Bibliografía

- [1] International Commission on Radiological Protection (1997). *Individual monitoring for internal exposure of workers*. Oxford: Pergamon Press; (ICRP) Publication 78; 1997.
- [2] López Fidalgo J, Rodríguez Díaz J.M., Sánchez G. and Santos Martín M.T. (2005) Optimal Designs for Compartmental Models with Correlated Observations. *Journal of Applied Statistics* **32**, 10, 1075-1088; 2005

Estimaciones para un conjunto de señales: Un estudio de simulación

I. Luengo¹, C.N. Hernández², P. Saavedra²

¹mluengo@dis.ulpgc.es, Departamento de Informática y Sistemas, ULPGC

²cflores@dma.ulpgc.es, saavedra@dma.ulpgc.es,
Departamento de Matemáticas, ULPGC

Abstract

A partir de un conjunto de señales medidas en los mismos instantes sobre una muestra aleatoria de una población, en [4] Luengo *et al* proponen un modelo de efectos aleatorios y estimaciones del patón y la componente individual. En este trabajo presentamos un estudio de simulación de la bondad de dichas estimaciones para distintas familias de procesos.

Claves: Series temporales, análisis espectral, estimaciones de núcleo.

1. Modelo

Partimos de r series medidas en los mismos N instantes sobre una muestra de una población $\{X_i(t_j)\}$ y sus correspondientes periodogramas $\{I_i(\omega_j) : j = 1, \dots, \nu = [N/2]\}$. Bajo las condiciones del teorema 6.2.2 de Priestley [5] tenemos

$$I_i(\omega_j) = Q_i(\omega_j) U_{ij} + R_{i,j} \quad (2.29)$$

donde $Q_i(\omega_j)$ representa la densidad espectral del proceso asociado al objeto i -ésimo y U_{ij} sigue asintoticamente una distribución exponencial de parámetro 1. Despreciando el término $R_{i,j}$ como en Franke y Härdle [3] tenemos

$$Y_{ij} = \mu(\omega_j) + Z_i(\omega_j) + e_{ij} \text{ for } i = 1, \dots, r; j = 1, \dots, \nu \quad (2.30)$$

siendo $Y_{ij} = \log I_i(\omega_j) - C_0$, $C_0 = -0,57$ es la constante de Euler, $\mu(\omega_j) = E[\log Q_i(\omega_j)]$, $Z_i(\omega_j) = \log Q_i(\omega_j) - E[\log Q_i(\omega_j)]$ y $e_{ij} = \log U_{ij}(\omega_j) - C_0$. Así se tiene $E[e_{ij}] \approx 0$ (vease Davis and Jones [1]) y $E[Z_i(\omega_j)] = 0$

2. Estimación de los parámetros

Sea el log-periodograma medio $y_{\cdot j} = (1/r) \sum_{i=1}^r y_{ij}$ para $j = 1, \dots, \nu$. Sea $K(\theta)$ una función de núcleo con las propiedades (K1) y (K2) descritas por Luengo *et al.* en [4], las estimaciones propuestas son:

$$\hat{\mu}(\omega) = \frac{1}{Nh} \sum_{j=-\nu}^{\nu} K\left(\frac{\omega - \omega_j}{h}\right) y_{\cdot j} \quad (2.31)$$

$$\widehat{Z}_i(\omega) = \frac{1}{N\lambda_i} \sum_{j=-\nu}^{\nu} K\left(\frac{\omega - \omega_j}{h}\right) (y_{ij} - \widehat{\mu}(\omega_j)) \quad (2.32)$$

donde h y λ_i son los diferentes anchos de banda.

3. Estudio de Simulación

Queremos hacer un estudio empírico de la bondad de las estimaciones definidas en la sección anterior, en términos visuales y calculando el MSE de las mismas. Hemos elegido dos familias de procesos: AR(2) y MA(2).

En ambos casos generamos $r = 1000$ series con $N = 100$ observaciones por serie.

Para el modelo AR(2) la serie i -ésima $X_i(t_j) = \epsilon_{ij} + a_i X_i(t_j - 1) + b_i X_i(t_j - 2)$. La especificidad del objeto i viene dada por (a_i, b_i) , b_i sigue una distribución uniforme en $(-0.4, 0.2)$, a_i una distribución uniforme en $(-|b_i - 1|, |b_i - 1|)$ y ϵ_{ij} es ruido blanco gaussiano con desviación típica 0.5.

Para el modelo MA(2) la serie i -ésima $X_i(t_j) = \epsilon(t_j) + a_i \epsilon(t_j - 1) + b_i \epsilon(t_j - 2)$ donde el par $\begin{pmatrix} a_i \\ b_i \end{pmatrix}$ sigue una distribución $N_2\left(\begin{pmatrix} 3 \\ 3 \end{pmatrix}, \begin{pmatrix} 2\gamma & \gamma \\ \gamma & 2\gamma \end{pmatrix}\right)$, $\gamma \geq 0$, y para cada i $\{\epsilon_i(t) : t \in \mathbf{Z}\}$ es un ruido blanco con distribución uniforme $[-\sqrt{3}, \sqrt{3}]$.

4. Bibliografía

- [1] Davis, H.T. and Jones, R.H. (1968). Estimation of the Innovation Variance of a Stationary Time Series. *Journal Amer. Statist. Assoc.*, 63, 141-149.
- [2] Diggle, P. J. and Al-Wasel, I. (1993). On Periodogram-Based Spectral Estimation for Replicated Time Series, in: Subba Rao (Ed), Developments in Time Series Analysis. *Chapman and Hall, Great Britain*, 341-354.
- [3] Franke, J. and Härdle, W. (1992). On Bootstrapping Kernel Spectral Estimates. *Ann. Statist.*, 20, 121-145.
- [4] Luengo, I., Hernández, C. N. y Saavedra, P. Spectral estimation in a Random Effect Model. *Preprint*.
- [5] Priestley, M.B. (1981). Spectral Analysis and Time Series. *Wiley, New York*.
- [6] Saavedra, P., Hernández, C.N. and Artiles, J. (2000). Spectral Analysis with Replicated Time Series. *Communications in Statistics Theory and Methods*, 29, 2343-2362.

Estimadores de Dependencia Fuerte con Anisotropía Espacio-Temporal

María P. Frías¹, María D. Ruiz-Medina², José M. Angulo²,
Francisco J. Alonso²

¹mpfrias@ujaen.es, Departamento de Estadística, Investigación Operativa,
Universidad de Jaén

²mruiz@ugr.es, jmangulo@ugr.es, falonso@ugr.es, Departamento de
Estadística, Investigación Operativa, Universidad de Granada

Abstract

En este trabajo se plantean diferentes métodos semiparamétricos para la estimación del vector de parámetros que caracteriza la estructura de dependencia fuerte en modelos espacio-temporales anisotrópicos.

Claves: Anisotropía, estructura de dependencia fuerte, modelos espacio-temporales.

1. Introducción

Es conocida la presencia de dependencia fuerte en datos medioambientales (ver, p.e., [1]). La mayor parte de los modelos derivados en este contexto son estacionarios con isotropía espacial (ver, p.e., [2]). Sin embargo, en la práctica, se presentan con frecuencia comportamientos anisotrópicos en el espacio. Se plantea así la necesidad de formular modelos que representen simultáneamente ambas características, dependencia fuerte y anisotropía. En [3] se introducen familias de modelos espacio-temporales con dependencia fuerte y anisotropía. En este trabajo se proponen diferentes métodos de estimación del vector de parámetros de dependencia fuerte en dichos modelos, basados en el comportamiento local de la densidad espectral integrada en entornos de cero.

2. Generación del Modelo

Se consideran modelos espacio-temporales anisotrópicos con dependencia de largo rango generados mediante la convolución, en media cuadrática, de un proceso de entrada Y con el núcleo espacio-temporal

$$r(t, \mathbf{z}) = |t|^{-1+\nu} \prod_{i=1}^d |z_i|^{-1+\beta_i}, \quad (t, \mathbf{z}) \in \mathbb{R}^{d+1}. \quad (2.33)$$

Se supone que el proceso de entrada Y verifica las condiciones necesarias para que el proceso de salida sea de segundo orden (ver[3]). La covarianza de dicho proceso será no separable si lo es la covarianza del proceso de entrada Y .

3. Métodos de Estimación

El proceso de salida X construido a partir del núcleo (2.33) y el proceso de entrada Y posee densidad espectral dada por

$$f_X(\omega, \boldsymbol{\lambda}) = f_Y(\omega, \boldsymbol{\lambda}) |\omega|^{-2\nu} \prod_{i=1}^d |\lambda_i|^{-2\beta_i},$$

donde f_Y representa la densidad espectral de Y . Suponiendo dependencia de corto rango respecto a cada coordenada en el proceso Y , se plantean los siguientes métodos de estimación:

Método I: Basado en la integración respecto a cada coordenada de la densidad espectral en una secuencia de entornos direccionales de cero.

Método II: Basado en la integración marginal de la densidad espectral sobre entornos temporales y entornos espaciales. La discriminación de los parámetros espaciales se efectúa en términos de ecuaciones adicionales, derivadas a partir de la proyección de la densidad espectral sobre bases de *wavelets*, construidas mediante el producto tensorial.

Método III: Basado en el cálculo de las funciones de distribución marginales espectrales en puntos próximos a la frecuencia cero.

Método IV: Basado en el cálculo de los valores de la distribución espectral conjunta en un entorno de cero.

La selección entre los métodos de estimación propuestos se realizará a partir de las características de segundo orden de Y y de la información disponible sobre las mismas.

4. Agradecimientos

Este trabajo ha sido parcialmente financiado por los proyectos MTM2005-08597 de la DGI, MEC, y P05-FQM-00990 de la CICE, Junta de Andalucía.

5. Bibliografía

- [1] Christakos G. (2000). *Modern Spatiotemporal Geostatistics*. Oxford University Press.
- [2] Kelbert M., Leonenko N. y Ruiz-Medina M.D. (2005). Fractional Random Fields Associated with Stochastic Fractional Heat Equation. *Advances in Applied Probability* 37, 108-133.
- [3] Frías M.P., Ruiz-Medina M.D., Angulo J.M. y Alonso F.J. (2005). Spatio-temporal Generation of Long-Range Dependence Models and Estimation. *Environmetrics*. Aceptado (doi:10.1002/env.756)

Estimadores robustos en la fase de ajuste de un gráfico de control

José Luis Alfaro Navarro¹, Juan Fco. Ortega Dato²

¹JoseLuis.Alfaro@uclm.es, Facultad de Ciencias Económicas y Empresariales de Albacete (Área de Estadística), Universidad de Castilla-La Mancha

²JuanFco.Ortega@uclm.es, Facultad de Ciencias Económicas y Empresariales de Albacete (Área de Matemáticas), Universidad de Castilla-La Mancha

Abstract

En el período base de elaboración de un gráfico de control, se estiman los parámetros del mismo mediante una muestra disponible de las variables en juego, bajo ciertos supuestos de los vectores de observaciones. La presencia de observaciones anómalas llevaría a estimaciones que pueden desviarse del valor original, incidiendo por tanto en el comportamiento del gráfico de control. En este trabajo se analiza el comportamiento de los gráficos de control ante la presencia de observaciones atípicas, proponiendo un método que presente un buen comportamiento ante dichas observaciones.

Palabras Clave: Control estadístico de la calidad, estimadores robustos, medias truncadas

1. Introducción

Dentro de las técnicas estadísticas de control de la calidad, los gráficos de control son los más utilizados en los procesos industriales para controlar la calidad de los productos. Estos gráficos son una representación de una o varias característica de calidad que se suelen establecer para algunas medidas de posición y de dispersión, de manera que sea posible controlar las variaciones del proceso en términos medios o en términos de variabilidad.

En la elaboración de un gráfico de control se pueden diferenciar dos etapas: *período base*, donde se estiman los parámetros del gráfico de control; y *período de vigilancia*, donde se utilizan dichas estimaciones para analizar el comportamiento del proceso.

Para estimar los parámetros en el período base se utilizan las muestras disponibles de las variables en juego, de manera que un problema que puede aparecer en dicha estimación es la presencia de ciertas observaciones que no sean genuina del proceso en estudio, llamadas observaciones atípicas. En este caso, si los métodos o estimadores utilizados en la determinación de los parámetros no son resistentes a estas observaciones indeseables, las contaminaciones nos llevarían a obtener unas estimaciones que puede desviarse del valor original, incidiendo por tanto en el comportamiento del gráfico de control.

En el presente trabajo se analiza el comportamiento de los gráficos de control multivariantes para observaciones individuales, proponiendo un procedimiento para estimar los parámetros de tal forma que presenten un buen comportamiento ante la presencia de observaciones atípicas.

2. Bibliografía

- [1] Alfaro, J.L. (2002). Métodos Multivariantes en Control Estadístico de la Calidad. *Documentos de Trabajo de la Facultad de CC. Económicas y Empresariales de Albacete* (Doc: 2/2002/2).
- [2] Ortega, J.Fco. (2000). *Nuevas Familias de Estimadores Robustos y Deteción de Observaciones Atípicas en Modelos Lineales*. Tesis Doctoral. Univ. de Castilla-La Mancha.
- [3] Ortega, J.Fco. (2003). Elección de un Nivel de Truncamiento óptimo para la familia de Medias Truncadas. Actas del la 27^a reunión de la *Sociedad de Estadística e Investigación Operativa* (SEIO).
- [4] Rocke, D.M. (1989). Robust control charts. *Technometrics*. Vol.31, No.2.
- [5] Vargas, J.A. (2003). Robust Estimation in Multivariate Control Charts for Individual Observation. *Journal of Quality Technology* 35, pp: 367-376.

Estimadores *shrinkage* basados en la descomposición atómica

Maria D. Ruiz-Medina¹, José M. Angulo¹

¹mruiz@ugr.es; jmangulo@ugr.es, Departamento de Estadística e Investigación Operativa, Universidad de Granada

Abstract

En la descomposición atómica de funciones se combinan elementos de las transformadas *wavelets* discreta y continua, encontrándose, pues, en dicha descomposición las ventajas de ambas transformaciones. En este trabajo, se formula una clase de estimadores *shrinkage*, que contiene a las clases previamente consideradas en relación con los *maxisets* asociados.

Claves: Densidad de probabilidad, estimación, modelo de ruido blanco.

1. Introducción

La transformada *wavelet* ha sido ampliamente utilizada en el diseño de estimadores *shrinkage* (ver, por ejemplo, [1], [2]). La calidad de dichos estimadores usualmente se mide en términos de su razón de convergencia en un espacio funcional dado, o bien en términos del *maxiset* asociado, espacio funcional (normalmente un espacio de Besov) donde el estimador *wavelet* alcanza una determinada razón de convergencia. En este trabajo se formula una nueva familia de estimadores *shrinkage* basados en la descomposición atómica de funciones en espacios de Besov fraccionarios (ver [5]) y multirraccionarios (ver [3], [4]). En el caso multirracionario, la familia de *maxisets* asociada se halla definida en términos de espacios de Besov de orden variable, conteniendo, en particular, a la familia de espacios de Besov de orden fijo, anteriormente considerada por otros autores en la construcción de dichos conjuntos.

La familia de estimadores *shrinkage* viene dada por

$$\mathcal{F} = \left\{ \hat{f}(\cdot) = \sum_j \sum_{\mathbf{k}} \gamma_{j\mathbf{k}} \hat{f}_{j\mathbf{k}} \psi_{j\mathbf{k}}(\cdot), \quad \gamma_{j\mathbf{k}} \in [0, 1] \right\}, \quad (2.34)$$

donde $f(\cdot) = \sum_j \sum_{\mathbf{k}} f_{j\mathbf{k}} \psi_{j\mathbf{k}}(\cdot)$, siendo $\{\psi_{j\mathbf{k}}, \mathbf{k} \in \mathbb{Z}^n, j \in \mathbb{Z}\}$ una base de funciones *wavelets*. Los pesos $\{\gamma_{j\mathbf{k}}, \mathbf{k} \in \mathbb{Z}^n, j \in \mathbb{Z}\}$ definen la regla de truncamiento. En general, el concepto de *maxiset* involucra: Una secuencia de modelos $M_n = \{P_\theta^n, \theta \in \Theta\}$, siendo $\{P_\theta^n, n \in \mathbb{N}\}$ distribuciones de probabilidad sobre $\{\Omega_n, n \in \mathbb{N}\}$ y Θ el espacio de parámetros; una secuencia de estimadores

\hat{f}_n de $f(\theta)$ junto con una función de pérdida $\rho(\hat{f}_n, f(\theta))$ y la razón de convergencia α_n tendiendo a cero. Se define entonces el *maxiset* asociado como $MS(\hat{f}_n, \rho, \alpha_n)(T) = \left\{ \theta \in \Theta, \sup_n E_\theta^n \rho(\hat{f}_n, f(\theta))(\alpha_n)^{-1} \leq T \right\}$, para T fijado.

2. Nueva familia de estimadores *shrinkage*

La descomposición átomica de $f \in H^{\sigma(\cdot)}(\mathbb{R}^n)$ viene dada por

$$\begin{aligned} f(\mathbf{x}) &= \sum_{\nu=0}^{\infty} \sum_{\mathbf{m} \in \mathbb{Z}^n} \lambda_{\nu\mathbf{m}} a_{\nu\mathbf{m}}(\mathbf{x}), \text{ con } a_{\nu\mathbf{m}}(\mathbf{x}) = \lambda_{\nu\mathbf{m}}^{-1} \psi_{\nu\mathbf{m}}(\mathbf{x}) k_N(2^{-\nu+1}, f)(\mathbf{x}), \\ \sum_{\mathbf{m} \in \mathbb{Z}^n} \psi_{\nu\mathbf{m}}(\mathbf{x}) &= 1, \quad \lambda_{\nu\mathbf{m}} = 2^{\nu \left(\inf_{\mathbf{x} \in Q_{\nu\mathbf{m}}} \sigma(\mathbf{x}) - n/2 \right)} \sup_{\mathbf{x} \in Q_{\nu\mathbf{m}}} |k_N(2^{-\nu+1}, f)(\mathbf{x})| \\ k_N(2^{-\nu+1}, f)(\mathbf{x}) &= \frac{1}{2^{n\nu-n}} \int_{\mathbb{R}^n} k_N \left(\frac{\mathbf{y} - \mathbf{x}}{2^{-\nu+1}} \right) f(\mathbf{y}) d\mathbf{y}, \quad \nu \in \mathbb{N}, \quad \mathbf{m} \in \mathbb{Z}^n \end{aligned} \quad (2.35)$$

siendo $H^{\sigma(\cdot)}(\mathbb{R}^n)$ el espacio de Sobolev de orden variable $\sigma(\cdot)$ sobre \mathbb{R}^n . La formulación general para espacios de Besov multifraccionarios $B_{p,q}^{\sigma(\cdot)}(\mathbb{R}^n)$ se obtiene considerando las normas L^p y L^q en lugar de la norma L^2 . La descomposición (2.35) sustituida en (2.34) proporciona una nueva familia de estimadores *shrinkage*. En este contexto, la reformulación de reglas de truncamiento conduce a la construcción de *maxisets* a partir de los espacios $B_{p,q}^{\sigma(\cdot)}(\mathbb{R}^n)$.

3. Agradecimientos

Este trabajo ha sido financiado en parte por los Proyectos MTM2005-08597 de la DGI, MEC, y P05-FQM-00990, Junta de Andalucía.

4. Bibliografía

- [1] Abramovich, F., Amato, U. y Angelini, C. (2004). On Optimality of Bayesian Wavelet Estimators. *Scandinavian Journal of Statistics* 31, 217-234.
- [2] Kerkyacharian, G. y Picard, D. (1993). Density Estimation by Kernel and Wavelet Methods: Optimality of Besov Spaces. *Statistics and Probability Letters* 18, 327-336.
- [3] Ruiz-Medina, M.D., Anh, V.V. y Angulo, J.M. (2004). Fractional Generalized Random Fields of Variable Order. *Stochastic Analysis and Applications* 22, 775-800.
- [4] Ruiz-Medina, M.D., Anh, V.V. y Angulo, J.M. (2005). Random Fields with Multifractional Regularity Order on Multifractals. (Sometido).
- [5] Triebel, H. (1997). *Fractals and Spectra*. Birkhäuser.

Estudio de la varianza poblacional en el muestreo en ocasiones sucesivas

Inmaculada Oña Casado¹, Amelia V. García Luengo¹, Eva M. Artés Rodríguez¹

¹ Departamento de Estadística y Matemática Aplicada.

Universidad de Almería

iocasado@ual.es, amgarcia@ual.es, eartes@ual.es

Abstract

En este trabajo, bajo un diseño de muestreo en dos ocasiones, estimamos la varianza poblacional en la segunda ocasión, combinando un estimador de regresión de doble muestreo de la varianza para la parte apareada de la muestra y un estimador simple de la varianza para la parte no apareada. Además, se estudia el estimador del cambio y la suma de varianzas.

Palabras Clave: muestreo sucesivo, varianza poblacional, ganancia en precisión.

AMS: Clasificación AMS (62D05)

1. Introducción

El uso de información auxiliar proporcionada por una variable x altamente correlada con la variable principal y es común, en la estimación de medias o totales poblacionales. Los estimadores de razón y regresión utilizan la información que proporciona la variable x para modificar los estimadores directos, consiguiendo estimadores más precisos del parámetro en cuestión. De una forma similar, es razonable suponer que estas técnicas de estimación se pueden utilizar, bajo las condiciones adecuadas, para proporcionar estimadores eficientes de la varianza.

2. Técnica

Presentamos la construcción de un estimador de la varianza poblacional en la segunda ocasión, combinando dos estimadores independientes de la varianza: un estimador de regresión de doble muestreo para la parte apareada de la muestra, $s_{ym}^{\prime 2}$ y un estimador simple para la parte no apareada, s_{yu}^2 , como se muestra en la tabla siguiente. La combinación de los dos estimadores independientes de la varianza, $s_{ym}^{\prime 2}$ y s_{yu}^2 , se lleva a cabo mediante una media aritmética ponderada, con pesos Q y $1 - Q$, inversamente proporcionales a las varianzas de los estimadores independientes y suma igual a la unidad.

$$\hat{S}_y^2 = Q s_{yu}^2 + (1 - Q) s_{ym}^{\prime 2}$$

$$\beta_2(y) = \frac{\mu_{40}}{\mu_{20}^2}; \quad \beta_2(x) = \frac{\mu_{04}}{\mu_{02}^2}; \quad \theta = \frac{\mu_{22}}{\mu_{20}\mu_{02}}; \quad \mu_{rs} = \frac{1}{N} \sum (y_j - \bar{Y})^r (x_j - \bar{X})^s$$

estimador de S_y^2	varianza
$s'_{ym}^2 = s_{ym}^2 + b(s_x^2 - s_{xm}^2)$	$S_y^4 \left[\frac{1}{m}(\beta_2(y) - 1) - \left(\frac{1}{m} - \frac{1}{n} \right) \left(\frac{(\theta - 1)^2(\beta_2(y) - 1)}{K} \right) \right]$
s_{yu}^2	$S_y^4 \frac{1}{u} (\beta_2(y) - 1)$

Teniendo en cuenta que

$$V(s_{yu}^2) = S_y^4 \frac{1}{u} (\beta_2(y) - 1); \quad V(s'_{ym}^2) = \frac{S_y^4}{m} (\beta_2(y) - 1) \left[1 - q \left(\frac{(\theta - 1)^2}{K} \right) \right]$$

Operando adecuadamente y siendo $K = (\beta_2(x) - 1)(\beta_2(y) - 1)$, se tiene que

$$V_{min} \left(\hat{S}_y^2 \right) = \frac{S_y^4}{n} (\beta_2(y) - 1) \frac{K - q(\theta - 1)^2}{K - q^2(\theta - 1)^2} = \frac{S_y^4}{n} (\beta_2(y) - 1) \frac{1 - q \frac{(\theta - 1)^2}{K}}{1 - q^2 \frac{(\theta - 1)^2}{K}}$$

3. Estimador del cambio

Un estimador del cambio, junto la expresión de su varianza es el siguiente

$$\Delta_2 = as_{xu}^2 + bs_{xm}^2 + cs_{ym}^2 + ds_{yu}^2; \quad V_{min}(\Delta_2) = \frac{2}{n} S_y^4 (\beta_2(y) - 1) \frac{(\beta_2(y) - 1) - (\theta - 1)}{(\beta_2(y) - 1) - q(\theta - 1)}$$

La varianza será mínima si las muestras en ambas ocasiones contienen las mismas unidades muestrales.

4. Estimador de la suma de varianzas

Un estimador de la suma, junto la expresión de su varianza es el siguiente

$$z = as_{xu}^2 + bs_{xm}^2 + cs_{ym}^2 + ds_{yu}^2; \quad V_{min}(z) = \frac{2}{n} S_y^4 (\beta_2(y) - 1) \frac{(\beta_2(y) - 1) + (\theta - 1)}{(\beta_2(y) - 1) + q(\theta - 1)}$$

La varianza será mínima si las muestras son independientes.

5. Bibliografía

- [1] García Luengo, A. V. (2003). *Mejora de estimadores en muestreo en ocasiones sucesivas*. Servicio de Publicaciones de la Universidad de Almería.
- [2] García Luengo, A. V., Artés Rodríguez, E. M., Oña Casado I. (2004). *Estimación de la Varianza poblacional en el Muestreo en Ocasiones Sucesivas*. Metodología de Encuestas, Volumen 6, No. 1, 9-22.

Estudio de supervivencia de pacientes con SIDA

E. Corpas Nogales¹, J. García Leal², A. Lara Porras², J. M. Quesada Rubio²

¹Escuela Andaluza de Salud Pública, Granada

²Departamento de Estadística e I.O., Universidad de Granada

e-mail: ecorpas@yahoo.es; juliagl@ugr.es; alara@ugr.es; quesada@ugr.es

Abstract

Estudiamos el tiempo de supervivencia en pacientes infectados de SIDA en Andalucía en el período 1982-2001. Este estudio lo hemos llevado a cabo basándonos en el modelo de regresión de Cox para datos censurados a la derecha, para lo cual hemos utilizado el paquete estadístico S-Plus.

Palabras Clave: Modelo de Cox, Supervivencia, SIDA, VIH.

AMS: 62N01; 62N02; 62N03; 62P10

1. Modelos de supervivencia

Para realizar un análisis de supervivencia se pueden utilizar diferentes métodos que han ido evolucionando a lo largo del tiempo, consiguiendo unas predicciones cada vez más precisas. El modelo de Cox o de riesgos proporcionales, Cox (1972) y Cox & Oakes (1984), es el más empleado en la actualidad. Este modelo nos permite comparar el tiempo de supervivencia entre diferentes grupos y analizar cuales son las covariables que influyen en dicho tiempo.

La función de riesgo para el modelo de Cox viene dada por la expresión $\lambda(t) = \lambda_0(t) \exp(\beta^t X)$, donde las covariables afectan a la función de riesgo base de forma multiplicativa. Para poder utilizar el modelo de Cox es necesario que los datos cumplan la propiedad de riesgos proporcionales.

2. Supervivencia en pacientes con SIDA en Andalucía

Se ha aplicado el modelo de Cox a datos de pacientes infectados de SIDA. Se han estudiado 8766 casos diagnosticados en Andalucía entre 1982 y 2001, estos datos han sido proporcionados por el Registro de SIDA de Andalucía. Las variables disponibles en el registro son: fecha de diagnóstico, fecha de muerte, estado vital, sexo, edad, periodo de diagnóstico, provincia de residencia y categoría de transmisión. El tiempo de supervivencia lo hemos definido como el tiempo transcurrido entre la fecha de diagnóstico de la enfermedad y la fecha de muerte, si ésta ha ocurrido antes del día 31 de mayo de 2005. Los individuos que han sobrevivido después de esa fecha se consideran censurados.

Para decidir cuales son las variables que influyen en el tiempo de supervivencia se ha utilizado el test de la razón de verosimilitud. Además de dicho test se

pueden utilizar otros, como los tests de Wald o score. Los tres test son asintóticamente equivalentes, aunque el test de la razón de verosimilitudes es el más fiable. Como resultado hemos obtenido que las covariables significativas en este modelo son: el sexo, la edad, el periodo de diagnóstico y la provincia de residencia.

En general es necesario comprobar que se verifica la condición de riesgos proporcionales en los datos y para ello hay varios procedimientos. Cuando el tamaño de la muestra es grande, como es nuestro caso, no es necesario probar dicha condición, ver Therneau & Grambsch (2000).

Los resultados del modelo seleccionado son los siguientes:

Variable	Coeficiente	e.e.	RR	I.C. RR	p-valor
Sexo					
Mujer	-	-	1	-	-
Hombre	0.17	0.040	1.19	(1.10-1.28)	<0.001
Edad					
Menores de 35 años	-	-	1	-	-
30-35 años	0.13	0.033	1.14	(1.07-1.22)	<0.001
Mayores de 35 años	0.34	0.036	1.41	(1.31-1.51)	<0.001
Periodo de diagnóstico					
1996-2001	-	-	1	-	-
1993-1995	0.66	0.034	1.94	(1.82-2.08)	<0.001
1982-1992	0.65	0.037	1.91	(1.76-2.05)	<0.001
Provincia					
Granada	-	-	1	-	-
Almería	0.22	0.071	1.24	(1.08-1.43)	0.002
Cádiz	0.02	0.057	1.02	(0.91-1.14)	0.770
Córdoba	0.16	0.0703	1.17	(1.02-1.34)	0.025
Huelva	0.14	0.072	1.15	(1.00-1.33)	0.046
Jaén	0.12	0.082	1.16	(0.96-1.32)	0.160
Málaga	0.24	0.054	1.28	(1.15-1.42)	<0.001
Sevilla	-0.07	0.058	0.93	(0.83-1.05)	0.240

Se pone de manifiesto que la introducción de la terapia antirretroviral de gran actividad (TARGA), en el año 1996, mejora la supervivencia de los pacientes de SIDA, ya que la supervivencia de los diagnosticados a partir de 1996 mejora un 91 % y un 94 % con respecto a los diagnosticados en 1982-1992 y 1993-1995 respectivamente.

3. Bibliografía

- [1] Cox DR. (1972). *Regression models and life-tables*. J.Roy.Statist.Soc. B 34, 187-220.
- [2] Cox, D.R. & Oakes, D. (1984). *Analysis of Survival Data*. Chapman and Hall, London.
- [3] Therneau, T; Grambsch, P. (2000). *Modeling Survival Data: Extending the Cox Model*. Springer-Verlag.

Estudio del Comportamiento Extremal de Sucesiones Periódicas

Manuel G. Scotto¹, Andreia Hall²

¹mscotto@mat.ua.pt, Departamento de Matemática, Universidade de Aveiro

²andreia@mat.ua.pt, Departamento de Matemática, Universidade de Aveiro

Abstract

El objetivo de este trabajo se centra en el estudio del comportamiento extremal de ciertas transformaciones asociadas a sucesiones periódicas. En particular se analizará la distribución asintótica del máximo y su correspondiente índice extremal. Este tipo de estudio es particularmente importante cuando se analizan procesos medio ambientales y financieros.

Claves: extremos, índice extremal, procesos puntuales.

AMS: 60G70

1. Introducción

Una de las formas más comunes de abordar el problema del estudio del comportamiento extremal de procesos no lineales, es a través del análisis de ecuaciones estocásticas del tipo

$$X_k = \sum_{j=1}^{\infty} \left(\prod_{s=1}^{j-1} A_{k-s} \right) B_{k-j}, \quad k \in \mathbb{N},$$

siendo $(A_k, B_k)_{k \in \mathbb{N}}$ pares de variables aleatorias independientes e idénticamente distribuidos (i.i.d) e independientes de X_0 . Ejemplos de aplicación se pueden encontrar en los trabajos de Dufresne [1], Stărică [4] y Mikosch [1]. Una cuestión importante desde el punto de vista práctico, que aún no ha sido estudiada, es analizar propiedades extremales de $\mathbf{X} = (X_k)_{k \in \mathbb{N}}$ asumiendo que los pares $(A_k, B_k)_{k \in \mathbb{N}}$ tienen una estructura periódica. Se dice que una sucesión $\mathbf{Y} = (Y_k)_{k \in \mathbb{N}}$ es periódica, si existe un valor entero $T \geq 1$ de manera que, para cualquier selección de enteros (i_1, \dots, i_n) los vectores $(Y_{i_1}, \dots, Y_{i_n})$ y $(Y_{i_1+T}, \dots, Y_{i_n+T})$ son idénticamente distribuidos. En este caso, diremos que la sucesión \mathbf{Y} es T -periódica. Si $T = 1$ la sucesión \mathbf{Y} es estacionaria. El análisis de las características extremales de sucesiones periódicas es especialmente importante en el estudio de procesos medio ambientales, tales como concentraciones diarias de ozono, caudales de ríos y valores de temperaturas diarias.

Formalmente, el problema que se pretende analizar es el siguiente: si $(A_k, B_k)_{k \in \mathbb{N}}$ forman una sucesión T -periódica de pares independientes en \mathbb{R}_+^2 , tal que $\bar{F}_r(x) =$

$P(B_r > x)$, $r = 1, \dots, T$, son de variación de regular del tipo

$$\bar{F}_r(x) = x^{-\alpha} L_r(x), \quad x > 0, \alpha > 0,$$

para $r = 1, \dots, T$,

$$\lim_{x \rightarrow \infty} \frac{\bar{F}_i(x)}{\bar{F}_r(x)} = \gamma_{i,r}, \quad r = 1, \dots, T, \quad i = 0, \dots, T-1$$

y (A_1, \dots, A_T) tienen distribuciones con colas menos pesada que las colas de las distribuciones de (B_1, \dots, B_T) tal que $EA_r^\alpha < 1$ ($r = 1, \dots, T$), se pretende analizar las propiedades extremales de \mathbf{X} . En particular se analizará la distribución asintótica del máximo y su correspondiente índice extremal, que se puede interpretar como el recíproco del tamaño medio de las agrupaciones de valores elevados. Estos resultados generalizan los obtenidos por Scotto [3] para el caso estacionario.

2. Bibliografía

- [1] Dufresne, D. (1991). The distribution of a perpetuity, with applications to risk theory and pension funding. *Scand. Actuar. J.* 39-79.
- [2] Mikosch, T. (2003). Modeling dependence and tails of financial time series. In: Finkenstaedt, B. and Rootzén, H. (eds.) *Extreme Values in Finance, Telecommunications, and the Environment*. Chapman and Hall, 185-286.
- [3] Scotto, M. G. (2005). Extremes of a class of deterministic sub-sampled processes with applications to stochastic difference equations. *Stoch. Process. Appl.* 115, 417-434.
- [4] Stărică, C. (1999). Multivariate extremes for models with constant conditional correlations. *J. Empirical Finance* 6, 515-553.

Evaluación de selectores del parámetro de suavizado mediante *SiZer Map*

M.D. Martínez Miranda¹, R. Raya Miranda¹, W. González Manteiga², A. González Carmona¹

¹mmiranda@ugr.es, rraya@ugr.es, andresgc@ugr.es; Departamento de Estadística e I. O., Universidad de Granada

²wenceslao@usc.es, Departamento de Estadística e I. O., Universidad de Santiago de Compostela

Abstract

Se modifica la herramienta gráfica denominada *SiZer Map* para evaluar las componentes de un modelo aditivo estimado usando distintos métodos que incluyen Backfitting e integración marginal. El objetivo es evaluar el comportamiento de varios selectores (locales y globales) del ancho de banda definidos en este contexto.

Claves: SiZer Map, Modelo aditivo, Parámetro de suavizado, Bootstrap.

AMS: 62G05

1. Introducción

Los modelos aditivos en regresión no paramétrica, se formulan como la suma de un conjunto de componentes, que se suponen funciones desconocidas univariantes suaves. Para estimar estos modelos se han propuesto diferentes métodos, entre los que destacan el algoritmo *backfitting* y el método de integración marginal, como los más importantes. En el artículo de Sperlich et al. (1999) se ofrece un estudio comparativo entre ambos métodos mostrando que, en general, el comportamiento del método depende del modelo e incluso del intervalo de estimación considerado. Posteriormente, se han definido otros métodos eficientes de estimación bajo aditividad que plantean la combinación de ambos métodos.

SiZer Map es una herramienta de tipo gráfico introducida por Chaudhuri y Marron (1999) en el ámbito unidimensional, que permite explorar las características del conjunto de datos que soporta el problema de estimación de curvas. El procedimiento se basa en la construcción de intervalos de confianza para la derivada, $m'(x; h)$, en el espacio definido por el parámetro de suavizado, h . La extensión de esta herramienta gráfica al caso multidimensional presenta serias dificultades que parten desde la propia representación gráfica de *SiZer Map*. Sin embargo, el problema puede atacarse de una forma más sencilla suponiendo aditividad en el modelo multidimensional ya que se pueden considerar por separado los efectos de las variables explicativas. La extensión para un modelo de

regresión aditivo consiste en realizar tantos *SiZer Map* como número de variables explicativas intervengan. Siguiendo esta línea Raya-Miranda et al. (2002) desarrollan las expresiones necesarias para la construcción del *SiZer Map* para un modelo aditivo estimado utilizando el algoritmo *backfitting*. Tomando de base este trabajo, se presentan las expresiones para el caso de estimar el modelo aditivo con el mét. eficiente de Kim et.al. (1999).

Sea $\{\hat{m}(\mathbf{x}; h_1, \dots, h_D) : h_d \in [h_{d,\min}, h_{d,\max}], d = 1, \dots, D\}$, una familia de suavizadores de tipo aditivo sobre la que se definen intervalos de confianza para las derivadas de las componentes y se expresan de la siguiente forma:

$$\hat{m}'_d(x_d; h_d) \pm q \sqrt{\text{Var}(\hat{m}'_d(x_d; h_d))},$$

donde el cuantil, q , de nuevo se calcula desde aproximaciones normales, o bien utilizando técnicas *bootstrap*. Se han derivado las expresiones para las derivadas de las componentes y las varianzas de estas derivadas.

Se ha realizado un estudio de simulación, en el que se han considerado tres modelos aditivos bivariantes. Se han creado los *SiZer Map* correspondientes a cada modelo y en ellos se han representado varios selectores (locales y globales) del ancho de banda. Entre ellos destacamos el selector de tipo local basado en la metodología *bootstrap* definido por Martínez-Miranda et al.(2005), que ha demostrado un buen comportamiento en todos los modelos considerados.

2. Bibliografía

- [1] Chaudhuri, P. y Marron, J.S. (1999), SiZer for exploration of structures in curves, *Journal of the American Statistical Association*, 94, 807–823.
- [2] Kim, W., Linton, O.B. y Hengartner, N.W. (1999), A computationally efficient oracle estimator for additive nonparametric regression with bootstrap confidence intervals, *Journal of Computational and Graphical Statistics*, 8, 278–297.
- [3] Martínez-Miranda, M.D., Raya-Miranda, R., González-Manteiga, W. y González-Carmona, A. (2005) SiZer Map for evaluating a bootstrap local bandwidth selector in nonparametric additive models. Technical Report 05-01. Universidad de Santiago de Compostela
- [4] Raya-Miranda, R., Martínez-Miranda, M.D. y González-Carmona, A. (2002) Exploring the structure of regression surfaces by using SiZer Map for additive models, in *Proceedings in Computational Statistics, COMPSTAT 2002*, eds. W. Härdle and B. Rönz, Humboldt University of Berlin, Berlin.
- [5] Sperlich, S., Linton, O.B. and Härdle, W. (1999) Integration and Backfitting Methods in Additive Models - Finite Sample Properties and Comparison. *Test*, 8, 419–458.

Exactitud de las Soluciones de MDS mediante técnicas de Remuestreo

Miguel Ángel Montero Alonso¹, José Fernando Vera Vera, Juan Antonio Muñoz Orellana, Rodrigo Macías Páez, Andrés González Carmona

Departamento de Estadística e Investigación Operativa,
Universidad de Granada ¹mmontero@ugr.es

Abstract

En el presente trabajo se pretende estudiar la precisión de la solución que se obtienen para multidimensional scaling, tanto en el modelo clásico como en diferencias individuales, utilizando técnicas de remuestreo no paramétrico.

Claves: Bootstrap, MDS, Regiones de Confianza.

AMS: Clasificación AMS 91C15

1. Introducción

Los procedimientos de remuestreo en general y las técnicas Bootstrap en particular son herramientas bien conocidas para el estudio de la exactitud de las soluciones del modelo en escalamiento Multidimensional. Varios procedimientos Bootstrap no-paramétricos han sido empleados para obtener regiones de confianza para las soluciones de MDS, algunos de los cuales, no resultan convenientes para tamaños pequeños de la muestra. En este trabajo discutimos la utilización de recientes procedimientos del remuestreo para estudiar la variabilidad de las soluciones y sus implicaciones en la interpretación de asociaciones con y sin hipótesis de normalidad. Para probar la eficacia de los procedimientos comparados, éstos han sido implementados MatLab y comparadas las regiones obtenidas. Para dicho estudio se han utilizado datos basados en una matriz triangular y datos en varias matrices de disimilitud, ofreciendo alternativas a los distintos estudios realizados.

2. Bibliografía

- [1] Bickel, P.J. and Freedman, D.A. (1981). *Some asymptotic theory for the bootstrap*. Ann. Statist., (9):1196–1217.
- [2] Booth, J.C. and Sarkar, S. (1998). *Monte carlo approximation of bootstrap variances*. Am. Statistic, (52):354–357.
- [3] Chernick, M.R. (1999) *Bootstrap Methods. A Practitioner's Guide*. John Wiley & Sons, Inc..

- [4] Cox, T.F. and Cox, M.A.A. (2001). *Multidimensional Scaling*. Chapman & Hall, second edition.
- [5] Davison, A.C. and Hinkley, D.V. (1997). *Bootstrap Methods and Their Applications*. Cambridge University Press.
- [6] Dudewicz, E.J. (1992) *The generalized bootstrap. In bootstrapping and related techniques. proceedings, trier, frg.* Lecture Notes in Economics and Mathematics system. Springer-Verlag. Berlin, (376):31–37.
- [7] Efron, B. and Tibshirani, R.J. (1998) *An Introducción To The Bootstrap*. Chapman & Hall, 2 edition.
- [8] Mammen, E. (1992). *When does the bootstrap work? asymptotic results and simulations*. Lecture Notes in Statistics. Springer-Verlag, Heidelberg, (77), 1992.
- [9] Mardia, K.V.; Kent, J.T. and Bibby, J.M. (1979). *Multivariate Analysis*. Academic Press.
- [10] Weinberg, S.L.; Carroll, J.D. and Cohen, H.S. (1984). *Confidence regions for indscal using the jackknife and bootstrap techniques*. Psychometrika, (49):475–491.

Explotación por horario: un modelo de gestión para Transports de Barcelona

Miquel Salicrú¹, Josep Maria Armengol²

¹msalicru@ub.edu, Dep. de Estadística, Universidad de Barcelona

²jmarmengol@tmb.net, Centre Regulació d'Operacions Transports Barcelona

Abstract

Desarrollamos un modelo de gestión para las líneas explotadas por puntualidad en base a fijar criterios para el diseño y construcción de horarios, identificar oportunidades y favorecer la mejora.

1. Introducción

En los últimos años se han producido una serie de cambios tecnológicos, sociales y políticos que han modificado los estilos de vida, han revolucionado las áreas económicas y en consecuencia, han cambiado las necesidades de movilidad de los ciudadanos –en la Ciudad de Barcelona y por extensión en la Región Metropolitana de Barcelona–. De ahí que, Transports Metropolitans de Barcelona, empresa que gestiona la red de autobuses –Transports de Barcelona SA– y la red de metro –Ferrocarril Metropolità de Barcelona– adapta su servicio a las necesidades cambiantes de los usuarios.

A nivel tecnológico, Transports de Barcelona –empresa que transporta casi 200 millones de pasajeros/año en sus 104 líneas que cubren una longitud aproximada de 800 Km– avanza en el desarrollo de un Sistema de Ayuda a la Explotación (SAE) soportado en sistemas de localización de la flota en tiempo real en base a tecnologías GPS, odómetros y giróscopos embarcados en los vehículos y cambia su modelo de gestión –íntegra por regularidad– a una explotación mixta con líneas explotadas regularidad y líneas explotadas por puntualidad –horario–.

En este marco se desarrolla la iniciativa estratégica: “Mejorar el modelo de gestión de las líneas explotadas por horario”, y nuestro esfuerzo se orienta a competuizar, sistematizar y automatizar buena parte de la actividad. Compartir nuestra experiencia es el objetivo de la comunicación.

2. Diseño y construcción de horarios

Optimizar la puntualidad en el cumplimiento de horarios requiere definir de forma previa la función objetivo, diseñar horarios –de acuerdo con las capacidades y necesidades– y cumplir la programación establecida.

Al objeto de maximizar la función objetivo: porcentaje de diferencias entre los tiempos reales de paso y los tiempos establecidos en el horario –Dif (TRP, TEH)– que se encuentran en el intervalo de tiempos $[-a, +b]$, hemos establecido

las siguientes líneas de actuación:

- *Filtrado y depuración de trayectos atípicos.* La necesidad de disponer de información fiable exige eliminar los tiempos de paso correspondientes a situaciones atípicas. En esta dirección, hemos establecido los criterios de normalidad en base a intervalos de confianza obtenidos por medianas.
- *Fijar las bandas horarias.* La relación entre el nivel de calidad establecido y la varianza máxima admisible en los tiempos de paso por parada en franja horaria, ha permitido identificar las horas de corte.
- *Fijar criterio para la construcción de las tablas de tiempo.* Maximizar la función objetivo –sin olvidar la deriva en el trayecto–, ha requerido fijar un criterio de tensado acorde con la varianza de las paradas en franja horaria –con mayor exigencia en paradas inicio de trayecto, ajustes de compatibilidad en paradas próximas y relajación en la parte final de trayecto–.
- *Ajuste y predicción.* Para el ajuste fino de bandas horarias y de tiempos de paso por parada, hemos establecido un algoritmo de simulación. Para cada una de las condiciones examinadas, obtenemos la predicción de resultados.
- *Construcción del horario.* En base a las tablas de tiempo, los tiempos de recuperación de incidencias, puntos de relevo, vehículos y personas disponibles, . . . , se establece el procedimiento para construir el horario.
- *Identificar oportunidades de mejora.* El análisis de los tiempos históricos de paso, nos permite identificar oportunidades y en algunos casos sus causas raíz. Así, en la parada inicio de trayecto contrastamos: fiabilidad de la información, tiempo de maniobras, actuación de las personas, . . . , en paradas intermedias evaluamos la deriva incorporada por cada tramo,
- *Validación del modelo.* El modelo obtenido es el resultado del desarrollo de planteamientos teóricos y contraste con la realidad. La validación efectiva la hemos realizado aplicando los criterios a varias líneas de la red.
- *Desarrollo de software.* La complejidad del proceso descrito, el volumen de la red de Transports de Barcelona S.A. y la necesidad de orientar el esfuerzo de los responsables de línea al trabajo de campo, a la gestión de incidencias y a la gestión de personas, obligan al desarrollo de un software específico, amable y transparente. En esta dirección, estamos trabajando.

3. Agradecimientos

Los autores del trabajo agradecen al Sr. J. Fernández Buil, al Sr. C. Saura, al Sr. R. Farré y al Sr. M. Gabal sus valiosas aportaciones y su contribución en la validación del modelo.

Extensión del espacio paramétrico de la distribución de Sichel

Xavi Puig¹, Marta Pérez-Casany², Josep Ginebra¹, Alex Riba¹

¹xavier.puig@upc.edu, Departament d'Estadística i Investigació Operativa,
Universitat Politècnica de Catalunya

²Departament de Matemàtica Aplicada II, Universitat Politècnica de
Catalunya

Abstract

En este trabajo se propone una ampliación de la distribución de Sichel obtenida extendiendo su espacio de parámetros. Dicha distribución se utiliza para el modelado de la frecuencia de la frecuencia de aparición de las palabras en el libro del Tirant lo Blanc, y se compara el ajuste obtenido con el correspondiente a otras distribuciones.

Palabras Clave: Estilometría. Mixtura de Poisson.

1. Introducción

El análisis estadístico del estilo literario busca características cuantificables de un texto, que el autor raramente controla, y las aprovecha para caracterizar la evolución del estilo. A menudo esto se hace a través de la frecuencia de la frecuencia de aparición de las palabras en el texto.

Sea V el número de palabras diferentes en un texto con un total de N palabras, sea V_r el número de palabras diferentes que aparecen r veces en el texto que en palabras de Good (1953) sería ‘la frecuencia de la frecuencia r ’, entonces, la proporción de palabras diferente de un texto que aparecen r veces es igual a $p_r = V_r/V$. La representación de p_r tiene forma de J invertida, con una cola muy larga hacia la derecha. Un modelo útil para modelar este tipo de datos debe cumplir que asigne probabilidad cero al cero y además que los valores esperados, $p(r/N)$ dependan de la longitud del texto, N .

Zipf (1932) sugiere modelar estos datos a través de la distribución de Riemann, pero se comprueba que sus ajustes no son buenos. Yule (1944), Good (1953), Herdan (1961) y Sichel (1975) sugieren modelarlos a través de mixturas de la distribución de Poisson truncada. El último autor utiliza como distribución de mixtura la distribución Inversa Gaussiana.

2. Objetivo del trabajo

Nuestro objetivo es poner de manifiesto que la distribución de Sichel puede considerarse en un dominio más amplio del que inicialmente supone su autor. Además, mediante el análisis de los datos del Tirant lo Blanc evidenciaremos que dicha ampliación es útil para el modelaje de frecuencias en capítulos cortos.

3. Distribución de Sichel ampliada

Bajo la distribución de Sichel, el valor esperado de la proporción de palabras que aparecen r veces se modela a través de:

$$p(r|N) = \frac{(2\alpha/\pi)^{1/2} \exp \alpha}{\exp(\alpha[1 - (1 - \theta)^{1/2}]) - 1} \frac{(0.5\alpha\theta)^r}{r!} k_{r-1/2}(\alpha), \text{ para } r = 1, 2, \dots, \infty,$$

con $\alpha > 0$ y $0 < \theta < 1$, y donde K es la función de Bessel de segundo tipo modificada.

A pesar de que Sichel sólo considera como válidos valores positivos de α , tiene sentido también considerar valores α entre menos uno y zero, puesto que las probabilidades resultantes son todas positivas y suman la unidad.

Al realizar el ajuste máximo verosímil observamos que en los capítulos cortos, valores de α ligeramente negativos ajustan mejor que el mejor de los valores positivos para α . Para valores de α entre -1 y 0, la expresión de $p(r|N)$ no corresponde a ninguna mezcla Poisson-Inversa Gausiana, pero esto no impide que pueda considerarse como distribución.

4. Bibliografía

- [1] Good, IJ (1953). The population frequencies of species and the estimation of population parameters. *Biometrika*, 40, 237-264.
- [2] Herdan, G (1961). A critical Examination of Simon's Model of Certain Distribution Functions in Linguistics. *Applied Statistics*, 10(2):65-76.
- [3] Sichel, HS (1975). On a distribution law for words frequencies. *J. Amer. Statist. Ass.* , **70**, 542-547.
- [4] Yule, GU (1944). The Statistical Study of Literary Vocabulary. Cambridge University Press, London.
- [5] Zipf, GK (1932). Selected Studies of the Principle of Relative Frequency in Language, Cambridge, Mass.: Harvard University Press.

Extremes of Periodic Integer-valued Moving Average Sequences

Manuel G. Scotto¹, Andreia Hall², João Pedro Cruz³

¹mscotto@mat.ua.pt, Departamento de Matemática, Universidade de Aveiro

²andreia@mat.ua.pt, Departamento de Matemática, Universidade de Aveiro

³jpedro@mat.ua.pt, Departamento de Matemática, Universidade de Aveiro

Abstract

The main purpose of this paper is to analyze the extremal properties of periodic integer-valued moving average sequences obtained as discrete analogues of conventional moving averages replacing scalar multiplication by binomial thinning. For these models, we find an unexpected phenomenon: when assessing their extremal properties, the extremal index seems not to be the object to look at.

Keywords: extremes, extremal index, periodic sequences

AMS: 60G70

1. Introduction

Recently, there has been considerable interest in integer-valued time series models and a sizeable volume of work is now available in specialized monographs and review papers . One way to obtain models for integer-valued data is to replace multiplication in conventional models by *binomial thinning*. The procedure of thinning is commoner in the study of point processes, but is also appropriate here as many discrete variate processes arise as aggregated point processes, i.e., counts of a point process in consecutive intervals of time. Models based on thinning were first considered by McKenzie [8] and has been elaborated on in subsequent papers by McKenzie [6] and Al-Osh and Alzaid ([1]) among other authors. Moreover, there have been some attempts to make the standard models based on thinning more attractive for economic applications by introducing covariates. Empirical applications are due to Blundell *et al.* [3] who studied the number of patents in firms, Rudholm [9] who analyzed competition in the generic pharmaceutical markets, and Brännäs *et al.* [4] who estimates a nonlinear integer-valued model for tourism demand.

Within the reasonably large spectrum of integer-valued models proposed in the literature, only a few have been already been studied with regard to their extremal properties. Anderson [2] gave a noticeable contribution to the study of the extremal properties of integer-valued independent and identically distributed (i.i.d.) sequences and, as an example of application, the author analyzed the behavior of the maximum queue length for $M/M/1$ queues. McCormick and Park [7] were the first to study the extremal properties of some

models obtained as discrete analogues of continuous models, replacing scalar multiplication by random thinning. More recently, Hall [5] provided results regarding the limiting distribution of the maximum of sequences within a generalized class of integer-valued moving averages driven by i.i.d. heavy-tailed innovations.

The primary objective of this paper is to provide a detailed analysis of the extremal properties of periodic integer-valued moving averages with regularly varying tails. For these models we manage to establish results which are similar to the ones obtained for the stationary case. In particular we prove that the extremal indices related with the stationary and the periodic case coincide.

2. Bibliography

- [1] Al-Osh, M. and Alzaid, A. (1987). First order integer-valued autoregressive INAR(1) process. *J. Time Series Anal.* 8, 261-275.
- [2] Anderson, C. W. (1970). Extreme value theory for a class of discrete distributions with applications to some stochastic processes. *J. Appl. Probab.* 7, 99-113.
- [3] Blundell, R., Griffith, R. and Windmeijer, F. (2002). Individual effects and dynamics in count data models. *Journal of Econometrics* 108, 113-131.
- [4] Brännäs, K., Hellström, J. and Nordström, J. (2002). A new approach to modelling and forecasting monthly guest nights in hotels. *International Journal of Forecasting* 18, 19-30.
- [5] Hall, A. (2001). Extremes of integer-valued moving averages models with regularly varying tails. *Extremes* 4, 219-239.
- [6] McKenzie, E. (1988). Some ARMA models for dependent sequences of Poisson counts. *Adv. Appl. Probab.* 20, 822-835.
- [7] McCormick, W. P. and Park, Y. S. (1992). Asymptotic analysis of extremes from autoregressive negative binomial processes. *J. Appl. Probab.* **29**, 904-920.
- [8] McKenzie, E. (1985). Some simple models for discrete variate time series. *Water Res. Bull.* 21, 645-650.
- [9] Rudholm, N. (2001). Entry and the number of firms in the Swedish pharmaceutical market. *Review of Industrial Organizations* 19, 351-364.

Factores determinantes en la supervivencia de la banca rusa al inicio de la transición

J. Orbe¹, V. Núñez-Antón¹

¹ Departamento de Economía Aplicada III, (Econometría y Estadística)
Universidad del País Vasco-Euskal Herriko Unibertsitatea

Abstract

Este trabajo toma como referencia el presentado por Carree (2003) y analiza los factores determinantes en la supervivencia de los bancos rusos, en el periodo inicial de la transición, desde un enfoque distinto y bajo una metodología alternativa y más general.

Claves: Supervivencia, Kaplan-Meier, Banca.

1. Introducción

El vertiginoso desarrollo de la banca comercial rusa tiene su punto de partida en el año 1988 con la reforma bancaria. Esto se refleja en las altas tasas de entrada, motivadas por las casi nulas barreras de entrada. Posteriormente, estas tasas de entrada disminuyen de forma importante acompañadas de altas tasas de salida, teniendo su punto más importante en la crisis interbancaria de mediados de 1995. La consecuencia de todo esto se manifiesta en que un número elevado de bancos se ven obligados a salir del mercado, y en bastantes casos, sin poder hacer frente a la devolución de los depósitos confiados por sus clientes. Por tanto, un estudio de gran interés para los potenciales clientes de la banca y para la propia banca consiste en analizar cuales son los factores que provocan que un banco cierre o salga del mercado. Este es el objetivo de Carree (2003) para el sistema bancario ruso al que estamos haciendo referencia y de Wheelock y Wilson (2000) para la banca en Estados Unidos.

2. Métodos

Este trabajo toma como referencia el presentado por Carree (2003) y realiza varias aportaciones a sus conclusiones utilizando un enfoque alternativo en el que se van a relajar varias restricciones o supuestos. Para comenzar señalar que, en lugar de analizar el efecto de los posibles factores sobre la función de riesgo o probabilidad condicional de fallo como se realiza en Carree (2003), abordamos el efecto directo sobre la variable de interés, la duración del banco, utilizando una nueva metodología que puede englobarse en el enfoque de los modelos de duración acelerada. Además de este enfoque alternativo, se realizan varias aportaciones, entre éstas hay dos que consideramos fundamentales. Por una parte, la relajación del supuesto de la distribución de probabilidad, Gompertz, dejando a ésta libre. Por otra parte, ante las dudas o desconocimiento de la

forma funcional de varios de los factores considerados claves para explicar el riesgo de salida del mercado, se deja que sean los propios datos los que modelen esa relación, en lugar de imponer o especificar una forma concreta.

3. Resultados

Los datos recogen información sobre bancos considerados activos en el mercado de depósitos en rublos a tres meses de Moscú. El comportamiento de estos bancos se puede tomar como representativo del comportamiento global de la banca rusa. Los resultados obtenidos refuerzan las conclusiones obtenidas por Carree (2003) y añaden nuevos resultados. Pero, lo que es más importante, es que para obtenerlos se ha utilizado un enfoque alternativo y bajo una modelización menos restrictiva. Así, obtenemos que aquellos bancos que tienen más experiencia y mayor tamaño sobreviven mejor. El tipo de interés ofrecido por los depósitos afecta a la duración del banco, siendo los que mayores niveles ofertan los primeros en salir del mercado. Además, se constata que el periodo de tiempo en el que el banco está en riesgo también es importante.

4. Contactar

Dirección: Jesus Orbe

Departamento de Economía Aplicada III (Econometría y Estadística),
Fac. de C. Económicas y Empresariales - Av. Lehendakari Agirre 83
48015 - Bilbao (Bizkaia)

Teléfono: 946013842 **Fax:** 946013754

e-mail: jesus.orbe@ehu.es

5. Agradecimientos

Este trabajo ha sido financiado por los proyectos BEC2003-02028 y MTM2004-00341 del Ministerio de Ciencia y Tecnología y UPV 00038.321-13631/2001 de la Universidad del País Vasco-Euskal Herriko Unibertsitatea. Los datos utilizados han sido gentilmente cedidos por M.A. Carree.

6. Bibliografía

- [1] Wheelock, D.C. y Wilson, P.W. (2000). Why do banks disappear? The determinants of US bank failures and acquisitions. *Review of Economics and Statistics* 82, 127-138.
- [2] Carree, M.A. (2003). A hazard rate analysis of Russian commercial banks in the period 1994-1997. *Economic Systems* 27, 255-269.

Fiabilidad de una medición en presencia de varios factores

Juan L. González Caballero¹, Juan L. Peralta Sáez², José Almenara Barrios³, Carolina Lagares Franco⁴

¹juanluis.gonzalez@uca.es, Dpto. de Estadística e Inv. Operativa, UCA

²jluis.peralta@uca.es, Dpto. de Estadística e Inv. Operativa, UCA

³jose.almenara@uca.es, Área de Med. Prev. y Salud Pública, UCA

⁴cama.lagarfranc@alum.uca.es, Becaria, UCA

Abstract

El coeficiente de correlación intraclasses permite medir la fiabilidad interobservadores y la fiabilidad interinformadores en variables cuantitativas, habiéndose propuesto varios tipos de coeficientes. Sin embargo, cuando se quiere analizar conjuntamente estos dos tipos de fiabilidad es necesario acudir a los coeficientes que propone la teoría de la generabilidad.

Palabras Clave: Fiabilidad, Coeficiente de correlación intraclasses, Teoría de la generabilidad

AMS: 97C40, 62H20

1. Introducción

La fiabilidad de una medición es una cuestión importante a la hora de obtener datos precisos en cualquier investigación. En Psiquiatría es frecuente analizar distintos tipos, entre las que se encuentran la fiabilidad interobservadores y la fiabilidad interinformadores. La teoría clásica (TC) define el coeficiente de correlación intraclasses (ICC) para medir este concepto en variables cuantitativas, habiéndose propuesto varios tipos por Shrout y Fleiss (1979) y siendo de nuevo tratados y extendidos por McGraw y Wong (1996). Sin embargo, cuando se quiere analizar conjuntamente estos dos tipos de fiabilidad es necesario acudir a la teoría de la generabilidad (TG), extensión de la teoría clásica que han propuesto Cronbach et al, (1972), en la que se analizan conjuntamente las fuentes de error que influyen en las observaciones de una medición mediante la descomposición de las componentes de la varianza y la construcción de coeficientes tipo ICCs.

2. Fiabilidad según el modelo clásico

El tipo de fiabilidad más estudiado es el denominado “acuerdo entre observadores” (fiabilidad interobservador) para el que se han definido distintos coeficientes de correlación intraclasses (ICC) (Barko, 1966) cuando se quiere evaluar el acuerdo entre 2 o más observadores en una variable cuantitativa, clasificado

en diferentes tipos por Shrout y Fleiss (1979) y posteriormente por McGraw y Wong (1996), según diversas situaciones de medición. No obstante, estos coeficientes sólo pueden contemplar, además de los sujetos experimentales, un factor de variación (observadores, fuentes, lugares, tiempos, ...)

3. Teoría de la generabilidad: Modelos de ICCs multifactoriales

Introducida por Cronbach et al. (1972), la TG es una técnica desarrollada para valorar la precisión, fiabilidad y validez, de cualquier proceso de medición, permitiendo al investigador examinar múltiples fuentes de error simultáneamente (p.ej.: diferentes ocasiones, instrumentos, observadores, etc) que pueden influir en el proceso de la medición y proporcionando modelos y coeficientes, análogos a los coeficientes de fiabilidad de la TC, que son capaces de analizar el grado de fiabilidad atribuible a varias fuentes de error.

4. Conclusiones

Con un cuestionario de utilización de servicios en pacientes esquizofrénicos, donde la misma información se ha recogido de tres fuentes distintas (paciente, familiar e historia clínica) y por dos observadores diferentes, este trabajo explora la posibilidad de analizar conjuntamente estos dos factores de influencia sobre el proceso de medición, utilizando para ello la teoría de la generabilidad.

5. Agradecimientos

Este trabajo ha sido financiado parcialmente a través de la red de excelencia RIRAG-SMD (G03/061) del Instituto de Salud Carlos III.

6. Bibliografía

- [1] Bartko, J.J. (1966). The intraclass correlation coefficient as a measure of reliability. *Psychological Reports*, 19, 3-11
- [2] Brennan, R.L. (2001). *Generalizability theory*. New York: Springer-Verlag.
- [3] Cronbach, L.J., Gleser, G.C., Nanda, H. y Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: Wiley.
- [4] McGraw, K.O. y Wong, S.P. (1996). Forming inferences about some intra-class correlation coefficients. *Psychological Methods*, 1, 30-46.
- [5] Shrout, P.E. y Fleiss, J.L. (1979). Intraclass correlation: Uses in assessing rater reliability. *Psychological Bulletin*, 86, 420-428.

Filtrado e inter/extrapolación funcional de procesos espacio-temporales fractales

Rosaura Fernández-Pascual¹, María D. Ruiz-Medina², José M. Angulo²

¹ rpascual@ujaen.es, Departamento de Estadística e Investigación Operativa,
Universidad de Jaén

² mruiz@ugr.es; jmangulo@ugr.es, Departamento de Estadística e
Investigación Operativa, Universidad de Granada

Abstract

Se considera el problema de filtrado e inter/extrapolación funcional en procesos de difusión anómala. En particular, se estudian modelos definidos mediante ecuaciones pseudodiferenciales fraccionarias parabólicas. Se deriva una solución estable en términos de la descomposición *wavelet-vaguelette* de los procesos involucrados.

Claves: Bases de wavelets ortonormales espacio-temporales, estimación funcional, expansiones ortogonales.

1. Introducción

En este trabajo se obtiene una extensión espacio-temporal de los resultados derivados en [2], [1] y [2] sobre filtrado e inter/extrapolación funcional de procesos espaciales. Más concretamente, se formula una solución funcional estable para el problema de filtrado e inter/extrapolación espacio-temporal asociado al proceso Z definido mediante la ecuación de evolución

$$\frac{\partial Z}{\partial t}(t, \mathbf{x}) = -L_{\mathbf{x}}Z(t, \mathbf{x}), \quad Z(0, \mathbf{x}) = U_0(\mathbf{x}), \quad (t, \mathbf{x}) \in \mathbb{R}_+ \times \mathbb{R}^n, \quad (2.36)$$

donde $L_{\mathbf{x}}$ es un operador pseudodiferencial fraccionario que refleja el ‘comportamiento anómalo’ del proceso de difusión Z . La condición inicial U_0 se define en términos de un proceso espacial que puede presentar características fractales. El proceso Z se define en sentido débil, en términos de un sistema apropiado de funciones *test*, que compensan la singularidad (fractalidad local) introducida en términos del operador $L_{\mathbf{x}}$. En particular, en este trabajo se considera la descomposición *wavelet-vaguelette* espacio-temporal derivada en [4], que proporciona un sistema de funciones *test* adecuado para la regularización del problema de estimación estudiado.

2. Estimación funcional espacio-temporal

Sea $Y(t, \mathbf{x}) = Z(t, \mathbf{x}) + N(t, \mathbf{x})$, para $t \in \mathbb{R}_+$, $\mathbf{x} \in D_Y^{OBS}$, el proceso observable definido a partir de Z en (2.36) y de N , ruido blanco espacio-temporal

incorrelado con Z , donde D_Y^{OBS} denota el conjunto de localizaciones observables. La aplicación débil, en términos de funciones *test*, del Teorema de la Proyección Ortogonal conduce a la siguiente expresión del estimador funcional lineal mínimo-cuadrático de Z :

$$\hat{Z}(t, \mathbf{x}) = \mathcal{P}_{t, \mathbf{x}} Y(t, \mathbf{x}) = [R_{ZY} R_{YY}^{-1} Y](t, \mathbf{x}), \quad (t, \mathbf{x}) \in \mathbb{R}_+ \times \mathbb{R}^n, \quad (2.37)$$

donde $\mathcal{P}_{t, \mathbf{x}}$ representa el operador que define la proyección ortogonal de la función aleatoria Z sobre el espacio de Hilbert $H(Y)$ generado por el proceso Y . Los operadores R_{YY} y R_{ZY} respectivamente denotan el operador de covarianza del proceso Y y el operador de covarianza cruzada del proceso Z con Y .

3. Simulación

Se desarrolla un estudio de simulación que ilustra la influencia del orden de fractalidad de Z y, en particular, de U_0 sobre la estabilidad de la solución (2.37), aproximada en términos de diferentes familias de *wavelets* Daubechies transformadas. En particular, se considera el proceso

$$Z(t, \mathbf{x}) = \int_{\mathbb{R}_+ \times \mathbb{D}} \frac{1}{(4\pi t)^{1/2}} \exp \left\{ -\frac{\|\mathbf{x} - \mathbf{y}\|^2}{4t} \right\} \|\mathbf{y} - \mathbf{z}\|^{-2+\alpha/2} \varepsilon(\mathbf{z}) d\mathbf{z} d\mathbf{y},$$

para $\alpha = 0,65, 0,75, 0,85$, siendo ε ruido blanco espacio-temporal y $\mathbb{D} \subset \mathbb{R}^2$.

4. Agradecimientos

Este trabajo ha sido financiado en parte por los Proyectos MTM2005-08597 de la DGI, MEC, y P05-FQM-00990 de la CICE, Junta de Andalucía.

5. Bibliografía

- [1] Fernández-Pascual, R., Ruiz-Medina, M.D. y Angulo, J.M. (2004). Wavelet-Based Functional Reconstruction and Extrapolation of Fractional Random Fields. *Test* 13, 1-28.
- [2] Fernández-Pascual, R., Ruiz-Medina, M.D. y Angulo, J.M. (2006). Estimation of Intrinsic Processes Affected by Additive Fractal Noise. *Journal of Multivariate Analysis*. Aceptado (doi:10.1016/j.jmva.2005.10.006).
- [3] Ruiz-Medina, M.D. y Angulo, J.M. (2002). Spatio-Temporal Filtering Using Wavelets. *Stochastic Environmental Research and Risk Assessment* 16, 241-266.
- [4] Ruiz-Medina, M.D. Angulo, J.M. y Fernández-Pascual, R. (2006). Spatio-Temporal Wavelet-Vaguelette Decomposition of Fractal Evolution Processes. (Sometido).

Fitting grouped Cox model using phi-divergences

M^a del Carmen Pardo

mcapardo@mat.ucm.es, Department of Statistics and Operational Research (I). Complutense University of Madrid

Abstract

El modelo de Cox agrupado modeliza la relación entre una variable respuesta categórica y un conjunto de variables explicativas. El objetivo de este trabajo consiste en proponer para este modelo nuevos procedimientos de estimación y contraste basados en la phi-divergencia que mejoren los procedimientos clásicos basados en la verosimilitud.

Keywords: grouped Cox model, phi-divergences.

1. Introduction

Consider a response random variable $\mathbf{Y} \in \{1, \dots, J\}$, which is observed together with a given vector $\mathbf{x}^T = (x_1, \dots, x_m)$ of explanatory variables. In this paper we suppose that the dependence between \mathbf{Y} and \mathbf{x}^T can be modeled by using the *grouped Cox model or proportional Hazards model*

$$\Pr(\mathbf{Y} \leq r | \mathbf{x}^T) = 1 - \exp(-\exp(\alpha_r + \mathbf{x}^T \beta)), \quad r = 1, \dots, J-1 \quad (2.38)$$

where $\beta^T = (\beta_1, \dots, \beta_m)$. We denote by $\gamma^T = (\alpha_1, \dots, \alpha_{J-1}, \beta^T)$ the vector of unknown parameters. McCullagh [1] discussed its use and its connections with related proportional hazards models for survival data.

2. Statistical Inference based on phi-divergences

To fit this model to the data, we would like to estimate and test the model parameters, select a parsimonious set of explanatory variables, test for goodness-of-fit and analyze residuals, just to mention some aspects of statistical inference.

An approach for solving the estimation and testing problem is the likelihood approach. The parameters are estimated by the maximum likelihood method and tested by likelihood ratio tests having asymptotic χ^2 distributions when the total number of observations tends to infinity. For fixed \mathbf{x} , the vector $\mathbf{n}(\mathbf{x}) = (n_1(\mathbf{x}), \dots, n_J(\mathbf{x}))$ of observed numbers of objects in response categories has the multinomial distribution which follows from the assumption of independence of the individual observations $Y_i, i = 1, \dots, N$, then the loglikelihood equals

$$l = \sum_{i=1}^N \sum_{j=1}^J n_j(\mathbf{x}_i) \log \pi_j(\mathbf{x}_i)$$

where $\pi_j(\mathbf{x}) = P(\mathbf{Y} = j \mid \mathbf{x})$, $j = 1, \dots, J$. If model (5.98) is combined with the loglikelihood l , then the model parameters can be estimated and tested.

Let us introduce the following two random vectors,

$$\widehat{\mathbf{p}} = \left(\frac{\mathbf{n}(\mathbf{x}_1)^T}{n}, \dots, \frac{\mathbf{n}(\mathbf{x}_N)^T}{n} \right)^T$$

$$\mathbf{p}(\gamma) = \left(\frac{m(\mathbf{x}_1)}{n} \pi(\mathbf{x}_1)^T, \dots, \frac{m(\mathbf{x}_N)}{n} \pi(\mathbf{x}_N)^T \right)^T$$

where

$$m(\mathbf{x}_1) = \sum_{j=1}^J n_j(\mathbf{x}_i), \quad n = \sum_{i=1}^N m(\mathbf{x}_i) \text{ and } \pi(\mathbf{x}_i)^T = (\pi_1(\mathbf{x}_i), \dots, \pi_J(\mathbf{x}_i)).$$

It is easy to see that maximize l is equivalent to minimize the Kullback-Leibler divergence measure between the probability vectors $\widehat{\mathbf{p}}$ and $\mathbf{p}(\beta)$. So, new procedures for solving the estimation and testing problem can be considering using instead of the Kullback-Leibler divergence measure every divergence measure. In this work we consider the known ϕ -divergence measure defined by

$$D_\phi(\widehat{\mathbf{p}}, \mathbf{p}(\gamma)) = \sum_{j=1}^J \sum_{i=1}^N \pi_j(\mathbf{x}_i) \frac{n(\mathbf{x}_i)}{n} \phi\left(\frac{n_j(\mathbf{x}_i)}{\pi_j(\mathbf{x}_i) n(\mathbf{x}_i)}\right)$$

with $\phi \in \Phi^*$ and Φ^* is the class of all convex functions $\phi(x)$, $x > 0$, such that at $x = 1$, $\phi(1) = 0$, $\phi''(1) > 0$, for $x = 0$, $0\phi(0/0) = 0$ and $0\phi(p/0) = \lim_{u \rightarrow \infty} \phi(u)/u$. For more details about ϕ -divergences see Pardo [2],

As this family of divergences contains the Kullback-Leibler divergence measure, the likelihood procedures will be a particular case of ours.

3. Acknowledgement

This work was supported by Grants BMF2003-00892 and UCM2005-910707.

4. Bibliography

- [1] McCullagh, P. (1980). Regression Models for Ordinary Data. *Journal of the Royal Statistical Society B* 42, 109-142.
- [2] Pardo, L. (2006). *Statistical Inference based on Divergence Measures. (Statistics: Series of Textbooks and Monographs)*. Chapman & Hall/CRC. New York.

Funcionamiento Diferencial de los Ítems y del Test (DFIT): Su aplicación en el QUALEFFO

M. Sánchez Barba¹, P. Vicente Galindo²

¹mersanbar@usal.es, ²purivg@usal.es

Departamento de Estadística, Universidad de Salamanca

Abstract

El objetivo de este trabajo es evaluar la posible presencia de funcionamiento diferencial de los ítems (DIF) y funcionamiento diferencial del test (DTF) en la escala QUALEFFO para examinar si los ítems y el test funcionan de forma distinta en hombres y en mujeres.

Palabras Clave: DIF, DTF, QUALEFFO

1. Introducción

La Osteoporosis constituye un problema de Salud Pública creciente y de primera magnitud, tanto por su incidencia, que continua creciente a medida que cree la esperanza de vida, como la gravedad y el coste de sus consecuencias en forma de fractura de cadera y columna, principalmente.

Según la OMS (Organización Mundial de la Salud), el 50 % de mujeres de más de 70 años padecen osteoporosis. En España, cada año, afecta a 3,5 millones de personas, se producen 33.000 fracturas de cadera, supone un coste sanitario de 442 millones de euros, por lo que la osteoporosis está considerada como la epidemia del siglo XXI.

En el estudio del funcionamiento diferencial (DF) se comparan las respuestas de sujetos de distintos grupos a un ítem o a un test. Se detecta DF en un ítem o en un test, cuando sujetos de un mismo nivel de habilidad tienen distinta probabilidad de obtener puntuaciones similares sobre el mismo ítem o el test, según al grupo al que pertenezcan *grupo focal* o *grupo de referencia*.

RAJU, VANDER LINDEN y FLEER (1995) propusieron un procedimiento paramétrico basado en la Teoría de Respuesta al Ítem, para evaluar el Funcionamiento Diferencial del Ítem (DIF) y el Funcionamiento Diferencial del Test (DTF) conocido como “Funcionamiento Diferencial de Ítems y de Test” (DFIT, Differential Functioning of Items and Tests).

En el marco del DFIT, el funcionamiento diferencial se define como una diferencia en el valor esperado del ítem, o del test, para los individuos con la misma situación en el constructo latente θ .

2. Material y Métodos

En este trabajo, evaluamos el funcionamiento diferencial de los ítems (DIF) y el funcionamiento diferencial del test (DTF) usando la metodología DFIT en el cuestionario QUALEFFO (Questionnaire of the European Foundation for Osteoporosis, BADIA & HERDMAN, 1999) el cual valora la Calidad de Vida en pacientes con osteoporosis y contiene 35 ítems que consideran 7 dimensiones de la calidad de vida relacionada con la salud. La respuesta de cada ítem viene dada en escala ordinal, desde 1 (ningún problema) a 5 (muchos problemas).

Para la estimación de los parámetros de los ítems para los hombres (grupo focal) y para las mujeres (grupo de referencia) por separado, utilizaremos el modelo de la TRI de Respuesta Graduada de SAMEJIMA (1969), implementado en el programa MULTILOG (THISSEN, 1991). Bajo este modelo, la probabilidad de seleccionar una determinada categoría para un cierto ítem, puede calcularse a partir de la función logística:

$$P_{ik}^*(\theta) = \frac{e^{[Da_i(\theta_s - b_{ik})]}}{1 + e^{[Da_i(\theta_s - b_{ik})]}} , \text{ donde :}$$

b_{ik} es el límite o umbral entre la categoría k y la categoría $k+1$ asociadas al ítem i ; a_i es la pendiente del ítem o parámetro de discriminación; θ_s es el parámetro del rasgo y D una constante.

3. Bibliografía

- [1] Badia, X., Herrdman, M. (1999). Adaptación transcultural al español de los cuestionarios OQLQ y QUALEFFO para la evaluación de la calidad de vida relacionada con la salud en mujeres con fractura vertebral osteoporótica. *Rev Esp Enf Met Oseas* 8, 136-140.
- [2] Raju, N., Van Der Linden, W., Fleer, P. (1995). IRT-based internal measures of differential functioning of ítems and test. *Applied Psychological Measurement* 19, 353-368.
- [3] Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph*. 17.
- [4] Thissen, D. (1991). MULTILOG: multiple category item análisis and test scoring using item response theory. Chicago: Scientific Software International, Inc.

Guía interactiva de autoaprendizaje de SPSS

J. García Leal, E. Hidalgo Ruiz, A. Lara Porras, J. M. Quesada Rubio, E. Ramos Ábalos, R. Raya Miranda, J. E. Ruiz Castro, I. R. Sánchez Borrego, J. Tarifa Blanco

Departamento de Estadística e Investigación Operativa, Facultad de Ciencias,
Universidad de Granada, Campus de Fuentenueva, s/n, 18071, Granada
e-mail: juliagl@ugr.es, alara@ugr.es, jeloy@ugr.es

Abstract

Se presenta una guía interactiva del programa SPSS, que consta de un portal Web y de un emulador del programa. El emulador de SPSS que se ha desarrollado guía en la realización de ejercicios de Estadística de forma interactiva. El alumno dispone de una ayuda complementaria a la del profesor que le facilita el aprendizaje de la Estadística con ordenador.

Palabras Clave: SPSS, portal WEB, emulador.

AMS: 97-04; 97U60; 97U70

1. Introducción

El desarrollo de esta guía interactiva se enmarca dentro de un proyecto más general que tiene como objetivo la reforma y mejora de la metodología utilizada en las dos asignaturas de estadística de la Licenciatura de Ciencias Biológicas de la Universidad de Granada. El proyecto surge de una realidad experimental ante las necesidades estadísticas de un licenciado en Ciencias Biológicas en el campo laboral y profesional. Durante distintos cursos académicos se ha ido modificando la metodología de la enseñanza de Estadística en la licenciatura de Biología. En una primera fase del trabajo se desarrolló nuevo material didáctico multimedia para el alumno, García Leal y otros (2003a, 2003b, 2004). Recientemente se ha creado un emulador de SPSS (nuestra Universidad posee la licencia del programa SPSS) que sirve de guía interactiva para el aprendizaje de la Estadística con ordenador.

2. Guía computacional interactiva

Se presenta una guía interactiva como apoyo docente para la realización de las prácticas de las asignaturas a las que va dirigido el proyecto.

La guía consta de dos partes:

1. *Un Portal Web* donde se le presentan al alumno las diferentes prácticas propuestas (basadas en ejercicios con SPSS). Cada práctica contiene una introducción de ésta y un ejercicio resuelto con SPSS utilizando un emulador del mismo. Finalmente se le plantea un ejercicio propuesto.

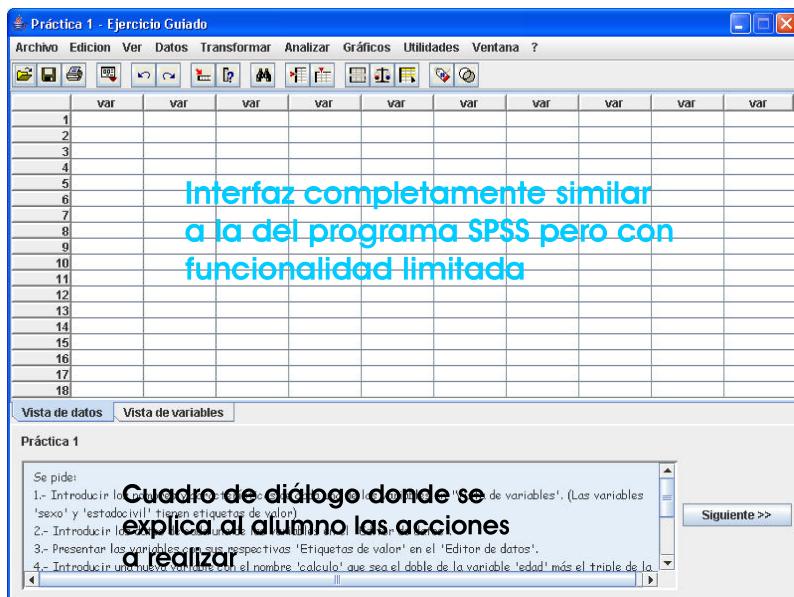


Figura 1: Emulador de SPSS

2. *Un emulador del programa SPSS.* Después de que el alumno haya leído todo la documentación e información referente a la práctica que esté realizando (todo esto desde el Portal Web), entra en juego un emulador del programa SPSS. Con este emulador el alumno empieza a tomar contacto con el programa, ya que su función es guiarlo en la realización de un ejercicio específico de forma interactiva. A continuación se muestra una vista del emulador indicando sus partes principales.

3. Bibliografía

- [1] García Leal, J; Lara Porras A.; Navarrete Álvarez, E.; Rico Castro, N.; Román Montoya, Y.; Ruiz Castro, J.E. y Sánchez Borrego, I. R. (2004). *Estadística para la Biología*. Granada. España.
- [2] García Leal, J; Lara Porras A.; Navarrete Álvarez, E.; Rico Castro, N.; Román Montoya, Y.; Ruiz Castro, J.E. y Sánchez Borrego, I. R. (2003a). *Introducción a la Estadística para Biología*. Granada. España.
- [3] García Leal, J; Lara Porras A.; Navarrete Álvarez, E.; Rico Castro, N.; Román Montoya, Y.; Ruiz Castro, J.E. y Sánchez Borrego, I. R. (2003b). *Técnicas Estadísticas para Investigación en Biología*. Granada (Spain).
- [4] Lara Porras, A.M. (2002). *Estadística para Ciencias Biológicas y Ciencias Ambientales. Problemas y Exámenes resueltos*. Proyecto Sur. Granada.

Identificación y caracterización de tipologías docentes. Aplicación a la UPV

Mónica Martínez¹, Jose Miguel Carot², José Jabaloyes Vivas³,
Pau Miró i Martínez⁴

¹ momargo@eio.upv.es, ²jcarot@eio.upv.es, ³jabaloye@eio.upv.es,
⁴pamimar@eio.upv.es

Departamento de Estadística, Investigación Operativa Aplicadas y Calidad
Universidad Politécnica De Valencia

Abstract

En el presente trabajo, tras haber obtenido los distintos grupos homogéneos mediante la aplicación del análisis de conglomerados, en función de las puntuaciones medias obtenidas en los ítems se procede a identificar tipologías docentes en función de diversas características descriptivas referentes a la asignatura y al profesor mediante las técnicas CHAID Exhaustivo y SIMCA.

Palabras Clave: Análisis encuestadas, Caracterización de perfiles, CHAID, SIMCA

AMS: 62H25

1. Introducción

El cuestionario de opinión de los estudiantes (CEDA), es la herramienta de evaluación de la actividad docente más generalizada en las universidades europeas. Con la utilización conjunta de las técnicas que se proponen a continuación, se pretende optimizar el uso del CEDA como herramienta de medida e indicador de calidad de la enseñanza en la universidad, para introducir actuaciones de mejora continua en los procesos educativos de la UPV.

En el presente trabajo, tras haber obtenido los distintos grupos homogéneos mediante la aplicación del análisis de conglomerados, en función de las puntuaciones medias obtenidas en los ítems (Martínez et al., 2004), se procede aplicar la técnica de árboles de clasificación CHAID (Chi-Squared Automatic Interaction Detector), con la finalidad segmentar los datos en función de determinadas variables predictoras de las asignaturas y/o de los profesores de las que disponíamos información. A continuación, se utilizará el método de clasificación SIMCA (Soft Independent Modelling of Class Analogy), no con la finalidad de obtener una modelización y clasificación de las observaciones, sino para obtener las variables de mayor poder discriminante entre los grupos formados utilizando los modelos construidos con cada uno de los nodos

2. Conclusiones

El método de clasificación CHAID se ha mostrado adecuado para identificar tipologías docentes en función de variables referidas al profesor y/o asignatura., aunque la batería de predictores podría ser mejorada. Se confirma que existen influencias jerarquizadas de las variables descriptivas o factores utilizados sobre la calidad docente percibida por los alumnos. Se pone de manifiesto que la calidad docente del profesorado valorada mediante el CEDA, no puede ser tratada como un grupo homogéneo sino que está integrado por subconjuntos diferenciados entre sí.

La aplicación del método SIMCA permitió detectar los ítems de mayor poder discriminante entre los modelos propuestos (correspondientes a cada uno de los nodos) y obtener las distancias obtenidas entre modelos. Este método multivariante se ha propuesto como una técnica muy adecuada para comparar las tipologías obtenidas en la segmentación proporcionada por el método CHAID. Sin embargo, este método por sí solo no proporciona resultados satisfactorios en la clasificación con las variables consideradas.

3. Bibliografía

- [1] Biggs, D., Ville, B. y Suen, E. (1991). A method of choosing multiway partitions for classification and decision trees. *Journal of Applied Statistics*, 18, (1), 49 -60.
- [2] Martínez, M; Carot Sierra, J.M; Jabaloyes, J.M.; Miró, P. *Comparación de distintos algoritmos de análisis clúster para identificar perfiles homogéneos de profesorado de la UPV*. Congreso Nacional de Estadística e Investigación Operativa. XXVIII Edición. 25-29 Octubre 2004.
- [3] Quinlan. J.R. (1986). Induction of decision trees. *Machine Learning*, 1, 81 -106.
- [4] Wold, S. (1976). Pattern recognition by means of disjoint Principal Components models. *Pattern Recognition*, 8, 127-139.

Indicador local de riqueza/pobreza: caracterización y relación con el indicador de renta municipal en Galicia

Patricia Sánchez Fernández¹, Esther López Vizcaíno², Antonio Vaamonde Liste³

¹patricia@uvigo.es, Departamento de Economía Financiera, Universidad de Vigo

²esther.lopez.vizcaino@xunta.es, Instituto Galego de Estatística (IGE)

³vaamonde@uvigo.es, Departamento de Estadística e Investigación Operativa, Universidad de Vigo

Abstract

La presente comunicación pretende desarrollar, a modo de propuesta, un nuevo método para la medida de la riqueza/pobreza procediendo, para ello, con un análisis desde el punto de vista de la dotación material de bienes y servicios. Junto a su desarrollo empírico se procede a buscar relación con un indicador de renta.

Palabras Clave: pobreza, riqueza, renta

1. Introducción

La pobreza, la desigualdad social, la discriminación o la marginalidad económica se configuran, ya no sólo desde una perspectiva económica, sino global como alguno de los más acuciantes problemas de la humanidad. Una de las principales dificultades para su solución radica en su conceptualización, esto es, en muchas ocasiones cuando se habla de pobre o de marginado no está claro a lo que nos estamos refiriendo. Ciertamente, este hecho condiciona de manera determinante la medición del problema y, por ende, la búsqueda de una solución al mismo.

El presente trabajo se enmarca dentro de la creciente preocupación por cuestiones sociales centrándonos en el caso de la Comunidad Autónoma de Galicia desde una perspectiva territorial. Así, su primera pretensión consiste en desarrollar un nuevo método de medida de la pobreza y, posteriormente, proceder con su aplicación empírica. Para ello, se analiza la riqueza y pobreza desde el punto de vista de la dotación de bienes y servicios de los hogares.

Esta perspectiva aporta una visión diferente a enfoques más tradicionales, al tiempo que posibilita su tratamiento como una realidad multidimensional en línea con los más recientes estudios en este campo¹. En este sentido conviene

¹Véase S. Bibi (2003).

señalar la utilización de técnicas estadísticas multivariantes para el tratamiento de la información y, de forma especial, el análisis discriminante.

Finalmente se analiza la relación existente entre el indicador propuesto con el Indicador municipal de la renta de los hogares elaborado por el Instituto Galego de Estatística (*IGE*) con el objetivo de establecer efectos de complementariedad o substitución entre dos conceptos vinculados como son la riqueza y la renta.

La organización de la comunicación responde a la siguiente estructura. En primer lugar se presenta el Indicador municipal de la renta de los hogares de Galicia para, a continuación, desarrollar un método para la medida de la pobreza. Aquí, se hace hincapié en sus características más relevantes, así como en los pasos en los que se materializa su desarrollo. Seguidamente se detallan los principales resultados obtenidos a lo largo de la investigación empírica llevada a cabo, para finalizar con los aspectos más destacables que pueden ser extraídos a modo de conclusión con especial atención entre los dos indicadores.

2. Indicador municipal de la renta de los hogares gallegos

Operación estadística del Instituto Galego de Estatística (*IGE*) en el que se ofrece una estimación de la renta dos fogares residentes en Galicia así como de su distribución a nivel municipal. El período temporal al que se refiere son los años 2000-2002 y enlaza con la primera información elaborada en este sentido del año 1996 consistente en un Indicador de renta municipal².

3. Método de medida de la pobreza: el Índice de Pobreza Material (IPM)

Se desarrolla un indicador para la medida de la pobreza que, siguiendo procedimientos análogos en estudios de esta índole se divide en cuatro partes: identificación y clasificación de los hogares, formulación del Índice (IPM), determinación de la línea de pobreza y cuantificación de la pobreza.

4. Principales resultados

Se detallan los resultados más relevantes de la aplicación empírica siguiendo la estructura secuencial reseñada anteriormente.

5. Bibliografía

- [1] Bibi S. (2003). Measuring poverty in a multidimensional perspective: A review of literature. *Working Paper Poverty Monitoring, Measurement and Analysis* (PMMA)

²Nótese que se produce un cambio de metodología ya que el actual hace referencia exclusivamente al sector de los hogares.

Influencia Local: estimadores de razón y de Horvitz-Thompson

I. Barranco Chamorro¹, A. Enguix González², J.L. Moreno Rebollo³⁴, J.M. Muñoz Pichardo⁵

¹chamorro@us.es, ²aenguix@us.es, ³jlmoreno@us.es, ⁵juanm@us.es

Dpto. de Estadística e Investigación Operativa, Universidad de Sevilla

⁴Centro Andaluz de Prospectiva

Abstract

El objetivo de este trabajo es el estudio de la influencia local sobre los estimadores de razón y de Horvitz-Thompson, supuesto un muestreo aleatorio simple y un diseño π ps, respectivamente. Analizamos dos perturbaciones sobre la variable de información auxiliar que modelan situaciones de interés práctico.

Claves: Influencia local, estimador de razón, estimador de Horvitz-Thompson.

AMS: 62D05, 62J20

1. Introducción

En la literatura se pueden encontrar un gran número de trabajos que abordan el problema de la influencia local, introducida por Cook (1986), en casi todas las áreas de la estadística.

En general, y a pesar de su reconocido interés, Smith (1987), el análisis de influencia en el muestreo en poblaciones finitas ha sido escasamente tratado en la literatura y, hasta donde nosotros sabemos, no existe ningún trabajo en el que se aborde la influencia local.

El objetivo de este trabajo es el análisis de influencia local sobre dos estimadores clásicos del total poblacional: el estimador de razón (\widehat{T}_R) bajo muestreo aleatorio simple, y el estimador de HT (\widehat{T}_π) bajo un diseño π ps.

2. Planteamiento, perturbaciones y análisis

Sea $U = \{1, \dots, N\}$ el conjunto de índices de una población finita, y_i y x_i los valores de la variable objetivo y de una variable auxiliar, respectivamente, para la unidad i . Denotemos por $T(y) = \sum_U y_i$ al total poblacional de la variable y , mediante s a una muestra aleatoria seleccionada de acuerdo a un diseño muestral y sea π_i la probabilidad de inclusión de primer orden asociada a i .

Para el desarrollo del trabajo es interesante destacar que tanto $\widehat{T}_R = \bar{y}/\bar{x} T(x)$ como $\widehat{T}_\pi = \sum_s y_i/\pi_i$ con $\pi_i \propto x_i$ hacen uso de la variable de información auxiliar (x) e implícitamente asumen un modelo de razón con varianza proporcional a x , $E_\xi(Y_i) = \beta x_i$, $V_\xi(Y_i) = \sigma^2 x_i$.

El análisis de influencia local requiere fijar una perturbación (ω), sobre el modelo o sobre los datos, y un estadístico sobre el que evaluar el efecto de la perturbación.

En este trabajo consideramos dos perturbaciones, una sobre el modelo $V_{\xi(\omega)}(Y_i) = \sigma^2 x_i \omega_i$ (P1) y otra sobre la variable auxiliar $x_i(\omega) = x_i \omega_i$ (P2). P1 es una perturbación de la estructura de varianzas del modelo y P2 es una perturbación que puede usarse para modelar situaciones en las que la variable auxiliar puede estar afectada por errores.

Realizamos el estudio de influencia local sobre las superficies $(\omega, \widehat{T}_R(\omega))$ y $(\omega, \widehat{T}_\pi(\omega))$ siguiendo la aproximación de segundo orden, Wu and Luo (1993), basada en el análisis de las direcciones correspondientes a las curvaturas máximas de los estimadores bajo las perturbaciones.

3. Agradecimientos

Este trabajo está parcialmente financiado por el proyecto MTM2004-01433 del Ministerio de Educación y Ciencia y por el grupo de investigación FQM153 de la Junta de Andalucía.

4. Bibliografía

- [1] Cook, R.D. (1986). Assessment of Local Influence. *Journal of the Royal Statistical Society B* 48, 133-169.
- [2] Smith, T.M.F. (1987). Influential observations in survey sampling. *Journal of Applied Statistics* 14, 143-152.
- [3] Wu, X. and Luo, Z. (1993). Second-order approach to local influence. *Journal of the Royal Statistical Society B* 55, 929-936.

Influencia macroeconómica en el mercado de la vivienda

Cristina Oyáñez, Patricia Rodríguez, Andrés Toledo

estudiosmercado@cohispania.com, Dpto. de Estudios, Cohispania S.A.

Abstract

En este estudio explicamos la evolución del precio de la vivienda libre nueva y usada en distintos países de Europa, entre ellos, España. Para ello definimos un modelo econométrico que permite explicar el comportamiento del precio de la vivienda a través de variables macroeconómicas, cuantificar su impacto y prever el comportamiento de esta variable en el futuro.

Palabras Clave: Econometría, Arima, Vivienda

1. Introducción

Dada la importancia que suscita el precio de la vivienda, no sólo en España sino en toda Europa, se ha considerado de interés estudiar este mercado en distintos países de este área geográfica. Presentamos los resultados obtenidos para España, Francia y Reino Unido.

El estudio explica, con la formulación dada por (1), la evolución del precio de la vivienda a partir de las variables macroeconómicas PIB, Índice Bursátil, Tipo de Interés, Tasa de Paro y N° de Hogares. Cuantificamos la contribución de cada una de ellas sobre la variable precio y presentamos previsiones a un horizonte temporal de dos años.

2. Modelo Econométrico

El modelo econométrico utilizado intenta conocer las relaciones dinámicas a largo plazo entre el precio de la vivienda y el resto de variables a través de la estimación de vectores de cointegración. En términos logarítmicos, la ecuación es la siguiente:

$$\begin{aligned} \log(P) = & \log(\mu) + \beta \cdot \log(IB) + \delta \cdot \log(PIB) \\ & + \gamma \log(TP) + \pi \cdot \log(Hog) + \alpha \cdot \log(VA) + \epsilon \cdot \log(e) \end{aligned} \quad (2.39)$$

o de forma equivalente,

$$P = \mu \cdot IB^\beta \cdot PIB^\delta \cdot TP^\gamma \cdot Hog^\pi \cdot VA^\alpha \cdot e^\epsilon$$

donde $P = Pm^2/IPC$ es el precio de la vivienda en términos reales; μ es la constante; IB es el índice bursátil; PIB es el producto interior bruto en términos

reales por hogar; TP es la tasa de paro; Hog es el número de hogares; VA es el valor actual calculado a partir del tipo de interés interbancario a doce meses; e es el residuo.

Se realizó una primera estimación en la que se observaron las elasticidades de cada variable explicativa sobre el precio de la vivienda. Posteriormente estimamos el residuo mediante un modelo ARIMA, obteniéndose un ruido blanco.

El modelo se ha aplicado a cada uno de los países estudiados, obteniendo resultados fiables.

3. Principales Resultados

La interacción de las variables económicas explicativas del precio de la vivienda constituye la primera conclusión del presente estudio. Se ha analizado y representado gráficamente el porcentaje de contribución sobre el precio de cada una de las variables macroeconómicas. Es de advertir, que la variable con mayor contribución sobre el crecimiento del precio de la vivienda en estos tres países es el PIB. En el caso de España y Francia la siguiente variable más contributiva es la Tasa de Paro y la que menor contribución aporta es el Valor Actual (Tipo de Interés). Sin embargo, en Reino Unido es, el Valor Actual (Tipo de Interés), la variable que más contribuye por detrás del PIB.

De la previsión del precio de la vivienda a dos años vista puede concluirse que, en los países europeos estudiados, hay un síntoma común de ralentización del crecimiento del precio de la vivienda.

En la caracterización del escenario futuro de las variables económicas determinantes en el precio de la vivienda en España se han utilizado previsiones publicadas por distintos institutos de reconocido prestigio (como, por ejemplo, FUNCAS).

Dirección:

COHISPANIA S.A.
Departamento de Estudios
C. Portillo del Pardo, 14 - 28023 Madrid, Spain

Teléfono: (+34) 913070035 **Fax:** (+34) 913071795

e-mail: estudiosmercado@cohispania.com **www:** <http://www.cohispania.com>

4. Bibliografía

- [1] Pindyck, R. S. And Rubinfeld, D. L. (2001). “*Econometría. Modelos y pronósticos*”. McGraw-Hill, Madrid.
- [2] Martínez Pagés, J. y Maza L.A. (2003). “*Análisis del Precio de la vivienda en España*”. Documentos de trabajo del Banco de España.

Intervalos de predicción en series temporales utilizando el método de Holt-Winters

José D. Bermúdez¹, José V. Segura², Enriqueta Vercher³

¹bermudez@uv.es, Departamento de Estadística e Investigación Operativa,
Universidad de Valencia

²jvsh@umh.es, Centro de Investigación Operativa, Universidad Miguel
Hernández de Elche

³vercher@uv.es, Departamento de Estadística e Investigación Operativa,
Universidad de Valencia

Abstract

En este trabajo se presenta una formulación del procedimiento Holt-Winters que permite obtener conjuntamente los estimadores máximo verosímiles de los parámetros de suavizado y de las condiciones iniciales. Se obtiene una nueva aproximación a los intervalos de predicción que mejora los resultados empíricos de las aproximaciones habituales.

Palabras Clave: Suavizado exponencial, Modelos Lineales, Simulación.

1. Introducción

Los métodos de suavizado exponencial son una de las técnicas de predicción más utilizadas en la pequeña y mediana empresa. Las competiciones de predicción [4] han demostrado las buenas predicciones de los métodos Holt-Winters, obtenidas mediante procedimientos completamente automatizados y con un mínimo esfuerzo computacional. Además, recientes desarrollos han proporcionado una clase de modelos de espacio de estados para los que los métodos de suavizado exponencial son óptimos [5, 3].

Aunque los métodos Holt-Winters proporcionan buenas predicciones puntuales, no ocurre lo mismo con sus intervalos de predicción. Se han propuesto distintas fórmulas para el cálculo de dichos intervalos, pero estudios empíricos y de simulación [6, 5] han mostrado que los intervalos propuestos tienden a ser demasiado estrechos, en el sentido que el porcentaje de observaciones que finalmente caen fuera de los intervalos es mucho mayor de lo esperado.

La estimación de los parámetros desconocidos del modelo es otra de las controversias que actualmente acompañan a los métodos de Holt-Winters. Habitualmente los parámetros de suavizado se estiman minimizando el error de predicción de una etapa, una vez que ya se han fijado las condiciones iniciales, lo que suele hacerse mediante algún algoritmo heurístico. Recientemente se ha propuesto estimar conjuntamente las condiciones iniciales y los parámetros de suavizado utilizando estimación máximo verosímil [5, 3] y otros métodos de estimación [1, 2].

En este trabajo consideramos el método Holt-Winters aditivo, introduciendo la componente estocástica del modelo mediante errores que se suponen aditivos, incorrelados, homocedásticos y normalmente distribuidos. La distribución conjunta del vector de datos, dados los parámetros de suavizado y las condiciones iniciales, resulta ser normal multivariante. Comprobamos que el método de Holt-Winters puede formularse como un modelo lineal heterocedástico, cuyos coeficientes dependen de las condiciones iniciales mientras que la matriz de covarianzas sólo depende de los parámetros de suavizado.

Esta formulación alternativa nos permite simplificar considerablemente los cálculos necesarios para la obtención conjunta de los estimadores máximo verosímiles de las condiciones iniciales y de los parámetros de suavizado. Además, facilita la obtención de intervalos de predicción más fiables, en el sentido de una mayor adecuación del contenido teórico de los intervalos al contenido empírico, lo que se comprueba mediante bancos de datos simulados.

2. Agradecimientos

Parcialmente subvencionado por el Ministerio de Educación y Ciencia, TIN2005-08404-C04-04.

3. Bibliografía

- [1] Bermúdez, J.D., Segura, J.V. and Vercher, E. (2006). Improving demand forecasting accuracy using non-linear programming software. *Journal of the Operational Research Society* 57, 94–100.
- [2] Bermúdez, J.D., Segura, J.V. and Vercher, E. (2006). A decision support system methodology for forecasting of time series based on Soft Computing. *Computational Statistics and Data Analysis*. (In press).
- [3] Hyndman, R.J., Koehler, A.B., Snyder, R.D. and Grose, S. (2002). A state space framework automatic forecasting using exponential smoothing. *International Journal of Forecasting* 18, 439-454.
- [4] Makridakis, S. and Hibon, M. (2000). The M3-competition: results, conclusions and implications. *Journal of Forecasting* 16, 451-476.
- [5] Ord, J. K., Koehler, A.B. and Snyder, R.D. (1997). Estimation and Prediction for a Class of Dynamic Nonlinear Statistical Models. *Journal of the American Statistical Association* 92, 1621-1629.
- [6] Yar, M. and Chatfield, C. (1990). Prediction Intervals for the Holt-Winters Forecasting Procedure. *International Journal of Forecasting* 6, 127-137.

Introducción al análisis de la covarianza con intervalos aleatorios

M. Montenegro¹, G. González-Rodríguez¹, A. Colubi¹, R. Coppi²

¹{mmontenegro, gil, colubi}@uniovi.es, Departamento de Estadística e I.O. y D.M., Universidad de Oviedo

²renato.coppi@uniroma1.it, Dipartimento di Statistica, Probabilità e Statistiche Applicate, Università degli Studi di Roma, “La Sapienza”

Abstract

En este trabajo se introduce un procedimiento asintótico, basado en técnicas bootstrap, para analizar el efecto de un factor sobre una variable respuesta con valores de intervalo que está relacionada (mediante un modelo lineal) con una variable explicativa cuyos valores son también intervalares. El modelo se expresa utilizando la aritmética habitual entre intervalos.

Claves: análisis de la covarianza, intervalos aleatorios, tests bootstrap.

1. Introducción

En Gil *et al.* (2006a) y en González-Rodríguez *et al.* (2006) se analiza el modelo de regresión lineal simple entre intervalos aleatorios, basado en la aritmética intervalar, desde un punto de vista inferencial. Por otro lado, en Gil *et al.* (2006b) se presenta un procedimiento de análisis de la varianza que puede particularizarse al caso en que los valores de las variables consideradas sean intervalos. En este trabajo se analiza el correspondiente problema del análisis de la covarianza. Los estudios previos sirven de apoyo para introducir un procedimiento asintótico que es la base para contrastar el efecto de los distintos elementos en la variable respuesta por técnicas bootstrap.

2. Preliminares

Sea $\mathcal{K}_c(\mathbb{R})$ la clase de intervalos compactos no vacíos de \mathbb{R} dotado con la estructura semilineal inducida por la suma de Minkowski y el producto por escalares. Si existe la diferencia de Hukuhara entre $A, B \in \mathcal{K}_c(\mathbb{R})$ se denotará por $A -_H B$. Dado un espacio de probabilidad (Ω, \mathcal{A}, P) , una aplicación $X : \Omega \rightarrow \mathcal{K}_c(\mathbb{R})$ es un *intervalo aleatorio*, si es $\mathcal{A}|\mathcal{B}_{d_H}$ -medible, (siendo d_H la métrica de Hausdorff). Nótese que X queda caracterizado por $(\text{mid } X, \text{spr } X)$ donde $\text{mid } A = (\sup A + \inf A)/2$ y $\text{spr } A = (\sup A - \inf A)/2$ denotan el punto medio y el radio del intervalo $A \in \mathcal{K}_c(\mathbb{R})$, respectivamente. Si X es un *intervalo aleatorio* con $E(\max\{\inf X, \sup X\}; P) < \infty$, el valor esperado de Kudō-Aumann, $E^A[X; P]$, coincide con $[E(\inf X; P), E(\sup X; P)]$. Como en los estudios previos se utilizará la métrica d_W (donde W una medida de probabilidad simétrica no degenerada en $([0, 1], \beta_{[0,1]})$), que se define como $d_W(A, B) =$

$\sqrt{\int_{[0,1]} [f_A(\lambda) - f_B(\lambda)]^2 dW(\lambda)}$ con $f_A(\lambda) = \lambda \sup A + (1-\lambda) \inf A$, para todo $A, B \in \mathcal{K}_c(\mathbb{R})$.

3. Análisis de la covarianza

En cada una de las poblaciones determinada por un factor que actúa en k niveles se obtiene una m.a.s. $\{(X_{ij}, Y_{ij})\}_{j=1}^n$ de un intervalo aleatorio bidimensional (X_i, Y_i) ($i \in \{1, \dots, k\}$). El modelo lineal que consideramos en este trabajo se puede escribir como $Y_{ij} = A_i + \beta X_{ij} + \epsilon_{ij}$ donde $A_i \in \mathcal{K}_c(\mathbb{R})$ y los errores aleatorios ϵ_{ij} , con valores de intervalo, tienen valor esperado $B \in \mathcal{K}_c(\mathbb{R})$ para todo $i = 1 \dots k$ y $j = 1 \dots n$. Si $\beta = 0$, el problema se reduce a un análisis de la varianza (ver Gil *et al.*, 2006a). Para contrastar si $\beta = 0$ se propone utilizar la información conjunta de las k muestras, basando el estadístico en las covarianzas medias entre las variables mid y spr y adaptando el método bootstrap en Gil *et al.* (2006b). Si $\beta \neq 0$ se debería eliminar la influencia de la covariable para determinar posteriormente si las poblaciones producen la misma respuesta media o no. En este trabajo proponemos utilizar el estadístico básico del ANOVA a partir de los residuos estimados

$$\frac{\frac{1}{k} \sum_{i=1}^k d_W^2 (\bar{Y}_{i \cdot} - H \hat{\beta} \bar{X}_{i \cdot}, \bar{Y}_{..} - H \hat{\beta} \bar{X}_{..})}{\frac{1}{k} \sum_{i=1}^k \frac{1}{n} \sum_{j=1}^n d_W^2 (Y_{ij} - H \hat{\beta} X_{ij}, \bar{Y}_{i \cdot} - H \hat{\beta} \bar{X}_{i \cdot})}$$

donde $\hat{\beta}$ es una estimación de β que se puede obtener a partir de las ideas en González-Rodríguez *et al.* (2006). Los análisis de la covarianza habituales evitan construir los estadísticos directamente a partir de los residuos estimados, porque se introducen grados de dependencia que dificultan la determinación de la distribución del estadístico, aunque esta opción es bastante intuitiva y sencilla de llevar a la práctica. Las complicaciones que conlleva la aritmética intervalar hacen que la opción sugerida parezca un punto de partida razonable.

4. Bibliografía

- [1] Gil, M.A., Montenegro, M., González-Rodríguez, G., Colubi, A., Casals, M.R. (2006a). Bootstrap approach to the Multi-Sample Test of Means with Imprecise Data. *Comp. Stat. Data Anal.* (to appear)
- [2] Gil, M.A., González-Rodríguez, G., Colubi, A., Montenegro, M. (2006b). Testing Linear Independence in Linear Models with Interval-Valued Data. *Comp. Stat. Data Anal.* (to appear)
- [3] González-Rodríguez, G., Colubi, A., Coppi, R., Giordani, P. (2006) On the estimation of linear models with interval-valued data. COMPSTAT 2006.

La envolvente convexa esperada

I. Cascos Fernández

Departamento de Estadística, Universidad Carlos III de Madrid,
e-mail: ignacio.cascos@uc3m.es

Abstract

Construimos una familia de regiones centrales para la distribución de probabilidad de un vector aleatorio con primer momento finito.

Claves: envolvente convexa, esperanza de Aumann, función de profundidad.

AMS: 62H05, 60D05

1. Introducción

Muchas de las técnicas de análisis de datos univariantes se basan en la relación de orden *menor o igual* existente sobre la recta real. En el espacio Euclídeo multidimensional, no existe una relación de orden natural y esto nos plantea problemas a la hora de trasladar algunas técnicas unidimensionales al contexto multivariante. Una solución es ordenar los puntos según su centralidad respecto de una distribución de probabilidad. Las *funciones de profundidad* nos proporcionan ‘grados de centralidad’ y a los conjuntos de puntos con un cierto ‘grado de centralidad’ (conjuntos de nivel de las funciones de profundidad) los llamaremos *regiones centrales*. Zuo y Serfling [3, 4] y Mosler [2] proponen varias funciones de profundidad y regiones centrales.

En este trabajo se utilizarán nociones relativas a los conjuntos aleatorios, en concreto la *esperanza de Aumann o de las selecciones*, ver Molchanov [1].

2. Regiones centrales de la envolvente convexa esperada

Estudiaremos la sucesión de conjuntos aleatorios cuyo n -ésimo elemento es la envolvente convexa de n copias independientes de un vector aleatorio con primer momento finito. La sucesión de sus esperanzas de Aumann es creciente (se trata, por tanto, de una familia anidada) y sus elementos son conjuntos compactos y convexos, además de satisfacer la propiedad de equivarianza afín. Serán, por tanto, una familia de *regiones centrales* para la distribución del vector aleatorio, ver Zuo y Serfling [4].

A partir de la sucesión de conjuntos anterior, podemos cuantificar el grado de centralidad o *profundidad* de un punto respecto de la distribución del vector aleatorio. Ésta vendrá dada por el inverso del número de copias independientes del vector aleatorio necesarias para que el punto pertenezca a su esperanza. Si, independientemente del número de copias que consideremos, el punto nunca

pertenece a dicha esperanza, la profundidad será 0. Esta aplicación cumple las propiedades de *equivarianza afín, maximalidad en la esperanza del vector aleatorio, cuasiconcavidad y se anula en el infinito* que la convierten en una *función de profundidad* según la definición de Zuo y Serfling [3].

3. Aplicaciones

Describiremos brevemente dos de las aplicaciones de la familia de regiones centrales anterior.

Por un lado, podemos utilizarla para comparar dos distribuciones de probabilidad multivariantes según su variabilidad. Así, la relación de contenido de las regiones construidas a partir de n copias independientes de un vector aleatorio dentro de las que están construidas partir de n copias independientes de otro, para todo n , da lugar a un *orden estocástico* más débil que el orden lineal convexo.

Por otra parte, podemos utilizar estas regiones para cuantificar la dispersión asociada a un vector aleatorio. Si nos restringimos a $n = 2$ copias de un vector aleatorio, la esperanza de un segmento cuyos extremos son copias independientes de un vector aleatorio es un *zonoide*. En Mosler [2] se describe la fórmula del volumen de un zonoide. Aplicando esta fórmula a la envolvente convexa esperada de dos copias independientes de un vector aleatorio, obtenemos un *índice de Gini multidimensional*.

4. Agradecimientos

Se agradece la ayuda económica recibida de los proyectos del Ministerio de Ciencia y Tecnología MCT-BFM2002-01057, MCT-BFM2002-03263 y MTM2005-02254.

5. Bibliografía

- [1] Molchanov, I. (2005). *Theory of Random Sets*. Springer-Verlag, London.
- [2] Mosler, K. (2002). *Multivariate dispersion, central regions and depth. The lift zonoid approach*. LNS **165**. Springer-Verlag, Berlin.
- [3] Zuo, Y. and Serfling, R. (2000). General notions of statistical depth function. *Ann. Statist.*, **28**, 461–482.
- [4] Zuo, Y. and Serfling, R. (2000). Structural properties and convergence results for contours of sample statistical depth functions. *Ann. Statist.*, **28**, 483–499.

La Inferencia Estadística en la PAU

M. Candelaria Espinel¹, Carmen Ramos², Rosa Ramos³

¹mespinel@ull.es, Dpto. de Análisis Matemático, Universidad de La Laguna.

²ceramos@ull.es, Dpto. de Estadística, I.O. y Comp., Univ. de La Laguna.

³rramos@mat.ucm.es. Dpto. de Estadística e I.O. Universidad Complutense de Madrid.

Abstract

En este trabajo se realiza un análisis del contenido sobre Inferencia Estadística en la PAU. El objetivo es detectar los errores más significativos con el fin de prestar una mayor atención a tales dificultades en el proceso de enseñanza aprendizaje.

Claves: Didáctica. Intervalos de Confianza. Contrastes de Hipótesis.

AMS: 97D70

1. Introducción

A partir de una muestra de exámenes de las Pruebas de Acceso a la Universidad (PAU) del Distrito de Canarias, en la especialidad de Matemáticas de Ciencias Sociales, hemos analizado las técnicas usadas por los alumnos para resolver los problemas de Inferencia Estadística, y los fallos más comunes que éstos suelen cometer. En primer lugar se recoge la estructura de la prueba y las preferencias del alumnado ante los ejercicios planteados. En este sentido se ha visto que los alumnos optan mayoritariamente por las preguntas de Inferencia Estadística, y más concretamente, por las de Contrastes de Hipótesis. Además, se presenta un análisis cuantitativo sobre las puntuaciones de los ejercicios, donde se refleja que la preguntas sobre Contrastes de Hipótesis, Intervalos de Confianza y Probabilidad tienen una puntuación más alta. De los datos se extrae que las técnicas de solución basadas en representaciones gráficas parece que facilitan la comprensión del problema y conducen a una solución adecuada del ejercicio. En el siguiente apartado se analiza de forma cualitativa una pregunta referida a Contrastes de Hipótesis. Aquí, nuestro interés se centra en llevar a cabo un conteo sistemático de los errores que encontramos al revisar la prueba. De esta forma el profesor de secundaria puede prestar mayor atención a los aspectos que más dificultad suponen al alumno y mejorar el proceso de enseñanza-aprendizaje de la Estadística, mediante técnicas alternativas que usen visualizaciones gráficas.

2. Estructura de la Prueba

La estructura de la prueba para la asignatura de Matemáticas en Ciencias Sociales se plantea con dos opciones excluyentes: Prueba A y Prueba B. Cada una

de estas pruebas consta de cinco ejercicios. El alumno debe elegir una de las opciones y dentro de ésta contestar a cuatro ejercicios. La prueba A se compone de dos ejercicios de Estadística (Intervalos y Contrastes), dos de Análisis (Funciones y Máximos) y una de Ecuaciones. La prueba B tiene tres ejercicios de Estadística (Probabilidades, Intervalos y Contrastes) una de Funciones y una de Programación Lineal. La muestra que hemos recogido consta de 272 exámenes de la Convocatoria de Junio de 2005 de la Universidad de La Laguna. De esta muestra 113 alumnos eligieron la prueba A y 159 la B. Nuestro estudio se centra en la opción B, en la que se refleja una clara tendencia por parte del alumnado a elegir las preguntas con contenidos estadísticos, siendo los porcentajes de elección superiores al 90 %. Mientras que la menos elegida es la pregunta sobre el estudio de funciones, con un 20 %. Señalar que las puntuaciones más altas se corresponden también con la preguntas de Estadísticas y las más bajas con las de Funciones.

En un análisis cualitativo de la pregunta correspondiente a los contrastes de hipótesis se observa que los alumnos contestan siguiendo la pauta de los libros de texto (Ramos [2]). Aproximadamente un 50 % de los alumnos obtienen la máxima nota en esta pregunta, sin embargo se contemplan algunas dificultades que se repiten de forma reiterada, como son: No distinguen si el contraste es unilateral o bilateral; Confunden la media muestral con el valor hipotético del parámetro a contrastar; Mal uso de las tablas, debido a una identificación equivocada del nivel de significación α , y la probabilidad que deja el punto crítico a la derecha; Mala interpretación de los resultados obtenidos, con frecuencia rechazan la hipótesis cuando habría que aceptarla.

Algunas de estas dificultades en el aprendizaje ya han sido señaladas por otros autores (Vallecillos [3]). El problema de las tablas, quizás se solventa con el uso de las nuevas tecnologías (applets) y parece razonable que desde la enseñanza se potencien los métodos gráficos en este bloque del currículo.

3. Bibliografía

- [1] Batanero C. (2000). Controversies around the role of statistical tests in experimental research. *Mathematical Thinking and Learning* 2(1-2), 75-98.
- [2] Ramos R. y Espinel C. (2003) Estimación Estadística: Algunas dificultades observadas en la PAU. *Actas de las XI Jornadas sobre Aprendizaje y Enseñanza de las Matemáticas*.
- [3] Vallecillos A. y Batanero C. (1997) Conceptos activados en el Contraste de Hipótesis Estadísticas y su comprensión por estudiantes universitarios. *Recherches en didactique des mathématiques* 17(1), 29-48.

La k -agrupabilidad: Un nuevo método para distinguir conjuntos agrupados y dispersos

Agusti Solanas

agusti.solanas@urv.cat, Departament d'Engenieria Informàtica i Matemàtiques (DEIM), Universitat Rovira i Virgili.

Abstract

El presente artículo introduce un nuevo concepto para clasificar conjuntos numéricos en función de la capacidad de sus elementos para ser agrupados: la k -agrupabilidad. Abordamos el problema de determinar el grado de agrupabilidad de un conjunto numérico.

Claves: Agrupación, Minería de datos

1. Introducción

Métodos estadísticos clásicos basados en el análisis de agrupación [3] han sido utilizados en áreas tan diversas como la inteligencia artificial [1] (*e.g.* Extracción de conocimiento [4]) o la protección de la privacidad individual (*e.g.* microagregación [2][5]). Los estudios dedicados al campo de la agrupación nos han legado una gran cantidad de técnicas para agrupar los elementos de un conjunto para, principalmente, clasificarlos. En general, estos métodos no disponen de información previa sobre la disposición natural de los elementos en subconjuntos puesto que, hasta el momento, tal información no podía obtenerse de forma sencilla. Este artículo presenta una técnica de fácil aplicación para determinar si un conjunto de elementos numéricos contiene subconjuntos agrupados de forma natural (*i.e.* es agrupado) o no (*i.e.* es disperso), la k -agrupabilidad.

2. Determinación de la k -agrupabilidad de un conjunto

Dada una función de semejanza F_s , la k -agrupabilidad es la capacidad natural de un conjunto numérico de contener subconjuntos de cardinalidad k .

Cálculo de la agrupabilidad para una k dada: Para calcular la agrupabilidad de un conjunto C se calcula la matriz de semejanzas M_s usando F_s . Seguidamente, para cada elemento $c_i \in C$ se seleccionan sus $k - 1$ elementos más semejantes y se construye un subconjunto $G_{c_i} \subset G$ de cardinalidad k , obteniendo al fin del proceso n_G subconjuntos de G de cardinalidad k . Definimos dos cálculos sobre G_{c_i} : a) Calcular su error cuadrático medio (ECM). b) Determinar el valor de la distancia máxima del elemento $c_i \in G_{c_i}$ a cualquier elemento $c_j \in G_{c_i}$, $v_{G_{c_i}}$. Tras realizar alguno de estos cálculos, se procede a obtener el valor medio de los mismos para una k dada. Nótese que el valor de ECM_k i v_k crece con k , lo cual es una condición para aplicar el método que se propone, puesto que éste se basa en el análisis de dicho crecimiento.

Estudio del crecimiento de v_k : Los cálculos anteriores se repiten para varias k en el entorno de la cardinalidad a estudiar. Para determinar la k -agrupabilidad de un conjunto, se aproxima una función logarítmica y una lineal

y se observa cual de las dos aproximaciones se adapta mejor a los valores obtenidos en los anteriores cálculos mediante el valor de R^2 . Podemos encontrarnos en dos situaciones: a) La aproximación lineal es mejor: El conjunto es candidato a contener subconjuntos agrupados para una cierta k . Necesitamos un posterior estudio que lo confirme y determine los valores de k para los que los subconjuntos son agrupables. b) La aproximación logarítmica es mejor: El crecimiento de v_k es cada vez más lento, con lo que se deduce la no existencia de subconjuntos agrupados, pues para que éstos existan es necesario un crecimiento que no disminuya con k . En el primer caso, proseguimos con la determinación de los k máximos. En el segundo, concluimos el análisis considerando que el conjunto es disperso. Para estudiar con más detalle el crecimiento de los valores v_k , calculamos las diferencias entre las medias de distancias máximas consecutivas dv_k . La existencia de subconjuntos de una cierta cardinalidad k se traduce en un crecimiento abrupto de la media de distancia máxima y, por consiguiente, en un valor máximo de dv_k , para la k que iguala a la cardinalidad del subconjunto.

3. Conclusiones

En este resumen hemos propuesto un nuevo método para la determinación de la agrupabilidad de un conjunto de datos. El método permite distinguir conjuntos dispersos de agrupados naturalmente. Además el método propuesto permite determinar la cardinalidad de los subconjuntos agrupados de forma natural. Los resultados experimentales presentes en el artículo completo muestran la aplicabilidad y utilidad de la nueva medida.¹

4. Bibliografía

- [1] Pavel Berkhin. Survey of clustering data mining techniques. Technical report, Accrue Software, San Jose, CA, 2002.
- [2] J. Domingo-Ferrer and J. M. Mateo-Sanz. Practical data-oriented microaggregation for statistical disclosure control. *IEEE Transactions on Knowledge and Data Engineering*, 14(1):189–201, 2002.
- [3] John A. Hartigan. *Clustering Algorithms*. John Wiley & Sons, Inc., New York, NY, USA, 1975.
- [4] Tomasz Imielinski and Heikki Mannila. A database perspective on knowledge discovery. *Commun. ACM*, 39(11):58–64, 1996.
- [5] Agusti Solanas, Antoni Martínez-Balleste, Josep M. Mateo-Sanz, and Josep Domingo-Ferrer. Multivariate microaggregation based on a genetic algorithm. *manuscript*, 2006.

¹ **Agradecimientos** Este trabajo está parcialmente financiado por la Generalitat de Catalunya con la ayuda 2005 SGR 00446, y por el Ministerio de Educación y Ciencia mediante el proyecto SEG2004-04352-C04-01 "PROPRIETAS". El autor agradece los amables comentarios y sugerencias recibidas por parte del Prof. Dr. Josep Domingo-Ferrer, el Dr. Josep M. Mateo-Sanz y el Dr. Blas Herrera.

Las regiones europeas: una perspectiva demográfica

A. Arroyo¹, S. Bermúdez¹, R. Blanquero¹, A. Pozo¹

¹{aarroyo,silvber,rblanquero,apozo}@us.es

Departamento de Estadística e Investigación Operativa
Universidad de Sevilla

Abstract

En este trabajo se realiza un análisis comparativo de las regiones europeas (nivel 2 de la clasificación NUTS) desde un punto de vista demográfico. A partir de un conjunto de variables referidas a población, mortalidad y fecundidad, se establecen afinidades entre las distintas regiones mediante el empleo de técnicas de clasificación no supervisada.

Claves: Demografía, Unión Europea, NUTS.

AMS: 62P25

1. Introducción

La Unión Europea, como bien se pone de manifiesto en su lema *Unida en la diversidad*, aglutina una familia de Estados de muy diversas características, con el compromiso común de trabajar por la paz y la prosperidad. Es precisamente esta diversidad el origen de desequilibrios en su seno, que han sido objeto de diferentes estudios a lo largo de su historia, en especial tras la ampliación llevada a cabo en 2004, y que continúan siendo de interés ante la inminente adhesión de nuevos miembros. No obstante, los estudios realizados hasta la fecha han estado encaminados a proporcionar una panorámica de las regiones europeas desde el punto de vista económico, poniendo de relieve las desigualdades existentes ([1]), mientras que los aspectos puramente demográficos no han sido tratados.

En este trabajo se realiza un análisis de las características demográficas de los Estados miembros de la Unión Europea, que pretende poner de manifiesto las diferencias y similitudes existentes entre ellos y proporcionar un mapa de la Unión bajo una perspectiva demográfica. El ámbito territorial considerado ha sido el nivel 2 de la clasificación jerárquica NUTS (Nomenclatura de las Unidades Territoriales Estadísticas) de la Unión Europea, correspondiente al nivel regional principal de los estados miembros; en el caso de España, está constituido por las 17 comunidades autónomas junto a las ciudades autónomas de Ceuta y Melilla.

Para la realización del estudio se ha hecho uso, en primera instancia, de la información proporcionada por Eurostat en su base de datos New Cronos, accesible en Internet a través de la página <http://europa.eu.int/comm/eurostat>.

De dicha fuente se han seleccionado datos absolutos e indicadores referidos a población, fecundidad y mortalidad, a partir de los cuales se han elaborado otros indicadores considerados de interés para el estudio. El análisis de las variables seleccionadas y la posterior aplicación de técnicas de clasificación no supervisada, ha permitido vislumbrar aquellas agrupaciones de regiones que, desde un punto de vista demográfico, presentan mayor grado de similaridad.

2. Agradecimientos

Este trabajo ha sido parcialmente financiado por los proyectos MTM2005-09362-C03-01 del Ministerio de Educación y Ciencia y FQM-329 del Plan Andaluz de Investigación.

3. Bibliografía

- [1] Auñóles Martín, J., Fernández Cuevas, M.C., Manzanera Díaz, E. (2002). La ampliación de la Unión Europea hacia el Este: posibles efectos regionales. *Papeles de Economía Española* 93, 215-228.
- [2] Reglamento (CE) nº 1059/2003 del Parlamento Europeo y del Consejo, de 26 de Mayo de 2003, por el que se establece una nomenclatura común de unidades territoriales estadísticas (NUTS).
- [3] *Regions: Statistical yearbook 2005*. Office for Official Publications of the European Communities (2005).

Mapa euclídeo de la Red Internet basado en medidas de retardo

Francisco Javier Delgado¹, Purificación Galindo², José Luis Vicente-Villardón³

¹j.delgado@ieee.org, ²pgalindo@usal.es, ³villardon@usal.es

Departamento de Estadística, Universidad de Salamanca

Abstract

El retardo en la transmisión de información en una red como Internet juega un importante papel en la caracterización de su calidad de servicio. La matriz de retardo, que es cuadrada y no necesariamente simétrica, ha sido analizada mediante métodos multivariantes (MDS métrico, ACP) para obtener un “mapa” de la red Internet que nos permita estimar las prestaciones de las diferentes alternativas disponibles. En este trabajo se analizarán los resultados obtenidos con las técnicas utilizadas en la literatura para poner de manifiesto sus ventajas y limitaciones y se aplicarán otras técnicas de análisis de datos que no han sido utilizadas en este contexto (BI PLOT, *unfolding*) y que complementan y/o mejoran los análisis tradicionales.

Palabras Clave: Internet, BI PLOT, Unfolding

AMS: 62H25, 94A99, 94C99

1. Introducción

El retardo se define como el tiempo que tarda la energía que transporta la información a través de un canal de comunicación en trasladarse de un extremo del canal a otro. En una época en la que la velocidad de transmisión de información cada vez es más elevada y las telecomunicaciones alcanzan todo el globo, de la mano de la red Internet, el retardo en la transmisión de información se ha convertido en un parámetro crítico de las redes para la prestación de determinados servicios de telecomunicación [1].

Ng y Zhang [2] proponen que el retardo entre dos nodos en la red Internet puede ser modelada, con buena precisión, como una distancia geométrica en un espacio euclídeo de dimensión mayor o igual que tres.

Se han aplicado diversos métodos estadísticos multivariantes para obtener un “mapa local” de Internet basado en el retardo entre nodos de la red. Las posibles aplicaciones de un “mapa” de la red Internet son múltiples, desde su utilización como “tomografía” de la red para la gestión de la misma, hasta la elección del servidor más idóneo para la prestación de servicios sensibles al retardo. Shavitt y Tankel [3] proponen alternativas a las técnicas de escalado

multidimensional (MDS métrico) y Lim *et al.* [4], plantean la utilización del Análisis de Componentes Principales (PCA) con similares planteamientos.

2. Objetivos

El objetivo del presente trabajo es doble: por un lado, la obtención de una representación de un conjunto de equipos conectados a la red Internet a partir de la inspección BI PLOT del retardo de la comunicación entre ellos que nos permita seleccionar el servidor desde el que recibir un determinado servicio. Por otro lado se pretende comparar los resultados obtenidos con un ACP (técnica utilizada en los trabajos más recientes) y con un MDS, y los obtenidos con un BI PLOT, Análisis de Coordenadas Principales y un Análisis *Unfolding*.

3. Metodología

El análisis que se propone constaría de dos fases, una inicial basada en matrices de retardo ficticias basadas en las características que les son propias y con las que se experimentaría mediante simulación para analizar comparativamente los resultados obtenidos. Esta simulaciones tratarían de comparar los métodos en función de las dimensiones de la matriz y del modelo obtenido, entre otras cuestiones relevantes para los objetivos propuestos. De esta manera se validaría inicialmente el método frente a los existentes en la bibliografía.

Posteriormente los resultados serán contrastados con matrices de retardo obtenidas de mediciones sobre redes reales.

4. Bibliografía

- [1] Kleinrock, L. (1992) The Latency/Bandwidth Tradeoff in Gigabit Networks. *IEEE Communications Magazine*, 36-40.
- [2] Ng, T.S.E. and Zhang, H. (2001) Toward Global Network Positioning. *ACM SIGCOMM Internet Measurements Workshop 2001*.
- [3] Shavitt, Y., Tanel, T. (2004). Big-bang simulation for embedding network distances in Euclidean space. *IEEE/ACM Transactions on Networking*, 993-1006.
- [4] Lim, H., Hou, J.C., and Choi, C. H. (2005). Constructing Internet Coordinate System Based on Delay Measurement. *IEEE/ACM Transactions on Networking*, 513-525.

Matriz de Dispersión Inducida

Andrés Jiménez Jiménez¹, José María Gutiérrez Pérez², Francisco Alvarez Gonzalez³

¹andres.jimenez@uca.es, CITI., Univ. Cádiz

²josema.gutierrez@uca.es, Dpto. de Estadística e IO, Univ. Cádiz

³francisco.alvarez@uca.es, Dpto. de Estadística e IO, Univ. Cádiz

Abstract

El presente estudio plantea una generalización de la matriz de covarianza asociada a muestras supervisadas que denominamos Matriz de Dispersión Inducida. La descomposición espectral de dicha matriz es local a cada punto del espacio y encuadra el Análisis de Componentes Principales clásico dentro del caso particular de existencia de un solo grupo.

Palabras Clave: Dispersión Inducida, Componentes de Influencia

AMS: 62H25, 62H30.

1. Dispersión Inducida

Sea una muestra $m = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\} \in I\!\!R^d$ cuyos elementos pertenecen a las clases $\{\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_c\}$ y $t(\cdot)$ la función indicador que asocia a cada elemento $\mathbf{x} \in m$ la clase \mathcal{C}_i a la que pertenece. Para cuantificar la influencia de cada grupo sobre \mathbf{x} se define el concepto de Dispersión Inducida.

DEFINICIÓN: Se denomina Función de Dispersión Inducida por la muestra m a una función continua tal que a cada $\mathbf{x} \in I\!\!R^d$ le asocia una matriz $\Delta_{\mathbf{x}}$ simétrica y semidefinida positiva verificando que si $\mathbf{x} \rightarrow \boldsymbol{\mu}_i$ entonces $\Delta_{\mathbf{x}} \rightarrow \Sigma_i$, siendo $\boldsymbol{\mu}_i = E[\mathbf{X}/\mathcal{C}_i]$ y $\Sigma_i = \text{cov}(\mathbf{X}/\mathcal{C}_i)$.

En particular usando una combinación convexa de los Σ_i , puede construirse una función de dispersión inducida:

$$\Delta_{\mathbf{x}} = \sum_{i=1}^c \alpha_i^{\mathbf{x}} \Sigma_i \quad \text{con} \quad \sum_{i=1}^c \alpha_i^{\mathbf{x}} = 1, \quad 0 \leq \alpha_i^{\mathbf{x}} \leq 1 \quad \text{y} \quad \Sigma_i = \text{cov}(\mathbf{X}/\mathcal{C}_i),$$

La matriz $\Delta_{\mathbf{x}}$, es una matriz virtual de dispersión, es simétrica y semidefinida positiva. A pesar de que $\Delta_{\mathbf{x}}$ no representa la dispersión de la muestra sí recoge la influencia que ejercen las clases sobre cada punto del espacio. Se plantearán distintas formulaciones de los $\alpha_i^{\mathbf{x}}$ consecuentes con las restricciones anteriores.

DEFINICIÓN: Se denomina Transformación por Componentes de Influencia en \mathbf{x} a la proyección $\mathbf{Y}_{\mathbf{x}} = \mathbf{X} \mathbf{U}_{\mathbf{x}}$, donde $\mathbf{U}_{\mathbf{x}} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_d]_{\mathbf{x}}$ es la matriz de autovectores de $\Delta_{\mathbf{x}}$.

Como éstas varían de forma continua de unos centroides a otros (por la continuidad de las transformaciones efectuadas) es de esperar que en cualquier punto \mathbf{x} las Componentes de Influencia expliquen en parte la variabilidad de la muestra y en parte la influencia de los grupos. Se probará que las Componentes Principales son un caso particular de las CI.

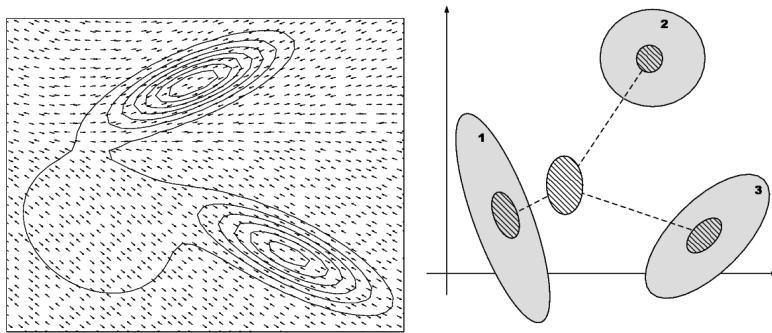


Figura 1: *Primer autovector de $\Delta_{\mathbf{x}}$ de una m.a.s. de una mixtura de tres distribuciones normales y formas adaptativas de las bolas en el método k-NN.*

1.1. Comportamiento de las CI

Se abordará el estudio de las CI en aquellos puntos del espacio donde $\Delta_{\mathbf{x}}$ manifieste alguna característica distintiva: a) zonas de máxima y mínima influencia de las clases, b) zonas de igual influencia de las clases, y c) zonas donde los autovectores de $\Delta_{\mathbf{x}}$ poseen capacidades explicativas análogas.

2. Variante del método k-NN empleando la métrica $\Delta_{\mathbf{x}}$

Se presentará una variante del método de clasificación de los *k vecinos mas cercanos* (*k*-NN) que denominaremos *k*-NN(Δ). La idea consiste en considerar la distancia asociada a la dispersión inducida, $d_{\mathbf{x}}(\mathbf{u}, \mathbf{x}) = (\mathbf{u} - \mathbf{x})\Delta_{\mathbf{x}}^{-1}(\mathbf{u} - \mathbf{x})^T$, manteniendo inalterado el resto del algoritmo. Como puede observarse en la figura 1, con esta variante la métrica no es constante y la forma del entorno de un punto depende de la proximidad a cada clase.

3. Bibliografía

- [1] Schott, J.R. (1997). Matrix Analysis for Statistics. Wiley & Sons.
- [2] Marchette D. y Poston W. (1999). Local Dimensionality Reduction using Normal Mixtures. Computational Statistics, 14, 469-489, 1999.

Maximal correlation between minima

A. Castaño Martínez¹, F. López Blázquez², B. Salamanca Miño²

¹antonia.castano@uca.es, Departamento de Estadística e Investigación Operativa, Universidad de Cádiz

²{lopez,bsm}@us.es, Departamento de Estadística e Investigación Operativa, Universidad de Sevilla

Abstract

En distribuciones absolutamente continuas presentamos una cota superior para la correlación máxima entre dos estadísticos ordenados mínimos obtenidos a partir de distintos tamaños muestrales. La cota se alcanza en un caso límite. Además dicha cota será válida para la razón de correlación.

Keywords: Order statistics, Correlation ratio, Laguerre polynomials

1. Introduction

The i th order statistic of a sample Z_1, \dots, Z_r of size r is denoted as $Z_{i:r}$. Let $X_1, \dots, X_n, X_{n+1}, \dots, X_{n+m}$ be iid random variables from an absolutely continuous distribution F with density f . The aim of this paper is to provide an upper bound for the maximal correlation between $X_{1:n}$ and $X_{1:n+m}$. The maximal correlation of the pair $(X_{1:n}, X_{1:n+m})$ is

$$\rho^* = \sup_{g,h} \rho(g(X_{1:n}), h(X_{1:n+m})),$$

where ρ denotes the correlation coefficient and g and h belong to the family of measurable functions such that $Eg^2(X_{1:n}) < \infty$ and $Eh^2(X_{1:n+m}) < \infty$ respectively.

Similar problems have been treated in the literature. The pioneering work in this area is that of Terrel (1983) who showed that for iid samples of size two from a continuos parent

$$\rho(X_{1:2}, X_{2:2}) \leq \frac{1}{2}. \quad (2.40)$$

Some years later, Szekely and Mori (1985) showed that for iid samples

$$\rho(X_{i:n}, X_{j:n}) \leq \left(\frac{i(n-j+1)}{j(n-i+1)} \right)^{1/2}, \quad 1 \leq i < j \leq n. \quad (2.41)$$

More recently, López-Blázquez and Castaño-Martínez (2006) have shown that the right hand side of (2.41) is also an upper bound for the correlation ratio

when the sampling is performed without replacement, namely,

$$R_{j|i}^2 = \frac{Var(E[X_{j:n} | X_{i:n}])}{Var(X_{i:n})} \leq \left(\frac{i(n-j+1)}{j(n-i+1)} \right)^{1/2}, \quad 1 \leq i < j \leq n. \quad (2.42)$$

The bounds given in (2.40), (2.41) and (2.42) are also valid for the maximal correlation in their respective cases. Some common features of the three above mentioned bounds are that they are attainable, the sample size is fixed and the fact that they are obtained using properties of certain families of orthogonal polynomials (Chebychev, Jacobi and Hahn polynomials respectively).

In this work we provide an upper bound for the maximal correlation between $X_{1:n}$ and $X_{1:n+m}$, so the sample size is not fixed. In this case the bound is not attainable, but in some sense the bound is sharp because it is attained in the limit. The proof of our result is based on properties of expansions of functions in terms of Laguerre polynomials.

2. Bibliography

- [1] López -Blázquez, F. (1998), A note on UMVU estimation in the transformed chi-square family, *Communications in Statistics, Theory and Methods*, 27, 2065-2073.
- [2] López -Blázquez, F., Castaño-Martínez, A. (2006). Upper and lower bounds for the correlation ratio of order statistics from a sample without replacement. *Journal of Statistical Planning and Inference*, 136, 43-52.
- [3] López -Blázquez, F., Salamanca-Miño, B. (1998), An upper bound for the correlation ratio of records, *Metrika*, 47, 165-174.
- [4] Nevzorov, V. B. (1992), A characterization of exponential distributions by correlations between records, *Math. Methods Statist.*, 1, 49-54.
- [5] Terrel, G. R. (1983), A characterization of rectangular distributions, *The Annals of Probability*, 11, 823-826.

ML solutions for a finite mixture discrete model

Javier Portela¹

¹jportela@estad.ucm.es, Dpto de Estadística e I. Operativa III, Universidad Complutense de Madrid

Abstract

The problem of mixtures of multinomial distributions is presented. A characterization of the space of maximum likelihood solutions is shown. This result is used through the EM algorithm to offer a new approach to the problem of selecting a suitable number of mixtures

Keywords: mixture models, multinomial, EM algorithm

1. Introduction

We study the case where data is supposed to come from a mixture of M multinomial distributions with dimension parameters K and n . That is,

$f_M(\mathbf{y}) = \pi_1 M n(n, \vec{p}_1, K) + \dots + \pi_M M n(n, \vec{p}_M, K)$ is the fitting mixture distribution. (1),

$$\text{with } \sum_{i=1}^M \pi_i = 1, \sum_{k=1}^K p_i^{(k)} = 1, \pi_i \geq 0, p_i^{(k)} \geq 0, \forall i = 1, \dots, M$$

The \mathbf{y} vector represents, for example, data from an individual that chooses between K alternatives, in n independent situations or time points. Data from N individuals is used to fit the mixture parameters through a Maximum Likelihood approach.

2. A characterization of the ML solutions

Definitions

Let $f_{M_0}(\mathbf{y}) = \pi_1 M n(n, \vec{p}_1, K) + \dots + \pi_{M_0} M n(n, \vec{p}_{M_0}, K)$ be the true mixture distribution. (2), with $\pi_i > 0, p_i^{(k)} > 0, \forall i = 1, \dots, M_0$.

Suppose $M > M_0$, that is, we are fitting a higher number of mixture components than the real number M_0 . Let $\psi_M = \{\pi_1, \vec{p}_1, \dots, \pi_M, \vec{p}_M\}$ be the $f_M(\mathbf{y})$ parameter space.

Let $\Omega_0 = \{\psi_M \mid f_M(\mathbf{y}) \stackrel{Id}{=} f_{M_0}(\mathbf{y})\}$ where $\stackrel{Id}{=}$ means functional identity. This functional identity only happens if $f_M(\mathbf{y})$ is reduced to $f_{M_0}(\mathbf{y})$ by these (not exclusive) means:

1. $\pi_i = 0$ for some $i = 1, \dots, M$

2. $p_i^{(k)} = 0$ for some $i = 1, \dots, M$, $k = 1, \dots, K$
3. $\vec{p}_i = \vec{p}_j$ for some i, j
4. Permutations of mixture components .

Theorem-.

Suppose $f_M(\mathbf{y})$ and $f_{M_0}(\mathbf{y})$ are mixtures as in (1) and (2), $M > M_0$.

Assume the following condition (C1):

There exist n_0 such that $\forall n \geq n_0$, $f_M(\mathbf{y}) = f_{M_0}(\mathbf{y}) \forall \mathbf{y} \in D_\mathbf{y} \iff f_M(\mathbf{y}) \stackrel{Id}{=} f_{M_0}(\mathbf{y})$.

Then, for mixtures of the form (1) and (2), with $n \geq n_0$, the Maximum Likelihood Estimator of ψ_M is consistent in Ω_0 , that is, $\lim_{N \rightarrow \infty} \widehat{\psi}_M$ is such that it has some components of the form presented in 1,2,3,4 with probability 1.

Corollary -. Under the hypothesis of Theorem 1, The EM algorithm solution $\widehat{\psi}_M^{EM}$ when fitting $f_M(\mathbf{y})$ to a true mixture $f_{M_0}(\mathbf{y})$, $M > M_0$, is such that $\lim_{N \rightarrow \infty} \widehat{\psi}_M^{EM} \in \Omega_0$.

3. Selecting the number of mixtures

When the number of mixtures component fitted is higher than the real number of components, the EM algorithm behaves creating groups of mixtures with vectors \vec{p}_k that are very close. A Monte Carlo Likelihood Ratio Test is created to test if two selected mixtures can be joint, leading to a model with one component left, without significantly affecting the likelihood. In this case the number of component of mixtures is decreased by one component. Otherwise it is increased by one component at a time.

Results show a good behaviour of the method, similar to the BIC method and slightly better to the AIC selecting method.

4. Bibliography

- [1] McLachlan, G., Peel, D. (2000) *Finite Mixture Models*. Wiley.
- [2] Titterington, D.(1990) Some recent research in the analysis of mixture distributions *Statistics*21, 619-641
- [3] Windham, M.P. and Cuttler, A. (1992).Information ratios for validating mixture analysis . *Journal of the American Statistical Association* 87,1188-1192

Modelización de datos de conteo multivariantes

J.M. Muñoz-Pichardo¹, J. García-Heras², M.D. Cubiles-Vega³, R. Pino-Mejías⁴

¹juanm@us.es, Departamento de Estadística e Investigación Operativa,
Universidad de Sevilla. ²heras@us.es, ³cubiles@us.es, ⁴rafaelp@us.es

Abstract

Este trabajo aborda el problema de la modelización de experiencias de datos de *conteo* multivariantes. Para ello, se define una distribución Poisson multivariante que admite correlaciones tanto positivas como negativas entre las componentes. Sobre ésta, se diseña un modelo lineal multivariante que puede ser útil en la modelización de experiencias basadas en el conteo de diversas especies.

Claves: Multivariate Poisson distributions, count data, conditional modelling

1. Descripción del trabajo

En diferentes áreas científicas (biomedicina, epidemiología, econometría, demografía, etc.), es de gran interés el estudio de experiencias que contienen conjuntos de datos de conteo multivariante. En la literatura científica se encuentran numerosos ejemplos de modelización de fenómenos discretos a través de la distribución univariante de Poisson o de algunas de sus extensiones. Mientras que el caso univariante ha sido ampliamente estudiado y aplicado, son escasos los trabajos que abordan su versión multivariante. Posiblemente, ello es debido a ciertas dificultades computacionales y, especialmente, al hecho de que las versiones multivariantes sólo contemplan correlaciones positivas entre las componentes.

En este trabajo se define una distribución Poisson multivariante basada en distribuciones condicionadas, extendiendo el caso bivariante propuesto por Berkhouit y Plug [1].

Se dice que el vector aleatorio p -dimensional $\underline{Y} = (Y_1, Y_2, \dots, Y_p)^t$ se distribuye según una distribución de Poisson log-lineal condicional (MLCP) si $Y_1 \sim P(\lambda_1)$ e $Y_j|Y_1, \dots, Y_{j-1} \sim P(\lambda_j(y_1 \dots y_{j-1}))$ para $j = 2, \dots, p$, tal que:

$$\begin{aligned} \log(\lambda_1) &= \eta_1, \\ \log(\lambda_j) &= \eta_j + \sum_{h=1}^{j-1} \alpha_{jh} y_h = \eta_j + \underline{\alpha}_j^t \underline{y}_{(j-1)} \quad j = 2, \dots, p. \end{aligned}$$

Se denotará por $MLCP_p(\underline{\eta}, \underline{\alpha}_2, \dots, \underline{\alpha}_p)$, donde $\underline{\eta} = (\eta_1, \dots, \eta_p)^t \in \mathbf{R}^p$, $\underline{\alpha}_j = (\alpha_{j1}, \dots, \alpha_{j(j-1)})^t \in \mathbf{R}^{j-1}$, para $j = 2, \dots, p$.

Se puede observar que esta distribución depende de $p(p + 1)/2$ parámetros reales que definen las combinaciones lineales que determinan los logaritmos de los parámetros de las distribuciones condicionadas. Además, la dependencia de la j^{th} variable Y_j de las primeras $(j - 1)$ variables se modeliza a través del parámetro de la distribución de Poisson, el cual es corregido por un factor multiplicativo que depende de estas variables:

$$\lambda_j = e^{\eta_j} f_j(Y_{(j-1)}) = e^{\eta_j} e^{\alpha_j^t Y_{(j-1)}}$$

Esta distribución multivariante puede ser usada para modelar experiencias descritas a través de datos de conteo multivariante en las que las variables están relacionadas entre sí con correlaciones positivas y/o negativas. En concreto, su aplicación al estudio del número de individuos de varias especies en cada unidad muestral puede permitir obtener conclusiones del tipo: posible relación de simbiosis entre dos especies (correlación positiva), posible relación de *competencia* entre dos especies (correlación negativa).

Posteriormente, se define un modelo lineal multivariante a través del cual poder explicar la dependencia estadística de un vector de variables respuesta siguiendo esta distribución, con un conjunto de covariantes o variables explicativas. Obviamente, su estudio lleva asociado diversos problemas:

- Estimación de parámetros, para lo cual se procederá a aplicar el método de máxima verosimilitud con los recursos computacionales necesarios.
- Estudio de la selección y adecuación del modelo.

2. Agradecimientos

Este trabajo está parcialmente financiado por el proyecto MTM2004-01433 del Ministerio de Educación y Ciencia y por el Grupo de Investigación FQM153 de la Junta de Andalucía.

3. Bibliografía

- [1] Berkhout, P. and Plug, E. (2004). A bivariate Poisson count data model using conditional probabilities. *Statistica neerlandica*, **58** (3), 349-364

Modelización temporal de los contaminantes registrados por la red automática de monitoreo en la ciudad de Valencia

A. Montesinos Guillot¹, C. Capilla Romá², S. Balasch Paris³

¹amontesi@eio.upv.es, ²ccapilla@eio.upv.es, ³sbalasch@eio.upv.es,

Departamento de Estadística e Investigación Operativa aplicadas y Calidad,
Universidad Politécnica de Valencia

Abstract

La red automática de monitoreo de la ciudad de Valencia registra datos de diversos contaminantes atmosféricos en cinco puntos de la ciudad. El presente trabajo realiza una modelización temporal de cada una de las series de contaminantes utilizando la metodología Box-Jenkins.

Palabras Clave: Contaminación atmosférica, series temporales

AMS: 62-07, 62M10

1. Introducción

La contaminación atmosférica es un problema que afecta a todos los grandes núcleos urbanos y que tiene graves consecuencias tanto para las personas como para el medio ambiente. Esto hace que en todas las grandes ciudades exista la necesidad de disponer de datos fiables que permitan evaluar las concentraciones de los contaminantes.

El objetivo del presente trabajo es analizar con modelos temporales, los datos de los distintos contaminantes registrados en cinco puntos de Valencia mediante la red automática de monitoreo del Ayuntamiento.

2. Material y Métodos

El Laboratorio Químico Municipal del Ayuntamiento de Valencia gestiona el funcionamiento de una red automática de toma de datos. Esta red cuenta con un total de cinco estaciones de medición situadas en distintos puntos de la ciudad que pueden ser críticos por su elevada densidad de tráfico. Las estaciones de monitoreo se encuentran situadas en las siguientes localizaciones: Av. Aragón, C/Linares, Nuevo Centro, Pista de Silla y Pza. de España.

Cada una de las estaciones registra los niveles de distintos contaminantes de acuerdo con la legislación vigente al respecto. El registro de datos de la red automática se realiza cada 15 minutos, siendo el valor registrado el promedio de las 180 medidas obtenidas cada 5 segundos dentro del intervalo citado de 15 minutos. Con esta frecuencia de muestreo, se detectaron porcentajes elevados de datos faltantes debidos a distintas causas. Por este motivo, en el primer

análisis temporal se agregaron las observaciones y se trabajó con promedios diarios.

Desde el año 1995 se dispone de datos diarios en las 5 estaciones. En el presente análisis se utilizan los datos disponibles desde el 1 de enero de 1995 hasta el 26 de febrero de 1997.

3. Resultados

Se aplicó un primer análisis descriptivo de los datos que permitió eliminar los datos anómalos. A continuación se realizó la modelización temporal de cada una de las series de contaminantes en cada una de las estaciones de esta red. Se ha observando en todas las situaciones que las desviaciones típicas están correlacionadas con las medias, por lo que ha sido necesario aplicar la transformación Box-Cox conveniente en cada caso para estabilizar la dispersión. El trabajo recoge los modelos ARIMA que mejor explican la variabilidad temporal en el periodo considerado de los contaminantes en cada estación.

4. Bibliografía

- [1] Box, G. E.P. y Cox, D.R. (1964). An analysis of transformations (with discussion). *Journal of the Royal Statistical Society*, 26, 211-252.
- [2] Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (1994), *Time Series Analysis: Forecasting and Control*, (3rd ed.). New Jersey: Prentice Hall.
- [3] Carroll , R.J., et al. (1997), Ozone Exposure and Population Density in Harris County, Texas". *Journal of The American Statistical Association*, Vol.92, 430, pp.392-415.
- [4] Guttorp, P., Meiring, W., and Sampson, P.D.(1994), A Space-Time Analysis of Ground-Level Ozone Data. *Environmetrics*, Vol.5, pp.241-254.
- [5] Little, R.J.A.; and Rubin, D.B., (1987), *Statistical Analysis with Missing Data*, New York: Wiley.
- [6] Montesinos, A., (2003), Statistical Analysis of Atmospheric Pollution in an Urban Area. ProQuest Information and Learning.

Modelización y optimización multiobjetivo de las condiciones de trabajo de un motor diesel

Andrea Conchado¹, José Jabaloyes¹, Jose Miguel Carot¹, Leonor Hernandez², Jose María Desantes², Mónica Martínez¹

¹Departamento de Estadística, Investigación Operativa Aplicadas y Calidad,
Universidad Politécnica De Valencia

²CMT, Máquinas y Motores, Universidad Politécnica de Valencia
anconpei@etsii.upv.es

Abstract

En este trabajo se presenta la comparación de distintas metodologías para determinar modelos matemáticos que permitan predecir simultáneamente cinco variables del comportamiento de un motor diesel en cuanto a emisiones contaminantes y consumo, con la finalidad de encontrar las condiciones óptimas de funcionamiento del motor de acuerdo con determinados objetivos como cumplir la futura normativa anticontaminante europea para los motores Diesel (EURO V). Se ha planteado la modelización del comportamiento de las variables respuesta mediante técnicas de superficies de respuesta y en base a los modelos obtenidos se ha realizado una optimización multiobjetivo de acuerdo a los valores establecidos por la legislación europea en cuanto a valores máximos de contaminantes. Se han utilizado diferentes técnicas de diseño de experimentos, en concreto los diseños centrales compuestos y Box-Behnken, y modelización mediante regresión múltiple. En el proceso de optimización se han comparado varias técnicas: SPQ (sequential quadratic programming), algoritmos genéticos y función de deseabilidad. Se trata de un problema de optimización multiobjetivo con restricciones. Se han realizado también experiencias de validación para evaluar la bondad del ajuste y la calidad de las predicciones según la metodología utilizada.

Palabras Clave: Diseño de experimentos, Motores diesel, modelización, optimización

Modelo basado en factores inobservados para la predicción de precios de la energía eléctrica a medio y largo plazo

Andrés M. Alonso¹, Carolina García-Martos², Julio Rodríguez³,
María Jesús Sánchez⁴

¹amalonso@est-econ.uc3m.es, Departamento de Estadística, Universidad Carlos III de Madrid

²cgarciam@etsii.upm.es, ³puerta@etsii.upm.es, ⁴mjsan@etsii.upm.es
Laboratorio de Estadística, Escuela Técnica Superior de Ingenieros
Industriales, Universidad Politécnica de Madrid

Abstract

En este trabajo se propone metodología para la predicción a medio y largo plazo de los precios de la energía eléctrica. El interés por este tipo de predicciones está relacionado con la negociación de contratos bilaterales entre clientes y comercializadoras.

Palabras Clave: Análisis factorial dinámico, bootstrap, contratos bilaterales

AMS: 37M10

1. Introducción

Actualmente en el mercado eléctrico se opera de dos formas: En el *pool* (Mercado Diario) y mediante los contratos bilaterales.

Mediante los contratos bilaterales las comercializadoras se comprometen a entregar a sus clientes determinadas cantidades de energía a un precio fijo, que queda establecido en un contrato que ambos suscriben por un plazo que suele ser de un año. La comercializadora debe adquirir a posteriori, hora a hora, la energía necesaria para servir a ese cliente, con un coste que vendrá determinado por el precio de mercado y la cantidad consumida. Puesto que este precio no se puede anticipar con certeza en el momento de la firma del contrato, se asume un riesgo, que disminuye si se dispone de unas predicciones más precisas de los precios.

2. Metodología

El objetivo de este trabajo es la obtención de predicciones para los precios de la energía eléctrica con un horizonte de predicción de entre 6 meses y un año, tema escasamente tratado hasta la fecha pues la mayoría de los trabajos existentes se centran en la predicción a corto plazo (horizonte de predicción entre 24 y 72 horas), ver Contreras et al. (2003).

Para ello se aplica y extiende la metodología desarrollada por Peña y Box (1987), Lee y Carter (1992) y Peña y Poncela (2004, 2006).

Se identifican y modelan los factores comunes inobservados a las 24 series

horarias de precios (obtenidos a partir de las relaciones existentes entre éstas), así como los factores específicos para cada una de las 24 series horarias.

Además se aplican técnicas bootstrap para la obtención de intervalos de confianza de las predicciones y se realiza inferencia para determinar la importancia que tiene cada una de las 24 series en la tendencia a largo plazo.

Los resultados obtenidos al aplicar la metodología que se propone son prometedores pues los errores de predicción están en torno al 20 % para un horizonte de entre 6 meses y un año. Con el objetivo de enmarcar estos resultados con respecto a resultados existentes hasta el momento hay que tener en cuenta que para horizontes de predicción de entre 24 y 72 horas se tienen errores de predicción en torno al 13 %.

3. Agradecimientos

A Pilar Poncela por su implementación en Matlab de la estimación del modelo factorial dinámico.

Andrés M. Alonso agradece la financiación del programa “Juan de la Cierva”.

Carolina García-Martos, Julio Rodríguez y María Jesús Sánchez agradecen la financiación del proyecto MTM2005-08897 del Ministerio de Educación y Ciencia.

4. Bibliografía

- [1] Contreras, J., Espínola, R. Nogales, F.J. y Conejo, A.J. *.^A RIMA Models to Predict Next-Day Electricity Prices*. IEEE Transactions on Power Systems, 18, 3, 1014-1020. (2003).
- [2] Lee R. D., Carter L. R. (1992). “Modeling and Forecasting U. S. Mortality”. *Journal of the American Statistical Association*. Vol 87, No. 419.
- [3] Peña D., Box G.E.P. (1987). “Identifying a Simplifying Structure in Time Series”. *Journal of the American Statistical Association*. Vol 82, No. 399.
- [4] Peña, D. and Poncela, P. (2004). “Forecasting with Nonstationary Dynamic Factor Models”. *The Journal of Econometrics*, 119, 291-321.
- [5] Peña, D. and Poncela, P. (2006) “Nonstationary Dynamic Factor Analysis” *Journal of Statistical Planning and Inference*, (en prensa).

Modelo bayesiano espacio-temporal para detectar actividad cerebral en imágenes de Resonancia Magnética funcional

Raquel Montes¹, Alicia Quirós²

¹raquel.montes@urjc.es, Departamento de Estadística e Investigación Operativa, Universidad Rey Juan Carlos

²alicia.quiros@urjc.es, Departamento de Estadística e Investigación Operativa, Universidad Rey Juan Carlos

Claves: fMRI, MCMC.

1. Resumen

Proponemos un modelo espacio - temporal para analizar imágenes de resonancia magnética funcional y de esta manera detectar regiones activas en el cerebro, basandonos en el modelo descrito por John Kornak. El modelo no hace ninguna suposición sobre la forma de las zonas activadas, excepto que se elevan con respecto a las no activadas y que forman regiones conexas ya que están caracterizadas a priori por un campo aleatorio de Markov (MRF).

La extensión mediante un único modelo bayesiano espacio - temporal permitirá obtener mayor información sobre la serie temporal que conforma, para un píxel, un estudio de resonancia magnética. A pesar del mayor coste computacional, un modelo espacio - temporal debería mejorar la capacidad de inferencia ya que tiene en cuenta, no solo la incertidumbre en la dimensión espacial sino también en la temporal.

Dada la complejidad del modelo, para hacer inferencia sobre los parámetros utilizamos métodos MCMC. Realizamos un estudio sobre datos simulados para comprobar la aplicabilidad del modelo a futuros datos reales y la sensibilidad del mismo. El principal benficio de este método es que las distribuciones a priori proporcionan un marco en el que detectar regiones activas más similar a los criterios del neurólogo; basado en la evidencia a posteriori sobre un ancho rango de escalas espaciales, considerando simultaneamente el valor de la magnitud en cada voxel y el tamaño del área activa. Además, al incorporar la información temporal en el modelo espacial (formando un modelo espacio - temporal) se incluye una fuente de información que muchos otros modelos no tienen en cuenta, sin aumentar demasiado el coste computacional.

Modelo de Zwinderman para evaluación del cambio en Calidad de Vida Relacionada con la Salud

Sara Morgado Nunes¹, M. Purificación Galindo²

¹sara@esg.ipcb.pt, Instituto Politécnico de Castelo Branco, Portugal

²pgalindo@usal.es, Departamento de Estadística, Universidad de Salamanca

Abstract

El estudio del cambio asume una importancia creciente en contextos de calidad de vida relacionada con la salud. En este trabajo se lleva a cabo un análisis crítico del modelo propuesto por Zwinderman (1990) para respuestas dicotómicas, el cual contiene la medida de la variable latente, un parámetro relacionado con el tiempo y otro relacionado con el tratamiento.

Palabras Clave: Calidad de Vida, Cambio, Modelo de Zwinderman

AMS: 62-07

1. Introducción

En los últimos años se ha asistido a un aumento significativo de publicaciones en los cuales se considera la temática de la Calidad de Vida Relacionada con la Salud (CVRS), especialmente en el contexto de ensayos clínicos en los cuales se pretende valorar aspectos relacionados con una determinada enfermedad o con las consecuencias de su tratamiento, como ocurre, a menudo, por ejemplo, en estudios con pacientes de cáncer que son sometidos a diversos tratamientos.

Una necesidad que surge con frecuencia en estudios de calidad de vida es la evaluación de cambios en los pacientes a lo largo del tiempo. Tradicionalmente, el análisis de este tipo de datos se basa en el estudio de las respuestas a los ítems con base en el ANOVA o MANOVA con medidas repetidas. Este tipo de análisis presenta algunos inconvenientes. En realidad, se está asumiendo que las variables latentes son constructos unidimensionales y los ítems sobre los cuales se calcula la suma de las puntuaciones se refieren al mismo constructo. Sin embargo, la CVRS, es un constructo de naturaleza dinámica y multidimensional que no puede ser observado o medido directamente.

2. Cambio en Calidad de Vida Relacionada con la Salud

Las valoraciones del constructo conseguidas por autoinforme presentan una problemática muy particular ya que en general no es posible aislar, ni evaluar, el verdadero cambio en la valoración del constructo cuando el individuo es sometido a una intervención experimental puesto que existen tres tipos de cambio tal como define Golembiewski *et al* (1976): cambio Alpha que es el verdadero cambio, cambio Beta debido a la posible recalibración de la escala de

medida del encuestado; y cambio Gamma el cual se produce cuando existe una reconceptualización del significado del constructo. Los cambios Beta y Gamma introducen un sesgo que necesariamente debe ser tenido en cuenta y aislado convenientemente, para poder medir los cambios reales.

En la literatura revisada son muy escasas las alusiones a este problema y mucho más escasos los trabajos que tratan de aportar alguna solución al respecto. Una de las aportaciones en el sentido de medir cambios en la calidad de vida en contextos de ensayos clínicos se debe a Zwinderman (1990), del cual se llevará a cabo un análisis crítico para poner de manifiesto sus ventajas y limitaciones.

3. El Modelo de Zwinderman

El modelo de Zwinderman (1990) se basa en el modelo de Rash (1961), según el cual las respuestas de N pacientes a k ítems son modelizadas por una curva logística.

Siendo $X_{\nu i1}$ la respuesta del paciente ν al ítem i en el instante T_1 y $X_{\nu i2}$ la respuesta del paciente ν al ítem i en el instante T_2 , se asume una variable latente θ_i para cada ítem y $\theta_{\nu i}$ representa la posición del paciente ν en el *continuum* latente θ_i en el instante T_1 . Debido a que se trata de un proceso temporal que envuelve uno o más tratamientos, $\theta_{\nu i}$ cambia según una cantidad δ_{ν} . Luego, la posición del paciente ν en el *continuum* latente θ_i en el instante T_2 es dada por $\theta_{\nu i2} = \theta_{\nu i} + \delta_{\nu}$. El cambio δ_{ν} es modelizado como una función del tiempo y del tratamiento que el paciente ν recibe $\delta_{\nu} = q_{\nu}\mu + \tau$ donde τ es el cambio inherente al proceso temporal, μ el cambio debido al tratamiento y q_{ν} la matriz de pesos.

El objetivo es estimar, de forma consistente, los parámetros μ y τ recurriendo al método de máxima verosimilitud de forma a evaluar el cambio observado.

4. Bibliografía

- [1] Golembiewski, R.T., Billingsley, K. and Yeager, S. (1976). Measuring change and persistence in human affairs: types of change generated by OD designs. *Journal of Applied Behavioural Science* **12**, 133-157.
- [2] Rash, G. (1961). On general laws and the meaning of measurement in psychology, *Berkeley symposium on mathematical statistics and probability*, University of California.
- [3] Zwinderman, A. (1990). The measurement of change of quality of life in clinical trials, *Statistics in Medicine* **9**, 931-942.

Modelo para la predicción de accidentes con participación de vehículos industriales en carreteras interurbanas

B. Arenas¹, E. Bernardos², F. Aparicio³, C. González⁴, M. J. Sánchez⁵, F. J. Páez⁶, A. Gómez⁷

¹barenas@etsii.upm.es, ²ebernardos@etsii.upm.es, ³faparicio@insia.upm.es,

⁶jpaez@insia.upm.es, ⁷givet@etsii.upm.es, Instituto Universitario de Investigación del Automóvil (INSIA), Cátedra de Transportes, E.T.S. de Ingenieros Industriales, Universidad Politécnica de Madrid

⁴camino@etsii.upm.es, ⁵mjsan@etsii.upm.es, Laboratorio de Estadística, E.T.S. de Ingenieros Industriales, Universidad Politécnica de Madrid

Abstract

Se han desarrollado modelos estadísticos para establecer la relación entre los accidentes con víctimas en los que se han visto involucrados vehículos industriales, con el tipo de vía y el volumen de tráfico de un determinado tramo de carretera y se han aplicado a la predicción de accidentes en algunos de los ejes de transporte de mercancías más importantes de España.

Palabras Clave: Regresión de Poisson, Accidentes, Vehículos industriales

AMS: 62J12, 62P30

1. Introducción

La estimación del número de accidentes que puede tener lugar en un tramo de carretera, es un asunto que ha generado gran interés en las últimas décadas. Como la mejora de la seguridad en las carreteras españolas es un objetivo prioritario por parte de las administraciones competentes y se enmarca en el objetivo europeo de reducción de un 50 % de las víctimas mortales hasta el año 2010, la influencia sobre los accidentes, de variables como el tipo de vía y el volumen de tráfico en las carreteras, constituyen un campo relevante de estudio y de plena actualidad.

Aunque los vehículos industriales suponen una pequeña parte de los vehículos implicados en accidentes, el número de víctimas mortales que este tipo de accidentes producen en los ocupantes de los vehículos contrarios, se revela muy significativo.

El objetivo de este trabajo es presentar el modelo ajustado, que permite cuantificar la influencia de las variables del tráfico, definidas en tramos homogéneos, sobre la tasa de accidentes con implicación de vehículos industriales.

2. Modelo de regresión de Poisson

El carácter aleatorio, discreto, no negativo y esporádico de los accidentes establece la regresión de Poisson [2], como el procedimiento más adecuado para modelar la tasa de accidentes λ_i definida por:

$$\lambda_i = \frac{\mu_i}{\nu_i}$$

siendo μ_i la media del número esperado de accidentes con vehículos industriales en un tramo i de carretera, y utilizando como medida de exposición ν_i la “exposición de vehículos industriales” (en millón vehículo industrial-kilómetro).

La selección del modelo empleado para la predicción y aplicación a ejes de transporte españoles se ha llevado a cabo en base a criterios estadísticos de bondad de ajuste (AIC, BIC, R²).

Los resultados y tendencias obtenidos muestran un ajuste adecuado con los valores observados y publicados en la literatura especializada. [1], [3] y [4].

3. Agradecimientos

Este trabajo ha sido financiado por el Ministerio de Fomento a través de la convocatoria 2003 de ayudas para la realización de proyectos de I+D en el área de transportes.

4. Bibliografía

- [1] Bauer, K. M, Harwood. D.W. (1994). *Statistical Models of Accidents on Interchange Ramps and Speed-Change Lanes*. Publication N° FWHA – RD- 97-106. U.S.DOT.
- [2] Dobson, A.J. (2002). *An Introduction to Generalized Linear Models*, Chapman & Hall.
- [3] Miaou, S.P., Hu, P.S., Wright, T., Rathin, A.K., Davis, S.C. (1992) Development of Relationships Between Truck Accidents and Highway Geometric Design: A Poisson Regression Approach. *Proceedings of the 71st Annual Meeting of the Transportation Research Board*. Washington. DC.
- [4] Vogt A. (1999). *Crash Models for Rural Intersections: Four Lane by Two Lane Stop Controlled and Two Lane by Two Lane Signalized*. Publication N° FWHA – RD- 99-128. U.S.DOT.

Modelos mixtos para la predicción de precios de la energía eléctrica a corto plazo

Carolina García-Martos¹, Julio Rodríguez², María Jesús Sánchez³

¹cgarciam@etsii.upm.es, ²puerta@etsii.upm.es, ³mjsan@etsii.upm.es

Laboratorio de Estadística, Escuela Técnica Superior de Ingenieros Industriales, Universidad Politécnica de Madrid

Abstract

En este trabajo se propone metodología para la construcción de modelos mixtos de predicción de precios a corto plazo que combinen las ventajas de diversos métodos que se han comparado para un largo periodo de tiempo. Además se realiza un estudio para determinar la longitud de serie con la que es preferible construir los modelos para predicción a corto plazo.

Palabras Clave: Análisis de series temporales, precio marginal, diseño de experimentos, predicción

AMS: 37M10

1. Introducción

La predicción de precios a corto plazo se ha convertido en algo necesario para poder planificar la generación de energía eléctrica en las distintas unidades de producción, pues es la base de cualquier estrategia para maximizar beneficios.

2. Metodología

El objetivo principal es la predicción de precios de la energía eléctrica a corto plazo. Las predicciones se calculan a partir de un modelo mixto construido combinando los aspectos más ventajosos de diversos modelos.

Se proponen dos nuevos modelos. El primero de ellos computa las predicciones para el precio en cada una de las 24 horas del día siguiente a partir del modelo ARIMA multiplicativo ajustado para cada una de las series horarias. Con el segundo modelo se calculan las predicciones para los días laborables utilizando las 24 series horarias de precios de este tipo de días, y se procede análogamente con los fines de semana.

Para determinar qué modelo conduce a unos menores errores de predicción se ha llevado a cabo un diseño de experimentos, en el que se han incluido factores relevantes en la predicción a corto plazo. Esto ha permitido determinar qué longitud de serie es preferible para construir los modelos con los que se calcularán las predicciones. Hasta ahora se habían tomado por defecto longitudes que varían entre 8 y 12 semanas.

Las predicciones se han calculado para varias combinaciones posibles de los factores cuya influencia en el error de predicción se estudia (Modelo y longitud

de la serie) para todos los precios horarios del periodo comprendido entre el 1 de enero de 1998 y el 31 de diciembre de 2003.

Dado el gran número de modelos a estimar (más de medio millón) para disponer de una gran muestra con la que llevar a cabo el diseño de experimentos, se ha procedido a la automatización del procedimiento de identificación y estimación de los modelos mediante la interfaz TS del programa TRAMO para Matlab.

El objetivo de este trabajo es más ambicioso que obtener predicciones para unos cuantos días o semanas, pues se pretende determinar los niveles de los factores que permitan minimizar los errores de predicción en futuras predicciones.

Se muestran resultados numéricos para el caso de mercado español pero la metodología que se propone en cuanto a construcción de modelos mixtos es aplicable a otros mercados.

3. Agradecimientos

A Juan Bógalo por las modificaciones en la interfaz TS del programa TRAMO para Matlab.

Los autores agradecen la financiación del proyecto MTM2005-08897 del Ministerio de Educación y Ciencia.

4. Bibliografía

- [1] Conejo A. J., Contreras J., Espínola R., Plazas M. A. (2005). *"Forecasting electricity prices for a day-ahead pool-based electric energy market"*. International Journal of Forecasting (2005).
- [2] Contreras, J., Espínola, R. Nogales, F.J. y Conejo, A.J. *"^ARIMA Models to Predict Next-Day Electricity Prices"*. IEEE Transactions on Power Systems, 18, 3, 1014-1020. (2003).
- [3] Crespo-Cuaresma J., Hlouskova, Kossmeier S., Obersteiner M. *"Forecasting electricity spot-prices using linear univariate time-series models"*. Applied Energy. Vol 77. (2004).
- [4] Nogales, F.J., Contreras, J., Conejo, A.J. y Espínola, R. *"Forecasting Next-Day Electricity Prices by Time Series Models"*. IEEE Transactions on Power Systems, 17, 2, 342-348. (2002).
- [5] Nogales F. J., Conejo A. J. *"Electricity Price Forecasting through Transfer Function Models"*. Journal of the Operational Research Society. (2006).

Monitorización on-line de procesos por lotes con el método de regresión sobre los scores truncados

Francisco Arteaga¹, Alberto Ferrer²

¹francisco.arteaga@ucv.es, Facultad de Estudios de la Empresa, Universidad Católica de Valencia San Vicente Mártir

²aferrer@eio.upv.es, Departamento de Estadística e I.O. Aplicadas y Calidad, Universidad Politécnica de Valencia

Abstract

En el presente trabajo se aborda el problema de monitorizar *on-line* procesos por lotes caracterizando la incertidumbre debida al desconocimiento que en cada instante se tiene de la trayectoria futura de las variables. Para ello se recurre a la distribución condicional de los estadísticos de monitorización dada una transformada lineal de la parte conocida del lote.

Palabras Clave: Control Estadístico Multivariante de Procesos por lotes, Análisis de Componentes Principales, TSR

1. Introducción

Entre las diferentes propuestas para la monitorización *on-line* de procesos por lotes las más empleadas consisten en asumir que en cada instante k el valor de las variables en los instantes que restan hasta la conclusión del lote son datos faltantes, recurriendo a métodos de imputación para su estimación. Con esto se consigue tener en cada instante k el lote “completo”, en el que una parte de los datos, los k primeros instantes del lote, son valores reales recogidos por los sensores y el resto son valores imputados.

Es práctica habitual que una vez imputadas las observaciones faltantes, la información del lote se trate estadísticamente como si procediera de un lote realmente completo (sin datos faltantes). Al proceder de este modo se ignora la incertidumbre que existe en cada instante k debida al desconocimiento del verdadero valor que van a tomar las variables en los instantes sucesivos.

En el presente trabajo se caracteriza la incertidumbre existente en cada instante a partir de la distribución condicional de los estadísticos de monitorización dada una transformación lineal del valor conocido de las variables durante los k primeros instantes del lote, basándose en una primera propuesta inicial de Nelson *et al.* (2006).

Esta forma de proceder proporciona intervalos de incertidumbre para los estadísticos de monitorización que permiten matizar la información obtenida de los métodos de imputación (valores esperados), permitiendo un mejor control

estadístico de este tipo de procesos. Los resultados se ilustran sobre un proceso industrial real.

2. Bibliografía

- [1] Nelson PRC, Taylor PA and MacGregor JF. Missing data methods in PCA and PLS: Score calculations with incomplete observations. *Chemometrics Intell. Lab. Syst.* 1996; **35**: 45-65.
- [2] Arteaga F and Ferrer A. Dealing with missing data in MSPC: several methods, different interpretations, some examples. *J. Chemometrics* 2002; **16**: 408-418.
- [3] Arteaga F and Ferrer A. Framework for regression-based missing data imputation methods in on-line MSPC. *J. Chemometrics*. En prensa. DOI: 10.1002/cem.946. Disponible *on-line*: www.interscience.wiley.com.
- [4] Nomikos P and MacGregor JF. Multivariate SPC charts for monitoring batch processes. *Technometrics* 1995; **41**: 41-59.
- [5] Nelson PRC, Taylor PA and MacGregor JF. The impact of missing measurements on PCA and PLS prediction and monitoring applications. *Chemometrics Intell. Lab. Syst.* 2006; **80**: 1-12.

Non-linear association on R^2 along a curve

Marcelo Smrekar¹, Pedro Delicado²

¹marcelo.smrekar@upc.edu, Departament d'Estadística i Investigació Operativa, Universitat Politècnica de Catalunya

²pedro.delicado@upc.edu, Departament d'Estadística i Investigació Operativa, Universitat Politècnica de Catalunya

Abstract

Se definen nuevas medidas de asociación entre dos variables aleatorias reales con distribución conjunta alrededor de una curva. Basándose en las expresiones de la covarianza y el coeficiente de correlación en términos de la primera componente principal, las nuevas definiciones se obtienen sustituyendo ésta por una curva principal.

Keywords: Measures of association; PCA; Principal Curves.

AMS: 62H20, 62G99, 62H25

1. Introduction

Correlation coefficient and covariance are appropriate association measures between two random variables when the underlying joint distribution is elliptical (and their association is then linear). However, it is very regular to find situations where the distribution is not elliptical. Several works have introduced measures of association between two random variables with non-elliptical joint distribution. Rényi [3] enunciated which properties any general association measure between two variables should verify. In this paper we propose two new association measures for two real random variables with joint distribution not necessarily elliptical. First we express correlation and covariance in terms of the first principal component. Then, we generalize them using principal curves (see Hastie and Stuetzle [2] or Delicado [1]).

2. Linear association in terms of Principal Components

Let $\mathbf{z} = [X, Y]$ be a two-dimensional random vector with variance matrix Σ . Let λ_1 and λ_2 be the greatest and the smallest eigenvalue of Σ , and let α be the angle between the eigenvector associated to λ_1 (PC1) and axis x . We can express variances, covariance and correlation as a function of λ_1 , λ_2 and α :

$$\sigma_X^2 = \lambda_1 \cos^2 \alpha + \lambda_2 \sin^2 \alpha, \quad \sigma_Y^2 = \lambda_1 \sin^2 \alpha + \lambda_2 \cos^2 \alpha,$$

$$\text{cov}(X, Y) = (\lambda_1 - \lambda_2) \cos \alpha \sin \alpha,$$

$$\rho_{(X,Y)} = (\lambda_1 - \lambda_2) \cos \alpha \sin \alpha / \{(\lambda_1 \cos^2 \alpha + \lambda_2 \sin^2 \alpha)(\lambda_1 \sin^2 \alpha + \lambda_2 \cos^2 \alpha)\}^{1/2}.$$

3. Non-linear association measures on R^2

Let $c : I \subseteq R \mapsto R^2$ be an one-dimensional smooth curve in the plane. Let $v(s), s \in I$, be and unitary vector field orthogonal to c (that is, $\langle c'(s), v(s) \rangle = 0$ for all $s \in I$). We define $\chi : I \times R \mapsto R^2$ by $\chi(s, t) = c(s) + tv(s)$. Assume that (S, T) has jointly distributed in $A \subseteq I \times R$ with density $h(S, T)$, that $E(T|_{S=s}) = 0$ and $V(S) > V(T|_{S=s})$, and that $\chi : A \mapsto \chi(A)$ is a one-to-one application. Then the random variable $(X, Y) = \chi(S, T) : A \mapsto R^2$ can be described as generated by a random point over the curve $c(I)$ plus an orthogonal random noise. So we can think of the curve $c(I)$ summarizes the structure of the (X, Y) distribution. For the particular case of $c(I)$ being a straight line, the statistical association between X and Y can be said to be linear, and it is well measured using covariance and correlation.

In order to generalize linear association measures, we start by defining local measures (of variance and covariance) around points $c(s)$ in R^2 :

$$V_X(s) = V(S) \cos^2 \alpha_{(s)} + V(T|_{S=s}) \sin^2 \alpha_{(s)},$$

$$V_Y(s) = V(S) \sin^2 \alpha_{(s)} + V(T|_{S=s}) \cos^2 \alpha_{(s)},$$

$$LGC_{(X,Y)}(s) = (E_S(S^2) - E_T(T^2|_{S=s})) \cos \alpha_{(s)} \sin \alpha_{(s)},$$

$$LGR_{(X,Y)}(s) = LGC_{(X,Y)}(s) / (V_X(s)V_Y(s))^{1/2}.$$

Two global measures of association (analogous to the squared values of covariance and correlation, respectively), are derived:

$$SGC(X, Y) = E_S[(LGC_{(X,Y)}(S))^2], \quad SGR(X, Y) = E_S[(LGR_{(X,Y)}(S))^2].$$

4. Acknowledgement

Research partially supported by the Spanish Ministry of Education and Science and FEDER, MTM2005-02370, and by the EU PASCAL Network of Excellence, IST-2002-506778.

5. Bibliography

- [1] Delicado, P. (2001). Another look at principal curves and surfaces. *Journal of Multivariate Analysis* 77, 84–116.
- [2] Hastie, T. and W. Stuetzle (1989). Principal curves. *Journal of the American Statistical Association* 84, 502–516.
- [3] Rényi, A. (1959). On measures of dependence. *Acta. Math. Acad. Sci. Hungar.* 10, 441–451.

Nonparametric forecasting in time series. A comparative study

Juan M. Vilar Fernández¹, Ricardo Cao²

¹eijvilar@udc.es ²rcao@udc.es

Departamento de Matemáticas, Facultad de Informática,
Universidad de A Coruña

Abstract

Se estudia el problema de predecir los futuros valores de una serie de tiempo utilizando la estimación núcleo de la autorregresión. Tres problemas son tratados: selección de los autorregresores, selección de la ventana y cálculo de intervalos de predicción por bootstrap. El algoritmo se aplica a 43 series y se compara con la predicción Box-Jenkins.

Keywords: time series, bootstrap, kernel regression estimation.

1. Formulation of the problem

Let us consider a strictly stationary univariate time series $\{Z_t : t \in Z\}$. We assume that the series has been observed in the time interval $1 \leq t \leq n$. The series is assumed to follow a Markov model of the form

$$Z_t = m(Z_{t-i_1}, Z_{t-i_2}, \dots, Z_{t-i_p}) + \varepsilon_t \quad (2.43)$$

where $m(\tilde{\mathbf{u}}) = E(Z_t / (Z_{t-i_1}, Z_{t-i_2}, \dots, Z_{t-i_p})) = \tilde{\mathbf{u}}$ with $\tilde{\mathbf{u}} \in R^p$, $0 < i_1 < i_2 < \dots < i_p$ and $\{\varepsilon_t\}$ is the error process, assumed to be independent on the past of Z_t and with zero conditional mean with respect to this past.

In order to predict Z_{t+1} we consider the data $\{(\tilde{\mathbf{X}}_j, Y_j) : j \in Z\}$, with values in $R^p \times R$, where $\tilde{\mathbf{X}}_j = (Z_{j+i_p-i_1}, Z_{j+i_p-i_2}, \dots, Z_j)$ and $Y_j = Z_{j+i_p}$. In other terms, we move from the observed time series data $\{z_t : 1 \leq t \leq n\}$ to the regression data $\{(\tilde{\mathbf{x}}_j, y_j) : 1 \leq j \leq n' = n - i_p\}$. The “new” data can be directly used to estimate the autoregression function $m(\cdot)$.

A predictor of Z_{t+1} is $\hat{m}(\tilde{\mathbf{x}}_{n-i_p+1}) = \hat{m}(z_{n+i_p-i_1}, z_{n+i_p-i_2}, \dots, z_n)$, where \hat{m} is an estimator of m .

Thus, the basic problem for predicting Z_{n+1} is finding some estimator $\hat{m}(\tilde{\mathbf{u}})$, with $\tilde{\mathbf{u}} = \tilde{\mathbf{x}}_{n-i_p+1}$, in a dependence context. Usual nonparametric estimators of $m(\tilde{\mathbf{u}})$ are of the general form

$$\hat{m}(\tilde{\mathbf{u}}) = \sum_{j=1}^{n'} W(\tilde{\mathbf{u}}, \tilde{\mathbf{x}}_j) y_j \quad (2.44)$$

where $W(\tilde{\mathbf{u}}, \tilde{\mathbf{x}}_j)$ are some smoothing positive weights with high values if $\tilde{\mathbf{u}}$ is close to $\tilde{\mathbf{x}}_j$ and values close to zero otherwise. Two different types of nonparametric kernel estimators are used in this paper: the Nadaraya-Watson estimator and the local lineal estimator.

The generalization to the problem of predicting Z_{t+l} (with $l \geq 1$) is straightforward, although presents an additional computational cost. After deciding the autoregressor variables that will play a role in this prediction problem the new sample $\{(\tilde{\mathbf{x}}_j, y_j)\}$ has to be redefined with an horizon of l instants ahead.

An alternative procedure consists in using a recursive algorithm. To do this, the one-instant ahead pointwise prediction of Z_{n+1} is obtained as detailed above. This will be denoted by $\hat{Z}_n(1)$. Now this prediction is inserted in the time series sample:

$$\{Z_1, Z_2, \dots, Z_n, \hat{Z}_n(1)\}$$

in order to predict Z_{n+2} one-instant ahead. This process is recursively repeated until the horizon prediction, l , is reached.

2. Nonparametric forecasts

There are three major issues in the implementation of these algorithms:

1. Selection of the autoregressor variables, in this work we use the procedure by Tjostheim & Auestad (1994). This method proposed to use a nonparametric version of the final prediction error (FPE) criterion and doing a sequence of searches. In each search a new lag is determined, with the lags found in the previous scans held fixed.
2. Smoothing parameter selection, the bandwidth selector that will be used for the two prediction algorithms proposed in this paper is the cross-validation method adapted to the presence of dependence.
3. Computing prediction intervals. We consider two different methods to construct prediction intervals. The first one is based in bootstrapping the residuals, while the second uses some estimation of the conditional distribution.

The performance of these nonparametric algorithms has been studied by applying them to a collection of 43 time series. Most of them are very well-known series from the books of Box & Jenkins (1976), Brockwell & Davis (1987), Pankratz (1983), Makridakis et al (1998) and Tong (1990). The majority of the 43 series are stationary or seasonally stationary and were also classified according to their linearity. The forecasts are computed using three different methods: a parametric ARIMA fit, the nonparametric Nadaraya-Watson estimator and the local linear procedure. Finally, the practical behaviour of the methods is also illustrated by a detailed analysis of two data sets.

On a variant of integral trimmed regions

Pedro Terán

teran@unizar.es, Dpto. Métodos Estadísticos, Universidad de Zaragoza

Abstract

A variant of the integral trimmed regions recently defined by Cascos and López-Díaz is proposed, its properties studied and compared to the original definition.

Keywords: Integral stochastic orders. Statistical depth. Trimmed regions.

AMS: 62H05

1. Introduction

Cascos and López-Díaz [1] have recently presented a general definition of *integral trimmed regions* for multivariate data analysis, as follows. Given a family \mathcal{F} of real functions on \mathbf{R}^d , and a value $\alpha \in (0, 1]$ (which controls the trimming intensity), define for a given probability measure P the region

$$D_{\mathcal{F}}^\alpha(P) = \{x \in \mathbf{R}^d \mid \exists Q \leq \alpha^{-1}P; \forall f \in \mathcal{F} f(x) \leq \int f dQ\}.$$

The definition is inspired by integral stochastic orders, in fact the trimmed region is formed by those points whose degenerate distribution is dominated, in the integral order associated to \mathcal{F} , by some distribution $Q \leq \alpha^{-1}P$ (denoted $Q \in P^\alpha$). One has an associated notion of *depth*: for each point $x \in \mathbf{R}^d$,

$$d_{\mathcal{F}}(x) = \sup\{\alpha \in (0, 1] \mid x \in D_{\mathcal{F}}^\alpha(P)\}.$$

Notice that $D_{\mathcal{F}}^\alpha(P)$ yields the contour of the depth function $d_{\mathcal{F}}$ at height level α . We propose the following variant:

$$T_{\mathcal{F}}^\alpha(P) = \{x \in \mathbf{R}^d \mid \forall f \in \mathcal{F} f(x) \leq \sup_{Q \in P^\alpha} \int f dQ\}.$$

It turns out to be simpler to prove good properties of $T_{\mathcal{F}}^\alpha$, and it often involves assumptions on the family \mathcal{F} which are less stringent and easier to check. The known classical examples of trimmed regions and depth functions that are covered by the definition of $D_{\mathcal{F}}^\alpha$ (like Tukey's half-space depth [3] and Koshevoy and Mosler's zonoid depth [2]) are also covered by $T_{\mathcal{F}}^\alpha$.

2. Properties of $T_{\mathcal{F}}^{\alpha}(P)$

The following condition, based on a minimax theorem, gives a sufficient condition for the identity of the two definitions.

(i) *If \mathcal{F} is formed by upper semicontinuous functions, then $D_{\mathcal{F}}^{\alpha}(P) = T_{\mathcal{F}}^{\alpha}(P)$.*

We will show, by a comparison of our results to those proven in [1], that proving a property for $T_{\mathcal{F}}^{\alpha}$ and then imposing upper semicontinuity almost always leads to the same or better conditions for $D_{\mathcal{F}}^{\alpha}$ than those derived directly by Cascos and López-Díaz.

Now we present sufficient conditions on \mathcal{F} so that $T_{\mathcal{F}}^{\alpha}$ enjoys some good properties proposed in [4, 5] for depth functions and trimmed regions.

(ii) *If \mathcal{F} is formed by lower semicontinuous functions, then $T_{\mathcal{F}}^{\alpha}(P)$ is closed.*

(iii) *If \mathcal{F} contains an integrable lower bounded function, then $T_{\mathcal{F}}^{\alpha}(P)$ is bounded.* (iv) *If \mathcal{F} is formed by quasiconvex functions, then $T_{\mathcal{F}}^{\alpha}(P)$ is convex.*

Other important properties, like *affine equivariance*, are obtained under conditions similar to those in [1]. Also, the asymptotic result for $D_{\mathcal{F}}^{\alpha}$ in [1] is valid for $T_{\mathcal{F}}^{\alpha}$, since the hypothesis include continuity. We note that their hypothesis that \mathcal{F} has an integrable envelope can be dropped.

3. Acknowledgement

Fruitful discussions with Nacho Cascos and Miguel López Díaz are acknowledged.

4. Bibliography

- [1] Cascos I. and López-Díaz M. (2005). Integral trimmed regions. *J. Multivar. Anal.* **96**, 404–424.
- [2] Koshevoy G. and Mosler K. (1997). Zonoid trimming for multivariate distributions. *Ann. Statist.* **25**, 1998–2017.
- [3] Tukey J. W. (1975). Mathematics and the picturing of data. *Proceedings of the International Congress of Mathematicians (Vancouver, B. C., 1974)*, 523–531. Canad. Math. Congress, Montreal.
- [4] Zuo Y. and Serfling R. (2000). General notions of statistical depth function. *Ann. Statist.* **28**, 461–482.
- [5] Zuo Y. and Serfling R. (2000). Structural properties and convergence results for contours of sample statistical depth functions. *Ann. Statist.* **28**, 483–499.

On ANOVA tests for a weighted sample of density functions

Pedro Delicado Useros¹

¹pedro.delicado@upc.edu, Departament d'Estadística i Investigació Operativa, Universitat Politècnica de Catalunya

Abstract

Se estudia el contraste ANOVA para datos funcionales cuando las observaciones son funciones de densidad con pesos distintos. Se adaptan a este tipo de datos dos tests genéricos y se introducen contrastes basados en distancias. Como aplicación se contrasta si la distribución de la renta relativa en regiones de Europa depende del país al que éstas pertenecen.

Keywords: Distance based inference; Functional Data Analysis; Nonparametric density estimation.

AMS: 62G07, 62H15

1. Introduction

In the last years, the joint development of real-time measurement instruments and data storage computer resources has made possible to observe and save complete functions as results of random experiments. Ramsay and Silverman [4] express it saying that random functions are the *statistical atoms* in these cases. Functional Data Analysis (FDA) deals with the statistical description and modelization of samples of random functions.

The one-way analysis of variance is one of the standard methods that have been generalized to be used in FDA, giving rise to *functional ANOVA*. It is assumed that n functions $f_{ri}(x)$ have been observed from the model

$$f_{ri}(x) = m_r(x) + e_{ri}(x), \quad i = 1, \dots, n_r, r = 1, \dots, k, x \in [a, b],$$

where $\sum_r n_r = n$ and $m_r(x)$ are k unknown mean functions and $e_{ri}(x)$ are independent trajectories drawn from a process with zero mean and covariance function $Cov(e_{ri}(x), e_{ri}(y)) = K(x, y)$. The hypothesis to be tested is

$$H_0 : m_1(x) = \dots = m_k(x), \quad \text{for all } x \in [a, b].$$

Two known methods for functional ANOVA are the proposals of Ramsay and Silverman [4] (based on F-ratio functions) and that of Cuevas, Febrero and Fraiman [1] (based on a functional Central Limit Theorem).

2. Functional ANOVA for weighted density functions

This work has two main objectives. The first one is to fit functional ANOVA to data being density functions. We analyze the applicability of the two known techniques, see [4] and [1], and how they can be adapted to the case of densities. Moreover we present a distance based ANOVA test working on pairwise distances between observed data (Gower and Krzanowski [3]). This device allows us to use the L_1 distance, that is the most natural one for density functions. The same distance based ANOVA procedure is also applicable to other distance definitions and to data not being density functions. The second objective is to generalize functional ANOVA tests for weighted samples. We establish results valid for weighted data that generalize those of [3] and [1].

The null distribution of any of the test statistics included in the paper is unknown. Thus, the use of permutations is required (see [3] and references therein). In this paper we consider two alternatives that we call *permuting observations* (the standard one) and *permuting residuals* (the model is estimated assuming the alternative hypothesis is truth, and then the artificial samples are defined functions as the sum of the global mean plus the permuted estimated residuals).

A simulation study is carried out to compare the practical behaviour of these tests. Moreover, for a collection of European regions we estimate the regional relative income densities and then we test the significance of the *country effect*. An extended version of this work is available in [2].

3. Acknowledgement

Research partially supported by the Spanish Ministry of Education and Science and FEDER, MTM2005-02370, and by the EU PASCAL Network of Excellence, IST-2002-506778.

4. Bibliography

- [1] Cuevas, A., M. Febrero, and R. Fraiman (2004). An anova test for functional data. *Computational Statistics and Data Analysis* 47, 111–122.
- [2] Delicado, P. (2005). Functional ANOVA when data are a weighted sample of density functions. *Preprint*.
[<www-eio.upc.es/~delicado/my-public-files/FANOVA.density.pdf>](http://www-eio.upc.es/~delicado/my-public-files/FANOVA.density.pdf)
- [3] Gower, J. C. and W. J. Krzanowski (1999). Analysis of distance for structured multivariate data and extensions to multivariate analysis of variance. *Journal of the Royal Statistical Society, Series B* 48, 505–519.
- [4] Ramsay, J. and B. W. Silverman (1997). *Functional Data Analysis*. New York: Springer.

Planes de muestreo de aceptación aproximados con censura progresiva

Carlos J. Pérez-González¹, Arturo J. Fernández²

¹cpgonzal@ull.es, ²ajfernán@ull.es, Departamento de Estadística e Investigación Operativa, Universidad de La Laguna

Abstract

El presente trabajo expone algunos diseños aproximados de planes de muestreo de aceptación en fiabilidad con censura progresiva tipo II, en el caso en que la variable de tiempo de vida sigue una distribución exponencial o una Weibull. Se muestra el procedimiento de obtención de dichos diseños así como una comparativa de los mismos con aquellos que se obtendrían de forma exacta aunque por métodos iterativos.

Claves: Fiabilidad, curva característica, tiempo de vida, distribuciones exponencial y Weibull.

1. Introducción

En análisis de fiabilidad, un plan de muestreo permite determinar la aceptabilidad de un producto sometido a una prueba con respecto a su duración o tiempo de fallo de acuerdo a las especificaciones fijadas por el productor y consumidor del mismo.

En este artículo se estudia el diseño de planes de muestreo en fiabilidad aproximados para varios supuestos de la distribución del tiempo de fallo, T . Se analiza el caso en que T sea exponencial, $Exp(\mu, \sigma)$, o Weibull, $W(exp(\mu), 1/\sigma)$, lo cual equivale a suponer que $\log(T)$ sigue una distribución del valor extremo, $EV(\mu, \sigma)$. Asimismo, se lleva a cabo una comparativa con resultados ya conocidos. En el caso de la Weibull la comparativa se lleva a cabo con respecto a resultados basados, también, en aproximaciones asintóticas.

2. Diseño de planes de muestreo en fiabilidad.

Considérese una muestra aleatoria de n unidades de producto sometidas a experimentación simultáneamente. Se observan los r primeros fallos, en cada uno de los cuales se retiran un número pre-especificado R_i de unidades funcionales del ensayo, es decir, se aplica un esquema $R = (R_1, \dots, R_r)$ de censura progresiva de tipo II. Por tanto, la proporción de censura total aplicada vendría dada por $q = 1 - r/n$, con $q \in [0, 1]$.

Supongamos que l es un límite de especificación superior asignado a T . Entonces la proporción de disconformes del producto es $p = \Pr(T < l)$. En el caso exponencial, el criterio para decidir sobre la aceptación de un lote de producto

es $V = \hat{\mu} - k\hat{\sigma} \geq l$, donde $\hat{\mu}$ y $\hat{\sigma}$ son los estimadores de máxima verosimilitud (MLEs) de μ y σ , mientras k es una constante de aceptación. En el caso Weibull (o del valor extremo), el criterio es $V = \tilde{\mu} - k\tilde{\sigma} \geq \ln(l)$, donde $\tilde{\mu}$ y $\tilde{\sigma}$ son los mejores estimadores lineales insesgados (BLUEs) de μ y σ .

La selección del plan de muestreo requiere de un acuerdo entre productor y consumidor de forma que el productor sólo aceptará los lotes con $p \leq p_\alpha$ con una probabilidad, al menos, de $1 - \alpha$, mientras que el consumidor rechazará los lotes con $p \geq p_\beta$ con, al menos, probabilidad $1 - \beta$. Por tanto, en la determinación del plan de muestreo es necesario conocer la expresión de la curva característica (curva OC) $L(p) = L(p; n, r, k, R) = \Pr(V \geq l | p, n, r, k, R)$, con lo cual los requerimientos anteriores se especifican como:

$$L(p_\alpha) = L(p_\alpha; n, r, k, R) \geq 1 - \alpha \quad \text{and} \quad L(p_\beta) = L(p_\beta; n, r, k, R) \leq \beta. \quad (2.45)$$

3. Determinación de los planes de muestreo aproximados.

Se supone, en primer lugar, que $T \sim Exp(\mu, \sigma)$. En caso que μ sea conocida (e igual a 0, sin pérdida de generalidad), se tiene que $U = 2r\hat{\sigma}/\sigma \sim \chi^2(2r)$ y la curva OC se expresa como:

$$L(p) = L(p; r, k) = \Pr(-kU > -\ln(1 - p)). \quad (2.46)$$

Si μ es desconocida, entonces $U_1 = 2n(\hat{\mu} - \mu)/\sigma \sim \chi^2(2)$ y $U_2 = 2r\hat{\sigma}/\sigma \sim \chi^2(2r - 2)$ son independientes, y la OC curva viene dada por:

$$L(p) = L(p; r, k) = \Pr(U_1 - kU_2 > -\ln(1 - p)). \quad (2.47)$$

A partir de estas expresiones será posible especificar, una vez fijado R (aunque bastará con fijar el nivel q en el caso de la distribución exponencial), el procedimiento para obtener el diseño (n , r y k) correspondiente al plan de muestreo de menor tamaño muestral que verifique las condiciones (2.45). Sin embargo, no es posible dar una expresión explícita general del mismo a partir de $L(p)$. Por tanto, se utilizan aproximaciones de varios tipos a la curva OC (asintótica, Wilson-Hilferty, Patnaik, etc.) que permiten deducir de forma inmediata algunos diseños aproximados.

También se analiza el caso en que $\log(T) \sim EV(\mu, \sigma)$. Bajo este supuesto no es posible obtener la distribución exacta de V y se requiere la utilización de aproximaciones a la curva OC. Se presenta un plan de muestreo obtenido a partir de una aproximación asintótica de V empleando estimadores BLUEs de μ y σ , y se lleva a cabo una comparativa con diseños aproximados ya conocidos para dicha distribución, los cuales utilizan estimadores MLEs.

Presmoothed estimation with left truncated and right censored data

M^a del Carmen Iglesias Pérez¹, M^a Amalia Jácome Pumar²

¹mcigles@uvigo.es, Departamento de Estadística e Investigación Operativa,
Universidade de Vigo

²majacome@udc.es, Departamento de Matemáticas, Universidade da Coruña

Abstract

In many situations, when studying survival times, the data may be incomplete because of left truncation and right censoring. In this work presmoothing ideas are proposed to estimate the cumulative hazard and the survival functions in this setup. An iid representation for both estimators is derived, from which the asymptotic distribution can be obtained.

Keywords: Almost sure representation; Presmoothing; Truncated and censored data.

1. Introduction

Let (Y, T, C) denote a random vector where Y is the variable of interest (lifetime) with distribution function (d.f.) F , T is the random left truncation time with d.f. L , and C is the random right censoring time with d.f. G . Left truncation arises when a subject is not included in the study because its lifetime origin precedes the starting time of the study and dies before this moment. Right censoring happens when the lifetime Y of a subject is not completely observed because a different event occurs before the end of the lifetime. So, in the random left truncation and right censoring (LTRC) model, one observes (T, Z, δ) , if $T \leq Z$, where $Z = \min(Y, C)$ and $\delta = \mathbf{1}\{Y \leq C\}$ indicates the cause of failure. When $Z < T$ nothing is observed. Let $\alpha = P(T \leq Z) > 0$ and H denote the d.f. of Z . It is assumed that Y and (T, C) are independent and $a_L \leq a_H$, where, for any d.f. W , $a_W = \inf\{y : W(y) > 0\}$.

Let (T_i, Z_i, δ_i) , $i = 1, 2, \dots, n$, be an i.i.d. sample from (T, Z, δ) which one observes (i.e., $Z_i \geq T_i$). The classical estimators of the cumulative hazard function Λ and F (see Tsai *et al.* (1987) among others) are:

$$\Lambda_n(y) = \sum_{i:Z_i \leq y} \frac{\delta_i}{nC_n(Z_i)} \text{ and } F_n(y) = \prod_{i:Z_i \leq y} \left(1 - \frac{\delta_i}{nC_n(Z_i)}\right)$$

where C_n is the empirical estimator of $C(y) = P(T \leq y \leq Z | T \leq Z)$.

2. The presmoothed estimators

Presmoothing ideas consist of replacing the censoring indicator variables δ_i in Λ_n and F_n with the values of a nonparametric estimator m_n of the conditional probability $m(y) = P(\delta = 1|Z = y, T \leq Z) = E(\delta|Z = y, T \leq Z)$ (without truncation see, for example, Cao *et al.* (2005)):

$$\Lambda_n^P(y) = \sum_{i:Z_i \leq y} \frac{m_n(Z_i)}{nC_n(Z_i)} \text{ and } F_n^P(y) = \prod_{i:Z_i \leq y} \left(1 - \frac{m_n(Z_i)}{nC_n(Z_i)}\right),$$

where m_n is the Nadaraya-Watson estimator with kernel K and bandwidth b .

Without truncation, and under some conditions, presmoothing has been proved to yield estimators more efficient than their classical counterparts.

3. Main results

In this work an almost sure i.i.d. decomposition of Λ_n^P and F_n^P is established. From it, the uniform consistency and a central limit result can be obtained. In addition, these new estimators are shown to have a good behavior with respect to Λ_n and F_n in a simulation study. Specifically, under certain assumptions we obtain for any $T < b_H = \sup\{y : H(y) < 1\}$ that:

$$\Lambda_n^P(y) - \Lambda(y) = \frac{1}{n} \sum_{i=1}^n [g_1^P(y, Z_i) - g_2^P(y, Z_i, T_i) + g_3^P(y, Z_i, \delta_i)] + S_n(y)$$

where

$$\begin{aligned} g_1^P(y, Z_i) &= \frac{m(Z_i)}{C(Z_i)} \mathbf{1}\{Z_i \leq y\} \\ g_2^P(y, Z_i, T_i) &= \int_{a_H}^y \frac{m(v)}{C^2(v)} \mathbf{1}\{T_i \leq v \leq Z_i\} dH^*(v) \\ g_3^P(y, Z_i, \delta_i) &= \int_{a_H}^y K_b(v - Z_i) \frac{\delta_i - m(v)}{C(v)} dv \end{aligned}$$

$$\text{and } \sup_{a_H \leq y \leq T} |S_n(y)| = O\left(\left[\left(\frac{\ln n}{nb}\right)^{1/2} + b^2\right]^2\right) \text{ a.s.}$$

with $H^*(y) = P(Z \leq y|T \leq Z)$. For F_n^P one has a similar result.

4. Bibliography

- [1] Cao R., López-de-Ullibarri I., Janssen P. and Veraverbeke N. (2005). Presmoothed Kaplan-Meier and Nelson-Aalen estimators. *Journal of Nonparametric Statistics* 17, 31-56.
- [2] Tsai W.Y., Jewely N.P. and Wang M.C. (1987). A note on the product limit estimator under right censoring and left truncation. *Biometrika* 74, 883-886.

Programa de simulación de superficies de respuesta

Josep Maria Mateo Sanz¹

¹josepmaria.mateo@urv.net, Área de Estadística e Investigación Operativa,
Universitat Rovira i Virgili

Abstract

Se presenta un archivo, para docencia, realizado en Excel en el cual se genera de manera aleatoria, pero controlada, una función de segundo orden con dos variables independientes y con un único punto óptimo. El alumno desconoce dicha función y dicho óptimo y, mediante las técnicas de los diseños factoriales simples y compuestos, debe encontrar ese punto óptimo.

Palabras Clave: Docencia, superficies de respuesta, simulación

1. Introducción

La motivación de este trabajo es presentar una herramienta para uso docente que facilite el aprendizaje integral de los contenidos propios de las técnicas de superficies de respuesta ([1], [2], [3]).

En mi labor docente he observado que los ejercicios de superficies de respuesta se presentan, a menudo, de manera parcial: hay que analizar un diseño factorial central para determinar si estamos en una zona de comportamiento lineal o cuadrático o se dan una serie de puntos y el alumno debe comprobar si están en la dirección de máxima pendiente o hay que analizar un diseño factorial central compuesto y, a partir de él, encontrar el punto óptimo. Muchas veces, no se ve el problema en conjunto.

También he observado que los alumnos tienen dificultades para decidir en qué valores de las variables independientes se deben realizar los experimentos ya que estos valores, y el resultado de la variable respuesta, se dan, generalmente, en los enunciados de los ejercicios.

El programa de simulación genera una función de segundo orden, en la que hay dos variables independientes, con un punto óptimo. El alumno desconoce cuál es dicha función y su punto óptimo y su objetivo es, a través de una hoja que permite generar valores de la variable respuesta a partir de valores de las variables independientes y del uso de las técnicas estadísticas de las superficies de respuesta, encontrar dicho punto óptimo.

Esta herramienta ha sido creada en Excel por ser un programa ampliamente usado en cualquier ámbito y que permite, a través de sus funciones básicas, generar y controlar todos los procesos que intervienen en un análisis de superficies de respuesta.

El archivo tiene una serie de celdas bloqueadas y ocultas con el objetivo que los alumnos no borren fórmulas accidentalmente y que no tengan acceso al resultado del punto óptimo que están buscando.

2. Uso del programa: visión alumno

Básicamente, el proceso a seguir por parte del alumno sería:

1. Crear un diseño factorial central alrededor de un punto central, propuesto aleatoriamente por el programa, y analizar los resultados obtenidos de dicho diseño. Dependiendo del resultado obtenido, el alumno puede optar por dos alternativas: paso 2) o paso 3).
2. Si en el paso 1) hay evidencias de comportamiento lineal, hay que realizar un estudio en la dirección donde se prevé que se encuentre el óptimo para fijar un nuevo punto central a partir del cual se volverá a la opción 1) y el alumno creará un nuevo diseño factorial central.
3. Si en el paso 1) no hay evidencia de comportamiento lineal y sí de comportamiento cuadrático, hay que crear un diseño factorial compuesto central a partir del diseño factorial central del paso 1). A partir del diseño factorial compuesto se obtendrá un punto óptimo aproximado al real.

Antes de empezar el proceso, el alumno debe indicar si desea buscar un máximo o un mínimo para que el programa genere la función adecuada. Para generar datos de la variable respuesta, el alumno dispone de una hoja de cálculo que genera dichos datos de manera aleatoria y controlada.

3. Uso del programa: visión profesor

El profesor puede escoger los valores de una serie de parámetros del proceso: en qué rango se pueden mover tanto los coeficientes de los términos cuadráticos como el punto óptimo como el coeficiente independiente de la función como el primer punto central a partir del cual se inicia todo el proceso. También puede decidir cuál es el paso que se usará para codificar las variables y la variabilidad que tendrá la generación de valores experimentales.

4. Bibliografía

- [1] Box, G.E.P., Hunter, W.G. and Hunter, J.S. (1999). *Estadística para Investigadores*. Ed. Reverté.
- [2] Montgomery, D.C. (2002). *Diseño y análisis de experimentos*. Limusa Wiley.
- [3] Prat, A., Tort-Martorell, X., Grima, P. y Pozueta, L. (1997). *Métodos Estadísticos. Control y Mejora de la Calidad*. Ed. UPC.

Proyecciones de la población española

Andrés M. Alonso¹, Daniel Peña², Julio Rodríguez³

¹andres.alonso@uc3m.es, Departamento de Estadística,
Universidad Carlos III de Madrid

²daniel.pena@uc3m.es, Departamento de Estadística,
Universidad Carlos III de Madrid

³puerta@etsii.upm.es, Departamento de Ingeniería de Organización,
Administración de Empresas y Estadística, Universidad Politécnica de Madrid

Abstract

En este trabajo se abordan las proyecciones de la población española por sexo y edad para el periodo 2005 a 2050. Estas proyecciones se realizan a partir de los pronósticos de nacimientos, defunciones y saldo migratorio. Los resultados obtenidos se comparan con las proyecciones de población realizadas por el Instituto Nacional de Estadística (2004).

Claves: demografía; modelo factorial dinámico; bootstrap.

AMS: 91D20; 62F40.

1. Introducción

El cambio demográfico que se ha observado en el último cuarto del siglo pasado se prevé que tenga numerosas consecuencias socioeconómicas. Este cambio ha involucrado las tres componentes básicas de la evolución de la población: (*i*) España ha pasado de tener una alta tasa de fecundidad, 2.8 hijos/mujer en 1975, a ser uno de los países del mundo con fecundidad más baja; (*ii*) entre 1975 y 2000 se incrementó la esperanza de vida en 5.9 años en mujeres y en 5.0 años en hombres; (*iii*) y en los últimos años ha experimentado un notable incremento en el flujo de inmigrantes.

En este trabajo aborda la proyección de la población española por sexo y edad para el periodo 2005 a 2050. Se combina el método de cohortes por componentes con el análisis de series temporales tal como proponen Lee y Carter (1992) y Lee y Tuljapurkar (1994). Concretamente, empleamos un modelo factorial dinámico similar al propuesto por Lee y Tuljapurkar (1994), pero introduciendo restricciones en los factores comunes para modelar las tasas de mortalidad y fecundidad por grupos de edades simples. Se propone un procedimiento bootstrap para modelo factorial dinámico. Una ventaja fundamental de este enfoque bootstrap es que como resultado se obtienen las funciones de distribución de las predicciones de las tasas de mortalidad y fecundidad. En lo que se refiere a la migración neta, proponemos analizar la inmigración mediante la selección de distintos modelos de crecimiento bajo

supuestos sobre las cotas superiores en el porcentaje de inmigrantes respecto de la población total.

2. Resultados

A modo de resumen de los resultados, en la figura 1 presentamos la pirámide de población obtenida con un flujo de inmigración intermedio donde se observa un envejecimiento progresivo de la población y pequeño ensanchamiento de la base de las pirámides que corresponde a la ligera recuperación de la fecundidad. El resto de los resultados pueden consultarse en Alonso et al (2005).

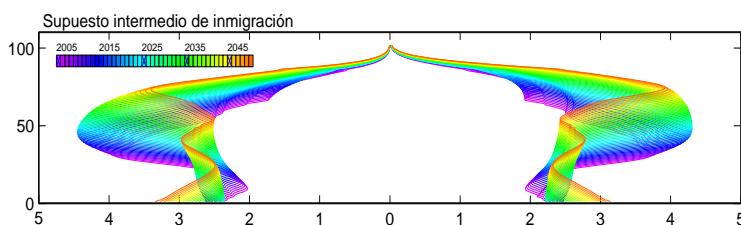


Figura 1: Pronósticos de las pirámides de población utilizando el modelo de cohortes por componentes. España 2005 - 2050.

3. Agradecimientos

Este trabajo ha sido financiado por el proyecto # 2233 de la Fundación BBVA. El primer autor agradece la financiación del programa “Juan de La Cierva” del Ministerio de Educación y Ciencia.

4. Bibliografía

- [1] Alonso, A.M., Peña, D. y Rodríguez, J. (2005). Proyecciones de la población española, Documento de trabajo. Disponible en: <http://halweb.uc3m.es/esp/Personal/personas/amalonso/esp/wp1.pdf>.
- [2] INE: Instituto Nacional de Estadística (2004). Proyecciones de la población de España calculadas a partir del censo de población de 2001, Madrid. Disponible en: <http://www.ine.es/inebase/cgi/um#5b>
- [3] Lee, R.D. y Carter, L. (1992). Modeling and forecasting the time series of U.S. mortality, *The Journal of the American Statistical Association*, 87, 659-671.
- [4] Lee, R.D. y Tuljapurkar, S. (1994). Stochastic populations forecast for the United States: beyond the high, medium and low, *The Journal of the American Statistical Association* 89, 1175-1189.

Quantile Curves and Dependence Structure for Bivariate Distributions

F. Belzunce¹, A. Castaño-Martínez², A. Olvera-Cervantes², A. Suárez-Llorens²

¹belzunce@um.es, Departamento de Estadística e Investigación Operativa,
Universidad de Murcia

²{antonio.castano, ana.olvera, alfonso.suarez}@uca.es, Departamento de
Estadística e Investigación Operativa, Universidad de Cádiz

Abstract

A partir de la definición de cuantil multivariante dada por Fernández-Ponce y Suárez-Llorens (2002), como curvas de nivel, estudiamos para distribuciones bivariantes, las relaciones entre las probabilidades acumuladas en las 5 regiones obtenidas a partir de las curvas y su estructura de dependencia. Propondremos un contraste de independencia.

Keywords: Bivariate Quantile, Copula, Test for Independence.

1. Introduction

For a one-dimensional probability distribution, the classical concept of quantile is frequently used in the construction of popular descriptive statistics and arises in all applied sciences. The usual definition of the univariate quantile function is given by the inverse of the distribution function and has the following well known characterizations in the literature:

- By accumulated probability. From the inverse of the distribution function given by

$$Q_X(u) = F_X^{-1}(u) = \inf\{x : F_X(x) \geq u\}, \forall u \in . \quad (2.48)$$

- By minimizing distances. Under some regular conditions, the quantile function is the argument in the following minimization problem

$$\inf_{\theta \in \mathbb{R}} E \left\{ \frac{|X - \theta| + (2\alpha - 1)(X - \theta)}{2} - \frac{|X| + (2\alpha - 1)X}{2} \right\}. \quad (2.49)$$

- By uniform transformation. Let U be a uniform random variable on $[0, 1]$. If the distribution function of X is strictly increasing then the quantile function is the only one increasing transformation of U which satisfies that $Q_X(U) =_{st} X$, where here st means the same distribution.

In addition, the concept of quantile is linked to the definition of the interquantile interval $IQ(p) = (Q_X(1-p), Q_X(p))$, for $1/2 < p < 1$ also called central region for univariate variables.

During many years authors have been concerned about the generalization of those concepts. Since 1976 we have in the literature various attempts towards both multidimensional generalization of univariate quantiles and the real interquantile interval. The majority of those generalizations have been developed applying the three characterizations mentioned above. Authors found difficulties in the choice of the shape of the central region which sometimes could not be appropriated for non-symmetrical distributions, the non parametric estimation of the new concepts and the study of the accumulated probability in the regions which most of the times is less obvious than in the real case.

As a natural generalization of the univariate quantile function Fernández-Ponce and Suárez-Llorens (2002) defined a multivariate quantile as a set of points which accumulate the same probability for a fixed orthant, they called it level curves or quantile curves. This definition also leads to a new multivariate central region concept. In particular those concepts solve some of the above described problems.

The purpose of this work is to use this concept of quantile curve in order to study the dependence structure -given by the concept of copula- of a bivariate distribution function, we analyze how the accumulated probability in the lateral regions is modified under a positive dependence structure. We use this result to present an independence test for bivariate distributions. Finally, we provide two numerical examples in order to illustrate our method.

2. Bibliography

- [1] Fernandez-Ponce, J. M. and Suárez-Llorens, A. (2002). Central Regions for Bivariate Distributions, *Austrian Journal of Statistics. -On Line Available-* 31, 141-156.
- [2] Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman and Hall, London.
- [3] Nelsen, R. B. (1999). *An Introduction to Copulas*. Springer-Verlag, New York, Inc.
- [4] De Martini, D. and Vespa E. (2005). Copula-Based Models for the Power of Independence Test. *Communications in Statistics- Theory and Methods*. 34, 2283-2297.

Rational bounds for the median of the $\Gamma(n + 1, 1)$ distribution

José A. Adell, P. Jodrá

Dpto. de Métodos Estadísticos, Universidad de Zaragoza, Spain
 {adell,pjodra}@unizar.es

Abstract

We develop a methodology to obtain lower and upper rational bounds for the median of the $\Gamma(n + 1, 1)$ distribution from its asymptotic expansion. An improved procedure to compute such bounds is outlined.

Keywords: Gamma distribution, median.

AMS: 60E05

1. Introduction

Denote by λ_n the median of the $\Gamma(n + 1, 1)$ distribution, $n = 0, 1, \dots$, and by $(N_t)_{t \geq 0}$ the standard Poisson process. Let $(C_k(t; x))_{k \geq 0}$ be the sequence of Charlier polynomials with respect to N_t . On the other hand, Ramanujan [3] defined the sequence $(\theta_n)_{n \geq 0}$ as the solution to the equation

$$\frac{e^n}{2} = \sum_{k=0}^{n-1} \frac{n^k}{k!} + \frac{n^n}{n!} \theta_n.$$

The asymptotic expansion of λ_n can be obtained from the asymptotic expansion of θ_n –an efficient method to obtain an arbitrary number of terms in the asymptotic expansion of θ_n is given in Marsaglia [2]– and by virtue of the following relation between θ_n and λ_n given in Adell and Jodrá [1]

$$\theta_n - 1 = \sum_{k=0}^{\infty} C_k(n; n) \frac{(n - \lambda_n)^{k+1}}{(k+1)!}, \quad n \in \mathbb{N} \setminus \{0\}. \quad (2.50)$$

In order to estimate the p first terms of the asymptotic expansion of λ_n , we shall equate the coefficients of n^{-k} , $k = 1, \dots, p$, in both sides of formula (2.50). To this end, the main term on the right-hand side of (2.50) is

$$B_p(n) := \sum_{k=0}^{2p} C_k(n; n) \frac{(n - \lambda_p(n))^{k+1}}{(k+1)!}, \quad (2.51)$$

where $\lambda_p(n) := n + 2/3 + \sum_{i=1}^p q_i n^{-i}$. We note that all of the $2p$ summands in (2.51) are meaningful, because each one contains terms of order n^{-k} , for

some $k \leq p$. However, formula (2.51) contains terms of order n^{-k} , for $k = 0, 1, \dots, 4p^3 - p$. This turns out to be the bottleneck of the procedure (see Remark below). To speed up the brute force algorithm, we shall obtain the term $\lambda_p(n)$ from $\lambda_{p-1}(n)$ by means of the following recurrence formula

$$B_p(n) = B_{p-1}(n) + C_{2p-1}(n; n) \frac{(n - \lambda_{p-1}(n))^{2p}}{(2p)!} + C_{2p}(n; n) \frac{(n - \lambda_{p-1}(n))^{2p+1}}{(2p+1)!}$$

$$+ \sum_{k=0}^{2p} \frac{C_k(n; n)}{(k+1)!} \left\{ \sum_{j=1}^{k+1} \binom{k+1}{j} (n - \lambda_{p-1}(n))^{k+1-j} (-q_p n^{-p})^j \right\},$$

where $B_0(n) := 0$ and $\lambda_0(n) := n + 2/3$.

Numerical computations show that the series $\{\lambda_p(n)\}_{p \geq 1}$ diverges for every n . As a consequence, the series behaves as asymptotic series often do; partial sums approach the true value of λ_n , reach the point of closest approach, then get increasingly worse. Results for a few n 's are given in the table below.

n	Closest partial sum	Number of correct significant digits
1	10th	3
2	16th	6
4	28th	11

Finally, using similar arguments as in Adell and Jodrá [1], the following lower and upper bounds for λ_n , $b_m(n) < \lambda_n < a_m(n)$, are obtained

$$b_m(n) := n + \frac{2}{3} + \sum_{i=1}^{m-1} q_i n^{-i}, \quad \text{and} \quad a_m(n) := b_m(n) + \frac{q_m}{n^m}, \quad \text{if } q_m, q_{m-1} > 0,$$

$$b_m(n) := n + \frac{2}{3} + \sum_{i=1}^m q_i n^{-i}, \quad \text{and} \quad a_m(n) := b_m(n) - \frac{q_m}{n^m}, \quad \text{if } q_m, q_{m-1} < 0,$$

whenever $m \geq 7$. These bounds are sharp.

Remark. It remains as an open problem to obtain an explicit expression for the aforementioned coefficients q_i , which allows us to develop an efficient algorithm to compute an arbitrary number of terms in the asymptotic expansion of λ_n .

2. Bibliography

- [1] J.A. Adell and P. Jodrá. On a Ramanujan's equation connected with the median of the gamma distribution. Preprint.
- [2] J.C.W. Marsaglia (1986). The incomplete Gamma function and Ramanujan's rational approximation to e^x . *J. Statis. Comput. Simul.* **24**, 163–169.
- [3] S. Ramanujan (1911). Question 294. *J. Indian Math. Soc.* **3**, 128.

Redes neuronales para la predicción del fracaso empresarial

Toni Moreno-Garcia

toni.moreno@upc.edu

Departamento de Organización de Empresas, Universidad Politécnica de Catalunya

Abstract

Este trabajo presenta el problema del diagnóstico del fracaso empresarial y expone diversos métodos para predecir la quiebra de una empresa a partir de datos históricos. Mediante técnicas basadas en redes neuronales se ha conseguido predecir correctamente hasta el 88.9 % de los datos de una muestra independiente.

Palabras Clave: redes neuronales, fracaso empresarial, clasificación

1. Introducción

La determinación de la salud financiera de una empresa es un problema muy importante para muchos actores de la escena económica. La predicción del fracaso empresarial, entendiendo como tal la situación jurídica de quiebra o suspensión de pagos, es un caso particular del problema de la determinación de la salud financiera de una empresa. Un sistema eficaz de predicción, entre otros beneficios, aportaría a los ejecutivos de la empresa una información muy útil para detectar los problemas que ponen en peligro su supervivencia.

Este estudio presenta el problema del fracaso empresarial y expone diversos métodos provenientes de la estadística, la investigación operativa y la inteligencia artificial aplicables a la resolución al problema de la predicción del fracaso empresarial, que se formula como un problema de clasificación con dos categorías: empresas quebradas y empresas no quebradas. Se han desarrollado, entre otros, algoritmos de teoría bayesiana de la decisión, de discriminación lineal, de regresión logística, de programación lineal y entera y de redes neuronales.

Los diferentes métodos se han comparado empíricamente sobre una muestra de cinco ratios financieros de 129 empresas, de las cuales 64 son empresas quebradas. La muestra se ha dividido en tres conjuntos independientes: conjunto de entrenamiento (74 empresas, 36 de ellas quebradas), que contiene los datos que utilizan los diferentes algoritmos para aprender la regla que permite clasificar según sean quebradas o no quebradas; conjunto de validación (28 empresas, 14 de ellas quebradas), que se utiliza en algunos algoritmos para asegurar que no se sobreentrena el sistema, interrumpiéndose el entrenamiento cuando se produce un descenso del rendimiento para las empresas de dicho conjunto; y conjunto de test (27 empresas, 14 de ellas quebradas), que se utiliza

para medir el rendimiento del sistema (es decir, su tasa de acierto) cuando funciona con datos nuevos.

Se ha constatado que las técnicas aplicadas son, en general, adecuadas en mayor o menor medida para tratar el problema de predicción del fracaso empresarial, teniendo en cuenta que en un problema de clasificación binaria cualquier técnica que ofrezca una tasa de acierto superior al 50 % (la que se obtendría con un decisor aleatorio) aporta valor a la situación preliminar.

Los mejores resultados se han obtenido con los métodos basados en redes neuronales, con los que se ha podido predecir el estado financiero de una empresa en el 100 % de los casos de la muestra de entrenamiento y en el 88.9 % de una muestra independiente.

2. Bibliografía

- [1] Y. Alici. *Neural networks in corporate failure prediction: The UK experience*. in Proc. Third Int. Conf. Neural Networks in the Capital Markets, A. N. Refenes, Y. Abu-Mostafa, J. Moody, and A. Weigend, London, UK, Oct. 1995, pp. 393–406.
- [2] E. Altman. *Financial ratios, discriminant analysis and the prediction corporate bankruptcy*. J. Finance, vol. 13, pp. 589–609, 1968.
- [3] Amir F. Atiya. *Bankruptcy Prediction For Credit Risk Using Neural Networks: A Survey And New Results*. IEEE Transactions On Neural Networks, Vol. 12, No. 4, July 2001.
- [4] R. Duda, P. Hart, D. Stork. *Pattern Classification. 2nd Edition*. John Wiley & Sons, 2001.
- [5] N. Freed, F. Glover. *Simple but powerful goal programming models for discriminant problems*. European Journal of Operational Research, 7:44-60, 1981.
- [6] M. Odom, R. Sharda. *A neural network model for bankruptcy prediction*. Proc. Int. Joint Conf. Neural Networks, San Diego, CA, 1990.
- [7] C. Serrano. *Self organizing neural networks for financial diagnosis*. Decision Support Systems, 1996.
- [8] L.C. Thomas, D. B. Edelman, J. N. Crook. *Credit Scoring & Its Applications*. Society for Industrial & Applied Mathematics, 2002.
- [9] P.D. Wasserman, T. Schwartz. *Neural Networks. Part 1*. IEEE Expert, Spring, 10-15, 1988.

Reglas de clasificación para poblaciones normales ordenadas

Cristina Rueda¹, Miguel Fernández, Bonifacio Salvador

¹Departamento de Estadística e Investigación Operativa
Universidad de Valladolid, crueda@eio.uva.es

Abstract

Consideramos el problema de clasificación en dos poblaciones cuando se dispone de información adicional de tipo ordenación estocástica sobre las distribuciones de las poblaciones. Construimos nuevas reglas y comprobamos mediante resultados teóricos y simulación que mejoran a la regla de Fisher. Ilustramos además su comportamiento en un caso real.

Claves: Restricciones de Orden, Análisis Discriminante, Regla de Bayes.

1. Introducción

Consideremos el problema clásico de discriminación: tenemos una observación del vector T y queremos clasificarla en Π_1 o Π_2 . Supongamos que las distribuciones de T en Π_1 y Π_2 son normales $N_k(\mu_1, \Sigma)$ y $N_k(\mu_2, \Sigma)$ respectivamente y que están estocásticamente ordenadas lo que implica que $\mu_1 \geq \mu_2$. En esta situación podemos definir nuevas reglas de Bayes de diferentes formas. En primer lugar sustituyendo el estimador habitual de las medias por un estimador restringido y en segundo lugar utilizando un espacio latente de observaciones en el que se define un problema de discriminación equivalente que permite modificar la regla teórica. En este trabajo se presentan las dos opciones y se aplican en la resolución de un caso real.

2. Reglas con estimadores restringidos de las medias

Es claro que para tener en cuenta la restricción entre las medias podría utilizarse directamente el estimador máximo verosímil restringido en la regla de Fisher pero hemos comprobado que este estimador puede mejorarse debido a que toma valores en la frontera del cono con probabilidad mayor que 0. En Fernández et al. (2006) definimos mediante un procedimiento iterativo una familia de estimadores de la diferencia de medias δ_γ^* con buenas propiedades. La regla de clasificación $R(\delta_\gamma^*)$ es

$$\text{Clasificar } T \text{ en } \Pi_1 \text{ si y solo si } (T - (c_1\hat{\mu}_1 + c_2\hat{\mu}_2) + c\delta_\gamma^*)' S^{-1} \delta_\gamma^* \geq 0.$$

donde $c_1 = \frac{n_1}{n_1+n_2}$, $c_2 = \frac{n_2}{n_1+n_2}$, $c = \frac{c_1-c_2}{2}$ y n_1 y n_2 son los tamaños de las muestras de entrenamiento.

3. Reglas a partir de un espacio latente

En este caso definimos un espacio latente de observaciones en una línea similar en cierto modo a la de Hoff (2003). Esto nos permite incluir la información sobre la ordenación estocástica entre las poblaciones directamente en la regla y no sólo a través de los estimadores de los parámetros desconocidos. Estos se consigue dando alta probabilidad a priori a ciertos conjuntos A_m , $m = 1, \dots, k$ construidos a partir de las restricciones.

4. Ejemplo. Diabetes en los indios Pima

Este conjunto de datos clásico en análisis discriminante puede encontrarse en el archivo del UCI. Consta de 768 observaciones y 8 variables. Los grupos se establecen dependiendo de si el paciente muestra signos de diabetes o no. Las restricciones usadas son las dadas por la muestra total. De este modo se supone que la población de pacientes con diabetes toma valores medios mayores en cada una de las variables. Se extrae una muestra de entrenamiento del 25 % del total que no cumple alguna de las restricciones.

	Muestra de entrenamiento						
	Tamaño	Fisher	$R(\delta_0^*)$	$R(\delta_{0,5}^*)$	$R(\delta_1^*)$	$m = 4$	$m = 3$
Grupo 0	129	108	105	103	101	98	103
Grupo 1	74	55	53	53	53	54	54
Global %		0,7980	0,7783	0,7684	0,7586	0,7438	0,7734

	Muestra test						
Grupo 0	371	291	294	298	298	296	301
Grupo 1	194	133	134	135	135	134	134
Global %		0,7504	0,7575	0,7663	0,7663	0,7611	0,7699

Cuadro 1: Datos Diabetes. Tabla de clasificación correcta ($p = 0,25$)

5. Bibliografía

- [1] Blake, C. L. and Merz, C. J. (1998). *UCI Repository learning databases*. [<http://www.ics.uci.edu/~mlearn/MLRepository.html>] Irvine, CA: University of California, Department of Information and Computer Science.
- [2] Fernández, M. A., Rueda, C. and Salvador, B. (2006). Incorporating additional information to normal linear discriminant rules. *Journal of the American Statistical Association*. To appear.
- [3] Hoff, P.D. (2003). Bayesian Methods for Partial Stochastic Ordering. *Biometrika* **90**, 303-317.

Regresión basada en distancias en presencia de heteroscedasticidad

Eva Boj¹, Josep Fortiana², M. Mercè Claramunt¹

¹evaboj@ub.edu, mmclaramunt@ub.edu, Departament de Matemàtica Econòmica, Financera i Actuarial, Universitat de Barcelona

²fortiana@ub.edu, Departament de Probabilitat, Lògica i Estadística, Universitat de Barcelona

Abstract

Deducimos las fórmulas de predicción para nuevos individuos en la regresión basada en distancias cuando aparece una ponderación, ya sea debida a heteroscedasticidad intrínseca de las respuestas o agregación de los datos.

Palabras Clave: Predicción Basada en Distancias, Datos Agregados, Seguros de Automóviles

AMS: 62H99, 62J99, 62P05

1. Introducción

En la tarificación *a priori* de los seguros no vida, por ejemplo el seguro del automóvil [1], se suele trabajar con datos agregados, a partir de los cuales construimos los grupos de tarifa. La *prima pura* se calcula como el producto de la esperanza del número de siniestros por la esperanza de la cuantía de un siniestro [2]. Ambas cantidades conllevan ponderación: Por un lado, el tratamiento del número de siniestros se realiza a través de la *frecuencia de siniestralidad* o número medio de siniestros por celda, ponderando las observaciones con el número de pólizas a partir de las cuales se ha calculado el promedio; por otro lado, el tratamiento de la cuantía de un siniestro se realiza a partir de la cuantía media por celda, ponderando con el número de siniestros. Tanto en un caso como en el otro, estas ponderaciones inducen heteroscedasticidad en las respuestas de las correspondientes regresiones. Los métodos basados en distancias, [4] [5], son particularmente apropiados para problemas de predicción cuyas covariables son de carácter mixto como el que nos ocupa. En trabajos anteriores, cfr. e.g., [3] hemos estudiado diferentes aspectos de esta aplicación. En este trabajo abordamos el problema de adaptar los cálculos de la regresión basada en distancias a observaciones con ponderación o heteroscedásticas.

2. Efecto de la ponderación en las predicciones

Consideramos m observaciones, a las que corresponde un vector de ponderaciones $\mathbf{w} = (w_1, \dots, w_m)'$. Los centrados deben hacerse según el proyector no

ortogonal (idempotente no simétrico, hablando algebraicamente)

$$\mathbf{J}_\mathbf{w} = \mathbf{I} - \mathbf{1} \cdot \mathbf{w}'.$$

Obtenemos las fórmulas adecuadamente adaptadas y los correspondientes procedimientos computacionales para las predicciones.

3. Contactar

Dirección: Eva Boj del Val,

Departament de Matemàtica Econòmica, Financera i Actuarial, Facultat de Ciències Econòmiques i Empresariais, Universitat de Barcelona,
Diagonal, 690, 08034 Barcelona, España.

Teléfono: (+34) 934035744 **Fax:** (+34) 934034892

e-mail: evaboj@ub.edu

4. Agradecimientos

Trabajo financiado en parte por el proyecto MTM 2005-02370 del Ministerio de Ciencia y Tecnología y FEDER.

5. Bibliografía

- [1] Baxter, L. A., S. M. Coutts and G. A. F. Ross (1980). Applications of Linear Models in Motor Insurance. *Transactions of the 21st International Congress of Actuaries* **2**, T11-29.
- [2] Boj, E., M. M. Claramunt, J. Fortiana y A. Vegas (2005). Bases de datos y estadísticas del seguro de automóviles en España: influencia en el cálculo de primas. *Estadística Española* **47**, 539–566.
- [3] Boj, E., M. M. Claramunt and J. Fortiana (2006). *Selection of Predictors in Distance-Based Regression* (submitted).
- [4] Cuadras, C. M. (1989). Distance Analysis in Discrimination and Classification using both Continuous and Categorical Variables. In: Dodge, Y. (ed.), *Recent Developments in Statistical Data Analysis and Inference*. Elsevier Science Publisher, North-Holland, Amsterdam, 459-474.
- [5] Cuadras, C. M., A. Arenas and J. Fortiana (1996) Some Computational Aspects of a Distance-Based Model for prediction. *Communications in Statistics B. Simulation and Computation* **25**, 593–609.

Restricted Minimum Phi-Divergence Estimators in Logistic Regression Models

Julio Angel Pardo

japardo@mat.ucm.es, Department of Statistic and Operational
Research (I). Complutense University of Madrid

Abstract

In this paper we introduce and study a new family of estimators for the problem of testing general linear hypotheses. This family is based on the ϕ -divergence measure and includes as special case the classical maximum likelihood estimator.

Keywords: Logistic regression, Restricted minimum ϕ -divergence estimator, General linear hypotheses.

1. Introduction

We consider the logistic regression model

$$\text{logit}(\pi(\mathbf{x}_i^T \boldsymbol{\beta})) = \beta_0 + \sum_{j=1}^k \beta_j x_{ij}, \quad i = 1, 2, \dots, I$$

where $\text{logit}(p) = \log(p/(1-p))$, $\mathbf{x}_i = (1, x_{i1}, \dots, x_{ik})^T$ is the vector of explanatory variables, $\boldsymbol{\beta} \in \Theta = \{(\beta_0, \dots, \beta_k)^T : \beta_j \in \mathbb{R}, j = 0, \dots, k\}$ and $\pi(\mathbf{x}_i^T \boldsymbol{\beta})$ is the probability of success in a Bernoulli probability. The explanatory “design” matrix is $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_I)^T$ and its rank is $k+1$.

We denote by $\hat{\mathbf{p}} = (n_{11}/N, n_{12}/N, n_{21}/N, n_{22}/N, \dots, n_{I1}/N, n_{I2}/N)^T$, where n_{i1} is the number of “success” associated with a binomial random variable of parameters n_i and $\pi(\mathbf{x}_i^T \boldsymbol{\beta})$, $i = 1, \dots, I$, $N = \sum_{j=1}^I n_j$, by

$$\mathbf{p}(\boldsymbol{\beta}) = (\pi(\mathbf{x}_1^T \boldsymbol{\beta})n_1/N, 1 - \pi(\mathbf{x}_1^T \boldsymbol{\beta})n_1/N, \dots, \pi(\mathbf{x}_I^T \boldsymbol{\beta})n_I/N, (1 - \pi(\mathbf{x}_I^T \boldsymbol{\beta}))n_I/N)^T,$$

and by $\boldsymbol{\beta}^0$ the true value of parameter. Pardo, J. A. et al. (2005) considered the problem of estimating the parameter $\boldsymbol{\beta} = (\beta_0, \dots, \beta_k)^T$ using the minimum ϕ -divergence estimator, i.e., $\hat{\boldsymbol{\beta}}_\phi \equiv \arg \min_{\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_k} D_\phi(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\beta}))$, with

$$\begin{aligned} D_\phi(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\beta})) &= \sum_{i=1}^I \frac{n_i}{N} \left\{ \pi(\mathbf{x}_i^T \boldsymbol{\beta}) \phi(n_{i1}/\pi(\mathbf{x}_i^T \boldsymbol{\beta})n_i) \right. \\ &\quad \left. + (1 - \pi(\mathbf{x}_i^T \boldsymbol{\beta})) \phi(n_{i2}/(1 - \pi(\mathbf{x}_i^T \boldsymbol{\beta}))n_i) \right\}, \end{aligned}$$

where ϕ is a convex function for $x > 0$ such that $\phi(1) = 0, \phi''(1) > 0$, $0\phi(0/0) = 0$ and $0\phi(p/0) = \lim_{x \rightarrow 0} x\phi(p)/0 = \lim_{u \rightarrow \infty} \phi(up)/u$. For more details about ϕ -divergences see Pardo (2006).

The main purpose of this work is to develop, on the basis of the ϕ -divergence, the theory for estimating the parameters under the general hypothesis

$$H_0 : \mathbf{K}^T \boldsymbol{\beta} = \mathbf{m}, \quad (2.52)$$

where $\boldsymbol{\beta}$ is the $(k+1)$ -vector of unknown parameters in the logistic regression model; \mathbf{K}^T is any matrix of r rows and $k+1$ columns that it must have full row rank and \mathbf{m} is a vector, of order r of specified constants.

2. Restricted phi-divergence estimators

For estimating $\boldsymbol{\beta}$ under the hypothesis given in (1) we can define the minimum ϕ -divergence estimator restricted to $\Theta_0 = \left\{ \boldsymbol{\beta} \in \Theta : \mathbf{K}^T \boldsymbol{\beta} = \mathbf{m} \right\}$ as $\widehat{\boldsymbol{\beta}}_{\phi}^{\Theta_0} \equiv \arg \min_{\boldsymbol{\beta} \in \Theta_0} D_{\phi}(\widehat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\beta}))$. The decomposition of $\widehat{\boldsymbol{\beta}}_{\phi}^{\Theta_0}$ is obtained in the following theorem,

Theorem 1. It holds,

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_{\phi}^{\Theta_0} = & \boldsymbol{\beta}^0 + \mathbf{H}(\boldsymbol{\beta}^0) (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \text{diag}((\mathbf{C}_i^0)_{i=1,\dots,I}^T) \\ & \times \text{diag}(\mathbf{p}(\boldsymbol{\beta}^0)^{-1/2})(\widehat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\beta}^0)) + o(\|\widehat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\beta}^0)\|) \end{aligned}$$

where $\mathbf{W} = \text{diag}((\mathbf{C}_i^0)_{i=1,\dots,I}^T) \text{diag}\left(((\mathbf{C}_i^0)_{i=1,\dots,I})\right)$,

$$\mathbf{C}_i^0 = \left(\frac{n_i}{N} \pi(\mathbf{x}_i^T \boldsymbol{\beta}^0) (1 - \pi(\mathbf{x}_i^T \boldsymbol{\beta}^0)) \right)^{1/2} ((1 - \pi(\mathbf{x}_i^T \boldsymbol{\beta}^0))^{1/2}, -\pi(\mathbf{x}_i^T \boldsymbol{\beta}^0)^{1/2})^T,$$

$$i = 1, \dots, I, \text{ and } \mathbf{H}(\boldsymbol{\beta}^0) = \mathbf{I} - (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{K} (\mathbf{K}^T (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{K})^{-1} \mathbf{K}^T.$$

Based on this result we have the asymptotic distribution of $\widehat{\boldsymbol{\beta}}_{\phi}^{\Theta_0}$,

Theorem 2.

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}_{\phi}^{\Theta_0} - \boldsymbol{\beta}^0) \xrightarrow[N \rightarrow \infty]{L} \mathcal{N}(\mathbf{0}, (\mathbf{X}^T \mathbf{W}^* \mathbf{X})^{-1} \mathbf{S} (\mathbf{X}^T \mathbf{W}^* \mathbf{X})^{-1}),$$

where $\mathbf{S} = (\mathbf{I} - \mathbf{K} (\mathbf{K}^T (\mathbf{X}^T \mathbf{W}^* \mathbf{X})^{-1} \mathbf{K})^{-1} \mathbf{K}^T)$ and $\mathbf{W}^* = \lim_{N \rightarrow \infty} \mathbf{W}$.

Acknowledgements. This work was supported by Grants BMF2003-00892 and UCM2005-910707.

3. Bibliography

- [1] Pardo, J.A., Pardo, L. and Pardo, M.C. (2005). Minimum ϕ -divergence estimator in logistic regression models. *Statistical Papers*, **47**, 91-108.
- [2] Pardo, L. (2006). *Statistical Inference Based on Divergence Measures*. Chapman & Hall..

Robustez bayesiana de la tasa de falsos positivos (positive false discovery rate)

Julián de la Horra¹

¹julian.delahorra@uam.es, Departamento de Matemáticas, Universidad Autónoma de Madrid

Abstract

La tasa de falsos positivos (positive false discovery rate) fue introducida por Storey (2003) como una alternativa a la probabilidad de obtención de uno o más falsos positivos (corrección por Bonferroni) para el caso en que estamos contrastando, simultáneamente, una gran cantidad de hipótesis. Esta tasa de falsos positivos tiene una bonita interpretación bayesiana (como fue probado por Storey (2003)) y aquí analizamos su robustez. La clase de distribuciones a priori de epsilon-contaminación es considerada (ya que ésta es una de las más utilizadas en los estudios de robustez bayesiana) obteniéndose falta de robustez cuando la a priori básica concentra la probabilidad sobre la hipótesis nula.

Palabras Clave: tasa de falsos positivos, robustez bayesiana, clase de epsilon-contaminación

AMS: 62F15

1. Bibliografía

- [1] Storey, J.D. (2003). The positive false discovery rate: a Bayesian interpretation and the q-value. *Annals of Statistics* 31, 2013-2035.

Selección bayesiana de modelos lineales espacio-temporales autorregresivos

M^a Asunción Beamonte¹, Pilar Gargallo², Manuel Salvador³

¹asunbea@unizar.es, ²pigarga@unizar.es, ³salvador@unizar.es

Facultad de Ciencias Económicas y Empresariales, Universidad de Zaragoza

Abstract

En este trabajo se plantea un proceso bayesiano de selección de modelos espacio-temporales autorregresivos (STAR) utilizando métodos MCMC. La metodología se ilustra mediante una aplicación al mercado inmobiliario de Zaragoza.

Palabras Clave: STAR, Inferencia Bayesiana, Selección de Modelos

AMS: 62F15, 91B72

1. Introducción

Una gran cantidad de datos suelen venir organizados por unidades de tiempo tales como meses, trimestres, años así como por zonas geográficas tales como provincias, regiones, países. Así, muchos de los datos económicos, médicos, medio-ambientales o sociológicos recogidos por los organismos oficiales caen dentro de esta categoría. La creciente capacidad de los sistemas de información y, especialmente, los sistemas de información geográficos (GIS) han facilitado la construcción de grandes bases de datos con información muy desagregada que tiene referencias espacio-temporales muy precisas.

En este contexto se hace necesario la implementación de técnicas estadísticas que permitan la construcción de modelos que describan, de forma adecuada y parsimoniosa, las dependencias espacio-temporales existentes en los datos. De este modo se podrán realizar inferencias más precisas para interpretar adecuadamente la información disponible así como elaborar predicciones más acuradas.

Dos son las dificultades más habituales con las que se enfrentan los analistas de este tipo de problemas: por un lado el alto grado de no linealidad asociado a los parámetros que describen las dependencias espaciales y, por el otro, los elevados tamaños muestrales de los conjuntos de datos que dificultan los procesos de estimación.

Recientemente Pace y otros [1] han propuesto una familia de modelos lineales autorregresivos que permite capturar, de forma parsimoniosa, una gran cantidad de efectos espacio-temporales de carácter local, cuya estimación no es computacionalmente muy costosa al estar basada en procedimientos OLS.

Dichos autores aplican su metodología al análisis de precios de viviendas obteniendo mejores resultados que técnicas más complejas que manejan un número de parámetros mucho mayor.

Sin embargo, su procedimiento de selección de modelos está basado en técnicas de estimación máximo-verosímil cuyos resultados asintóticos no suelen ser aplicables (Stein [2]). En este trabajo abordamos dicho problema utilizando técnicas bayesianas de selección de modelos (Banerjee y otros [3]) basadas en la utilización de métodos MCMC. Estas técnicas no dependen de resultados asintóticos para su implementación y tienen en cuenta el grado de parsimoniosidad de los modelos propuestos, lo cual redunda en una mayor robustez y acuracidad de los resultados obtenidos.

La metodología se ilustra analizando datos del Mercado Inmobiliario de Zaragoza.

2. Bibliografía

- [1] Pace R.K.; Barry, R.; Gilley, O.W. y Sirmans, C.F. (2000). A method for spatial-temporal forecasting with an application to real estate prices. *International Journal of Forecasting* 16, 229-246.
- [2] Stein, M.L. (1999). Predicting Random Fields with Increasingly Dense Observations. *Annals of Applied Probability*, 9, 242-273.
- [3] Banerjee, S.; Carlin, B.P. y Gelfand, A.E. (2003). *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall.

Selección de transformaciones en la predicción de series temporales con el modelo de Holt

Jose D. Bermúdez¹, Ana Corberán², Enriqueta Vercher³

¹bermudez@uv.es, Departamento de Estadística e Investigación Operativa,
Universidad de Valencia

²ana.corberan@uv.es, Departamento de Estadística e Investigación Operativa,
Universidad de Valencia

³Enriqueta.Vercher@uv.es, Departamento de Estadística e Investigación
Operativa, Universidad de Valencia

Abstract

El objetivo del trabajo es obtener buenas predicciones de valores futuros de una serie temporal sin componente estacional mediante el modelo de Holt, realizando una estimación conjunta de los parámetros de suavizado y condiciones iniciales y seleccionando previamente la transformación de los datos a emplear para obtener las predicciones mediante técnicas jack-nife.

Palabras Clave: Predicción, Método de Holt, Transformaciones.

1. Introducción

Los procedimientos de suavizado exponencial son reconocidos como una de las herramientas más empleadas en la predicción a corto plazo de series temporales univariantes.

La elección adecuada de los valores de los parámetros de suavizado y de las condiciones iniciales es crucial en la obtención de predicciones, sin embargo, su obtención es una cuestión que todavía está por determinar. Recientemente, se ha propuesto la estimación conjunta de los parámetros de suavizado y de las condiciones iniciales mediante la estimación máximo verosímil (Ord et al., 1997; Hyndman et al., 2002) y otros métodos de estimación (Harvey, 1989; Segura and Vercher, 2002; Bermúdez et al., 2006).

En el trabajo se presenta una nueva formulación del modelo de Holt con errores aditivos, donde la componente estocástica es introducida mediante errores independientes, homocedásticos y con distribución Normal. El vector de datos sigue una distribución Normal multivariante, con vector de medias y matriz de varianzas-covarianzas dependiendo de los parámetros de suavizado y de las condiciones iniciales. Es entonces cuando puede mostrarse que el modelo de Holt puede formularse como un modelo lineal heterocedástico, cuyos coeficientes vienen dados por las condiciones iniciales y con matriz de varianzas-covarianzas dependiendo de los parámetros de suavizado. Esta formulación permite optimizar conjuntamente tanto los parámetros de suavizado como las condiciones iniciales.

Puede ocurrir en algunas ocasiones que el modelo de Holt con errores aditivos no sea adecuado para los datos que forman la serie temporal, pero sí lo sea para alguna transformación de éstos, por ejemplo la logarítmica en el caso en que los datos se ajusten al modelo de Holt con errores multiplicativos. En estos casos, resultaría más conveniente obtener predicciones de la serie transformada y a continuación, aplicar la transformación inversa, obteniendo así predicciones en la escala original. Cuando es aconsejable trabajar con transformaciones de los datos es necesario disponer de un mecanismo que permita decidir cual es la transformación a aplicar.

El procedimiento que generalmente se ha estado empleando hasta el momento es seleccionar aquella transformación con menor error de ajuste en la zona de estimación, el cual es obtenido a partir de los errores de predicción de una etapa. Este mecanismo no siempre selecciona la transformación adecuada, pues la transformación con menor error en la zona de estimación no es siempre la que proporciona las predicciones más precisas.

Nos hemos planteado desarrollar un mecanismo alternativo en la selección de la transformación a emplear para obtener predicciones basado en técnicas jackknife, en concreto la conocida como leaving-one-out: Supondremos conocidos todos los datos de la serie temporal excepto el que ocupa la posición m . A continuación, veremos cual sería su predicción con el método de Holt y con los datos disponibles hasta ese momento, $y_{ii=1,\dots,m-1}$, obteniendo así el error de predicción $e_m = y_m - \hat{y}_m$. Haciendo variar m en el rango $n-k+1, \dots, n$, es decir, repitiendo el proceso para los últimos k datos, obtendremos k errores de predicción La transformación de los datos que tomaremos para obtener las predicciones será aquella cuyo error de predicción de los k datos supuestos desconocidos sea menor.

Notar que para obtener la predicción del dato m , supuesto desconocido, necesitamos un procedimiento que nos permita previamente obtener las estimaciones de los parámetros del modelo de Holt cuando hay datos faltantes. Dicha herramienta será el algoritmo EM (Dempster, Laird y Rubin, 1977).

Por último, aplicaremos este procedimiento para predecir las 645 series anuales de la competición M3 con el fin de valorar las predicciones obtenidas mediante nuestro esquema de trabajo y compararlas con otros resultados ya publicados (Makridakis and Hibon, 2000).

2. Agradecimientos

Parcialmente subvencionado por el Ministerio de Educación y Ciencia TIN2005, 08404-C04-04 y por la Generalitat Valenciana CTBPRB/2005/006.

Selection of Variables for Multivariate Control Charts

Isabel González¹, Ismael Sánchez²

¹Departamento de Ing. Mecánica, Universidad Carlos III de Madrid

²Departamento de Estadística, Universidad Carlos III de Madrid

Abstract

En este trabajo propone una método para seleccionar las variables que han de monitorizarse en un gráfico de control multivariante. El método selecciona las variables que son más sensibles en términos de capacidad de detección de una situación fuera de control.

Keywords: Factor analysis; Process simulation; Quality control; Statistical process control, Variable selection

1. Introduction

In order to assure the quality of a manufactured article, a final inspection is usually established. In this final inspection, a set of key variables of the article are evaluated. We will denote this set of key variables as final inspection variables (FIV). The definition of the FIV should be made under agreement between the manufacturer and the client who will eventually use the manufactured article. The definition of the FIV is also made in conjunction with their tolerance limits, beyond them the article can be rejected. The definition of the FIV is then an important issue, since they must reflect the real quality of the article. As a result, the selection of the FIV is both a managerial and a technical problem. Depending on the complexity of the article, the dimension of the FIV (i.e., the number of variables to be inspected) can be large, which in turn will determine the frequency and the cost of the final inspection.

On the other hand, in order to perform an efficient monitoring of the process, a number of variables of the article should also be selected. We will denote this variables as process control variables (PCV). The PCV do not need to be the same as the FIV. Although process control and quality control are closely related, they have to satisfy different goals. In process control, the goal is to maintain the production process at some stationary conditions. Then, the selection of the PCV should guarantee the detection of changes in the process, even though such changes do not necessarily lead to the production of defective articles. Both the design of the process control and the definition of the PCV are, in a high degree, statistical problems. Of course, PCV and FIV will not be completely independent, since a satisfactory process control should guarantee a successful final inspection. Therefore, it would be reasonable, that the PCV are some selected subset of the FIV.

In this article, we just propose a methodology to select a reduced number of variables out of the FIV to be used as PCV. In order to assure that the client agrees with the final quality, it is quite reasonable to measure as many variables as possible to avoid omitting any that might prove critical. As a results, the FIV would involve much more variables than strictly necessary. Consequently, the dimensionality of the FIV can be reduced easily without disturbing the overall features of the inspected article. It is just this kind of dimension reduction what we pursue in order to choose the PCV from the FIV. This selection will allow a faster and a cheaper data collection than the required with the full FIV. Then, the selected PCV will be amenable to be used for statistical process control (SPC) charts.

The proposed variable selection methodology can be summarized in two steps: (1) selection step, and (2) evaluation step. In the selection step we will have to decide the subset of $p \leq K$ variables that will be used to build the control chart. The principal component analysis (PCA), factor analysis (FA) and multivariate regression are used in order to select the PCV. In the evaluation step, we will estimate the performance of the control chart based on those p variables, comparing it with the performance of using all the K variables. The performance of a control chart will be measured using the so called Average Run Length (ARL)

2. Contact

Address:

Ismael Sánchez
 Department of Statistics, University Carlos III of Madrid
 Av. Universidad 30
 28911 - Leganés (Madrid)
 Spain

Telephone: (+34) 916249179

Fax: (+34) 916249430

e-mail: ismael@est.econ.uc3m.es

Sensibilidad en Redes Bayesianas Gausianas mediante una Medida de Divergencia

Miguel A. Gómez-Villegas¹, Paloma Maín², Rosario Susi³

¹ma gv@mat.ucm.es, Departamento de Estadística e Investigación Operativa,
Universidad Complutense de Madrid

²pmain@mat.ucm.es, Departamento de Estadística e Investigación Operativa,
Universidad Complutense de Madrid

³rsusi@estad.ucm.es, Departamento de Estadística e Investigación Operativa
III, Universidad Complutense de Madrid

Claves: Red Bayesiana Gausiana, Divergencia de Kullback-Leibler, Análisis de Sensibilidad.

AMS: 62F15.

Abstract

En este trabajo se desarrolla un método para realizar un análisis de sensibilidad de una Red Bayesiana Gausiana. Se define una medida basada en la divergencia de Kullback-Leibler, que servirá para evaluar el impacto, sobre la variable de interés, de cambios en la red inicial.

1. Introducción

Una Red Bayesiana (RB) es un modelo gráfico probabilístico formado por dos partes, una parte cualitativa y otra cuantitativa. Así, se define una RB como un par $(\mathcal{G}, \mathcal{P})$ en el que \mathcal{G} es un grafo acíclico dirigido (DAG) cuyos nodos representan las variables aleatorias del modelo $\mathbf{X} = \{X_1, \dots, X_n\}$ y las aristas representan las dependencias probabilísticas entre las mismas (parte cualitativa) y \mathcal{P} es un conjunto de densidades de probabilidad condicionadas $p(x_i|pa(x_i))$ donde $pa(x_i)$ es el conjunto de padres del nodo X_i en \mathcal{G} (parte cuantitativa). Además la densidad de probabilidad conjunta de \mathbf{X} se obtiene como el producto de las densidades de probabilidad condicionadas.

Una RB Gausiana se caracteriza por tener una densidad de probabilidad conjunta, asociada al conjunto de variables \mathbf{X} , normal multivariante $N(\mu, \Sigma)$ donde μ es el vector de medias n -dimensional y Σ es la matriz de covarianzas.

Es habitual que en una RB exista una variable de interés X_i . Así, cuando la RB está definida y se tiene evidencia acerca de alguna otra variable $X_e \in \mathbf{X}$ se propaga la evidencia por la red y se estudia el impacto de dicha evidencia en la variable de interés. Se han diseñado diversos algoritmos para propagar la evidencia en RB. Una forma sencilla consiste en obtener la densidad de probabilidad condicionada dada la evidencia; así, sea \mathbf{X} el conjunto de variables del problema y $X_e \in \mathbf{X}$ la variable evidencial, tras propagar la evidencia

$E = \{X_e = e\}$, la densidad de probabilidad conjunta de $\mathbf{Y} = \mathbf{X} \setminus E$ es normal multivariante de parámetros $\mu^{\mathbf{Y}|E=e} = \mu_{\mathbf{Y}} + \Sigma_{YE} \frac{1}{\sigma_e}(e - \mu_E)$ y $\Sigma^{\mathbf{Y}|E=e} = \Sigma_{\mathbf{YY}} - \Sigma_{YE} \frac{1}{\sigma_e} \Sigma_{EY}$ donde $\frac{1}{\sigma_e}$ es la inversa de la varianza de X_e .

2. Análisis de sensibilidad

Definir la parte cuantitativa de una RB es una tarea complicada, debido al gran número de parámetros del problema, y en ocasiones, se puede estar incurriendo en una asignación inadecuada de los mismos. Para conocer el efecto de una asignación incorrecta de los parámetros es fundamental realizar un análisis de sensibilidad. Autores como Laskey (1995), que desarrolla un análisis de sensibilidad en RB Discretas, o Castillo, et al. (2003), que describen la propagación simbólica en RB Gausianas, han estudiado la sensibilidad de los resultados a imprecisiones en los parámetros.

En este trabajo se estudia el efecto de parámetros inadecuados sobre la densidad de la variable de interés tras propagar la evidencia. Para ello, se define la inexactitud de un parámetro, antes de propagar la evidencia, añadiendo un *factor de desviación* δ al parámetro inadecuado. Tras la propagación, se evalúa el efecto de δ , comparando la densidad a posteriori de la variable de interés sin introducir el factor δ en la red inicial $f(x_i|e)$ e introduciéndolo $f(x_i|e, \delta)$, mediante la *medida de sensibilidad*, definida apoyándonos en la divergencia de Kullback-Leibler (1951). Se obtiene la medida de sensibilidad considerando inexacto cada uno de los parámetros de la red y se estudian los casos en los que X_i se ve afectado por δ . Se comparan gráficamente las medidas obtenidas en función de δ , de forma que si el valor de la medida de sensibilidad es grande se ha de revisar la información acerca del parámetro, y en caso contrario, el resultado no es sensible a dicha imprecisión.

Una aplicación práctica ilustra los conceptos y métodos presentados.

3. Bibliografía

- [1] Castillo, E. and Kjaerulff, U. (2003). Sensitivity analysis in Gaussian Bayesian networks using a symbolic-numerical technique. *Reliability Engineering and System Safety* 79,139-148.
- [2] Kullback, S. and Leibler, R.A. (1951). On Information and Sufficiency. *Annals of Mathematical Statistics* 22,79-86.
- [3] Laskey, K.B. (1995). Sensitivity Analysis for Probability Assessments in Bayesian Networks. *IEEE transactions on systems, man, and cybernetics*, 25(6),901-909.

Sistema integrado de control estadístico y regulación automática de un proceso continuo de polimerización. Análisis de robustez del controlador predictivo DMC.

S. Barcelo¹, J. Sanchis², A. Ferrer³

¹sbarcelo@eio.upv.es, Departamento de Estadística e Investigación Operativa Aplicadas y Calidad, Universidad Politécnica de Valencia

²jsanchis@isa.upv.es, Departamento de Ingeniería de Sistemas y Automática, Universidad Politécnica de Valencia

³aferrer@eio.upv.es, Departamento de Estadística e Investigación Operativa Aplicadas y Calidad, Universidad Politécnica de Valencia

Abstract

Se analiza la robustez de un controlador DMC integrado en un sistema de control estadístico y regulación automática de un proceso continuo de polimerización, frente a la presencia de discrepancias en el modelo utilizado para diseñarlo. Se implementa un sistema de monitorización estadística multivariante para detectar estas discrepancias eficientemente.

Palabras Clave: Control estadístico multivariante de procesos; Control ingenieril multivariante de procesos; Robustez de controladores DMC.

AMS: 62M10, 62H25, 62P30

1. Introducción

El MESPC (*Multivariate Engineering Statistical Process Control*) es una estrategia de control que integra el control automático o ingenieril de procesos (MEPC) y el control estadístico de procesos (MSPC) en un contexto multivariante. Ambas estrategias de control, aunque se han desarrollado de forma independiente, son en realidad complementarias. El MEPC se utiliza para reducir la variabilidad debida a causas comunes ante observaciones dinámicamente relacionadas mediante ajustes en variables compensatorias, basándose en algoritmos de control que aprovechan la relación dinámica que existe entre las variables del proceso. Por el otro lado, el MSPC tiene como objetivo general reducir la variabilidad mediante la detección, identificación y eliminación de causas especiales de variabilidad cuando las distribuciones de las variables a controlar son estables.

Se ha desarrollado un sistema de control integrado MESPC [1] tomando ideas de ambas filosofías de control en un proceso de polimerización en continuo (polietileno de alta densidad) correspondiente a una industria petroquímica española con el objetivo de asegurar tanto la optimización como la mejora del proceso. Se ha identificado y estimado un modelo que permite caracterizar la dinámica del proceso a partir de datos obtenidos en bucle cerrado. Se

ha aplicado el MPC (*Model Predictive Control*) para desarrollar la parte de regulación automática del proceso y se ha derivado el controlador DMC (*Dynamic Matrix Control*) [2].

2. Análisis de robustez del controlador DMC

En el contexto de la filosofía MESPC, el presente trabajo se centra en el estudio de la robustez y de las prestaciones del controlador predictivo DMC, que dependen de la capacidad del modelo utilizado para reflejar el comportamiento dinámico del proceso. En una situación real, ningún modelo puede exactamente imitar al proceso físico real, y por lo tanto, debido a las diferencias entre el proceso y el modelo pueden aparecer problemas relacionados con la estabilidad y prestaciones del sistema. Se pretende determinar la región de estabilidad y el comportamiento del controlador ante las discrepancias mencionadas. Para ello se simularán cambios en el modelo del proceso y se analizará la robustez.

3. Sistema de monitorización estadística del proceso

Aunque la robustez del controlador sea buena y la regulación automática del proceso acabe compensando las discrepancias surgidas entre el modelo y el proceso, si éstas permanecen sin detectarse, el error cuadrático medio puede aumentar y puede producirse producto fuera de especificaciones. Por estas razones se propone implementar un sistema de control estadístico de proceso que permita detectar estos cambios rápidamente. Dicho sistema se ha desarrollado aplicando métodos proyectivos multivariantes [3]. Para ello se ha estimado un modelo PCA a partir de variables registradas en el proceso cuando éste funciona bajo control y se han construido gráficos de control, cuyos estadísticos de monitorización son la T^2 de Hotelling y el SPE (*Squared Prediction Error*) que permiten monitorizar el comportamiento futuro del proceso. En el trabajo se muestra cómo esta componente del sistema de control integrado MESPC permite detectar estos cambios eficientemente.

4. Bibliografía

- [1] Barceló, S.; Sanchis, J.; Ferrer, A. (2005). An integrated multivariate engineering statistical process control system in a continuous polymerization process. 5th European Network Business and Industrial Statistics Conference. Newcastle (UK).
- [2] Cutler, C. R.; Ramaker, B. L. (1979). "Dynamic Matrix Control- a Computer Control Algorithm". AICHE annual meeting. Houston, TX.
- [3] Kourtzi, T.; MacGregor, J. F. (1996). "Multivariate SPC Methods for Process and Product Monitoring". *Journal of Quality Technology* 28 (4), 409-428.

Sobre los estimadores intrínsecos

G. García¹, J.M. Oller²

¹gloriam.garcia@upf.edu, Departamento de Economía y Empresa, Universidad Pompeu Fabra

²joller@ub.edu, Departamento de Estadística, Universidad de Barcelona

Abstract

En este trabajo se aplican algunos resultados básicos del análisis intrínseco de la estimación estadística a diferentes ejemplos. Se examina el estimador intrínseco obtenido minimizando el riesgo Bayesiano cuando se considera el cuadrado de la distancia informativa como pérdida intrínseca y a la uniforme de Jeffreys como distribución a priori de referencia, comparando los resultados obtenidos con los estimadores intrínsecos desarrollados por Bernardo [1].

Claves: Análisis Intrínseco, Métrica Informativa.

1. Introducción

La estimación estadística tiene como objetivo determinar un mecanismo probabilístico que explique las observaciones asociadas a un cierto fenómeno. La estimación estadística paramétrica asume estas probabilidades como elementos de una familia indexadas por un parámetro.

Sin embargo, un procedimiento de estimación no debería depender de la parametrización elegida para el modelo. En otras palabras: sería conveniente que todo estimador cumpliera la propiedad de *invariancia funcional*, siendo de esta forma un estimador real del mecanismo probabilístico generador de las observaciones.

2. Propiedades intrínsecas de un estimador

El sesgo y el error cuadrático medio son ampliamente aceptados como medidas de la bondad de un estimador. Sin embargo estos criterios son dependientes de la parametrización escogida para el modelo y, en consecuencia, las propiedades de mínimo riesgo y ausencia de sesgo son *no intrínsecas*.

Resulta necesario examinar la bondad de un estimador a través de criterios *intrínsecos*: este es el objeto del *Análisis intrínseco de la estimación estadística* introducido en Oller y Corcuera [3] (ver también Oller [4] y [5]). Estos trabajos consideran medidas intrínsecas del error cuadrático y del sesgo basadas en la estructura Riemanniana de un modelo estadístico paramétrico suficientemente regular. Concretamente, la pérdida cuadrática considerada es el cuadrado de la correspondiente distancia Riemanniana y el sesgo se redefine a través de un campo vectorial basado en las propiedades geométricas del modelo.

En el trabajo que se presenta, junto a una revisión de los elementos básicos de la estimación intrínseca, se examinan diferentes estimadores obtenidos tras minimizar el riesgo Bayesiano cuando la función de pérdida elegida es el cuadrado de la distancia informativa. La distribución a priori escogida es la uniforme de Jeffreys; esta distribución a priori, habitualmente impropia, es proporcional al volumen Riemanniano, ver Jeffreys [2]. En cada caso se compara el correspondiente estimador con el obtenido por Bernardo en [1]: en este trabajo se introduce el concepto de estimación intrínseca considerando el estimador que minimice el riesgo Bayesiano en base a una a priori de referencia que coincide, en algunas situaciones, con la a priori uniforme de Jeffreys y estableciendo como pérdida una versión simetrizada de la divergencia de Kullback-Leibler.

3. Contactar

Dirección: Gloria García

Departamento de Economía y Empresa
 Edificio Jaume I - Ramon Trias Fargas, 25-27
 08005 - Barcelona
 España

e-mail: gloriam.garcia@upf.edu

4. Bibliografía

- [1] Bernardo, J.M. and Juárez, M.A. (2003). Intrinsic Estimation. In: Bernardo, Bayarri, Berger, Dawid, Hackerman, Smith and West (eds.), *Bayesian Statistics 7*. Oxford University Press, 465-476.
- [2] Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proc. Royal Soc.* 186 A, 453-461.
- [3] Oller, J.M. and Corcuera, J.M. (1995). Intrinsic analysis of statistical estimation. *Ann. Statist.* 23(5), 1562-1581.
- [4] Oller, J.M. (1993). On an Intrinsic Bias Measure. In: Kalasnikov and Zolotarev (eds.), *Stability Problems for Stochastic Models*. Springer-Verlag, 134-158.
- [5] Oller, J.M. (1993). On an Intrinsic analysis of statistical estimation. In: Cuadras and Rao (eds.), *Multivariate Analysis: Future Directions 2*. Elsevier science publishers B. V., 421-437.

Software JavaScript para comparación de r curvas de supervivencia

Emilio Letón¹, Pilar Zuluaga²

¹Departamento de Estadística, Universidad Carlos III de Madrid,
eleton@est-econ.uc3m.es

²Departamento de Estadística e I.O., Facultad de Medicina Universidad
Complutense de Madrid, pilarzul@med.ucm.es

Abstract

En este artículo se presenta un software en JavaScript que calcula diversos tests de Puntuaciones y tests Ponderados para comparar r curvas de supervivencia. Además, se comparan dichos tests para el caso general de empates con desbalanceo entre los grupos.

Palabras Clave: Software, tests de puntuaciones, tests ponderados.

AMS: 62N03

1. Introducción

Los paquetes estadísticos más habituales para comparar r curvas de supervivencia sólo incluyen algunos tests de la familia de los Ponderados. Por otra parte la nomenclatura que aparece en el software no está unificada denominando con distinto nombre a un mismo test. Hay situaciones, donde es más aconsejable utilizar un test de la familia de Puntuaciones frente a uno Ponderado (Letón y Zuluaga (2001) y Letón y Zuluaga (2002)), por ello hemos elaborado un software que calcula 10 de ellos (entre Ponderados y Puntuaciones) y que unifica la nomenclatura. También presentamos su comparación mediante simulación, la cuál nos dará pautas de utilización.

2. Contrastes de hipótesis

Los tests Ponderados y de Puntuaciones para r grupos aparecen descritos entre otros en Peto y Peto (1972), Tarone y Ware (1977), Lee (1992), Letón y Zuluaga (2002) y Desu y Raghavarao (2004). En Letón y Zuluaga (2002) se puede ver cómo reescribir la habitual expresión matricial de dichos tests en otra no matricial, también se presentan las relaciones entre cinco tests Ponderados y sus correspondientes tests de Puntuaciones. Como caso particular de uno de estos tests aparece el tests de Kruskal-Wallis. Otros tests reciben el nombre de Wilcoxon generalizados, la justificación de ello es fácil de ver siguiendo las relaciones dadas en Letón y Zuluaga (2005) para dos grupos. Estas relaciones justifican las denominaciones de los tests en el software comercial.

3. Resultados computacionales

En Letón y Zuluaga (2002) se puede ver una comparación de los tests del apartado anterior para el caso de igual tamaño muestral y r grupos. En este trabajo presentamos simulaciones adicionales para el caso de desbalanceo entre grupos que permiten identificar escenarios de idoneidad de utilización de cada test.

4. Software JavaScript

Sólo una parte de los tests estadísticos (tests Ponderados) del apartado anterior aparecen en el software comercial habitual, por ello hemos elaborado programas en JavaScript para Intenet que permiten calcular todos ellos sin más que incorporar los datos desde un fichero ASCII (dichos programas están a disposición de los interesados, bajo petición directa a los autores). Los resultados del apartado anterior nos guiarán a la hora de elegir el más adecuado cuando se resuelve un caso real.

5. Bibliografía

- [1] Letón, E. and Zuluaga, P. (2001). Equivalence between score and weighted tests for survival curves. *Communications in Statistics: Theory and Methods* 30 , 591-608.
- [2] Letón, E. and Zuluaga, P. (2002). Survival tests for r groups. *Biometrical Journal* 44 , 15-27.
- [3] Peto, R. and Peto, J. (1972). Asymptotically efficient rank invariant test procedures (with discussion). *Journal of the Royal Statist.Soc A* 135, 185-207.
- [4] Tarone, R.E. and Ware, J. (1977). On distribution-free tests for equality of survival distributions. *Biometrika* 64, 156-160.
- [5] Lee, E.T. (1992). *Statistical methods for survival data analysis*. John Wiley and Sons.
- [6] Desu, M.M. and Raghavarao D. (2004). *Nonparametric statistical methods for complete and censored data*. Chapman and Hall.
- [7] Letón, E. and Zuluaga, P. (2005). Relationships among tests for censored data. *Biometrical Journal* 47, 377-387.

Some Recent Contributions to Spatio-Temporal Modelling

Emilio Porcu¹

¹ Departamento de Matemáticas, Universitat Jaume I, Campus Riu Sec,
12071 Castellón de la Plana.
email:porcu@mat.uji.es

Abstract

In this work we propose new classes of nonseparable, stationary or nonstationary covariance functions for space time data. We show that through partial differentiation and integration of covariance functions it is possible to obtain new classes of space-time covariances.

Keywords: Completely Monotone Functions, Space-Time Covariance Functions, Space-Time Random Fields.

1. Introduction

The last years have confirmed the importance of construction of new classes of covariance functions for space-time data. Several contributions, coming from fields as diverse as mathematics, geology, geography, physics, statistics, are concerned with criteria of construction of functions which are positive definite on a subset of $\mathbb{R}^d \times \mathbb{R}$. In particular, recent literature has been mainly focused on *nonseparable* space-time covariance functions, which allow for interaction between the spatial and temporal component and avoid some unpleasant features concerned with *separable* ones, such us the fact that *separable* covariances allow for a very poor BLUP and suffer of a lack of differentiability (in the mean square sense) away from the origin, fact which causes unpleasant discontinuities of the associated correlation function.

In modern literature we can find several approaches to the construction of nonseparable covariances. In the first contributions (see, for instance, Kyriakidis and Journel, 1999) the focus is on extending spatial or temporal methods to spatio-temporal ones, considering the spatio-temporal dependence separately in most of the cases. These extension approaches could be classified into two main categories: the former considers a pure extension of multivariate spatial or temporal models. The latter is based on considering a univariate spatio-temporal process. Even if there are many points in common between the two approaches, some distinctions are necessary in order to see the opportunity of applying an approach despite to the other.

In this work we propose new classes of covariance functions taking into account some crucial aspects, such as anisotropy and nonstationarity. Besides we

inspect some important properties of some linear operators whenever applied to space-time, isotropic or anisotropic, covariance functions. As far as the new classes of covariance functions are concerned, we propose four main classes which can be classified as follows:

- 1** Nonseparable stationary space-time covariance functions obtained through mixture modelling;
- 2** Nonseparable, anisotropic stationary covariance functions obtained through Archimedean functionals;
- 3** Nonseparable stationary covariances in which zonally anisotropy is obtained through componentwise isotropy;
- 4** Nonseparable covariances which are stationary in time and nonstationary in space.

As far as the linear operators are concerned, we show that through partial derivation and integration of space time covariances it is possible to generate new models of nonseparable space time covariance functions. This results extend the celebre works of Matheron (1965) called *la Montée* and *la Descente*. In order to obtain these results, we decompose the spatial lag vector in a arbitrary number of zones of isotropy, using the same technique as in Fernández-Casal *et al.* (2003). The work ends with a critical discussion about the properties of the proposed classes of covariance functions.

2. Bibliography

1. Fernández-Casal, R., González-Manteiga, W. and Febrero-Bande, M. (2003). Flexible spatio-temporal stationary variogram models. *Statistics and Computing*, **13**, 127-136.
2. Kyriakidis, P.C. and Journel, A.G. (1999). Geostatistical space-time models: A review. *Mathematical Geology*, **31**, 651-684.
3. Matheron, G. (1965). *Les Variables Régionalisées et leur Estimation*, Masson, Paris.

Stability under products of sufficient and minimal sufficient σ -fields

J.E. Chacón, J. Montanero, A.G. Nogales, P. Pérez

Dpto. de Matemáticas, Universidad de Extremadura

Avda. de Elvas, s/n, 06071-Badajoz, SPAIN.

e-mail: nogales@unex.es

Abstract

Se prueba en esta comunicación que la clase de las sub- σ -álgebras suficientes (y la de las minimal suficientes) es estable frente al producto. Los resultados pueden usarse para construir ejemplos que pueden arrojar alguna luz sobre la relación entre los conceptos de suficiencia minimal e invarianza, problema propuesto en Hall, Wijsman, Ghosh (1965). El caso de estructuras exponenciales será considerado especialmente.

Keywords: minimal sufficiency, invariance, completeness.

AMS: 62A05

1. Introduction

From now on $(\Omega, \mathcal{A}, \mathcal{P})$ will denote a statistical experiment. Let \mathcal{B} be a sub- σ -field of \mathcal{A} .

\mathcal{B} is said to be minimal sufficient if it is sufficient and it is \mathcal{P} -contained in any other sufficient σ -field \mathcal{C} . In the dominated case it is known that there exists a privileged dominating probability (i.e., a probability measure of the form $P^* = \sum_{n=1}^{\infty} c_n P_n$ such that $c_n \geq 0$, $\sum_n c_n = 1$, $\{P_n : n \in \mathbb{N}\} \subset \mathcal{P}$ and $P \ll P^*$, $\forall P \in \mathcal{P}$); in this case, a sub- σ -field is minimal sufficient if and only if it is the least σ -field making measurable some versions of the densities dP/dP^* , $P \in \mathcal{P}$.

The σ -field \mathcal{B} is said to be complete if every real and \mathcal{B} -measurable statistic $g : (\Omega, \mathcal{B}) \rightarrow \mathbb{R}$ such that $E_P(g) = 0$, for all $P \in \mathcal{P}$, is \mathcal{P} -equivalent to 0 (i.e., $P(g \neq 0) = 0$, $\forall P \in \mathcal{P}$).

It is well known that every sufficient and complete σ -field is minimal sufficient. Landers and Rogge (1976) shows the stability of the class of complete σ -fields under products. These results can be found, for example, in Pfanzagl (1994) or Lehmann (1986), where the reader is referred for other concepts and results to be used below. Our theorem shows that the classes of sufficient and minimal sufficient σ -fields exhibit the same property.

Theorem 1 For $1 \leq i \leq n$, let \mathcal{B}_i be a sufficient (resp., minimal sufficient) σ -field in the statistical experiment $(\Omega_i, \mathcal{A}_i, \mathcal{P}_i)$. Then, the product σ -field $\prod_{i=1}^n \mathcal{B}_i$

is sufficient (resp., minimal sufficient) for the product statistical experiment

$$\left(\prod_{i=1}^n \Omega_i, \prod_{i=1}^n \mathcal{A}_i, \prod_{i=1}^n \mathcal{P}_i \right).$$

2. Contact

Address:

A.G. Nogales
 Departamento de Matemáticas
 Avda. de Elvas s/n
 06071 - Badajoz
 Spain

Telephone: (+34) 924289561

Fax: (+34) 924272911

e-mail: nogales@unex.es

www: <http://kolmogorov.unex.es/> nogales

3. Acknowledgement

This work has been partially supported by the Spanish Ministerio de Educación y Ciencia under the project MTM2005-06348.

4. Bibliography

- [1] Andersen E.B. (1967). On partial sufficiency and partial ancillarity. *Scand. Aktuarietidskr.*, 137–152.
- [2] Chacón J.E., Montanero J., Nogales A.G. and Pérez P. (2006). A note on minimal sufficiency. *Statistica Sinica* **16** (1).
- [3] Hall W.J., Wijsman R.A. and Ghosh J.R. (1965). The relationship between sufficiency and invariance with applications in sequential analysis. *Ann. Math. Statist.* **36**, 575–614.
- [4] Landers D. and Rogge L. (1976). A note on completeness. *Scand. J. Statist.* **3**, 139.
- [5] Lehmann E.L. (1986). *Testing Statistical Hypotheses*. 2nd ed. Wiley, New York.

Statistical Inference in Poisson Loglinear Models: An approach using Phi-divergences

Nirian Martín¹, Leandro Pardo²

¹nirian@estad.ucm.es, Statistics and O.R. III, Complutense U. of Madrid

²lpardo@mat.ucm.es, Statistics and O.R. I, Complutense U. of Madrid

Abstract

En este artículo se presentan procedimientos inferenciales para modelos loglineales de Poisson basados en medidas de Phi-divergencia.

Keywords: Poisson loglinear models, Minimum ϕ -divergence estimators, ϕ -divergence test statistics.

1. Introduction

Consider a table with k cells, containing counts, $\mathbf{n} \equiv (n_1, \dots, n_k)^T$. In what is to follow, we assume that \mathbf{n} is the result of Poisson sampling. Define $\mathbf{m}(\boldsymbol{\theta}) \equiv (m_1(\boldsymbol{\theta}), \dots, m_k(\boldsymbol{\theta}))^T \equiv E[\mathbf{n}]$. If we denote by $\log \mathbf{m}(\boldsymbol{\theta})$ the vector $(\log(m_1(\boldsymbol{\theta})), \dots, \log(m_k(\boldsymbol{\theta})))^T$, the loglinear model is defined by $\log \mathbf{m}(\boldsymbol{\theta}) = \mathbf{X}\boldsymbol{\theta}$, where \mathbf{X} is a known $k \times t$ full rank matrix such that $t < k$ and $\boldsymbol{\theta} \in \mathbb{R}^t$. The range of models is summarized by,

$$\mathcal{C}(\mathbf{X}) \equiv \left\{ \log \mathbf{m}(\boldsymbol{\theta}) : \log \mathbf{m}(\boldsymbol{\theta}) = \mathbf{X}\boldsymbol{\theta}; \boldsymbol{\theta} \in \mathbb{R}^t \right\}. \quad (2.53)$$

We can observe that $\mathcal{C}(\mathbf{X})$ represents the span of the columns of the matrix \mathbf{X} . An important assumption is that the $k \times 1$ dimensional vector $(1, \dots, 1)^T \in \mathcal{C}(\mathbf{X})$. For Poisson sampling, the total $\sum_{i=1}^k n_i$ is random; in this case, we denote $N = E\left[\sum_{i=1}^k n_i\right]$, the expected total.

2. Statistical Inference based on phi-divergences

It can be seen in Christensen (1997) that the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ in Poisson loglinear models can be obtained maximizing

$$\ell(\boldsymbol{\theta}, \mathbf{m}(\boldsymbol{\theta})) = \sum_{i=1}^k n_i \log m_i(\boldsymbol{\theta}) - \sum_{i=1}^k m_i(\boldsymbol{\theta}). \quad (2.54)$$

For more details see formula (5) in Christensen (1997, p. 399).

Now we denote by

$$D_\phi(\mathbf{n}, \mathbf{m}(\boldsymbol{\theta})) = \sum_{i=1}^k m_i(\boldsymbol{\theta}) \phi\left(\frac{n_i}{m_i(\boldsymbol{\theta})}\right), \quad \phi \in \Phi, \quad (2.55)$$

the ϕ -divergence measure between the expected vectors \mathbf{n} and $\mathbf{m}(\boldsymbol{\theta})$, where Φ is the class of all convex and differentiable functions $\phi : [0, \infty) \rightarrow \mathbb{R} \cup \{\infty\}$, such that at $x = 1, \phi(1) = \phi'(1) = 0, \phi''(1) > 0$, and at $x = 0, 0\phi(0/0) \equiv 0$ and $0\phi(p/0) \equiv \lim_{u \rightarrow \infty} \phi(u)/u$. More details about ϕ -divergence measures can be seen in Pardo (2006). The function $\phi(x) = x \log x - x + 1 \in \Phi$ gives the Kullback-Leibler divergence, $D_{Kull}(\mathbf{n}, \mathbf{m}(\boldsymbol{\theta}))$. It is immediate to see that

$$D_{Kull}(\mathbf{n}, \mathbf{m}(\boldsymbol{\theta})) = \sum_{i=1}^k n_i \log n_i - \sum_{i=1}^k n_i - \ell(\boldsymbol{\theta}, \mathbf{m}(\boldsymbol{\theta})).$$

Based on this expression maximizing $\ell(\boldsymbol{\theta}, \mathbf{m}(\boldsymbol{\theta}))$ is equivalent to minimizing $D_{Kull}(\mathbf{n}, \mathbf{m}(\boldsymbol{\theta}))$ and therefore we can define the maximum likelihood estimator, $\hat{\boldsymbol{\theta}}$, by

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta} \in \mathbb{R}^t} D_{Kull}(\mathbf{n}, \mathbf{m}(\boldsymbol{\theta})). \quad (2.56)$$

In this paper we study the behaviour of $\hat{\boldsymbol{\theta}}$ as well as the ϕ -divergence test statistics based on it to study the classical problems in Poisson loglinear models: goodness of fit, model checking and testing a nested sequence of hierarchical models.

3. Acknowledgement

This work was supported by Grants BMF2003-00892 and UCM2005-910707.

4. Bibliography

- [1] Christensen, R. (1997). *Log-Linear Model and Logistic Regression*. Springer-Verlag, New York.
- [2] Pardo, L. (2006). *Statistical Inference Based on Divergence Measures*. Statistics: Textbooks and Monographs. Chapman & Hall/CRC, New York.

STATMEDIA II: Sistema de elaboración y gestión de prácticas multimedia individualizadas de estadística

*M. Calvo, A. Villarroya, A. Miñarro, A. Arcas, S. Vives,
E. Vegas, F. Reverter,*

Departamento de Estadística. Universidad de Barcelona
Avda. Diagonal 645. 08028 Barcelona, España
mcalvo@ub.edu, aarcas@ub.edu

Abstract

El proyecto Statmedia II consiste en un sistema de elaboración y gestión de sesiones prácticas individualizadas de Estadística implementadas sobre un servidor Web, que se pueden realizar parcial o totalmente fuera de clase y/o del campus. El proyecto se inició el año 2003 y actualmente ya se ha utilizado en la docencia de las asignaturas de estadística que imparte nuestro departamento en las licenciaturas de Biología, Geología, Ciencias Ambientales y en la Diplomatura de Estadística.

Palabras Clave: Estadística Multimedia, Programación Java.

AMS: 97U50

1. Introducción

La finalidad principal del proyecto ha consistido en la elaboración de un sistema que permite diseñar, publicar y gestionar por internet una colección de prácticas de estadística en formato multimedia, utilizando básicamente las tecnologías HTML y Java.

A partir de la experiencia didáctica y técnica de un proyecto anterior (STATMEDIA I) se planteó el objetivo de conseguir un material guía exclusivamente orientado a prácticas. En este caso, se trata de que el alumno se pueda acercar a situaciones reales y/o simuladas dejando más en segundo plano lo que serían aspectos conceptuales (estos ya se abordan en STATMEDIA I). Como objetivo, un tanto más secundario, se desea mostrar el interés que tiene la estadística en ciencias en donde el análisis de datos resulta vital para obtener conclusiones. En este sentido, cada práctica se ha concebido como un verdadero caso de análisis aplicado de datos.

Podríamos esquematizar lo que aporta STATMEDIA II en los siguientes puntos:

1. Es una herramienta útil para diseñar prácticas de análisis de datos
2. Complementa Statmedia I

3. Facilita la implementación del sistema de créditos ECTS
4. Permite la semi-presencialidad
5. Pueden ser prácticas individualizadas
6. El sistema es fácilmente adaptable a muchas asignaturas.
7. El profesor tiene un control exhaustivo del desarrollo de las prácticas.

2. Aspectos técnicos

La parte técnica se basa en la utilización de la tecnología Servlet para permitir un flujo de información entre el cliente y el servidor. La respuesta típica es la creación dinámica de una página de código HTML. La generación de páginas HTML dinámicas se ve facilitada por la utilización de páginas JSP, en realidad un tipo especial de servlets orientado a generar el texto de la interfaz gráfica.

Todo el material ha sido dotado de la posibilidad de interacción entre las páginas de hipertexto que desarrollan los temas o las prácticas con las herramientas informáticas que permiten su resolución.

3. Bibliografía

- [1] Becker, B.J. (1996) *“A Look at the Literature (and Other Resources) on Teaching Statistics”*. Journal of Educational and Behavioral Statistics, 21, 71-90.
- [2] Garfield, J., Ahlgren, A. (1988) *“Difficulties in learning basic concepts in statistics: Implications for research”*. Journal for Research in Mathematics Education, 19, 1988, 44-63.
- [3] Houldsworth, E. (1994) *Multimedia in management education- theoretical framework and methodological choices*. Henley Management College working paper, HW P9420.
- [4] M. Calvo, A. Arcas, A. Miñarro, A. Villarroya, S. Vives (2003) *Statmedia: Un curso multimedia de Estadística*. Publicaciones de la Universidad de Barcelona.

Tablas de mortalidad dinámicas. Modelos edad-periodo-cohorte.

A. Debón¹, F. Montes², F. Martínez-Ruiz³

¹andeau@eio.upv.es, Departamento de Estadística, Investigación Operativa Aplicadas y Calidad, Universidad Politécnica de Valencia

²montes@uv.es, Departament de Estadística i Investigació Operativa, Universitat de València

³pmartinez@uv.es, Oficina de Estadística, Ayuntamiento de Valencia

1. Resumen

Las tablas dinámicas surgieron de la necesidad de incorporar el efecto del periodo (año de muerte) en la estimación de la medida mortalidad. Han supuesto un avance importante respecto a las tablas estáticas a la hora de capturar tendencias históricas en la mortalidad. Por esta razón, se han convertido en herramienta gran utilidad para la ciencia actuaria. Prueba de ello es el texto de [13], donde se describen modelos desarrollados recientemente y se subraya la necesidad de adoptar por los actuarios modelos que ya han dado buenos resultados en otros campos, como por ejemplo la epidemiología. En [8] se recopilan las más recientes contribuciones con especial énfasis en analizar el riesgo que para el asegurador tiene la longevidad. Algunos de los modelos más destacados, y también más frecuentemente utilizados por los actuarios, son los modelos basados en factores de reducción [2], funciones Gompertz-Makeham dinámicas [12] y, más recientemente, el modelo de Lee-Carter [7], [1], o el modelo basado en p-splines [5].

Los modelos edad-periodo-cohorte constituyen una evolución de los modelos dinámicos incorporando la influencia del año de nacimiento (cohorte). Recientemente [11] estudia la flexibilidad para extender los modelos basados en factores de reducción de la mortalidad [10, 9], a la modelización y proyección de efectos históricos de la edad, el periodo y la cohorte.

La propuesta de nuestro trabajo es utilizar el modelo sugerido por [6] y [3, 4] con muestras de datos pertenecientes a períodos recientes de Países desarrollados (España, Suecia y Checoslovaquia), analizando los datos por edad (A), año de muerte (P) y año de nacimiento (C) con especial énfasis en el efecto cohorte en los diferentes países.

Claves: Edad-Periodo-Cohorte, Mortalidad, Longevidad.

2. Agradecimientos

Este trabajo ha sido parcialmente financiado por la Facultad de Administración y Dirección de Empresas de la Universidad Politécnica de Valencia y mediante el proyecto MTM-2004-06231 del Ministerio de Educación y Ciencia.

3. Bibliografía

- [1] Brouhns, N., Denuit, M., y Vermunt, J. (2002). A poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics & Economics*, 31(3),373-393.
- [2] Bureau, C. M. I. (1999). *Continuous Mortality Investigation Reports CMIR 17*. The Institute of Actuaries and Faculty of Actuaries, UK.
- [3] Clayton, D. y Schifflers, E. (1987). Models for temporal variation in cancer rates. I: Age-Period and Age-Cohort models. *Statistics in medicine*, 6(4),449-467.
- [4] Clayton, D. y Schifflers, E. (1987). Models for temporal variation in cancer rates. II: Age-period-cohort models. *Statistics in medicine*, 6(4),449-467.
- [5] Currie, I., Durban, M., y Eilers, P. (2004). Smoothing and forecasting mortality rates. *Statistical Modelling*, 4(4),279-298.
- [6] Holford, T. (1983). The estimation of age, period and cohort effects for vital rates. *Biometrics*, 39,311-324.
- [7] Lee, R. y Carter, L. (1992). Modelling and forecasting U. S. mortality. *Journal of the American Statistical Association*, 87(419),659-671.
- [8] Pitacco, E. (2004). Survival models in dynamic context: a survey. *Insurance: Mathematics & Economics*, 35(2),279-298.
- [9] Renshaw, A. y Haberman, S. (2003a). Lee-Carter mortality forecasting with age specific enhancement. *Insurance: Mathematics & Economics*, 33(2),255-272.
- [10] Renshaw, A. y Haberman, S. (2003b). On the forecasting of mortality reduction factors. *Insurance: Mathematics & Economics*, 32(3),379-401.
- [11] Renshaw, A. y Haberman, S. (2006). A cohort-based extension to the lee-carter model for mortality reduction factors. *Insurance: Mathematics & Economics*.
- [12] Renshaw, A., Haberman, S., y Hatzopoulos, P. (1996). The modelling of recent mortality in United Kingdom male assured lives. *British Actuarial Journal*, 2(II),449-477.
- [13] Tabeau, E., van den Berg Jeths, A., y (Eds), C. H. (2001). *A Review of Demographic Forecasting Models for Mortality. Forecasting in Developed Countries: From description to explanation*. Kluwer Academic Publishers.

TAID: Algoritmo de segmentación basado en árboles ternarios

P. Galindo¹, C. R. Castro²

¹pgalindo@usal.es, Dpto. de Estadística, Univ. Salamanca. España

²castro@uv.mx, Facultad de Estadística e Informática. Universidad Veracruzana, México

Abstract

Presentamos un algoritmo de segmentación en base a la creación de árboles ternarios con variable respuesta latente. Se aborda la descomposición factorial de la predictividad mediante el Análisis No Simétrico de Correspondencias. Se presenta una aplicación del Algoritmo propuesto a datos de marketing político en las elecciones americanas.

Palabras Clave: Clases Latentes, Índice de predictividad, CHAID

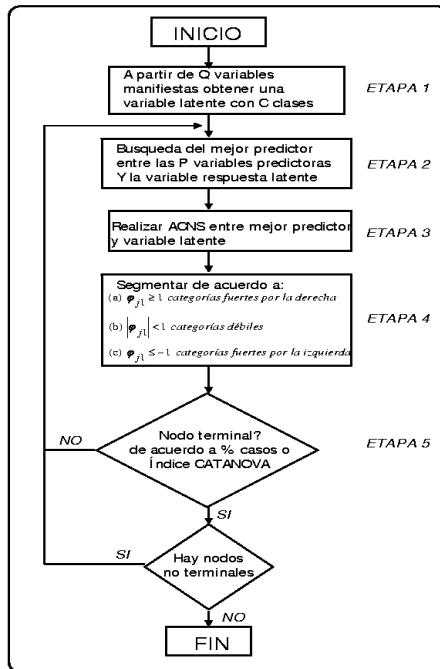
1. Introducción

Métodos como el CHAID (*Chi-square Automatic Interaction Detection*), propuesto por Kass (1980), o el método CART (Clasification And Regression Trees), propuesto por Breiman, y col. (1984), utilizan un conjunto de variables predictoras y una variable respuesta; sin embargo, hay un gran número de situaciones en donde una o más variables respuesta necesitan ser consideradas para definir segmentos homogéneos en relación a esa, o esas, variables respuesta. Repetir el análisis para cada variable respuesta conlleva, cuando menos, un importante incremento en el riesgo Tipo I. El test chi-cuadrado en el que se basan los test, no considera un rol entre variable predictora y variable respuesta y evalúa solamente si las variables consideradas covarian. Otra cuestión clave que puede invalidar los análisis es que en cada paso del algoritmo de segmentación se cruza un par de variables sin tener en cuenta las otras, lo cual solo es posible llevarlo a cabo cuando se haya demostrado que la tabla de datos es *colapsable* (Asmussen y Edwards, 1983) sobre dos cualesquiera de las variables. Avila (1996) y Dordo (1998) demostraron que en el algoritmo CHAID, la condición de colapsabilidad no se verifica en general. Además, los algoritmos utilizados tradicionalmente solo contemplan una variable respuesta.

2. Algoritmo

La metodología aquí propuesta tiene como objetivo segmentar n casos o individuos y formar un árbol ternario de segmentación, tomando como base un criterio de segmentación para formar tres subgrupos disjuntos que se buscan sean internamente homogéneos, tanto como esto sea posible; esto se realiza en base a la predictividad de las categorías de la variable predictora. (*categorías fuertes y débiles*), siguiendo las ideas de Siciliano y Mola (1997). Está basada en la idea de utilizar coeficientes de asociación que tengan en cuenta el claro carácter asimétrico de la información contenida en los datos. En aquellos casos

en los que existan varias variables respuesta, el primer paso será definir una variable latente que recoja el carácter multivariante de la respuesta.



3. Bibliografía

- [1] Asmussen, S y Edwards, D (1983): Collapsibility and Response Variables in Contingency tables. *Biometrika*. 70, p: 567-578.
- [2] Avila, C.A. (1996). Una Alternativa al Análisis de Segmentación Basada en el Análisis de Hipótesis de Independencia Condicionada. *Tesis Doctoral, Universidad de Salamanca*.
- [3] Breiman, L., Friedman, J., Olsen, R. y Stone, C. (1984): *Classification and Regression Trees*. Chapman and Hall.
- [4] Dorado, A. (1998): Métodos de Búsqueda de Variables Relevantes en Análisis de Segmentación: Aportaciones desde una Perspectiva Multivariante. *Tesis Doctoral, Universidad de Salamanca*.
- [5] Kass, G. (1980): An Exploratory Technique for Investigating Large Quantities of Categorical Data. *Applied Statistics*, 29, 2, 199-127.
- [6] Siciliano, R. y Mola, F. (1997): Ternary Classification Trees: A Factorial Approach. En *Visualization of Categorical Data*. Cap 22, 311-323, Academic Press.

Técnicas paramétricas de clasificación automática para datos composicionales: resultados preliminares

J.A. Martín Fernández¹, J. Daunis i Estadella², G. Mateu Figueras³

¹josepantoni.martin@udg.es, ²josep.daunis@udg.es, gloria.mateu@udg.es,
Departament d'Informàtica i Matemàtica Aplicada, Universitat de Girona

Abstract

Los últimos avances en el campo de los datos composicionales nos aseguran que el simplex tiene estructura de espacio Euclíadiano y que la transformación log-cociente isométrica nos permite trabajar en coordenadas ortonormales. En este trabajo mostramos los resultados preliminares cuando se aplican estas nuevas propiedades en una clasificación paramétrica de datos composicionales.

Palabras Clave: log-ratio, datos de proporciones, cluster.

AMS: 62H30, 62P99

1. Introducción

Un dato composicional expresa las proporciones de las partes respecto a un todo. El espacio muestral de los datos composicionales ([1]) es el simplex S^D definido por $S^D = \{x = [x_1, \dots, x_D] : x_i > 0, x_1 + \dots + x_D = c\}$, donde c puede valer 1, 100, 10^6 o cualquier otro valor constante que refleja las unidades de medida. Por lo tanto, los datos composicionales se pueden representar por vectores cuyas componentes son positivas y tienen suma constante. Como se constató en el último CoDaWork [2] este tipo de datos está presente en multitud de problemas de diferentes áreas como son, entre otras: arqueometría, geología, economía, biomedicina, y ciencias del espacio. En lo que se refiere a nuestro trabajo, la cuestión *clave* es qué distribuciones de probabilidad pueden ser apropiadas para este tipo de datos. Desde [1] la comunidad científica acepta mayoritariamente que los datos composicionales reflejan magnitudes relativas. En consecuencia, las distribuciones de probabilidad definidas sobre el simplex deben ser compatibles con la estructura de S^D [4] y deben modelar estos datos teniendo en cuenta que se está interesado en los cambios relativos y no en los cambios absolutos, por lo tanto deberán construirse en función de los logaritmos de los cocientes de las componentes ([1], [5]).

En [4] se realiza una presentación de la metodología para las técnicas paramétricas de clasificación de conjuntos de datos composicionales. La metodología presentada en ese trabajo, si bien de carácter general, se basó en las técnicas de clasificación que combinan el algoritmo EM y la mezcla de distribuciones. Las distribuciones de probabilidad definidas en el simplex que han proporcionado los resultados más razonables se definen a través de

las transformaciones alr, clr e ilr, todas ellas biyectivas pero con diferencias importantes. En [1] se introduce las transformaciones *additive log-ratio* (alr) y *centred log-ratio* (clr)

$$\text{alr}(x) = \left[\log \frac{x_1}{x_D}, \dots, \log \frac{x_{D-1}}{x_D} \right] , \quad \text{clr}(x) = \left[\log \frac{x_1}{g(x)}, \dots, \log \frac{x_D}{g(x)} \right] ,$$

donde $g(x)$ es la media geométrica de la composición x . En [3] se introduce la transformación *isometric log-ratio* (ilr), $\text{ilr}(x)=y \in \mathbb{R}^{D-1}$, donde

$$y_i = \frac{1}{\sqrt{i(i+1)}} \log \left(\frac{\prod_{j=1}^i x_j}{(x_{i+1})^i} \right) , \quad i = 1, \dots, D-1.$$

En este trabajo mostramos los resultados preliminares que se obtienen cuando se aplican las técnicas paramétricas de clasificación usuales en combinación con las transformaciones alr, clr e ilr. Ponemos especial énfasis en describir el comportamiento de las tres transformaciones en relación a las diferentes parametrizaciones posibles de la matriz de covarianzas en el método EM. En particular, analizamos la invarianza por permutaciones de las componentes para el caso de la transformación alr y la singularidad de la matriz de covarianzas para el caso de la transformación clr.

2. Agradecimientos

Este trabajo ha recibido financiación mediante el proyecto BFM2003-05640/MATE (Dir. General de Investigación; Minis. de Ciencia y Tecnología).

3. Bibliografía

- [1] Aitchison J. (1986) *The Statistical Analysis of Compositional Data*. London: Chapman and Hall. Reprinted in 2003 by Blackburn Press.
- [2] CODAWORK'05 (2005) Second Compositional Data Analysis Workshop, Mateu G. and Barceló C. (Eds.), October 19-21, University of Girona (Spain), CD-ROM, ISBN: 84-8458-222-1 [<http://ima.udg.es/Activitats/CoDaWork05/>].
- [3] Egozcue, J. J., Pawlowsky-Glahn V., Mateu-Figueras G., Barceló-Vidal C. (2003) Isometric logratio transformations for compositional data analysis. *Math. Geol.* 35 (3): 279-300.
- [4] Martín-Fernández, J. A., Daunis-i-Estadella, J., and Mateu-Figueras, G., 2004, Clasificación paramétrica de datos composicionales: aproximación metodológica. In: Dept. EIO, UCA (Eds.), XXVIII Congreso Nacional de Estad. e Invest. Oper., Cádiz (E), CD-ROM, ISBN: 84-689-0438-4, 18 p.
- [5] Mateu-Figueras G. (2003) Distribution Models on the Simplex; ISBN: 84-688-6734-9, *Ph.D. thesis*, Universitat Politècnica Catalunya [<http://www.tdcat.cesca.es/TDX-0427104-170301/index.html>], 202p.

Test de bondad de ajuste multivariante basado en profundidad

Ángel López¹, Juan Romo²

¹anlopez@est-econ.uc3m.es, Departamento de Estadística, Universidad Carlos III de Madrid

²romo@est-econ.uc3m.es, Departamento de Estadística, Universidad Carlos III de Madrid

Abstract

Se introduce un nuevo test de bondad de ajuste multivariante para contrastar si una muestra, de dimensión uno o superior, procede de una distribución completamente especificada o una familia de distribuciones. El estadístico del contraste se construirá a partir de ordenaciones hechas por medio de funciones de profundidad estadística.

Claves: tests de bondad de ajuste; multivariante; profundidad.

1. Introducción

En este trabajo de investigación se presenta un contraste de bondad de ajuste distribucional para su aplicación sobre datos con dimensión mayor o igual que uno. Este caso no cuenta en la actualidad con un gran número de contrastes aplicables a cualquier tipo de distribución. Así pues, aunque pueden encontrarse numerosos trabajos orientados a la distribución Normal Multivariante, lejos de ésta continúa la búsqueda de contrastes orientados a la resolución de este problema.

El estadístico de contraste está construido a partir de la idea de profundidad estadística de datos. Ésta consiste en medir la distancia de un punto cualquiera con el centro de la distribución pero intentando respetar la forma de la distribución o de la nube de puntos si es una muestra. La cuantificación de este concepto se realiza a través de las funciones de profundidad. De las funciones de profundidad propuestas en la literatura, véanse [2] y [3], salvo algunas excepciones se respeta de forma razonable la forma de los datos cuando ésta no es elíptica o simétrica.

Asignar un valor a cada observación dentro de un conjunto de datos se traduce en la posibilidad de ordenarlos parcialmente desde el centro hacia fuera. Desde puntos muy profundos hasta puntos menos profundos (de centrales a externos).

2. Estadístico del contraste

La profundidad de un punto x con respecto a una función de distribución teórica o empírica, F y F_n respectivamente, se denota como $D(x; F)$ y $D(x; F_n)$. La

idea que se propone es trabajar con estos valores como variables aleatorias con el fin de estudiar la concordancia entre la muestra y la hipótesis nula. Suponemos que queremos contrastar $H_0 : X \sim F_0$ frente a $H_1 : X \not\sim F_0$, y que nuestro conjunto de observaciones, y_1, y_2, \dots, y_n , se distribuye según G .

Para medir la concordancia entre muestra e hipótesis nula se usarán las siguientes posibilidades de cálculo de profundidad: $D(X; F_0)$, $D(X; G_n)$, $D(y_i; F_0)$ y $D(y_i; G_n)$.

Se define el estadístico $CP(p)$ con $p \in [0, 1]$, o concordancia en profundidad, para la proporción p como el mínimo de dos probabilidades que miden concordancia de hipótesis nula con muestra (P_1) y viceversa (P_2):

$$CP(p) = \min(P_1(p), P_2(p)),$$

donde

$$\begin{aligned} P_1(p) &= P_{F_0}(x : D(x; G_n) \geq D_p(y_i; G_n) \text{ y } D(x; F_0) \geq D_p(x; F_0)) \\ P_2(p) &= P_{G_n}(y_i : D(y_i; F_0) \geq D_p(x; F_0) \text{ y } D(y_i; G_n) \geq D_p(y_i; G_n)), \end{aligned}$$

y $D_p(X; F_0)$ y $D_p(y_i; G_n)$ son los percentiles p -ésimo de $D(X; F_0)$ y $D(y_i; G_n)$.

El estadístico de discrepancia se define como el área entre $CP(p)$ y su esperanza.

3. Simulaciones de potencia

Se presentarán resultados de simulación sobre distintas funciones de distribución, tanto en el caso unidimensional como el multidimensional, comparando los resultados con otros contrastes como, por ejemplo, Kolmogorov-Smirnov en una y varias variables, véanse respectivamente [4] y [1].

4. Bibliografía

- [1] Justel A., Peña D. y Zamar R. (1997). A Multivariate Kolmogorov-Smirnov test of goodness of fit. *Statistics & Probability Letters* 35, 251-259.
- [2] Liu R.Y., Parelius J.M. y Singh K. (1999). Multivariate Analysis by Data Depth: Descriptive Statistics, Graphics and Inference. *The Annals of Statistics* 27, 783-858.
- [3] López-Pintado S. y Romo J. (2005). Depth-based classification for functional data. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, American Mathematical Society.
- [4] Massey F.J. (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association* 46, 60-78.

Time-varying joint distribution through copulas

Maria Concepción Ausín¹, Hedibert Freitas Lopes²

¹mausin@udc.es, Departamento de Matemáticas, Universidade da Coruña

²hlopes@gsb.uchicago.edu, Graduate School of Business, University of Chicago

Abstract

This paper deals with the analysis of temporal dependence in multivariate high-frequency time series data. The dependence structure between the marginal series is modelled through the use of copulas, which are assumed to be time dependent. We develop full Bayesian inference where the whole set of model parameters is estimated simultaneously. This represents an essential difference with previous approaches in the literature where the marginal and the copula parameters are estimated separately in two consecutive steps. The proposed approach is illustrated with simulated data and a real bivariate time series on exchange rates.

Claves: Bayesian inference, dynamic copulas, GARCH models.

1. Introducción

It is well known that financial volatilities move frequently together through time and then, it is important to find appropriate models to understand and predict the temporal dependence in multivariate high-frequency data. Although widely used, the correlation matrix is not a well measure of dependence when normality assumptions do not hold and it can be misleading for financial investors.

Copula is one of the most promising tools to describe the dependence between random variables, see e.g. Nelsen (1999) for an detailed introduction to copulas. The main idea behind copula is to separate the marginal behaviour and dependence structure from joint distribution. This is a great advantage in order to define time-varying variances models because one can specify separately, on the one hand, the individual conditional variances, and on the other hand, the conditional correlation matrix or another measure of dependence between the individual series.

The use of copulas in financial economics is quite new and it is a fast growing field of research. Although copulas were introduced in the fifties, they are not applied to finance until Embrechts et al. (1999). Furthermore, the application of copulas in finance have been mainly devoted to multivariate static problems and few articles have considered the time varying dependence structure in multivariate time series, see e.g. Hu (2004), Patton (2005) and Jondeau and Rockinger (2005). Other areas of copula application in finance include credit risk modelling, multivariate option pricing, portfolio and market Value-at-Risk

calculations, default and credit risk dependence, contagion, tail dependence, etc.

Previous works using copulas in time-varying variance models does not account for the parameter uncertainty simultaneously. Proposed methods are generally based on a two-stage approach where, in the first step, the marginal series are estimated assuming independence and then, in a second step, these estimations are plugged in the copula function so that the copula parameters can be estimated. Two-steps approaches can be easily implemented because the joint density is decomposed as the product of the copula function and the marginal densities. Different two-step procedures are considered in the references cited above using maximum likelihood and non-parametric estimations. It can be shown that two-stage ML approaches lead to consistent but not efficient estimators, see Patton (2005).

In this work, we propose a full Bayesian approach where all the model parameters are estimated simultaneously using the whole likelihood function. To better illustrate the methodology, we consider a bivariate copula-GARCH model with univariate t-GARCH(1,1) marginals and a t-copula function, but the same approach can be extended to other dynamic models. Bayesian inference and prediction is implemented via the Markov Chain Monte Carlo (MCMC) methodology. Furthermore, the copula dependence parameter is assumed to vary through time according to certain equations which are also estimated. We illustrate our approach with some simulated data and a real high-frequency bivariate time series consisting in daily Deutsche Mark - U.S. Dollar and Japanese Yen - U.S. Dollar exchange rates.

2. References

- Embrechts, P., McNeil, A., and Straumann, D. (1999). "Correlation: pitfalls and alternatives". *Risk* 5, 69-71.
- Hu, L., (2004). "Dependence Patterns across Financial Markets: A Mixed Copula Approach," *Applied Financial Economics*, forthcoming.
- Jondeau, E., and Rockinger, M., (2005). "The Copula-GARCH model of conditional dependencies: An international stock-market application". *Journal of International Money and Finance*, forthcoming.
- Nelsen, R. B., (1999). *An Introduction to Copulas*. Springer-Verlag, New York.
- Patton, A. (2005). "Estimation of Multivariate Models for Time Series of Possibly Different Lengths". *Journal of Applied Econometrics*, forthcoming.

TIRPE, un tutor informatizado para la resolución de problemas de Bioestadística

Pedro Femia Marzo, Juan de Dios Luna del Castillo

pfemia@ugr.es, jdluna@ugr.es

Departamento de Estadística e I.O.

Facultad de Medicina. Universidad de Granada

Abstract

TIRPE es un Tutor Informatizado para la Resolución de Problemas de (Bio)Estadística. Se trata de un sistema informático que permite la resolución interactiva de problemas de Estadística aplicada.

1. Introducción

TIRPE es un Tutor Informatizado para la Resolución de Problemas de (Bio)Estadística. Se trata de un sistema informático que permite la resolución interactiva de problemas de Estadística aplicada. Originalmente está concebido para estudiantes de titulaciones vinculadas a las Ciencias de la Salud que tienen la Bioestadística como asignatura troncal.

Desde el punto de vista didáctico, TIRPE induce al usuario a resolver los problemas reproduciendo el que debe ser el esquema metodológico que conduce al encuadre del problema, la identificación de la información suministrada, el reconocimiento de los principios a considerar, la ejecución de los cálculos necesarios y la elaboración de una la conclusión final.

Desde el punto de vista informático, TIRPE es una aplicación cliente-servidor. Del lado del cliente consiste en un programa ejecutable sobre plataforma Win-32 con componentes de navegación web. En el lado del servidor se trata de una base de datos MySQL. El usuario puede descargar el programa cliente y seleccionar un problema, bien correspondiente a un capítulo concreto o bien de forma aleatoria. Esto genera una consulta a la base de datos del servidor de forma que se transfiere la información relativa al problema seleccionado al cliente. La resolución del problema se va haciendo a través de un proceso de preguntas con respuesta de elección múltiple. Los errores de contestación dan lugar a la aparición de ventanas emergentes que asesoran al usuario. La selección de las sucesivas respuestas correctas supone la progresiva resolución del problema hasta obtener un documento final consistente en el enunciado, cada una de las etapas implicadas en su análisis y la conclusión final. El formato del documento final es HTML, por tanto el usuario lo puede archivar o incluso editar mediante aplicaciones ofimáticas de uso frecuente.

Un enfoque bootstrap para la estimación no paramétrica de una función de regresión a partir de tiempos de recurrencia

José A. Cristóbal¹, Pilar Olave², Tomás Alcalá³

¹cristo@unizar.es, Dpto. de Métodos Estadísticos, Fac. Ciencias
(Matemáticas) Universidad de Zaragoza

²polave@unizar.es, Dpto. de Métodos Estadísticos, Fac. Ciencias Económicas
y Empresariales. Universidad de Zaragoza

³jtalcala@unizar.es, Dpto. de Métodos Estadísticos, Fac. Ciencias
(Matemáticas) Universidad de Zaragoza

Abstract

Se construye un estimador no paramétrico de una función de regresión cuando la variable respuesta, no observable, es el tiempo de espera entre dos eventos consecutivos de un proceso de renovación. Se aplica un procedimiento bootstrap para obtener valores artificiales de la variable original, a partir de una muestra que está sesgada por longitud y censurada multiplicativamente.

Palabras Clave: Bootstrap. Sesgo por longitud. Tiempo de recurrencia.

AMS: 62G08, 62G09

1. Metodología

En este trabajo se analiza el problema de la estimación no paramétrica de una función de regresión cuando la variable respuesta X no es directamente observable, con respecto a un cierto vector de covariables Z . Los datos provienen de un muestreo sesgado por longitud sobre la variable X , obteniéndose la variable sesgada X^w . Los valores de esta variable sesgada X^w tampoco se pueden medir, y en lugar de ello sólo es posible observar una variable Y , la cual es una censura multiplicativa de X^w , es decir $Y = U.X^w$, donde U es una variable uniforme en $[0,1]$, independiente de X^w .

Dado que la función de densidad de Y debe ser no creciente, podemos utilizar el estimador no paramétrico de máxima verosimilitud (NPMLE) de esta función, que se deduce del menor mayorante cóncavo de la función de distribución empírica de la muestra $y_{i1,n}$, y que, como probó Grenander, es consistente para todo $y > 0$. A partir de dicha densidad estimada, se propone un estimador para la función de distribución condicional de $X^w|Y$, y en consecuencia, otro estimador para la distribución condicional de $X^w|Z$.

De este modo, podemos obtener muestras bootstrap de la variable condicional $X^w|Z$. Para cada una de estas muestras X^w* , es posible utilizar el método

de Cristóbal y Alcalá (2000) para construir un estimador consistente de la media condicional $E(X/Z)$ a partir de la muestra $(X_i^{(w)}, Z_i)_{i=1,n}$. Este método está basado en el cálculo del estimador lineal local, con pesos $(X_i^{(w)})^{-1}$, de la función de regresión $m(z)=E(X/z)$. El estimador final se obtiene calculando la media de todas las funciones anteriores cuando se replica B veces el procedimiento bootstrap.

La justificación teórica se basa en el análisis del comportamiento asintótico, obteniéndose la razón de convergencia del estimador propuesto. Por otra parte, el comportamiento para muestras de tamaño finito se analiza a través de un proceso de simulación, bajo diferentes distribuciones y formas de la función de regresión.

2. Agradecimientos

Este trabajo ha sido subvencionado parcialmente por el Ministerio de Educación y Ciencia (proyecto MTM2005-01464), así como por el Gobierno de Aragón (Grupo Consolidado 2005 - S11).

3. Bibliografía

- [1] Asgharian M, Wolfson DB (2005) Asymptotic behavior of the unconditional NPMLE of the length-biased survivor function from right censored prevalent cohort data. *Ann. Statist.* 33: 2109-2131
- [2] Cristóbal JA, Alcalá JT (2000) Nonparametric regression estimators for length-biased data. *J. Statist. Plann. Inference* 89: 145-168
- [3] Groeneboom P (1985) Estimating a monotone density. In: LeCam L, Olshen R (eds) *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer* 2: 539-555
- [4] Van Es B, Klaassen CAJ, Oudshoorn K (2000) Survival analysis under cross-sectional sampling: length bias and multiplicative censoring. *J. Statist. Plann. Inference* 91: 295-312
- [5] Woodroffe M, Sun J (1993) A penalized maximum likelihood estimate of $f(0+)$ when f is non-increasing. *Stat. Sinica* 3: 501-515

Un Estudio Comparativo de Técnicas de Clasificación para la Diagnosis del Alzheimer Mediante SPECT y CAMCOG

J.M. Fernández-Ponce, M.T. Gómez-Gómez, J.L. Pino-Mejías,
A. Ruiz-Gándara

Departamento de Estadística e I.O.

Universidad de Sevilla

ferpon@us.es, terego@us.es, jlpin@us.es, africaruiz@us.es

Abstract

En los últimos diez años el desarrollo del Alzheimer junto con otras enfermedades, han visto en expansión un nuevo campo de investigación neurológica: crear un mapa funcional del cerebro humano. Han surgido muchas preguntas dirigidas a este área de investigación, algunas de ellas podrían ser contestadas usando técnicas de medicina nuclear como el *Single Photon Emission Computed Tomography* (SPECT). Relativamente pocos estudios de activación SPECT han sido realizados (ver entre otros a [1,2,3,4]). También se estudió el conocimiento de las reacciones que llega a sufrir un cerebro humano ante un determinado estímulo y distinguir en esas ocasiones un cerebro normal de uno demente (ver [6]). En este trabajo, se pretende comparar métodos de clasificación para imágenes cerebrales SPECT que sean capaz de distinguir entre pacientes con demencia y pacientes no dementes (para más detalles sobre técnicas estadísticas usadas ver [5]). En nuestro estudio tenemos un conjunto de 57 pacientes de los que se obtiene el estudio basal y de activación, tomando como referencia en cada estudio los 11 cortes transaxiales realizados mediante imágenes SPECT, a estos cortes (tras realizar una cuadriculación de los mismos), se les asigna un número correspondiente al grado de color de cada cuadrícula del corte, una vez realizado esto se obtienen los datos, para cada paciente, resultantes de realizar la sustracción (Estudio de activación - Estudio basal). Aplicando a estos el análisis adecuado seleccionaremos una serie de cuadrículas y cortes que son los que nos darán la información necesaria para poder encuadrar a cada paciente en su correspondiente grupo (Alzheimer, demencia mixta y sin deterioro).

Claves: Análisis Discriminante; Árboles de decisión; CAMCOG; Redes Neuronales; Regresión Logística; SPECT.

AMS: 62P10, 62H30

Dirección:

Departamento de Estadística e Investigación Operativa
 Facultad de Matemáticas - Av. Reina Mercedes s/n
 41012 - Sevilla
 Spain

Teléfono: (+34) 954557927 **Fax:** (+34) 954622800

web <http://destio.us.es/>

1. Bibliografía

- [1] Adriaan A. Lammertsma. (2001). PET/SPECT: functional imaging beyond flow. *Vision Research*. 41, 1277-1281.
- [2] Borroni, B., Perani, D., Broli, M., Colciaghi,F., Garibotto,V., Paghera, B., Agosti,C., Giubbini, R., Di Luca, M. and Padovani, A. (2005). Pre-clinical diagnosis of Alzheimer disease combining platelet amyloid precursor protein ratio and rCBF spect analysis. *Journal of Neurology* ISSN 1432-1459.
- [3] Cardebat,D., Demonet,J.F., Celsis, P. and Puel, M. (1996). Living/non-living dissociation in a case of semantic dementia: a SPECT activation study. *Neuropsychologia* 34, 1175-1179.
- [4] Cardebat, D., Dèmonet, J.F., Puel, M., Agniel, A., Viallard, G. and Celsis, P. (1998). Lagrangean Decomposition. Brain correlates of memory processes in patients with dementia of Alzheimer's type: a SPECT activation study. *Journal of Cerebral Blood Flow and Metabolism* 18, 457-462.
- [5] Higdon, R., Foster,N.L., Koepp, R.A., DeCarli, C.S., Jagust, W.J., Clark, C.M., Barbas, N.R., Arnold, S.E., Turner, R.S., Heidebrink, J.L. and Minoshima, S. (2004). A Comparison of classification methods for differentiating fronto-temporal dementia from Alzheimer's disease using FDG-PET imaging. *Statistics in Medicine*. 23, 315-326.
- [6] Shapiro, P.A., Sloan, R.P., Bagiella, E., Kuhl,J.P., Anjilvel, S. and Mann J. (2000). Cerebral activation, hostility, and cardiovascular control during mental stress *Journal of Psychosomatic Research*. 48, 485-491.

Un experimento computacional para la simulación estocástica de la dinámica de burbujas

Javier Jiménez-Fernández¹, Julio Rodríguez², José Mira³, Camino González⁴

¹jajimenez@etsii.upm.es, Departamento de Ingeniería Energética y Fluido mecánica

²puerta@etsii.upm.es, ³jmira@etsii.upm.es, ⁴camino@etsii.upm.es

Laboratorio de Estadística

Escuela Técnica Superior de Ingenieros Industriales, Universidad Politécnica de Madrid

Abstract

Este trabajo es un experimento computacional para la simulación del comportamiento dinámico de burbujas en fluidos viscoelásticos; se varían parámetros del modelo determinista y se añade una componente estocástica igual a un proceso AR(1) continuo para estudiar la sensibilidad de la respuesta.

Palabras Clave: Rayleigh-Plesset, AR(1), Experimento computacional

1. Introducción

La dinámica de burbujas gaseosas en medio viscoelástico es un campo de interés creciente con numerosas aplicaciones. Las burbujas constituyen, por ejemplo, una herramienta de gran valor como agentes de contraste en diagnóstico médico con ultrasonidos.

Es bien conocido que, bajo la acción de un campo acústico, las burbujas de radio pequeño experimentan una expansión hasta tamaños iguales varias veces su valor inicial, seguida de un rápido colapso. Este proceso, llamado cavitación inercial, puede ocurrir durante el proceso de diagnóstico médico, y dañar los tejidos del paciente debido a las altas presiones y temperaturas que se alcanzan en la burbuja al final del colapso. Para bajas frecuencias, la cavitación inercial aparece a partir de un valor umbral de la amplitud de la onda acústica de presión.

Se trata de estudiar la probabilidad de que se alcance dicho umbral cuando la excitación (onda de presión acústica) tiene una componente estocástica y se varían parámetros del proceso.

2. Metodología

El modelado determinista se hace a través de la ecuación de Rayleigh-Plesset que cuantifica la evolución temporal del radio de la burbuja, a partir del radio

se determina la presión.. En este trabajo se añade a la excitación senoidal una componente estocástica consistente en un proceso AR(1) continuo. Además se varían una serie de parámetros del modelo determinista para estudiar la sensibilidad de la respuesta.

El modelado estadístico de los experimentos computacionales surge a finales de los 80 con los trabajos pioneros de Sacks et al (1989) y Currin et al (1991). Los diseños experimentales para este tipo de modelos que se aplican en este trabajo están descritos en Koehler y Owen (1996).

3. Agradecimientos

Javier Jiménez-Fernández agradece la financiación del proyecto DP12003-5643 del Ministerio de Educación y Ciencia. Los demás autores agradecen la financiación del proyecto MTM2005-08897 del Ministerio de Educación.

4. Bibliografía

- [1] Jimenez-Fernández, J. y Crespo, A. (2005), "*Bubble oscilation and inertial cavitation in viscoleastic fluids*". Ultrasonics, 43, 643-651.
- [2] Currin, C., Mitchell, T., Morris, M., y Ylvisaker, D., (1991), "*Bayesian Prediction of Deterministic Functions with Applications to the Design and Analysis of Deterministic Functions*". Journal of the American Statistical Association, vol 86, 416.
- [3] Sacks, J., Welch, W., Mitchell, T. y Wynn., H., (1989), "*Design and Anañysis of Computer Experiments*", Statistical Science, vol. 4, num 4, 409-435.
- [4] Koehler, J.R., Owen, A.B., (1996) "*Computer Experiments*", Handbook of Statistics, vol 13, 261-308.

Un modelo híbrido para la detección precoz del abandono de clientes

*Susana San Matías¹, Ángeles Caldúch², Vicent Giner³,
Miguel Ángel Edo⁴*

¹ssanmat@eio.upv.es, ²mcalduch@eio.upv.es, ³vigibos@eio.upv.es,
⁴miedap@eio.upv.es, Departamento de Estadística e Investigación Operativa
Aplicadas y Calidad, Universidad Politécnica de Valencia

Abstract

La identificación de clientes en riesgo de abandono es un objetivo fundamental dentro de cualquier estrategia de marketing. En este trabajo abordamos dicho problema desde la minería de datos, proponiendo una metodología basada en la aplicación combinada de dos modelos que permite detectar la desvinculación de un cliente antes de que se produzca.

Claves: Minería de datos, series temporales, fidelización de clientes.

AMS: 62H12, 62M10, 62P25

1. Introducción

Un aspecto muy importante en la estrategia de marketing de cualquier empresa es la fidelización de los clientes, sobre todo en sectores con un alto grado de competitividad, como pueden ser finanzas o telecomunicaciones, y en los que no se sabe realmente cuándo un cliente se ha desvinculado hasta que es demasiado tarde.

Hay que tener en cuenta que cuesta mucho esfuerzo y dinero tanto hacer nuevos clientes como recuperar los que han dejado de serlo. Por ello, el adelantarse al momento del abandono del cliente es esencial.

Por otra parte, el porcentaje de abandonos en este tipo de negocios suele ser pequeño en comparación con los clientes que permanecen (entre un 1% y un 5% por mes o trimestre), lo cual dificulta la identificación de un perfil o patrón de dichos abandonos.

2. Nuestra propuesta

En este trabajo se propone una metodología que puede aplicarse a empresas que dispongan de datos individuales de sus clientes, en lo que respecta a sus características y su comportamiento o hábitos, como pueden ser entidades bancarias, empresas de telefonía, supermercados, grandes almacenes...

El objetivo general que abordamos en este trabajo es utilizar la minería de datos para establecer una metodología que permita determinar qué clientes son susceptibles de desligarse de la empresa y en qué medida.

Así, proponemos una estrategia para la detección precoz del abandono de clientes. Los resultados de esta detección serían utilizados por los departamentos comerciales o de marketing, que deberían diseñar acciones de retención para los clientes en los que se ha detectado riesgo de transferencia.

La metodología que proponemos para cuantificar el riesgo de abandono del cliente consiste en un doble enfoque combinado, ya que empleamos tanto un modelo de predicción del abandono como un análisis de evolución personal, para obtener finalmente el riesgo de abandono para cada cliente.

En el modelo de predicción del abandono, se analizan las características comunes de los clientes que abandonan, se compara el perfil de cada cliente con el grupo de abandono, y se asigna una probabilidad o propensión al abandono para cada cliente.

En el análisis de evolución personal, se selecciona una característica vinculada al abandono, se estudia la progresión de dicha característica para cada cliente individualmente, y se realiza una predicción de dicha característica en cada cliente, utilizando series temporales.

En resumen, nuestra propuesta consiste en:

1. Evaluar el riesgo de abandono de un cliente a partir de su situación actual en relación al grupo total de clientes
2. Evaluar el riesgo de abandono del cliente a partir de su evolución en el tiempo de forma individual
3. Combinar las predicciones obtenidas por ambos sistemas para clasificar al cliente en una única categoría de riesgo

3. Agradecimientos

Este trabajo ha sido parcialmente subvencionado por un proyecto de investigación de la Universidad Politécnica de Valencia (PPI-06-05)

Un nuevo estadístico para detectar el desequilibrio de ligamiento global entre loci multialélicos

A. Cancela¹, C. Carollo², C. Zapata³

¹infacc00@ucv.udc.es, Departamento de tecnologías de la Información y las Comunicaciones, Universidad de A Coruña

²mcarollo@usc., Departamento de Estadística e I.O. , Universidad de Santiago de Compostela

³ bfcazaba@usc.es, Departamento de Genética, Universidad de Santiago de Compostela

Abstract

La existencia del desequilibrio de ligamiento global entre pares de loci con múltiples alelos suele detectarse mediante el test χ^2 de Pearson. En el presente estudio, se desarrolla un test alternativo basado en el coeficiente D que se define como el sumatorio de las covarianzas interalélicas en valor absoluto. La potencia de este nuevo test es muy superior a la del test de Pearson bajo condiciones experimentales.

Palabras Clave: desequilibrio de ligamiento global, loci multialélicos, tests de hipótesis

El análisis de la asociación no aleatoria entre alelos de loci distintos (desequilibrio de ligamiento, DL) es muy útil para conocer la estructura genética y la historia evolutiva de las poblaciones. Asimismo, el análisis del DL es una herramienta eficaz para la localización cromosómica aproximada de genes implicados en enfermedades (clonación posicional) [3]. La teoría de estimación del DL entre pares de loci multialélicos (con m y n alelos, respectivamente) está mucho menos desarrollada que la análoga para el sistema genético, relativamente más simple, de dos loci bialélicos [4]. La significación estadística del desequilibrio global entre todos los alelos de dos loci multialélicos se establece generalmente en base al estadístico Q de razón de verosimilitudes. El estadístico Q sigue, asintóticamente, una distribución χ^2 con $(m - 1)(n - 1)$ grados de libertad, bajo la hipótesis nula de independencia. Este test de hipótesis es básicamente equivalente al chi-cuadrado de Pearson para tablas de contingencia de dimensión $m \times n$.

En el presente trabajo hemos desarrollado una aproximación alternativa (*D_{estadístico}*) para contrastar la hipótesis de independencia entre loci multialélicos basada en un estadístico de asociación (D) definido como el sumatorio de las covarianzas interalélicas en valor absoluto. La varianza asintótica de muestreo de D se obtuvo mediante el método delta. Mediante simulaciones Monte Carlo se verificó que su distribución de muestreo se ajusta, generalmente, a la

distribución normal para tamaños de muestra y frecuencias alélicas experimentales. Por lo tanto, $D_{estadistico} = (D - D_0)^2 / \sigma_{\infty 0}^2$ sigue, asintóticamente, una distribución χ^2 con 1 grado de libertad, siendo D_0 y $\sigma_{\infty 0}$, respectivamente, la media y varianza de D bajo la hipótesis de independencia. Se obtuvieron los errores de Tipo I para el test propuesto en una gran variedad de combinaciones de tamaños muestrales, número de alelos en cada locus y frecuencias alélicas, al nivel de significación $\alpha = 0,05$. Se observó que los errores de Tipo I alcanzan valores generalmente inferiores o muy próximos al error nominal de significación del 5 %. Por último, se realizó un análisis comparativo de la potencia estadística de $D_{estadistico}$ con la potencia del test χ^2 de Pearson en un gran número de casos con diferentes intensidades de DL global, diferente número y frecuencias de los alelos y distintos tamaños de muestra ($\alpha = 0,05$). Los resultados muestran que la potencia de $D_{estadistico}$ es generalmente muy superior a la del test de Pearson, especialmente cuando la intensidad del DL es débil. El incremento promedio en la potencia es de $28.22\% \pm 41.36\%$.

Las investigaciones realizadas sugieren que el test χ^2 de Pearson, utilizado frecuentemente para detectar el DL global entre loci multialélicos, está probablemente infraestimando su frecuencia en los genomas y que el nuevo test, $D_{estadistico}$, manifiesta una mayor capacidad para detectar las asociaciones no aleatorias entre genes.

1. Agradecimientos

Este trabajo ha sido financiado parcialmente a través de un proyecto de investigación (BMC2001-3444) concedido por el Ministerio de Educación y Ciencia.

2. Bibliografía

- [1] Hinds, D. A., Stuve, L. L., Nilsen, G. B., Halperin, E., Eskin, E. *et al.* (2005). Whole-genome patterns of common DNA variation in three human populations. *Science* **307**: 1072- 1079.
- [2] Zapata, C., Carollo, C. & Rodríguez, S. (2001). Sampling variance and distribution of the D' measure of overall gametic disequilibrium between multiallelic loci. *Ann. Hum. Genet.* **65**: 395-406.
- [3] Hill, W. G. (1975). Tests for association of gene frequencies at several loci in random mating diploid populations. *Biometrics* **31**: 881-888.
- [4] Hedrick, P. W. (2005). *Genetics of Populations*. Jones and Bartlett Publishers, Massachusetts.

Un nuevo procedimiento de tratamiento de la autocorrelación en control estadístico de la calidad

José Luis Alfaro Navarro¹, José Mondéjar Jiménez, Manuel Vargas Vargas²

¹*JoseLuis.Alfaro@uclm.es, Facultad de Ciencias Económicas y Empresariales de Albacete, Universidad de Castilla-La Mancha.*

²*Jose.Mondejar@uclm.es, ManuelVargas@uclm.es, Facultad de Ciencias Sociales de Cuenca, Universidad de Castilla-La Mancha.*

Abstract

La literatura en control estadístico de la calidad analiza gráficos de control multivariantes para procesos independientes y gráficos de control univariantes para procesos autocorrelacionados de forma separada. Este documento combina las dos situaciones y propone un gráfico de control para procesos multivariantes autocorrelacionados mediante la modelización de la información con modelos en espacio de los estados.

Palabras Clave: Control estadístico de la calidad, modelización en espacio de los estados.

1. Introducción

En muchos procesos de producción está presente la existencia de correlación entre las observaciones, debida bien a la propia naturaleza del proceso o al esquema muestral adoptado. Sin embargo, los desarrollos clásicos de los gráficos de control están basados en la suposición de independencia, abordándose el efecto de la autocorrelación sólo desde una óptica univariante, es decir, tratando de forma independiente las características analizadas.

El tratamiento de la autocorrelación en control estadístico de la calidad se ha centrado en el uso de límites de control empírico o el filtrado de la información mediante modelos de series temporales. En esta segunda alternativa, los desarrollos multivariantes existentes están basados en la modelización mediante series temporales univariantes de las variables de forma individual o en la aplicación de una modelización VAR, propuesta en el trabajo de Pan (2002). En el primer caso, el análisis de la estructura de correlación cruzada entre las variables, gran aportación del enfoque multivariante, queda desvirtuado. Además, en la aplicación de un modelo VAR, si la correlación cruzada original no es fuerte, el proceso de filtrado podría modificar las relaciones más débiles. En estos casos, resulta interesante disponer de resultados comparativos sobre el comportamiento del ARL entre los gráficos originales y de los residuos. Además, la autocorrelación puede ser reflejo también de la existencia de algún

“factor latente” o “inercia” del proceso. En este caso, el análisis de los residuos pondrá de manifiesto el comportamiento “filtrado” del proceso, es decir, el comportamiento de la parte no previsible o innovacional, sin profundizar en el análisis del “factor inercial”.

En este trabajo se propone intentar eliminar el efecto de la autocorrelación mediante la modelización en espacio de los estados. De esta forma obtenemos estimaciones eficientes de la componente residual, las innovaciones del modelo, y del “factor inercial”, los estados. Este planteamiento permite construir un gráfico de estados para analizar el comportamiento general del proceso, considerando la autocorrelación propia del mismo, que complementa la información dada por el gráfico de control para los residuos, constituyendo así una extensión multivariante de la propuesta desarrollada por Alwan y Roberts (1988).

2. Bibliografía

- [1] Alwan, L. C. & Roberts, H. V. (1988). Time-series modelling for statistical process control. *Journal of Business and Economic Statistics* 6, pp: 87-95.
- [2] Alfaro, J. L. (2005): *Control estadístico de la calidad en procesos multivariantes autocorrelacionados. Una aplicación en la industria cuchillera de Albacete*. Tesis Doctoral. Universidad de Castilla-La Mancha.
- [3] Montgomery, D. C. (2004). *Introduction to statistical quality control(5th ed)*. John Wiley & Sons.
- [4] Pan, X. (2002). Var chart: a quality control chart monitoring multivariate autocorrelated processes. *Decision Science Institute 2002 Annual Meeting Proceedings*, pp: 2377-2382.
- [5] Vargas, M. (1999). *Modelización de series temporales multiples en espacio de estados. Análisis de procesos no estacionarios y cointegración*. Tesis Doctoral. Universidad de Castilla-La Mancha.

Un refinamiento del algoritmo de k-medias mediante profundidad y bootstrap

Aurora Torrente¹ y Juan Romo²

¹etorrent@est-econ.uc3m.es, Departamento de Estadística, Universidad Carlos III de Madrid

²juan.romo@uc3m.es, Departamento de Estadística, Universidad Carlos III de Madrid

Abstract

Se proponen dos procedimientos sencillos y computacionalmente muy rápidos para mejorar el comportamiento del procedimiento de k-medias. La primera técnica consiste en alternar el algoritmo de k-medias con la búsqueda del punto más profundo de cada conglomerado y la segunda combina el bootstrap con la obtención del punto más profundo en el espacio de centros de conglomerados. Ambos métodos se aplican tanto a datos reales como simulados.

Claves: k-medias; profundidad; bootstrap.

1. Introducción

Muchas técnicas iterativas como *k-medias* o el algoritmo *EM* son particularmente sensibles a las condiciones iniciales: habitualmente el proceso sólo puede alcanzar uno de los óptimos locales del problema, lo que puede llevar a conclusiones muy diferentes al analizar el mismo conjunto de datos. En la literatura pueden encontrarse distintos procesos de inicialización, que He et al. (2004) agrupa en tres categorías.

Jörnsten (2004) introduce el algoritmo robusto *DDclust* para resolver el problema de minimizar la suma de distancias L_1 desde las observaciones a los centros de los conglomerados más cercanos. En este trabajo, se combinan ideas de Bradley y Fayyad (1998) y Jörnsten (2004) para mejorar el algoritmo de *k-medias*.

2. Los algoritmos

El primer algoritmo procede de la siguiente manera. Supongamos que el conjunto X ha sido dividido en k grupos. Sean C_1, C_2, \dots, C_k los centros de los grupos y sea Δ la función objetivo o distorsión. La primera propuesta consiste en aplicar *k-medias* a X y asignar cada punto al centro más cercano. A continuación, se halla el punto más profundo de cada conglomerado y se utiliza de nuevo *k-medias* comenzando ahora con los puntos más profundos. Si la distorsión Δ es menor que la anterior, se toman los nuevos centros como valores iniciales y, en caso contrario, termina el procedimiento.

El segundo algoritmo propuesto combina remuestreo y la noción de profundidad. Se inspira en la idea de bagging de Breiman (1996) (véase también Bradley y Fayyad, 1998) y en el planteamiento de Leisch (1999). El algoritmo comienza generando una replicación bootstrap X_b del conjunto original de datos X . A continuación, se aplica k -medias a X_b y se repite el procedimiento B veces. Después se aplica k -medias a estos puntos y se obtienen k conglomerados de centros. Se obtiene el más profundo de cada uno de ellos y se aplica k -medias con estos puntos como valores iniciales.

3. Simulaciones

Se han utilizado datos reales y simulados para comparar *PAM*, *CLARA*, *DDclust* y los dos procedimientos propuestos. En todos los casos, los dos nuevos algoritmos presentan un buen comportamiento.

4. Bibliografía

- [1] Bradley P.S. y Fayyad U. (1998). Refining initial points for k-means clustering. In J. Shavlik, editor, *Proceedings of the Fifteenth International Conference on Machine Learning (ICML '98)*, 91–99, San Francisco.
- [2] Breiman L. (1996). Bagging predictors. *Machine Learning* 24 (2), 123-140.
- [3] He J., Lan M. Tan C.L. Sung S.Y. y Low H.B. (2004). Initialization of cluster refinement algorithms: A review and comparative study. *International joint conference on neural networks*, Budapest (Hungary).
- [4] Leisch F. (1999). Bagged clustering. Working Paper 51, SFB “*Adaptive Information Systems and Modeling in Economics and Management Science*”.
- [5] Jörnsten R. (2004). Clustering and classification based on the L_1 -data depth. *Journal of Multivariate Analysis* 90 (1), 67-89.

Una alternativa al precontrol clásico basada en redes neuronales artificiales

Vicent Giner¹, Susana San Matías², Ángeles Caldúch³

¹vigibos@eio.upv.es, ²ssanmat@eio.upv.es, ³mcalduch@eio.upv.es,
Departamento de Estadística e Investigación Operativa Aplicadas y Calidad,
Universidad Politécnica de Valencia

Abstract

El precontrol es una técnica de control estadístico de la calidad que permite evaluar la capacidad de un proceso para generar piezas dentro de especificaciones. En este trabajo construimos una red neuronal artificial que imita el comportamiento de la fase de calificación del precontrol y presentamos resultados comparativos preliminares.

Claves: Control de procesos, precontrol, redes neuronales artificiales.

1. Introducción

El precontrol es una técnica que se engloba dentro de los métodos de control estadístico de procesos. Fue desarrollado en 1954 por un grupo de consultores dirigido por Satterthwaite [3], como alternativa a los gráficos de control \bar{X} - R de Shewhart. A diferencia de éstos, se trata de un método cuyo objeto es supervisar un proceso teniendo en cuenta directamente las especificaciones del producto.

La versión original del precontrol, publicada por Shainin y Shainin [4], propone la delimitación dentro del rango de tolerancias de una zona *verde*, centrada en el valor nominal de la característica de calidad que se desea controlar, y una zona *amarilla*, más alejada del valor nominal, siendo el ancho de la zona verde la mitad del ancho total del rango de tolerancias. Asimismo, los valores fuera del intervalo de especificaciones constituyen la zona *roja*.

En la fase inicial del precontrol, cuyo objetivo es la calificación del proceso, se muestran piezas sucesivamente hasta obtener 5 consecutivas dentro de la zona verde, en cuyo caso se considera que el proceso es correcto, o bien 2 piezas amarillas consecutivas o una roja, caso en el cual se concluye que el proceso necesita ser revisado.

Han surgido en la literatura diferentes alternativas y mejoras al precontrol clásico, con el objeto de mejorar su potencia y disminuir su tasa de falsas alarmas. Entre ellas cabe destacar la propuesta de San Matías et al. [2] y el posterior desarrollo de la misma por Giner y San Matías [1], consistentes en liberar ciertos parámetros del precontrol y encontrar una configuración óptima de dichos parámetros.

2. Un precontrol basado en redes neuronales

En este trabajo presentamos un modelo basado en redes neuronales artificiales que intenta imitar en cierto modo el comportamiento del precontrol. Concretamente, proponemos la construcción de una red de tipo perceptrón multicapa que será usada como alternativa a la fase de calificación del precontrol.

Al igual que el precontrol —en cualquiera de sus versiones—, este procedimiento basado en un modelo neuronal recibirá de alguna manera como *inputs* las sucesivas extracciones de piezas del proceso, y deberá ser capaz de producir como resultado una evaluación sobre la aptitud del proceso para generar unidades dentro de especificaciones. El modelo así construido ha de ser evaluable en términos de la tasa de falsas alarmas y de la capacidad para detectar desviaciones inaceptables del proceso, con el fin de poder establecer una comparación con los métodos propuestos en la literatura.

Teniendo en cuenta todas estas condiciones, presentamos diferentes alternativas para la arquitectura del modelo neuronal, así como para los valores de la capa de entrada. Mostramos algunos resultados preliminares, obtenidos tras entrenar los diferentes modelos con datos de procesos generados mediante simulación. Por último, realizamos una comparación entre los modelos obtenidos mediante este procedimiento y los generados de manera óptima siguiendo la metodología descrita por Giner y San Matías [1].

3. Bibliografía

- [1] Giner V. and San Matías S. (2003). A Methodology for the Optimal Determination of Parameters in Quality Control Tools. *EURO/INFORMS 5th EURO/INFORMS Joint International Meeting. New Opportunities For Operations Research. Abstract Book*. Estambul. Boğaziçi Üniversitesi, 109.
- [2] San Matías S., Jabaloyes J. and Carrión A. (2004). Some modifications of the Classical Pre-Control technique. *Quality and Reliability Engineering International* 20 (1), 47-60.
- [3] Satterthwaite F.E. (1954). *A simple, effective process control method*. Report 54-1. Rath & Strong, Inc.
- [4] Shainin D. and Shainin P. (1989). Pre-control versus \bar{X} & R charting: Continuous or immediate quality improvement? *Quality Engineering* 1, 419-429.

Una aplicación de la regresión logística a la recuperación de imágenes

G. Ayala¹, E. De Ves², J. Domingo², T. León¹, P. Zucarello²

¹Departamento de Estadística e Investigación Operativa

²Departamento de Informática

Universidad de Valencia

Abstract

En este trabajo abordamos un problema de recuperación de imágenes de una base de datos. Proponemos un esquema iterativo basado en el análisis de regresión logística para ordenar un conjunto de imágenes en orden decreciente de sus probabilidades de relevancia.

Palabras Clave: Regresión logística, recuperación de imágenes.

1. Planteamiento del problema y metodología.

El problema en el que estamos interesados es el de recuperación de imágenes en una base de datos general. El usuario desea encontrar una o varias imágenes de cierto tipo, algunas de las imágenes que se encuentran en la base de datos son ‘relevantes’, es decir, se ajustan a su petición hasta cierto punto y otras no lo son en absoluto. Asociada a cada imagen I , podemos considerar la probabilidad de relevancia $p(I)$ como una cantidad que refleja la estimación de la relevancia de la imagen para el usuario. Inicialmente todas las imágenes serán igualmente relevantes, sin embargo, cuando se va obteniendo información acerca de las preferencias del usuario, la medida de probabilidad se irá concentrando en un subconjunto de la base de datos.

El análisis de regresión logística permite calcular las probabilidades de relevancia y ordenar las imágenes de la base de datos de acuerdo con ellas. Esta ordenación cambia de una iteración a la siguiente mediante un mecanismo de “feedback”. En cada etapa el usuario inspecciona una muestra de imágenes y hace selecciones positivas o negativas dependiendo de la relevancia de cada imagen. Esta información se recoge en una variable binaria que indica si una imagen de la muestra es clasificada como “ejemplo” o como “contraejemplo”.

Suponiendo que cada imagen I_j se representa mediante un vector de características K -dimensional x_j la probabilidad de relevancia $p(I_j)$ es una función $p(x_j)$ del vector de características. Estas probabilidades pueden estimarse mediante un modelo de regresión logística.

Para una variable de respuesta binaria Y y p variables explicativas X_1, \dots, X_p , el modelo para $\pi(x) = P(Y=1/x)$ en los valores $x = (x_1, \dots, x_p)$ es $\text{logit } (\pi(x)) = \alpha + \beta_1 x_1 + \dots + \beta_p x_p$, donde $\text{logit } (\pi(x)) = \ln(\pi(x)) / (1-\pi(x))$.

Los parámetros del modelo se obtienen mediante la resolución numérica de las ecuaciones de verosimilitud.

En las primeras etapas del problema es bastante difícil ajustar un modelo de regresión global que tenga en cuenta simultáneamente todas las variables porque el número de datos suele ser inferior al de características. En este caso el modelo de regresión ajustado tendría tantos parámetros no nulos como datos y podría ser que algunas variables importantes no apareciesen en el modelo. Tampoco es realista, para diseñar una aplicación, pedirle al usuario que haga una gran número de selecciones positivas y negativas desde el principio. Para resolver esta cuestión, hemos optado por ajustar varios sub-modelos de regresión más pequeños: cada sub-modelo considera sólo un subconjunto de variables o características semánticamente relacionadas y por tanto ajusta una probabilidad de relevancia diferente a cada imagen. Así pues, se nos planteaba el problema de combinar las distintas probabilidades de relevancia que cada sub-modelo ajusta para cada imagen. Hemos acudido a la metodología de los Conjuntos Difusos que trata de cómo agregar información y hemos utilizado diferentes operadores de agregación.

Hemos trabajado con una base de datos de imágenes y los resultados obtenidos son bastante prometedores.

2. Bibliografía

- [1] Agresti, A. Categorical data analysis, 2nd. Edition. Wiley Series in Probability and Statistics. 2002.
- [2] Berman, A.P. and Shapiro, L.G. "A flexible image database system for content-based retrieval", Computer Vision and Image Understanding, 1999, pp. 175-195.
- [3] Das, M. and Riseman, E.M. "Feature Selection for Robust Color Image Retrieval", Proc. of the ARPS Image Understanding Workshop, vol. 1, 1997", pp. 701-706.
- [4] Veltkamp, R. and Tanase, M., "Content-Based Image Retrieval Systems: A Survey", url = citesear.csail.mit.edu/veltkamp00contentbased.html, institution = "Dept. of Computing Science, Utrecht University", 2000.

Una aproximación logística para modelizar la fecundidad

Pedro César Alvarez Esteban, Cristina Rueda Sabater

Departamento de Estadística e Investigación Operativa
Universidad de Valladolid

Paseo Prado de la Magdalena, s/n. 47005 Valladolid.
Correo electrónico: pedroc@eio.uva.es; crueda@eio.uva.es

Abstract

En este trabajo introducimos un modelo logístico diseñado para estimar la curva de fecundidad. El modelo se formula como un modelo lineal del logaritmo de las odds de la fecundidad y se define en términos de tres parámetros que pueden ser interpretados como medidas de nivel, localización y forma del calendario de la fecundidad.

Claves: Modelo logístico, calendario de fecundidad, quantum.

AMS: 62P25, 62H17, 62J12

1. Introducción

En demografía, las tasas de fecundidad son utilizadas frecuentemente como indicador ya que no están afectadas por cambios en el tamaño y distribución de la población, y además son uno de los datos de entrada de los métodos de proyección de la población más populares. Una característica, común a todos los países, de las tasas de fecundidad por edad es que forman una curva suave. Es el llamado calendario de la fecundidad.

Desde la introducción de los métodos estocásticos en la modelización de la fecundidad, distintos puntos de vista han sido adoptados para el análisis, estimación y proyección de la fecundidad. En primer lugar podemos mencionar los métodos basados en el ajuste de una curva paramétrica al calendario de la fecundidad. Los más destacados en este apartado son los modelos de Coale y Trussell (1974), el modelo basado en ajustar una curva Gamma (Thompson et al (1989)) y recientemente el modelo de Schmertmann (2003), basado en el ajuste de splines cuadráticos con restricciones. En segundo lugar están los métodos basados en técnicas de reducción de la dimensión que transforman linealmente las tasas de fecundidad. El más popular entre estos es el método de Lee y Carter (1992). Una extensión interesante de este último trabajo puede encontrarse en Hyndman y Ullah (2005). Finalmente encontramos los métodos basados en obtener una descomposición quantum-tempo que permitirían calcular indicadores del nivel de la fecundidad ajustados por los cambios en el tiempo de la misma (ver Kohler y Ortega(2002a,b)).

En este trabajo proponemos una nueva forma de modelizar y ajustar el calendario de la fecundidad basado en la utilización de un modelo logístico sobre una tabla de contingencia.

2. Modelo logístico para tablas de contingencia de fecundidad

Si d es el número de grupos de edad de la madre, consideraremos la tabla de contingencia de tamaño $2 \times d$ dada por las frecuencias por filas $n_{1j}(t)$ y $n_{2j}(t)$, donde $n_{1j}(t) = n^o$ de mujeres en el grupo de edad j con nacimientos en el periodo $[t, t+1]$ y $n_{2j}(t) = n^o$ de mujeres en el grupo de edad j que no han sido madres en el periodo $[t, t+1]$. Si $O_j(t)$ son los odds ratio de esta tabla, entonces consideraremos el siguiente modelo:

$$\log O_j = \beta_0 + A_{1j}\beta_1 + A_{2j}\beta_2 = \mathbf{A}'\beta \quad j = 1, \dots, d$$

que denominaremos modelo *LO* y donde \mathbf{A} es una matriz $d \times 3$ con columnas ortogonales y cuyos valores se determinarán de forma apropiada.

3. Comparación con el modelo de Schmertmann.

En la figura se muestran los ajustes utilizando el modelo LO y el modelo propuesto por Schmertmann (modelo QS) para los perfiles de fecundidad de cuatro países diferentes.

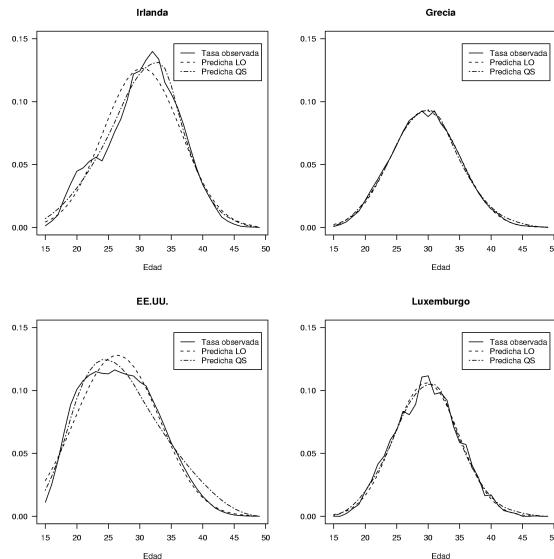


Figura 1: Ajuste de las curvas de fecundidad usando los modelos LO y QS.

Una comparativa entre el wild bootstrap y el bootstrap de los residuos suavizados

Tomás R. Cotos-Yáñez¹

¹cotos@uvigo.es, Departamento de Estadística e Investigación Operativa,
Universidad de Vigo

Abstract

En este trabajo se comparan, por simulación, el remuestreo de los residuos suavizado y el wild bootstrap en modelos heterocedásticos de regresión. La información adicional, sobre la distribución del error, del remuestreo de los residuos suavizados nos permite la construcción de intervalos de predicción bootstrap para la variable respuesta.

Claves: remuestreo bootstrap, intervalos de predicción.

1. Introducción

En la construcción de intervalos de predicción bootstrap en modelos de regresión es necesario conocer la distribución del error. En modelos del tipo

$$y_i = m(x_i) + \epsilon_i \quad 1 \leq i \leq n,$$

con $\epsilon_i \sim F$, es decir, la distribución del error es la misma en todos los puntos, con lo cual el remuestreo uniforme aplicado a los residuos recenterados nos da información sobre esta distribución. Cuando $\epsilon_i \sim F_{x_i}$, por ejemplo en modelos heterocedásticos $\epsilon_i = v(x_i)\epsilon, \epsilon \sim F$, el wild bootstrap solo proporciona información sobre la distribución del error en los puntos $x_i, 1 \leq i \leq n$, por lo que únicamente podemos construir intervalos de predicción bootstrap en estos puntos. Con el remuestro de los residuos suavizados podemos obtener información sobre la distribución del error en puntos distintos a los de la muestra.

El objetivo de este estudio es la comparación, para tamaños de muestras pequeñas, de la estimación de la función del error en modelos heterocedásticos de regresión entre el wild bootstrap y el mecanismo de remuestreo de los residuos suavizados Cotos-Yáñez [1]. Dicha estimación es crucial en la construcción de intervalos de predicción.

2. Remuestreos bootstrap

Sea una muestra aleatoria simple $\{(\mathbf{X}_i, Y_i)\}_{i=1}^n$ del vector (\mathbf{X}, Y) $(r+1)$ -dimensional, donde \mathbf{X}_i es un vector r -dimensional e Y_i la variable respuesta asociada. Supongamos que (\mathbf{X}, Y) sigue el modelo de regresión

$$Y = m(\mathbf{X}) + \epsilon,$$

donde el error se supone de media nula. El wild bootstrap (Wu [5] y Härdle y Mammen [4]) se aplica en modelos heterocedásticos donde se estima cada una de las n funciones de distribución del error con su propio residuo ($\hat{F}_{\mathbf{x}_i}$). Monografías sobre el planteamiento bootstrap en general y específicamente en modelos de regresión pueden verse en Efron [2] y Hall [3].

2.1. Bootstrap basado en los residuos suavizados

Condicionalmente a la muestra, se propone extraer réplicas de una estimación de la función de distribución del error en $\mathbf{X}_0 = \mathbf{x}_0$, dada por

$$\hat{F}_{\mathbf{H}}(r | \mathbf{X} = \mathbf{x}_0) = \sum_{i=1}^n w_i(\mathbf{x}_0, \mathbf{H}) 1_{(\tilde{r}_i \leq r)},$$

con $\{w_i(\cdot, \mathbf{H})\}$ una secuencia de pesos no negativos que sumen 1, \mathbf{H} una matriz de parámetros de suavizado, $\tilde{r}_i = r_i - \sum_{i=1}^n w_i(\mathbf{x}_0, \mathbf{H}) r_i$ y $r_i = y_i - \hat{m}(\mathbf{x}_i)$

3. Conclusiones

A la vista de los resultados, se puede apreciar un comportamiento superior del remuestreo basado en los residuos suavizados frente al wild bootstrap, con independencia de que este último únicamente es aplicable en la obtención de la distribución del error en los puntos de la muestra.

4. Agradecimientos

El autor agradece la financiación del Proyecto MTM2005-01274 (fondos FEDER incluidos) del Ministerio de Educación y Ciencia.

5. Bibliografía

- [1] Cotos-Yáñez, T.R. (2004). *Predicción puntual y por intervalos con modelos parcialmente lineales y aditivos de regresión: Aspectos computacionales y aplicación a datos medioambientales*. Tesis Doctoral. Santiago de Compostela.
- [2] Efron, B. (1993). *An introduction to the bootstrap*. Chapman and Hall.
- [3] Hall, P. (1992). *The bootstrap and Edgeworth expansion*. Springer series in Statistics. Springer.
- [4] Härdle, W. y Mammen, E. (1993). Comparing non parametric versus parametric regression fits. *Annals of Statistics*, **21**, 1926-1947.
- [5] Wu, C.F.J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis (with discussions). *Annals of Statistics* **14**, 1261-1350.

Una mejora del gráfico de control U mediante la aplicación de la metodología de Daudin

Elena Pérez Bernabeu¹, José Jabaloyes Vivas², Andrés Carríon García³

¹elenapb@eio.upv.es, ²jabaloye@eio.upv.es, ³acarrion@eio.upv.es,
Departamento de Estadística e Investigación Operativa Aplicadas y Calidad,
Universidad Politécnica de Valencia

Abstract

En este trabajo se muestra un comparativa entre el gráfico U clásico y un nuevo gráfico U propuesto por los autores [2], basado en artículo de J.J.Daudin [1] que hace lo propio para los gráficos de la media. Se centra en la optimización en un punto y para obtener las soluciones óptimas se ha hecho uso de algoritmos genéticos.

Palabras Clave: Gráfico de control U , Daudin, doble muestreo

1. Introducción

Existen varios tipos de gráficos de control por atributos. Los gráficos np y p se utilizan para controlar la proporción de piezas defectuosas que genera el proceso; el primero exige que el tamaño de muestra sea constante mientras que el segundo no. Los gráficos c y u sirven para controlar el número de defectos; el primero exige que el tamaño de muestra sea constante, mientras que esta exigencia no es necesaria en el gráfico u . Este último gráfico de control que sirve para controlar los defectos por unidad es en el que se va a centrar la aplicación de la metodología de Daudin.

2. Gráfico de Control DS-U

La filosofía propuesta por Daudin [1] consiste en introducir el doble muestreo, de forma se tendrán dos etapas. La primera de ellas, que tendrá lugar siempre, tiene un Límite de Atención Superior (LAS) y otro Inferior (LAI), un Límite de Control Superior (LCS) y otro Inferior (LCI). La segunda etapa tiene dos límites de control, el Superior (LCS1) y el Inferior (LCI1), y sólo se pasará a esta etapa en el caso de que el valor representado en la primera etapa procedente de una muestra n_1 se encuentre entre LCI y LAI, o bien entre LAS y LCS; en ese caso se toma una segunda muestra n_2 y se analiza, tomándose la decisión de si el proceso está bajo control o no. En la figura 1 queda reflejado un ejemplo.

La propuesta realizada en este trabajo consiste en aplicar la metodología de J.J. Daudin al gráfico de control u de defectos, aplicando el doble muestreo para tratar de mejorar las cualidades del gráfico.

3. Resultados y conclusiones

Al aplicar la metodología de Daudin a este tipo de gráficos, las curvas de

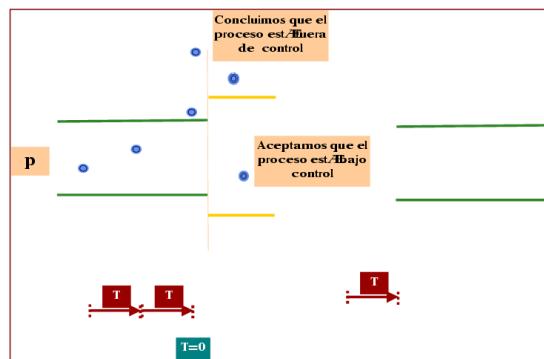


Figura 1: Gráfico DS-U

potencia nos ofrecen información variada según qué punto estemos intentando optimizar. Por ejemplo, para un caso particular en el que el gráfico clásico tiene un tamaño de muestra de 8, un $\alpha=0.064$ y un $p_0=1$, para la optimización en $p_1=0.2$ se obtiene una mejora de la potencia para valores de $p < 1$, pero no así en el resto, donde la curva de potencia del gráfico clásico es mejor que el propuesto. En otros puntos, como en $p_1=1.9$, la mejoría se extiende a todos los valores de p . En este trabajo se analiza detalladamente alguno de estos casos.

4. Bibliografía

- [1] Daudin, J.J. (1992). *Double Sampling X-bar Charts*. Journal of Quality Technology. Vol. 24, 78-87.
- [2] Pérez Bernabeu, E., Jabaloyes Vivas, J.M. y Carrión García, A. (2003). *Aplicación de la Metodología de Daudin a los Gráficos por Atributos*. 27 Congreso Nacional de Estadística e Investigación Operativa. Lleida, 8-11 de abril de 2003.
- [3] Carot Alonso, V. (1998). *Control Estadístico de Calidad*. Servicio de Publicaciones de la Universidad Politécnica de Valencia. SPUPV-98.4007

Una metodología para la aplicación del análisis conjunto al marketing bancario

Ángeles Caldúch¹, Vicent Giner², Susana San Matías³

¹mcalduch@eio.upv.es, ²vigibos@eio.upv.es, ³ssanmat@eio.upv.es,
Departamento de Estadística e Investigación Operativa Aplicadas y Calidad,
Universidad Politécnica de Valencia

Abstract

El análisis conjunto (conjoint analysis) es una técnica usada en investigación de mercados para determinar las preferencias de los clientes ante una serie de atributos de un producto dado. En este trabajo proponemos una nueva forma de aplicar el análisis conjunto en el contexto del marketing bancario.

Claves: Análisis conjunto, marketing bancario, análisis de preferencias.

1. Introducción

El análisis conjunto (conjoint analysis) es una técnica utilizada en marketing para conocer las preferencias de los clientes ante una serie de atributos de un determinado producto, mediante la valoración que ellos realizan de diferentes combinaciones de dichos atributos. Fue introducido por Green y Srinivasan [1] a finales de los años 70, y desde entonces ha sido usado con profusión en investigación de mercados, orientado sobre todo al diseño de nuevos productos.

El modo en que habitualmente se aplica esta técnica es el siguiente. Se extrae un subconjunto de todas las combinaciones posibles de los atributos de manera que forme un diseño ortogonal. Seleccionado un grupo representativo de clientes, se muestra a cada uno de ellos las distintas configuraciones del producto que conforman el diseño ortogonal, y se le pide que les dé una puntuación o que las ordene por grado de preferencia. Una vez recogidos y analizados los datos mediante herramientas de diseño de experimentos, el análisis conjunto permite conocer qué atributos influyen o no en la decisión del cliente, medir la importancia o el peso relativo de cada uno de ellos, e incluso estimar la valoración que obtendría cualquier configuración del producto no incluida en el diseño ortogonal.

2. Nuestra propuesta

En este trabajo abordamos la aplicación del análisis conjunto en el ámbito del marketing bancario. En un área tan competitiva como es actualmente el mundo de los servicios financieros, y dentro de una estrategia de marketing relacional continuo (CRM), es fundamental para una entidad bancaria detectar cuáles son las preferencias de sus clientes y diseñar nuevos productos que cubran dichas

expectativas. Por tanto, el análisis conjunto es una herramienta útil en este contexto.

Sin embargo, aplicar el análisis conjunto clásico en el ámbito bancario entraña un riesgo de imagen. Dar a elegir diferentes opciones a un mismo cliente puede confundirlo y llegar a dañar su confianza en la entidad. Los intentos de aplicar el análisis conjunto en su versión clásica al ámbito bancario (véase por ejemplo [2]) se limitan a analizar preferencias expresadas por los clientes en un sondeo. Nosotros planteamos ir más lejos y trabajar directamente con decisiones reales de compra.

Para ello, y dado que no es posible ofrecer a un mismo cliente *todas* las ofertas que componen el diseño ortogonal, proponemos dividir aleatoriamente el público objetivo de la acción en tantos subgrupos como versiones diferentes del producto haya en dicho diseño. A cada una de estas muestras se le ofrece una variante del producto a través de una acción comercial directa. Como resultado de dichas acciones comerciales, obtenemos la valoración que los clientes público objetivo realizan de cada oferta, medida a través de la eficacia (proporción de clientes que adquirieron el producto) de cada acción.

Una vez analizados e interpretados estos datos, el análisis conjunto permite conocer qué atributos del producto son o no relevantes en la decisión de compra del cliente, y en qué medida, así como qué microcampañas (de todas las posibles) habrían sido las más exitosas, obteniendo una estimación de su eficacia. Si previamente a la aplicación del análisis conjunto se ha realizado una segmentación apropiada del público objetivo, se podrán además obtener resultados particularizados para cada colectivo. Toda esta información, complementada con un adecuado análisis de rentabilidad, puede ser usada para configurar la versión o versiones *óptimas* del producto y servir de base para diseñar posteriores acciones comerciales.

3. Agradecimientos

Este trabajo ha sido parcialmente subvencionado por un proyecto de investigación de la Universidad Politécnica de Valencia (PPI-06-05)

4. Bibliografía

- [1] Green P.E. and Srinivasan V. (1978). Conjoint Analysis in Consumer Research: Issue and Outlook. *Journal of Consumer Research* 5, 103-123.
- [2] Kara A., Kaynak E. and Kucukemiroglu, O. (1994). Credit Card Development Strategies for the Youth Market: The Use of Conjoint Analysis. *The International Journal of Bank Marketing* 12, 30-36

Una métrica entre distribuciones basada en una representación ‘fuzzy’

Gil González-Rodríguez, Ana Colubi, Manuel Montenegro, María Ángeles Gil

{gil,colubi,mmontenegro,magil}@uniovi.es,

Departamento de Estadística, I.O. y D.M., Universidad de Oviedo

Abstract

Sobre la base de una representación ‘fuzzy’ de tipo caracterizador de una variable aleatoria y de una métrica entre números ‘fuzzy’, se establece en esta comunicación una distancia entre distribuciones de probabilidad.

Claves: métrica entre distribuciones, métrica entre números ‘fuzzy’, representación ‘fuzzy’ de una variable aleatoria.

1. Introducción

En González-Rodríguez et al. [1] se ha introducido una familia de representaciones ‘fuzzy’ de una variable aleatoria. Una representación ‘fuzzy’ de una variable aleatoria consiste en una variable aleatoria ‘fuzzy’ (en el sentido de Puri and Ralescu [3]) obtenida por composición de cierta función (de variable real con valores ‘fuzzy’) con la variable original. La familia anterior contiene varias subfamilias de tipo caracterizador. Más concretamente, representaciones ‘fuzzy’ cuyo valor esperado (función de variable real con valores en el intervalo unidad) caracteriza la distribución de la variable original.

Por otro lado, existen diversas métricas entre números ‘fuzzy’ que se han introducido con el fin de desarrollar diferentes procedimientos estadísticos con datos ‘fuzzy’.

Dadas dos variables aleatorias, puede construirse una distancia entre sus distribuciones combinando una de las métricas precedentes con representaciones ‘fuzzy’ de dichas variables. Esta distancia podrá aplicarse, entre otros estudios, al desarrollo de procedimientos inferenciales.

2. Preliminares

Sea $\mathcal{F}_c(\mathbb{R})$ la clase de los números ‘fuzzy’ (NFs), es decir, $\mathcal{F}_c(\mathbb{R}) = \{A : \mathbb{R} \rightarrow [0, 1] \mid A_\alpha = \text{intervalo compacto no vacío } \forall \alpha \in [0, 1]\}$, con $A_\alpha = \{x \in \mathbb{R} \mid A(x) \geq \alpha\}$ if $\alpha \in (0, 1]$, $A_0 = \text{cl}\{x \in \mathbb{R} \mid A(x) > 0\}$.

Dado un espacio de probabilidad (Ω, \mathcal{A}, P) , una aplicación $\mathcal{X} : \Omega \rightarrow \mathcal{F}_c(\mathbb{R})$ es una *variable aleatoria ‘fuzzy’* (VAF) (Puri & Ralescu [3]) si y sólo si para todo $\alpha \in [0, 1]$ la aplicación $\mathcal{X}_\alpha : \Omega \rightarrow \mathcal{P}(\mathbb{R})$ tal que $\mathcal{X}_\alpha(\omega) = (\mathcal{X}(\omega))_\alpha$ para todo $\omega \in \Omega$ es un conjunto aleatorio compacto y convexo.

Si $d_H(\mathcal{X}_0, \{0\}) \in L^1(\Omega, \mathcal{A}, P)$ (con d_H = distancia de Hausdorff), el *valor esperado ‘fuzzy’* de \mathcal{X} (cf. [3]) es el único NF $\tilde{E}(\mathcal{X}; P)$ tal que $(\tilde{E}(\mathcal{X}; P))_\alpha$ = integral de Aumann de $\mathcal{X}_\alpha = [E(\inf \mathcal{X}_\alpha), E(\sup \mathcal{X}_\alpha)]$ para todo $\alpha \in [0, 1]$.

Körner and Näther [4] han definido la *métrica genérica* tal que para $A, B \in \mathcal{F}_c(\mathbb{R})$:

$$[D_K(U, V)]^2 = \int_{\mathbb{S}^2 \times [0, 1]^2} (s_A(x, \alpha) - s_B(x, \alpha))(s_A(y, \beta) - s_B(y, \beta)) dK(x, \alpha, y, \beta),$$

donde $\mathbb{S} = \{-1, 1\}$ y s es la función soporte en $\mathcal{F}_c(\mathbb{R})$ (es decir, $s_A(x, \alpha) = \sup_{z \in A_\alpha} \langle x, z \rangle$ para $x \in \mathbb{S}, \alpha \in [0, 1], \langle \cdot, \cdot \rangle$ = producto escalar).

3. Métrica entre distribuciones de variables aleatorias

González-Rodríguez *et al.* [1] han introducido una familia de aplicaciones $\gamma^c : \mathbb{R} \rightarrow \mathcal{F}_c(\mathbb{R})$ tales que dos variables aleatorias cualesquiera X e Y asociadas a un espacio de probabilidad (Ω, \mathcal{A}, P) tienen la misma distribución si y sólo si $\tilde{E}(\gamma^c \circ X; P) = \tilde{E}(\gamma^c \circ Y; P)$.

Si P_X y P_Y denotan las distribuciones inducidas por X e Y , respectivamente, se define la $D_K^{\gamma^c}$ -distancia entre P_X y P_Y como:

$$\mathcal{D}_K^{\gamma^c}(P_X, P_Y) = D_K(\tilde{E}(\gamma^c \circ X; P), \tilde{E}(\gamma^c \circ Y; P)).$$

4. Observaciones finales

La métrica introducida en la sección precedente puede servir de base para el desarrollo de procedimientos de contraste/estimación sobre la distribución de una variable aleatoria. Entre éstos, la aplicación de técnicas de contraste sobre la media de una VAF (como la que se presenta en Montenegro *et al.* [2]) conducirá a técnicas de bondad de ajuste.

5. Bibliografía

- [1] González-Rodríguez G., Colubi A. and Gil M.A. (2006). A fuzzy representation of random variables: an operational tool in exploratory analysis and hypothesis testing. *Computational Statistics and Data Analysis* (to appear).
- [2] Montenegro M., Colubi A., Casals, M.R. and Gil, M. A. (2004). Asymptotic and Bootstrap techniques for testing the expected value of a fuzzy random variable. *Metrika* 59, 31–49.
- [3] Puri M.L. and Ralescu D.A. (1986). Fuzzy random variables. *Journal of Mathematical Analysis and Applications* 114, 409–422.
- [4] Körner R. and Näther, W. (2002). On the variance of random fuzzy variables. In: Bertoluzza C., Gil M.A. and Ralescu, D.A. (eds.), *Statistical Modeling, Analysis and Management of Fuzzy Data*. Physica-Verlag, Heidelberg, pp. 22–39.

Una propuesta para Análisis Cluster Robusto con grupos heterogéneos

L.A. García-Escudero¹, A. Gordaliza²

Departamento de Estadística e Investigación Operativa, Univ. de Valladolid

¹lagarcia@eio.uva.es, ²alfonsog@eio.uva.es

Abstract

La estimación correcta de las escalas juega un papel trascendental en la obtención de métodos robustos de Análisis Cluster. Esto se hace especialmente presente en algunos métodos recientes que hacen uso de los llamados “concentration steps” (Rousseeuw and van Driessen 1999).

El nuevo método que presentamos, llamado SSC, está basado en combinar adecuadamente un procedimiento de estimación robusta de la covarianza, como es el MCD (Mínimo Determinante de la Covarianza) introducido en Rousseeuw (1985), con los procedimientos de recorte imparcial introducidos en Cuesta-Albertos et al. (1997) en el contexto del Análisis Cluster robusto. El método propuesto parte de un nivel de recorte alto que permite la eliminación de todas las observaciones atípicas y posteriormente desarrolla un proceso iterativo que permite estimar correctamente las escalas de los distintos grupos. El método proporciona también estimaciones de los pesos de los grupos y del nivel de contaminación presente.

Claves: Robusez; Análisis Cluster ; MCD; recorte; parámetro de escala.

AMS: 62H30.

1. Introducción

Una de las técnicas estadísticas robustas más utilizadas en Análisis Multivariante es el estimador “Mínimo Determinante de la Covarianza”, MCD, introducido por Rousseeuw (1985). Este método se basa en la búsqueda de la submuestra de tamaño $[n(1 - \alpha)]$ para la que el determinante de la matriz de covarianzas es mínimo. Este procedimiento recibió un gran impulso a raíz de la publicación del algoritmo “fast-MCD” (Rousseeuw and van Driessen 1999) que permite obtener aproximaciones precisas de este estimador con un bajo coste computacional. El algoritmo fast-MCD se basa en ordenar las observaciones x_1, \dots, x_n por medio de las distancias de Mahalanobis definidas como

$$d_{\Sigma_n}^2(x_i, m) = (x_i - m_n)' \Sigma_n^{-1} (x_i - m_n),$$

donde m_n y Σ_n son estimadores de localización y escala. Estos estimadores se van mejorando en cada paso al tomar la media y covarianza muestrales de las $[n(1 - \alpha)]$ observaciones con menores distancias. Estos pasos (“concentration

steps") hacen disminuir el determinante de la matriz de covarianzas recortada. Para evitar mínimos locales, el algoritmo necesita extraer aleatoriamente varios valores iniciales de m_n y Σ_n . Este procedimiento tiene similitudes con el algoritmo clásico para calcular las k -medias, el cual procede en cada paso a agrupar aquellas observaciones con menores distancias Euclídeas a cada uno de los k centros establecidos en el paso anterior. Este algoritmo también necesita extraer aleatoriamente varios valores iniciales para los centros.

Por otra parte, numerosos autores actuales consideran que el Análisis Cluster y la Estimación Multivariante Robusta son temas íntimamente relacionados, de modo que algunas ideas y técnicas pueden ser útiles simultáneamente en ambos campos. Esto nos lleva a pensar en trasladar al Análisis Cluster robusto los "concentration steps". En el trabajo se discuten distintas posibles implementaciones de esta idea y se comprueba que la existencia de diferentes escalas para los grupos tiene un efecto muy perjudicial sobre una posible implementación directa de la misma. Por ello, presentaremos un procedimiento iterativo diseñado para evitar este problema. El procedimiento comienza con una solución inicial obtenida a partir del algoritmo de Gallegos (2002) aplicado con un nivel de recorte alto y posteriormente se van añadiendo observaciones de manera paulatina, recalculando los pesos y las escalas en cada paso. La monitorización de estos estimadores permite obtener de una manera gráfica un valor aproximado para la proporción de contaminación presente en los datos.

2. Agradecimientos

Este trabajo ha sido parcialmente financiado por el *Ministerio de Educación y Ciencia y FEDER*, Proyecto de Investigación MTM2005-08519-C02-01 y por la *Consejería de Educación y Cultura de la Junta de Castilla y León*, Proyecto de Investigación PAPIJCL VA074/03.

3. Bibliografía

- [1] Cuesta-Albertos, J.A., Gordaliza, A. and Matrán, C. (1997), "Trimmed k -means: An attempt to robustify quantizers", *Ann. Statist.*, **25**, 553-576.
- [2] Gallegos, M.T. (2002), "Maximum likelihood cluster with outliers", in Jajuga, K., Sokolowski, A. and Bock, H.-H. (eds.): *Classification, Clustering and Data Analysis*, Springer-Verlag, Berlin, 247-255.
- [3] Rousseeuw, P.J. (1985), "Multivariate Estimation with High Breakdown Point", in *Mathematical Statistics and Applications*, edited by W. Grossmann, G. Pflug, I. Vincze, and W. Wertz, Reidel Publishing Company, Dordrecht, 283-297.
- [4] Rousseeuw, P.J. and Van Driessen, K. (1999), "A Fast Algorithm for the Minimum Covariance Determinant Estimator", *Technometrics*, **41**, 212-223.

Una variante del análisis multicanónico para tablas múltiples heterogéneas con grupos de variables cuantitativas y grupos de datos textuales. Aplicación en marketing

Mónica Bécue-Bertaut¹, Jérôme Pagès², Ramón Álvarez Esteban³,
José Luis Vázquez Burguete⁴

¹monica.becue@upc.edu, Statistics and Operations Research Department,
Universitat Politècnica de Catalunya, Spain

²jerome.pages@agrocampus-rennes.fr, Agrocampus-Rennes, France

³dderae@unileon.es, Statistics and Operations Research Department,
Universidad de León, Spain

⁴ddejvb@unileon.es, Marketing Depart., Universidad de León, Spain

Abstract

We propose an original multicanonical method able to integrate groups of continuous variables and textual data, taking advantage of the possibility of extending CA to different metrics. This method is applied to build up a global score, compromise between the mark and the free comments, in the case of the evaluation of wines of Castile and Leon by an expert.

Keywords: Canonical correlation analysis; Multiple factorial analysis; Textual data

1. Introducción

Proponemos un método de correlación canónica que integra grupos de datos textuales. Este método presenta un interés particular en marketing donde es frecuente evaluar productos mediante puntuaciones clásicas pero completadas con comentarios libres. Se aplica el método resultante a vinos de Castilla y León evaluados por una puntuación numérica y notas de cata en forma de comentarios libres.

2. Tabla múltiple heterogénea a analizar

Como es usual en el análisis estadístico de datos textuales [1], los comentarios libres llevan a construir una tabla de contingencia, o tabla léxica, que contiene la frecuencia con la cual cada palabra (en columna) se utiliza para cualificar a un producto (en fila). Las puntuaciones constituyen variables cuantitativas. Por tanto, se obtiene una tabla múltiple heterogénea en la cual los productos-fila están descritos por un grupo cuantitativo y un grupo textual, de tipo tabla de contingencia.

3. Metodología propuesta

El objetivo de construcción de una puntuación global, compromiso entre la puntuación numérica y las notas de cata, se corresponde con el objetivo de

varios métodos, reagrupados bajo el nombre de análisis multicanónico. Dichos métodos buscan identificar la variable general (o una sucesión de variables generales no correlacionadas) que tenga la mayor correlación con los distintos grupos de variables. Esta variable general puede considerarse como el mejor compromiso entre los grupos de variables.

La metodología que proponemos es una extensión de uno de dichos métodos, el análisis factorial múltiple AFM [2]. Dicho método mide la relación entre una variable y un grupo de variables mediante una variante del índice de redundancia [3]. La introducción de un grupo de datos textuales plantea varios problemas que se resuelven mediante:

- la extensión del AFM a tablas de contingencias (análisis factorial múltiple para tablas de contingencia AFMTC [4]);
- la modificación del peso de las filas (métrica en el espacio de las columnas) inducido por la tabla de contingencia (correspondientes al margen sobre las filas) para conservar a dichas filas un peso uniforme (o el peso inducido por el método de muestreo).

4. Resultados

Se obtiene una puntuación global para los vinos muy fuertemente correlacionada con la puntuación numérica ($r=0.95$). Los dos tipos de evaluación contribuyen de manera equilibrada a su varianza (48.3 % aportado por las notas textuales y 51.7 % por la puntuación numérica). Se puede cuantificar la aportación de cada palabra, lo que constituye una información de interés para los enólogos. El método permite unir la solidez de la apreciación cuantitativa y la riqueza de los comentarios libres.

5. Agradecimientos

Este trabajo ha sido parcialmente financiado por la Junta de Castilla y León.

6. Bibliografía

- [1] Lebart, L., Salem A. and Bécue M. (2000). *Análisis estadístico de textos*. Milenio.
- [2] Escofier, B. and Pagès, J. (1994). Multiple factor analysis: AFMULT package. *Comp. Statistics & Data Analysis* 18, 121-140.
- [3] Van den Wollenberg, A.L. (1977). Redundancy analysis: an alternative for canonical correlation. *Psychometrica* 42, 207-219.
- [4] Bécue, M. and Pagès J. (2004). A principal axes method for comparing contingency tables: MFACT. *Comp. Statistics & Data Analysis* 45(3), 481-503.

Valores composicionales por debajo del límite de detección: un reemplazamiento paramétrico

Javier Palarea Albaladejo¹, Josep Antoni Martín-Fernández²

¹jpalarea@pdi.ucam.edu, Depto. de Informática de Sistemas, Universidad Católica San Antonio

²josepantoni.martin@udg.es, Depto. de Informática y Matemática Aplicada, Universidad de Girona

Abstract

En este trabajo se presenta un reemplazamiento paramétrico para sustituir valores por debajo del límite de detección en muestras de datos composicionales. Siempre en comparación con otros métodos existentes, se analizan sus propiedades y se evalúa su rendimiento mediante la aplicación del reemplazamiento a un conjunto de datos reales.

Palabras Clave: datos composicionales, algoritmo EM, ceros por redondeo

1. Introducción

Un dato composicional es la realización de un vector aleatorio D-dimensional $\mathbf{x} = (x_1, \dots, x_n)$ en el que cada parte o componente del mismo verifica que

$$x_i > 0 \quad \forall i = 1, \dots, D \quad \text{y} \quad x_1 + \dots + x_D = 1.$$

En [1] se establece que el espacio muestral para este tipo de datos es el simplex S^D definido como

$$S^D = \{(x_1, \dots, x_D) : x_1 > 0, \dots, x_D > 0; x_1 + \dots + x_D = 1\}.$$

Las operaciones perturbación, $\mathbf{p} \oplus \mathbf{x} = C(p_1 x_1, \dots, p_D x_D)$ definida sobre $S^D x S^D$, y la transformación por potencia, $\alpha \otimes \mathbf{x} = C(x_1^\alpha, \dots, x_D^\alpha)$ definida sobre $R x S^D$, inducen una estructura de espacio vectorial sobre S^D . El operador clausura $C(\mathbf{x})$ se define como $C(\mathbf{X}) = (x_1 / \sum x_i, \dots, x_d / \sum x_i)$.

Aitchison ([1]) establece que un análisis adecuado de este tipo de datos debe basarse en el estudio de las magnitudes relativas y no en las magnitudes absolutas. En consecuencia, cualquier aseveración sobre una composición debe realizarse en términos de cocientes entre las componentes. Los log-cocientes $\ln(x_i/x_j)$ son más fáciles de manejar matemáticamente que los cocientes, y además la transformación log-cociente proporciona una correspondencia uno a uno entre vectores composicionales y vectores log-cociente transformados en un espacio real. Esto evita el problema de un espacio muestral original restringido,

el simplex, pasando a trabajar en un espacio real multivariante no restringido, lo que permite entonces emplear las técnicas multivariantes disponibles para espacios reales. En consecuencia, la metodología log-cociente se ha convertido en el enfoque estándar para el análisis estadístico de datos composicionales.

Ninguna transformación log-cociente puede aplicarse si el vector composicional contiene partes con valores nulos, y esto justifica la necesidad de métodos de imputación para reemplazarlos. La presencia de valores nulos en una muestra de datos composicionales se debe habitualmente a que la parte correspondiente toma valores por debajo del límite de detección de los aparatos de medida, y es frecuente referirse a ellos como ceros por redondeo. Para éstos pueden emplearse métodos de imputación, ya que en esencia un cero por redondeo es como un valor perdido que no ha podido observarse ([3]).

En este trabajo proponemos un enfoque paramétrico para reemplazar los ceros por redondeo basado en la transformación log-cociente aditiva. Este enfoque tiene en cuenta la naturaleza especial tanto de los datos composicionales como de los ceros por redondeo. Nuestro algoritmo es una modificación del conocido algoritmo EM ([2]) el cual, a partir de la información contenida en la parte observada de la muestra de datos, converge hacia los estimadores máximo-verosímiles de los parámetros desconocidos del modelo para los datos completos. En primer lugar, revisaremos las estrategias existentes en la literatura para reemplazar ceros por redondeo, tanto desde un punto de vista paramétrico como no paramétrico ([3]). A continuación presentaremos nuestro reemplazamiento y analizaremos sus propiedades desde una perspectiva composicional. Finalmente, mediante el estudio de un conjunto de datos de un caso real, compararemos su rendimiento frente a otros métodos con el fin de extraer conclusiones y marcar las futuras líneas de investigación.

2. Bibliografía

- [1] Aitchison, J. (1986). *The statistical analysis of compositional data*. Chapman and Hall, London. Reprinted in 2003 by Blackburn Press.
- [2] Dempster, A. P., Laird N. M. y Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *J. Royal Statist. Soc., Series B* 39, 1–38.
- [3] Martín-Fernández, J. A., Barceló-Vidal, C., y Pawlowsky-Glahn, V. (2003). Dealing with zeros and missing values in compositional data sets. *Mathematical Geology* 35, 253–278.

Variables auxiliares óptimas para la estimación de cuantiles

M.D. Martínez Miranda¹, M. Rueda¹, A. Arcos¹

¹mmiranda@ugr.es; Departamento de Estadística e I. O., Universidad de Granada

Abstract

Se trata el problema de seleccionar variables auxiliares óptimas para la estimación de los cuantiles de una población finita. Bajo una perspectiva asintótica se obtienen condiciones de optimalidad para un diseño cualquiera y una amplia clase de estimadores. Los asintóticos se evalúan en la práctica mediante simulaciones con poblaciones reales.

Claves: Cuantil, población finita, diseño muestral, información auxiliar.

AMS: 62G05

1. Introducción

En la práctica del muestreo de poblaciones finitas es habitual disponer de información auxiliar cuyo empleo conveniente en la fase de estimación, permite obtener ganancias sustanciales en la precisión de los estimadores de los parámetros.

El problema de estimar parámetros lineales, tales como medias o totales, ha sido un tópico muy estudiado en la literatura existente en muestreo, no ocurre igual con la estimación de parámetros no lineales, como cuantiles, aquí la literatura es más reciente y fundamentalmente motivada por la extensión y adaptación de los métodos de estimación existentes para el caso de parámetros lineales, tales como los estimadores diferencia (ver por ejemplo Rueda et al., 2003, 2004) o en general los denominados métodos indirectos de estimación.

Si se dispone de una única variable auxiliar x , relacionada con la variable de interés y , y se conocen sus cuantiles poblacionales, estos pueden ser utilizados en la construcción del estimador dando lugar a importantes mejoras con respecto a los métodos directos (que no usan información auxiliar). Cuando la información auxiliar disponible es relativa a varias variables (x_1, \dots, x_p), se han propuesto extensiones multivariantes de los estimadores indirectos, tales como los estimadores de razón y diferencia (Rueda y Arcos, 2002). Sin embargo, aunque son posibles tales estimaciones, al menos en teoría, resulta innegable el hecho de que utilizar demasiadas variables auxiliares puede resultar no sólo complejo sino también ineficiente, puesto que pueden existir algunas redundantes.

Recientemente Agarwal et al. (2001) han estudiado el problema de seleccionar (automáticamente) entre dos variables auxiliares disponibles para construir estimadores univariantes con precisión óptima. Estos autores, bajo un planteamiento mucho más restrictivo que el que asumimos en este trabajo (sólo consideran la estimación de parámetros lineales, en concreto el total poblacional de y y como método de estimación el de tipo razón) obtienen condiciones teóricas para descubrir y seleccionar qué variable es la óptima, que vienen referidas al coeficiente de correlación entre las variables y los coeficientes de variación.

En este trabajo, considerando unos planteamientos iniciales similares a los de Agarwal et al. (2001), se estudia el problema de elegir variables auxiliares óptimas para la estimación, desde una amplia clase de estimadores indirectos, de los cuantiles de la población finita. Las condiciones obtenidos se pueden particularizar a estimadores que son óptimos dentro de la clase y a diseños muestrales sencillos y resultan interesantes en cuanto que son fácilmente evaluables en la práctica y además permiten una gran intuición desde el punto de vista del investigador.

Se incluye un estudio de simulación usando diversas poblaciones reales para evaluar cómo tales condiciones, obtenidas bajo supuestos asintóticos, se mantienen en la práctica.

2. Bibliografía

- [1] Agarwal, S.K., Al-Mutairi, D.K. y Kumar, P. (2001) Selecting between two concomitant variables in ratio method of estimation. *Applied Mathematics and Computation*, 124, 227–234.
- [2] Rueda, M. y Arcos, A. (2002) The use of quantiles of auxiliary variables to estimate medians. *Biometrical Journal*, 44, 619–632.
- [3] Rueda, M., Arcos, A. y Martínez-Miranda, M.D., 2003, Difference estimators of quantiles in finite populations. *Test*, 12, 481–496.
- [4] Rueda, M., Arcos, A., Martínez-Miranda, M.D. y Román, Y. (2004) Some improved estimators of finite population quantile using auxiliary information in sample surveys. *Computational Statistics and Data Analysis*, 45, 825–848.

Escala y evaluación de modelo Poisson-GPD

M.I. Ortego¹, J.J. Egozcue²

¹ma.isabel.ortego@upc.edu, Departamento de Matemática Aplicada III,
Universitat Politècnica de Catalunya

²juan.jose.egozcue@upc.edu, Departamento de Matemática Aplicada III,
Universitat Politècnica de Catalunya

Abstract

Se estudia un problema de peligrosidad de precipitación, dada en escala logarítmica y ordinaria. Se establece un modelo Poisson-GPD. La estimación de parámetros es bayesiana. Se establecen las hipótesis *a priori*, suponiendo que se trata de un fenómeno finito. Mediante contrastes bayesianos se evalúa la compatibilidad del modelo y los datos.

Claves: escala logarítmica, bondad de ajuste, p-valor bayesiano.

1. Introducción

La selección de una escala adecuada para la descripción de un fenómeno es un elemento importante en la modelización estadística. Muchas características físicas, económicas o sociales se miden por magnitudes positivas que muestran una escala relativa (la magnitud nula es inalcanzable o corresponde a fenómenos diferentes; los errores relativos tienen más sentido que los absolutos). Sin embargo, es frecuente relegar esta parte del modelo suponiendo que la escala de los datos no tiene consecuencias relevantes; y así se toma como escala *natural* de los datos aquella en que fueron presentados inicialmente, sin otras consideraciones.

Se discute los problemas de escala en el contexto del análisis de peligrosidad, *hazard*, debido a un fenómeno asociado a sucesos que ocurren como un proceso de Poisson. Estos sucesos se evalúan con una magnitud. Ejemplos típicos son los sismos, vientos, la precipitación y el oleaje. Las magnitudes, frecuentemente positivas, pueden ser transformadas a una escala relativa tomando logaritmos ([1], [4]). Los resultados que se obtienen con los datos en escala logarítmica parecen mejorar los que se obtienen en la escala original. Estas mejoras pueden cuantificarse en lo relativo al ajuste del modelo a los datos. Por ello las medidas de bondad de ajuste y consistencia del modelo son esenciales para la validación de la escala de los datos.

Se consideran sucesos de Poisson, cuyas magnitudes tienen excesos sobre un umbral de referencia con una distribución de Pareto Generalizada (GPD). Además los excesos son independientes entre si y del proceso de Poisson. El resultado es un modelo con tres parámetros (tasa de Poisson, parámetros de forma y escala de GPD) considerados aleatorios, a los que debe añadirse un

umbral de referencia. Por otra parte, en los estudios de peligrosidad la escasez de datos es habitual. De aquí que la utilización de información *a priori* y el control de la incertidumbre de la estimación sean esenciales. Por ello, los modelos bayesianos son muy útiles ([2]). Esto introduce en el modelo la caracterización del *priori* de los parámetros. El modelo completo queda descrito por la escala y el umbral de los datos, los parámetros ya mencionados y su modelo *a priori*.

Dos puntos críticos pueden utilizarse para discernir sobre la escala adecuada de los datos: el ajuste de la GPD a los excesos observados y la consistencia global del modelo con los datos. El primer punto puede cuantificarse mediante un test de bondad de ajuste en contexto bayesiano. El segundo reclama una inspección del modelo global que incluye las distribuciones utilizadas, el umbral y las hipótesis *a priori*.

Se presenta un ejemplo de precipitación diaria. Una hipótesis (*a priori*) realista es suponer que la distribución de excesos debe tener el soporte de los datos limitado superiormente (dominio de atracción de Weibull) puesto que ciertas precipitaciones son físicamente imposibles. Esto debe ser válido tanto para las precipitaciones como para sus respectivos logaritmos. En estas condiciones se utiliza un test de bondad de ajuste predictivo para comparar el ajuste de la GPD a ambos conjuntos de datos.

También se diseña una discrepancia, relacionada con el parámetro de forma de la GPD, y con ella se procede a calcular un p-valor Bayesiano ([3]). Mediante este p-valor se evalúa la compatibilidad del modelo global con los datos en ambas escalas. El modelo resulta incompatible con los datos debido especialmente a la elección del *priori* (limitación a GPD de soporte finito). Para los datos en escala ordinaria, el p-valor indica incompatibilidad, mientras que se mantiene la compatibilidad para los datos en escala logarítmica.

2. Bibliografía

- [1] Egozcue, J.J., Pawlowsky-Glahn, V. and Ortego, M.I. (2005). Wave-height hazard analysis in Eastern Coast of Spain. Bayesian approach using generalized Pareto distribution. *Advances in Geosciences* 2, 25-30.
- [2] Egozcue, J.J. and Ramis, C. (2001). Bayesian hazard analysis of heavy precipitation in Eastern Spain, *Int. J. Climatology* 21, 1263–1279.
- [3] Gelman, A., Meng, X.L. and Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies, *Statistica Sinica* 6, 733–807.
- [4] Pawlowsky-Glahn, V., R. Tolosana-Delgado and J.J. Egozcue. (2005). Scale effect in hazard assessment. Application to daily rainfall, *Advances in Geosciences*, 2, 117-121.

¿Cómo obtener razones de fallo con forma de bañera?

Jorge Navarro Camacho¹, Pedro J. Hernández Martínez²,
Eshetu T. Wondmagegnehu³

¹jorgenav@um.es, Departamento de Estadística, Investigación Operativa y
Computación, Universidad de Murcia

²pedroj.hernandez@upct.es, Departamento de Métodos Cuantitativos e
Informáticos, Universidad Politécnica de Cartagena

³eshetu.wondmagegnehu@wmich.edu, Department of Statistics, Western
Michigan University.

Abstract

Las distribuciones cuyas razones de fallo tienen forma de bañera (BFR) tienen gran importancia práctica en Fiabilidad y Supervivencia. En este trabajo mostramos cómo, desde el punto vista práctico, se pueden obtener distribuciones BFR mediante mixturas de dos distribuciones con razón de fallo creciente.

Claves: Fiabilidad, razón de fallo, mixturas.

AMS: 60E15, 60K10

1. Introducción

Si la variable aleatoria positiva X representa la duración de un componente industrial o un determinado ser vivo, la razón de fallo (o de riesgo) de X se define por $r(t) = f(t)/R(t)$ para valores reales t tales que $R(t) > 0$, donde $R(t) = Pr(X \geq t)$ es la función de fiabilidad y $f(t) = R'(t)$.

La función de fallo sirve para medir el envejecimiento (es decir el efecto del paso del tiempo) de la unidad representada por X . Es bien conocido que el modelo exponencial tienen razón de fallo constante por lo que, en este contexto, representa a unidades que carecen de envejecimiento.

La función de fallo también es utilizada para caracterizar, comprar y clasificar modelos de distribuciones de probabilidad. Así, por ejemplo, se dice que una distribución es IFR (DFR) si tiene razón de fallo creciente (decreciente) en sentido amplio. De esta forma, el modelo exponencial es, a la vez, IFR y DFR.

Se dice que una distribución tiene razón de fallo con forma de bañera (bathtub shaped failure rate o BFR) si existen puntos $0 < t_1 \leq t_2$ tales que r es estrictamente decreciente en $(0, t_1)$, es constante en $[t_1, t_2]$ y es creciente en (t_2, ∞) . Los puntos t_1 y t_2 se denominan puntos de cambio.

En fiabilidad y supervivencia existen numerosos conjuntos de datos con razones de fallo (estimadas) con forma de bañera. El periodo inicial donde la razón

de fallo es decreciente se interpreta como un periodo de ‘quemado’ (burn-in) en donde se detectan las unidades con defectos de fabricación. Este periodo es de gran importancia para establecer los periodos de prueba y garantía. El segundo periodo con razón de fallo aproximadamente constante representa la vida útil del componente y el último, con razón de fallo creciente, el de desgaste o envejecimiento.

Durante los últimos 20 años se han propuesto diversos métodos para obtener distribuciones BFR (véase [1] y las referencias incluidas en dicho artículo).

En este trabajo mostramos como, desde un punto de vista práctico, las distribuciones BFR se pueden obtener mediante mixturas de dos distribuciones IFR. Este método explicaría cómo las distribuciones BFR provienen de la mezcla de dos tipos de componentes, los fabricados correctamente y los defectuosos.

Concretamente, en este trabajo consideramos mixturas de una distribución exponencial (representando a las unidades defectuosas) y una distribución Weibull con razón de fallo creciente (representando a las unidades sin defectos). Para este tipo de mixturas demostramos, bajo condiciones razonables, que su distribución es, desde el punto de vista práctico BFR. Pueden obtenerse resultados similares para mixturas de otros tipos de distribuciones IFR. Por ejemplo, en [2] se estudian mixturas de distribuciones normales truncadas o con razones de fallo lineales. Los resultados de este trabajo están incluidos en [3].

2. Agradecimientos

Parcialmente financiado por el Ministerio de Ciencia y Tecnología, proyecto BFM2003-02947 y la Fundación Seneca de la Comunidad Autónoma de la Región de Murcia, proyecto 00698/PI/04.

3. Bibliografía

- [1] Lai, C.D., Xie M. y Murthy, D.N.P. (2001). Bathtub-Shaped failure rate life distributions. *Advances in Reliability*, 69-104. North-Holland.
- [2] Navarro, J. y Hernández, P.J. (2004). How to obtain bathtub shaped failure rate models from normal mixtures. *Probab. Eng. Inform. Sci.* 18(4), 511-531.
- [3] Wondmagegnehu, E., Navarro, J. y Hernández, P.J. (2005). Bathtub shaped failure rates from mixtures: a practical point of view. *IEEE Transactions on Reliability* 54(2), 270-275.

Sección 3
Estadística Pública

Comparación de soluciones a la multicolinealidad en regresión logística funcional

M. Escabias¹, A.M. Aguilera², M.J. Valderrama³

¹escabias@ugr.es, ²aaguiler@ugr.es, ³valderra@ugr.es

Departamento de Estadística e I.O. Universidad de Granada

Abstract

Dos métodos se comparan para la solución del problema de multicolinealidad en regresión logística funcional: uno basado en análisis en componentes principales y otro basado en regresión PLS.

Claves: Regresión logística funcional, ACP funcional, Regresión PLS.

1. Introducción

La regresión logística funcional es uno de los métodos más utilizados para predecir una variable escalar dicotómica a partir de una variable funcional (ver Escabias *et al.* (2004)). La estimación del modelo pasa por la consideración de que los elementos funcionales del mismo (curvas muestrales y función parámetro) pertenecen a un mismo espacio generado por una base de funciones (Ramsay and Silverman, 2005), en cuyo caso el modelo funcional se convierte en un modelo logístico múltiple con gran multicolinealidad. Para evitar los problemas que provoca la multicolinealidad (estimaciones poco precisas de los parámetros), se han propuesto diversas soluciones en la literatura. En este artículo se pretenden comparar los métodos clásicos basados en componentes principales, con otros desarrollados en los últimos años basados en regresión PLS.

2. El modelo logístico. Soluciones a la multicolinealidad

Para formular el modelo logístico funcional, consideraremos una muestra de curvas $\{x_i(t) : i = 1, \dots, n, t \in T\}$, de una variable funcional $X(t)$ y asociada a cada curva, una observación de una variable respuesta binaria que resumiremos en el vector $Y = (y_1, \dots, y_n)'$ con $y_i \in \{0, 1\}$, $i = 1, \dots, n$. Entonces el modelo se formula como $y_i = \pi_i + \varepsilon_i$, $i = 1, \dots, n$, siendo ε_i los errores independientes centrados y de varianza $\pi(1 - \pi)$, y π_i la probabilidad de que la i -ésima observación de la variable respuesta tome el valor 1 dada la observación $x_i(t)$ de la variable funcional. Dicha probabilidad se expresa en términos de las curvas muestrales como $\pi_i = \exp\{l_i\} / (1 + \exp\{l_i\})$ con $l_i = \alpha + \int_T x_i(t) \beta(t) dt$ y $\alpha, \beta(t)$ parámetro y función parámetro a estimar. Las cantidades l_i se conocen como transformaciones logit.

Ante la imposibilidad de estimar la función parámetro del modelo por máxima verosimilitud (ver Ramsay and Silverman (2005)) es usual considerar que tanto

$\beta(t)$ como $x_i(t)$ pertenecen a un espacio generado por una base de funciones de manera que $x_i(t) = a'_i \Phi(t)$ y $\beta(t) = \beta' \Phi(t)$ con $\Phi(t) = (\phi_1(t), \dots, \phi_k(t))'$ el vector de elementos básicos y β', a'_i vectores de coeficientes básicos de la función parámetro y las curvas muestrales respectivamente. Bajo estas condiciones, el modelo logístico funcional se convierte en un modelo múltiple con gran multicolinealidad, expresándose el vector de las transformaciones logit $L = (l_1, \dots, l_n)'$ en términos matriciales como $L = A\Psi\beta$ con A la matriz que tiene por columnas los coeficientes básicos de las curvas muestrales, y $\Psi = (\psi_{ij})$ la que tiene como entradas los productos escalares entre las funciones básicas $\psi_{ij} = \int_T \phi_i(t) \phi_j(t) dt$.

La estimación de los parámetros de este modelo múltiple es poco precisa debido a la gran multicolinealidad existente en el modelo. En este artículo nosotros comparamos dos soluciones a este problema: una basada en ACP funcional y otra basada en regresión PLS. En ambas situaciones se sustituye la matriz $A\Psi$ por una factorización en términos de variables incorreladas $A\Psi = \Gamma V$, en cuyo caso podríamos dar una estimación precisa del vector de coeficientes básicos de la función parámetro de la forma $\beta = V'\gamma$, siendo γ el vector de parámetros del modelo que tiene como covariables las columnas de Γ . Los métodos que se comparan son las formas de factorizar la matriz $A\Psi$:

- utilizando el clásico ACP funcional (ver por ejemplo Ramsay and Silverman (2005)),
- utilizando el algoritmo de obtención de componentes PLS para regresión logística propuesto por Bastien *et al.* (2005).

3. Agradecimientos

Esta investigación ha sido financiada con el proyecto MTM2004-5992 de la Dirección General de Investigación, Ministerio de Ciencia y Tecnología.

4. Bibliografía

- [1] Escabias, M., Aguilera, A.M. and Valderrama, M.J. (2004). Principal component estimation of functional logistic regression: discussion of two different approaches. *Journal of nonparametric statistics* 16 (3-4), 365-384.
- [2] Ramsay, J. O. and Silverman, B. W., (2005). *Functional Data Analysis* New York Springer-Verlag.
- [3] Bastien, P., Esposito-Vinzi, V. and Tenenhaus, M., (2005). PLS generalised linear regression. *Computational Statistics and Data Analysis*, 48 (1): 17-46.

Datos faltantes en estudios longitudinales: un nuevo enfoque de análisis

Wilmer Fermín¹, Purificación Galindo², Javier Martín²

¹wilmer_fermin@hotmail.com, Departamento de Estadística, Universidad de Oriente-Venezuela

²pgalindo@usal.es, Departamento de Estadística, Universidad de Salamanca-España

Abstract

Se propone un método para analizar datos longitudinales de respuestas continuas con abandonos. Éste consiste en estratificar la población mediante un “desvío de abandono” y usar un modelo multinivel para el análisis. El método propuesto se compara con otros; los resultados demuestran que nuestra propuesta es más sensible en la detección de efectos de tratamientos.

Palabras Clave: Tiempo de abandono, desvío, modelo multinivel

1. Introducción

Un problema frecuente en estudios longitudinales es el abandono, un tipo de dato faltante en el que uno o varios sujetos dejan el estudio prematuramente, después de haber participado en una o más ocasiones consecutivas. El abandono puede sesgar severamente las estimaciones de los efectos de las covariables del estudio. Existen tres tipos abandono: abandono completamente aleatorio (MCAR), abandono aleatorio (MAR) y abandono no-aleatorio (MNAR) (ver [1]). Estadísticamente el más problemático es el MNAR, puesto que no se puede ni excluir del análisis a los sujetos que lo presentan ni imputar algún valor sin introducir sesgos en los parámetros estimados del modelo. Los modelos de patrones mezclados son un enfoque para modelizar datos longitudinales con abandonos MNAR [2]. Recientemente se han desarrollado varias enfoques para tratar abandonos no-aleatorios orientados en los modelos de patrones mezclados (ver [3, 4, 5]).

La estratificación con el tiempo de abandono produce grupos heterogéneos internamente y, a la vez, homogéneos entre ellos ya que, en un tiempo específico es posible encontrar MCAR, MAR y MNAR simultáneamente y, como el comportamiento de éstos es diferente, (ver [2, 6]), la posible ocurrencia simultánea hace difícil detectar efectos de tratamientos. Se propone clasificar los sujetos que abandonan de acuerdo al comportamiento de su respuesta en relación con los de su tratamiento que si completan el estudio. Para el j -ésimo sujeto dentro del k -ésimo tratamiento, que abandona en la r -ésima ocasión, se define su “desvío de abandono” mediante:

$$D_{jk} = Y_{rjk} - \bar{Y}_{rk}^C ; \quad j = 1, \dots, m_k; \quad k = 1, \dots, K \text{ y } r = 1, \dots, n - 1$$

La magnitud de D_{jk} indica que tan discrepante es la respuesta antes del abandono. Una vez agrupados los sujetos de acuerdo al desvió de abandono, se realiza un análisis mediante un modelo multínivel de tres niveles (ver [7]) donde los grupos de abandono representan el tercer nivel de jerarquía. El método propuesto se compara, mediante un estudio de simulación, con el análisis de casos completos, el análisis de casos disponibles y un modelo multínivel con el tiempo de abandono. Los resultados demuestran que nuestra propuesta es más sensible en detectar efectos de tratamientos cuando la base de datos tiene simultáneamente abandonos MCAR, MAR y MNAR o cuando hay únicamente abandonos MNAR.

2. Bibliografía

- [1] Diggle, P.J. and Kenward, M.G. (1994). Informative Dropout in Longitudinal Data Analysis (with Discussion). *Applied Statistics*, 43, 49-93.
- [2] Little, R. J. A. (1995). Modeling the Drop-out Mechanism in Repeated-Measures Studies. *Journal of the American Statistical Association*, 90, 1112-1121.
- [3] Hogan, J.W. and Laird, N.M. (1997). Mixture Models for the Joint Distribution of Repeated Measures and Event Times. *Statistics in Medicine*, 16, 239-257.
- [4] Fermín, W., Galindo, P. and Javier, M. (2004). Multilevel Strategy for Detecting Behaviours of Dropout Cases in the Analysis of Longitudinal Data. Salamanca Statistics seminar V. Advances in Descriptive Multivariate Analysis.
- [5] Demirtas, H. and Schafer, J. L. (2003). On the Performance of Random-Coefficient Pattern-Mixture Models for Non-Ignorable Drop-out. *Statistics in Medicine*, 22, 2553-2575.
- [6] Daniels, M.J. and Hogan, J.W. (2000). Reparameterizing the Pattern Mixture Model for Sensitivity Analyses Under Informative Dropout. *Biometrics*, 56, 1241-1248.
- [7] Goldstein, H. (1995). Multilevel Statistical Model. Second Edition. London: Edward Arnold.

DELFOS: explotación de datos estadísticos vía web

*Sergio Alonso¹, Carlos Duque-Gómez², Manuel García-Román³,
Carlos Pérez-González¹*

¹{salonso, cpgonzal}@ull.es, Departamento de Estadística, Investigación Operativa y Computación, Universidad de La Laguna

²cduqgom@gobiernodecanarias.org, Instituto Canario de Estadística

³mroman@ull.es, Departamento Matemática Fundamental, Universidad de La Laguna

Abstract

La cesión de micro-datos estadísticos a grupos de investigación de diversas universidades por parte de institutos estadísticos es una actividad necesaria en el desarrollo de proyectos de colaboración, en la justificación de nuevas metodologías y resultados en trabajos de investigación, y, en definitiva, para incrementar la rentabilidad del esfuerzo público en las operaciones estadísticas. El proyecto DELFOS del Instituto Canario de Estadística, ISTAC, tiene como objetivo facilitar el acceso a estas fuentes de datos y su explotación estadística a través de un portal que se convierte es un paquete estadístico *on-line*. Es un primer esfuerzo para incluir la actividad estadística en la Web 2.0.

Palabras Clave: Bancos de datos estadísticos, Web 2.0

AMS: 68U35, 62-07

1. Introducción

En el año 2003, el ISTAC firmó sendos convenios de colaboración con las dos universidades canarias en los que, entre otros objetivos, se proponía la cesión de fuentes de datos estadísticos de interés para los investigadores canarios a través de un centro documental. La implementación de dicho centro es el origen del proyecto DELFOS.

El desarrollo de iniciativas web, donde se traslada al portal el conjunto de servicios dados al usuario, como pueden ser la gestión de correo electrónico, plataformas de contenidos educativos, o, próximamente, procesadores de texto, u hojas de cálculo, empiezan a configurar lo que se ha venido en llamar la web 2.0. Enmarcamos este proyecto bajo esta misma filosofía.

2. El proyecto DELFOS

Un servidor web dinámico será la puerta de acceso y gestión del banco de datos. Partirá de una versión inicial, a la que se irán añadiendo mejoras en la gestión de usuarios y datos, así como la posibilidad de tratar estos últimos con nuevas

y específicas metodologías estadísticas. Este objetivo convertirá al servidor web del banco de datos en un entorno completo de un paquete estadístico.

Son dos las líneas maestras en torno a las que gira el desarrollo del proyecto. La primera es la administración, entendiendo por tal, la acciones que relacionan el sistema con los usuarios. Para lograr cuenta de acceso, se tramitará una solicitud en la que se justifique la necesidad de uso de los datos, se compruebe la personalidad del peticionario y se le comprometa al cumplimiento de las condiciones establecidas para cada grupo de usuarios. El sistema hará distinción de los usuarios dotándolos de perfiles diferenciados que les permitirán acceder, mediante la gestión de privilegios, a distintos conjuntos de servicios y acceso a diferentes fuentes de datos. Los requerimientos de seguridad podrán incluir además, por ejemplo, la comprobación de pertenencia a rangos de números IP del ordenador solicitante, o claves de acceso mediante certificado digital u otras medidas seguras.

Una vez que el usuario esté dado de alta, y acceda al sistema, éste pondrá a su disposición sus servicios: herramientas de manipulación controlada de datos y de desarrollo y cálculo estadístico. En el primer caso, cada fuente de información disponible según el perfil del usuario y convenientemente anonimizada, informará de su metodología de origen, listado de cada variable, tipificación y comentarios acerca de su rango de valores. Para el desarrollo e implementación de los servicios estadísticos, usaremos el sistema, GNU R, un derivado del S-PLUS pero de desarrollo y uso libre. Estará integrado en el servidor web, analizará las demandas de los usuarios y dará respuesta vía páginas dinámicas a sus consultas.

Estimación de errores cuadráticos medios de estimadores de áreas pequeñas en la encuesta de población activa de Canarias

*Jorge Saralegui¹, Montserrat Herrador¹, Domingo Morales², M^a
Dolores Esteban², Ángel Sánchez², Laureano Santamaría²,
Yolanda Marhuenda², Agustín Pérez²*

¹herrador@ine.es, Instituto Nacional de Estadística.

²d.morales@umh.es, Centro de Investigación Operativa, Universidad Miguel Hernández de Elche.

Abstract

En este trabajo se describen tres procedimientos para la estimación de errores cuadráticos medios de estimadores de áreas pequeñas en la encuesta de población activa y se aplican a datos de Canarias en el segundo trimestre de 2003. El primer método usa fórmulas directas, el segundo remuestreo bootstrap bietápico y el tercero muestras jackknife. Finalmente se analizan los resultados y se dan recomendaciones.

Palabras Clave: Encuesta de población activa, estimación en áreas pequeñas, error cuadrático medio.

AMS: 62E30, 62J12

1. Introducción

Las técnicas de estimación en áreas pequeñas se han desarrollado típicamente para ser utilizadas bajo muestreo aleatorio simple y no tienen en cuenta los problemas derivados de la falta de respuesta o de la necesaria consistencia de los datos publicados por las oficinas de Estadísticas en los distintos niveles de agregación territorial. En este trabajo se presentan tres procedimientos, adaptados a la Encuesta de Población Activa (EPA), para la estimación de errores cuadráticos medios de estimadores basados o asistidos por modelos. Los procedimientos se aplican a estimadores de totales de parados, ocupados e inactivos y a estimadores de tasas de paro para ambos sexos por comarcas en la comunidad autónoma de Canarias.

El trabajo comienza introduciendo la notación empleada y describiendo algunos detalles técnicos de la EPA, así como del estimador de razón separado aplicado a la estimación de totales. Se plantea el principio de consistencia de estimaciones de totales entre distintos niveles de agregación territorial. Se ilustra además el procedimiento implementado para conseguir que la suma en la provincia de los totales estimados de sus comarcas, por cualquier procedimiento de estimación, coincida con el total que el estimador EPA da para la provincia.

En un segundo apartado se introducen los estimadores de áreas pequeñas utilizados en las aplicaciones así como los estimadores directos de sus errores cuadráticos medios (véase [1] y [2]). Se dan algunas características técnicas de las aplicaciones realizadas tales como el universo investigado, el ajuste de modelos, el dominio o área pequeña de interés, las variables auxiliares agregadas y no agregadas y los modelos utilizados en el cálculo de los distintos estimadores de parados y ocupados. Se analizan los problemas relativos a la selección de variables explicativas.

En un tercer apartado se describen los procedimientos de remuestreo bootstrap bietápico y jackknife para la estimación de varianzas (véase [3]). El método bootstrap genera muestras bootstrap (remuestras) a partir de la muestra EPA original. El bootstrap requiere evaluar los estimadores en las remuestras de la misma forma que se hizo en la muestra original. Aquí aparecen los problemas de calibración, resueltos de forma simplificada (sin usar el software CALMAR) en este trabajo. El método jackknife genera remuestras quitando de la muestra original una sección censal cada vez y recalcando los pesos calibrados originales (factores de elevación de la EPA).

De la observación de los resultados numéricos se pueden extraer las siguientes conclusiones: (1) las estimaciones de errores cuadráticos medios (ECMs) obtenidas para los tres métodos tienden a coincidir en aquellos dominios con mayor tamaño muestral, (2) el método bootstrap proporciona en general estimaciones más elevadas de ECMs, lo que hace sospechar que produce sesgos por exceso, (3) el método directo es fuertemente criticable al estar basado en suposiciones que no se verifican en la práctica, 4) el método jackknife evita los problemas teóricos del método directo y la dificultad de implementación del método bootstrap. En consecuencia, éste método es un candidato de consenso para la estimación de ECMs, siempre y cuando no se utilice en dominios con pocas secciones censales.

2. Agradecimientos

Los autores agradecen el apoyo recibido de los miembros de la Subdirección de Metodología y Técnicas Estadísticas del Instituto Nacional de Estadística.

3. Bibliografía

- [1] Rao, J.N.K. (2003). *Small area estimation*. John Wiley.
- [2] Särndal, C.-E., Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling*, Springer-Verlag, New York.
- [3] Shao J. y Tu D. (1995). *The Jackknife and Bootstrap*. Springer-Verlag, New York.

Estimación en áreas pequeñas en la Encuesta Industrial de la C.A. de Euskadi

Josu Iradi¹, Haritz Olaeta², Cristina Prado³

¹Josu.Iradi@eustat.es, Director General, Eustat

²H_Olaeta@eustat.es, Metodología, Eustat

³cristina_prado@eustat.es, Subdirectora, Eustat

Abstract

Eustat ha publicado por primera vez en el 2005 estimaciones comarcas de la Encuesta Industrial de los años 2002 y 2003 obtenidas mediante la aplicación de modelos de áreas pequeñas, siendo también la primera vez que se utiliza esta metodología para difundir datos en la estadística oficial de la C.A. de Euskadi. En este trabajo se resumen los modelos utilizados y se presentan las estimaciones publicadas.

Palabras Clave: Áreas Pequeñas, Encuesta industrial

1. Introducción

Eustat, consciente de la creciente demanda de estimaciones de estadísticas de calidad cada vez más desagregadas, ha publicado por vez primera estimaciones basadas en modelos de áreas pequeñas de algunas de las variables más relevantes de la Encuesta Industrial para las veinte comarcas de la C.A. de Euskadi correspondientes a los años 2002 y 2003.

2. Estimadores de áreas pequeñas

Los modelos de áreas pequeñas suponen la existencia de un modelo subyacente que siguen todos los datos de la población, pero que se estima con los datos de la muestra (Rao, 2003). Eustat utiliza para la obtención de estimaciones comarcas en la Encuesta Industrial dos tipos de modelos: el modelo lineal de efectos fijos y el modelo de regresión lineal con efectos fijos y aleatorios, llamado también modelo mixto.

En el modelo mixto el predictor consta de un término común de efectos fijos y otro diferenciado para los elementos de cada comarca d ($d = 1, \dots, t$). Este término diferenciado está formado por los efectos aleatorios (v_d), de modo que todos los datos de la misma comarca comparten el mismo efecto aleatorio. En el caso del modelo de efectos fijos no existen términos diferenciados para cada comarca, ya que la parte sistemática ($X\beta$) es común para todas las comarcas. Sin embargo, la especificidad se consigue al proyectar el coeficiente común (β) a la información auxiliar específica (X_d) de cada comarca.

3. Publicación de estimaciones

Las macromagnitudes de la Encuesta Industrial de la C.A. de Euskadi correspondientes a los años 2002 y 2003 cuyas estimaciones comarcales se han publicado han sido el valor añadido a coste de factores, las ventas netas, los costes de personal, el excedente bruto de explotación, la inversión y el resultado antes de impuestos. Además, se han publicado también las estimaciones del personal por comarcas dado que ha sido ésta la variable exógena utilizada en los modelos de áreas pequeñas. Todas las estimaciones de totales se han acompañado de sus respectivos coeficientes de variación (del error cuadrático medio) estimados.

4. Conclusiones

Debido a la calidad de los resultados obtenidos, Eustat ofrecerá a partir de ahora estimaciones comarcales de las principales macromagnitudes de la Encuesta Industrial de la C.A. de Euskadi anualmente. La metodología descrita en este trabajo se aplicará, además, a otras encuestas económicas del Instituto. Además, se está trabajando en la actualidad en la utilización de modelos de áreas pequeñas para encuestas socio-demográficas tomando como punto de referencia la encuesta de la Población en relación a la actividad (PRA).

5. Bibliografía

- [1] Rao, J.N.K. (2003). Small Area Estimation. Wiley Series in Survey Methodology.

Estimadores en Áreas Pequeñas aplicados a la Estadística Pública Canaria

Alberto González-Yanes¹, Miguel A. González-Sierra², Enrique González-Dávila², Roberto Dorta-Guerra², Sergio Alonso², Domingo Morales³

¹jesusalberto.gonzalezyanes@gobiernodecanarias.org, Instituto Canario de Estadística

²magsierr@ull.es, egonzale@ull.es, rodorta@ull.es, salonso@ull.es, Departamento de Estadística, Investigación Operativa y Computación, Universidad de La Laguna

²d.morales@umh.es, Centro de Investigación Operativa, Universidad Miguel Hernández de Elche

Abstract

La demanda de estadísticas oficiales a nivel de islas y/o comarcas (Zonas Estadísticas Intermedias), cuando las encuestas de referencia se prevén para dominios superiores, motiva el desarrollo de diversas líneas de investigación. En este trabajo se exponen los principales resultados del proyecto CANAREA 2005 y líneas futuras de investigación.

Claves: Encuesta de Población Activa (EPA); Estimación en áreas pequeñas; Error cuadrático medio.

AMS: 62J07; 62J10; 62H12

1. Introducción

La demanda de estadísticas oficiales a nivel de dominios pequeños se ha consolidado estos últimos años entre los objetivos prioritarios de los sistemas de estadística oficial (Herrador y Saralegui (2003)). El Instituto Canario de Estadística (ISTAC) se incorporó al grupo de trabajo de estimadores de áreas pequeñas del Instituto Nacional de Estadística (INE) desde sus inicios en abril de 2004, en el que participan también otros institutos autonómicos. Durante el año 2005, el ISTAC con la colaboración de un grupo de investigadores de la Universidad de La Laguna y Miguel Hernández de Elche firman un proyecto titulado "Estimadores en Áreas Pequeñas aplicados a la Estadística Pública Canaria, CANAREA 2005" con la intención de valorar y aportar procedimientos que resuelvan la estimación en áreas pequeñas de las variables más significativas de la Encuesta de Población Activa (EPA), así como los requerimientos del grupo de trabajo del INE.

La definición de dominio o área pequeña debe ir ligada a la encuesta de referencia. Para la EPA, la cual publica datos por estimación directa para

provincias, se consideró como áreas de interés las islas y los futuros niveles NUT IV (Comarcas). En el caso de Canarias, estas comarcas están constituidas por la unión de municipios con un total de 27 comarcas, 12 en la provincia de Las Palmas y 15 en la provincia de S/C de Tenerife.

2. CANAREA 2005

El objetivo principal del proyecto CANAREA 2005 fue la valoración de diferentes estimadores basados en el diseño de la encuesta de referencia, la EPA, al incluir información suministrada por la toma de muestra suplementaria a nivel de isla o comarca. Para ello, y usando como fuente primaria el Censo 2001 y simulación de MonteCarlo, se implementó la metodología del diseño muestral de la EPA. La muestra suplementaria se obtuvo al tomar muestreo uniforme por islas o comarcas con selección de las viviendas por muestreo aleatorio simple. Para la valoración del estudio se contemplan estadísticas descriptivas, errores cuadráticos medios relativos y sesgos relativos de diferentes estimadores. Los estimadores estudiados han sido el estimador directo, el posestratificado según grupo de edad (estimador GREG con variable auxiliar edad), el estimador sintético basado en la provincia y cuatro estimadores compuestos dependientes del diseño formados como combinación del sintético y posestratificado (ver Rao (2003)).

Como resultado de este proyecto, se constata que el estimador sintético basado en la provincia, aunque es con diferencia el que tiene menor coeficiente de variación, presenta los mayores sesgos, agravándose incluso esto para comarcas relativamente grandes dentro de las islas (o islas grandes no capitalinas frente a la provincia). Este sesgo se reduce con la utilización de los estimadores compuestos, a costa de aumentar la variabilidad. Aún así, en la mayoría de las comarcas el error cuadrático medio relativo para el sintético sigue siendo de los más pequeños, incluso con la suplementación de la muestra duplicada.

Las líneas futuras de investigación se centran en la valoración de estimadores sintéticos basados en regiones con características más homogéneas que la provincia, y que permitan una reducción del error cuadrático medio debida principalmente a la disminución de los sesgos. También se probarán estimadores basados en modelos siguiendo las directrices del proyecto europeo EURAREA donde ha participado el INE.

3. Bibliografía

- [1] Herrador, M. y Saralegui, J. (2003). El problema de la estimación en áreas pequeñas para la estadística oficial. Recientes progresos en España. *27 Congreso Nacional de E. e I.O., Lleida*, 103-121.
- [2] Rao, J.N.K. (2003). *Small Area Estimation*. John Wiley & Sons, New Jersey.

Estimation of Unemployment Rates in Small Areas by a Multinomial Logit Mixed Model

Isabel Molina¹, Ayoub Saer², María J. Lombardía³

¹isabel.molina@uc3m.es, Departamento de Estadística,
Universidad Carlos III de Madrid

²axs96@socsci.soton.ac.uk, Department of Social Statistics,
University of Southampton

³mjoselc@usc.es, Departamento de Estadística e Investigación Operativa,
Universidade de Santiago de Compostela

Abstract

Se obtienen estimadores de tasas de paro en áreas pequeñas a través de un modelo logístico multinomial con efectos aleatorios de las áreas. Además, se proponen tres aproximaciones del error cuadrático medio, una analítica y dos basadas en bootstrap. Se aplican los resultados a datos de la EPA de Gran Bretaña y se realiza un estudio de simulación.

Keywords: Multinomial logit mixed model, Small area estimation.

AMS: AMS Classification: 62J12, 62H12.

1. Introduction

There is an increasing demand from both private and official institutions on information related to small geographical areas. For instance, there is special interest of the Office for National Statistics in UK (ONS) and its equivalent in Spain (INE) in obtaining estimates of unemployment rates and annual family income in certain established small areas, where direct estimates lack of precision.

In this work, the variable of interest takes one of the values “unemployed”, “employed” and “not in labor force”. Assuming that this variable follows a multinomial logit model with area-specific random effects, estimators of small area unemployment rates are constructed and three methods for approximating the mean squared errors are given.

2. Description of data and proposed model

The available data file contains labor force data of small areas in UK from year 2000 aggregated by sex-age groups. An incidence matrix is created from data. Let $\mathbf{x}_{di} = (x_{di1}, \dots, x_{di23})$ be the row of the incidence matrix corresponding to area d and sex-age group i , by u_d the random effect of area d , by y_{di1} , y_{di2} and y_{di3} respectively the number of unemployed, employed and inactive in the sample within sex-age group i and area d , and by p_{di1} , p_{di2} and

$p_{di3} = 1 - p_{di1} - p_{di2}$ the corresponding probabilities of unemployed, employed and inactive. Let $m_{di} = \sum_{j=1}^3 y_{dij}$ be the sample size. The vectors (y_{di1}, y_{di2}) given u_d are assumed to be independent with probability mass function

$$f(y_{di1}, y_{di2}|u_d) = \frac{m_{di}!}{y_{di1}! y_{di2}! y_{di3}!} p_{di1}^{y_{di1}} p_{di2}^{y_{di2}} p_{di3}^{y_{di3}}. \quad (3.57)$$

Further, the following model is assumed for the the probabilities (p_{di1}, p_{di2}) ,

$$\begin{aligned} \log p_{dij}/p_{di3} &= \mathbf{x}_{di}\beta_j + u_d, \quad j = 1, 2, \quad i = 1, \dots, 6, \\ u_d &\stackrel{iid}{\sim} N(0, \varphi), \quad d = 1, \dots, 406, \end{aligned} \quad (3.58)$$

where $\beta_j = (\beta_{1j}, \dots, \beta_{23j})^t$ denotes the vector of coefficients of the explanatory variables for the multinomial category j , $j = 1, 2$.

3. Parameters, estimators and errors

The first goal of this work is to estimate the totals of unemployed δ_{d1} , employed δ_{d2} , and the unemployment rate ur_d , for each area $d = 1, \dots, 406$. Estimators $\hat{\delta}_{dj}$, $j = 1, 2$ and \hat{ur}_d are obtained by fitting the model.

The second goal is to estimate the mean squared error of the estimators \hat{ur}_d . For this, one approach is to use an adaptation of the well-known Prasad-Rao formula to estimate the mean squared errors of $\hat{\delta}_{dj}$, $j = 1, 2$ and the mean crossed product error of $\hat{\delta}_{d1}$ and $\hat{\delta}_{d2}$. From them, by the method of linearization of Taylor, an approximation of the mean squared error of \hat{ur}_d can be obtained. Another approach is to use a resampling procedure. Here we propose two bootstrap estimators of the MSE of the unemployment rates \hat{ur}_d which avoid Taylor linearizations. The simulations carried out show the good performance of the bootstrap approximations in comparison with the analytical one.

4. Acknowledgement

This work has been supported by grants BMF2003-04820, 06/HSE/0181 / 2004, PGIDIT03PXIC20702PN and MTM2005-00820.

Evaluación empírica de los estimadores de áreas pequeñas en la estimación del desempleo

Antonio Vaamonde Liste¹, Esther López Vizcaíno²

¹vaamonde@uvigo.es, Departamento de Estadística e Investigación Operativa,
Universidad de Vigo

²esther.lopez.vizcaino@xunta.es, Instituto Galego de Estatística

Abstract

Los estimadores en áreas pequeñas se utilizan, a menudo, para estimar variables en áreas geográficas que no fueron tenidas en cuenta en el diseño muestral de una encuesta. En este trabajo se pretende evaluar la precisión de una serie de estimadores en áreas pequeñas utilizados en la literatura.

Palabras Clave: áreas pequeñas, modelos lineales mixtos, desempleo

1. Introducción

Las encuestas por muestreo probabilístico están diseñadas para obtener estimaciones fiables de ciertas variables de investigación con un mínimo de precisión prefijada en agregados poblacionales definidos a priori. Pero a menudo es frecuente que se demanden datos para agregados poblacionales de magnitud inferior. El problema que se plantea, entonces, es el de la insuficiencia de tamaño muestral, e incluso ausencia total en algunos casos, para llevar a cabo tales estimaciones. Ante esta situación se puede ampliar el tamaño muestral, lo que no es siempre aconsejable e incluso algunas veces inviable desde el punto de vista económico, o utilizar estimadores especialmente concebidos para dar estimaciones en áreas pequeñas, con ayuda de información auxiliar ajena a la encuesta.

Para realizar las estimaciones en un área pequeña es necesario que otras áreas relacionadas con ella le presten la información para construir así estimaciones indirectas que incrementen el tamaño efectivo de la muestra e incrementar, por tanto, la precisión. Tales estimadores indirectos están basados en modelos implícitos o explícitos que relacionan las áreas pequeñas a través de información auxiliar. La información auxiliar utilizada para dar estimaciones en las áreas pequeñas puede proceder de la propia encuesta, de otras encuestas, de censos o de registros administrativos. Los estimadores indirectos basados en modelos explícitos incluyen el sintético, mientras que aquellos basados en modelos explícitos incorporando efectos específicos de área incluyen al mejor predictor empírico lineal insesgado (EBLUP) y a los estimadores bayesianos.

En este trabajo se va a realizar una evaluación empírica de la precisión de la estimación del desempleo en 50 unidades geográficas artificiales creadas a partir

de los datos que la Encuesta de Población Activa ofrece para toda España, utilizando para ello estimadores en áreas pequeñas.

2. Métodos

A partir de los microdatos de la EPA se seleccionaron un total de 200 muestras diferentes de viviendas, utilizando muestreo aleatorio simple estratificado.

Los estimadores que se emplearán serán el estimador directo, sintético, estimadores basados en modelos lineales de efectos fijos, estimadores EBLUP basados en modelos lineales mixtos y estimadores bayesianos basados en modelos a nivel de unidad.

Las medidas para valorar el comportamiento de los estimadores serán el sesgo relativo medio (ARB) y la raíz cuadrada del error cuadrático medio (RMSE) asociados al área pequeña d :

$$\text{ARB}_d = \frac{1}{K} \sum_{k_1}^K \left(\frac{\hat{Y}_d(k)}{Y_d} - 1 \right) 100$$

$$\text{RMSE}_d = \frac{100}{Y_d} \sqrt{\frac{1}{K} \sum_{k=1}^K (Y_d(k) - Y_d)^2}$$

donde $\hat{Y}_d(k)$ es la estimación del total de parados en el área pequeña d en la k -ésima selección de la muestra, siendo K el número de muestras seleccionadas e Y_d es el número de parados totales en el área d

3. Bibliografía

- [1] Ghosh, M and Rao, J.N.K (1994). Small Area Estimation: An Appraisal. *Statistical Science*, vol 9, nº 1, 55-93
- [2] Rao, J.N.K. (2003). *Small Area Estimation*. USA: Wiley Series in Survey Methodology

Fay–Herriot Models with Nonparametric Estimation of Heteroscedasticity

M.J. Lombardía¹, W. González-Manteiga¹, I. Molina², D. Morales³, L. Santamaría³

¹Dpto. de Estadística e Inv. Oper., Universidad de Santiago de Compostela

²Dpto. de Estadística, Universidad Carlos III de Madrid

³Centro de Investigación Operativa, Universidad Miguel Hernández de Elche

Abstract

Los modelos tipo Fay-Herriot (1979) son frecuentes en estimación de áreas pequeñas. Debido a la falta de datos desagregados se asumen que las varianzas de los errores del modelo son conocidas. En este trabajo se proponen estimadores utilizando inferencia no paramétrica, y se estudian sus propiedades asintóticas. Se obtienen estimadores de áreas pequeñas, y sus errores cuadráticos medios se aproximan mediante bootstrap.

Keywords: Bootstrap; kernel-based estimator; linear mixed model.

AMS: 62D05, 62J05.

1. Introduction

In small area estimation, estimators of small areas (geographical regions or domains) of a population are required, but the available data was typically obtained through a survey that was planned to provide reliable direct estimators for larger areas of the population. As direct estimator we consider any estimator calculated just using the sample data belonging to the target area. In this situation, because of the lack of observations in the target area, direct estimators have great sampling errors. A way of increasing the precision of direct estimators is by the usage of models that link all the sample observations through the inclusion of auxiliary variables.

In the context of estimation of per-capita income for small places in U.S., Fay and Herriot (1969) assumed that the mean per-capita income of small place d , μ_d , is linearly related to the values of p auxiliary variables $\mathbf{x}_d = (x_{d1}, \dots, x_{dp})'$ plus an error term u_d , that is,

$$\mu_d = \mathbf{x}_d\beta + u_d, \quad \text{where } u_d \stackrel{iid}{\sim} N(0, \sigma_u^2), \quad d = 1, \dots, D, \quad (3.59)$$

where the variance σ_u^2 is an unknown parameter. They further considered that, given u_d , the direct estimators y_d of mean per-capita income are realizations of independent variables with normal distribution centered in μ_d and with variances w_d , that is,

$$y_d | u_d \stackrel{ind}{\sim} N(\mu_d, w_d), \quad d = 1, \dots, D, \quad (3.60)$$

where the variances w_d were assumed to be known. However, in their application, they found that the standard deviation $w_d^{1/2}$ increased with the expected value. When the conditional variances w_d are function either of the design points \mathbf{x}_d or of the mean response $\mathbf{x}_d\beta$, we speak about heteroscedasticity.

Carrol (1982) already proposed a kernel-based estimator of the error variance for linear regression models. Here we extend the results of Carroll (1982) to a twofold extension of the Fay-Herriot model (3.59)–(3.60). On the one hand, the normality assumption is eliminated from the model random effects and errors and on the other hand, heteroscedasticity is introduced by considering that the error variance is an unknown function of the design x_d ($p = 1$) or of the mean response $\mathbf{x}_d\beta$ ($p > 1$). We apply the obtained results to the estimation of small-area parameters, studying their asymptotic behaviour and assessing in a simulation study the gain in precision of the derived small-area estimators with respect to the estimator that ignores the heteroscedasticity. Further, we compare a naive approximation of the mean squared error of the small-area estimator obtained through direct application of the results of Prasad and Rao (1990) to the extended model with a bootstrap-based approximation.

2. Acknowledgement

This work has been supported by projects BMF2003-04820, MTM2005-00820, PGIDIT03PXIC20702PN and MTM2005-00820.

3. Bibliography

- [1] Carroll, R.J. (1982). Adapting for heteroscedasticity in linear models *The Annals of Statistics* 10, 1224–1233.
- [2] Fay, R. E. and Herriot, R. A. (1979). Estimation of Income from Small Places: An Application of James-Stein Procedures to Census Data. *Journal of the American Statistical Association* 74, 269–277.

La Encuesta Continua de Presupuestos Familiares 2006 (ECPF 2006)

Carmen Ureña Ureña¹, Encarnación Martín Bernia²

¹curena@ine.es, Jefe de Área de EPF y Condiciones de Vida del INE

²enmartin@ine.es, Consejero Técnico de EPF del INE

Abstract

En este trabajo se describe las principales características técnicas de la Encuesta Continua de Presupuestos Familiares 2006 elaborada por el Instituto Nacional de Estadística.

Palabras Clave: Diseño y metodología de encuestas, Muestreo

AMS: 62D05, 62Q05

1. Introducción

Las EPF ofrecen información sobre los gastos de consumo de los hogares privados, siendo unas de las operaciones estadísticas con mayor tradición en el INE.

Desde su implantación en 1958, en el INE se han realizado dos tipos de EPF: Cada ocho o diez años, las estructurales o básicas, consideradas como encuestas de objetivos múltiples y que se han utilizado para obtener estimaciones del nivel y la estructura del consumo anual de los hogares así como sobre la estructura de ponderaciones del gasto para los cambios de base del IPC. Por otro lado, se han realizado las encuestas continuas o de tipo coyuntural, siendo su principal objetivo ofrecer estimaciones del gasto trimestral de los hogares, así como del cambio interanual de dicho gasto.

Las EPF son las operaciones más complejas que existen en cualquier sistema estadístico, metodológicamente y por la dificultad que suponen para los entrevistadores y fundamentalmente para las familias informantes. Todo esto da lugar a que se obtengan unos errores ajenos al muestreo elevados en comparación con los de otras encuestas dirigidas a hogares.

A lo largo del tiempo, los dos tipos de encuestas han sido objeto de reformas metodológicas referidas a: tipo de población considerada, tamaño muestral, variables objeto de investigación, nivel de desagregación de los gastos, periodicidad, sistemas de recogida de la información, etc.

La última reforma se implantó en el año 1997, dando lugar a la ECPF 1997, de periodicidad trimestral y con la que se unificaron los dos tipos de encuestas que hasta ese momento se venían realizando.

El INE ha implantado en enero 2006 una nueva EPF que sustituye a la que estaba en vigor desde 1997.

Los principales planteamientos que han impulsado este cambio son: necesidad de conseguir una mayor simplicidad en el diseño metodológico de este tipo de operación, reducción de la carga de trabajo de los hogares, satisfacción de nuevas exigencias por parte de los diferentes usuarios y de organismos internacionales, principalmente Eurostat.

2. Principales características

Las principales características y elementos diferenciadores de esta nueva encuesta respecto a encuestas anteriores son los siguientes:

Periodo de estudio y tipo de encuesta: encuesta anual de carácter continuo, en forma de panel rotante, renovándose cada año la mitad de los hogares de la muestra.

Tamaño de la muestra: aproximadamente 24.000 hogares al año (similar al tamaño de las básicas). En la anterior encuesta se entrevistaban aproximadamente a 8.000 hogares cada trimestre.

Colaboración de los hogares: cada hogar de la muestra colabora a lo largo 14 días al año durante dos años consecutivos (en todas las encuestas anteriores la colaboración eran siete días).

Desagregación geográfica y funcional: se obtendrán estimaciones del gasto anual de los hogares a nivel de cinco dígitos de la clasificación de gastos para el conjunto nacional y a cuatro dígitos para las comunidades autónomas.

Periodos de referencia: se amplía el número de períodos de observación de los gastos a cinco (bisemanal, mensual, trimestral, anual y último recibo).

En definitiva, se pretende que la nueva EPF sea un potente instrumento para el análisis de la estructura de las pautas de consumo de los hogares residentes en España, además de un elemento básico para su utilización como fuente principal en estadísticas como la Contabilidad Nacional o el Índice de Precios de Consumo (IPC).

3. Bibliografía

- [1] Eurostat (2003). Household Budget Surveys in the EU: Methodology and Recommendations for harmonization.

Posprocesado sobre microagregación multivariante: un enfoque con grupos de cardinalidad variable

Josep Maria Mateo Sanz¹, Agustí Solanas², Josep Domingo Ferrer²

¹josepmaria.mateo@urv.net, Área de Estadística e Investigación Operativa,
Universitat Rovira i Virgili

²agusti.solanas, josep.domingo@urv.net, Dept. de Ingeniería Informática y
Matemáticas, Universitat Rovira i Virgili

Abstract

Se propone un método de posprocesado sobre la microagregación de un conjunto multivariante de microdatos con la particularidad de que los grupos resultantes pueden tener cardinalidades diferentes según la suma de cuadrados entre grupos que provoque la agrupación de los registros. El método se basa en la reasignación de los registros que provocan una suma de cuadrados alta.

Palabras Clave: Microagregación, confidencialidad

1. Introducción

Cuando se trabaja en Estadística Oficial, suele plantearse el problema de la protección de datos (confidencialidad) a nivel individual (microdatos). Para proteger los microdatos existen dos grandes grupos de métodos de control de la revelación estadística: los métodos que suprimen datos y los métodos perturbativos. En este trabajo nos vamos a centrar en la mejora de la microagregación, que es un método perturbativo, mediante un método de posprocesado que nos permita obtener grupos de registros más homogéneos que los obtenidos por la aplicación directa de cualquier método de microagregación.

2. Microagregación: antecedentes

Dado un conjunto de microdatos con variables numéricas, la microagregación consiste en agrupar registros de manera que en cada grupo resultante haya un número, como mínimo, k de registros y que estos grupos sean lo más homogéneos posible. Una vez construidos los grupos, se publican los valores medios de cada grupo. El valor de k se escoge para proveer protección sobre los individuos. La homogeneidad de los grupos se mide mediante de la suma de cuadrados dentro de los grupos, SCDG; cuanto menor sea este valor más homogéneos serán los grupos que provocan esa SCDG.

Los métodos de microagregación existentes están basados en la distancia existente entre los registros, ya sea a través de alguna proyección de los datos [1] o calculando directamente distancias entre los registros multivariantes [2] y, la mayoría de ellos, forma grupos de tamaño fijo k de manera que, a veces, la

agrupación obtenida no refleja la estructura de los datos (puede haber grupos poco homogéneos).

3. Método de posprocesado de la microagregación.

El método que se propone parte de los grupos resultantes obtenidos al aplicar cualquiera de los métodos existentes de microagregación. El resultado que se obtendrá será una nueva agrupación de registros con una SCDG menor o igual que la obtenida en primera instancia y manteniendo la condición de que en cada grupo haya un mínimo de k registros; por lo tanto es un método de microagregación que se obtiene posprocesando una microagregación original.

Cuando tenemos un conjunto de microdatos y aplicamos microagregación, se obtiene una agrupación de registros en la que cada grupo tiene su SCDG concreta. En el método que proponemos, partimos de dicha agrupación, ordenamos los grupos según su SCDG y escogemos los grupos que tienen una SCDG más alta. Para cada grupo escogido, deshacemos dicho grupo y reasignamos cada registro que formaba ese grupo al grupo más cercano. La distancia de un registro a un grupo la mediremos según el incremento en la SCDG que implique la inclusión de dicho registro en ese grupo. Si, una vez realizada la reasignación, la SCDG total ha disminuido respecto de la que teníamos antes de la reasignación, nos quedamos con la nueva agrupación obtenida; en caso que la SCDG total no haya disminuido, dejamos los grupos tal como estaban antes de la reasignación. Este proceso lo podemos aplicar a un porcentaje de grupos (los que tengan una SCDG más alta) o a todos los grupos siguiendo el orden decreciente de su SCDG.

Al final del proceso, obtendremos una agrupación de registros que tendrá una SCDG menor o igual que la que teníamos al aplicar la microagregación inicial; será igual si no se ha deshecho ninguno de los grupos iniciales, es decir, si coincide el resultado de la microagregación inicial con el de la microagregación posprocesada. Otra característica es que la cardinalidad de cada grupo puede ser variable; sólo aseguramos que será mayor o igual que k , con lo que los grupos resultantes se pueden adaptar a la estructura de los datos iniciales.

4. Agradecimientos

El trabajo ha sido subvencionado por el proyecto SEG2004-04352-C04-01 “PROPRIETAS” y la ayuda 2005 SGR 00446 de la Generalitat de Catalunya.

5. Bibliografía

- [1] Anwar, N. (1993). *Micro-aggregation – The Small Aggregates Method*, internal report. Eurostat.
- [2] Mateo-Sanz, J.M., Solanas, A., Martínez-Ballesté, A., Domingo-Ferrer, J., Torres, A. *Microaggregation methods in large databases: a complexity analysis*, manuscrito.

Solución de sistemas de las ecuaciones simultáneas por métodos de altas prestaciones

José Juan López Espín¹, Domingo Jiménez Cánovas²

¹jlopez@umh.es, Departamento de Estadística, Matemáticas e Informática,
Universidad Miguel Hernández

²domingo@dif.um.es, Departamento de Informática y Sistemas, Universidad
de Murcia

Abstract

En este trabajo, se desarrollan métodos para encontrar la solución de modelos de ecuaciones simultáneas que exigen el uso de grandes recursos computacionales en sistemas de altas prestaciones. El trabajo consiste en el estudio y desarrollo de algoritmos que calculen los estimadores de Mínimos Cuadrados, Mínimos Cuadrados Indirectos y Mínimos Cuadrados en dos etapas.

Palabras Clave: Modelos de Ecuaciones Simultáneas, Econometría.

1. Introducción

En algunas aplicaciones de econometría es necesario resolver sistemas de ecuaciones simultáneas compuestas por miles de ecuaciones y variables.

Una posibilidad es resolver este tipo de sistemas con algún método que necesite muchos recursos computacionales. Por otro lado, también es posible resolver los sistemas usando métodos estadísticos cuyas estimaciones son sesgadas. Generalmente se usa la segunda opción por la falta de un software eficiente para computadoras actuales de altas prestaciones.

En el presente trabajo se desarrollan métodos capaces de resolver modelos de ecuaciones simultáneas de forma eficiente, es decir, desarrollando algoritmos capaces de aprovechar sistemas de altas prestaciones y trabajar así en paralelo para resolver el modelo en el menor tiempo posible. Esto hace que se puedan resolver sistemas de miles de ecuaciones y variables en muy poco tiempo, y por lo tanto, no sea necesario el uso de otros métodos de resolución, que aunque sesgados, eran considerados por la rapidez con la que concluían sus cálculos.

Se han desarrollado algoritmos paralelos en memoria compartida y en memoria distribuida. Para su implementación se han usado los lenguajes de programación de OpenMP (para compartida) y MPI (para distribuida) y las bibliotecas básicas de álgebra lineal (BLAS, ATLAS y LAPACK).

Finalmente se estudian y comparan tiempos de ejecución medidos en las pruebas experimentales hechas en distintas máquinas: Los algoritmos en OpenMP se han medido en HPC160 (32GFlops, 4 nodos con 4 microprocesadores por

nodo a 1Ghz). Y los algoritmos en MPI en un cluster de 10 biprocesadores Intel Xeon 2 con conexión SCI.

2. Bibliografía

- [1] "Econometría", Damodar Gujarati Ed: Mc Graw Hill. Gujarati, D. N. (1995), "Basic Econometrics", McGraw Hill
- [2] "Parallel Algorithms for Linear Models: Numerical Methods and Estimation Problems" Aut. Erricos John Kontoghiorghes. Ed: Kluwer Academic Publishers 2000
- [3] "Análisis Econométrico", William Greene. Ed: Prentice-Hall. "Econometric Analysis", third edition. Prentice Hall 1998.

TEIDE: Una nueva herramienta para depurar encuestas

Sergio Delgado Quintero¹, Juan José Salazar González², Fayna Álamo Santana³

{¹sdelquin,²jjsalaza}@ull.es, Departamento de Estadística, Investigación Operativa y Computación, Universidad de La Laguna

³fayna.alamosantana@gobiernodecanarias.org, Estadísticas Demográficas y Sociales, Instituto Canario de Estadística

Abstract

El problema de depurar una encuesta consiste en detectar errores dentro de un conjunto de datos y modificarlos para que el resultado sean datos coherentes. En términos estadísticos se habla del problema de *edición e imputación de datos*. Se presenta una herramienta para automatizar estas tareas aportando soluciones de buena calidad.

Claves: Software, Edición e Imputación.

AMS: 68N99, 68U99

1. Introducción

La depuración de encuestas es un problema que deben afrontar los institutos de estadística antes de poder llevar a cabo medidas sobre los datos recogidos. Durante mucho tiempo este problema se abordó de forma manual, lo que provocaba graves retrasos temporales y económicos en la explotación de la encuesta. Presentamos aquí una herramienta software para ayudar al técnico estadístico en la depuración de forma automática.

2. Características de TEIDE

TEIDE viene de “Técnicas de Edición e Imputación de Datos Estadísticos”. Se trata de una aplicación informática capaz de procesar datos, detectar los errores que existen en ellos y corregirlos mediante ciertos criterios de calidad. Estos criterios de calidad son principalmente dos. Por un lado minimizar el número de cambios que se efectúan sobre los datos de entrada, y por otro no modificar en exceso las distribuciones de frecuencia de las variables de la encuesta. Para cumplir con estas restricciones, la aplicación se basa en los estudios desarrollados por Fellegi y Holt (ver [1]) y en la metodología del registro donante.

La aplicación es capaz de trabajar tanto con variables discretas como con variables continuas. Además admite cualquier tipo de reglas de coherencia sobre un mismo registro. Partiendo del esquema “if (...) then (...)” se puede expresar cualquier condición necesaria. Los algoritmos utilizados para resolver

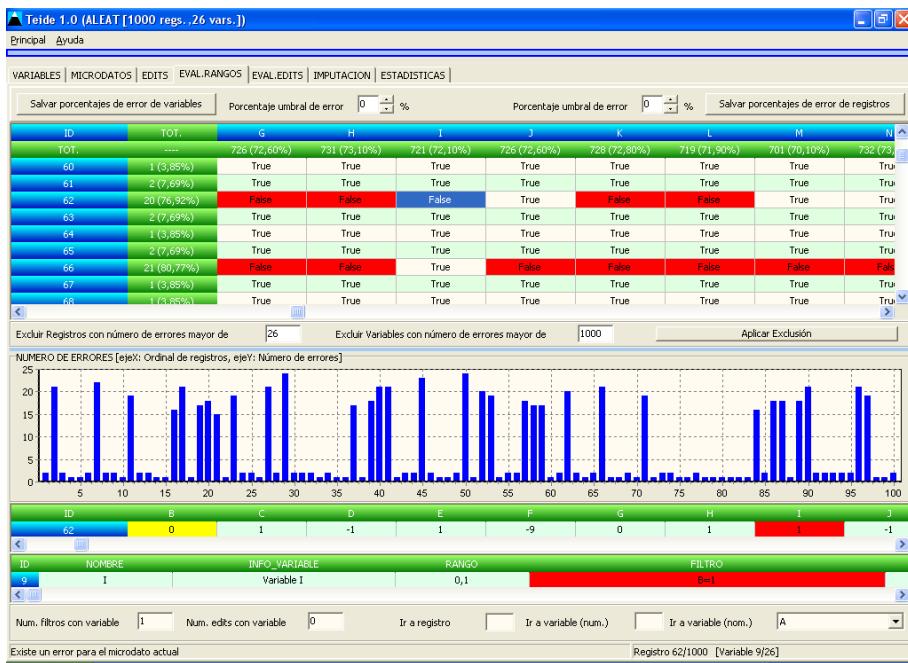


Figura 1: Captura de pantalla de TEIDE

los problemas combinatorios se basan en técnicas modernas de Optimización Matemática. La aplicación tiene un marcado enfoque visual con una potente interfaz de usuario (ver Figura 1). Ha sido desarrollada en *C++* y corre sobre *Windows*. Además de ser un software de edición e imputación de datos, lo podríamos ver como una herramienta de apoyo en el ciclo de vida de una encuesta, ya que proporciona una visión global de los datos a través de gráficas y resultados agregados. TEIDE ha sido y está siendo evaluado sobre encuestas reales.

3. Agradecimientos

En el desarrollo de TEIDE ha colaborado el Instituto Canario de Estadística (ISTAC).

4. Bibliografía

- [1] Felleji, I.P., Holt, D., (1976). A systematic approach to automatic edit and imputation. *Journal of the American Statistical Association* 71, 17-35.

Sección 4

Investigación Operativa

A conic IPM decomposition approach for large scale SDP

Kartik K. Sivaramakrishnan¹, Gema Plaza², Tamás Terlaky³

¹kksivara@ncsu.edu, Department of Mathematics,
North Carolina State University

²gema.plaza@ua.es, Departamento de Estadística e Investigación Operativa,
Universidad de Alicante

³terlaky@mcmaster.ca, Department of Computing & Software,
McMaster University

Abstract

Se describe un método para resolver problemas de programación semi-definida (SDP) de gran tamaño. El método se inspira en los métodos interiores y los métodos de planos de corte. También se discuten asuntos relacionados con una implementación eficiente del algoritmo y se presentan resultados numéricos para familias conocidas de problemas de SDP.

Keywords: Semidefinite Programming, Column and Matrix Generation, Interior Point Methods.

AMS: 90C22, 90C51, 49M27.

1. Introduction

Semidefinite programming (SDP) has been one of the most exciting and active research areas in optimization lately. This tremendous activity was spurred by the discovery of efficient interior point algorithms for solving SDP problems, and important applications of the SDP in developing approximation algorithms for various combinatorial optimization problems. However, although interior point methods (IPM's) are a great theoretical tool, and offer polynomial time complexity, they are fairly limited in the size of SDP problems they can handle. To overcome this shortcoming we present a conic interior point based decomposition methodology to solve a semidefinite program whose primal feasible region is bounded.

SDP is a generalization of linear programming (LP) where the nonnegative orthant is replaced by the cone of positive semidefinite matrices. Any SDP problem whose primal feasible set is bounded can be formulated as

$$\begin{array}{ll}
 (P) \quad \min & C \bullet X \\
 \text{s.t.} & \mathcal{A}(X) = b, \\
 & I \bullet X = 1, \\
 & X \succeq 0,
 \end{array}
 \quad
 \begin{array}{ll}
 (D) \quad \max & b^T y + z \\
 \text{s.t.} & \mathcal{A}^T y + zI + S = C, \\
 & S \succeq 0.
 \end{array}$$

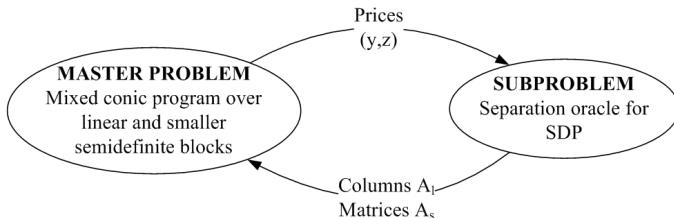
Here $X, S \in \mathcal{S}^n$, the space of real symmetric $n \times n$ matrices, $b \in \mathbb{R}^m$, $C \bullet X = \text{trace}(CX) = \sum_{i,j=1}^n C_{ij}X_{ij}$ is the Frobenius inner product of matrices in \mathcal{S}^n . The linear operator $\mathcal{A}: \mathcal{S}^n \rightarrow \mathbb{R}^m$ and its adjoint $\mathcal{A}^T: \mathbb{R}^m \rightarrow \mathcal{S}^n$ are:

$$\mathcal{A}(X) = [A_1 \bullet X, \dots, A_m \bullet X]^T \text{ and } \mathcal{A}^T y = \sum_{i=1}^m y_i A_i,$$

where the matrices $A_i \in \mathcal{S}^n$, $i = 1, \dots, m$, and $C \in \mathcal{S}^n$ are the given problem parameters. The constraints $X \succeq 0$, $S \succeq 0$ ($X \succ 0$, $S \succ 0$) are the only nonlinear constraints in the problem requiring that X and S be symmetric positive semidefinite (positive definite) matrices.

2. The approach

We describe in our paper an approach for solving large scale semidefinite programs. The approach solves a semidefinite program over one semidefinite cone in an iterative fashion between a master problem and a subproblem. Our master problem is a mixed conic program over linear and semidefinite cones and the subproblem is a simple separation oracle for the semidefinite cone. The following diagram shows the interaction between the two problems:



Depending on the multiplicity of the minimum eigenvalue of the dual slack matrix, the oracle returns either a column or a matrix in the primal master problem. The master problem is solved with a conic solver in a feasible primal-dual IPM setting. We also discuss several issues required for an efficient implementation of the approach, including a) a stabilized matrix generation approach for improving convergence, b) a technique to warm-start the master problem after matrix generation, and c) a procedure to detect and aggregate unimportant linear and semidefinite cones in the master problem. Our computational experiments indicate that our algorithm has the potential to outperform IPMs on large and dense semidefinite programs.

3. Bibliography

- [1] Sivaramakrishnan K., Plaza G. and Terlaky T. (2005). A conic IPM decomposition approach for large scale SDP, Submitted.
- [2] Dantzig G.B. and Wolfe P. (1960). The decomposition principle for linear programming, *Operations Research* 8, 101-111.

A family of cooperative games obtained by delegation

Rafael Amer¹, José Miguel Giménez²

¹rafel.amer@upc.edu, Department of Applied Mathematics II, Universitat Politècnica de Catalunya

²jose.miguel.gimenez@upc.edu, Department of Applied Mathematics III,
Universitat Politècnica de Catalunya

Abstract

Los semivalores forman una amplia familia de soluciones para los juegos cooperativos. Para cada juego y cada pareja de jugadores, se introduce una familia llamada de los juegos de delegación. Estos juegos permiten caracterizar todos aquellos semivalores cuyos coeficientes de ponderación son estrictamente positivos (semivalores regulares).

Keywords: Cooperative game, Semivalue, Characterization.

AMS: 91A12

1. Introduction

The Shapley value [3] and the Banzhaf value [2, 8] are two of the most important solution concepts for cooperative games. A generalization of both solutions is formed by the family of semivalues [4], allocation rules that assign to each player in a given game a weighted sum of the marginal contributions of the player in the game. The weight system depends only on the coalition sizes and all weighting coefficients are nonnegative. We speak of *regular semivalues* (Carreras and Freixas, 1999) [3] if all weighting coefficients are positive. In this case, all marginal contributions of each player matter when we consider the allocation by such a semivalue. The Shapley value and the Banzhaf value are regular semivalues.

Different authors have provided systems of properties in order to axiomatically characterize the Shapley value or the Banzhaf value. For instance, Shapley [3], Hart and Mas-Colell [6] or Feltkamp [5] for the Shapley value, and Lehrer [7] or Feltkamp [5] for the Banzhaf value. The aim of the present work is to state an axiomatic characterization for each regular semivalue.

In the characterization of the Banzhaf value obtained by Lehrer [7], a cooperative game essential to the axiomatic system is introduced: the quotient game in the case where only two players join together. The procedure followed here is a generalization of Lehrer's method. It requires the introduction of a particular game for each regular semivalue.

The specific games introduced in [1] are called *delegation games*. Given a cooperative game on a finite set of players and two distinct players i and j , a delegation game from player j to player i attributes to every coalition S containing player i the utility that this coalition would obtain increased by the contribution marginal of player j to coalition $S \cup \{j\}$ weighted by a certain coefficient. If player i does not belong to S , this coalition loses the marginal contribution of player j weighted by another coefficient.

2. Acknowledgement

Research partially supported by Grant BFM 2003-01314 of the Spanish Ministry of Science and Technology and the European Regional Development Fund.

3. Bibliography

- [1] Amer, R. and Giménez, J.M. (2005). A system of specific properties for each semivalue on cooperative games with positive weighting coefficients. Working paper MA2-IR-05-00013. Universitat Politècnica de Catalunya.
- [2] Banzhaf, J.F. (1965). Weighted voting doesn't work: A mathematical analysis. *Rutgers Law Review* 19, 317–343.
- [3] Carreras, F. and Freixas, J. (1999). Some theoretical reasons for using (regular) semivalues. In: Swart, H. (ed.), *Logic, Game Theory and Social Choice*. Tilburg University Press, 140–154.
- [4] Dubey, P., Neyman, A., Weber and R.J. (1981). Value theory without efficiency. *Mathematics of Operations Research* 6, 122–128.
- [5] Feltkamp, V. (1995). Alternative axiomatic characterizations of the Shapley and Banzhaf values. *International Journal of Game Theory* 24, 179–186.
- [6] Hart, S. and Mas-Colell, A. (1989). Potential, value and consistency. *Econometrica* 57, 589–614.
- [7] Lehrer, E. (1988). An axiomatization of the Banzhaf value. *International Journal of Game Theory* 17, 89–99.
- [8] Owen G. (1975). Multilinear extensions and the Banzhaf value. *Naval Research Logistics Quarterly* 22, 741–750.
- [9] Shapley, L.S. (1953). A value for n-person games. In: Kuhn H.W. and Tucker A.W. (eds.), *Contributions to the Theory of Games II*. Princeton University Press, 307–317.

A GRASP algorithm for VRPTW

F. Cano Sevilla¹, C. Simón de Blas², S. Cano Alsúa³, A. Felipe Ortega⁴

¹cano@mat.ucm.es, Departamento de Estadística e Inv. Operativa, UCM

²clara.simon@urjc.es, Departamento de Estadística e Inv. Operativa, URJC

³scano@pas.ucm.es, Servicio de Apoyo a la docencia y la investigación, UCM

⁴felipe@mat.ucm.es, Departamento de Estadística e Inv. Operativa, UCM

Abstract

We propose a two phase algorithm based in the generation of a good initial set of feasible solutions and a GRASP algorithm for the **V**ehicle **R**outing **P**roblem with **T**ime **W**indows (**VRPTW**) problem. The main idea of the algorithm is to take advantage of the location of the client respect to the supplier and the time window disposal.

Keywords: Vehicle routing; time windows. **Heuristics.**

A Greedy approach to solve the Multidimensional Assignment Problem

F. Perea¹, H. W. de Waard²

¹federico.perea@nl.thalesgroup.com ²huub.dewaard@nl.thalesgroup.com
AWS Department, Thales Naval Nederland B.V.

Abstract

El problema de asignación multidimensional (MDA) es un problema de optimización combinatoria, de complejidad NP. En este trabajo se presenta un algoritmo tipo Greedy para su resolución. Se estudia la eficiencia de este algoritmo como aplicación a problemas en los que diferentes sensores han de clasificar objetos en el espacio.

Keywords: Multidimensional assignment, tracking, approximation

1. Introduction

The objective of target tracking is to obtain an estimate of the state (position, velocity, ...) of a moving object (from now on called *target*) using one or more sensors at fixed locations or on moving platforms, a so-called Multi-Target Multi-Sensor Tracking (MTMST) problem [1]. Each sensor collects a set of *observations*, also called *plots* or *measurements*, which originates from an existing target or a false alarm. The widespread use and increasing sophistication of surveillance systems, both military and civilian, has generated a great deal of interest in algorithms capable of tracking large numbers of targets using measurement data from many and possibly diverse sensors.

Suppose that a sensor starts observing the airspace periodically at a time $y_0 = 0$. During the first *scan*, corresponding with the time interval $[y_0, y_1)$, the first set of measurements is received. Variable y_1 denotes the point in time in which the first scan of the sensor finishes and the second one begins. Analogously, the second set of measurements, corresponding with the second scan, is received within the time interval $[y_1, y_2)$ and so on.

Let $Z(k)$ be the set of measurements collected at scan k . The cumulative data set for N scans is denoted by $Z^N = \{Z(1), \dots, Z(N)\}$. A *track hypothesis* t is defined as a set of measurements of Z^N such that it contains at most one measurement of each scan and consists of at least one measurement. This can be mathematically expressed as:

$$t \subset Z^N : |t| \geq 1, |t \cap Z(k)| \leq 1 \quad \forall k = 1, \dots, N. \quad (4.61)$$

Assigning plots to targets can be formulated as a MDA problem [3, 3]. In order to express MTMST problems as MDA problems, a linear objective function has

to be defined. To obtain such a function, a quality measure $Q(t)$ is assigned to each track hypotheses t , which expresses how well each measurement of t fits the target's assumed dynamical model [4].

Due to the NP-hardness of MDA problems, the search for fast approximation algorithms for solving them is of crucial importance [5].

2. The Greedy approach

For each track t , the effectiveness of t , $\alpha(t)$, is defined as the result of dividing its quality measure $Q(t)$ by the number of measurements that constitute t . The algorithm proposed in this paper runs as follows. First it chooses the track hypothesis with the highest effectiveness as the first element of the solution. At this point it removes all the track hypotheses that have at least one measurement in common with the previously selected one. Then, it iteratively performs the same two steps: 1) select the track hypothesis t with maximum effectiveness from the remaining set, 2) remove all track hypotheses intersecting that contain at least one of the measurements in t .

The algorithm runs until the set of available tracks is empty, obtaining this way the first solution. In order to obtain a larger variety of solutions, the algorithm is carried out a required number of times, starting always from a track hypothesis that has not been included in any previous solution.

3. Properties

The complexity of the proposed approximation algorithm is proven to be polynomial, $O(n \log n)$. This algorithm is an $O(\log n)$ factor approximation algorithm. Based on simulations performed for MTMST problems, some results on its accuracy and speed are provided.

4. Bibliography

- [1] Bar-Shalom, Y., Li, X-R. (1986). *Multitarget-Multisensor Tracking: Principles and Techniques*. Storrs, CT: YBS Publishing.
- [2] W. Pierskalla, (1968) The multidimensional assignment problem. *Operations Research* 16, 422-431.
- [3] Poore, A.B. (1994). Multidimensional Assignment Formulation of Data Association Problems Arising from Multitarget and Multisensor Tracking. *Computational Optimization and Applications* 3, 27-57.
- [4] Reid, D.B. (1979) An Algorithm for Tracking Multiple Targets. *IEEE Transactions on Automatic Control* vol. AC-24, N. 6.
- [5] Vazirani, V.V. (2001) *Approximation Algorithms*. Springer.

A value for games with asymmetric cooperation among the players

Rafael Amer¹, José Miguel Giménez², Antonio Magaña³

¹rafel.amer@upc.edu, Department of Applied Mathematics II, Universitat Politècnica de Catalunya

²jose.miguel.gimenez@upc.edu, Department of Applied Mathematics III, Universitat Politècnica de Catalunya

³antonio.magan@upc.edu, Department of Applied Mathematics II, Universitat Politècnica de Catalunya

Abstract

Diversas situaciones económicas y sociales presentan posibilidades asimétricas de cooperación entre sus actores. El valor que se introduce pretende conjugar la importancia de cada jugador en un juego cooperativo junto a su posición estratégica en una red orientada de comunicación que pueda modelizarse mediante un digrafo.

Keywords: Cooperative game, Digraph, Shapley value.

AMS: 91A12

1. Introduction

The aim of this work consists of allocating a value that allows us to emphasize the importance of each player in a cooperative game when the cooperation among the players have asymmetric possibilities. In a classical cooperative game every subset of players or *coalition* gets a utility as a result of the social or economical power of its members, which are supposed to behave as a block. Initially, any coalition is likely to form, and the set of the utilities of all of them forms the game's characteristic function.

It seems, therefore, convenient to widen the analysis to situations of cooperation where the order in which players constitute the coalitions should be taken into account. In fact, oriented networks are nowadays present in many economic and social situations. It is enough to think about hierarchy relations, sports competitions, paired comparisons, mutual citations among journals or, in general, oriented channels of communication with a predefined sense.

A digraph is a directed graph consisting of a finite set N of *nodes* and a collection of ordered pairs of these nodes that we call *oriented edges*. An oriented edge from one node i to another node j supposes that it is possible to accede to the node j from the node i . In the framework of accessibility [1], node j has a better position than node i .

In this work, we consider cooperative games modified by structures of *oriented cooperation*, that we model using digraphs, as the same way that Myerson [2] considered symmetric partial cooperation modeled by means of graphs. We think that it is interesting to allocate to every player a payment that reflects two aspects: on the one hand, the importance in the considered game and, on the other hand, the importance of its geometric position in the oriented structure represented by the digraph. Hence, the main purpose is to establish a new rule that allows to order the nodes of an oriented network following two initial considerations: the marginal contribution of every node to the different coalitions according to the fixed game, and the possible ordered sequences of nodes depending on the geometry of the oriented network.

Diverse properties of this value for cooperative games modified by structures of oriented cooperation will be proved. In particular, this value provides an extension of the Shapley value [3]: when the complete digraph is considered, the value agrees with the Shapley value of each cooperative game. In addition, we offer explicit expressions of the value when the digraphs are oriented paths, considered as elementary digraphs. A set of these properties allows us to determine completely this value on games modified by structures of oriented cooperation.

2. Acknowledgement

Research partially supported by Grant BFM 2003–01314 of the Spanish Ministry of Science and Technology and the European Regional Development Fund.

3. Bibliography

- [1] Amer, R., Giménez, J.M. and Magaña, A. (2004). A value for cooperative games in generalized characteristic function form induced by oriented structures. Working paper MA2–IR–04–00004. Universitat Politècnica de Catalunya.
- [2] Myerson, R.B. (1977). Graphs and cooperation in games, *Mathematics of Operations Research* 2, 225–229.
- [3] Shapley, L.S. (1953). A value for n-person games. In: Kuhn H.W. and Tucker A.W. (eds.), *Contributions to the Theory of Games II*. Princeton University Press, 307–317.

Agregación de preferencias y DEA: Un análisis de los métodos propuestos para discriminar candidatos eficientes

Bonifacio Llamazares¹, Teresa Peña²

¹boni@eco.uva.es, Dpto. de Economía Aplicada, Universidad de Valladolid

² maitepe@eco.uva.es, Dpto. de Economía Aplicada, Universidad de Valladolid

Abstract

En este trabajo se realiza un análisis de algunos procedimientos que han sido propuestos en la literatura para discriminar entre los candidatos eficientes que resultan al agregar las preferencias ordinales de un conjunto de votantes mediante el empleo del análisis envolvente de datos (DEA).

Claves: Sistemas de puntuación posicionales, DEA, discriminación de candidatos eficientes.

AMS: 90B50, 91B12, 90C90

1. Introducción

En muchos sistemas de votación es habitual considerar que las preferencias de los votantes sobre un conjunto de candidatos se representan mediante órdenes lineales. Uno de los procedimientos más conocidos para agregar dichas preferencias ordinales y elegir al ganador es asignar puntuaciones fijas a las diferentes posiciones que ocupan los candidatos y seleccionar aquellos que obtengan la máxima puntuación (sistemas de puntuación posicionales). Así, si se considera un conjunto de m candidatos entre los que cada votante selecciona k de ellos, la puntuación obtenida por el candidato $i \in \{1, \dots, m\}$ es $Z_i = \sum_{j=1}^k w_j v_{ij}$, donde v_{ij} es el número de veces que dicho candidato ocupa la j -ésima posición y (w_1, \dots, w_k) es el vector de puntuaciones utilizado. Entre estos sistemas, los más conocidos son el método de pluralidad y la regla de Borda.

A pesar de que el uso de estos sistemas está muy extendido en algunos contextos decisionales, el hecho de utilizar un vector fijo plantea un problema desde el punto de vista de los candidatos: existen situaciones en las que un candidato que no resulta ganador con el vector de puntuaciones fijado inicialmente podría serlo si se utilizase otro distinto. Para evitar este problema, Cook y Kress [1] proponen evaluar a cada candidato con el vector de puntuaciones que más le favorezca. Para ello introdujeron en este contexto el análisis envolvente de datos. El modelo DEA/AR planteado por estos autores es el siguiente:

$$\begin{aligned}
 Z_o^* &= \max \sum_{j=1}^k w_j v_{oj}, \\
 \text{s.a. } &\sum_{j=1}^k w_j v_{ij} \leq 1, \quad i = 1, \dots, m, \\
 &w_j - w_{j+1} \geq d(j, \epsilon), \quad j = 1, \dots, k-1, \\
 &w_k \geq d(k, \epsilon),
 \end{aligned} \tag{4.62}$$

donde $d(j, \epsilon)$ es una función positiva y no decreciente en ϵ . Sin embargo, cabe destacar que con este procedimiento pueden resultar eficientes varios candidatos, es decir, que alcancen la máxima puntuación $Z_o^* = 1$. Por este motivo, han aparecido en la literatura distintos métodos para discriminar entre los candidatos eficientes. Así, Cook y Kress [1] proponen maximizar en el modelo (4.62) la diferencia entre pesos consecutivos. Green *et al.* [2] plantean discriminar entre los candidatos eficientes utilizando una matriz de evaluación cruzada. Hashimoto [3] propone eliminar del conjunto de restricciones del problema (4.62) la relativa al candidato que está siendo evaluado y añade una nueva restricción para que la secuencia de pesos sea convexa. Finalmente, Obata y Ishii [4] consideran que lo adecuado es comparar la máxima puntuación obtenida por cada candidato utilizando vectores de puntuaciones del mismo tamaño, por lo que proponen normalizar los vectores de puntuaciones más favorables para cada uno de los candidatos eficientes.

En este trabajo se muestran algunos problemas que presentan los métodos anteriormente citados. En este sentido, se analizan distintos supuestos como el de independencia de candidatos ineficientes y se muestran diversas inconsistencias que presentan algunos de estos métodos en la ordenación final de los candidatos.

2. Bibliografía

- [1] Cook W.D. and Kress M. (1990). A data envelopment model for aggregating preference rankings. *Management Science* 36 (11), 1302–1310.
- [2] Green R.H., Doyle J.R. and Cook W.D. (1996). Preference voting and project ranking using DEA and cross-evaluation. *European Journal of Operational Research* 90, 461–472.
- [3] Hashimoto A. (1997). A ranked voting system using a DEA/AR exclusion model: A note. *European Journal of Operational Research* 97, 600–604.
- [4] Obata T. and Ishii H. (2003). A method for discriminating efficient candidates with ranked voting data. *European Journal of Operational Research* 151, 233–237.

Algoritmo Genético (GA) para una línea de producción con posibilidad de resecuenciar considerando almacenes limitados

Gerrit Färber, Anna M. Coves Moreno

Instituto de Organización y Control de Sistemas Industriales (IOC),
Universidad Politécnica de Cataluña (UPC), España.
gerrit_faerber@gmx.de, anna.maría.coves@upc.es

Abstract

¹Este trabajo presenta un Algoritmo Genético (GA) del problema de secuenciar unidades en una línea de producción. Se tiene en cuenta la posibilidad de cambiar la secuencia de piezas mediante estaciones con acceso a un almacén intermedio o centralizado. El acceso al almacén además está restringido, debido al tamaño de las piezas.

Palabras Clave: Algoritmo Genético; Non-permutation flowshop; Almacén limitado, intermedio y centralizado.

1. Introducción

El problema de secuenciar unidades en una línea aparece cuando se producen variaciones o modelos del mismo producto básico en la misma línea de producción. Estas variaciones de producto implican que el tiempo de proceso en una estación difiera, dependiendo de las unidades. La gran mayoría de trabajos se limita a soluciones con una permutación (permutation flowshop). En [7] se estudia el beneficio de usar una línea de flujo no-permutación. Existen además varios diseños de líneas que permiten la resecuenciación, desacoplando una parte del resto de la línea [5], almacenes con algunas plazas fuera de la línea [4], líneas híbridas o flexibles [9], la división y unión de líneas [2] o el cambio de los atributos de las piezas en vez de la posición en la secuencia, [8]. La resecuenciación de piezas en la línea llega a ser más efectiva cuando se presenta setup-time y setup-cost, [1].

2. Descripción de la línea

El trabajo realizado presenta un Algoritmo Genético (basado en [6]) para el problema [3] donde se considera una línea de flujo en la cual las unidades pasan consecutivamente por las estaciones. Además, después de determinadas estaciones, se permite resecuenciar las unidades mediante un almacén, colocado fuera de la línea. Los almacenes disponen de varias plazas y para obtener un diseño reducido, cada plaza está restringida por el tamaño de la unidad que pueda ser almacenada temporalmente. En un primer paso se colocan los almacenes entre estaciones consecutivas. Después, para un beneficio adicional,

¹Este trabajo se deriva de la participación en el proyecto de investigación financiado por el Ministerio de Ciencia y Tecnología con cofinanciación del Fondo de Desarrollo Regional con referencia DPI2004-03472, titulado “Diseño y equilibrio de líneas de montaje en entornos realistas”.

se utiliza un solo almacén centralizado, con acceso desde varias estaciones, mientras se mantienen las restricciones de tamaño.

3. Modelo y resolución

El Algoritmo Genético propuesto utiliza una cascada de dos pasos. Primero, mediante cromosomas reducidos, se limita el algoritmo a soluciones de permutación que sirven de soluciones iniciales para el segundo paso, en el cual se amplia las soluciones a no-permutación. Se utiliza como operadores genéticos la herencia, dos tipos de cruce y tres tipos de mutación. Se incluye como función objetivo principalmente la minimización del makespan, el setup-cost y el setup-time. Además se presenta como incorporar una penalización para soluciones prometedores pero no factibles, teniendo en cuenta el tamaño físico de las unidades con respecto al tamaño físico de la plaza disponible del almacén intermedio o centralizado. El modelo se ha resuelto de forma efectiva mediante programación en C.

4. Bibliografía

- [1] Bolat A. (1994). Sequencing jobs on an automobile assembly line: objectives and procedures. *Int. J. Prod. Res.*, 32, 5, 1219–1236.
- [2] Engström T., Jonsson D. and Johansson, B. (1996). Alternatives to line assembly: Some swedish examples. *International Journal of Industrial Ergonomics*, Vol.17, No.3, 235–245.
- [3] Färber G. and Coves A.M. (2005). Modelo de programación lógica de restricciones (CLP) para una línea de producción con posibilidad de resecuenciar considerando almacenes limitados. *IX Congreso de Ingeniería de Organización*, Gijón, Spain
- [4] Lahmar M., Ergan H. and Benjaafar S. (2003). Resequencing and feature assignment on an automated assembly line. *IEEE Transactions on Robotics and Automation*, Vol.19, No.1, 89–102.
- [5] Lee H. and Schaefer S. (1997). Sequencing methods for automated storage and retrieval systems with dedicated storage. *Computers and Industrial Engineering*, Vol.32, No.2, 351–362.
- [6] Michalewicz Z. (1996). Gentic Algorithms + Data Structures = Evolution Programs, Springer Verlag, 3a edición.
- [7] Potts C., Shmoys D. and Williamson D. (1991). Permutation vs. non-permutation flow shop schedules. *Operations Research Letters*, Vol. 10, No. 5, pp. 281–284.
- [8] Rachakonda P. and Nagane S. (2000). Simulation Study of Paint Batching Problem in Automobile Industry, <http://sweb.uky.edu/pkrach0/Projects/MFS605Project.pdf>, consultado 14.07.2004
- [9] Sawik T. (2000). Mixed integer programming for scheduling flexible flow lines with limited intermediate buffers. *Mathematical and Computer Modelling*, Vol.31, No.13, 39–52.

Algoritmos heurísticos para secuenciación de proyectos con costes sobre los recursos

R. Alvarez-Valdés¹, G. Martín², J. M. Tamarit¹

¹Depto. de Estadística e Investigación Operativa, Universidad de Valencia

²Universidad Católica de Valencia

Abstract

Estudiamos el problema de secuenciación de proyectos cuando los recursos pueden considerarse ilimitados, disponibles a un cierto coste, y se ha de minimizar el coste asociado al máximo consumo de dichos recursos. Hemos desarrollado algoritmos metaheurísticos, basados en GRASP e Iterated Local Search, y las soluciones obtenidas se combinan en una fase de Path Relinking.

Palabras Clave: Secuenciación de proyectos, Nivelación de recursos, Algoritmos metaheurísticos

1. Introducción

El problema de secuenciación de proyectos con costes sobre los recursos (RACP) es el problema en que el tiempo para finalizar el proyecto está limitado, pero los recursos pueden considerarse ilimitados con costes no decrecientes. El objetivo es minimizar el coste máximo de los recursos necesarios para completar el proyecto a tiempo. Se trata de un caso particular del problema de nivelación de recursos, estudiado por Neumann et al. (2002).

RACP fue inicialmente propuesto por Möhring (1984) para resolver un problema de construcción de un puente. Posteriormente, se han propuesto procedimientos exactos (Demeulemeester, 1995), cotas inferiores (Drexl & Kimms, 2000) y heurísticos basados en Scatter Search (Yamashita et al., 2006).

En este trabajo describimos el problema, presentamos una formulación entera, y desarrollamos dos algoritmos metaheurísticos, basados en *GRASP* y en *ILS*. Las soluciones obtenidas por estos dos métodos se combinan en una segunda fase basada en *Path Relinking*.

2. Algoritmo GRASP

Los tres elementos básicos de un algoritmo GRASP son el procedimiento constructivo *greedy*, la aleatorización y la fase de mejora. La solución se construye utilizando una adaptación de los procedimientos de secuenciación en serie, con un criterio de prioridad basado en el mínimo aumento de los máximos consumos de recursos. La elección de la actividad a secuenciar y el tiempo en el que procesarse se elige mediante un procedimiento de elección sesgada sobre una lista restringida de candidatos. En cuanto a la fase de mejora, como el movimiento simple, en el que una actividad se traslada a un nuevo tiempo, es muy limitado, hemos desarrollado un movimiento más complejo en el que

se intenta desplazar simultáneamente una actividad implicada en el máximo consumo de un recurso y sus antecesoras (o sucesoras).

3. Iterated Local Search (ILS)

Los elementos básicos de este procedimiento son la perturbación y la mejora local. La solución se perturba eligiendo al azar una actividad y desplazándolo al máximo a la izquierda (o derecha), junto con todos sus antecesores (sucesores). La mejora local utiliza el movimiento descrito en el GRASP.

4. Path Relinking

Las 5 mejores soluciones obtenidas por el GRASP y las 5 mejores de ILS se utilizan en un procedimiento de Path Relinking en el que todas las parejas de soluciones son consideradas por turno y cada una de ellas es, primero, la solución guía, mientras la otra es la solución inicial, y luego al contrario. Las soluciones intermedias se obtienen imponiendo a la solución inicial los tiempos de proceso de las actividades en la solución guía, uno en cada iteración, y mejorando la solución obtenida en un proceso similar al ya explicado en los procedimientos anteriores.

5. Resultados computacionales

Los algoritmos desarrollados han sido probados sobre los dos conjuntos de problemas test generados por Drexel y Kimms (2000) y Yamashita et al. (2006). Los resultados obtenidos por el momento muestran que la combinación de algoritmos propuesta en este trabajo consigue resultados similares a los mejores resultados publicados hasta el momento.

6. Agradecimientos

Este trabajo ha sido financiado en parte por el M.E.C., DPI2005-04796.

7. Bibliografía

- [1] Demeulemeester, E.L. (1995) Minimizing resource availability costs in time-limited project networks, *Management Science* **41**, 1590-1598.
- [2] Drexel, A., Kimms, A. (2000) Optimization guided lower and upper bounds for the resource investment problem, *Journal of the Operational Research Society* **52**, 340-351.
- [3] Möhring, R.H. (1984) Minimizing costs of resource requirements in project networks subject to a fixed completion time, *Operations Research* **32**, 89-120.
- [4] Neumann, K, Schwindt, C., Zimmermann, J. (2002) *Project scheduling with time windows and scarce resources*, Springer Verlag, Berlin.
- [5] Yamashita, D.S., Armentano, V.A., Laguna, M. (2006) Scatter search for project scheduling with resource availability cost, *European Journal of Operational Research* **169**, 623-637.

Algoritmos para la construcción del conjunto de árboles generadores bajo un sistema de preferencia: condiciones y diseño

Miguel Ángel Domínguez-Ríos, Sergio Alonso, Marcos Colebrook,
Antonio Sedeño-Noda

{madoming, salonso, mcolesan, asedeno}@ull.es,
Departamento de Estadística, Investigación Operativa y Computación,
Universidad de La Laguna

Abstract

Presentamos aquí la fase final del estudio de las condiciones necesarias y suficientes a exigir a los sistemas de preferencia definidos sobre subconjuntos de aristas de un grafo conexo, para que los árboles generadores más preferidos verifiquen ciertas condiciones de optimalidad. El cumplimiento de tales condiciones facilita la labor de diseño de algoritmos para la construcción de los árboles generadores más preferidos y constituye una elegante generalización del problema del árbol generador de mínimo coste.

Palabras Clave: Sistemas de preferencias, Árboles generadores de mínimo coste, Árboles generadores multicriterio

1. Introducción

Sea $G = (V, E)$ un grafo no dirigido y conexo, donde denotamos por $V(G)$ al conjunto de n nodos de G y por $E(G)$ al conjunto de aristas de G . Un árbol generador de G es un subgrafo que conecta con $n-1$ aristas los n nodos de $V(G)$. Denotamos por $\sigma(G)$ al conjunto de todos los árboles generadores de G .

Dado un par de nodos, $i, j \in V(G)$, y $T \in \sigma(G)$, denotamos por $path\langle T, (i, j) \rangle$ al conjunto de aristas de T que conecta los nodos i y j . Asimismo, dado $T \in \sigma(G)$ y una arista $e \in E(G)$, denotamos por $cut\langle T, e \rangle$ al conjunto de aristas de $E(G)$ que conectan las componentes conexas de $T \setminus \{e\}$. Un estudio profundo y serio sobre las condiciones de optimalidad para el problema escalar, el árbol generador de mínimo coste, puede encontrarse en [1].

La relación de preferencia sobre la que trabajamos, denotada por \succ , está definida entre pares de subconjuntos de aristas. Le pedimos inicialmente cierta racionalidad, es decir, asimetría y transitividad. Sobre esta relación de preferencia, construimos la indiferencia \sim y con ella la relación \gtrsim .

Bajo esta estructura, numerosos problemas de optimización multicriterio pueden ser estudiados. En particular, cuando el sistema de preferencia se basa en órdenes de Pareto (véase [3]) o lexicográfico.

Finalmente, decimos que un árbol generador $T \in \sigma(G)$ verifica la condición de optimalidad del *corte* generalizada si $\forall e \in T, \forall f \in \text{cut}\langle T, e \rangle$, si $f \succsim e$ entonces $f \sim e$. El conjunto de árboles que verifica tal propiedad lo denotamos por $\text{OptCUT}(G)$. Análogamente, decimos que un árbol generador $T \in \sigma(G)$ verifica la condición de optimalidad de la *ruta* generalizada si $\forall f = (i, j) \in E(G), \forall e \in \text{path}\langle T, (i, j) \rangle$, si $f \succsim e$, entonces $f \sim e$. El conjunto de árboles que verifica tal propiedad lo denotamos por $\text{OptPATH}(G)$. Se demuestra que $\text{OptCUT}(G) = \text{OptPATH}(G)$.

2. Diseño de algoritmos

La propiedad fundamental en la que nos basamos para la conclusión de este trabajo y el diseño efectivo de algoritmos es la siguiente:

Proposición. Consideremos un grafo no dirigido y conexo $G = (V, E)$ con un sistema de preferencia racional \sim definido sobre los subconjuntos de aristas $E(G)$. Sea $T \in \text{OptCUT}(G)$ y sea F un subconjunto de aristas de T . Denotamos por S los nodos de cualquiera de las componentes conexas de F . Si \sim es negativamente transitiva, entonces algún árbol generador de $\text{OptCUT}(G)$ contiene las aristas de F y una cualquiera de las aristas más preferidas en el corte $[S, \bar{S}]$.

Con esta propiedad aplicadas a las clásicas estrategias de Prim y de Kruskal para el problema del árbol generador de mínimo coste, se deducen sendos algoritmos y sus condiciones para la construcción del conjunto de árboles en $\text{OptCUT}(G)$. Dichos algoritmos mejoran los propuestos en [2].

3. Bibliografía

- [1] R. K. Ahuja, T. L. Magnanti and J. B. Orlin, *Network Flows. Theory, Algorithms and Applications*. Prentice Hall, 1993.
- [2] P. Perny and O. Spanjaard, “A preference-based approach to spanning trees and shortest path problems”, in *European Journal of Operational Research*, vol. 162, issue 3, pages 584-601, May 2005.
- [3] R. M. Ramos, S. Alonso, J. Sicilia and C. González, “The problem of the optimal biobjective spanning tree” in *European Journal of Operational Research*, vol. 111(3), pages 617-628, 1998.

An evolutionary algorithm for the resource-constrained project scheduling problem subject to temporal constraints

Vicente Valls¹, Francisco Ballestín², Agustín Barrios³

¹vicente.valls@uv.es, Departamento de Estadística e Investigación Operativa,
Universidad de Valencia,

²francisco.ballestin@uv.es, Departamento de Estadística e Investigación
Operativa, Universidad Pública de Navarra,

³abarrios@uninorte.edu.co, Departamento de Matemáticas y Física,
Universidad del Norte, Colombia

Abstract

En este trabajo presentamos un algoritmo evolutivo para el problema RCPSP/max. Los resultados computacionales comparativos muestran su gran eficiencia.

Keywords: project scheduling, evolutionary algorithms.

1. Introduction

The Resource-Constrained Project Scheduling Problem with maximum and minimum time lags (RCPSP/max) is an important extension of the classic RCPSP in which generalised precedence constraints are allowed. Problem RCPSP/max, as an extension of the RCPSP, is NP-hard, and even the question whether a problem instance has a feasible solution is NP-hard (Bartusch et al. 1988).

In this paper we present an evolutionary algorithm (EVA) for solving the RCPSP/max problem. EVA works on a population consisting of several *distance-order-preserving activity lists* (*L*) representing feasible or infeasible schedules. EVA uses the *conglomerate-based crossover operator* (*CONG*), the objective of which is to exploit the knowledge of the problem to identify and combine those good parts of the solution that have really contributed to its quality. In a recent paper, Valls et al. (2005) showed that incorporating justification in RCPSP heuristic algorithms can produce a substantial improvement in the results obtained. EVA also applies two double justification operators *DJmax* and *DJU* adapted to the specific characteristics of problem RCPSP/max. Given a priority rule, the serial generation scheme with unscheduling (*SGSU*) proceeds similarly to the SGS for the RCPSP except when an activity cannot be feasibly scheduled. At this instant, SGSU first unschedules some activities and then re-schedules part of them at later time instants.

2. The EVA algorithm

1. Apply pre-processing procedure

2. Generate initial population $POP = L_1, L_2, \dots, L_{POPsize}$
3. $POPnew = POP$. While no stopping rule is fulfilled:
4. For $i = 1$ to $POPsize$:
 - 4.1. Randomly select L' and L'' from POP
 - 4.2. $L = CONG(L_i; L', L'')$
 - 4.3. $S = SGSU(L)$
 - 4.3.1. If S feasible then $L^* = DJU(S)$
 - 4.3.2. If S infeasible then $L^* = DJmax(S)$
 - 4.4. $POPnew = POPnew \cup L^*$
5. $POP = POPsize$ best activity lists in $POPnew$. Go to 3.
6. Return the best solution found

3. Experimental performance analysis

We have performed a computational study concerning the algorithm EVA and compare it to state-of-the-art heuristic algorithms. The results are based on two test sets used by many researches: UBO100 (Franck et al. (2001), 360 instancias) y CD (Dorndorf et al. (2000), Cesta et al. (2002), 1080 instancias). The analysis of the computational results indicates that:

- DJ is a profitable technique also for the RCPSP/max.
- EVA yields the lowest average deviation from the lower bound.
- EVA is one of the fastest, if not the fastest, heuristic algorithms for problem RCPSP/max.
- EVA obtains feasible schedules in all feasible instances.

4. Bibliography

- [1] Bartusch, M., Möhring, R. H., Radermacher, F. J. (1988). Scheduling project networks with resource constraints and time windows. *Ann. Oper. Res.* 16, 201–240.
- [2] Cesta, A., Oddi, A., Smith, S. (2002). A Constraint Based Method for Project Scheduling with Time Windows. *Journal of Heuristics* 8(1), 109 - 136.
- [3] Dorndorf, U., Pesch, E., Phan-Huy, T., (2000). A time-oriented branch-and-bound algorithm for resource-constrained project scheduling with generalised precedence constraints. *Management Science* 46, 1365–1384.
- [4] Franck, B., Neumann, K., Schwindt, Ch., (2001). Truncated branch and bound, schedule construction, and schedule improvement procedures for resource constrained project scheduling. *OR Spektrum*, Springer - Verlag, 23, 297 - 324.
- [5] Valls, V., Ballestín, F., Quintanilla, S., (2005). Justification and RCPSP: a technique that pays. *European Journal of Operational Research* 165, 375-386.

Análisis del efecto de las incertidumbres en la optimización de los intervalos de pruebas y mantenimiento bajo criterios de coste y riesgo

Sofía Carlos¹, Ana Sánchez², Sebastián Martorell³

¹scarlos@iqn.upv.es, Departamento de Ingeniería Química y Nuclear,
Universidad Politécnica de Valencia

²aisanche@eio.upv.es, Departamento de Estadística e Investigación Operativa
Aplicadas y Calidad , Universidad Politécnica de Valencia

¹smartore@iqn.upv.es, Departamento de Ingeniería Química y Nuclear,
Universidad Politécnica de Valencia

Abstract

La optimización de las actividades de vigilancia y mantenimiento realizadas en los diferentes sistemas de una planta industrial constituyen un tema de gran interés puesto que diferentes aspectos tanto económicos como funcionales de las mismas depende en gran medida de las actividades planificadas. En esta ponencia se presenta una optimización multiobjetivo de los intervalos de vigilancia y mantenimiento utilizando algoritmos genéticos y teniendo en cuenta las incertidumbres asociadas tanto a la selección del modelo de mantenimiento imperfecto como a la efectividad del mismo.

Palabras Clave: Mantenimiento imperfecto, incertidumbres, algoritmos genéticos

1. Introducción

La optimización de las actividades de vigilancia y mantenimiento realizadas en los diferentes sistemas de una planta industrial constituye un tema de gran interés puesto que la disponibilidad y coste de las mismas depende en gran medida de las actividades planificadas. Tradicionalmente, se han considerado dos tipos de modelos, deterministas y probabilistas, para simular el efecto que las pruebas y el mantenimiento tienen sobre el coste e indisponibilidad de los sistemas. Ambos tipos de modelos involucran incertidumbres, las cuales pueden ser clasificadas en aleatorias o epistémicas. Las incertidumbres epistémicas aplican cuando existe un conocimiento limitado sobre el modelo apropiado para representar el problema y/o los valores asociados a dichos modelos.

En los últimos años la incorporación de incertidumbres en el análisis de sistemas complejos se ha convertido en un área de gran interés. Así, diferentes métodos se han propuesto para considerar los dos tipos de incertidumbres y un considerable número de estudios se han publicado en la última década relacionados con el tratamiento de incertidumbres de tipo epistémico.

2. Objetivo

La presente ponencia se centra en la optimización de los intervalos de vigilancia y mantenimientos basado sobre criterios de coste y riesgo y considerando incertidumbres de tipo epistémico. Concretamente el tipo de incertidumbre que se considera es el asociado con el efecto que las actividades de mantenimiento tienen sobre el componente. Dicho efecto puede ser simulado mediante un modelo de mantenimiento imperfecto, existiendo en la literatura diferentes expresiones que permiten formular dicho mantenimiento, siendo la selección del modelo más adecuado dependiente del juicio de expertos. Adicionalmente, los modelos de mantenimiento imperfecto asumen que cada actividad mejora el estado del componente en función de un parámetro denominado efectividad, cuyo valor se estima en base a datos reales o juicio de expertos.

Sobre la base anterior en esta ponencia se presenta la optimización de los intervalos de vigilancia y mantenimiento basado en criterios de riesgo y coste y considerando incierto tanto el modelo de mantenimiento imperfecto como la efectividad asociada a las diferentes actividades realizadas. Este problema de optimización constituye un problema de toma de decisiones multicriterio donde la indisponibilidad y coste actúan como criterios de decisión inciertos y en conflicto y los intervalos de pruebas y mantenimiento actúan como variables de decisión.

El tratamiento de dichas incertidumbres se realiza mediante una aproximación basada en el intervalo de tolerancias. Dicha aproximación se ha implementado en una algoritmo genético multiobjetivo que permite llevar a cabo la optimización de los intervalos de pruebas y mantenimiento.

Finalmente, se ha realizado un caso de aplicación centrado en la optimización de los intervalos de pruebas y mantenimiento de un sistema de seguridad de una central nuclear. Los resultados obtenidos en dicho caso de aplicación y los diferentes análisis de sensibilidad realizados muestran la importancia de considerar incertidumbres en el modelo de mantenimiento imperfecto y en la efectividad durante el proceso de optimización, puesto que las soluciones óptimas encontradas son considerablemente diferentes modificando la toma de decisiones final.

Aplicación de ecuaciones estructurales para la valoración de sistemas de armas y materiales en un simulador de combate

M. Molero, A. Planells, O. Muñoz, J. Padura, A. Castellote, F. Bargueño

Unidad Central de Investigación Operativa Cuartel General del Ejército de Tierra. 28004 Madrid

Abstract

El trabajo responde a la necesidad de establecer un criterio para la valoración de los sistemas de armas y materiales en un sistema de simulación de combate que permita la introducción de nuevos elementos en la acción.

Palabras Clave: Análisis multivariante, Ecuaciones estructurales, Simulación.

AMS: 62J12, 62J99

1. Introducción

Los sistemas de simulación de combate para el entrenamiento de tripulaciones o cuadros de mando se han convertido en un elemento esencial para la instrucción de personal en los ejércitos modernos. Los altos costes del material, la sofisticación de los sistemas de armas y comunicaciones y la necesidad de ajustar los gastos de la instrucción a presupuestos limitados, han hecho de la simulación un elemento imprescindible en la formación del personal militar en todo el mundo.

Para que el desarrollo de un sistema de simulación sea rentable, debe ser lo suficientemente flexible como para poder incorporar aquellas variaciones que se produzcan por la necesaria modernización del material o por la introducción de elementos nuevos en el campo de batalla. Por ello, para hacer frente a tales vicisitudes, los sistemas de simulación del campo de batalla deben implementar un mecanismo de validación de la valoración de los materiales y sistemas de armas que permita la inclusión de los últimos avances en armamento y material, asignándoles puntuaciones según criterios preferiblemente objetivos de forma que no se desvirtúen los resultados de los ejercicios simulados.

La existencia de un sistema de puntuaciones establecido, en versiones anteriores del sistema de simulación, según criterios no documentados obligan a establecer un modelo teórico “universal” para asignar valores a los nuevos elementos del modelo, respetando la necesaria coherencia con los datos existentes. Tras un primer análisis exploratorio se confirma la imposibilidad de establecer un modelo de relaciones lineales mediante técnicas de regresión. Las ecuaciones

estructurales permiten relajar los supuestos de los modelos de regresión, al tiempo que favorecen la introducción de variables latentes que, en este caso, tienen una relación causal mucho más estrecha y comprensible con las características que se desean medir que las variables observables que pueden ser empleadas en un modelo de regresión, permitiendo identificar los efectos directos e indirectos entre las variables que definen el modelo.

2. Bibliografía

- [1] Arbuckle, J. L. & W. Wothke (1999). Amos 4.0 User's Guide. Chicago: SmallWaters Corporation.
- [2] Bollen, (1989). Structural Equations with Latent Variables. New York: John Wiley & sons.
- [3] Ruiz, M.A. (2000). Introducción a los modelos de ecuaciones estructurales. Madrid: Ediciones UNED.
- [4] Saris, W. E. & L. H. Stronkhorst (1984). Causal Modelling in Non-Experimental Research. Amsterdam: Sociometric Research.

3. Contactar

Dirección:

Unidad Central de Investigación Operativa. (UCIO)

Sección de Técnicas de Apoyo a la Decisión.

SUBAT-JCISAT

C/Prim núm 6, 8.

28004 – Madrid

ESPAÑA

Teléfono: (+34) 917802138 (+34) 917802211 (+34) 917802491

Fax: (+34) 917802596

Aplicación de Inteligencia Artificial en un Simulador Táctico de Combate para Adiestramiento (STCA)

F. Bargueño¹, A. Castellote¹, M. Molero¹, O. Muñoz¹, J. Padura¹, A. Planells¹.

¹Unidad Central de Investigación Operativa del Ejército de Tierra.

Abstract

El Ejército Español viene utilizando de forma creciente los medios y sistemas de simulación para paliar, en la medida de lo posible, los incrementos en los costes de la preparación que traen consigo unas Fuerzas Armadas modernas, preservar los materiales más costosos, acercar la enseñanza a los usuarios y no perder el tren de la tecnología.

Actualmente el Ejército de Tierra dispone en servicio o en investigación y desarrollo de Simuladores de Adiestramiento desde nivel Grupo Táctico (GT) a División.

Palabras Clave: Simulación, Inteligencia Artificial, Sistemas Expertos.

AMS: 68U20, 68T99

1. Introducción

Un Ejército que no se encuentre comprometido en una guerra, deberá estar entrenándose para la guerra y todo entrenamiento fuera del combate es una simulación. La tecnología empleada en las simulaciones de combate ha mejorado notoriamente en las dos décadas pasadas. Antes del año 1990, los ejercicios de entrenamiento se realizaban en tablas de combate que cubrían el suelo y las acciones de combate se resolvían echando los dados. Con las simulaciones asistidas por computadora, se aceleró la velocidad del combate y se intensificó la naturaleza impredecible de sus resultados; sin embargo, la simulación del realismo del campo de batalla aún depende de las aptitudes del observador/controlador encargado del desarrollo del ejercicio.

Los sistemas de simulación han demostrado ser una efectiva forma de mejorar la coordinación, comunicaciones y procesos de toma de decisiones, antes de ejecutar las planificaciones en el terreno con unidades militares reales. Estos sistemas de entrenamiento son conocidos en el contexto internacional como *sistemas constructivos*, en donde modelos estocásticos implementados en el computador permiten que dos voluntades representadas por unidades interactúen en un terreno, clima y armamento virtual. La simulación constructiva tiene por objetivo el entrenamiento, permitiendo que los participantes conduzcan sus unidades y se enfrenten al adversario en el computador, accionando sólo el teclado y ratón.

Los sistemas constructivos actuales mantienen una alta fidelidad computacional de lo que podría suceder en el campo táctico, sin embargo, estos sistemas no mantienen herramientas que permitan entre otros factores evaluar la conducta en tiempo real de los usuarios, o bien permitir que uno de los bandos (la computadora) resuelva una acción táctica, fundamentada en la misión, terreno, clima, enemigo, etc., entre muchas otras variables del sistema computacional.

2. Simulador Táctico de Combate para Adiestramiento (STCA)

Presentamos un prototipo de Simulador Táctico de Adiestramiento (STAC), de nivel básico, en tiempo real (RTS), que sobre una representación cartográfica digital en dos dimensiones (2D), permite que en un entorno de red multiusuario interactúen dos bandos, pudiendo ser manual-manual o bien manual-computadora, en el desarrollo de una acción táctica. Dicho simulador incorpora herramientas de Inteligencia Artificial para:

- La inteligencia del contrincante y habilidad para desarrollar estrategias.
- El terreno debe añadirse a la inteligencia en el desarrollo de estrategias.
- Aprendizaje por la computadora, en base a *lecciones aprendidas*.
- Módulos que permitan la evaluación de las decisiones tomadas (juicio crítico).

3. Contactar

Dirección:

Unidad Central de Investigación Operativa. (UCIO)
 Sección de Técnicas de Apoyo a la Decisión.
 SUBAT - JCISAT
 c/. Prim núm 6, 8.
 28004 - Madrid
 ESPAÑA

Teléfono: (+34) 917802138 (+34) 917802211 (+34) 917802491

Fax: (+34) 917802596

Aplicaciones de la Teoría de Juegos a sistemas de comunicaciones multislot

Javier Gozámez¹, Natividad Llorca², M^a Carmen Lucas³, Manuel A. Pulido⁴, Joaquín Sánchez-Soriano⁵

¹j.gozalvez@umh.es, Área de Teoría de la Señal y Comunicaciones,
Universidad Miguel Hernández de Elche

²nllorca@umh.es, CIO y Departamento de Estadística, Matemáticas e
Informática, Universidad Miguel Hernández de Elche

³maria.lucas@alu.umh.es, Área de Teoría de la Señal y Comunicaciones,
Universidad Miguel Hernández de Elche

⁴manpul@umh.es, CIO y Departamento de Estadística, Matemáticas e
Informática, Universidad Miguel Hernández de Elche

⁵joaquin@umh.es, CIO y Departamento de Estadística, Matemáticas e
Informática, Universidad Miguel Hernández de Elche

Abstract

En este trabajo abordamos el estudio de la gestión eficiente de los recursos en los sistemas de comunicaciones basados en las nuevas tecnologías 2.5/3G. En particular se analiza, desde el punto de vista de la Teoría de Juegos, el diseño de protocolos de asignación dinámica de *slots* a los distintos usuarios que requieren cantidades diferentes de recurso en función del servicio que demandan.

Palabras Clave: Teoría de juegos, comunicaciones móviles.

AMS: 91A80

1. Agradecimientos

Este trabajo ha sido parcialmente subvencionado por la Generalitat Valenciana (GV05/189) y por el Ministerio de Educación y Ciencia (MTM2005-09184-C02-02 y TEC2005-08211-C02-02) con aportación de los Fondos Europeos FEDER.

Asignación de Turnos de Riego. Parte II: Modelo cuadrático separable con variables 0-1 para la asignación de turnos de riego

M. Almiñana, L.F. Escudero, M. Landete, J.F. Monge, A. Rabasa, J. Sánchez-Soriano

Centro de Investigación Operativa.

Universidad Miguel Hernández de Elche, España

{marc, escudero, landete, monge, a.rabasa,joaquin } @umh.es

Abstract

En este trabajo se presenta un modelo cuadrático separable con variables 0-1 para la planificación y optimización de los recursos hídricos de regadío en agricultura. Se resuelve el modelo cuadrático de forma iterativa mediante la resolución de la aproximación lineal con variables 0-1 al problema cuadrático.

Claves: Planificación de recursos hídricos, programación cuadrática 0-1.

1. Introducción

En este trabajo se presenta un modelo de programación 0-1 mixto cuadrático para resolver la asignación de turnos de riego en una red de regadío. Se obtiene una planificación diaria de los turnos de riego que satisface la demanda de agua solicitada por cada regante y las restricciones técnicas de la red, presión y velocidad en las canalizaciones. Dado el gran número de parcelas de regadío (hidrantes), el problema puede tener gran número de variables (0-1 y continuas), pero con una estructura especial que posibilita aplicar técnicas de descomposición y aproximación lineal del problema cuadrático.

2. Descripción del Problema

El problema consiste en regar cierta área donde el agua fluye desde un embalse hasta cada hidrante a través de una estructura arborescente de tuberías. Los hidrantes están agrupados en sectores. La presión en cada boca de riego (hidrante) es una función de la presión en la cabecera de sector, más la diferencia de altura del hidrante con la cabecera de sector, menos la pérdida de carga que se produce por el rozamiento del agua en la tubería. El conjunto de hidrantes a los cuales se les proporciona agua en un periodo determinado deben satisfacer la demanda de agua solicitada y las restricciones técnicas, presión mínima en cada hidrante y velocidad en cada segmento de tubería. El objetivo del problema consiste en seleccionar los hidrantes que dispondrán de riego en cada periodo de forma que se optimice la función de prioridades de asignación de turnos de riego y se satisfagan las restricciones técnicas y logísticas.

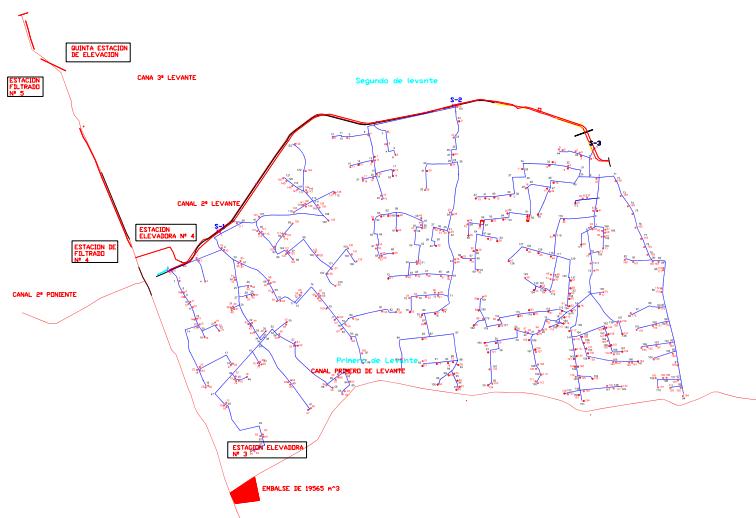


Figura 1: Topología del caso

3. Experiencia Computacional

El modelo presentado en este trabajo ha sido aplicado en La Comunidad de Regantes, Riegos de Levante, Canal Segundo, Elche (Alicante). El área de regadío está compuesta por 2188 Has distribuidas en 20 sectores, con un total de 2025 hidrantes. Es necesaria una planificación diaria en 5 intervalos de tiempo de 4 horas cada uno, reservando un periodo de 4 horas para tareas de mantenimiento en la red. El agua fluye desde un embalse de capacidad 13 Hm³, a través de una red de tuberías con estructura arbórea, ver figura 1.

4. Bibliografía

- [1] J. Andreu, J. Capilla and E. Sanchis. AQUATOOL, a generalized decision-support system for water-resources planning and operational management. *Journal of Hydrology* 177 (1996) 269-291.
- [2] J.W. Labadie, I.E. Brazil, I. Corbu and L.F. Johnson, editors. *Computerized Decision Support Systems for Water Managers*. American Society of Civil Engineers, 1989.

Asignación de Turnos de Riego. Parte I: Descripción de la aplicación y Generación de Prioridades

M. Almiñana, L.F. Escudero, M. Landete, J.F. Monge, A. Rabasa, J. Sánchez Soriano

Centro de Investigación Operativa. Universidad Miguel Hernández de Elche
{marc, escudero, landete, monge, a.rabasa, joaquin}@umh.es

Abstract

En este trabajo se presenta una aplicación informática para la planificación y asignación de turnos de riego sobre una red para uso agrícola con estructura arborescente. La asignación final de un determinado turno a un cierto hidrante se realiza atendiendo tanto a las limitaciones físicas de la red, como a una escala de prioridades basada en el aprovechamiento de asignaciones anteriores.¹

Palabras Clave: Planificación, recursos hídricos.

1. Introducción

El problema de abastecimiento de agua para uso agrícola en el Sureste español está motivando la modernización de los sistemas de planificación y control de recursos hídricos. La Comunidad General de Regantes, Riegos de Levante, Elche (Alicante), está implantando un sistema de monitorización en tiempo real y telecontrol de la red de abastecimiento de su canal denominado Segundo de Levante.

Uno de los objetivos de dicho proyecto es la planificación de los turnos diarios de riego de 2025 hidrantes que se distribuyen en una red de estructura arborescente, a lo largo de una superficie regable de 2188 Has. Existen 5 turnos diarios de riego, de 4 horas cada uno y cada hidrante puede demandar entre 1 y 4 de estos turnos en función del número de Has. que abarca. No obstante, la concesión de las solicitudes, y en definitiva la asignación de turnos se realiza teniendo en cuenta tanto las limitaciones físicas de la red (caudales y presiones) como las prioridades de las demandas en base a criterios tales como el aprovechamiento de demandas anteriores y atención equilibrada a todos sus abonados.

2. Aplicación informática

La aplicación recibe como entradas los siguientes ficheros: la topología de la red; las solicitudes de turnos de cada hidrante y los datos de telemetría (caudal acumulado y presión media) recogidos por un sistema SCADA en tiempo real, con una frecuencia de media hora.

¹Investigación parcialmente subvencionada por los proyectos MTM2004-01095 del Ministerio de Educación y Ciencia, y GV04B/655 de la Generalitat Valenciana.

Los módulos principales del programa son: Topología (lectura de la estructura de la red), Telemetría (lectura, preprocesado [1] y estadísticas sobre los datos recogidos por el sistema SCADA), Asignación de Prioridades (con posibilidad de ajustar los pesos de cada uno de los factores que se consideran) y Asignación de Turnos (donde se tienen presente las limitaciones físicas de la red y las prioridades calculadas en el módulo anterior).

La aplicación funciona bajo dos modos posibles: el modo “Automático”, donde dichos módulos son llamados secuencialmente sin intervención del usuario, hasta proporcionar el fichero de Asignaciones de Turnos; y el modo “Paso-a-paso”, donde el usuario puede interactuar con el sistema ajustando parámetros y analizando resultados parciales.

El código está escrito en C++, y la herramienta elegida fue C++ Builder de Borland, por su potencia en el desarrollo de interfaces gráficas de usuario.

3. Asignación de prioridades

Cuando un hidrante solicita un turno de riego para un determinado día, el programa debe asignarle una cierta prioridad (en función de su comportamiento anterior [2]) que será tenida en cuenta por el módulo de Asignación de turnos. El cálculo de la prioridad de un hidrante para regar en un turno responde a la siguiente ecuación general:

$$P(h, t) = \sum_{i=1}^n F_i * V_{i,h,t}$$

Siendo:

$P(h, t)$: Prioridad de un hidrante h , de regar en un turno t . $P(h, t) \in [0, 1]$

$V_{i,h,t}$: Variable i -ésima que incide en la asignación de la prioridad de que el hidrante h riegue en turno t

F_i : Factor de incidencia de la variable i . $F_i \in [0, 1]$; $\sum_{i=1}^n F_i = 1$

La aplicación permite que el usuario defina n variables de incidencia en la prioridad, V_i , y su factor de incidencia, F_i . La matriz de prioridades así calculada se emplea en el módulo final de Asignación de Turnos.

4. Bibliografía

- [1] Chen, Z. (2001). Data Mining and Uncertaining Reasoning. An Integrated Approach. Wiley Interscience.
- [2] Solomatine D.P. (2004). Innovative Approaches to Flood Forecasting Using Data Driven and Hybrid Modelling. World Scientific Publishing Co.

Assignment situations with multiple ownership and their games

Silvia Miquel¹, Bas van Velzen, Herbert Hamers, Henk Norde

¹Departament de matemàtica. Universitat de Lleida
smiquel@matematica.udl.es

Abstract

Como generalización del modelo de Shapley y Shubik (1972), se introducen las situaciones de asignación con propiedad múltiple: Cada objeto puede ser propiedad de varios jugadores y cada jugador puede participar en la propiedad de más de un objeto. Juegos asociados a estas situaciones son los juegos de asignación simples y sus relajaciones. Se establecen condiciones para su equilibrio.

Keywords: Assignment situations, matchings, balancedness

1. Introduction

This paper analyzes assignment situations in which an object can be owned by several agents and where agents can participate in the ownership of more than one object. Similar to Shapley and Shubik (1972), we define simple assignment games, a class of cooperative games arising from these situations and a non-negative reward matrix. In a simple assignment game, each coalition can only match objects completely owned by this coalition. Moreover, if a player is participating in the ownership of more than one object, at most one of these objects can be matched. In the relaxations of simple assignment games, if a player is participating in more than one object, at most two (three,...,etc.) of these objects can be matched.

The house market is still a nice example of this model. The properties of the sellers can be shopping malls which usually are owned by several agents. Moreover, it can be that some agents are involved in the ownership of several shopping malls. The buyers can be viewed as a group of investors. The possible restrictions on the number of shopping malls that can be sold or bought can be imposed by governmental rules to resist monopolies in city areas of owners.

2. The assignment situation with multiple ownership and its game

An assignment situation $((M_1, M_2), (A_1, A_2), (S_1, S_2))$ with multiple ownership consists of two finite and disjoint agent sets M_1 and M_2 , two finite and disjoint sets of objects A_1 and A_2 and functions $S_i : A_i \rightarrow 2^{M_i}$, $i = 1, 2$, that describe the set of agents that own some object. If an object of A_1 is matched with

an object of A_2 , there can be obtained some reward. A non-negative matrix $R \in \mathbf{R}^{A_1 \times A_2}$ expresses this reward.

The simple assignment game, $(M_1 \cup M_2, v)$, that arises from an assignment situation with multiple ownership, $((M_1, M_2), (A_1, A_2), (S_1, S_2))$, is defined for all $T_1 \subset M_1$, $T_2 \subset M_2$ by

$$v(T_1 \cup T_2) = \max \left\{ \sum_{(a,b) \in \mu} R_{ab} : \mu \in \mathcal{A}(T_1 \cup T_2) \right\},$$

where $\mathcal{A}(T_1 \cup T_2)$ is the set of admissible matchings with respect to coalition $T_1 \cup T_2$. It implies that for each player at most one object is matched where he is participating in.

3. Balanced assignment situations

An assignment situation with multiple ownership is called *balanced* if for any reward matrix R the corresponding simple assignment game is balanced.

Proposition An assingment situation with multiple ownership is balanced if and only if it satisfies the partitioning property or the intersection property.

4. k -Assignment games with multiple ownership

In this section we discuss k -assignment games with multiple ownership, which are relaxations of simple assignment games.

Proposition A simple assingment game is balanced if and only if the solution of the ILP-problem of the gran coalition coincides with the solution of its LP-relaxation.

Proposition Let $(M_1 \cup M_2, v_k)$ be a k -assignment game with multiple ownership. If the optimum of the ILP-problem of the grand coalition coincides with the optimum of its LP-relaxation, then v_k is balanced.

5. Bibliography

- [1] Deng, X., Ibaraki,T. and Nagamochi, H. (1999). Algorithmic aspects of the core of combinatorial optimization games. *Mathematics of Operations Research* 24, 751-766
- [2] Shapley, L.S. and Shubik, M. (1971). The assignment game I: The core. *International Journal of Game Theory* 1, 111-130.
- [3] Quint, T. (1991). Characterization of cores of assignment games. *International Journal of Game Theory* 19, 413-420.

Búsqueda de óptimos globales en problemas de localización discreta mediante un algoritmo heurístico

P. Fernández¹, B. Pelegrín¹, J.L. Redondo², I. García², P.M. Ortigosa²

¹pfdez,pelegrin@um.es, Dpto. Estadística e Investigación Operativa,
Universidad de Murcia

²juani,inma,pilar@ace.ual.es, Dpto. de Arquitectura de Computadores y
Electrónica, Univ. de Almería

Abstract

Muchos problemas de localización discreta son formulados como problemas de P.L.E. y resueltos mediante optimizadores estándar que proporcionan un óptimo global, aún en el caso de existir más. Proponemos un algoritmo genético, GASUB, que es capaz de obtener un número prefijado de óptimos globales, si los hay. Este algoritmo se compara con MSH, un conocido método heurístico de sustitución multi-start. Los resultados son validados resolviendo los problemas objeto de estudio mediante Xpress-MP.

Palabras Clave: Optimos globales, algoritmo genético, heurísticos

1. Introducción

Existe una gran cantidad de problemas de localización discreta, en los cuales una firma tiene que seleccionar un número dado s de localizaciones para sus nuevos centros en un conjunto que contiene m posibles localizaciones. Algunos de estos problemas pueden ser formulados como modelos de P.L.E., y por tanto son resueltos utilizando alguno de los optimizadores disponibles en el mercado (Xpress-MP, CPLEX, LINDO, ...). Sin embargo, este tipo de software sólo genera un óptimo global. Para muchos de estos problemas suelen existir varios óptimos globales, por lo que sería deseable disponer del mayor número de ellos para poder tener otros criterios alternativos en la decisión final. Por otra parte, cuando hay un gran número de posibles localizaciones y de puntos de demanda, dichos optimizadores no pueden gestionar el tamaño de las matrices que definen su correspondiente formulación como problemas de P.L.E., y por tanto no son capaces de generar una solución.

En este trabajo se presenta un nuevo algoritmo, GASUB, que es capaz de generar un número predeterminado de óptimos globales, siempre que existan. Dicho algoritmo es de tipo genético multimodal, lo cual significa que invierte la mayor parte del tiempo en descubrir múltiples y diferentes zonas de búsqueda, más que en explorar una de ellas múltiples veces.

Para comprobar que dicho algoritmo es capaz de obtener óptimos globales, hemos seleccionado varias instancias de tres problemas de localización discreta:

s-MEDIANA [1], elegido entre los que estudian modelos sin competencia, y cuyo objetivo es minimizar el coste total de transporte entre los puntos de demanda y los centros de servicio, de manera que para satisfacer la demanda, cada cliente debe ser servido por el centro más cercano; MAXCAP [3], problema de localización competitiva cuyo objetivo es la maximización de la cuota de mercado, ofreciendo el mismo precio en todos los centros y siendo los clientes los que se desplazan a comprar el producto; y MAXPROFIT [2], también de localización competitiva, pero cuyo objetivo es la maximización del beneficio total, ofreciéndose un precio para cada área de mercado, el cual incluye el precio del producto y el coste de transporte.

Para los estudios computacionales, hemos considerado tres conjuntos de puntos contenido 1046, 1273 y 1671 ciudades de España como puntos de demanda, siendo su población la demanda de cada uno. Los datos de coordenadas GPS y de población de las ciudades utilizadas han sido obtenidas de las páginas <http://www.terra.es/personal/GPS.2000> y <http://www.ine.es> utilizando entre puntos la distancia Euclídea, considerando a todos como puntos posibles para nuevas localizaciones, y ejecutándolos bajo Linux en un Pentium IV con 3GHz y 2GB de memoria.

Todos los problemas tests, obtenidos de la combinación de los diferentes parámetros, han sido resueltos por GASUB y sus resultados comparados por los obtenidos por MSH, un procedimiento heurístico ampliamente utilizado en problemas de localización combinatoria, y para verificar si los resultados obtenidos por ambos heurísticos son efectivamente óptimos globales, todos los problemas tests han sido óptimamente resueltos por *Xpress-MP*.

Los resultados muestran que GASUB es capaz de encontrar el 100 % de los casos al menos un óptimo global, mientras que MSH en algunos tests no llega a encontrar un óptimo global.

2. Agradecimientos

Este trabajo ha sido financiado por el Ministerio de Ciencia y Tecnología de España bajo el proyecto SEJ2005-06273/ECON.

3. Bibliografía

- [1] Daskin, M.S. (1995). *Network and discrete location: models, algorithms and applications*. Wiley.
- [2] Fernández, P., Pelegrín, B., García, M.D. and Peeters, P. (2005). A discrete long-term location-price problem under the assumption of discriminatory pricing: formulations and parametric analysis. *EJOR*, (aceptado).
- [3] Serra, D. and ReVelle, C. (1995). Competitive location in discrete space. In: Z. Drezner (ed), Facility location: a survey of applications and methods, Springer, 367-386.

Cálculo de valores extremos en teoría de colas

J.R. Artalejo

jesus_artalejo@mat.ucm.es, Departamento de Estadística e Investigación Operativa , Universidad Complutense de Madrid

Abstract

En este trabajo se estudia la distribución del máximo nivel visitado durante un período de ocupación de una cadena de Markov bidimensional definida en tiempo continuo. Presentamos un análisis de naturaleza algorítmica que es válido para cadenas estructuradas por bloques.

Claves: máximo nivel visitado, cadenas de Markov

AMS: 90B22, 60K25

1. Introducción

Las cadenas de Markov en tiempo continuo proporcionan el marco matemático apropiado para modelizar estocásticamente muchos modelos de colas, inventarios, sistemas de producción, etc. En este contexto es típico asumir las siguientes hipótesis: (i) el régimen estacionario, (ii) la homogenidad del sistema, y (iii) estructuras particulares (QBD, M/G/1 o GI/M/1) para los generadores infinitesimales. En ausencia de estas hipótesis el cálculo de la distribución estacionaria del estado del sistema resulta extremadamente complicado e, incluso, intratable.

La consideración de la distribución del máximo nivel visitado (o, con más generalidad, de valores extremos) proporciona un descriptor alternativo que es fácilmente computable en ausencia de las hipótesis anteriormente indicadas. Nuestros resultados pueden ser aplicados a una gran variedad de modelos de colas [1], ‘call centers’ [2], etc.

2. El modelo matemático

Consideramos la cadena (irreducible y regular) $\{(N(t), C(t)); t \geq 0\}$ con espacio de estados $S = \cup\{(i, j) : i \geq 0, 0 \leq j \leq L_i\}$. Dado el estado (i, j) , la primera coordenada i se denomina el nivel del estado. El estado $(0, 0)$ será denotado como 0^* . El resto de los estados son particionados en vectores:

$$\begin{aligned}\underline{0} &= \{(0, j) : 1 \leq j \leq L_0\}, \\ \underline{i} &= \{(i, j) : 0 \leq j \leq L_i\}, \quad i \geq 1.\end{aligned}$$

El número de estados de cada nivel, L_i , es finito. El generador infinitesimal

$Q = [q_{(i,j)(m,n)}]$ ($q_{(i,j)} = -q_{(i,j)(i,j)}$), viene dado por

$$Q = \begin{pmatrix} q_{0^*} & A_{0^*,0} & A_{0^*,1} & A_{0^*,2} & \cdots \\ A_{0,0^*} & A_{0,0} & A_{0,1} & A_{0,2} & \cdots \\ A_{1,0^*} & A_{1,0} & A_{1,1} & A_{1,2} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

3. Análisis del máximo nivel visitado durante un periodo de ocupación

Definimos el período de ocupación de la cadena descrita en la Sección 2 como el intervalo de tiempo que comienza cuando el proceso abandona el estado 0^* y finaliza en el primer instante en el que el proceso visita 0^* nuevamente. También definimos N_{\max} como el máximo nivel visitado por la cadena durante el período de ocupación. Observamos que

$$P(N_{\max} = k) = \sum_{(m,n) \in S_0} \frac{q_{0^*,(m,n)}}{q_{0^*}} P_{(m,n)}(N_{\max} = k), \quad k \geq 0,$$

siendo $S_0 = \{(m, n) : q_{0^*,(m,n)} > 0\}$, es decir, S_0 representa el subconjunto de estados que el proceso puede visitar justo al comenzar el período de ocupación.

La probabilidad $P_{(m,n)}(N_{\max} < k)$ corresponde con la probabilidad del evento consistente en que la cadena visite el estado 0^* antes de alcanzar el nivel \underline{k} , $k \geq 1$, dado que el estado inicial es (m, n) . Por ello, el cálculo de $P(N_{\max} < k)$, $k \geq 1$, se reduce a encontrar la probabilidad de absorción en una cadena finita auxiliar con estados absorbentes 0^* y k^* (estado que engloba las transiciones infinitesimales a los niveles i , $i \geq k$).

4. Bibliografía

- [1] Artalejo, J.R. and Chakravarthy, S.R. (2006). Algorithmic analysis of the maximum level length in general-block two-dimensional Markov processes, aparecerá en *Mathematical Problems in Engineering*.
- [2] Artalejo, J.R., Economou, A. and Gomez-Corral, A. (2006). Applications of maximum queue length to call centers management, aparecerá en *Computers and Operations Research*.

Coalition formation and partnerships in the Catalonia Parliament

Francesc Carreras¹, Maria Dolores Llongueras², Maria Albina Puente³

¹francesc.carreras@upc.edu, Department of Applied Mathematics II and Industrial Engineering School of Terrassa, Technical University of Catalonia

²m.dolors.llongueras@upc.edu, Department of Applied Mathematics II and Industrial Engineering School of Terrassa, Technical University of Catalonia

³m.albina.puente@upc.edu, Department of Applied Mathematics III and School of Engineering of Manresa, Technical University of Catalonia

Abstract

Se aplican los semivalores binomiales al estudio de la formación de coaliciones en el Parlamento de Cataluña y la influencia, sobre el resultado final, de un consorcio previo entre determinados partidos.

Keywords: coalition formation, partnership, binomial semivalue.

AMS: 91A12

1. Introduction

We study here the coalition formation in the current Catalonia Parliament, where no party is holding an absolute majority, and discuss whether the existence of a previous partnership may influence the issue. Pure and coalitional binomial semivalues are used, to this end, as power measures.

2. Simple games, partnerships and binomial semivalues

A cooperative game v in N is *simple* if it is monotonic, $v(T) = 0$ or 1 for all T , and $v(N) = 1$. A simple game v is a *weighted majority game* if there are nonnegative *weights* w_1, w_2, \dots, w_n allocated to the players and a positive *quota* q such that $v(S) = 1$ iff $\sum_{i \in S} w_i \geq q$. We then write $v = [q; w_1, w_2, \dots, w_n]$.

Definition 2.1 (Kalai and Samet [2]) A coalition $S \subseteq N$ is a *partnership* in a game v if $v(R \cup T) = v(R)$ for all $T \subset S$ and all $R \subseteq N \setminus S$.

In the case where only one coalition S turns into a partnership, the partnership formation is stated as follows (Carreras [1]). Let v a game in N and $\emptyset \neq S \subseteq N$. The *partnership game* v^S in N (where S clearly becomes a partnership) is defined by $v^S(T) = v(T)$ if $S \subseteq T$ and $v^S(T) = v(T \setminus S)$ otherwise.

Definition 2.2 (Puente [4]) Let $p \in [0, 1]$ and take $p_k = p^k(1 - p)^{n-k-1}$ for $k = 0, \dots, n - 1$, with $0^0 = 1$ in the extreme cases ($p = 0$ and $p = 1$). Then

$\{p_k\}_{k=0}^{n-1}$ is a weighting vector and defines a semivalue, which will be denoted as ψ^p and called the *p-binomial semivalue*. Then the *p-binomial semivalue* of player i in game v is $\psi_i^p[v] = \sum_{T \subseteq N: i \in T} p^{t-1}(1-p)^{n-t}[v(T) - v(T \setminus \{i\})]$, where $t = |T|$. Of course, $\psi^{1/2} = \beta$, the Banzhaf value.

Definition 2.3 (Carreras and Puente [2]) Let $p \in [0, 1]$. The *symmetric coalitional p-binomial semivalue* of player i in a game with a coalition structure $[v; P]$ is defined by

$$\Omega_i^p[v; P] = \sum_{R \subseteq M \setminus \{k\}} \sum_{T \subseteq P_k \setminus \{i\}} p^r(1-p)^{m-r-1} \frac{1}{p_k \binom{p_k-1}{t}} [v(Q \cup T \cup \{i\}) - v(Q \cup T)]$$

where M is the quotient set, $P_k \in P$ is the union such that $i \in P_k$, $r = |R|$, $p_k = |P_k|$, $Q = \bigcup_{r \in R} P_r$ and, again, $0^0 = 1$ when necessary.

3. The Catalonia Parliament, Legislature 2003-2007

The weighted majority game [68; 46, 42, 23, 15, 9] describes the current Catalonia Parliament. The politically likely coalitions are $\{1, 3\}$ and $\{2, 3, 5\}$. Parameter p defines the tendency of the parties as to coalition formation. Depending on the value of p , party 3 would prefer entering one coalition or the other. However, a previous partnership $\{2, 5\}$ could have been formed. We discuss, in terms of p , the convenience of parties 2 and 5 to do so in order to force party 3 to join them instead of party 1.

4. Acknowledgement

Research partially supported by Grant BFM 2003-01314 of the Science and Technology Spanish Ministry and the European Regional Development Fund.

5. Bibliography

- [1] Carreras, F. [1996]: On the existence and formation of partnerships in a game. *Games and Economic Behavior* 12, 54–67.
- [2] Carreras, F. and Puente, M.A. [2006]: A parametric family of mixed coalitional values. In: Recent Advances in Optimization (A. Seeger, ed.), Lecture Notes in Economics and Mathematical Systems, Springer–Verlag, 323–339.
- [3] Kalai, E. and Samet, D. [1987]: On weighted Shapley values. *International Journal of Game Theory* 16, 205–222.
- [4] Puente, M.A. [2000]: Contributions to the representability of simple games and to the calculus of solutions for this class of games (in Spanish). Ph.D. Thesis. Technical University of Catalonia, Spain.

Coloreando grafos con redes neuronales

Javier Yáñez¹, Pedro Martínez Talaván²

¹jayage@mat.ucm.es, Departamento de Estadística e Investigación Operativa
I, Universidad Complutense de Madrid

²ptalavan@ine.es, Instituto Nacional de Estadística, Madrid

Abstract

El modelo de Hopfield de redes neuronales se pueden resolver determinados problemas de optimización combinatoria. Por medio de una función de energía que generaliza la utilizada por Hopfield para resolver el problema del viajante, se pueden resolver otros problemas de optimización combinatoria, en concreto se utilizará para resolver el problema de coloración.

Palabras Clave: Redes neuronales, modelo de Hopfield, Problema de coloración de grafos

1. Introducción

El modelo continuo de Hopfield (CHN) fue introducido por Hopfield y Tank [1] para resolver el problema del viajante. Sin embargo, hasta el trabajo de Talaván y Yáñez [2], no se puede hablar de un procedimiento sistemático de parametrización que garantice la factibilidad de la solución obtenida. Por otra parte, una nueva función de energía ha sido propuesta por Talaván y Yáñez [3] con el fin de que pueda adaptarse a otros problemas de optimización combinatoria. Los buenos resultados computacionales obtenidos al resolver el problema del viajante con el algoritmo de Yáñez y Talaván [4] han animado a los autores a abordar otros problemas de optimización diferentes. En este sentido se ha optado por el *Problema de coloración de grafos*, conocido por ser uno de los más difíciles y estimulantes para los investigadores operativos.

2. El problema de coloración como un problema de programación 0-1

Sea un grafo $G = (V, E)$, siendo $V = \{1, 2, \dots, n\}$ el conjunto de vértices y E el conjunto de aristas o pares no ordenados de vértices. El problema de coloración de grafos busca el menor entero $c \in \{1, 1, \dots, n\}$ para el que existe una asignación de colores $\text{col} : V \longrightarrow \{1, 2, \dots, n\}$ verificando para cualquier arista $e = \{x, y\} \in E$ que $\text{col}(x) \neq \text{col}(y)$

Este problema se puede plantear como un problema de optimización cuadrática generalizado considerando como variable de decisión binaria $v_{i,k} = 1$ si y sólo

si $col(i) = k$:

$$\text{Min}_v \left\{ \sum_{k=1}^n (n/2 + 2k/n)v_{0,k} \right\}$$

sujeto a:

$$\sum_{k=1}^n (1 - v_{0,k}) \sum_{i=1}^n v_{i,k} = 0; \quad \sum_{k=1}^n \sum_{i=1}^n \sum_{j=1}^n v_{i,k} b_{i,j} v_{j,k}; \quad \sum_{k=1}^n v_{i,k} = 1 \quad \forall i = 1, \dots, n$$

siendo la matriz $B = (b_{i,j})$ la matriz de adyacencia del grafo.

3. Proyectando el problema en el modelo continuo de Hopfield

La red neuronal tiene $n \times (n + 1)$ neuronas cada una de ellas asociada a la variable $v_{i,k}$ con $i = 0, 1, \dots, n$ $k = 1, \dots, n$. Los pesos de las conexiones, a diferencia de otros modelos de redes neuronales que modifican su valor en la fase de aprendizaje, hay que fijarlos para conseguir que la ecuación diferencial asociada al modelo de Hopfield converja a una solución factible y con una función objetivo aceptablemente buena cuando alcanza el mínimo local de la función de energía.

Las conexiones entre neuronas y la entrada exógena a cada neurona son del tipo:

$$\begin{aligned} T_{i,k;j,l} &= -(1 - \delta_{i,j})\delta_{k,l}b_{i,j}\rho - \delta_{i,j}\phi + 2\delta_{i,j}\delta_{k,l}\gamma \\ i_{i,k} &= -\alpha(n/2 + 2k/n) - \gamma \end{aligned}$$

siendo ρ, ϕ, γ parámetros que se pueden determinar con el fin de garantizar una solución apropiada.

4. Bibliografía

- [1] Hopfield, J.J., Tank, D.W. (1985). Neural computation of decisions in optimization problems. *Biological Cybernetics* **52**, 1-25.
- [2] M.Talaván, P., Yáñez, J. (2002). Parameter Setting of the Hopfield Network applied to TSP. *Neural Network* **15** (3), 363-373.
- [3] M.Talaván, P., Yáñez, J. (2005). A Continuous Hopfield Network Equilibrium Points Algorithm. *Computers and Operations Research* **32**(8), 2179-2196.
- [4] M.Talaván, P., Yáñez, J. (2006). The Generalized Quadratic Knapsack Problem. A Neuronal Network Approach. *Neural Network* (*aceptado*).

Competencia espacial duopolística con demanda elástica y política de precios en destino

P. Dorta González¹, D.R. Santos Peñate¹, B. Pelegrín Pelegrín²,
R. Suárez Vega¹

¹[pdorta,drsantos,rsuarez]@dmc.ulpgc.es, Departamento de Métodos

Cuantitativos en E. y G., Universidad de Las Palmas de Gran Canaria

²pelegrin@um.es, Departamento de Estadística e Investigación Operativa,
Universidad de Murcia

Abstract

Se estudia un duopolio de Bertrand espacial donde dos firmas compiten en precios por suministrar un bien no esencial en n mercados separados espacialmente y localizados en los vértices de una red. La demanda de cada mercado es elástica al precio y se considera una política de precios en destino.

Claves: Localización competitiva, precio en destino.

AMS: 90B10

1. Introducción

Este problema es modelado como un juego en dos etapas, en la primera de las cuales las firmas deciden sus localizaciones en la red, y en la segunda, conocidas éstas, deciden simultáneamente los precios en destino. Para localizaciones fijadas en la primera etapa del juego, se prueba la existencia de un único equilibrio de Nash en precios para el escenario competitivo y se indica el precio de mercado para el escenario colusivo. Asumiendo que las firmas elegirán una de las políticas de precios anteriores, se prueba la optimalidad nodal de las localizaciones de equilibrio de Nash.

2. Modelo

Sea $N = (V, E)$ una red conexa con un conjunto finito de vértices $V = \{v_1, \dots, v_n\}$ y un conjunto de aristas E . En cada vértice v_k se localiza un mercado donde se vende un bien homogéneo al precio unitario p^k . Este precio es una función lineal decreciente de la cantidad total q^k ofertada en v_k , $p^k(q^k) = a_k - b_k q^k$, $0 \leq q^k \leq \frac{a_k}{b_k}$, con $a_k, b_k > 0$.

Supuesto 1 El coste de transporte unitario de la firma i , $i = A, B$, entre la localización x y el mercado k se denota por $t_i^k(x)$ y es positivo, cóncavo y creciente con la distancia.

Supuesto 2 El coste de producción marginal no depende de la cantidad producida y se denota por $C_i(x)$, $i = A, B$. Además, es una función positiva y cóncava cuando x se desplaza a lo largo de cualquier arista de la red.

El coste marginal en destino de la firma i en el mercado k es $c_i^k(x) = C_i(x) + t_i^k(x)$ y, bajo los supuestos 1 y 2, es una función cóncava.

Supuesto 3 (*Regla de elección*) i. Los clientes compran a la firma que ofrece el precio más bajo. ii. Los empates en precios se rompen en favor de la firma más eficiente (menor c_i^k). iii. Si $c_i^k = c_j^k$ la demanda de k se reparte en partes iguales.

Supuesto 4 La disposición a pagar de los consumidores cumple $a_k > c_i^k(x)$, $\forall x \in N$, $k = 1, \dots, n$, $i = A, B$.

El precio ofrecido por la firma i en el mercado k es p_i^k y $\mathbf{p}_i = (p_i^1, \dots, p_i^n)$.

3. Problema de precios de la segunda etapa

Sean x_A y x_B un par de localizaciones fijas para las empresas.

Proposición 1 Bajo los supuestos 2, 3, dadas dos localizaciones tales que $a_k > \max\{c_A^k, c_B^k\}$, $k = 1, \dots, n$, existe un único equilibrio de Nash $(\mathbf{p}_A, \mathbf{p}_B)$. Además, si $c_i^k \leq c_j^k$, entonces $(p_i^k, p_j^k) = \begin{cases} \left(\frac{1}{2}(a_k + c_i^k), c_j^k\right) & \text{si } a_k < 2c_j^k - c_i^k \\ (c_j^k, c_j^k) & \text{si } a_k \geq 2c_j^k - c_i^k \end{cases}$

Este resultado indica precios de monopolio y de punto base. Consideremos ahora un escenario de colusión en precios. Ambas firmas se ponen de acuerdo para maximizar el beneficio conjunto en esta segunda etapa. Cada mercado es cubierto por la firma más eficiente. En caso de igualdad, ambas firmas se reparten en partes iguales la demanda y fijan precio de monopolio.

Proposición 2 Bajo el supuesto 2, dadas dos localizaciones tales que $a_k > \max\{c_A^k, c_B^k\}$, $k = 1, \dots, n$, el precio colusivo en el mercado k que maximiza el beneficio conjunto de las empresas es $p^k = \frac{1}{2}(a_k + \min\{c_A^k, c_B^k\})$.

4. Problema de localización de la primera etapa

Sean π_i y π_i^c las funciones de beneficio para el escenario competitivo y colusivo en precios, respectivamente. Bajo los supuestos 1 al 4, $\pi_i(x_i, x_j)$ y $\pi_i^c(x_i, x_j)$ son funciones cuasiconvexas en x_i cuando x_i se desplaza a lo largo de una arista de la red y x_j no cambia.

Proposición 3 Bajo los supuestos 1 al 4, tanto en el escenario de competencia como en el de colusión en precios, cada firma maximiza su beneficio en algún vértice de la red, independientemente de donde se localice su competidora. Además, si existe algún equilibrio entonces un equilibrio consistente en un par de vértices está asegurado.

Condiciones de optimalidad de orden superior para mínimos estrictos

César Gutiérrez¹, Bienvenido Jiménez², Vicente Novo³

¹cesargv@mat.uva.es, Departamento de Matemática Aplicada, E.T.S.I. Informática, Universidad de Valladolid

²bjimenez@ind.uned.es, Departamento de Matemática Aplicada, E.T.S.I. Industriales, Universidad Nacional de Educación a Distancia

³vnov@ind.uned.es, Departamento de Matemática Aplicada, E.T.S.I. Industriales, Universidad Nacional de Educación a Distancia

Abstract

En este trabajo, se estudia un problema de optimización no regular con restricciones de desigualdad generalizadas y con una restricción de conjunto arbitrario. Se presentan condiciones necesarias para que un punto sea un minimizador local de orden k en términos de las derivadas de Studniarski superior e inferior de orden superior y del cono contingente del conjunto restricción. En la misma línea, cuando el espacio inicial es de dimensión finita, se desarrollan condiciones suficientes de optimalidad. También se desarrollan condiciones de minimizador de orden k usando la derivada de Studniarski inferior de la función lagrangiana del problema. Se presta atención especial a los minimizadores de orden dos, usando en este caso una derivada de segundo orden que se reduce a la derivada Fréchet para funciones diferenciables.

Claves: Condiciones de optimalidad, mínimo estricto de orden superior.

AMS: 90C29, 90C46.

1. Introducción

La noción de minimizador estricto de orden k ha resultado ser muy útil en optimización. Recordemos que dados un espacio vectorial normado X , una función $f : X \rightarrow \mathbb{R}$ y un conjunto $M \subset X$, se dice que el punto $x_0 \in M$ es un minimizador local estricto de orden k ($k \geq 1$ un entero), denotado $x_0 \in \text{Estl}(k, f, M)$, para el problema de optimización

$$\text{Min}\{f(x) : x \in M\}$$

si existen $\alpha > 0$ y un entorno U de x_0 tales que

$$f(x) > f(x_0) + \alpha \|x - x_0\|^k \quad \forall x \in M \cap U \setminus \{x_0\}.$$

Esta noción (sin un nombre específico) fue manejada por Hestenes (1966, 1975) para los valores $k = 1$ y $k = 2$ para demostrar condiciones suficientes de optimalidad.

Cromme (1978) usó esta noción en el contexto de la convergencia de procedimientos numéricos iterativos. Auslender (1984) estudió la minimalidad estricta para desarrollar condiciones de estabilidad en optimización no regular.

Studniarski (1986) extendiendo algunos resultados de Auslender proporcionó condiciones necesarias y suficientes de minimizador estricto de orden k para cualquier función f y un conjunto arbitrario M de un espacio de dimensión finita. Para este fin, usó derivadas direccionales que son generalizaciones de las derivadas Hadamard superior e inferior. Ward (1994) sigue la línea de Studniarski usando otras derivadas y conos tangentes. En (1997) Studniarski considera funciones localmente Lipschitz y la subdiferencial de Clarke. Esta noción (con el nombre de k -único) fue también estudiada en programas semi-infinitos lineales por Goberna, López y Todorov (1995).

Jiménez (2002) extiende la noción de minimizador estricto a problemas de optimización vectorial, esto es, para una función de X en otro espacio normado Y , parcialmente ordenado por un cono convexo. En varios artículos, Jiménez y Novo (2003, 2003a, 2003b, 2004) desarrollan una teoría de minimizadores estrictos de orden k considerando diferentes contextos.

En este trabajo, se discute un problema de programación matemática (escalar) particular, definido por restricciones de desigualdad generalizadas y por una restricción de conjunto arbitrario. Se proporcionan condiciones necesarias y suficientes de optimalidad formuladas a través de las derivadas de Studniarski de modo que no haya salto entre ellas. También, se presentan condiciones suficientes de mínimo estricto de orden k usando la función Lagrangiana del problema. Además, se establece una condición suficiente de segundo orden usando una derivada particular que se convierte en la derivada Fréchet de segundo orden cuando las funciones involucradas son dos veces diferenciables Fréchet. Los resultados obtenidos generalizan los de Studniarski (1986), Ward (1994), Qi (2001) y Huang (2005). Se proporcionan también ejemplos que ilustran los resultados obtenidos.

Construcción de paneles de datos con información de fuentes administrativas a partir de muestreo. aplicación al panel de datos del impuesto sobre la renta de las personas físicas del instituto de estudios fiscales

César Pérez López¹, Santiago Díaz de Sarralde²

¹cesar.perez@ief.minhac.es, Departamento de Estadística e Investigación Operativa III, Universidad Complutense de Madrid e Instituto de Estudios Fiscales

²santiago.diaz@ief.minhac.es, Instituto de Estudios Fiscales y Universidad Rey Juan Carlos de Madrid

Abstract

El presente trabajo tiene por objeto realizar un análisis exhaustivo de la problemática en la construcción de paneles de datos con información procedente de registros administrativos a partir de muestras. Como aplicación práctica se presenta la construcción del panel de datos del Impuesto sobre la Renta de las Personas Físicas del Instituto de Estudios Fiscales 1999-2004.

Palabras Clave: Paneles, Muestreo, Registros administrativos

1. Contenido

El trabajo comienza analizando la problemática de la construcción de un panel de datos a partir de fuentes administrativas.

En la aplicación práctica relativa a la construcción del Panel del Impuesto sobre la Renta de las Personas Físicas del Instituto de Estudios Fiscales 1999-2004, se considera la disponibilidad de marcos muestrales extensos y precisos como base para la elaboración de un panel de calidad.

Se elegirá el año base del panel (1993) con criterios precisos y se tendrá en cuenta un diseño de muestreo adecuado para el año base con selección estratificada sin reposición.

Asimismo se cuantificarán los tamaños muestrales necesarios para que los errores de las estimaciones habituales se acoten adecuadamente.

Se tendrá presente el tipo de afijación de mínima varianza para minimizar el error de muestreo, factor importante para la extensión del panel a los años anteriores y posteriores al año base.

Se analizan las implicaciones de considerar un panel extendido en lugar de un panel puro persiguiendo suavizar la *attrition* (desgaste) del panel.

Finalmente se hace hincapié en la importancia práctica de este panel para tomar decisiones de política fiscal favorecidas por la posibilidad del análisis de la evolución a lo largo del tiempo de variables como la presión fiscal, los índices de progresividad y redistribución y el análisis de los efectos de las reformas impositivas.

Coordination in discrete review inventory situations

Luis A. Guardiola

luis.guardiola@alu.umh.es, Departamento de Estadística, Matemáticas e Informática, Universidad Miguel Hernández de Elche

Abstract

En este trabajo introducimos una nueva clase de juegos cooperativos que surgen de problemas de pedidos e inventario, donde varios agentes tienen que cubrir sus demandas en un horizonte finito de tiempo, permitiéndose períodos de escasez. Para esta clase de juegos, estudiaremos algunas subclases interesantes y demostramos que son totalmente equilibrados

Keywords: inventory situations, cooperative games and totally balanced games

1. Introduction

In this work we analyze situations where a group of firms order an indivisible good over a finite planning horizon to cover a known demand. We consider that there exist four different types of costs: purchasing, inventory holding, backlogging and setup. The goal is to satisfy firms' entire demand over the planning horizon at a minimum operation cost. This model is the backlogging extension of the well-known Wagner & Whitin model which Zangwill (1969) solved by dynamic programming techniques.

Our approach assumes the above mentioned framework when several firms try to coordinate their actions in order to reduce their operation costs. This coordination makes that firms share their holding technologies and ordering channels. Therefore, when a coalition of firms is to form (joint venture) each firm works with the best holding technology and ordering channels among the members of the coalition. This means that the members of that coalition purchase, hold inventory, pay backlogged demand and make orders at the minimum cost of the coalition members. It is clear that the above coordination process induces savings and therefore, studying the problem of how to allocate the overall saving among the firms is a meaningful problem. This allocation problem can be modelled by a transferable utility cooperative game. In this game the characteristic value of each coalition of firms is obtained solving the combinatorial optimization problem that results from Zangwill's model induced by the members of the coalition.

It is clear by the above description that we deal with a new class of cooperation in inventory situations. Inventory games were introduced in Meca et al. (2003, 2004). These games arise when studying the coordination in inventory problems where a group of firms dealing with the ordering and holding of a certain

commodity (every individual agent's problem being an Economic Production Quantity and Economic Order Quantity problem respectively), either decide to cooperate and make their orders jointly or consider coordination with regards to ordering and holding costs.

A recent paper by Van Den Heuvel et al. (2006) studies coordination in economic lot sizing situations (henceforth, ELS-situations). In that finite horizon model, players should satisfy the demand in each period by producing in that period or carrying inventory from previous periods, backlogging is not allowed. The main difference between that model and the one given in this work is that the former considers setup costs but assumes that costs are the same for all players in every period. The main result in Van Den Heuvel et al. (2006) is that ELS-games (games induced by ELS-situations) are balanced.

In this work we study a new model of coordination in inventory problems where a group of firms place orders of indivisible goods instead of producing them, therefore a new cost is needed to be considered: setup cost. It leads to a new class of totally balanced combinatorial optimization games called setup-inventory games (henceforth, SI-games). SI-games extends ELS-games since all costs considered can be different for several players in every period and backorders are allowed. We prove in this work that SI-games are totally balanced and we study a subclass of SI-games without setup costs. In addition, we prove for this subclass of games that the Owen set (set of allocations which are achievable through dual solutions, see Gellekom et al. (2000)) reduces to a singleton. Moreover, it is always a core-allocation and it can be reached through a population monotonic allocation scheme. This fact motivates the name of Owen point for this core-allocation on SI-games without setup costs. Later, we investigate the structure of the core for this subclass of SI-games determining its algorithmic complexity and carrying out an axiomatic characterization of the Owen point.

2. Bibliography

- [1] Gellekom JRG, Potters JAM, Reijnierse JH, Engel MC, Tijs SH (2000). *Characterization of the Owen Set of Linear Production Processes*. Games and Economic Behavior 32:139-156.
- [2] Meca A, García-Jurado I and Borm PEM (2003). *Cooperation and Competition in Inventory Games*. Math. Meth. Op. Res. 3: 481-493.
- [3] Meca A, Timmer J, García-Jurado I and Borm PEM (2004). *Inventory Games*. European Journal of Operations Research 156: 127-139.
- [4] Van den Heuvel W, Borm P and Hamers H (2006). *Economic lot-sizing games*. European Journal of Operational Research (to appear).
- [5] Zangwill WI (1969). *A backlogging model and multi-echelon model of a dynamic economic lot size production system- a network approach*. Management Sci., 15, 9: 506-527.

Cualificación de restricciones en programación multiobjetivo

Luis Coladas Uría¹

¹coladas@usc.es, Departamento de Estadística e Investigación Operativa,
Universidad de Santiago de Compostela

Abstract

En un contexto multiobjetivo estudiaremos condiciones bajo las cuales las cualificaciones de restricciones clásicas de la optimización mono-objetivo garantizan la verificación de las condiciones de Kuhn-Tucker.

Palabras Clave: Optimización multiobjetivo. Cualificación de restricciones.

AMS: 90C29

1. Introducción

Consideraremos problemas de optimización multiobjetivo en los que el sistema de preferencias del decisor puede representarse aproximadamente por un cono convexo y la región factible viene definida mediante desigualdades.

Si bien muchos resultados habituales de la optimización mono-objetivo pueden generalizarse a la multiobjetivo no siempre el comportamiento de ambos tipos de problemas es análogo. En particular esta situación se presenta a la hora de utilizar en la programación multiobjetivo las cualificaciones de restricciones (CR) habituales en la programación mono-objetivo. Véase por ejemplo Wang-Yang [10].

Es bien conocido el papel que juegan las CR para garantizar que una solución óptima verifique no solo las condiciones de Fritz-John sino también las de Kuhn-Tucker. Son numerosas las CR que pueden utilizarse en el caso mono-objetivo con funciones diferenciables, siendo la más débil entre las *clásicas* la de Guignard. Un estudio completo de esta situación puede verse en Bazaraa-Shetty [1].

En el caso multiobjetivo diversos autores han utilizado condiciones más fuertes como pueden ser la CR de Kuhn-Tucker (Sawaragi-Nakayama-Tanino [9]) o la, más débil, de Abadie (Coladas [3]). En los últimos años se ha estudiado profusamente el tema para el caso de las soluciones de Pareto. Así, Maeda [6] introduce generalizaciones a la CR de Guignard y a otras clásicas y da condiciones para su verificación. Este trabajo da lugar a su vez a numerosas generalizaciones de las que, por brevedad, citaremos solo algunas: Preda-Chitescu [8] las dan para el caso semidiferenciable y Li [5] para el caso *no regular* de funciones localmente Lipschitz. Destacaremos también las aportaciones de Jiménez y Novo, incluidas o citadas en [4] en donde estudian los casos Dini y

Hadamard diferenciables. Volviendo al caso diferenciable, son muy interesantes las condiciones de segundo orden estudiadas por Bigi-Castellani [2]. Finalmente, de nuevo Maeda [7] estudia también condiciones de segundo orden en casos *no regulares*.

En nuestro trabajo retomaremos la situación clásica de la programación monoobjetivo, considerando funciones diferenciables y estudiando bajo distintas hipótesis la verificación de las condiciones de Kuhn-Tucker en un contexto multiobjetivo bajo la CR de Guignard. La eliminación de dichas hipótesis suplementarias no parece posible, confirmando la existencia de un *gap* entre la optimización multiobjetivo y la escalar.

2. Bibliografía

- [1] Bazaraa, M.S. and Shetty, C.M. (1976). *Foundations of Optimization*, Springer-Verlag.
- [2] Bigi, G. and Castellani, M. (2000). Second order optimality conditions for differentiable multiobjective problems. *RAIRO Oper. Res.* 34, 411-426.
- [3] Coladas, L. (1981). Soluciones no dominadas en problemas multiobjetivo. *Trabajos Estadíst. Investigación Oper.* 32, 3-12.
- [4] Giorgi, G., Jiménez, B. and Novo, V. (2004). On constraint qualifications in directionally differentiable multiobjective optimization problems. *RAIRO Oper. Res.* 38, 255-274.
- [5] Li, X.F. (2000). Constraint qualifications in nonsmooth multiobjective optimization. *J. Optimization Theory Appl.* 106, 373-398.
- [6] Maeda, T. (1994). Constraint qualifications in multiobjective optimization problems: Differentiable case. *J. Optimization Theory Appl.* 80, 483-500.
- [7] Maeda, T. (2004). Second-order conditions for efficiency in nonsmooth multiobjective optimization problems. *J. Optimization Theory Appl.* 122, 521-538.
- [8] Preda, V. and Chitescu, I. (1999). On constraint qualification in multiobjective optimization problems: Semidifferentiable case. *J. Optimization Theory Appl.* 100, 417-433.
- [9] Sawaragi, V., Nakayama, H. and Tanino, T. (1985). *Theory of Multiobjective Optimization*, Academic Press.
- [10] Wang, S.Y. and Yang, F.M. (1991). A gap between multiobjective optimization and scalar optimization. *J. Optimization Theory Appl.* 68, 389-391.

Definición y uso de un problema Core para el problema de transporte con costos fijos

Jesús Sáez Aguado

jsaez@eio.uva.es, Departamento de Estadística Investigación Operativa,
Universidad de Valladolid

Abstract

En este trabajo se presenta un nuevo algoritmo para el problema de transporte con costos fijos (FCTP), basado en la definición de un problema Core de dimensiones reducidas, y en la aplicación de una nueva relajación y heurística Lagrangeana al problema Core.

Claves: Programación Entera, Problema Core, Heurística Lagrangeana.

AMS: Clasificación AMS (Opcional)

1. Introducción

Se considera el problema de transporte con costos fijos, cuya formulación como un problema de programación entera mixta es la siguiente:

$$\begin{array}{ll} \text{Minimizar} & \sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij} + \sum_{i=1}^m \sum_{j=1}^n f_{ij} y_{ij} \\ \text{sujeto a} & \\ & \sum_{j=1}^n x_{ij} = s_i \quad i = 1, \dots, m \\ & \sum_{i=1}^m x_{ij} = d_j \quad j = 1, \dots, n \\ & x_{ij} \leq u_{ij} y_{ij} \quad i = 1, \dots, m, j = 1, \dots, n \\ & x_{ij} \geq 0 \quad i = 1, \dots, m, j = 1, \dots, n \\ & y_{ij} \in \{0, 1\} \quad i = 1, \dots, m, j = 1, \dots, n \end{array}$$

donde m es el número de orígenes, n el número de destinos, s_i la oferta del origen i , d_j la demanda del destino j y c_{ij} , f_{ij} los costes unitario y fijo, respectivamente, del par ij .

Es muy bien conocida la gran dificultad computacional de este problema, lo que hace que los métodos de resolución más habituales, fundamentalmente los algoritmos branch and bound, necesiten enormes tiempos de cálculo. En este trabajo se presenta un nuevo algoritmo heurístico para este problema basado en tres componentes:

1. Relajación Lagrangeana. Los siguientes conjuntos son relajaciones para el conjunto de variables binarias del problema:

$$\blacksquare Y_r = \{y \in \{0, 1\}^{nm} : \sum_{j=1}^n u_{ij} y_{ij} \geq s_i, i = 1, \dots, m\}$$

■ $Y_c = \{y \in \{0,1\}^{nm} : \sum_{i=1}^n u_{ij}y_{ij} \geq d_j, j = 1, \dots, n\}$

La optimización sobre Y_r o Y_c se reduce a una serie de subproblemas separables tipo mochila 0-1. Utilizando estas estructuras se define una relajación Lagrangeana y una heurística Lagrangeana que mejoran las desarrolladas por ejemplo en [3].

2. Definición de un problema Core. Utilizando los costes reducidos obtenidos después de aplicar la relajación Lagrangeana anterior, se define un problema Core que incluye sólo un subconjunto de todas las variables, pero donde se espera que estén todas las variables de interés. Esta metodología está siendo actualmente aplicada con éxito en diferentes problemas de localización, como en [1]. El problema Core que se propone, depende de la elección de dos parámetros (k, f) , donde k es el mínimo número de variables (i, j) para cada destino j , y f es un factor de compensación entre los costos fijos y los variables.
3. Resolución del problema Core. Sobre el problema Core puede aplicarse cualquier algoritmo exacto o aproximado. En este trabajo se aplica de nuevo la relajación y heurística Lagrangeana anterior a los problema Core obtenidos para diferentes valores de los parámetros k y f . El algoritmo resultante encuentra soluciones al problema original que mejoran enormemente a las obtenidas por Cplex 9.1, y son competitivas con las soluciones basadas en el reciente procedimiento Parametric Host Image Process ([2]).

2. Bibliografía

- [1] P. Avella, M. Boccia, A. Sforza, I. Vasil'ev (2003), An effective heuristic for large-scale Capacitated Plant Location problems, revised version submitted for publication in *Journal of Heuristics*.
- [2] Glover, F., M. Amini and G. Kochenberger (2005). Parametric ghost image processes for fixed-charge problems: a study of transportation networks. *Journal of Heuristics* 11, 307-226.
- [3] Wright, D. D., C. Haehling and von Lanzenauer (1989). Solving the fixed charge problem with Lagrangian relaxation and cost allocation heuristics, *European Journal of Operational Research* 42, 305-312.

Determinación del precio y del tamaño del lote para una curva demanda-precio tipo racional

L.A. San-José¹, J. Sicilia², J. García-Laguna³

¹Departamento de Matemática Aplicada, Universidad de Valladolid

²Departamento de Estadística, I.O. y Computación, Univ. de La Laguna

³Departamento de Estadística e Investigación Operativa, Univ. de Valladolid

¹augusto@mat.uva.es, ²jsicilia@ull.es, ³laguna@eio.uva.es

Abstract

Se estudia un modelo de optimización para el análisis de un mercado en el que intervienen las ganancias asociadas con la venta de un artículo y los costes asociados a la gestión del inventario de dicho artículo. Se considera que la demanda y el precio de venta están inversamente relacionados a través de una curva tipo racional. Bajo las hipótesis anteriores, se calculan el precio final de venta y el tamaño óptimo del lote.

Claves: modelos EOQ, demanda dependiente, precio.

AMS: 90B05.

1. Introducción

En la literatura sobre Gestión de Inventarios se pueden encontrar interesantes trabajos donde se formulan modelos de optimización para la reposición de artículos. En ellos, el objetivo se centra en determinar la política óptima de reposición de forma que se minimicen los costes relativos a la gestión del inventario. Generalmente se admite que el precio de venta de los artículos está previamente determinado y que la cuantía de los ingresos depende principalmente de la demanda de los clientes. Sin embargo, en muchos casos, el vendedor tiene opción de modificar el precio de venta, influyendo de esta forma en el valor económico de lo vendido. El problema de determinar el precio de venta a los consumidores finales, cuando para tales consumidores existe una relación demanda-precio lineal fue tratado por Porteus (1985). Abad (1988) estudia un problema similar considerando una función de demanda más general, cuando el minorista ofrece descuentos cuantitativos a los consumidores. Posteriormente, Abad (1996) considera el caso de artículos perecederos. Parlar y Wang (1994) y Weng (1995) estudian un problema similar, pero centrado en la relación entre el minorista y sus proveedores.

Recientemente, Lau y Lau (2003) han investigado un modelo demanda-precio en ausencia de costes asociados a la gestión del inventario. Una de sus aportaciones consiste en señalar que la naturaleza de la interrelación demanda-precio puede tener efectos significativos en los resultados del modelo concreto a estudiar.

Teniendo en cuenta que no es fácil de obtener la estimación de la función que relaciona la demanda con el precio, a la hora de analizar modelos donde interviene tal relación, un enfoque interesante se tiene al considerar una curva tipo racional que relaciona la demanda con el precio y, a partir de ahí, determinar el precio final de venta y la reposición óptima.

2. Agradecimientos

Este trabajo ha sido parcialmente subvencionado por el *Ministerio de Educación y Ciencia*, Proyecto de Investigación MTM2004-07550, del Plan National de Investigación Científica, Desarrollo Tecnológico e Innovación. También, ha sido parcialmente subvencionado por el proyecto OTP/04/03 del Gobierno Regional de la *Comunidad Autónoma de Castilla y León*. Una parte de esta investigación ha sido realizado durante una estancia del segundo autor en la Universidad de Valladolid (Beca de Estancia de la Dirección General de Universidades e Investigación del Gobierno de la *Comunidad Autónoma de Canarias*).

3. Bibliografía

- [1] Abad P.L. (1988). Determining optimal selling price and lot size when the supplier offers all-unit quantity discounts. *Decisions Science* 19, 622-634.
- [2] Abad P.L. (1996). Optimal pricing and lot-sizing under conditions of perishability and partial backordering. *Management Science* 42, 1093-1104.
- [3] Lau A.H.L. y Lau H-S. (2003). Effects of a demand-curve's shape on the optimal solutions of a multi-echelon inventory-pricing model. *Production, Manufacturing and Logistics* 147, 530-548.
- [4] Parlar M. y Wang Q. (1994). Discounting decisions in a supplier-buyer relationship with a linear buyer's demand. *IIE Transactions* 26 (2), 34-41.
- [5] Porteus E.L. (1985). Investing in reduced set-ups in the EOQ model. *Management Science* 31 (8), 998-1010.
- [6] Weng Z.K. (1995). Channel coordination and quantity discounts. *Management Science* 41, 1509-1522.

Diseño de rutas de recogida de residuos sólidos urbanos en el área metropolitana de Barcelona

Joaquín Bautista¹, Jordi Pereira², Elena Fernández³

¹joaquin.bautista@upc.edu, Nissan Chair UPC, Universitat Politècnica de Catalunya

²jorge.pereira@upc.edu, Dpto Organizació d'Empreses, Universitat Politècnica de Catalunya

³e.fernandez@upc.edu, Departament d'Estadística i Investigació Operativa, Universitat Politècnica de Catalunya

Abstract

Los problemas asociados a la recogida de residuos sólidos urbanos son muy variados. En este trabajo se presenta el problema de diseño de itinerarios de recogida y se muestran los resultados ofrecidos por un procedimiento basado en colonias de hormigas a la recogida en un núcleo urbano del Área Metropolitana de Barcelona.

Palabras Clave: Recogida de residuos sólidos urbanos, itinerarios, metaheurísticas

1. Introducción

Durante los últimos años, la preocupación social sobre el medio ambiente ha adquirido una gran relevancia. En la Unión Europea un cuerpo creciente de directivas comunitarias basadas en el entorno de trabajo fijado por las directivas 75/442/CEE, 91/156/CEE y 94/62/CEE obliga a los miembros de la unión a recuperar y reciclar muchos productos y componentes como cubre en el caso español los decretos reales del 15 de Julio, 6/1993 y 21 de Abril 10/1998. Como el ámbito de las directivas está ampliándose, es fácil predecir que los años venideros verán la creación de nuevas empresas dedicadas al transporte, almacenamiento y distribución específicas de residuos. En el marco impuesto, la recogida selectiva de residuos urbanos se convierte en uno de los procesos más importantes del ciclo generación, transformación y eliminación de residuos, [1].

En el caso de los residuos urbanos son, por adjudicación, organizaciones privadas o municipales las encargadas de realizar la recogida selectiva de los residuos, siendo responsabilidad suya la correcta recuperación a mínimo coste.

Un factor clave para la correcta operación de estos sistemas de recogida es el diseño de rutas de recogida para cada fracción.

La naturaleza básica del problema tratado es la de un problema de itinerarios con demanda asociada a los arcos del grafo de circulación, aunque presenta diversas características que lo apartan de los modelos presentes en la literatura

[2]. Estas características especiales son, por una parte, derivadas de las señales de circulación, y por otra, de la naturaleza mixta del grafo de transporte.

El presente trabajo se centra en el diseño de rutas de recogida tal como se encuentran en el Área Metropolitana de Barcelona, donde cada tipo de residuo se recoge de forma separada con una periodicidad fija establecida por el municipio. Inicialmente se construye un modelo para el problema de itinerarios con servicio asociado a arcos en un grafo mixto, donde se tiene en cuenta las señales de circulación en el conjunto de restricciones del problema. Para su resolución se transforma el modelo original en un problema de itinerarios con demanda asociada a los vértices y se resuelve mediante dos heurísticas basadas en colonias de hormigas, [3].

La aplicación de este modelo a la recogida de residuos sólidos urbanos en la fracción de rechazo en el municipio de Sant Boi de Llobregat, genera una instancia con 220 aristas y 460 arcos con demanda, que representan los diversos puntos de recogida de residuos de unos 75000 habitantes de la población. Dicha recogida se realiza con periodicidad diaria por cinco camiones de recogida.

Comparando el resultado obtenido por los procedimientos propuestos con la solución provista por el Ayuntamiento, los algoritmos implementados obtienen un ahorro en la longitud de cada ruta de entre el 15 % y el 30 %, lo que representa un ahorro total de entre 15 y 37 kilómetros diarios. Esta reducción de longitud está directamente asociada no sólo al coste de operaciones sino también a la calidad de servicio percibida por los usuarios al reducir el tiempo al que los ciudadanos se ven sometidos a las molestias asociadas por la recogida de residuos.

2. Agradecimientos

Esta investigación ha sido parcialmente financiada por la ayuda BEC2003-03809 del Gobierno de España.

3. Bibliografía

- [1] Bautista, J. (2001). *Proyecto integral de gestión de residuos urbanos en el municipio de Sant Boi de Llobregat*. CPDA.
- [2] Dror, M., editor (2000) *Arc Routing: Theory, Solutions and Applications*. Kluwer.
- [3] Bautista, J.; Pereira, J. (2004) Ant Algorithms for Urban Waste Collection Routing. *Lecture Notes in Computer Science*, 3172:302-309.

Distribución del número de reintentos efectuados por un cliente

M.J. López-Herrero

Lherrero@estad.ucm.es, Departamento de Estadística e Investigación Operativa III, Escuela Universitaria de Estadística, Universidad Complutense de Madrid

Abstract

Consideramos un modelo de colas sin línea de espera en el que los clientes bloqueados intentan acceder al servidor cada cierto tiempo para conseguir ser atendidos. En este trabajo se obtiene la distribución del número total de reintentos efectuados por un cliente marcado hasta que consigue acceder al servicio.

Claves: cola con reintentos, cliente marcado, tiempo de espera.

AMS: 90B22, 60K25

1. Introducción

En modelos de colas con reintentos un cliente que a su llegada encuentra todos los servidores ocupados debe abandonar la zona de servicio y repetir su demanda algún tiempo aleatorio después. Durante este tiempo se dice que el cliente está en "órbita". Las colas con reintentos se utilizan para modelizar problemas asociados a sistemas telefónicos, informáticos y, en general, a sistemas de comunicación.

El número de reintentos efectuados por un cliente hasta conseguir ser atendido por el servidor, R , es una medida equivalente, desde un punto de vista discreto, al tiempo de espera de un cliente, W . El análisis del tiempo de espera constituye una parte fundamental de la literatura en teoría de colas y su estudio en modelos con reintentos es más difícil que el estudio del estado del sistema. Ello se debe a que los clientes se sirven de acuerdo con un orden aleatorio, siendo posible que un cliente que se une a la órbita en cierto momento sea atendido por el servidor con anterioridad a otro cliente más antiguo.

El principal objetivo del trabajo es determinar numéricamente la distribución estacionaria del número de reintentos efectuados por un cliente durante su tiempo de espera.

2. Descripción del modelo

Consideramos un sistema de colas con un único servidor al que los clientes llegan siguiendo un proceso de Poisson de intensidad λ . Cualquier cliente que encuentra el servidor ocupado se une a la órbita. Cada cliente en órbita genera, independientemente de los demás, un flujo de demandas repetidas para

conseguir ser atendido de tipo Poisson, con tasa μ . La duración del servicio de cualquier cliente sigue una distribución general con función de distribución $B(x)$ ($B(0) = 0$) y momento de orden k dado por β_k . Se supone que el flujo de llegadas, los intervalos entre sucesivos reintentos y la duración de los tiempos de servicio son variables mutuamente independientes.

En cualquier instante t el estado del sistema puede representarse mediante un proceso bidimensional $Y(t) = (C(t), N(t))$, donde $C(t)$ es 0 ó 1 dependiendo de si el servidor está vacío u ocupado y $N(t)$ contabiliza el número de clientes en la órbita en el instante t . La condición de estabilidad del sistema viene dada por $\rho = \lambda\beta_1 < 1$ y bajo esta hipótesis definimos las probabilidades estacionarias:

$$p_{in} = \lim_{t \rightarrow \infty} P\{C(t) = i, N(t) = n\}, \quad 0 \leq i \leq c, \quad n \geq 0.$$

3. Estudio del número de reintentos en un sistema $M/G/1$

En primer lugar marcamos un cliente a su llegada al sistema. Para determinar la distribución asociada a R , número de reintentos que realiza para acceder al servidor, observamos la situación del sistema cuando el cliente marcado llega. Puede ocurrir que el servidor esté libre, en este caso, su servicio comienza y $R = 0$. En otro caso el cliente ha llegado durante un tiempo de servicio y hasta que consiga acceder al servidor deberá unirse a la órbita al menos durante el tiempo residual del servicio que está llevándose a cabo. En concreto la distribución de R satisface:

$$P\{R = 0\} = \sum_{n=0}^{\infty} p_{0n} = 1 - \rho, \quad (1)$$

$$P\{R = i\} = \int_0^{\infty} \sum_{n=0}^{\infty} p_{1n}(x) \sum_{j=0}^{i-1} \sum_{k=0}^{\infty} e_{j,k}(x) x_{0,n+k+1}^{i-j} dx, \quad i \geq 1. \quad (2)$$

donde $x_{r,n}^i$ es una probabilidad condicionada a los reintentos previamente realizados por el cliente marcado y $e_{j,k}(x)$ es la probabilidad de efectuar j reintentos y que lleguen k clientes durante el servicio residual.

Divergencia no probabilística entre conjuntos borrosos

Jorge Jiménez¹, Susana Montes²

¹meana@uniovi.es, Departamento de Matemáticas, Universidad de Oviedo

²montes@uniovi.es, Departamento de Estadística, Investigación Operativa y Didáctica de la Matemática, Universidad de Oviedo

Abstract

En este trabajo se introduce el concepto de medida de divergencia no probabilística de tipo local, como forma de comparar dos conjuntos borrosos mediante medidas que sean operativas en la práctica. Para dichas medidas se analizan además sus principales propiedades.

Claves: medida de divergencia no probabilística, t-norma.

1. Introducción

En muchas ocasiones es necesario comparar dos conjuntos que no están claramente determinados, es decir, sobre los que existe una falta de información en forma de imprecisión. Dichos conjuntos pueden surgir, por ejemplo, como resultado de un experimento que ya ha sido realizado. Así pues, el resultado es conocido pero, por alguna razón, no puede ser descrito con precisión. En tales situaciones nos encontraremos ante conjuntos borrosos o difusos (ver, por ejemplo, [1]), y por lo tanto será necesario una herramienta para comparar dichos tipos de conjuntos. Con este propósito surgen las medidas de divergencia no probabilísticas ([3]), basadas inicialmente en las clásicas medidas de divergencia (ver, por ejemplo, [5]) y que posteriormente han sido axiomatizadas ([4]). Dichas medidas son muy generales y, aunque tienen algunas interesantes propiedades, poco manejables desde un punto de vista práctico o computacional. Para resolver este problema, se ha trabajado principalmente con una subfamilia de medidas de divergencia no probabilística: las medidas de divergencia locales. En este trabajo se considera una definición más general de medidas de divergencia locales, pero no por ello menos operativa. Para esta nueva familia se estudian sus principales propiedades y características.

2. Conceptos previos y resultados

Si un conjunto \tilde{A} no está definido con total precisión, ciertos elementos del espacio de referencia o muestral Ω no sabremos con claridad si están o no en dicho conjunto. En tales casos, el grado de pertenencia del elemento ω de Ω al conjunto \tilde{A} no podremos decir si es 1 (está en el conjunto) o 0 (no está en el conjunto), con lo que le asignaremos un valor en el intervalo (0, 1), en función del grado de imprecisión que tengamos (el valor 1/2 se consideraría como imprecisión máxima). En tales casos, estamos pues trabajando con un

conjunto borroso o difuso, que viene caracterizado por la función que nos mide el grado de pertenencia de cada elemento de Ω a él. Dicha función será denotada por $\tilde{A}(\omega)$ y se denomina **función de pertenencia**.

Dados dos conjuntos borrosos \tilde{A} y \tilde{B} , el grado de divergencia o diferencia entre ellos puede determinarse a través de una aplicación D , llamada **medida de divergencia no probabilística**. Dicha medida queda caracterizada por los siguiente axiomas naturales: (A1) $D(\tilde{A}, \tilde{B}) = D(\tilde{B}, \tilde{A}), \forall \tilde{A}, \tilde{B} \in \mathcal{F}(\Omega)$; (A2) $D(\tilde{A}, \tilde{A}) = 0, \forall \tilde{A} \in \mathcal{F}(\Omega)$; (A3) $\max\{D(\tilde{A} \cup_S \tilde{C}, \tilde{B} \cup_S \tilde{C}), D(\tilde{A} \cap_T \tilde{C}, \tilde{B} \cap_T \tilde{C})\} \leq D(\tilde{A}, \tilde{B}), \forall \tilde{A}, \tilde{B}, \tilde{C} \in \mathcal{F}(\Omega)$; donde $\mathcal{F}(\Omega)$ representa el conjunto de partes borrosas de Ω , \cap_T la intersección entre conjuntos borrosos definida a partir de la norma triangular T y \cup_S la unión definida mediante la conorma triangular S ([2], [6]). Si no hay lugar a confusión con las medidas de divergencia probabilísticas o clásicas, a la aplicación D se le llama simplemente medida de divergencia. Un tipo particular de medidas de divergencia son las **locales**, en las cuales la divergencia entre dos conjuntos se puede descomponer como la suma de la divergencia entre ellos para cada elemento de Ω . Dichas medidas presentan importantes propiedades y además son muy adecuadas desde el punto de vista computacional. Se presenta ahora una generalización de esta idea, considerando cualquier conorma triangular para componer la divergencia en cada punto, en lugar de considerar simplemente la suma. Se propone pues el caso más general de medidas de divergencia componibles y se analizan sus principales propiedades.

3. Agradecimientos

Esta investigación está subvencionada por el Proyecto MTM2004-01269.

4. Bibliografía

- [1] Gil M.A. (1993). Análisis y tratamiento estadístico de elementos difusos en experimentos aleatorios. *Estadística Española* 35(134), 477-525.
- [2] Klement E.P., Mesiar R. y Pap E. (2000). *Triangular Norms*. Kluwer Academic Publishers.
- [3] Montes S., Couso I. y Bertoluzza C. (1998). Some classes of fuzziness measures from local divergences. *Belg. J. Oper. Res. Statist. Comput. Sci.* 38, 37-49.
- [4] Montes S., Couso I., Gil P. y Bertoluzza C. (2002). Divergence measure between fuzzy sets. *Int. J. Approximate Reasoning* 30, 91-105.
- [5] Pardo L. (1997). *Teoría de la Información estadística*. Ed. Hesperides.
- [6] Schweizer B. y Sklar A. (1960). Statistical metric spaces. *Pacific J. Math.* 10, 313-334.

EAGH (Empirically Adjusted Greedy Algorithms) para la determinación de secuencias en una máquina multiproducto

Albert Corominas¹, Rafael Pastor¹, Ana Sánchez¹

¹albert.corominas@upc.edu, rafael.pastor@upc.edu, ana.sanchez@upc.edu

Universidad Politécnica de Cataluña

Abstract

Se considera una máquina capaz de producir distintas piezas simultáneamente y se desea minimizar sus costes de almacenamiento y de carencia. En este trabajo se propone resolver el problema aplicando EAGH, como un procedimiento de resolución basado en calibrar un polinomio de los atributos del problema.

Palabras Clave: Prioritized Hedging Points, EAGH, 1-máquina

1. Definición del problema

Se considera una máquina capaz de producir, de forma simultánea, n tipos de pieza distintos con unas tasas de demanda conocidas. Dicha máquina está sujeta a fallos y el objetivo consiste en minimizar la esperanza matemática de los costes de almacenamiento y de carencia.

El problema se formaliza en [1] dónde los autores proponen seguir una política de tipo PHP (*prioritized hedging points*) para su resolución: respetando un orden de prioridad previamente establecido, se va asignando la capacidad disponible de la máquina a los diferentes tipos de pieza hasta que el stock de dichas piezas alcanza una determinada cota llamada *hedging point* (que depende del tipo de pieza y de su prioridad y cuya expresión queda establecida); una vez alcanzada dicha cota, se mantiene un nivel de producción igual a la demanda de la pieza para mantener constante el nivel de stock.

En [2] se propone un procedimiento de resolución exacto (que implica determinar la secuencia óptima de prioridad de las piezas) y se observa que los tiempos de cálculo crecen exponencialmente con el tamaño del problema. Esto hace necesario procedimientos de secuenciación alternativos como los procedimientos heurísticos propuestos en [3], una gran parte de los cuáles se basa en la ordenación de las piezas en función del valor de diferentes polinomios formados por los datos característicos de cada tipo de pieza y que provienen de un análisis exhaustivo del problema.

2. Calibración de heurísticas

Una alternativa más elaborada es la calibración un polinomio (en este caso mediante el algoritmo de Nelder y Mead) para su posterior utilización como un procedimiento heurístico de ordenación (véase [4] para mayor detalle sobre

EAGH). Se desarrollan, para el problema de estudio, tres procedimientos basados en esta metodología: EAGH-1, EAGH-N y EAGH-N*m.

En EAGH-1, para cada ejemplar y partiendo del polinomio que en promedio proporciona mejores resultados en [3], se aplica el algoritmo de Nelder y Mead a modo de procedimiento de optimización local. En EAGH-N, se aplica el algoritmo de Nelder y Mead a un conjunto de ejemplares de “entrenamiento” para obtener un polinomio, que se utiliza posteriormente para la secuenciación en un nuevo conjunto de ejemplares de validación. EAGH-N*m es similar a EAGH-N, con la única diferencia que se consideran polinomios distintos para determinar cada una de las posiciones de la secuencia (un polinomio para determinar la primera posición, un polinomio distinto para determinar la segunda posición, etc.).

3. Experiencia computacional

La experiencia computacional consiste en la resolución de los mismos 1400 ejemplares que en [2] y [3] con cada uno de los tres procedimientos heurísticos desarrollados. Posteriormente, se comparan los resultados obtenidos (entre sí y respecto a los resultados de las heurísticas de [3]) y se mide la calidad de los procedimientos desarrollados calculando el promedio de la discrepancia respecto a las soluciones óptimas.

4. Perspectivas

Resolución mediante el uso de cotas en un programa dinámico (de tipo Branch & Bound) o mediante metaheurísticas.

5. Bibliografía

- [1] Shu, C.; Perkins, J. (2001). Optimal PHP production of multiple part-types on a failure-prone machine with quadratic buffer costs. *IEEE Trans. Automat. Contr.*, 46 541-549.
- [2] Corominas, A.; Griñó, R.; Pastor, R. (2004). Determinación de secuencias óptimas en una máquina multiproducto sujeta a fallos y con costes cuadráticos. *XXVIII Congreso Nacional de Estadística e Investigación Operativa*, Cádiz.
- [3] Corominas, A.; Pastor, R.; Sánchez, A. (2006). Procedimientos heurísticos para la determinación de secuencias en una máquina multiproducto sujeta a fallos y con costes cuadráticos. *XXIX Congreso Nacional de Estadística e Investigación Operativa*, Tenerife.
- [4] Corominas, A. (2005). Empirically Adjusted Greedy Algorithms (EAGH): A new approach to solving combinatorial optimisation problems. *Working paper IOC-DT-P-2005-22*. Universidad Politécnica de Cataluña. Barcelona.

Eficiencia de diferentes modelos de PLE para resolver SALBP

Rafael Pastor¹, Laia Ferrer²

¹rafael.pastor@upc.edu, ²laia.ferrer@upc.edu

Instituto de Organización y Control, Univ. Politécnica de Cataluña

Abstract

El Simple Assembly Line Balancing Problem puede modelizarse matemáticamente de distintos modos. En un contexto en el que la resolución óptima mediante programación matemática está adquiriendo relevancia, se compara la eficacia de los 3 modelos destacados por Scholl (1999) y se identifica el más eficiente para dos variaciones del problema.

Palabras Clave: Equilibrado de líneas de montaje

1. Introducción

El Simple Assembly Line Balancing Problem (SALBP) consiste en asignar una serie de tareas a estaciones de trabajo de modo que se optimice una medida de eficiencia, respetando restricciones de precedencia. Si el objetivo es minimizar el numero de estaciones m dado un tiempo de ciclo tc , el problema se denomina SALBP-1; si el objetivo es minimizar tc dado m , se denomina SALBP-2; SALBP-F consiste en obtener una solución factible dado tc y m . Para modelizar el problema matemáticamente, Scholl (1999) destaca 3 modelos básicos alternativos para SALBP-F.

Tradicionalmente, la resolución exacta de SALBP con programación matemática y software estándar se ha considerado ineficiente para resolver ejemplos de tamaño real. Aun así, recientes mejoras en la potencia de cálculo en hardware y en software, han reducido en un factor de 1.000.000 el tiempo de cálculo para resolver modelos matemáticos (Bixby, 2002). En este contexto, analizar la mejor modelización con programación matemática está ganado relevancia. Se realiza una amplia experiencia computacional para comparar la eficiencia de las 3 modelizaciones presentadas por Scholl (1999) del SALBP-1 y SALBP-2, y que, según nuestro conocimiento, no han sido estudiadas.

2. Modelos de PLE para SALBP

Scholl (1999) destaca 3 modelos, que se diferencian por utilizar:

- **Var. binarias de impulso:** x_{ij} toma valor 1 si y solo si la tarea i se asigna a la estación j .
- **Var. binarias de salto:** x_{ij} toma valor 1 si y solo si la tarea i se asigna a la estación j o a una estación anterior.
- **Var. mixtas:** z_i toma el valor de la estación a la que se asigna la tarea i . La adaptación de la formulación de SALBP-F a SALBP-1 y SALPB-2 utiliza los resultados de Pastor et al (2004), para las restricciones de precedencia, y el tipo de linealización presentado en Lusa et al (2006).

3. Experiencia computacional

La experiencia computacional resuelve ejemplares presentados en www.assembly-line-balancing.de. Los resultados muestran que el modelo de variables mixtas es claramente menos eficaz que los otros dos, tanto para SALBP-1 como para SALBP-2. Para SALBP-1, el modelo de variables de impulso es ligeramente más eficaz que el de variables de salto, obtiene más soluciones óptimas (138 vs. 124) y no encuentra solución factible en menos ejemplares (86 vs 116), sobre un total de 259 ejemplares. En cambio para SALBP-2, el modelo de variables de salto resulta considerablemente más eficaz que el de variables de impulso: sobre un total de 293 ejemplares, encuentra muchas más soluciones óptimas (122 vs. 84) y siempre encuentra solución factible (0 vs. 23 ejemplares sin solución).

4. Conclusiones

El Simple Assembly Line Balancing Problem (SALBP) puede modelizarse matemáticamente de distintos modos. Aunque tradicionalmente se ha considerado ineficaz, la resolución de este problema con software estándar y el estudio de sus modelos más eficaces está adquiriendo relevancia, gracias a recientes mejoras en hardware y en software. Este trabajo realiza una amplia experiencia computacional que compara a eficiencia de 3 modelizaciones alternativas destacadas en Scholl (1999). Los resultados muestran que, para SALBP-1, el modelo más eficaz es el que utiliza variables impulso, y que, para SALBP-2, el de variables de salto resulta considerablemente más eficiente.

5. Agradecimientos

Financiado por el proyecto MCYT DPI2004-03472, cofinanciado por FEDER.

6. Bibliografía

- [1] Bixby, R.E. (2002). Solving real-world linear programs: a decade and more of progress. *Operations Research* 50, 3-15.
- [2] Lusa, A., Pastor R., Corominas, A. (2006). Determining the most appropriate set of weekly working hours for planning annualised working time. IOC-DT-I-2006-03
- [3] Pastor, R., Corominas, A., Lusa, A. (2004). Different ways of modelling and solving precedence and incompatibility constraints in the assembly line balancing problem. *Frontiers in Artificial Intelligence and Applications*, 113, 359-366.
- [4] Scholl, A. 1999. *Balancing and sequencing of assembly lines*. Physica-Verlag Heidelberg: Germany. 2nd edition.

El comportamiento emprendedor en la cultura empresarial: una aplicación del modelo de Rasch

J. Febles Acosta¹, J.R. Oreja Rodríguez²

¹jfebles@ull.es, ²joreja@ull.es

Departamento de Economía y Dirección de Empresas, Universidad de La Laguna

Abstract

En el presente estudio se lleva a cabo una aproximación de carácter exploratorio del papel que juega el comportamiento emprendedor en la cultura empresarial canaria. A través del análisis factorial se identifica el predominio de cuatro dimensiones culturales en las empresas analizadas: colaboración, compromiso, flexibilidad y formación. Posteriormente mediante el modelo de Rasch se analiza la presencia de las variables representativas del comportamiento emprendedor dentro de las dimensiones culturales determinadas.

Palabras Clave: Comportamiento emprendedor, cultura y estrategia

AMS: 62H25

Introducción

En un mercado altamente complejo, dinámico e interrelacionado como es el actual (Nicholson, 1995), las empresas están obligadas a adoptar un enfoque proactivo que les permita adaptarse en todo momento de forma rápida y flexible al entorno, o incluso influir sobre el mismo (Cuervo, 1989; Pumpin y García, 1988). En este contexto la cultura empresarial juega un importante papel en la toma de decisiones estratégicas (Mukherji y Hurtado, 2001).

El descubrimiento de oportunidades y la creación de nuevas empresas ha sido el criterio más utilizado para estudiar el proceso emprendedor. Los directivos determinan con su gestión la habilidad de la empresa para responder a las amenazas y oportunidades, no sólo en sus inicios sino también durante todo el ciclo de vida. Sobre esta capacidad de respuesta juega también un importante papel la cultura empresarial, sobre cuya formación y mantenimiento influye abiertamente el estilo directivo, y que a su vez determina el comportamiento de los líderes.

El crecimiento de las empresas también suele relacionarse con el espíritu emprendedor mediante dos perspectivas: los estudios que presentan la figura del emprendedor como un atributo a difundir en la cultura de la propia empresa así como la actitud emprendedora de empresarios y directivos a lo largo del tiempo (Sharma y Crisman, 1999).

De este modo el fenómeno emprendedor se contempla no sólo en términos de creación de nuevas empresas independientes, sino también en función de su incorporación a las propias estrategias corporativas y la orientación a la innovación y explotación de nuevas oportunidades de negocio.

En este trabajo abordamos el estudio de la presencia del comportamiento emprendedor en la cultura empresarial canaria, mediante el análisis factorial y el método Rasch. A través del análisis factorial se identifica cuatro dimensiones culturales en las empresas analizadas: colaboración, compromiso, flexibilidad y formación. Con el modelo de Rasch se jerarquizan las variables de la cultura empresarial, lo cual nos permite discernir la presencia de las variables relacionadas con el comportamiento emprendedor, dentro de las dimensiones culturales establecidas en el análisis factorial previo.

1. Bibliografía

- [1] Cuervo, A. (1989). Bases para el diseño organizativo de la empresa, Papeles de Economía Española, 39.
- [2] Nicholson, N. (Ed.) (1995). Encyclopedic dictionary of organizational behavior, Blackwell Business, Cambridge, Massachussets.
- [3] Mukherji, A. y Hurtado, P. (2001). Interpreting, categorizing and responding to the environment: the role of culture in strategic problem definition, Management Decisión, Vol. 38, 2.
- [4] Pumpin, C. y García, S. (1988). Cultura empresarial, Díaz de Santos, Madrid.
- [5] Rasch, G. (1980). Probabilistic models for some intelligence and attainment tests (expanded ed.). University of Chicago Press.
- [6] Sharma, P. y Chrisman,J.(1999). Toward a reconciliation of the definitional issues in the field of corporate entrepreneurship, Entrepreneurship Theory and Practice, vol.23, Nº3.

El concepto de KT-invexidad en un problema de control

M. Arana-Jiménez¹, R. Osuna-Gómez², A. Rufián-Lizana³, G. Ruiz-Garzón⁴

¹majerez@ono.com, SIES Vega del Guadalete, S. José del Valle, 11580, Cádiz

²rafaela@us.es, Departamento de Estadística e I.O., Universidad de Sevilla

³rufian@us.es, Departamento de Estadística e I.O., Universidad de Sevilla

⁴gabriel.ruiz@uca.es, Departamento de Estadística e I.O., Universidad de Cádiz

Abstract

En este trabajo introducimos una nueva condición sobre las funciones de un Problema de Control, estas condiciones nos definirán el llamado Problema de Control KT-invex. Probamos que un Problema de Control KT-invex queda caracterizado porque todo punto de Kuhn-Tucker es una solución óptima.

Claves: problema de control, invexidad, condiciones de Kuhn-Tucker.

AMS: 34H05, 49K30

1. Introducción

En este trabajo consideraremos la siguiente formulación del Problema de Control matemático:

$$(CP) \quad \begin{array}{l} \text{Minimizar} \int_a^b f(t, x, u) dt \\ \text{sujeto a :} \end{array}$$

$$x(a) = \alpha, \quad x(b) = \beta$$

$$g(t, x, u) \leq 0, \quad t \in I$$

$$h(t, x, u) = \dot{x}, \quad t \in I$$

Donde $I = [a, b]$ es un intervalo real y $f : I \times R^n \times R^m \rightarrow R$, $g : I \times R^n \times R^m \rightarrow R^k$, $h : I \times R^n \times R^m \rightarrow R^n$ son funciones continuamente diferenciables. Es conocido también que:

Definición 1 El punto $(\bar{x}, \bar{u}) \in K$ se dice que es un punto de Kuhn-Tucker si existe unas funciones continuas a trozos $\lambda : I \rightarrow R^k$ y $\mu : I \rightarrow R^n$ satisfaciendo lo siguiente:

$$f_x(t, \bar{x}, \bar{u}) + \lambda(t)^T g_x(t, \bar{x}, \bar{u}) + \mu(t)^T h_x(t, \bar{x}, \bar{u}) + \dot{\mu}(t) = 0 \quad (4.63)$$

$$f_u(t, \bar{x}, \bar{u}) + \lambda(t)^T g_u(t, \bar{x}, \bar{u}) + \mu(t)^T h_u(t, \bar{x}, \bar{u}) = 0 \quad (4.64)$$

$$\lambda(t)^T g(t, \bar{x}, \bar{u}) = 0 \quad (4.65)$$

$$\lambda(t) \geq 0 \quad (4.66)$$

para todo $t \in I$, excepto en las discontinuidades.

En este trabajo introduciremos la siguiente definición:

Definición 2 El Problema de Control (CP) se dice que es *KT-invex* en $(\bar{x}, \bar{u}) \in K$ si para todo $(x, u) \in K$, y para todas $\lambda : I \rightarrow R^k$, que verifica (3) y (4), y $\mu : I \rightarrow R^n$ funciones continuas a trozos, existen funciones vectoriales diferenciables $\eta(t, \bar{x}, x, \bar{u}, u, \lambda, \mu)$ y $\xi(t, \bar{x}, x, \bar{u}, u, \lambda, \mu)$ con $\eta(a, \bar{x}, x, \bar{u}, u, \lambda, \mu) = 0 = \eta(b, \bar{x}, x, \bar{u}, u, \lambda, \mu)$ y $\xi(a, \bar{x}, x, \bar{u}, u, \lambda, \mu) = 0 = \xi(b, \bar{x}, x, \bar{u}, u, \lambda, \mu)$, tal que

$$F(x, u) - F(\bar{x}, \bar{u}) < 0 \Rightarrow$$

$$\begin{aligned} & \int_a^b ((f_x(t, \bar{x}, \bar{u}) + \lambda(t)^T g_x(t, \bar{x}, \bar{u}) + \mu(t)^T h_x(t, \bar{x}, \bar{u}))\eta(t, \bar{x}, x, \bar{u}, u, \lambda, \mu) \\ & - \mu(t)^T \dot{\eta}(t, \bar{x}, x, \bar{u}, u, \lambda, \mu) \\ & + ((f_u(t, \bar{x}, \bar{u}) + \lambda(t)^T g_u(t, \bar{x}, \bar{u}) + \mu(t)^T h_u(t, \bar{x}, \bar{u}))\xi(t, \bar{x}, x, \bar{u}, u, \lambda, \mu)) dt < 0 \end{aligned}$$

Teorema 1 El problema (CP) es *KT-invex* si y sólo si todo punto de Kuhn-Tucker es solución óptima para el problema (CP).

2. Agradecimientos

Este trabajo está financiado por el Ministerio de Ciencia y Tecnología, dentro del Plan Nacional de I+D+I 2003-06, código de proyecto: 06579.

3. Bibliografía

- [1] Arana M., Osuna R., Ruiz G. y Rojas M. (2005). On Variational Problems: Characterization of Solutions and Duality. *J. Math. Anal. Appl.* 311, 1-12
- [2] Martin D.M. (1985). The essence of Invexity. *J. Optim. Theory Appl.* 47, (1), 65-76.

El plan de extensión óptimo de la red ferroviaria de alta velocidad

V. Blanco¹, J. Puerto²

¹vblanco@us.es, ²puelto@us.es

Departamento de Estadística e Investigación Operativa
Universidad de Sevilla

Abstract

Este trabajo presenta un modelo para la expansión de cualquier red de transportes con aplicación en el proyecto actual de expansión de la red de alta velocidad española redactado en el PEIT. La naturaleza NP-dura del problema hace necesario desarrollar algoritmos heurísticos que reduzcan los tiempos de ejecución, y en este sentido presentamos un algoritmo no exacto basado en la descomposición de Benders.

Claves: Programación entera y combinatoria, Diseño de redes.

AMS: 90C11, 90C35

1. Introducción

En el Plan Estratégico de Infraestructuras y Transportes aprobado el pasado año por el Consejo de Ministros, el ferrocarril aparece como el eje principal, pues se tiene previsto desarrollar una ambiciosa Red de Altas Prestaciones que cubrirá todo el territorio nacional.

Según las directrices establecidas, las disposiciones legales aprobadas recientemente y las manifestaciones públicas realizadas por los responsables del Ministerio de Fomento, el escenario que se plantea en el ámbito de los ferrocarriles para el año 2020 tras la aplicación del PEIT será el siguiente:

1. Se habrá multiplicado por diez la red de Alta Velocidad, que pasará de los 1.031 kilómetros actuales a 10.000 kilómetros.
2. En los próximos 15 años se habrán construido 9.000 kilómetros de Alta Velocidad, frente a los 1.000 kilómetros construidos en los últimos 15 años.
3. El 90 % de la población estará a menos de 50 kilómetros de una estación de Alta Velocidad y todas las capitales de provincia tendrán acceso directo a la red de Alta Velocidad.

El objetivo genérico de nuestra propuesta será entonces, la implementación de una herramienta de soporte a la toma de decisiones para el diseño de un plan

general para la expansión de cualquier red de transportes, basada en objetivos similares a los anteriores marcados por el PEIT.

Metodológicamente esto se concreta en un algoritmo para la construcción de un modelo de red. En nuestro análisis utilizaremos, por tanto, herramientas de Programación Lineal Entera Mixta, resaltando su naturaleza multietápica, pues la expansión se realiza de forma progresiva.

Asimismo, en el trabajo realizamos un análisis comparativo de la red obtenida como solución óptima de nuestro programa entero, para el caso español y la propuesta del Ministerio de Fomento presentada en el PEIT. Tal modelo se ha implementado en XPRESS, software que nos facilita información gráfica de la red obtenida, haciendo posible comparar de forma visual las soluciones de ambas propuestas.

Otro de los objetivos de este trabajo es realizar un estudio computacional del modelo general variando diversos parámetros del mismo. Con este fin se ha implementado en CPLEX el modelo propuesto y reportamos los datos presentados en este estudio. Sin embargo, dado que estos problemas pertenecen a los problemas de complejidad algorítmica NP-dura, desarrollamos algoritmos alternativos que mejoren la eficiencia. En este sentido, hemos implementado un algoritmo basado en la descomposición de Benders ([2]). Presentamos este algoritmo alternativo, así como el análisis computacional comparativo con el algoritmo exacto resuelto con CPLEX, mostrándose que nuestro heurístico permite resolver instancias de estos problemas en menor tiempo y con suficiente acuracidad.

2. Bibliografía

- [1] Assad, A. A. (1980), Models for rail transportation, *Transpn. Res.-A*, 14A (3) 205-220.
- [2] Benders, J.F. (1962), Partitioning procedures for solving mixed integer variables programming problems. *Numerische Mathematik* 4 238-252
- [3] M. Grötschel, C.L. Monma, M. Stoer, (1995), Design of Survivable Networks in Magnanti, C.L. Monma, G.L. Nemhauser (eds), *Handbooks in Operations Research And Management Science*, 7: Network Models, North-Holland.
- [4] Hinojosa, Y., J. Puerto, F. R. Fernández, (2000), A multiperiod two-echelon multicommodity capacitated plant location problem. *European Journal of Operational Research* 123 271-291.
- [5] Puerto, J., A. Tamir, (2004), Locating tree-shaped facilities using the ordered median objective, *Mathematical Programming* 102 313-338.
- [6] Sprague, R.H. and H.J. Watson, (1986), *Decision Support Systems: Putting Theory into Practice*, Prentice-Hall.

El problema de la mediana ordenada con normas

 l_p

I. Espejo Miranda, A.M. Rodríguez-Chía, C. Valero Franco

Departamento de Estadística e Investigación Operativa
Universidad de Cádiz

Abstract

En este trabajo se presenta un procedimiento para resolver la versión convexa del problema de la mediana ordenada con distancias l_p . Dicho procedimiento, basado en una modificación del método del gradiente, genera una sucesión que desciende en valor objetivo y que converge a la solución óptima del problema para $p \in [1, 2]$. Finalmente, se muestra un análisis de los tiempos computacionales necesitados para resolver estos tipos de problemas.

Claves: Problema de la mediana ordenada, método del gradiente.

AMS: 90B85

1. Introducción

En los últimos años la obtención de modelos flexibles que se ajusten a diferentes tipos de situaciones reales con la máxima precisión ha atraído la atención de gran número de investigadores. En este sentido, el problema de la mediana ordenada presenta una estructura común a la mayor parte de los modelos clásicos de localización, además de proporcionar una formulación a modelos no considerados anteriormente en la literatura (ver [1, 2, 3] para más detalles).

En este trabajo presentamos un procedimiento de resolución al problema ordenado de la mediana donde las distancias vienen medidas a través de la norma l_p . Debido a la gran versatilidad del modelo, este procedimiento de resolución proporciona una herramienta robusta para abordar una gran variedad de problemas de localización continua.

2. El modelo

Sea $A = \{a_1, \dots, a_M\}$ un conjunto de puntos de demanda en el plano y $\Omega := \{\omega_1, \dots, \omega_M\}$ y $\Lambda := \{\lambda_1, \dots, \lambda_M\}$ dos conjuntos de escalares no negativos, con $\lambda_1 \leq \dots \leq \lambda_M$. Los elementos ω_i representan la importancia de los puntos de demanda a_i , $i \in \{1, \dots, M\}$ y los elementos de Λ diseñan la función objetivo.

Dada una permutación σ del conjunto $\{1, \dots, M\}$ verificando

$$\omega_{\sigma_1} \|x - a_{\sigma_1}\|_p \leq \dots \leq \omega_{\sigma_M} \|x - a_{\sigma_M}\|_p,$$

donde $\|\cdot\|_p$ denotan la norma l_p , se define

$$d_{(i)}(x) := \omega_{\sigma_i} \|x - a_{\sigma_i}\|_p.$$

La formulación del problema de la mediana ordenada viene dada por:

$$\min_{x \in \mathbb{R}^2} F(x) = \sum_{i=1}^M \lambda_i d_{(i)}(x). \quad (4.67)$$

Como los pesos λ vienen dados en orden no decreciente, la función objetivo $F(x)$ es convexa (ver [1] para más detalles). Se observa además que dicha función, dependiendo del punto evaluado, puede tener diferentes expresiones como suma ponderada de las distancias.

El objetivo de este trabajo consiste en desarrollar un procedimiento para resolver el problema dado en (4.67). Teniendo en cuenta que la función objetivo no es diferenciable en los puntos de demanda, se recurrirán a las aproximaciones hiperbólicas (ver [4] para más detalles). Dicho procedimiento está basado en una modificación del método del gradiente. La sucesión generada por el algoritmo propuesto desciende en valor objetivo y converge a la solución óptima del problema para $p \in [1, 2]$.

Finalmente, se muestra un análisis de los tiempos computacionales necesitados para resolver estos tipos de problemas.

3. Bibliografía

- [1] Rodríguez-Chía, A. M., Nickel, S., Puerto, J. and Fernández, F. R. (2000). *A flexible approach to location problems*. Mathematical Methods of Operations Research 51, 69-89.
- [2] Nickel, S., Puerto, J., Rodríguez-Chía, A. M., Weissler, A. (2005) *Multicriteria Planar Ordered Median Problems*. Journal of Optimization Theory and Applications, Vol.126 N. 3, 657-683.
- [3] Puerto, J., Fernández, F.R.(1995) *The symmetrical single facility location problem*. Prepublicación de la Facultad de Matemáticas. Universidad de Sevilla.
- [4] Frenk, J. B. G., Melo, M. T., Zhang, S. (1994) *A Weiszfeld method for a generalized l_p distance minisum location model in a continuous space*. Location Science, 2, 111-127.

El problema de localización con distancias esperadas

*Concepción Valero Franco¹, Antonio M. Rodríguez Chía,
Inmaculada Espejo Miranda*

Dpto. de Estadística e Investigación Operativa, Universidad de Cádiz

¹concepcion.valero@uca.es

Abstract

En este trabajo se considera un problema de localización donde la demanda se produce en una región con una determinada distribución de probabilidad. El objetivo es ubicar un servicio que minimice la distancia esperada en norma l_p al área de demanda. Para su resolución se presenta un algoritmo que converge a la solución óptima del problema para $p \in [1, 2]$. El trabajo concluye con un estudio de los tiempos computacionales.

Claves: Distancia esperada, Problema de Weber.

1. Introducción

De forma general un problema de localización consiste en localizar uno o varios servicios que satisfagan la demanda de un conjunto de puntos, cuya ubicación es conocida, optimizando alguna medida de efectividad. Sin embargo, existen situaciones donde se desconoce la posición exacta desde donde se produce la demanda. Este puede ser el caso de localizar una estación de bomberos que cubra una determinada área expuesta al riesgo de incendios o el caso de un centro de reparto de comida rápida, casos donde a priori no se sabe con exactitud dónde se va a producir el fuego o desde dónde se va a solicitar un pedido de comida. Una forma de modelar este tipo de situaciones reales con mayor exactitud es a través de los modelos de localización con demanda regional, es decir, modelos donde la demanda viene dada por una variable aleatoria con una determinada probabilidad. Dichos problemas han sido estudiados con anterioridad por Drezner (1986), Carrizosa et al. (1995-1998), Muñoz-Márquez (1995), Fernández-Palacín et al. (2003), entre otros, exponiendo algunos esquemas de solución.

2. El Modelo

Una posible formulación del problema planteado, donde las distancias son medidas con norma l_p , viene dado por:

$$\min_{x \in \mathbb{R}^n} F(x) = \bar{d}_p(x, \mathcal{A}) = \sum_{i=1}^M \frac{w_i}{\Delta(R_i)} \int_{R_i} \|x - r_i\|_p dr_i,$$

donde \mathcal{A} es la región de demanda, definida por un conjunto finito de rectángulos R_i y $\Delta(R_i)$ es el área del rectángulo R_i con peso asociado w_i . Un esquema iterativo para la resolución del mismo fue presentado en Fernández-Palacín et al. (2003).

En este trabajo se prueba que dicho esquema iterativo está bien definido para $p \in (1, 2]$ y que genera una sucesión convergente a la solución óptima del problema. Se proporciona una expresión analítica para el caso $p = 2$ que evita el cálculo de las integrales en cada iteración, obteniendo así mejores resultados computacionales al resolver este problema.

Debido a las indefiniciones que presenta el esquema propuesto para el caso $p = 1$, se presentan resultados que simplifican el cálculo de la distancia esperada en norma l_1 , proporcionando un esquema alternativo donde se evitan dichas singularidades. Para este nuevo procedimiento se demuestra la convergencia a la solución óptima del problema y se proporciona un estudio de los tiempos computacionales necesarios para resolverlo.

3. Bibliografía

- [1] E. CARRIZOSA, E. CONDE, M. MUÑOZ-MÁRQUEZ, AND J. PUERTO. The generalized Weber problem with expected distances. *Operations Research*, Vol.29, No.1: 35-57, 1995.
- [2] E. CARRIZOSA, M. MUÑOZ-MÁRQUEZ, AND J. PUERTO. The Weber problem with regional demand. *European Journal of Operational Research* 104: 358-365, 1998.
- [3] Z. DREZNER. Location of regional facilities. *Naval Research Logistic Quarterly*, 33: 523-529, 1986.
- [4] F. FERNÁNDEZ PALACÍN, M. MUÑOZ-MÁRQUEZ, A.M. RODRÍGUEZ CHÍA Y C. VALERO FRANCO. Resolución práctica del Problema de Weber. *XXVII Congreso Nacional de Estadística e Investigación Operativa*, Lleida 2003.
- [5] M. MUÑOZ-MÁRQUEZ. El Problema de Weber Regional. *Memoria de Tesis. Universidad de Sevilla*, 1995.
- [6] M. MUÑOZ-MÁRQUEZ, A.M. RODRÍGUEZ CHÍA Y C. VALERO FRANCO. El algoritmo de Weiszfeld en el problema de localización regional. *XXVI Congreso Nacional de Estadística e Investigación Operativa*, Ubeda, 2001.

Equilibrado de la línea de montaje del motor del modelo Pathfinder en la planta de NISSAN de Barcelona

Joaquín Bautista Valhondo¹, Jordi Pereira Gude²

¹joaquin.bautista@upc.edu, Nissan Chair UPC, Universitat Politècnica de Catalunya

²jorge.pereira@upc.edu, Dpto. Organització d'Empreses, Universitat Politècnica de Catalunya

Abstract

El presente trabajo se centra en la resolución del equilibrado de cargas entre las diferentes estaciones de la línea de montaje del motor del Nissan Pathfinder en la planta de Barcelona de Nissan Motor Ibérica. Se introduce el origen del problema y se exponen los resultados obtenidos.

Palabras Clave: Equilibrado de Líneas de Montaje, Optimización, Heurísticas

1. Introducción

En un sistema productivo con una o diversas líneas de montaje, el problema de equilibrado de las líneas aparece frecuentemente. Este es el caso de la industria de automoción. Generalmente, la asignación de tareas a las estaciones se realiza teniendo en cuenta los tiempos de proceso de las tareas, las relaciones de precedencia entre ellas y la tasa de producción. En nuestro problema, el equilibrado depende, además, de los espacios requeridos por los componentes a montar dispuestos en uno o en ambos lados de la línea y de los espacios disponibles en ella. Entre las causas más comunes de problemas de equilibrado que deben considerar restricciones de espacio, podemos destacar:

(1) Longitud de línea limitada. Cuando la unidad se mueve a lo largo de la línea con velocidad constante, usando una cinta transportadora por ejemplo, los operarios inician sus tareas en el punto inicial de su estación de trabajo y las efectúan en sintonía con el movimiento del producto; los operarios están limitados en sus desplazamientos hacia los elementos que contienen los componentes y en su recorrido a orillas de la línea. Esta limitación también aparece cuando el producto no se mueve y se transfiere de una estación a la siguiente, ciclo tras ciclo, en tiempo despreciable.

(2) Las herramientas y componentes requeridos deben estar distribuidos a lo largo de la línea. Es importante mantenerlos lo más cerca posible del lugar de trabajo, así como evitar su emplazamiento en doble fila. Adicionalmente, en la industria del automóvil, ciertas operaciones sólo pueden realizarse a un lado de la línea, condicionando el espacio físico donde las herramientas y los materiales pueden localizarse, lo cual limita más todavía el espacio disponible.

(3) En las líneas de modelos mixtos las restricciones espaciales son todavía más importantes. Cuando se resuelve el problema de equilibrado tradicional, es habitual usar una duración ponderada de las tareas, pero cada estación requerirá el espacio de los componentes y de las herramientas asociadas a cada variante de producto a montar en la estación.

(4) Otra fuente habitual de limitaciones de espacio procede de los cambios de producto. Cuando se lanza un nuevo modelo de coche y se mantiene la planta de producción aparecen nuevos requerimientos de espacios.

Los modelos clásicos de equilibrado de líneas de montaje [1] no tienen en cuenta dichas restricciones, lo que obliga a la elaboración de nuevos modelos y procedimientos de resolución.

2. Modelización y procedimiento de resolución

El modelo, [2], plantea la asociación de un requerimiento de espacio lineal a cada tarea, asociando un espacio disponible máximo a cada estación. Para su resolución se utiliza un algoritmo basado en colonias de hormigas que incorpora ideas procedentes de otros procedimientos para la resolución de problemas de equilibrado.

3. Resultados

Los resultados ofrecidos por el procedimiento para la instancia procedente de Nissan Spain muestran la mayor dificultad del problema al añadir las restricciones de espacio aumentando en cuatro el número de estaciones necesarias para cumplir con las nuevas restricciones.

4. Agradecimientos

Agradecemos a Spanish Operation of Nissan y a la Cátedra Nissan UPC el apoyo dado al presente trabajo. Este trabajo ha sido parcialmente financiado por el proyecto PROTHIUS DPI2004-03475 del Gobierno de España.

5. Bibliografía

- [1] Scholl, A. (1999). *Balancing and Sequencing of Assembly Line*. Physica Verlag. Heidelberg.
- [2] Bautista, J.; Pereira J. (2006). Ant algorithms for a time and space constrained assembly line balancing problem. *European Journal of Operational Research* disponible en: doi:10.1016/j.ejor.2005.12.017

Equilibrado de Líneas con Alternativas de Montaje

Liliana Capacho Betancourt¹, Rafael Pastor Moreno²

¹liliana.capacho@upc.edu, Departamento de I.O. y CESIMO, Facultad de Ingeniería, Universidad de Los Andes, Mérida, Venezuela

²rafael.pastor@upc.edu, ^{1,2}Instituto de Organización y Control de Sistemas Industriales, UPC, Barcelona, España

Abstract

Este trabajo presenta una nueva variante del problema de equilibrado de líneas: se dispone de diagramas de precedencias alternativos que definen distintas posibilidades de proceso de montaje de un producto. Para resolver este problema se desarrolla un modelo de programación matemática que considera, de manera simultánea, la selección de una alternativa de montaje y el equilibrado de la línea.

Palabras Clave: Equilibrado de líneas de montaje

1. Introducción

Los problemas de equilibrado de líneas consisten básicamente en asignar un conjunto de tareas a un grupo de estaciones de trabajo, tratando de optimizar cierta medida de eficiencia (ej. minimizar el número de estaciones). La mayoría de los problemas tratados en la literatura consideran el caso más simple en el que se asume que se dispone de una única línea, se procesa un único producto, todas las estaciones son idénticas y no se tienen otras restricciones a parte de las relaciones de precedencia (véase por ejemplo [1] y [2]). Las referencias que tratan casos más generales, como por ejemplo problemas con estaciones diferentes, líneas paralelas o líneas mixtas, son más escasas en la literatura (ver [3]). Aún más limitados son los trabajos que incorporan restricciones como incompatibilidades entre las tareas o restricciones de capacidad, o que consideran tiempos de proceso dependientes de la secuencia, alternativas de equipos de manufactura o de proceso.

Por otra parte, es habitual que en los problemas de equilibrado de líneas se considere un diagrama de precedencias único y predeterminado, que representa a todas las relaciones entre las tareas de montaje. Sin embargo, en problemas de la realidad industrial se presentan casos en los que se dispone de variantes de montaje que son alternativas entre sí. En este trabajo se define y formaliza un nuevo problema de equilibrado de líneas, denominado ASALBP (*Alternative Subgraphs Assembly Line Balancing Problem*), en el que se consideran las variantes alternativas que puede admitir un proceso de montaje de un producto. En este caso, cada alternativa consiste en un orden de proceso particular de un subconjunto de tareas. El problema también considera que los tiempos de proceso de las tareas no son fijos sino dependientes de la alternativa de montaje seleccionada.

2. Problema de equilibrado con alternativas de montaje (ASALBP)

El problema ASALBP implica resolver simultáneamente la selección de una alternativa de montaje, que determina la forma en que los productos son manufacturados y los tiempos de proceso de las tareas, así como el equilibrado de la línea, que asigna las tareas a las estaciones. Habitualmente, este problema se suele resolver seleccionando a priori la alternativa de montaje y posteriormente equilibrando la línea. Este procedimiento puede sub-optimizar el sistema porque no considera las repercusiones que podría tener cada alternativa de montaje sobre el equilibrado de la línea. Considerense los resultados óptimos que se muestran en la tabla 1 obtenidos al equilibrar un proceso con dos alternativas de montaje. En dicho ejemplo posiblemente se habría seleccionado a priori la alternativa 1, por ser la que requiere el menor tiempo total de proceso; sin embargo, si se considera la selección de la alternativa de montaje junto al equilibrado de la línea, la alternativa 2 proporciona una mejor solución, puesto que sólo requiere 3 estaciones en lugar de las 4 que requiere la alternativa 1.

Alter- nativa	Tareas por estación (tiempo de estación)				Tiempo total de proceso	Número de estaciones
	I	II	III	V		
1	A (11)	B,C (9)	D,E (14)	F (7)	41	4
2	A,C (15)	D,E (14)	B,F (14)	-	43	3

Tabla 1: Resultados de equilibrar dos problemas de montaje alternativos

3. Modelo matemático del problema de equilibrado con alternativas de montaje

Para resolver el problema ASALBP se ha desarrollado un modelo de programación matemática, en el que cada alternativa es referida como una ruta global de montaje y está representada por un subgrafo de precedencias. Así, se dispone de tantas rutas globales como combinaciones posibles de los subgrafos de montaje disponibles. El modelo considera que los tiempos de proceso de las tareas son dependientes de la ruta seleccionada, por lo que se puede presentar un tiempo total de proceso diferente para cada ruta alternativa. El modelo, que intenta minimizar el número de estaciones dado el tiempo ciclo, selecciona la ruta global de montaje a la vez que asigna las tareas a las estaciones de trabajo. A través de un experimento computacional, usando ILOG CPLEX 9.0 y generando ejemplares de problemas que contemplen rutas de montaje, se muestra que únicamente se pueden resolver en tiempos razonables problemas de tamaño pequeño.

4. Bibliografía

- [1] Baybars, I. (1986). A survey of exact algorithms for the simple assembly line balancing problem. *Management Science*, 32, 909-932.
- [2] Becker, C. and Scholl, A. (2006). A survey on problems and methods in generalized assembly line balancing. *European J. of Op. Res.*, 168, 694-715.
- [3] Scholl, A. and Becker, C. (2006). State-of-the-art exact and heuristic solution procedures for simple assembly line balancing. *Eur. J. of Op. Res.*, 168, 666-693.

Equilibrado de líneas con setup. Un modelo matemático

Rafael Pastor¹, Carlos Andrés², Cristóbal Miralles³

¹rafael.pastor@upc.es, Instituto de Organización y Control Industrial,
Universidad Politécnica de Cataluña

²candres@omp.upv.es, ³cmiralles@omp.upv.es,
Departamento de Organización de Empresas, Universidad Politécnica de
Valencia

Abstract

En este artículo se presenta un modelo de Programación Entera Mixta para una nueva variante del problema de equilibrado de líneas monodelo donde se aborda la posibilidad de que existan tiempos de ajuste entre las tareas a asignar.

Palabras Clave: assembly line balancíng problem, setup

AMS: 68M20

1. Introducción

El problema de equilibrado líneas es un problema clásico de diseño de sistemas de fabricación en flujo que ha sido estudiado desde hace más de 40 años. El objetivo de equilibrar una línea es la asignación de un conjunto de tareas ligadas entre sí por relaciones de precedencia a estaciones de trabajo de tal forma que no se sobrepease un límite de asignación marcado por el tiempo de ciclo de la línea, adicionalmente es conveniente que la suma de los tiempos de las tareas asignadas a cada estación sea lo más parecida posible.

En las configuraciones en línea o flujo, las tareas se deben de desarrollar sucesivamente y conducen a la obtención de un producto final. En este tipo de problemas, las relaciones de precedencia entre las tareas imponen un orden parcial, y obligan a que algunas tareas deban ser completadas antes de otras. Dentro de cada estación el producto debe de permanecer un tiempo máximo determinado por el tiempo de ciclo de la línea. Cuando se parte de las hipótesis de un único producto a fabricar y todos los datos son conocidos se denomina problema simple de equilibrado de líneas SALBP.

Cuando no se cumplen las hipótesis simplificativas del SALBP se plantea el problema de equilibrado general o GALBP. El problema GALBP ha sido menos estudiado que SALBP, aunque es más realista.

El problema presentado en este escrito añade ciertos supuestos basados en la consideración de que existen tiempos de preparación dependientes entre cada par de tareas. Así, entre cada par de tareas asignadas a una estación

aparece un tiempo de preparación dependiente de las dos tareas que puede corresponder a la situación real en donde aparece un tiempo para dejar una herramienta y preparar la siguiente de tal forma que este tiempo dependa de las operaciones realizadas. En el siguiente apartado se describe brevemente el modelo matemático desarrollado.

2. Modelo matemático

En el trabajo completo se presentará el modelo matemático desarrollado así como las experiencias computacionales planteadas. En el modelo se trata de minimizar el número de estaciones utilizadas sujeto a las siguientes restricciones:

- La primera serie de restricciones imponen que toda tarea debe asignarse y a una única posición de la secuencia de una única estación.
- A continuación se plantea unas restricciones para obligar a que, en cada posición de la secuencia de toda estación, como máximo se asigne una tarea.
- Después se consideran las restricciones que imponen que las tareas se asignen en posiciones crecientes (en orden) en las secuencias de las estaciones y otras que imponen la condición de precedencia entre parejas de tareas, precedencias referentes a las estaciones donde son asignadas y, si las tareas se asignan a la misma estación, referentes a las posiciones en la secuencia.
- El siguiente juego de restricciones aseguran que el tiempo de trabajo en las estaciones más el tiempo de cambios no sea mayor que el tiempo ciclo (el primer sumando es el tiempo de trabajo, el segundo es el tiempo de cambio, sin el cambio de ciclo, y el tercero es el tiempo de cambio entre la última tarea de un ciclo y la primera de ciclo siguiente).
- Las últimas restricciones obligan a que se cumplan las relaciones de precedencia entre las tareas asignadas a la estación.

3. Planteamiento del trabajo

El modelo matemático ha sido resuelto mediante CPLEX. En el trabajo completo se mostrarán los resultados y conclusiones sobre el estudio realizado.

4. Agradecimientos

Este trabajo ha sido realizado en el marco del proyecto DELIMER:DISEÑO Y EQUILIBRADO DE LÍNEAS DE MONTAJE EN ENTORNOS REALISTAS (DPI2004-03472).

Esqueletos Orientados a Objeto para la técnica de Ramificación y Acotación

Isabel Dorta¹, Coromoto León¹, Carlos Segura¹

¹isadorta@ull.es, cleon@ull.es, carlossegurag@gmail.com, Departamento de Estadística, Investigación Operativa y Computación, Universidad de La Laguna

Abstract

La librería MaLLBa proporciona una interfaz de usuario y esqueletos para la técnica de Ramificación y Acotación en los lenguajes de programación C++ y Java. En este trabajo se describe la interfaz de usuario. Como caso de estudio, se presenta el problema de la Mochila 0/1 y se realiza un estudio del rendimiento de las implementaciones con ambos lenguajes.

Claves: Ramificación y acotación, Programación Orientada a Objetos.

AMS: 68Q01

1. Descripción del esqueleto

La librería MaLLBa [1] proporciona esqueletos de técnicas algorítmicas para la resolución de problemas de optimización. Este trabajo se centra en el esquema proporcionado para el paradigma de Ramificación y Acotación: *MaLLBa::BnB*. En la implementación se pueden distinguir: *Interfaces* y *Clases*. Las *Interfaces* definen lo que deben implementar los usuarios de la librería para caracterizar su problema (interfaces requeridas), así como lo que deben implementar las clases que resuelven el esquema, y que usarán los usuarios finales (interfaces proporcionadas). La *interfaces* requeridas son las siguientes: La interfaz *Problem* permite definir un problema. La interfaz *SubProblem* se usa para representar cada uno de los subproblemas en que se irá dividiendo el problema. Los métodos que se han de implementar son: *Solution solve(Problem p)*, *void branch(Problem sp, Solver s)*, *double lower_bound(Problem p)* y *double upper_bound(Problem p)*. La interfaz *Solution* debe ser implementada por la clase que representa la solución al problema. La única interfaz proporcionada es *Solver*. Los métodos de esta interfaz son: *Solution run()* y *void insert(SubProblem sp)*. Las *Clases* proporcionadas son las clases que implementan los métodos que resuelven el problema. Las clases proporcionadas son *SolverSeqStack* y *SolverSeqQueue*. Estas clases resuelven el problema de ramificación y acotación a través de un esquema secuencial, usando un recorrido en profundidad y un recorrido en anchura, respectivamente.

núm. obj.	1000	5000	10000	50000
Java	0,812	21,117	106,84	-
Java-O1	1,101	4,669	183,51	-
Java-O2	0,185	0,498	3,028	44,967
C++	1,897	40,980	197,18	5728
C++-O2	0,234	4,830	24,594	1144,4

Cuadro 1: Resultado para el Problema de la Mochila 0/1

2. Conclusiones

Los resultados que se presentan en la tabla 1 son para el Problema de la Mochila 0/1. La ejecuciones se han realizado en una máquina con un Procesador AMD Duron a 750 MHz, 256 Mbyte de memoria y 32 Gbyte de disco duro. El sistema operativo es Debian Linux versión 2.2.19. El compilador de C++ usado es GNU gcc versión 2.7.2.3 y el de Java J2SE 5.0. La librería `Mallba::BnB` proporciona varios esqueletos en C++ y Java. En la fila denominada Java se muestran los tiempos de ejecución para el esqueleto Java que realiza un recorrido en anchura del espacio de soluciones. Las filas con etiquetas C++ y C++-O2 muestran los tiempos de ejecución para un esqueleto C++ que implementa el mismo algoritmo que el Java, en el primer caso compilado sin optimizaciones y en segundo con optimizaciones hasta nivel 2. En la fila Java-O1 se muestran los resultados de un esqueleto Java en el que se hace un recorrido en profundidad y se modifica la forma en que se calculan las cotas inferiores, y en la fila Java-O2 se realiza una única copia de la solución en la ramificación de un subproblema en dos subproblemas “hijos” y se copia la solución en el momento de analizar el subproblema. Nótese que las versiones en Java son menos escalables, presentando más problemas al aumentar el tamaño de los datos, por ello se introducen una serie de mejoras que permiten realizar menos cálculos, y disminuir el consumo de memoria.

3. Agradecimientos

Este trabajo se ha realizado con fondos del proyecto de TIC2005-08818-c04-04 del ‘Plan Nacional de I+D+i’ del Ministerio de Educación, Cultura y Deportes. El trabajo de Carlos Segura ha estado subvencionado por una beca de Colaboración del curso 2005-2006.

4. Bibliografía

- [1] E. Alba, et. al. (2002). MALLBA: A library of skeletons for combinatorial optimisation. *Proceedings of the Euro-Par. Lecture Notes in Computer Science. Springer-Verlag* 2400, 927-932.

Estimación de Índices de Eficiencia de Bibliotecas Universitarias mediante Análisis Envolvente de Datos

J. García-Laguna¹, B. Salvador², J. A. Tapia³

¹laguna@eio.uva.es, ²bosal@eio.uva.es, ³tapia@eio.uva.es

Dpto. de Estadística e Investigación Operativa, Univ. de Valladolid

Abstract

En este trabajo se analiza la eficiencia de las bibliotecas universitarias españolas en distintos escenarios, elegidos considerando el tamaño de la universidad, los inputs y los outputs utilizados. Los índices de eficiencia, las holguras y los índices de supereficiencia se han calculado usando los modelos BCC y FDH del DEA. Los datos se han tomado del Anuario Estadístico de REBIUN para el año 2004.

Palabras Clave: DEA, eficiencia, ítem, input, output, biblioteca

AMS: 90C90

1. Introducción

Actualmente se están realizando numerosos esfuerzos para evaluar la calidad en los servicios públicos, en particular en el ámbito de la educación. De un modo general, puede decirse que para obtener información sobre la calidad de los servicios que presta la Universidad ó detectar los puntos fuertes y débiles de la institución, así como de sus posibilidades y propuestas de mejora, es necesario disponer de información actualizada y relevante sobre el sistema. Las bibliotecas, en general, y las universitarias en particular, recogen datos en relación a sus fondos bibliográficos, infraestructuras, servicios prestados, etc., con el objeto de resumir su funcionamiento.

Partiendo de esos datos, en este trabajo nos hemos centrado en medir el rendimiento relativo de una biblioteca con indicadores de naturaleza estrictamente cuantitativa. Más concretamente, dado que empleamos el enfoque del Análisis Envolvente de Datos (DEA), analizamos el servicio de bibliotecas desde una perspectiva de items (unidades a analizar), inputs (recursos introducidos en el sistema) y outputs (servicios prestados por las bibliotecas). Para tal fin, nos hemos apoyado en los datos publicados por REBIUN en el anuario de 2004. Es de destacar que toda esta información se recoge con certificación de veracidad, avalada por el responsable del servicio de bibliotecas de cada Universidad.

El intercambio de opiniones con profesionales del servicio de bibliotecas ha conducido a la necesidad de construir diversos escenarios. En cada uno de ellos

se ha medido la eficiencia, las holguras y la supereficiencia de cada universidad. El escenario más amplio ha sido el constituido por todas las universidades. Otros escenarios han surgido de la estratificación de las universidades por su tamaño (número de usuarios potencial) con el objeto de aumentar la homogeneidad, así como de la decisión de qué inputs y outputs considerar. En este trabajo utilizamos los modelos BCC y el FDH. El primero por ser el más ampliamente usado en la literatura y el segundo por su sencillez y facilidad de interpretación.

2. Agradecimientos

Este trabajo ha sido parcialmente subvencionado por el proyecto OTP/05/16 de la Consejería de Hacienda del Gobierno Regional de la *Comunidad Autónoma de Castilla y León*.

3. Bibliografía

- [1] Álvarez Pinilla A. (2001). La medición de la eficiencia y la productividad. Pirámide.
- [2] Andersen P. y Petersen N. C. (1993). A procedure for ranking efficient units in Data Envelopment Analysis. Management Science 39, 10, 1261-1264.
- [3] Chen T-Y. (19997). An evaluation of the relative performance of university libraries in Taipei. Library Review, 46, 3, 190-201.
- [4] Cooper W.W., Seiford L. M. y Tone K. (2000). Data Envelopment Analysis. Kluwer Academic Publishers.
- [5] REBIUN (2003). Anuario estadístico de las bibliotecas universitarias y científicas españolas 2001. Madrid: Servicio de Biblioteca UAM. (Documentos de trabajo REBIUN).
- [6] Stancheva N. y Angelova V. (2004). Measuring the efficiency of university libraries usin Data Envelopment Analysis. INFORUM 2004. Praga.
- [7] Sutter, Eric. "La démarche qualité en bibliothèque: questiones-réponses". Bulletin des Bibliothèques de France, 1998; 43(1):20-23.

Estudio Experimental sobre el Comportamiento de Reglas de Despacho en un Taller de Máquinas Paralelas no Relacionadas con Llegada Dinámica de Trabajos y Setup Dependiente de la Secuencia

Larissa Stania¹, Pedro Gómez¹, Carlos Andrés¹

¹pgomez@cigip.upv.es, Centro de Investigación de Gestión e Ingeniería de la Producción (CIGIP), Universidad Politécnica de Valencia

Abstract

En este trabajo se ha realizado el estudio experimental sobre un caso real de secuenciación de trabajos en un taller cerámico con máquinas paralelas no relacionadas usando reglas de despacho. Se ha establecido como objetivo identificar la influencia de diversas variables sobre el rendimiento del taller.

Palabras Clave: Secuenciación, Maquinas Paralelas, Dinámico.

AMS: 68M20

1. Introducción

Este trabajo está basado en datos de un caso real de la industria cerámica. En concreto se centra en la etapa de prensado-secado, correspondiente a la primera fase en la producción de pavimento y revestimiento cerámico. Esta etapa está caracterizada por la llegada de trabajos de forma dinámica con tiempos de proceso conocidos y tiempos de cambio de partida o setup dependientes de la secuencia. El objetivo es identificar la influencia del estado del sistema productivo y las características de los órdenes de fabricación sobre los resultados obtenidos en la función objetivo.

2. Estado del Arte

Desde [4] en cuya comunicación se realizó la primera prueba de competitividad a un algoritmo dinámico para un problema de programación, hasta [6] en cuyo artículo se realiza un profundo análisis de los problemas de programación en línea, el número de referencias que se han ocupado del problema de programación dinámica con y sin incertidumbre en un modelo de taller de máquinas paralelas no es demasiado extenso.

Casi no existen referencias que además del dinamismo contemplen el caso de las máquinas paralelas no relacionadas. [1][5] se aproximan al problema proponiendo diversas soluciones. No obstante, ninguno de ellos ha enfocado el problema teniendo en cuenta las características reales de un problema como el de la industria cerámica.

3. Estudio Experimental

El estudio experimental ha consistido en la simulación del comportamiento de la etapa productiva de prensas-secado, asumiendo las restricciones de [2], realizando el análisis dos medidas de eficiencia regular: La media del retraso \bar{T} y la varianza del mismo $\sigma^2 T$. Previamente, se ha realizado un estudio sobre la influencia de la consideración o no de los tiempos de setup en los cálculos de la reglas de despacho. Las secuencias se han generado usando el algoritmo de [3]. Se han analizado tres factores la premura en la fecha de entrega, la tasa de llegada de trabajos (número de trabajos) y las reglas de despacho usadas (SPT, LPT, AT-RPT, EDD, MDD, FIFO, RANDOM, SLACK, SLACK ratio, CR, S/RMWK y SST).

4. Conclusiones

El estudio preliminar ha demostrado las reglas de despacho ofrecen mejores resultados sobre \bar{T} y $\sigma^2 T$ si se incluyen en sus cálculos los tiempos de setup de forma explícita. Esta influencia es mayor cuando el número de trabajos es elevado.

Una vez modificadas las reglas de despacho para incluir el tiempo de setup dentro del proceso de cálculo, se ha observado que \bar{T} y $\sigma^2 T$ están fundamentalmente influidas en primer lugar por la fecha de entrega y en segundo lugar por el número de trabajos en el sistema. No se observan diferencias significativas entre las diversas reglas de despacho para \bar{T} y siendo estas mayores $\sigma^2 T$. SST y SPT presentan los mejores resultados.

5. Agradecimientos

Este estudio se ha realizado dentro del marco de la Red de Investigación subvencionada por la Generalitat Valencia en Project Management and Scheduling (PMSVALNET)- IIARC0/2004/40.

6. Bibliografía

- [1] Azar, Y. y Epstein, L. (2002), On-line scheduling with precedence constraints. *Discrete Applied Mathematics*, 119(1-2): 169-180.
- [2] Conway, R. W., Maxwell, W. L., y Miller, L. W. (1967), “Theory of Scheduling”, Addison –Wesley.
- [3] Giffler B. y Thompson G. (1960): “Algorithms for solving production scheduling problems”, *Operacitions Research*, 8, 487-503.
- [4] Graham, R. L. (1969) Bounds on Multiprocessing Timing Anomalies. *Siam Journal on Applied Mathematics*, 17(2): 416.
- [5] Megow,N., Uetz,M., y Vredeveld,T. (2004c) Stochastic Online scheduling on parallel machines. *Approximation and Online Algorithms*, 3351: 167-180.
- [6] Sgall, J. (1998), On-line scheduling. *Online Algorithms*, 1442: 196-231.

Evaluación y optimización de recursos sanitarios en el Hospital San Juan de Dios de Pamplona, Colombia

Aurora Inés Gáfaro Rojas¹, Cristina Azcárate Camio², M^a Luisa Eraso Goicoechea³

¹gafaro.44150@e.unavarra.es, ²cazcarate@unavarra.es, ³leraso@unavarra.es,
Departamento de Estadística e Investigación Operativa, Universidad Pública
de Navarra

Abstract

Este trabajo analiza un problema de gestión de recursos sanitarios en el Hospital San Juan de Dios de Pamplona, Colombia. Para su resolución se emplean técnicas de optimización con simulación.

Palabras Clave: Servicios de salud, Simulación, Optimización

AMS: 90B90, 37M05, 80M50

1. Introducción

Este trabajo analiza un problema relacionado con la gestión de recursos sanitarios. Algunas publicaciones sobre problemas de diseño y gestión de sistemas de salud se recogen en [1], [2] y [4].

El estudio está relacionado con el Hospital San Juan de Dios de Pamplona, Colombia, donde se proporciona una amplia gama de servicios médicos:

- El área de Servicios Ambulatorios integra: consulta médica general, consulta odontológica general, consulta especializada, anestesiología, cirugía general, ginecología y obstetricia, medicina interna, pediatría, neurología, oftalmología, ortopedia, neurocirugía y dermatología.
- En el área de Servicios de 24 horas cabe señalar: atención de urgencias, laboratorio clínico, imaginología (Rayos X), servicio de ecografías, hospitalización, especialidades (pediatría gineco-obstetricia, medicina interna, quirúrgicas).
- Otros servicios: trabajo social, intervención en salud y medio ambiente, educación en salud, actividades de detección temprana y protección específica.

Los datos actuales sobre el Hospital indican un aumento de la demanda de estos servicios y una falta de capacidad para atenderla, lo que está ocasionando

excesivas colas y un incremento considerable en los tiempos de permanencia en el sistema.

Los pacientes que acuden a este centro, para ser atendidos, tienen que pasar por una oficina de facturación, para asignarles una cita y para cancelar el correspondiente copago. La segunda cola se genera en el archivo, para buscar la historia clínica y la tercera cola se genera en la sala de espera de los consultorios.

El objetivo de este estudio es proponer cambios en la planificación de recursos en este centro, para disminuir las colas y el tiempo permanencia de los pacientes en el sistema.

Para ello, se recogió y analizó la información existente en el hospital y la recopilada mediante encuestas a los trabajadores del centro. Se representó el sistema mediante un modelo matemático dinámico y multiobjetivo. Se combinaron técnicas de optimización con simulación de eventos discretos ([3]) para la evaluación de diferentes escenarios y la determinación de la mejor configuración.

La alternativa propuesta mejora la gestión de los recursos humanos y materiales, reduce significativamente el tiempo de espera en cola de los pacientes, los tiempos de atención de las consultas y, por tanto, el tiempo de permanencia de los usuarios en el sistema.

2. Bibliografía

- [1] Brailsford, S.; Schmidt, B. (2003). Towards incorporating human behaviour in models of health care systems: An approach using discrete event simulation. *European Journal of Operational Research* 150, 19-31.
- [2] Taylor, S.; Kuljis, J. (1998). Simulation in health care management: modelling an outpatient clinic. An example of who simulation can assist in the improvement of health care management. *OR Insight* vol 11, Issue 3.
- [3] Vanda de Angelis; Felici, G.; Impelluso, P. (2003). Integrating simulation and optimisation in health care centre management. *European Journal of Operational Research* 150, 101-114.
- [4] Walczak, S.; Pofahl, W. E.; Scorpio, R. J. (2002). A decision support tool for allocating hospital bed resources and determining required acuity of care. *Decision Support Systems* 34, 435-446.

Exceso de Información en Optimización Lineal Paramétrica

Miguel Ángel Goberna, Valentín Jornet, Mariola Molina¹

Departamento de Estadística e Investigación Operativa, Universidad de
Alicante

¹mariola.molina@ua.es

Claves: Programación Semi-Infinita Lineal, Programación Lineal, Redundancia

AMS: 90C34;90C05

Abstract

Consideramos un problema de optimización lineal paramétrica y su correspondiente problema dual. El objetivo es caracterizar las restricciones primales y las variables duales que pueden ser eliminadas sin modificar el conjunto factible, el conjunto óptimo o el valor de los respectivos problemas, así como encontrar las relaciones entre ellas.

1. Introducción

Consideramos un sistema lineal en \mathbb{R}^n , $\sigma = \{a'_t x \geq b_t, t \in T\}$, donde T es un conjunto arbitrario de índices con cardinalidad $2 \leq |T| \leq \infty$. Asociamos con σ los dos siguientes problemas de optimización lineal paramétrica, uno dual del otro,

$$P(c) : \text{Inf } c'x \quad \text{s.a. } a'_t x \geq b_t, \quad t \in T$$

$$D(c) : \text{Sup } \Psi(\lambda) := \sum_{t \in T} \lambda_t b_t \quad \text{s.a. } \sum_{t \in T} \lambda_t a_t = c, \quad \lambda \in \mathbb{R}_+^{(T)},$$

donde el parámetro es $c \in \mathbb{R}^n$ mientras que $X^{(T)}$, con $0 \in X \subset \mathbb{R}$, denota el conjunto de aplicaciones $\lambda : T \rightarrow X$ que se anulan en todo $t \in T$ excepto a lo sumo en un conjunto finito de T , $\text{supp } \lambda = \{t \in T \mid \lambda_t \neq 0\}$, llamado soporte de T . Denotamos por F , $F^*(c)$ y $v(c)$ el conjunto factible, el conjunto óptimo y el valor óptimo de $P(c)$, y denotamos por $\Lambda(c)$, $\Lambda^*(c)$ y $v^D(c)$ el conjunto factible, el conjunto óptimo y el valor óptimo de $D(c)$, respectivamente.

Consideramos el conjunto factible de $P(c)$, F (que puede ser visto como una multifunción constante), las multifunciones $F^* : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ y Λ , $\Lambda^* : \mathbb{R}^n \rightrightarrows \mathbb{R}_+^{(T)}$, y las funciones ordinarias v , $v^D : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$. De igual forma, dado un índice $s \in T$, asociamos con su *sistema relajado* $\sigma_s := \{a'_t x \geq b_t, t \in T \setminus \{s\}\}$ sus correspondientes problemas paramétricos, $P_s(c)$ y $D_s(c)$, y sus correspondientes aplicaciones: F_s , F_s^* , Λ_s , Λ_s^* , v_s y v_s^D .

2. Principales conceptos y resultados

Decimos que una restricción $a'_s x \geq b_s$ (o la variable λ_s) es superflua con respecto a alguna de las aplicaciones asociadas con $P(c)$ ($D(c)$, respectivamente) si su eliminación no modifica la correspondiente aplicación. Por simplificar, decimos entonces que $s \in T$ es superfluo en ambos casos.

En particular, s es *superfluo con respecto al conjunto factible primal* si $F_s = F$ (i.e., s es redundante), s es *superfluo con respecto al valor primal* (PVS) si $v_s = v$, y s es *superfluo con respecto al conjunto óptimo primal* (POS) si $F_s^* = F^*$. De la misma forma, diremos que s es *superfluo con respecto al valor óptimo dual* (DVS) si $v_s^D = v^D$ y s es *superfluo con respecto al conjunto óptimo dual* (DOS) si $\overline{\Lambda}_s^* = \Lambda^*$, donde $\overline{\Lambda}_s^* := \left\{ \overline{\gamma} \in \mathbb{R}_+^{(T)} \mid \gamma \in \Lambda_s^* \right\}$, con

$$\overline{\gamma}_t = \begin{cases} \gamma_t, & t \in T \setminus \{s\}, \\ 0, & t = s, \end{cases}$$

no existiendo índices superfluos con respecto al conjunto factible dual.

Las restricciones e índices superfluos están relacionados también con otros índices que proporcionan información no necesaria, como son los índices *inesenciales* y los índices *extraños*.

Nuestro trabajo consiste en la caracterización de estas ocho clase de índices, así como en encontrar las relaciones existentes entre ellos.

En relación a este último objetivo, destacamos que si $|T| < \infty$ y σ es consistente, existen dos grupos de propiedades equivalentes:

(A) índices redundantes, PVS, POS, DVS, y extraños.

(B) índices fuertemente redundantes, DOS, fuertemente extraños, no activos, e inesenciales (añadiendo la condición de que F no contenga rectas).

3. Bibliografía

- [1] Karwan, M., Lofti, V., Telgen, J., and Zionts, S., 1983, *Redundancy in Mathematical Programming*. (Berlin: Springer).
- [2] Goberna, M.A., Jornet, V., and Molina, M., 2003, Saturation in Linear Optimization. *Journal of Optimization Theory and Applications*, 117(2), 327-348.
- [3] Goberna, M.A., Jornet, V., and Molina, M., 2006, Excess information in Parametric Linear Optimization, *Optimization* (aceptado, pendiente de publicación).

¹This work is partially supported by the Ministry of Science and Technology, Spain, and the funding for regional research DPI2004-03472

Extended Classification for Flowshops with Resequencing

Gerrit Färber, Anna M. Coves Moreno

Instituto de Organización y Control de Sistemas Industriales (IOC)

Universidad Politécnica de Cataluña (UPC), Barcelona, Spain

gerrit_faerber@gmx.de, anna.maría.coves@upc.es

Abstract

Este trabajo presenta una clasificación extendida de líneas de flujo no-permutación. Se consideran las múltiples opciones que se presentan al incluir la posibilidad de reseguenciar piezas en la línea. Se ha visto que en la literatura actual no se ha clasificado con profundidad este tipo de producción.

This paper presents an extended classification for non-permutation flowshops. The versatile options which occur with the possibility of resequencing jobs within the line are considered. The literature review shows that no classification exists which considers extensively this type of flowshop.

Keywords: Classification, Non-permutation Flowshop, Resequencing

1. Introduction

In the classical flowshop M stations are arranged in series, according to the technological sequence of the operations and a set of N jobs has to be processed on these stations, see e.g. [3]. Each of the N jobs has the same ordering of stations for its processing sequence. Each job can be processed on one and only one station at a time, each job is processed only once on each station, and each station can process only one job at a time. Jobs may bypass other jobs only between stations.

The above definition gives a general frame for the classification of non-permutation flowshops, but due to the numerous variations which occur in real production lines, a more extended classification is required. There exist various classifications, as for example Pinedo [7], Herrmann et al. [2], and Lageweg et al. [4], none of which considers exhaustively the possibilities of resequencing jobs within the production line, such as: using large buffers (Automatic-Storage-and-Retrieval-System) which decouple one part of the line from the rest of the line [6]; buffers which are located off-line [5]; hybrid or flexible lines [1]; and more seldomly, the interchange of job attributes instead of physically changing the position of a job within the sequence [8].

2. Classification

The classification of non-permutation flowshops, proposed in this paper, is based on the classification scheme by Pinedo, [7], and is intended to

adequately classify flowshop problems with particular consideration of options and restrictions which appear when resequencing of jobs takes place. The notation contains the triplet $\alpha|\beta|\gamma$ and helps classifying sequencing and scheduling problems. The triplet determines a problem as: α describes the station environment; β provides details on characteristics and constraints for the processing of the jobs; γ contains information on the objectives of the optimization.

Even though references on specific characteristics are given, this paper does not pretend to provide an exhaustive survey on solution techniques. It is intended to generate an instrument for properly categorizing the diversity of flowshop problems, in order to simplify their comparison and to improve the possibilities to finding new configurations that are not yet investigated and may lead to further optimization.

3. Bibliography

- [1] T. Engström, D. Jonsson, and B. Johansson. Alternatives to line assembly: Some swedish examples. *International Journal of Industrial Ergonomics*, 17(3):235–245, 1996.
- [2] J.W. Herrmann, L. Chung-Yee, and J.L. Snowdon. A classification of static scheduling problems. *Complexity in Numerical Optimization*, P.M. Pardalos, Editor, World Scientific Publishing Co., Singapore, 203-253, 1993.
- [3] S.M. Johnson. Optimal two and three stage production schedules with setup times included. *Naval Research Logistics Quaterly*, 1(1):61–68, 1954.
- [4] B.J. Lageweg, J.K. Lenstra, and K. Rinnooy. A general bounding scheme for the permutation flowshop problem. *Operations Research*, 26:53–67, 1978.
- [5] M. Lahmar, H. Ergan, and S. Benjaafar. Resequencing and feature assignment on an automated assembly line. *IEEE Transactions on Robotics and Automation*, 19(1):89–102, 2003.
- [6] H.F. Lee and S.K. Schaefer. Sequencing methods for automated storage and retrieval systems with dedicated storage. *Computers and Industrial Engineering*, 32(2):351–362, 1997.
- [7] M. Pinedo. *Scheduling. Theory, Algorithms and Systems*. Prentice Hall Inc., Englewood Cliffs, New Jersey, 1 edition, 1995.
- [8] P. Rachakonda and S. Nagane. Simulation study of paint batching problem in automobile industry. <http://sweb.uky.edu/~pkraach/Projects/MFS605Project.pdf>, 2000. consulted 14.07.2004.

Familia de medidas de divergencia entre conjuntos borrosos

Inés Couso¹, Susana Montes¹

¹couso,montes@uniovi.es, Departamento de Estadística, Investigación Operativa y Didáctica de la Matemática, Universidad de Oviedo
E.U.I.T. Industrial, Campus de Viesques, 33203 Gijón

Abstract

Introducimos de forma axiomática, una familia de medidas de divergencia entre conjuntos borrosos. Éstas toman, a su vez, valores borrosos. Enumeramos algunas propiedades adicionales de estas medidas.

Claves: Conjunto borroso, medida de divergencia.

1. Introducción, conceptos previos y notación

El estudio de las diferencias existentes entre dos conjuntos borrosos es útil en algunos problemas de clasificación y, por tanto, en el reconocimiento de imágenes y la teoría de cuestionarios con datos imprecisos. Este hecho ha llevado recientemente a muchos autores a establecer medidas de comparación entre conjuntos borrosos (ver, por ejemplo, [1], [5]). Nosotros hemos tomado como punto de partida dichos trabajos, y hemos introducido ([2, 4]) una definición axiomática del concepto de medida de divergencia. En este trabajo, proponemos una nueva definición axiomática de medida de divergencia borrosa basada en un trabajo inicial [2]. La familia de funciones resultante nos permite, para algunos índices, captar diferencias que las medidas anteriores no captaban.

Dado un universo de discurso Ω , identificaremos cada conjunto borroso, \tilde{A} , con su función de pertenencia $\tilde{A} : \Omega \rightarrow [0, 1]$. Denotamos por \tilde{A}_α a su α -corte: $\tilde{A}_\alpha = \{\omega \in \Omega \mid \tilde{A}(\omega) \geq \alpha\}$, $\forall \alpha \in [0, 1]$. Denotaremos por $\mathcal{F}(\Omega)$ a la clase de sus subconjuntos borrosos. Una medida de divergencia [4] es una función $D : \mathcal{F}(\Omega) \times \mathcal{F}(\Omega) \rightarrow R$ que satisface las siguientes propiedades:

(D1) $D(\tilde{A}, \tilde{B}) = D(\tilde{B}, \tilde{A}) \quad \forall \tilde{A}, \tilde{B} \in \mathcal{F}(\Omega)$.

(D2) $D(\tilde{A}, \tilde{A}) = 0 \quad \forall \tilde{A} \in \mathcal{F}(\Omega)$.

(D3) $\max\{D(\tilde{A} \cup \tilde{C}, \tilde{B} \cup \tilde{C}), D(\tilde{A} \cap \tilde{C}, \tilde{B} \cap \tilde{C})\} \leq D(\tilde{A}, \tilde{B}), \quad \forall \tilde{A}, \tilde{B}, \tilde{C} \in \mathcal{F}(\Omega)$.

En [4] se detallan las razones por las que se exigen estos axiomas, y también se muestran otras propiedades interesantes que se derivan de ellos.

2. Definición axiomática y propiedades adicionales

Las funciones anteriores tienen utilidad en los campos descritos en la introducción, pero no permiten, en general, distinguir si las diferencias entre dos conjuntos están en los grados de pertenencia altos o bajos. Para ello, es necesario extender el rango de valores de la divergencia a la clase de los

conjuntos reales borrosos, como se propone a continuación. Llamamos *medida de divergencia borrosa* a toda función, $\tilde{D} : \mathcal{F}(\Omega) \times \mathcal{F}(\Omega) \rightarrow \mathcal{F}(R)$ que cumple:

(DB1) $\tilde{D}(\tilde{A}, \tilde{B}) : R \rightarrow [0, 1]$ es una función decreciente.

(DB2) Existen una medida de divergencia D y una función creciente $f : R^{[0,1]} \rightarrow R^+$ tales que, para cada $\alpha \in [0, 1]$, $\sup[\tilde{D}(\tilde{A}, \tilde{B})]_\alpha = f(g_{(\tilde{A}, \tilde{B})}^D(\alpha, \cdot))$, donde $g_{(\tilde{A}, \tilde{B})}^D(\alpha, \cdot) : [0, 1] \rightarrow R$ se define de la forma $g_{(\tilde{A}, \tilde{B})}^D(\alpha, \beta) = 0$, si $\beta \leq \alpha$ y $g_{(\tilde{A}, \tilde{B})}^D(\alpha, \beta) = D(\tilde{A}_\beta, \tilde{B}_\beta)$, si $\beta \geq \alpha$. Además $f(0_{R^{[0,1]}}) = 0$.

Dada una divergencia borrosa \tilde{D} y un par de conjuntos borrosos (\tilde{A}, \tilde{B}) la cantidad $\tilde{D}(\tilde{A}, \tilde{B})(d)$ representa la máxima distancia a la que se observa al menos una divergencia d . Así, el axioma 1 indica que, cuanto más nos alejamos, menos diferencias somos capaces de percibir. Por su parte, el valor $\sup[\tilde{D}(\tilde{A}, \tilde{B})]_\alpha$ representa la máxima divergencia que se observa a distancia α . El axioma 2 indica que éste sólo depende de las diferencias que captamos a distancias superiores a α . Se puede comprobar que toda función que satisface los dos axiomas anteriores cumple también los axiomas D1-D3 (utilizando aritmética borrosa). Además, se puede construir una función, F , que, a cada divergencia nítida, D , le asocie una divergencia borrosa, $F(D) = \tilde{D}$. Esto no significa que cada divergencia nítida contenga la misma información acerca de cada par (\tilde{A}, \tilde{B}) que la correspondiente medida borrosa. En otras palabras, no es cierto, en general, que $\tilde{D}(\tilde{A}, \tilde{B})$ se pueda expresar como $F(D(\tilde{A}, \tilde{B}))$. F está definida entre dos clases de funciones y no es una función real de variable real.

3. Agradecimientos

Esta investigación está subvencionada con el Proyecto MTM2004-01269.

4. Bibliografía

- [1] Bouchon-Meunier B., Rifqi M. y Bothorel. S. (1996). Towards general measures of comparison of objets. *Fuzzy Sets and Systems* 84, 143-153.
- [2] Couso I. Janis V. y Montes S. (2004). Cuantificación borrosa de la divergencia entre conjuntos *Actas de ESTYLF*.
- [3] Montes S., Couso I. y Bertoluzza C. (1998). Some classes of fuzziness measures from local divergences. *Belg. J. Oper. Res. Statist. Comput. Sci.* 38, 37-49.
- [4] Montes S., Couso I., Gil P. y Bertoluzza C. (2002). Divergence measure between fuzzy sets. *Int. J. Approximate Reasoning* 30, 91-105.
- [5] Wang X., De Baets B. y Kerre E. (1995). A comparative study of similarity measures. *Fuzzy Sets and Systems* 73, 259-268.

Fiabilidad de un sistema bajo reparaciones imperfectas gobernado por distribuciones tipo-fase

D. Montoro-Cazorla¹, R. Pérez-Ocón², M. Carmen Segovia³

¹dmontoro@ujaen.es, Departamento de Estadística e Investigación Operativa,
Universidad de Jaén

^{2,3}rperez@ugr.es, Departamento de Estadística e Investigación Operativa,
Universidad de Granada

Abstract

Presentamos un sistema sometido a reparaciones imperfectas antes de que ocurra una reparación perfecta. Los tiempos operativos y de reparación siguen distribuciones tipo-fase. Probamos que el modelo que gobierna el sistema es un proceso de Markov, cuya estructura es determinada varias medidas de fiabilidad son calculadas en régimen transitorio y estacionario. Este modelo extiende otros previamente estudiados en la literatura.

Claves: Reliability; Markov process; Phase-type distribution; Imperfect repair.

AMS: 90B25; 62N05

1. Introducción

En el presente trabajo se estudia un sistema sometido a reparaciones imperfectas. Los sucesivos tiempos operativos decrecen y los sucesivos tiempos de reparación crecen. Después de un número de reparaciones imperfectas, el sistema es reparado perfectamente. Todos estos tiempos siguen distribuciones aleatorias tipo-fase. El proceso que gobierna el sistema es un proceso de Markov generalizado, con estados vectoriales. Se estudia el sistema en régimen transitorio. Se obtienen expresiones explícitas para las probabilidades de transición, y a partir de estas se calcula la disponibilidad, la razón de ocurrencia de fallos, y la fiabilidad en un intervalo. Las expresiones obtenidas se presentan en forma bien estructurada y son computacionalmente implementadas. El modelo extiende otros previamente publicados. La siguiente lista resume trabajos previos relacionados con este artículo: Biswas et al. (2000); Pérez-Ocón et al. (2002), Montoro-Cazorla et al. (2006).

2. El modelo

Se resumen las hipótesis y las medidas de fiabilidad calculadas.

#1. Se denota por X_i el tiempo de fallo del sistema después de completar $(i-1)$ reparaciones imperfectas. La variable aleatoria X_i sigue una distribución tipo-fase F_{i+1} con representación $(\alpha(i), T(i))$ de orden $m(i)$, $i = 1, 2, \dots, k + 1$. Se denota por Y_i , $i = 1, 2, \dots, k$ el tiempo de reparación imperfecta del fallo i -ésimo, esta variable aleatoria sigue una distribución tipo-fase G_{i+1} con representación $(\beta(i), S(i))$ de orden $n(i)$.

#2. Suponemos que F_i y G_i son procesos geométricos de razones $a \leq 1$ y $b \geq 1$, respectivamente, $i = 1, \dots, k - 1$.

#3. La reparación perfecta sigue una distribución tipo-fase con representación (γ, L) de orden r .

#4. Todas las distribuciones son independientes.

#5. Cuando el sistema sufre k reparaciones imperfectas es reparado perfectamente al siguiente fallo.

Disponibilidad: $A(t) = \alpha(1) \sum_{i=1}^{k+1} P_i(t)e$,

Razón de ocurrencia de fallos: $\nu(t) = \alpha(1) \sum_{i=1}^{k+1} P_i(t)T^{(0)}(i)$.

Disponibilidad en el intervalo $[t, t+\tau]$: $R(t, \tau) = \sum_{i=1}^{k+1} \alpha(1)P_i(t) \exp\{a^{i-1}T\tau\}e$.

3. Aplicación numérica

Una aplicación numérica ilustra los cálculos efectuados a lo largo del trabajo. Tomamos el caso particular: $k = 3, a = 1, b = 0,75, \alpha(i) = \alpha = (1, 0, 0)$, $T(i) = T, i = 1, \dots, 4$, $\beta(i) = \beta = (1, 0, 0)$, $S(i) = b^{i-1}S, i = 1, \dots, 3$, y las matrices T y S y se obtiene la Tabla 1.

$$T = \begin{pmatrix} -3 & 3 & 0 \\ 0 & -2 & 2 \\ 0 & 0 & -1 \end{pmatrix}, \quad S = \begin{pmatrix} -27 & 27 & 0 \\ 3 & -18 & 15 \\ 0 & 1 & -9 \end{pmatrix}.$$

Tabla 1. Medidas de fiabilidad

k	$A(10)$	$v(10)$	$R(10, 15)$
0	0.8488	0.4630	0.0093
1	0.6651	0.4718	0.4415
2	0.8593	0.4690	0.6015
3	0.8531	0.4648	0.5823

4. Agradecimientos

Este trabajo ha sido parcialmente financiado por el Ministerio de Educación y Ciencia, España, Proyecto MTM2004-03672.

5. Bibliografía

- [1] Biswas, A., Sarkar, J. (2000). Availability of a system maintained thorough several imperfect repairs before a replacement or a perfect repair. *Statist. Probab. Lett.* 50, 105-114.
- [2] Montoro-Cazorla, D., Pérez-Ocón, R. (2006). Maintenance of systems by means of replacements and repairs: The case of phase-type distributions. *Asian-Pacific J. Oper. Res.* En prensa.
- [3] Pérez-Ocón, R., Torres-Castro, I. (2002). A reliability semi-Markov model involving geometric processes. *Appl. Stoch. Models in Bus. and Indust.* 18, 157-170.

Flexible Discrete Location Problems with Capacity Constraints

¹*J. Kalcsics,* ²*S. Nickel,* ³*J. Puerto,* ⁴*A.M. Rodríguez-Chía*

¹j.kalcsics@orl.uni-saarland.de, Universidad de Saarbrucken

²s.nickel@orl.uni-saarland.de, Universidad de Saarbrucken

³puerto@us.es, Universidad de Sevilla,

⁴antonio.rodriguezchia@uca.es, Universidad de Cádiz

Abstract

En este trabajo se introducen modelos flexibles para los problemas de localización discretos con capacidades. Estos modelos proporcionan tres formulaciones alternativas que corresponden a las diferentes puntos de vista desde el que se desea optimizar el problema de localización en el sistema logístico. Además, se presenta un análisis preliminar de los tiempos computacionales necesarios para resolverlos, mostrando la potencialidad y límites de estas formulaciones.

Claves: Localización discreta con capacidades, Problemas ordenados.

AMS: 90B80

1. Introducción

Location Analysis is one of the most active fields in Operations Research. Especially discrete location problems have been widely studied due to their importance in practical applications and also due to their interesting structural properties. Discrete location problems typically involve a finite set of sites at which facilities can be located, and a finite set of clients, which demands for service or goods have to be fulfilled by the facilities. The most simplest and well studied discrete location problems are the discrete p-median problem and the Uncapacitated (and Capacitated) Facility Location Problem. Evidently many extensions of these basic location problems have been developed.

Another important aspect of a location model is the right choice of the objective function and in most classical location models the objective function is the main differentiator. Therefore a great variety of objective functions have been considered. The median objective is to minimize the sum of the costs of fulfilling all demand requests from the clients. The center objective is to minimize the maximum cost of fulfilling the demand of a client, from amongst the sites chosen, over all clients. The centdian objective is a convex combination of the median and center objectives; it aims to keep both the average cost behavior as well as the highest cost in balance. Despite the fact that all three objectives (and some more) are frequently encountered in the literature, not much has been done in the direction of a unified framework for handling all these objectives.

The increasing need for discrete location models in strategic supply chain planning has made it necessary to develop new and flexible location models. To that end, [1] introduced a new type of objective function which generalizes the most popular objective functions mentioned above. This objective function applies a penalty to the cost of fulfilling the demand of a client which is dependent on the position of that cost relative to the costs of supplying other customers. For example, a different penalty might be applied if the cost of supplying the client was the 5th-most expensive cost rather than the 2nd-most expensive. It is even possible to neglect some customers by assigning a zero penalty. This adds a “sorting”-problem to the underlying facility location problem, making formulation and solution much more challenging.

In this paper we want to extend the basic DOMP model to cope with actual requirements from logistics as described above. We will present models taking fix costs as well as capacities into account. Moreover, we will have to consider different models depending on which member of the logistics network is the driving force of the planning process. Typically the client driven model is discussed in location models. The customer is assumed to choose the cheapest service to satisfy his demand. However, it might as well be the case that the customer just needs to know that he receives the service, but the client is chosen by the supplier who operates one or more warehouses. When using a median type objective these two models more or less coincide. Looking at more advanced objective functions we will see that one really get different outcomes. Finally, it might be that a 3rd party logistics provider is involved as well. He has not an accumulated point of view. Instead he wants to look at each transport relation.

2. Bibliografía

- [1] S. Nickel (2001). Discrete Ordered Weber Problems. In: B. Fleischmann and R. Lasch and U. Derigs and W. Domschke and U. Rieder (eds.), *Operations Research Proceedings 2000*, Springer, 71-76.
- [2] S. Nickel and J. Puerto (2005). *Location Theory — A Unified Approach*, Springer.
- [3] A. Rodríguez-Chía and S. Nickel and J. Puerto and F.R. Fernández (2000). A flexible approach to location problems. *Mathematical Methods of Operations Research* 51, 69-89.

Funciones e-convexas: Propiedades

V. Jornet¹, M. Rodríguez²

¹vjornet@ua.es, Departamento de Estadística, Investigación Operativa,
Universidad de Alicante

²marga.rodriguez@ua.es, Departamento de Estadística e Investigación
Operativa, Universidad de Alicante

Abstract

Un subconjunto de R^n se dice e-convexo si es intersección de una familia (posiblemente vacía) de semiespacios abiertos. En este trabajo consideramos las funciones cuyo epígrafo es e-convexo (funciones e-convexas) y estudiamos las propiedades de esta clase de funciones convexas que contiene a la subclase de las funciones convexas cerradas.

Claves: funciones e-convexas, conjuntos e-convexos.

Gestión de tablas de decisiones óptimas

Juan A. Fernández del Pozo, Concha Bielza

{jafernandez, mcbielza}@fi.upm.es, Decision Analysis and Statistics Group,
Departamento de Inteligencia Artificial, Universidad Politécnica de Madrid,
Campus de Montegancedo, Boadilla del Monte, 28660 Madrid, Spain

Abstract

El análisis de las tablas de decisiones óptimas que obtenemos mediante la evaluación de diagramas de influencia facilita una explicación del razonamiento útil para la validación y el análisis de sensibilidad del modelo. Estas tareas presentan dificultades debido al tamaño grande de las tablas y a la estructura compleja de relaciones entre los atributos.

Claves: Representación del conocimiento, Sistemas de ayuda a la decisión, Explicación del razonamiento bajo incertidumbre.

AMS: Clasificación AMS: 68T30, 68T35, 68T37.

1. Introducción

Consideramos los sistemas de ayuda a la decisión (SAD) estructurados mediante un modelo gráfico probabilístico conocido como diagrama de influencia (DI)[1]. El DI es una herramienta útil para la adquisición de conocimiento, representación del problema (modelo) con las relaciones cualitativas (independencias condicionales y dependencias informativas) y cuantitativas (probabilidades y utilidades) y evaluación de las políticas de decisión óptimas.

Esta última tarea permite construir una base de conocimiento sobre el problema que se almacena en las tablas de decisiones óptimas (TDO). El usuario recupera de las TDO las decisiones recomendadas (la alternativa de máxima utilidad esperada) por el SAD en un caso. La combinatoria subyacente a las TDO con un conjunto amplio de variables produce unas tablas son enormes.

2. Gestión de TDO e interfaz de usuarios

Las listas KBM2L [2] resumen las TDO y son introducidas en la tesis doctoral del primer autor [3]. La aplicación que describimos a continuación implementa la lista KBM2L y diversos procedimientos para minimizar su tamaño, visualizar su contenido, extraer explicaciones, realizar consultas y análisis de sensibilidad (AS). La solución software propuesta la denominamos *kbm32* y está realizada en Java 1.5.0_06. La documentación técnica del código y ejemplos se pueden consultar en: <http://www.dia.fi.upm.es/~jafernан/kbm32docs/index.html>. La Figura 1 muestra las ventanas principal, de control y del espectro de una lista KBM2L.

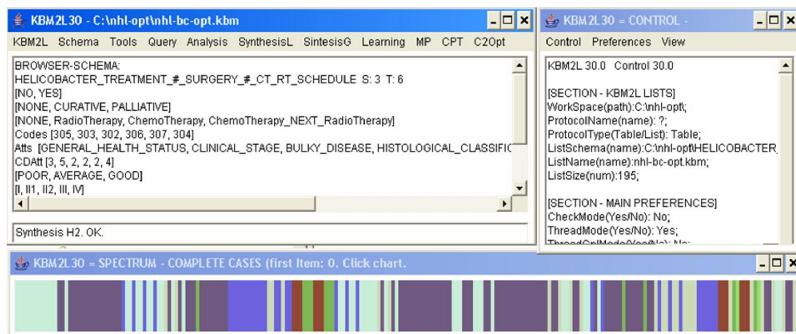


Figura 1: Ventanas de la aplicación *kbm32*.

3. Conclusiones y continuación

kbm32 es una herramienta que complementa el modelo del DI y permite trabajar con TDO muy complejas extrayendo conocimiento. Además de ciertas mejoras vinculadas a la implementación podemos ampliar las funcionalidades relacionadas con la explicación, las consultas y el AS.

4. Agradecimientos

Los autores agradecemos la ayuda de los proyectos de la CICYT (TSI2004-06801-C04-04, TIN2004-21428-E, MTM2004-21099-E), CAM (S-0505/TIC/0230).

5. Bibliografía

- [1] Howard, R.A. and Matheson, J.E. (1984). Influence Diagrams, *The Principles and Applications of Decision Analysis*, Howard, R.A. y Matheson, J.E., Strategic Decisions Group, Menlo Park (CA), 2, 719–762.
- [2] Fernández del Pozo, J.A., Bielza, C., Gómez, M.: A List-Based Compact Representation for Large Decision Tables Management. European Journal of Operational Research 160 (2005) 638-662.
- [3] Fernández del Pozo, J.A.: *Listas KBM2L para la síntesis de conocimiento en sistemas de ayuda a la decisión*, Tesis doctoral. Departamento de Inteligencia Artificial. Facultad de Informática. Universidad Politécnica de Madrid (2006)

GRASP para el Diseño de Territorios de Ventas

Elena Fernández¹, Roger Z. Ríos Mercado²

¹e.fernandez@upc.es, Departamento de Estadística e Investigación Operativa,
Universitat Politècnica de Catalunya

²roger@mail.uanl.mx, Programa de Posgrado en Ingeniería de Sistemas,
Universidad Autónoma de Nuevo León

Abstract

En este trabajo estudiamos un problema de diseño de territorios de ventas que surge en una empresa de productos embotellados. El trabajo incluye una implementación de una heurística constructiva basada en GRASP. Los resultados empíricos muestran una mejora considerable en la calidad de las soluciones con respecto a la práctica común en la empresa.

Claves: Diseño territorial, metaheurísticas, GRASP.

1. Introducción

En industrias de distribución y venta de productos, como las embotelladoras y repartidoras, por ejemplo, se cuenta típicamente con una red de distribución de tal producto, la cual es conformada por miles de puntos de venta. Desde la perspectiva logística existen varios problemas de interés, cada uno de éstos en diferentes etapas del proceso. En particular, el problema estudiado en este trabajo ocurre en una etapa previa al ruteo del producto y se describe como sigue. Con la finalidad de dar un servicio eficiente a un número relativamente grande de clientes, se desea agrupar o dividir a este conjunto de clientes en varios territorios o subconjuntos de tal forma que las tareas dentro de cada territorio compartan ciertos atributos. Entre estos atributos figuran la situación geográfica, por ejemplo. Esto se hace con el fin de delimitar zonas geográficas que puedan ser aprovechadas en una posterior etapa de ruteo del producto. Entre las restricciones propias del sistema figuran los balances de volúmenes de venta, de carga de trabajo y de tiempo de entrega. Además existen también ciertas condiciones de contigüidad que deben ser tomadas en cuenta. Finalmente, otra propiedad importante es la de crear territorios en que los clientes no estén muy separados el uno del otro por cuestiones estratégicas (compacidad).

2. Descripción del Problema

El problema se modela mediante un grafo $G = (V, E)$, donde V representa el conjunto de clientes y el arco $(i, j) \in E$ si los clientes i y j se localizan en manzanas colindantes. Cada cliente $i \in V$ tiene ciertos parámetros asociados como sus coordenadas geográficas $(x(i), y(i))$ y tres actividades medibles. Sea

w_v^a el valor de la actividad a en el nodo v , donde $a = 1$ (demanda del producto), $a = 2$ (carga de trabajo), y $a = 3$ (tiempo de entrega). El número de territorios es un valor fijo conocido con certeza dado por p . El problema consiste en encontrar una partición de V con las siguientes características. Sea B_k el k -ésimo territorio. Se pretende que los territorios estén balanceados con respecto a cada una de las actividades nodales. Esto se logra imponiendo la restricción que de que el tamaño de cada territorio k con respecto a la actividad a (dado por $w^a(B_k) = \sum_{v \in B_k} w_v^a$), se mantenga dentro de una cierta distancia del promedio (dado por $\mu^a = w^a(V)/p$). La condición de contigüidad se logra imponiendo que cada B_k induzca un subgrafo conexo. Finalmente, el criterio de compacidad se logra incorporando como función objetivo a ser minimizada la medida de distancia dada por $D = \sum_{k=1}^p \sum_{j \in B_k} d_{kj}$, donde d_{kj} denota la distancia Euclíadiana del nodo j al centro del territorio j .

3. Método de Solución

La motivación principal en hacer uso de GRASP en este problema en particular es que parece más adecuada para manejar la restricción de conectividad de los territorios que las técnicas actuales de asignación-localización [1], las cuales presentan dificultades al intentar manejar la conectividad. La idea de la fase constructiva de nuestra propuesta es construir los territorios añadiendo un cliente a la vez. En cada iteración, las restricciones de asignación, que aseguran una partición factible, y las de conectividad se satisfacen siempre. La función de mérito, la cual es una combinación ponderada de la función objetivo original (que mide la compacidad de los territorios) y del grado de violación de las restricciones de balanceo, pondera el costo/beneficio de asignar un nodo al territorio activo o “cerrar” dicho territorio y añadirlo a uno nuevo. La fase de postprocesamiento intenta tanto mejorar la calidad de la solución como recuperar la factibilidad de las posibles restricciones de balanceo no satisfechas al final de la fase constructiva. La experiencia empírica en una serie de instancias generadas aleatoriamente a partir de datos proporcionados por la industria muestran la significativa mejora del procedimiento propuesto al compararlo con la práctica actual.

4. Agradecimientos

Este trabajo ha sido financiado por el Ministerio de Educación y Ciencia de España bajo el Programa de Apoyo a Profesores Visitantes (SAB2004-0092) y por la Comisión de Ciencia y Tecnología (TIC2003-05982-C05/4).

5. Bibliografía

- [1] J. Kalcsics, S. Nickel y M. Schröder. (2005). Toward a unified territorial design approach: Applications, algorithms, and GIS integration. *Top*, 13(1):1-74.

Heurística para la resolución de un problema de líneas de montaje

Alberto Gómez, David de la Fuente, José Parreño, Isabel Fernández

{agomez,david,parreno,bel}parreno@epsig.uniovi.es

Departamento de Administración de Empresas y Contabilidad, Universidad de Oviedo

Abstract

En este trabajo se resuelve mediante diferentes heurísticas un problema de líneas de montaje en el sector de la automoción. Este problema fue propuesto por Renault dentro del marco del Challenge ROADEF'2005.

Palabras Clave: heurística, optimización, ensamblaje

1. Introducción

A continuación se describen las características del problema a resolver:

Variables de decisión: Los tipos de vehículos a fabricar en un día así como su número n se saben por adelantado. Las decisiones que se tomarán se relacionan con solamente el orden en el cuál se fabrican los diferentes vehículos.

Matemáticamente, si $v = 1 \dots n$ indica los números de coches y $l = 1 \dots n$ su posición en las cadenas, programar el vehículo v con la posición l se puede representar por el x_{vl} variable = 1 si v están con la posición l y 0 si no.

Limitaciones: Los vehículos son sometidos a una serie de operaciones que condicionan el orden de fabricación.

La pintura: cada vez que un vehículo debe ser pintado con un color diferente del anterior se necesita una purga (lavado de la pistola de pintura). Por otra parte, después de un número máximo de pintura del mismo color, en una serie de vehículos, También se necesaria purgar la pistola de pintura.

Equipamiento: Si un vehículo lleva muchos equipamientos, se le clasifica como “vehículo difícil”, por necesitar gran cantidad de mano de obra y de maquinaria. Además, en determinados equipamientos se necesita realizar ajustes en la cadena de montaje. Es deseable un alisado en la carga, es decir distribuir los esfuerzos autorizados en el tiempo, de forma que después del montaje de un “vehículo difícil” se procesa a montar un “vehículo fácil”. Es decir, se necesita el mayor espacio posible entre los vehículos difíciles.

Las distancias entre vehículos diferente se traducen en un cociente N/P. Es decir, en un tramo de P vehículos en la cadena de vehículos, no más de N están

autorizados para tener dificultades de montaje.

Objetivos: La purgar es un coste económico pues el disolvente necesario para la purgar es costoso. Se pretende reducir al mínimo el número de purgas para así reducir los costes de la fabricación. Luego, uno de los objetivos es reducir al mínimo el número de purgar. Además, debido al hecho de que el mismo vehículo presenta varias características, respetar un mínimo de espacio entre los vehículos difíciles, puede ser incompatible con la minimización del número de purgas. Luego es necesario llegar a un compromiso entre la necesidad de reducir al mínimo el número de purgas y la necesidad de mantener la regla adecuada entre vehículos difíciles. Este compromiso consiste en aceptar, en caso de necesidad, de las violaciones del espaciamiento. El segundo objetivo del problema es reducir al mínimo estas violaciones. El problema es así multiobjetivo con dos objetivos principales:

- minimización del número de purgas
- minimización del número de la violación del cociente N/P.

Para hacer la optimización posible, es necesario establecer el orden de los diferentes objetivos. En principio, se establece que el orden de un primer objetivo no debe ser cambiado por la optimización del segundo.

2. Resolución

En el trabajo se analiza el problema con diferentes tipos de heurísticas de diseño propio y después se mejoran las soluciones de las heurísticas con ayuda de algoritmos genéticos. Se realiza una aproximación a los dos problemas planteados, en primer lugar se resuelve primero el problema de las purgas y después se minimiza las violaciones del coeficiente N/P y, en segundo lugar se resuelve primero las violaciones del coeficiente N/P y después se resuelve el problema de las purgas.

3. Bibliografía

- [1] Descripción del problema a resolver:
<http://www.prism.uvsq.fr/%7Evdc/ROADEF/CHALLENGES/2005/#sujet>.
- [2] Davis, L. (1991). *Handbook of Genetic Algorithms*. Van Nostrand Reinhold.
- [3] Goldberg, D. (1989). *Genetic Algorithms in Search, Optimization & Machine Learning*. Addison-Wesley Co., Inc, Reading, MA.

Heurísticas constructivas y de mejora local para la secuenciación de modelos mixtos de vehículos en líneas de montaje

Joaquín Bautista Valhondo¹, Jaime Cano Belmán²

¹joaquin.bautista@upc.edu, Nissan Chair ETSEIB, Universitat Politècnica de Catalunya

²jaime.cano-belman@upc.edu, Departament d'Estadística e Investigació Operativ, Universitat Politècnica de Catalunya

Abstract

Este trabajo se centra en la secuenciación de productos mixtos en líneas de montaje con el propósito de minimizar la sobrecarga total de trabajo. Se parte de cuatro heurísticas constructivas y se proponen procedimientos de mejora con diferentes tipos de vecindario. En la experiencia computacional se muestra cual de los vecindarios ofrece mejores resultados para el problema.

Palabras Clave: Secuenciación, heurísticas, vecindarios

1. Introducción

Uno de los enfoques del problema de secuenciación de unidades mixtas en líneas de montaje es el propuesto por Yano y Rachamadugu [5]. Los autores usan el criterio de minimización de la sobrecarga de trabajo generada por las unidades que entran en secuencia a la línea. Dicha sobrecarga puede asimilarse al trabajo no realizado o al trabajo extra que deben hacer los denominados operarios utilitarios [3]. En [5], los autores se centran en ejemplares con una sola estación y dos tipos de producto y proponen una heurística de tipo greedy para ejemplares con K estaciones, cuya complejidad computacional es $O(KN)$.

Bolat y Yano [2] extienden el trabajo anterior con la propuesta de tres procedimientos: (1) el primero determina una subsecuencia cíclica (la situación del operario al inicio y final del esquema es la misma) que se replica, para construir la solución; (2) el segundo es un algoritmo greedy que procura, paso a paso, evitar la sobrecarga, combinando unidades; y (3) el tercer algoritmo, también greedy, intenta reducir el tiempo ocioso.

Tsai [4] plantea dos objetivos: minimizar el desplazamiento máximo del trabajador desde el origen de la estación y minimizar el trabajo no realizado. Considera, además, el tiempo que emplea el operario en desplazarse. El procedimiento propuesto tiene una complejidad computacional $O(\log N)$.

Los procedimientos incluidos en [2], [4] y [5] consideran sólo dos tipos de producto (un tipo con un tiempo de proceso mayor que el ciclo y otro con tiempo de proceso menor que el ciclo). Los procedimientos propuestos obtienen soluciones óptimas sólo en determinadas condiciones. En nuestro

trabajo consideramos más de dos productos y extendemos el análisis a múltiples estaciones, realizando una experiencia computacional con instancias tomadas de [3].

2. Procedimientos propuestos

Partiendo del procedimiento greedy propuesto en [1] (denominado *updown*) cuyo propósito es minimizar la sobrecarga, proponemos tres extensiones de éste para el caso de múltiples productos y múltiples estaciones. La propuesta incluye los conceptos de multi-ciclos (con desplazamientos alternativos del operario hacia el inicio y final de la estación) , regeneración (recuperación de la posición inicial del operario en cada ciclo) y regularidad (mantener una carga homogénea a lo largo del tiempo).

3. Procedimientos de mejora

A los resultados obtenidos con los procedimientos constructivos se aplica un procedimiento de mejora local. Se han considerado dos tipos de vecindarios para analizar el efecto en la mejora. Para cada tipo de vecindario se hacen pruebas con diferentes valores para los parámetros.

4. Agradecimientos

Agradecemos a Spanish Operation of Nissan y a la Cátedra Nissan UPC el apoyo dado al presente trabajo. Este trabajo ha sido parcialmente financiado por el proyecto PROTHIUS DPI2004-03475 del Gobierno de España y por la CONACYT del Gobierno de México.

5. Bibliografía

- [1] Bautista, J., Cano J. y Pereira J. (2004). Minimización del trabajo perdido en líneas de montaje con productos mixtos y múltiples estaciones. SEIO 2004, Cádiz.
- [2] Bolat, A. y Yano, C. (1992). Scheduling algorithms to minimize utility work at a single station on paced assembly line. Production Planning and Control 3, 393-405.
- [3] Scholl, A., Klein, R. y Domschke W. (1998). Pattern based vocabulary building for effectively sequencing mixed-model assembly lines. Journal of Heuristics 4, 359-381.
- [4] Tsai, L.H. (1995). Mixed-model sequencing to minimize utility work and the risk of conveyor stoppage. Management Science 41, 485-495.
- [5] Yano, C.A. y R. Rachamadugu (1991). Sequencing to minimize work overload in assembly lines with product options. Management Science 37, 572-586.

Heurísticas para determinar secuencias en una máquina multiproducto sujeta a fallos y con costes cuadráticos

Albert Corominas¹, Rafael Pastor¹, Ana Sánchez¹

¹albert.corominas@upc.edu, rafael.pastor@upc.edu, ana.sanchez@upc.edu
IOC. Universidad Politécnica de Cataluña

Abstract

Se considera una máquina capaz de producir distintas piezas de forma simultánea y se desea minimizar sus costes de almacenamiento y de carencia. En la literatura se plantean métodos de resolución exactos pero los tiempos de cálculo pueden llegar a ser prohibitivos. En este trabajo se proponen diversos procedimientos heurísticos y se comentan los resultados obtenidos.

Palabras Clave: Prioritized Hedging Points, heurística, 1-máquina.

1. Definición del problema

Se considera una máquina capaz de producir, de forma simultánea, n productos o tipos de pieza distintos, con unas tasas de demanda conocidas. La máquina está sujeta a fallos, por lo que puede adoptar dos estados, funcionamiento o paro, cuyos valores siguen distribuciones exponenciales de parámetros conocidos. Se trata de minimizar la esperanza matemática de los costes de almacenamiento y de carencia, por unidad de tiempo, en régimen permanente.

El problema se formaliza en [1], como una extensión de trabajos previos. En dicho artículo, los autores proponen seguir un orden de prioridad previamente establecido e ir asignando la capacidad de la máquina a los diferentes tipos de pieza para que el stock de dichas piezas alcance una determinada cota (que depende del tipo de pieza y de su posición en el orden de prioridad y cuya expresión se proporciona) llamada *hedging point*; una vez alcanzada dicha cota se mantiene un nivel de producción igual a la demanda de dicho tipo de pieza para mantener constante el nivel de stock. En [2] se propone un procedimiento para determinar la secuencia óptima de prioridad de las piezas, se realiza una amplia experiencia computacional y se observa que los tiempos de cálculo crece exponencialmente con el valor n . Dicha conclusión pone de manifiesto la necesidad, para valores altos de n , de procedimientos de resolución heurísticos.

2. Heurísticas

Las heurísticas desarrolladas para determinar el orden de prioridad de las piezas se basan en secuenciar las piezas según los siguientes criterios:

- Ordenar las piezas según el valor de diferentes polinomios, cuyos términos están formados por distintas combinaciones de las demandas de las piezas y los costes unitarios de almacenamiento: heurísticas H1 a H6.

- Calcular, en cada iteración, el coste para todas las parejas de piezas (considerando únicamente aquellas piezas que no han sido ordenadas) y fijar la primera pieza de la pareja de menor coste; se realiza el proceso sucesivamente hasta ordenar todas las piezas: heurística H7.
- Calcular, en una matriz, el coste mínimo de asignar cada pieza a cada posición y aplicar posteriormente un algoritmo de asignación: heurística H8.
- Ordenar las piezas en función del incremento de coste de almacenamiento mínimo que supondría no colocarlas en una determinada posición: heurística H9.
- Asociar a cada pieza al vértice de un grafo y a cada dominancia (definidas en [1]), un arco. Posteriormente formar caminos siguiendo ciertos criterios: heurísticas H10 y H11.

3. Experiencia computacional

Para realizar la experiencia computacional se ha utilizado un ordenador Dual Pentium III a 450 MHz, y, para cada heurística, se han resuelto los mismos 1400 ejemplares (con $n \in [10,23]$). Los tiempos de resolución han sido muy pequeños (≤ 0.02 segundos) y no se observa ninguna tendencia clara en cuanto a su evolución. Para medir la calidad de las heurísticas, se ha comparado la solución obtenida con la solución óptima (disponible al utilizar un método exacto) y se ha podido observar, en promedio, las heurísticas que funcionan mejor (aunque todas ellas, proporcionan cierta discrepancia respecto a las soluciones óptimas). Las discrepancias medias respecto a las soluciones óptimas, en porcentaje, se muestran en el cuadro 1.

	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11
% discrepancia	138,7	18,6	1,8	886,1	342,6	6,6	636,5	21,2	936,1	146,6	269,1

Cuadro 1: Discrepancias respecto a los resultados óptimos

4. Perspectivas

Resolución mediante el uso de cotas en un programa dinámico (de tipo Branch & Bound) y mediante metaheurísticas.

5. Bibliografía

- [1] Shu, C.; Perkins, J. (2001). Optimal PHP production of multiple part-types on a failure-prone machine with quadratic buffer costs. *IEEE Trans. Automat. Contr.*, 46 541-549.
- [2] Corominas, A.; Griñó, R.; Pastor, R. (2004). Determinación de secuencias óptimas en una máquina multiproducto sujeta a fallos y con costes cuadráticos. *XXVIII Congreso Nacional de Estadística e Investigación Operativa*, Cádiz.

Heurísticas para el Problema de Secuenciación de Flujo Regular de Permutación con Máquinas No Disponibles

Paz Pérez González, José Manuel Framiñán Torres

pazperez@esi.us.es, jose@esi.us.es

Grupo de Organización Industrial, Escuela Superior de Ingenieros,
Universidad de Sevilla

Abstract

La indisponibilidad de las máquinas desde el instante inicial de secuenciación es un problema que se plantea en numerosos entornos de fabricación. En este trabajo se estudia en el caso particular del flujo regular de permutación en el que todas las máquinas no están disponibles al inicio del procesamiento de los trabajos.

Palabras Clave: Secuenciación, Flow Shop, disponibilidad

1. Introducción

El problema de secuenciación de trabajos en un flujo regular (flowshop) es uno de los más estudiados en la literatura sobre secuenciación. Sin embargo, todas las referencias asumen que las máquinas están disponibles desde el instante cero. En este trabajo se analiza este problema considerando que cada máquina está disponible a partir de un instante conocido, con el objetivo de minimización del tiempo máximo de terminación, y se analizan y comparan diversas heurísticas adaptadas de casos similares.

2. Adaptación de heurísticas para el problema de indisponibilidades en tiempo cero

Aunque este problema no ha sido abordado, algunas referencias de problemas relacionados son [1-5]. Para el caso de dos máquinas, se demuestra que la regla de Johnson es óptima [1]. En el caso de más de dos máquinas, se asume que los instantes de disponibilidad crecen con el orden de éstas, ya que si alguna máquina verifica lo contrario, esto hace que la siguiente no influya en el problema. Sin pérdida de generalidad, se supone que la primera máquina está disponible en $t=0$.

Para resolver este problema se han adaptado diversas heurísticas de problemas similares, como la de [6]. Los datos del problema se generan a partir de la batería de Taillard [7]. Como no existen baterías de problemas para comparar se han desarrollado distintas formas para generar un orden al cual aplicar las heurísticas adaptadas, basado en el perfil que ofrecen los instantes de disponibilidad de cada máquina. Los resultados muestran que la ordenación de trabajos inicial para la heurística de [6], que consigue los mejores resultados,

se obtiene con un primer trabajo para el que la suma de las diferencias de cuadrados de los tiempos de procesamiento en cada máquina y los intervalos de indisponibilidad de la siguiente máquina es menor, y el resto de los trabajos secuenciados por orden creciente de las sumas de sus tiempos de proceso en todas las máquinas.

3. Contactar

Dirección:

Paz Pérez González y José Manuel Framiñán Torres

Departamento de Organización Industrial y Gestión de Empresas

Escuela Superior de Ingenieros Av. de los Descubrimientos s/n

41092 Sevilla (Spain)

Teléfono: (+34) 954486042 y (+34) 954487214

e-mail: pazperez@esi.us.es, jose@esi.us.es **www:** <http://taylor.us.es>

4. Bibliografía

- [1] Lee, C. Y. (1997). Minimizing the makespan in the two-machine flowshop scheduling problem with an availability constraint. *Operations Research Letters*, 20 (3), 129 - 139.
- [2] Cheng, T. C. E. y Wang, G. (2000). An improved heuristic for the two-machine flowshop scheduling with an availability constraint. *Operations Research Letters*, 22 (5), 223 - 229.
- [3] Breit, J. (2006). A polynomial-time approximation scheme for the two-machine flow shop scheduling problem with an availability constraint. *Computers & Operations Research*, 33 (8), 2143 - 2153.
- [4] Allaoui, H. et al. (2006). Scheduling of a two-machine flowshop with availability constraints on the first machine. *International Journal of Production Economics*, 99 (1-2), 16 - 27.
- [5] Aggoune, R. y Portmann, M. C. (2006). Flow shop scheduling problem with limited machine availability: A heuristic approach. *International Journal of Production Economics*, 99 (1-2), 4 - 15.
- [6] Nawaz, M.; Enscore, J. E. E.; y Ham, I. (1983). A heuristic algorithm for the m-machine, n-job flow-shop sequencing problem. *Omega*, 11 (1), 91 - 95.
- [7] Taillard, E. (1993). Benchmarks for basic scheduling problems. *European Journal of Operational Research*, 64 (2), 278 - 285.

Heurísticas para talleres de flujo sin esperas y tiempos de cambio

Rubén Ruiz¹, Ali Allahverdi²

¹rruiz@eio.upv.es, Departamento de Estadística e Investigación Operativa Aplicadas y Calidad, Universidad Politécnica de Valencia

² allahverdi@kuc01.kuniv.edu.kw, Dept. of Industrial and Management Systems Engineering, Kuwait University, Safat, Kuwait

Abstract

En este resumen se muestran los resultados de métodos heurísticos y de búsqueda local iterativa aplicados al problema de programación de la producción en talleres de flujo sin esperas y con tiempos de cambio separables e independientes de la secuencia. Los métodos propuestos muestran un rendimiento muy superior a las técnicas existentes.

Claves: algoritmos heurísticos, secuenciación, taller de flujo.

AMS: 90B35

1. Introducción

En un taller de flujo se procesan n trabajos en m máquinas donde cada trabajo visita las máquinas en el mismo orden. Este problema combinatorio suscita enorme interés dentro del área de programación de la producción o *scheduling*. En este resumen añadimos dos importantes características al problema. La primera es la restricción de que las tareas de un trabajo no pueden esperar y por tanto, cuando se procesa un trabajo se hace seguido en todas las máquinas (*no-wait*). La segunda es que tenemos tiempos de cambio de partida en las máquinas (*setups*) que con el *no-wait*, se deben realizar siempre antes de que el trabajo llegue a la máquina. Una secuencia de ejemplo, junto con la notación se muestra en la Figura 1. En este resumen mostramos los resultados de heurísticas básicas de intercambio y métodos de búsqueda local iterativa donde el objetivo de optimización es la minimización del tiempo total de proceso o *total completion time*.

2. Heurísticas y Resultados

Las heurísticas se basan en intercambios y en una regla de dominancia que se puede consultar en [3]. Concretamente se proponen las heurísticas H1-H6 que se diferencian en la intensidad de la búsqueda local. Se muestra también una heurística de búsqueda local iterativa o ILS. Los resultados se comparan con el algoritmo TRIPS de [1] y el algoritmo ACO_SLY basado en colonias de hormigas de [2]. Todos los algoritmos se implementan en Delphi 7.0 y se ejecutan en un PC Pentium IV a 3.0 GHz con las mismas condiciones. El

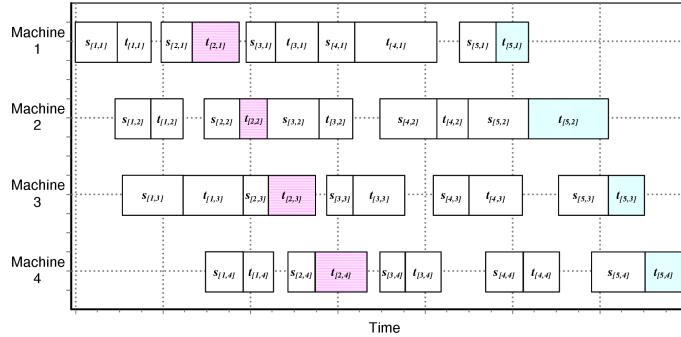


Figura 1: Ejemplo de secuencia sin esperas y con setups.

Cuadro 1 muestra la desviación porcentual sobre la mejor solución conocida de todos los métodos anteriormente comentados sobre dos grupos de 5.400 instancias. Las instancias y las mejores soluciones se pueden descargar de <http://www.upv.es/gio/rruiz>. Como se puede observar, el rendimiento de

Tamaño	H1	H2	H3	H4	H5	H6	ILS	ACO_SLY	TRIPS
Pequeñas	8,25	3,38	2,38	1,77	1,38	0,18	0,11	3,83	2,22
Grandes	35,46	6,90	2,68	2,03	1,12	0,52	0,47	6,81	3,08

Cuadro 1: Incremento porcentual medio sobre la mejor solución para los métodos propuestos y existentes

las heurísticas H1-H6 y del algoritmo ILS es muy bueno, sobre todo este último, que además es muy sencillo de codificar en comparación con TRIPS o ACO_SLY.

3. Bibliografía

- [1] Brown, S. I., McGarvey, R. and Ventura, J. A. (2004). Total flowtime and makespan for a no-wait m-machine flowshop with set-up times separated, *Journal of the Operational Research Society*, 55 (6):614-621.
- [2] Shyu, S. J., Lin, B. M. T. and Yin, P. Y. (2004). Application of ant colony optimization for no-wait flowshop scheduling problem to minimize the total completion time, *Computers & Industrial Engineering*, 47 (2-3):181-193.
- [3] Ruiz, R. and Allahverdi, A. (2006). Some Effective Heuristics for No-Wait Flowshops with Setup Times to Minimize Total Completion Time, *Annals of Operations Research*, Aceptado para su publicación.

Heurísticos para la Secuenciación de Proyectos con Recursos Parcialmente Renovables

R. Álvarez-Valdés¹, E. Crespo², J. M. Tamarit³, M. F. Villa⁴

¹ramon.alvarez@uv.es, ²jose.tamarit@uv.es Departamento de Estadística e Investigación Operativa, Universitat de València

³enric.crespo@uv.es, Departament de Matemàtiques para la Economía y la Empresa, Universitat de València

⁴fvilla@florida-uni.es, Departamento de Contabilidad y Finanzas, Florida Universitaria

Abstract

En este artículo se aborda la problemática de la secuenciación de proyectos con recursos parcialmente renovables (RCPS/II) y se proponen diferentes algoritmos basados en las metodologías GRASP y Scatter Search.

Palabras Clave: Recursos Parcialmente Renovables, GRASP, Scatter Search

1. Recursos Parcialmente Renovables

Nuestro problema consiste en un proyecto formado por un conjunto de actividades que están ligadas por relaciones de precedencia y que durante su ejecución consumen recursos limitados. Nuestro objetivo es secuenciar estas actividades de tal manera que se minimice la duración del proyecto, respetando las relaciones de precedencia y restricciones de recursos.

La particularidad de nuestro problema es que los recursos son “parcialmente renovables”. Este tipo de recursos se caracteriza porque su cantidad disponible está restringida a un conjunto de períodos. Las actividades que se ejecuten dentro de estos períodos consumirán ese recurso, pero si la actividad se realiza fuera no consume recurso. Otra característica de este tipo de recursos es la existencia de un tiempo máximo para llevar a cabo el proyecto. Por esta razón no es trivial la existencia de una solución posible.

Esta clase de recursos tiene aplicabilidad a problemas relacionados con *horarios, turnos...*

2. Resolución del Problema RCPS/II

Antes de intentar resolver el problema mediante cualquiera de los heurísticos diseñados, éste es sometido al “Preproceso”. Este proceso tiene dos objetivos principales: detectar problemas que no tienen una solución posible y, disminuir las dimensiones del problema, para que el heurístico se centre en los elementos del problema que realmente son restrictivos. Todos los estudios realizados concluyen que el preproceso logra los objetivos previstos.

Para resolver el problema hemos utilizado las metodologías del GRASP y del Scatter Search. Para generar soluciones, en ambos métodos, utilizamos un

algoritmo constructivo aleatorizado basado en el esquema en serie que Schirmer (2000) adaptó a este tipo de recursos. Además, nosotros proponemos un proceso “reparador de soluciones”. El objetivo de este proceso consiste en mover las actividades ya secuenciadas a otro tiempo anterior o posterior al asignado inicialmente, compatible con las restricciones de precedencia, con la finalidad de liberar recursos suficientes para que alguna actividad pueda llevarse a cabo. Para mejorar las soluciones proponemos un movimiento *simple* y otro *doble*. El objetivo final de ambos es adelantar las actividades que son las responsables de la duración del proyecto. Además de la versión básica del GRASP hemos desarrollado un algoritmo que combina GRASP y preprocesso de una forma más agresiva (GRASP_Agresivo). A las dos versiones del GRASP les hemos añadido Path Relinking.

En cuanto al algoritmo Scatter Search, para generar la población inicial utilizamos la versión básica del GRASP. Para crear y regenerar el conjunto de referencia utilizamos criterios de calidad y diversidad. La actualización de este conjunto es estática. Además, hemos analizado varias versiones con diferentes opciones para combinar soluciones y con varias estrategias en cuanto a la posibilidad de regenerar o no el conjunto de referencia. Cada vez que generamos una nueva solución mediante una combinación, si no es factible es “reparada” y, posteriormente, sometida a la fase de mejora mencionada.

3. Conclusiones

Todos los algoritmos diseñados han resultado ser muy eficientes a la hora de resolver las diferentes baterías de problemas de las que disponíamos. La versión del GRASP_Agresivo con Path Relinking es imbatible en tiempos de ejecución cortos. Sin embargo, tiene un techo y darle mucho más tiempo no mejora sus resultados. Un procedimiento más complejo, como Scatter Search, permite encontrar soluciones mejores si dispone de suficiente tiempo de computación.

4. Bibliografía

- [1] Festa, P. Resende, M.G.C. (2001): GRASP: An annotated bibliography in: M.G.C. Resende, P. Hansen (Eds.), Essays and Surveys in Metaheuristics, Kluwer Academic Press, Boston, pp. 325-367.
- [2] Martí, R. Laguna, M. (2004): “Scatter Search”. Kluwer Academic Publishers, Boston.
- [3] Schirmer, A. (2000): “Project Scheduling with scarce resources: models, methods and applications”, Verlag Dr. Kovac, Hamburg.

Información de segundo orden en optimización no lineal con cotas simples

Javier M. Moguerza¹, Alberto Olivares¹

¹{javier.moguerza,alberto.olivares}@urjc.es, Departamento de Estadística e Investigación Operativa, Universidad Rey Juan Carlos

Abstract

Se describe la implementación eficiente de un algoritmo de punto interior para la resolución de problemas no convexos con cotas simples utilizando direcciones de curvatura negativa mejorada. Estas direcciones deben garantizar convergencia a puntos que satisfagan las condiciones KKT de segundo orden, al mismo tiempo que mejoran la eficiencia computacional del procedimiento.

Claves: Punto-interior, métodos primal-dual, curvatura negativa

AMS: 49M37, 65K05, 90C30

1. Descripción

El objeto principal de este estudio es el desarrollo de un algoritmo que calcule soluciones locales de un problema de optimización no lineal, y posiblemente no convexo, de la forma

$$\begin{aligned} \min_x & f(x) \\ s.a & x \geq 0, \end{aligned} \tag{4.68}$$

donde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ es una función continua y, al menos, dos veces diferenciable. En concreto, se desea calcular puntos que satisfagan las condiciones KKT necesarias de primer y segundo orden para el problema (4.68), así como comprobar el papel fundamental que juega el uso de direcciones de curvatura negativa mejorada en dicho cálculo.

El algoritmo sigue una estrategia de búsqueda lineal que combina una dirección de Newton modificada y una dirección de curvatura negativa, siguiendo una línea semejante a la propuesta en Moguerza y Prieto [2]. Las direcciones requeridas para actualizar los iterandos se construyen utilizando una metodología de punto interior. En este sentido, se transforma el problema (4.68) en una sucesión de problemas sin restricciones del tipo

$$\min_x L_B(x; \mu) = f(x) - \sum_{i=1}^n \mu_i \log x_i.$$

De esta manera se incorporan a la función objetivo las restricciones de no negatividad y se genera una sucesión de problemas barrera sin restricciones de modo que $\mu \rightarrow 0$. Las direcciones de búsqueda se calculan para aproximar las soluciones de cada uno de estos subproblemas. Durante todo el proceso se usa un vector de parámetros barrera $\mu \in \mathbb{R}^n$, es decir, un parámetro distinto para cada cota simple $x \geq 0$. En cada iteración será necesario actualizar el vector de parámetros μ .

Uno de los puntos cruciales es el planteamiento de una metodología de búsqueda lineal sobre la parábola

$$x_{k+1} = x_k + \alpha_k^2 s_k + \alpha_k d_k,$$

donde α_k debe garantizar que $f(x_{k+1}) < f(x_k)$, s_k es una dirección de Newton modificada y d_k una dirección de curvatura negativa. Otro punto clave se encuentra en el desarrollo de algoritmos eficientes que mejoren la información de segundo orden siguiendo la línea propuesta en Olivares [3].

Finalmente, con el fin de comprobar la eficiencia práctica del algoritmo propuesto, se han llevado a cabo una serie de experimentos numéricos sobre dos conjuntos diferentes de problemas. Por una parte, se ha utilizado una colección de problemas cuadráticos generados de forma automática y controlada para garantizar la existencia de curvatura negativa significativa. Por otro lado, se ha considerado la colección de problemas CUTER descrita en [1].

2. Agradecimientos

Trabajo parcialmente financiado por el proyecto TIC2003-05892-C05-05 del MCYT.

3. Bibliografía

- [1] Gould, N. I. M., Orban, D. y Toint, Ph. L., *CUTER (and SifDec), A Constrained and Unconstrained Testing Environment, revisited*, Technical Report TR/PA, 2004
- [2] Moguerza, J.M. y Prieto, F. J., *An Augmented-Lagrangian interior-point method using directions of negative curvature*, Mathematical Programming, 95, pp. 573-616, 2003
- [3] Olivares, A. *Nonconvex Optimization and Directions of Negative Curvature: an Efficient Approach*, Ph. D. Thesis, Rey Juan Carlos University, 2005

Interval values for strategic games in which players cooperate

Luisa Carpente¹, Balbina Casas-Méndez², Ignacio García-Jurado³, Anne van den Nouweland⁴

¹luisacar@udc.es, Departamento de Matemáticas, Universidade da Coruña

²eigradu3@usc.es, Departamento de Estatística e IO, Universidade de Santiago de Compostela

³igjurado@usc.es, Departamento de Estatística e IO, Universidade de Santiago de Compostela

⁴annev@uoregon.edu, Department of Economics, University of Oregon

Abstract

En este trabajo se propone un nuevo método que asocia un juego coalicional de intervalo a cada juego en forma estratégica. Este método se basa en los valores inferiores y superiores de juegos bipersonales finitos de suma nula. El principal resultado que se presenta es la caracterización axiomática de este método.

Keywords: Strategic Games, Coalitional Interval Games, Axiomatic Characterization

AMS: 91A12

1. Introduction

The problem of associating a coalitional game with every strategic game was already addressed by von Neumann and Morgenstern in their pioneer 1944 book. As they write in section 25.2.1, their purpose is “to determine everything that can be said about coalitions between players, compensations between partners in every coalition, mergers or fights between coalitions, etc., in terms of the characteristic function $v(S)$ alone”. They define $v(S)$ using a zero-sum game between coalition S on one hand, and coalition $N \setminus S$ consisting of the other players on the other hand, and taking into account that the mixed extension of this game “has a well defined value” (see section 25.1.2).

The literature drifted away from the zero-sum game approach to this problem and toward a Nash-equilibrium approach. See, for example, Bergantiños and García-Jurado (1995). Recently, the zero-sum game approach was revived in Carpente et al. (2005). In that paper, the same zero-sum games as in von Neumann and Morgenstern (1944) are used, but attention is limited to coordinated actions by coalitions of players rather than mixed strategies. They take a conservative or pessimistic point of view and define $v(S)$ to be the worth that the players in S can guarantee for themselves no matter what actions the

players in the complementary coalition play. Hence, they use the lower value of the zero-sum game between coalition S and coalition $N \setminus S$. A more optimistic point of view would be to consider the worth of a coalition to be the amount that it can obtain for its members by correctly anticipating the actions of the complementary coalition and to use the upper value of the zero-sum game between coalition S and coalition $N \setminus S$.

In general, the pessimistic and optimistic methods lead to different predictions. Rather than choosing between them, in the current paper we simply acknowledge the fact that for every coalition S in a strategic game there is a reasonable *range* of worths that it can expect to obtain. This range is bounded from below by the pessimistic prediction obtained using the lower value of the associated zero-sum game and it is bounded from above by the optimistic prediction obtained using the upper value of that game. In doing so, we associate with each strategic game a *coalitional interval game* (see Branzei et al. (2003)).

In this paper we provide axiomatic characterizations of this new method which associates a coalitional interval game with every strategic game, and discuss its interest and its main properties.

2. Acknowledgement

The authors acknowledge the financial support of *Ministerio de Educación y Ciencia, FEDER*, and *Xunta de Galicia* through projects BEC2002-04102-C02-02, SEJ2005-07637-C02-02, and PGIDIT03PXIC20701PN.

3. Bibliography

- [1] Bergantiños G, García-Jurado I (1995) Estudio comparativo de diversas funciones características asociadas a un juego en forma normal. *Investigaciones Económicas* 19: 127-138.
- [2] Branzei R, Dimitrov D, Tijs S (2003) Shapley-like values for interval bankruptcy games. *Economics Bulletin* 3: 1-8.
- [3] Carpente L, Casas-Méndez B, García-Jurado I, van den Nouweland A (2005) Values for strategic games in which players cooperate. *International Journal of Game Theory* 33: 397-419.
- [4] Von Neumann J, Morgenstern O (1944) *Theory of games and economic behavior*. Princeton University Press.

KT-invexidad en optimización escalar no regular

Beatriz Hernández Jiménez¹, Rafaela Osuna Gómez², Marko A. Rojas-Medrano³, Antonio Beato Moreno⁴

¹beatriz.hernandez@dmat.uhu.es, Departamento de Matemáticas, Universidad de Huelva

²rafaela@us.es, Departamento de Estadística e Investigación Operativa, Universidad de Sevilla

³marko@ime.unicamp.br, IMECC, Universidad de Campinas (Brasil)

⁴beato@us.es, Departamento de Estadística e Investigación Operativa, Universidad de Sevilla

Abstract

La convexidad juega un papel decisivo en las condiciones de optimalidad, y puede sustituirse por la KT-invexidad en optimización regular escalar para establecer condiciones de optimalidad de tipo Kuhn-Tucker.

Extenderemos este resultado de Martin a la programación escalar no regular, definiendo un concepto análogo, denominado 2-KT-invexidad.

Claves: optimización no regular, KT-invexidad, condiciones de optimalidad.

1. Introducción

En la teoría de la optimización clásica escalar, donde tenemos funciones diferenciables y los problemas son regulares, las condiciones necesarias de optimalidad de tipo Kuhn-Tucker (KT), son suficientes también si las funciones que intervienen en el problema son convexas. En los últimos años se han concentrado los esfuerzos en debilitar esa hipótesis de convexidad.

Hanson [3] definió las funciones invex, bautizadas así por Craven [2], que permitieron el uso de las condiciones necesarias de optimalidad de tipo KT como suficientes en problemas escalares regulares con un número finito de restricciones. Más tarde, Martin [6] demostró que las hipótesis de invexidad no sólo eran suficientes, sino también necesarias en problemas escalares sin restricciones.

Para problemas regulares con restricciones, la invexidad definida por Hanson era una condición suficiente, pero no necesaria para que un punto estacionario de Kuhn-Tucker fuese un mínimo global. Martin definió una noción de invexidad más débil, denominada KT-invexidad, que era necesaria y suficiente para establecer condiciones de optimalidad de tipo KT.

En un problema de optimización no regular, no son aplicables las condiciones clásicas de optimalidad de primer orden de Karush-Kuhn-Tucker, pero sí las dadas por Izmailov [4], basadas en la caracterización de 2-regularidad del cono tangente en un punto del conjunto factible.

En este trabajo, extendemos el resultado de Martin de caracterización de la KT-invexidad a los problemas no regulares, definiendo para ello un concepto análogo, la 2-KT-invexidad, basándonos en las condiciones de optimalidad dadas para problemas escalares no regulares por Izmailov.

2. Bibliografía

- [1] Ben-Israel, B. and Mond, B., *What is invexity?*. Journal of the Australian Mathematical Society, **28 B** (1986), 1-9.
- [2] Craven, B. D., *Invex Functions and Constrained Local Minima*. Bulletin of Australian Mathematical Society, **24** (1981), 357-366.
- [3] Hanson, M.A. *On sufficiency on Kuhn-Tucker Conditions* . J. Math. Anal. Appl. **80**, (1981), 545-550.
- [4] Izmailov, A.F.. *Optimality conditions for degenerate extremun problems with inequality-type constraints*. Comp. Maths Math. Phys., **34**, Nº 6, (1994), 723-736 .
- [5] Mangasarian, O.L.. *Nonlinear Programming*. Ed. MacGraw-Hill (1969).
- [6] Martin, D.M.. *The essence of invexity*. Journal of Optimality Theory and Applications, **17**, Nº 1, (1985), 65-76.
- [7] Minoux, M.. *Mathematical Programming. Theory and Algorithms*. John Wiley and Sons, (1986).
- [8] Osuna-Gómez, R., Rufián-Lizana, A. and Ruiz-Canales, P.. *Invex Functions and Generalized Convexity in Multiobjective Programming*. Journal of Optimality Theory and Applications, **98**, Nº 3, (1998), 651-661.

La evolución histórica de la “robustez”: entre la Decisión Multicriterio y los Métodos Bayesianos

Maria del Carmen Escrivano Ródenas¹, Gabriela Mónica Fernández Barberis²

Departamento de Departamento de Métodos Cuantitativos, Universidad San Pablo-CEU

¹escrod@ceu.es, ²ferbar@ceu.es

Abstract

La robustez se ha venido definiendo y desarrollando de manera muy diferente en las diferentes disciplinas y en particular, en la T^a de la Decisión, y más concretamente en sus dos grandes pilares, los Métodos Bayesianos y la Decisión Multicriterio. Debido a la gran importancia de este tipo de análisis y a las confusiones, es necesario recurrir a la historia del concepto y dejar bien delimitados la interpretación que cada rama científica utiliza.

Palabras Clave: Robustez, Métodos Bayesianos, Decisión Multicriterio.

AMS: 91B06

1. Introducción

El término “robustez” se ha venido definiendo y desarrollando de manera muy diferente en las diferentes disciplinas científicas y en particular, en la Teoría de la Decisión, y más concretamente en sus dos grandes pilares, los Métodos Bayesianos y la Decisión Multicriterio. Sin embargo el análisis de la robustez ha adquirido una gran importancia en todos los ámbitos, pues la incertidumbre afecta e inunda todo el ambiente de decisión. No se puede obviar, ni omitir, ni relegar, es imprescindible su tratamiento desde el principio, tanto desde el punto de vista del analista, como del decidor. Al mismo tiempo, este análisis de la robustez ha dado lugar a confusiones y errores en su lectura y comprensión, ya que parte de supuestos conceptuales muy distintos.

Por lo tanto es necesario y aconsejable recurrir a la historia del concepto y dejar bien delimitados la interpretación que cada rama científica utiliza del mismo término robustez.

2. Recorrido histórico de la robustez dentro de la Decisión Multicriterio

La robustez se puede utilizar en relación con las soluciones, con los métodos, con los procesos y/o con las conclusiones. El recorrido histórico por la evolución del término robustez a lo largo del tiempo es un proceso necesario para comprender el concepto.

3. Recorrido histórico de la robustez en los Métodos Bayesianos

El análisis de sensibilidad está reconocido como imprescindible y reforzado por el hecho de la fuerte dosis de subjetividad y complejidad en la evaluación de los juicios, opiniones y preferencias del decidor.

4. Conclusiones

Al utilizar el término robustez es imprescindible sumergirse en el ámbito particular de cada caso, y comenzar con los puntos de partida propios de cada campo científico específico, ya que ni siquiera dentro del ámbito de la Teoría de la Decisión, en sentido amplio, tiene la misma definición.

5. Bibliografía

- [1] Berger, J.O.; Ríos Insúa, D.; Ruggeri, F. (2000). Bayesian Robustness. In Ríos Insúa, D.; Ruggeri, F., eds.: *Robust Bayesian Analysis*. Springer-Verlag, New York. USA.
- [2] Escribano Ródenas, M.C.; Fernández Barberis, G. (2003). La robustez y su aplicación en el ámbito de la Decisión Multicriterio. *Actas XXVII Congreso Nacional de Estadística e Investigación Operativa*, Lérida.
- [3] Fernández Barberis, G.; Escribano Ródenas, M.C. (2004). La problemática de la robustez: estado de la cuestión en el ámbito de la Decisión Multicriterio. *Actas del XXVIII Congreso Nacional de Estadística e Investigación Operativa*, Cádiz.
- [4] Hites, R.; De Smet, Y.; Rissee, N.; Salazar-Neumann, M.; Vincke, Ph. (2000). A Comparison between Multicriteria and Robustness Frameworks. *Service de Mathématiques de la Gestion*. Université Libre de Bruxelles. Belgique.
- [5] Ríos Insúa, D. (2004). Robustness Analysis: A Bayesian point of view. *EWG-MCDA*, Serie 3, Nº 9, pp. 6-8.
- [6] Rosenhead, J. (2002). Robustness Analysis. London. England. *EWG-MCDA Newsletter*.
- [7] Roy, B. (2002). Robustesse de quoi et vis-à-vis de quoi mais aussi robustesse pourquoi en Aide à la Décision?. *EGW-MCDA*. Université Paris-Dauphine.
- [8] Vincke, Ph. (2003). About Robustness Analysis. *EWG-MCDA*, Serie 3, Nº8, pp. 7-9.

Location and size selection for new competing centres on a network

R. Suárez Vega, D.R. Santos Peñate, P. Dorta González

{rsuarez, drsantos, pdorta}@dmc.ulpgc.es

Dept. de Métodos Cuantitativos en Ecomía y Gestión
Universidad de Las Palmas de Gran Canaria

Abstract

En este trabajo se emplea un algoritmo de optimización global basado en ramificación y acotación para resolver el problema de determinar tanto las localizaciones como las características (tamaño) de unos nuevos centros en un mercado. Los consumidores eligen el centro más atractivo, considerando para ello sus características y los desplazamientos.

Keywords: Competitive location, binary preferences, attraction functions

AMS: 90B85

1. Introduction

The follower location problem on networks was formalized by Hakimi (1983) who introduced the term $(r|X_p)$ -medianoid. This problem consists of determining the optimal locations of r facilities belonging to a firm (the follower) which enters a market where a competitor (the leader) has p facilities. The demand is concentrated at the nodes and the facilities may be located at any point on the network.

In this paper, we consider the $(r|X_p)$ -medianoid problem assuming inelastic demand and binary preferences, that is consumers patronize the facility which exerts the biggest attraction to them. We formulate the attraction as a function directly proportional to the facility quality and inversely proportional to an increasing function of the distance between customers and facilities.

In our model, the entry firm wants to determine both the locations and the quality levels of the new facilities, in order to maximize profits (income minus costs). This joint problem was solved by Suárez-Vega, Santos-Peña and Dorta-González (2004) in the discrete space but the network problem, to the authors knowledge, has not been solved yet.

After presenting the model and some discretizations results we formulate a global search algorithm based on a branch and bound procedure to solve the facility location-quality problem. We apply this procedure to locate new hypermarkets in Grand Canary Island.

2. The branch and bound algorithm

As Suárez-Vega et al (2004) proved, given the quality levels for the new facilities (A_r), their optimal locations, $Y_r^*(A_r)$, can be found evaluating a finite set of points. Taking into account this discretization result, the resolution of the location-size problem is equivalent to find the r -vector of quality levels $A_{Y_r}^*$ that verifies

$$\Pi(A_{Y_r}^*) = \max_{A_{Y_r} \in [L, U]^r} \Pi(A_{Y_r})$$

where $\Pi(A_{Y_r})$ is the follower's maximum profit when the quality levels for the new facilities are A_{Y_r} . Therefore, the solution is the pair $(Y_r^*(A_{Y_r}^*), A_{Y_r}^*)$.

The branch and bound method consists of generating a partition of the feasible set for the quality levels, $[L, U]^r$, to determine the subset where the optimal solution is. For computational efficiency, the branch and bound algorithm provides an ε -optimal solution, with prescribed $\varepsilon > 0$. This is not trouble because the solution to the quality problem must also be understood in terms of ε -optimal quality levels (See Suárez-Vega, et al (2004)).

3. Acknowledgement

This paper is partially financed by Ministerio de Ciencia y Tecnología (Spain) and FEDER, grant MTM2005-09362-C03-03

4. Bibliography

- [1] Hakimi, S.L. (1983), On locating new facilities in a competitive environment. *European Journal of Operational Research* 12, 29-35.
- [2] Suárez-Vega, R., Santos-Peña, D.R. and Dorta-González, P. (2004). Discretization and resolution of the $(r|X_p)$ -medianoid problem involving quality criteria. *TOP* 12(1), 111-133.

Mal planteamiento primal-dual en optimización lineal semi-infinita: Una aplicación a los problemas cónicos

M.J. Cánovas¹, M.A. López², J. Parra¹, F.J. Toledo¹

¹cánovas@umh.es, parra@umh.es, javier.toledo@umh.es

Centro de Investigación Operativa, Universidad Miguel Hernández

²marco.antonio@ua.es

Dpto. de Estadística e Investigación Operativa, Universidad de Alicante

Abstract

Consideramos un problema de optimización lineal con una cantidad arbitraria de restricciones, así como su problema dual. Caracterizamos aquellos problemas que están mal planteados con respecto a cierta propiedad, a saber, consistencia, resolubilidad, dual-consistencia, . . . , y proporcionamos las correspondientes distancias al mal planteamiento. Finalmente mostramos un nuevo enfoque para el estudio del mal planteamiento de problemas cónicos.

Palabras Clave: Distancia al mal planteamiento, programación semi-infinita

AMS: 65F22, 90C34, 90C05, 15A39, 49J53

1. Introducción

Consideramos el par de problemas duales de optimización lineal, en \mathbf{R}^n ,

$$\begin{aligned}\pi : \inf\{c'x : a_t'x \geq b_t, t \in T\} & \quad (\text{P}) \\ \pi^d : \sup\{\sum_{t \in T} \lambda_t b_t : \sum_{t \in T} \lambda_t a_t = c, \lambda \in R_+^{(T)}\} & \quad (\text{D})\end{aligned}$$

donde $c, x, a_t \in \mathbf{R}^n$, $b_t \in \mathbf{R}$, $t \in T$, y T es un conjunto arbitrario de índices, posiblemente infinito. $R_+^{(T)}$ es el cono de todas las funciones $\lambda : T \rightarrow \mathbf{R}_+$ que toman valores positivos sólo en una cantidad finita de puntos de T . El espacio paramétrico de todos los problemas π en la forma (P) se denota por Π .

Π_c , Π_i , Π_b , Π_s , y Π_c^d , denotan los subconjuntos de Π formados por los problemas consistentes (con soluciones factibles), inconsistentes, acotados (con valor óptimo finito), resolubles (con soluciones óptimas), y dual-consistentes (problema dual (D) con soluciones factibles), respectivamente.

Dadas dos normas arbitrarias en \mathbf{R}^n y en \mathbf{R}^{n+1} , denotadas ambas por $\|\cdot\|$, consideramos la distancia extendida $\delta : \Pi \times \Pi \rightarrow [0, +\infty]$, dada por

$$\delta(\pi, \pi_1) := \max\{\|c - c^1\|, \sup_{t \in T} \left\| \begin{pmatrix} a_t^1 \\ b_t^1 \end{pmatrix} - \begin{pmatrix} a_t \\ b_t \end{pmatrix} \right\| \}$$

Un programa lineal cónico en forma dual se puede reformular como un problema en la forma (P). En [3, Sección 4] mostramos un nuevo enfoque de la estabilidad de los programas cónicos a través de nuestros resultados.

2. Tipos de mal planteamiento: caracterizaciones y distancias

Entendemos que un problema está mal planteado con respecto a cierta propiedad si perturbaciones arbitrariamente pequeñas de sus vectores de coeficientes pueden dar lugar a problemas con y sin dicha propiedad. Así pues, $bd(\Pi_c)$, $bd(\Pi_s)$, $bd(\Pi_c^d)$, ..., constituyen diferentes conjuntos de problemas mal planteados. La distancia a cada uno de estos conjuntos se llama “distancia al mal planteamiento con respecto a la correspondiente propiedad”.

En lo que sigue $\Pi_\infty := \pi \in \Pi : \delta(\pi, bd(\Pi_c)) = +\infty$.

Teorema: Si $\pi \in \Pi \setminus \Pi_\infty$. Se tienen:

- (i) $\pi \in int(\Pi_c) \Leftrightarrow 0_{n+1} \in ext(H)$;
- (ii) $\pi \in ext(\Pi_c) \Leftrightarrow 0_{n+1} \in int(H)$;
- (iii) $\pi \in bd(\Pi_c) \Leftrightarrow 0_{n+1} \in bd(H)$;
- (iv) $\delta(\pi, bd(\Pi_c)) = d(0_{n+1}, bd(H))$,

donde $H :=$

$$conv \left(\left\{ \begin{pmatrix} a_t \\ b_t \end{pmatrix}, t \in T \right\} \right) + R_+ \left\{ \begin{pmatrix} O_n \\ -1 \end{pmatrix} \right\}$$

Teorema: Sea $\pi \in \Pi$. Se tienen:

- (i) $\pi \in int(\Pi_c^d) \Leftrightarrow 0_n \in int(Z^-)$;
- (ii) $\pi \in ext(\Pi_c^d) \Leftrightarrow 0_n \in ext(Z^-)$;
- (iii) $\pi \in bd(\Pi_c^d) \Leftrightarrow 0_n \in bd(Z^-)$;
- (iv) $\delta(\pi, bd(\Pi_c^d)) = d(0_n, bd(Z^-))$,

donde $Z^- := conv(\{a_t, t \in T; -c\})$.

Teorema: Si $\pi \in int(\Pi_c)$, entonces:

- (i) $\pi \in int(\Pi_s) \Leftrightarrow 0_n \in int(Z^-)$;
- (ii) $\pi \in ext(\Pi_s) \Leftrightarrow 0_n \in ext(Z^-)$;
- (iii) $\pi \in bd(\Pi_s) \Leftrightarrow 0_n \in bd(Z^-)$.

Teorema: Sea $\pi \in bd(\Pi_c)$. Entonces $\pi \in bd(\Pi_s)$ si y solo si:

- (i) $\pi \in cl(bd(\Pi_c) \cap \Pi_c)$, o bien
- (ii) $0_n \in bd(Z^+)$,

donde $Z^+ := conv(\{a_t, t \in T; +c\})$.

Teorema: Sea $\pi \in \Pi \setminus \Pi_\infty$. Si se tiene alguna de las siguientes afirmaciones:

- (i) $\pi \in cl(\Pi_s)$;
- (ii) $d(0_{n+1}, bd(H)) \neq d(0_n, bd(Z^-))$;
- (iii) $d(0_{n+1}, bd(H)) = d(0_n, bd(Z^-)) \geq \|c\|$.

Entonces $\delta(\pi, bd(\Pi_s)) = \min d(0_{n+1}, bd(H)), d(0_n, bd(Z^-))$

3. Bibliografía

- [1] Cánovas, M.J., López, M.A., Parra, J., Toledo, F.J. (2005). Distance to ill-posedness and the consistency value of linear semi-infinite inequality systems. *Math. Program.*, 103A, 95-126.
- [2] Cánovas, M.J., López, M.A., Parra, J., Toledo, F.J. (2006). Distance to solvability/unsolvability in linear optimization, *SIOPT*, 16(3), 629-649.
- [3] Cánovas, M.J., López, M.A., Parra, J., Toledo, F.J. (2005). Ill-posedness with respect to the solvability in linear optimization, *Linear Algebra and Its Appl.*, to appear.
- [4] Cánovas, M.J., López, M.A., Parra, J., Toledo, F.J. (2005). Sufficient conditions for total ill-posedness in linear optimization, *EJOR*, to appear.
- [5] Núñez, M.A. (2002). A characterization of ill-posed data instantaneos for convex programming, *Math. Program.*, 91, 375-390.

Métodos de Optimización Indirecta en Aplicaciones Biotecnológicas

A. Marín-Sanguino¹, C. González-Alcón², N.V. Torres Darias¹

¹Dpto. de Bioquímica y Biología Molecular

²Dpto. de Estadística, Investigación Operativa y Computación

Universidad de La Laguna

amarin@ull.es, cgalcon@ull.es, ntorres@ull.es

Claves: Linear Programming, Geometric Programming, Biochemistry.

AMS: Clasificación AMS:

90C90, 92C40, 92C35.

El desarrollo de potentes tecnologías de manipulación genética ha abierto un nuevo campo para la biotecnología y la biomedicina. Pero junto con la posibilidad de alterar el funcionamiento de las células a nivel molecular han aparecido una amplia gama de interrogantes sobre cómo y cuándo hay que manipular dichos sistemas. La complejidad del metabolismo y el alto grado de interdependencia de los constituyentes celulares hacen muy difícil establecer pautas para modificar la función celular con fines prácticos. Entre las aproximaciones más prometedoras a este problema está el desarrollo de modelos matemáticos basados en ecuaciones diferenciales y su posterior optimización.

El uso de formalismos matemáticos facilita considerablemente la matematización del problema. Uno de los más populares es el llamado formalismo en ley de potencias (power law) que combina la flexibilidad de los sistemas no lineales con una regularidad estructural que simplifica su análisis.

En este trabajo se presentará la familia de Métodos de Optimización Indirecta (IOM). Desarrollados para la optimización de sistemas biotecnológicos en representación power law, han sido utilizados con éxito en una gran variedad de contextos diferentes, incluyendo biorreactores, sistemas metabólicos, sistemas de expresión genética y enzimas aisladas.

Podríamos plantear una clase importante de problemas biotecnológicos como:

$$\begin{array}{ll} \text{max o min} & f(X) \\ \text{sujeto a:} & \text{Estado Estacionario} \\ & \text{Eficiencia energética} \\ & \text{Viabilidad celular} \end{array}$$

$f(X)$ suele ser un flujo de producción pero también puede ser un índice de productividad, un término económico, una propiedad dinámica, etc.

Las restricciones del problema son características comunes en bioquímica o microbiología y pueden representarse mediante ecuaciones en ley de potencias:

- La operación sostenible del sistema suele garantizarse haciéndolo funcionar en condiciones de estado estacionario. Modelizando la red de reacciones químicas e interacciones moleculares mediante un modelo matemático con la siguiente estructura:

$$\frac{dX_i}{dt} = \alpha_i \prod_j X_j^{g_{i,j}} - \beta_i \prod_j X_j^{h_{i,j}} \quad i = 1, \dots, n,$$

la condición de estado estacionario se obtiene igualando las derivadas a cero.

- La eficiencia energética suele estar condicionada por la síntesis de proteínas y la carga que esta supone para el metabolismo celular. Limitando la concentración total (o un conjunto) a una cota determinada, podemos controlar el coste energético de las soluciones.

$$\sum_{i \in I} X_i \leq T_I$$

- Para evitar la acumulación de sustancias tóxicas o la falta de precursores esenciales, se limita el rango de las variables:

$$X_i^L \leq X_i \leq X_i^U \quad i = 1, \dots, n.$$

La regularidad estructural de los formalismos empleados permiten reducir las ecuaciones presentadas o variantes de las mismas a problemas solubles mediante métodos eficientes como programación lineal o geométrica.

1. Bibliografía

- [1] Alvarez-Vasquez, F., González-Alcón, C.M. and Torres, N.V. (2000). Metabolism of citric acid production by *Aspergillus niger*. Model definition, steady state analysis, dynamic behavior and constrained optimization of citric acid production rate. *Biotechnology and Bioengineering* 70(1), 82-108.
- [2] Marín-Sanguino A. and Torres, N.V. (2002). Modelling, steady state analysis and optimization of the catalytic efficiency of the triosephosphate. *Bulletin of Mathematical Biology* 64 (2), 301-326.
- [3] Torres, N.V. and Voit, E.O. (2002). *Pathway Analysis and Optimization in Metabolic Engineering*. Cambridge University Press.

Minimización del tiempo total de retraso en el taller de flujo de permutación

Eva Vallada, Rubén Ruiz, Gerardo Minella

{evallada,rruiz,mgerar}@eio.upv.es, Departamento de Estadística e Investigación Operativa Aplicadas y Calidad, Universidad Politécnica de Valencia

Abstract

En este trabajo se proponen dos algoritmos genéticos y un algoritmo iterativo voraz para el problema del taller de flujo de permutación con el objetivo de minimizar el retraso total de los trabajos. Los algoritmos se han evaluado y comparado entre ellos y con otras técnicas propuestas en la literatura proporcionando los primeros mejores resultados.

Claves: algoritmos genéticos, secuenciación, taller de flujo.

AMS: 90B35

1. Introducción

El problema del taller de flujo de permutación es un problema de optimización combinatoria donde tenemos un conjunto de N trabajos ($N = 1, \dots, n$) que hay que procesar en un conjunto de M máquinas ($M = 1, \dots, m$). El criterio de optimización más utilizado en la literatura para este problema es la minimización de la fecha máxima de finalización de los trabajos o *Makespan* (C_{max}). Otro objetivo más realista es la minimización del tiempo total de retraso de los trabajos o *tardiness*. Dada una fecha máxima de finalización o *due date* para cada uno de los trabajos ($d_j, j \in N$), el objetivo será minimizar la suma de las diferencias entre el instante real de finalización de los trabajos ($C_j, j \in N$) y su fecha máxima de finalización. Por tanto, la tardanza total o *tardiness* de una secuencia de producción se calcula como $\sum_{j=1}^n \max\{C_j - d_j, 0\}$. El interés de este objetivo radica en el hecho de la gran importancia que tiene en el mundo real que los trabajos estén finalizados en la fecha establecida con los clientes (*make-to-order*).

En este trabajo se presentan y evalúan dos algoritmos genéticos y un algoritmo iterativo voraz para el problema del taller de flujo de permutación con el objetivo de minimizar el retraso total de los trabajos. Los resultados obtenidos indican que las técnicas propuestas presentan mejores resultados que los obtenidos con otras técnicas metaheurísticas propuestas recientemente para este objetivo.

2. Experimento Computacional y Resultados

Los algoritmos propuestos están basados en el algoritmo genético y el algoritmo genético híbrido que podemos encontrar en [1] y el algoritmo iterativo voraz

descrito en [7]. Los algoritmos se han evaluado con los 120 problemas de Taillard [3] que se han ampliado calculando las fechas de finalización o *due dates* de cada trabajo utilizando el procedimiento descrito en [4]. Los dos algoritmos genéticos y el algoritmo iterativo voraz presentados se comparan entre ellos y con el algoritmo basado en Simulated Annealing de [4]. Nos referiremos a estos algoritmos como GA, HGA, IG_LS y SRH respectivamente. Todos los métodos han sido programados en Delphi 2005 y se han ejecutado cinco veces de manera independiente en cada uno de los 120 problemas y se ha medido la eficacia de los algoritmos mediante el incremento porcentual medio (IPM) sobre la mejor solución obtenida para cada problema.

El criterio de parada para todos los métodos es el tiempo transcurrido de CPU siguiendo la expresión $n \cdot (m/2) \cdot t$ milisegundos, donde t puede tomar diferentes valores. Las pruebas se han llevado a cabo en un ordenador PC Pentium IV 3.0Ghz con 1 Gbyte de RAM. En el Cuadro 1 vemos los resultados obtenidos para cada algoritmo fijando el valor de t en la expresión del criterio de parada a 60. Podemos observar que el IG_LS es el algoritmo más eficaz de los comparados, seguido del HGA. Con ambos algoritmos se obtienen resultados mucho mejores que con el SRH.

	GA	HGA	SRH	IG_LS
IPM	2.78	0.74	2.13	0.59

Cuadro 1: Incremento porcentual medio sobre la mejor solución obtenida con $t = 60$

3. Bibliografía

- [1] Ruiz R., Maroto C. and Alcaraz J. (2006). Two new robust genetic algorithms for the flowshop scheduling problem. *Omega, The International Journal of Management Science* 34, 461-476.
- [2] Ruiz R. and Stützle T. (2006). A simple and effective iterated greedy algorithm for the permutation flowshop scheduling problem. *European Journal of Operational Research*. En prensa.
- [3] Taillard, E. (1993). Benchmarks for basic scheduling problems. *European Journal of Operational Research*, 64, 278-285.
- [4] Hasija S. and Rajendran C. (2004). Scheduling in flowshops to minimize total tardiness of jobs. *International Journal of Production Research* 42(11), 2289-2301.

Modelado de redes inalámbricas ad hoc: aspectos energéticos

David Soler¹, José Albiach¹, Eulalia Martínez¹, Pietro Manzoni²

¹dsoler@mat.upv.es, jalbiach@mat.upv.es, eumarti@mat.upv.es,
Departamento de Matemática Aplicada, Universidad Politécnica de Valencia

²pmanzoni@disca.upv.es, Departamento de Informática de Sistemas y
Computadores, Universidad Politécnica de Valencia

Abstract

Modelizamos una red inalámbrica ad hoc mediante un grafo dirigido que contiene toda la información necesaria de la red (periodos de activación de los ordenadores, accesibilidad entre ellos, etc.) para conocer si la información emitida por un ordenador llega o no a su destino y en qué momento, y por tanto, poder medir la eficacia de determinados protocolos de encaminamiento.

Palabras Clave: redes ad hoc.

AMS: 05C20, 05C85, 68M10, 68M12

1. Introducción

Las redes ad hoc, también conocidas como MANETs, son redes inalámbricas que no requieren ningún tipo de infraestructura fija ni administración centralizada. Los ordenadores que las constituyen deben proporcionar también servicios de encaminamiento, retransmitiendo paquetes entre aquellos ordenadores que no tienen conexión inalámbrica directa.

Las redes ad hoc presentan cambios de topología frecuentes e impredecibles debido a la movilidad de sus estaciones. Estas características impiden la utilización de protocolos de encaminamiento desarrollados para redes cableadas y crean nuevos retos de investigación que permitan ofrecer soluciones de encaminamiento eficientes que superen problemas tales como topología dinámica, recursos de ancho de banda y energéticos limitados.

Es fundamental tener alguna metodología formal para poder comparar las diferentes propuestas en este campo y poder determinar la efectiva bondad de dichas soluciones. Hasta el momento se ha utilizando la simulación para validar y comparar diferentes soluciones (ver [1] por ejemplo), pero falta todavía una metodología que dada toda la información sobre la evolución de la red permita decidir cuan lejos de la solución optima está la propuesta en fase de evaluación. Siguiendo el trabajo en esta dirección iniciado en [2], intentamos definir aquí un modelo basado en grafos y que permita representar aspectos como la posibilidad de apagar las estaciones para ahorrar energía.

2. El modelo

El problema a tratar es el siguiente: en una red ad hoc, el ordenador m_i manda al ordenador m_f un mensaje dividido en varios paquetes, cada uno de los cuales se envía en unidades de tiempo consecutivas, utilizando un protocolo de encaminamiento determinado. Supuesto que conocemos en un periodo suficiente de tiempo todos los datos de la red y del protocolo: ubicación en cada instante de los ordenadores, activación o no de los mismos, número de veces que reenvían un mismo paquete, número de paquetes que pueden recibir simultáneamente sin perder información, radio de alcance de las emisiones de cada ordenador en cada momento, etc., se trata de saber con garantía cuántos de estos paquetes le llegarán al destinatario, en qué momento y los caminos que han seguido desde su origen.

Con todos los datos necesarios, construimos un grafo dirigido G donde básicamente cada vértice v_m^h se corresponde con un ordenador m y un instante de tiempo h en que m reenvía o reconoce un paquete de los enviados por m_i y que ha recibido de forma directa a través de otro ordenador m' . El grafo contendrá entonces un arco $(v_{m'}^{h'}, v_m^h)$ con coste (tiempo) $h-h'$. En esta construcción se ha tenido en cuenta que no hayan interferencias y el tope máximo de reenvíos dispuesto a hacer por cada ordenador.

Se demuestra que la construcción de este grafo requiere un tiempo pseudopolinomial en el número de ordenadores de la red ad hoc y que en este grafo existe un camino de un vértice $v_{m_i}^t$ a un vértice $v_{m_f}^{t'}$ si y sólo si con el protocolo usado, el paquete enviado por el ordenador m_i en el instante t , llegará a su destinatario m_f al menos en el instante $t'-r_f$, siendo r_f una constante de tiempo de procesado del paquete por m_f . El mínimo de los t' bajo estas condiciones determinará la primera vez (camino más corto) en que llegará dicho paquete a su destino.

3. Agradecimientos

Este trabajo ha sido subvencionado por el Ministerio de Educación y Ciencia, proyecto *TIN2005-07705-C02-01*.

4. Bibliografía

- [1] Cano, J.C. and Manzoni, P. (2001). Evaluating the energy-consumption reduction in a MANET by dinamically switching-off network interfaces. *Proceedings of the IEEE Symposium on Computers and Communications, Hammamet, Tunisia*, July 2001.
- [2] Soler, D. and Manzoni, P. (2005). A novel energy model for MANETs based on graphs with activation windows. *Proceedings of the 13th Euromicro Conference on Parallel, Distributed and Network-Based Processing*, 126-133. Lugano, Switzerland, February 2005.

Modelización estadística para la predicción de fallos en los conductos de agua potable

Andrés Carrión García¹, Ana Debón Aucejo², Liney A. Manjarres³

¹acarrion@eio.upv.es, ²andeau@eio.upv.es, ³limanhe1@doctor.upv.es

Departamento de Estadística e Investigación Operativa Aplicadas y Calidad,
Universidad Politécnica de Valencia

Abstract

La situación actual de las redes de suministro de agua potable en muchos núcleos urbanos de España, con una edades altas de las instalaciones y unos costes de mantenimiento que no dejan de crecer, hace de especial interés el uso de técnicas estadísticas avanzadas para poder predecir los riesgo de fallo y actuar en consecuencia para garantizar el servicio y reducir costes.

Palabras Clave: Análisis de Supervivencia, Fiabilidad

AMS: 62N05

1. Introducción

En España buena parte de las tuberías de las redes de distribución de agua sobrepasan la edad media de 50 años. Esta situación se convierte en un factor de riesgo que incrementa la probabilidad de fugas o roturas en las tuberías, suponiendo con ello un sobre coste económico considerable a la empresas gestoras y a la sociedad en general, además de una baja en la calidad de un servicio público tan vital como es éste.

2. Objetivo del trabajo

El objetivo de este trabajo es elaborar un modelo estadístico que permita prever o anticipar, en un periodo dado, el comportamiento de las instalaciones a través una serie de indicadores como el número de roturas probables que las tuberías pudieran sufrir, la probabilidad de rotura en un periodo dado, el coste de reparación/mantenimiento de la red, etc. Tal modelo permitirá adoptar una estrategia de gestión proactiva.

3. Metodología

Con el objetivo descrito, se explotará información procedente de compañías de suministro de aguas de la zona mediterránea española. Dicha información hace referencia a dos áreas: características de la red, factores de uso y ambientales; y fechas de fallo, tipo y causas de fallo.

Las características de los datos permitirán el uso del análisis de datos de supervivencia, utilizando modelos que caracterizan la aparición de los fallos, tanto en tiempo como en probabilidad. Se están empleando tanto modelos

paramétricos como no paramétricos cuyos primeros resultados son presentados en este trabajo.

Se recurrirá tanto a modelos paramétricos como no paramétricos, en función del objetivo en cada momento perseguido y de los datos disponibles.

4. Primeros resultados

Se dispone en el momento actual de amplia información referida a una ciudad de la costa mediterránea española, en base a la cual se está aplicando la metodología descrita

La aplicación de este modelo puede contribuir al mejoramiento de la gestión de las redes de agua potable, disminuyendo los costes, reduciendo las pérdidas de aguas subsiguientes a las averías y mejorando los estándares de calidad.

Se están empleando estimadores no paramétricos (Kaplan-Meier) para la evaluación de la función de supervivencia, observándose una sensible coherencia en el comportamiento de los diferentes grupos de elementos de estudio que se han realizado.

Las variables que en los primeros análisis parecen estar más relacionadas con la probabilidad de fallo son las relacionadas con la edad de la instalación, el material de la tubería y el nivel de stress a que ésta se encuentra sometida. El modelo de Cox produce resultados interesantes, que la obtención de nuevos datos sobre la redes deberían mejorar.

A este respecto, debemos citar los problemas que se están encontrando para disponer de datos fiables sobre variables ambientales (tráfico rodado soportado, presión de funcionamiento, lecho de instalación, etc. Están retrasando la obtención de modelos más ajustados.

5. Agradecimientos

Este proyecto se está desarrollando con una ayuda del Ministerio de Educación y Ciencia.

6. Bibliografía

- [1] Eissenbeis, P. (1999). Modélisation statistique de la prévision des défaillances sur les conduits d'eau potable. Doctoral thesis, University Louis Pasteur of Strasbourg, collection Etudes Cemagref n° 17, 190 pp. + appendix, 1999.
- [2] Herz, R. (1998). Exploring rehabilitation needs and strategies for water distribution networks. *J. Water SRT-Aqua*, 47, 1-9.

Modelo de PLEM para la asignación de tareas al personal y su aplicación a un caso real

Ericka Zulema Rodríguez Calvo¹, Anna Maria Coves Moreno²

¹ericka.rodriguez@upc.edu, Instituto de Organización y Control de Sistemas Industriales, Universidad Politécnica de Cataluña

²anna.maria.coves@upc.edu

Abstract

En este trabajo se presenta un modelo de PLEM, para obtener la asignación de tareas óptima, al personal de un centro de servicio, considerando múltiples criterios basados en las prioridades de la organización y del personal. Se ha modelizado un caso real y la experimentación se ha realizado en ILOG-OPL-STUDIO.

Palabras Clave: Problema de Asignación, Optimización Multicriterio, Gestión de Personal.

1. Introducción

Debido a las actuales condiciones del mercado, en que los cambios son constantes, las organizaciones requieren adaptarse para seguir siendo competitivas. Lo que ha hecho que en los últimos años se le dé más importancia a la organización del tiempo de trabajo (OTT), respaldado esto con la adaptación laboral que posibilita la flexibilidad laboral que permite a las organizaciones adaptarse a esos cambios constantes del mercado.

La asignación de tareas (AT) forma parte de la OTT, considerando que la AT parte de un conocimiento previo de la cantidad de capacidad, que se dispone en cada instante de tiempo y de la demanda en las diferentes tareas [1],[2],[3]. Para llevar a cabo la AT se siguen ciertas pautas, las cuales son originadas por aspectos económicos, ergonómicos, sociales, sistemas de trabajo, condicionantes legales, prioridades de la organización, necesidades del personal, etc.

En este trabajo el problema que se modeliza, es el de buscar la asignación óptima del personal a las tareas, tratando de satisfacer la demanda de cada una de las tareas en los diferentes intervalos de tiempo que comprende el horizonte de asignación. La asignación óptima se evalúa considerando diferentes criterios basados en las prioridades de la organización y del personal.

2. Descripción del caso real

El problema corresponde a la asignación del personal a las tareas para una semana de trabajo de un centro de servicio, considerando que ya se dispone de la cantidad de personal (capacidad) en cada uno de los intervalos de tiempo

que componen el horizonte de asignación y no existe la posibilidad de cambiar la cantidad de personal disponible, porque los horarios de trabajo ya fueron establecidos con anterioridad [4].

El horizonte de asignación corresponde a una semana de trabajo, donde se trabaja de lunes a sábado con un horario de 7:00 a 22:00 horas. El personal del que dispone es totalmente polivalente (puede realizar todas las tareas) y está clasificado en categorías. Para la experimentación del problema se han considerado tres distintos niveles de categorías, tres distintos tipos de tareas y cuatro tamaños de personal.

3. Modelo y resolución

La optimización multicriterio se presenta como un modelo PLEM. La función objetivo está formada por cuatro criterios de evaluación, uno se refiere a la utilización del personal en las tareas para las que mejor esté capacitado, un segundo criterio representa la satisfacción de la capacidad ideal, otro busca la satisfacción acorde a la demanda mínima requerida y un cuarto criterio busca la asignación que mejor se aadecue a las preferencias del personal. Se resuelve el problema mediante ILOG-OPL-STUDIO versión 3.7 (librerías de ILOG-CPLEX 9.0). En total se han experimentado 264 ejemplares diferentes del caso.

4. Bibliografía

- [1] Corominas A. and Pastor R., (2000). Un mètode per a la planificació d 'horaris i d 'activitats en serveis amb demanda estacional. *Working paper*, IOC-DT-P-200-10, IOC'UPC, Universidad Politécnica de Cataluña.
- [2] Corominas A., Pastor R. and Rodríguez E., (2005). Rotational allocation of tasks to multifunctional workers in a service industry. *International journal of Production economics*. En prensa.
- [3] Lusa A., (2003). Planificación del tiempo de trabajo con jornada anualizada. *Tesis Doctoral*. Universidad Politécnica de Cataluña.
- [4] Ojeda R.J, (2004). Programación de horarios semanales de trabajadores polivalentes en un centro de servicios. *Tesis Doctoral*. Universidad Politécnica de Cataluña.

Modelos de localización líder-seguidor con demanda elástica

Dolores R. Santos Peñate, Rafael Suárez Vega, Pablo Dorta
González

Departamento de Métodos Cuantitativos en Economía y Gestión,
Universidad de Las Palmas de Gran Canaria
{drsantos, rsuarez, pdorta}@dmc.ulpgc.es

Abstract

Se estudia un modelo de localización competitiva líder-degidor. Se comparan escenarios donde la demanda es inelástica con otros donde ésta es elástica. Se presentan algunos resultados relativos a la discretización del problema del seguidor y a las localizaciones óptimas del líder.

Claves: Localización competitiva, modelo líder-seguidor.

AMS: 90B80, 90B85

1. El modelo estático

Sea F el conjunto de localizaciones candidatas para abrir centros en un mercado M cuya demanda se distribuye en el conjunto P . Dos firmas, A y B , operan en el mercado con centros localizados en los puntos de los conjuntos $X \subset F$ e $Y \subset F$, respectivamente. La demanda captada por las firmas A y B se denota $W_A(X|Y)$ y $W_B(Y|X)$, respectivamente. Cada firma desea maximizar su cuota de mercado o demanda captada.

■ *El problema del seguidor*

La firma A , el líder, opera en el mercado con p centros ubicados en los puntos $X_p = \{x_1, \dots, x_p\} \subset F$. La firma B , el seguidor, quiere entrar en el mercado con r centros localizados en el conjunto $Y_r = \{y_1, \dots, y_r\} \subset F$ que maximiza su cuota de mercado. El problema del seguidor es determinar el conjunto $Y_r^* = \{y_1^*, \dots, y_r^*\} \subset F$ tal que

$$W_B(Y_r^*|X_p) = \max_{Y_r \subset F} W_B(Y_r|X_p).$$

■ *El problema del líder*

No existen centros operando en el mercado. La firma A , el líder, quiere entrar en el mercado abriendo p centros en el conjunto de puntos $X_p = \{x_1, \dots, x_p\} \subset F$ que maximiza su cuota de mercado, considerando que una firma competidora B entrará en el mercado actuando como

un seguidor. El problema del líder es encontrar el conjunto $X_p^* = \{x_1^*, \dots, x_p^*\} \subset F$ tal que

$$W_A(X_p^*|Y_r^*(X_p^*)) = \max_{X_p \subset F} W_A(X_p|Y_r^*(X_p))$$

donde $Y_r^*(X_p)$ es una solución óptima para el seguidor dado que el líder tiene p centros localizados en X_p .

El par $(X_p^*, Y_r^*(X_p^*))$ es una *solución líder-seguidor o solución de Stackelberg*.

El cliente elige el centro más próximo y los empates en la distancia al centro más cercano se resuelven a favor del seguidor. Con el fin de evitar las soluciones triviales, se establece que en una localización no puede abrirse más de un centro. El mercado está representado por una red, $M = F = N(V, E)$, siendo $P = V = \{v_i\}_{i=1}^n$ el conjunto de nodos y E el conjunto de aristas. La demanda en $v \in V$ es $w(v) (\geq 0)$ y $l(e) (\geq 0)$ es la longitud de la arista. Para cada par de conjuntos finitos $X, Y \subset N(V, E)$, $D(v, X)$ denota la distancia entre el nodo v y el conjunto X , dada por $D(v, X) = \min\{d(v, x) : x \in X\}$. Se define $V(Y|X) = \{v \in V : D(v, Y) \leq D(v, X)\}$, y $V(Y|X)^c = V \setminus V(Y|X)$. Si A y B abren centros en X e Y , respectivamente, la demanda captada por las firmas A y B es

$$W_A(X|Y) = \sum_{v \in V(Y|X)^c} w(v) G_v(D(v, X))$$

$$W_B(Y|X) = \sum_{v \in V(Y|X)} w(v) G_v(D(v, Y))$$

donde $G_v(\delta)$ es una función decreciente tal que $0 < G_v(\delta) \leq 1$, $\forall \delta \geq 0$, $\forall v \in V$. Si $G_v(\delta) = 1$, $\forall \delta \geq 0$, $\forall v \in V$, se obtiene el caso de demanda inelástica.

2. Agradecimientos

Este trabajo ha sido parcialmente financiado por el Ministerio de Educación y Ciencia (España) y FEDER. Proyecto MTM2005-09362-C03-03.

3. Bibliografía

- [1] Hakimi, S.L. (1990). Location with spatial interactions: Competitive locations and games. In: Mirchandani P.B. and Francis R.L. (eds.), *Discrete Location Theory*. Wiley-Interscience, 439-478.

Modelos de subasta en un duopolio eléctrico con costes correlados

Estrella Alonso Pérez¹, Juan Tejada Cazorla²

¹ealonso@upcomillas.es, Universidad Pontificia Comillas de Madrid, Spain

²jtejada@mat.ucm.es, Universidad Complutense de Madrid, Spain

Abstract

Se calculan los equilibrios Bayesianos de Nash estrictamente monótonos, ingresos esperados por las empresas y el pago que el Operador del Mercado espera realizar en una familia paramétrica de modelos de subasta aplicables al Mercado Eléctrico. Dicha familia contiene, en particular, a los modelos Uniforme, de Vickrey y Discriminatorio.

Palabras Clave: costes estocásticos, mercado eléctrico

1. Introducción

Se van a suponer las siguientes hipótesis en el presente trabajo:

- Hay dos empresas con sendas unidades de producción, neutrales al riesgo $i = 1, 2$, interesadas en el despacho de unidades de electricidad en algún Mercado Eléctrico. Supondremos que ambas empresas tienen capacidades idénticas $k = 1$ y perfectamente divisibles.

- La función de coste de la empresa i es $g(q_i, \theta_i)$, $i = 1, 2$, donde q_i es la cantidad generada por i (la producción por encima de la capacidad es imposible) y θ_i representa el tipo de la empresa i , que es información privada de ésta. Se supondrá, sin pérdida de generalidad, que $g(q_i, \theta_i) = c(q_i)\theta_i$ donde $c(q)$ es una función creciente y diferenciable.
- Se supondrá que los tipos $\theta_i \in [a_1, a_2]$ dependen del precio de la materia prima. Para modelizar la incertidumbre que las empresas tienen sobre dicho precio, se considerará una variable aleatoria continua S , con función de densidad f_S definida en $\Omega_S \subset \mathbb{R}$, que representa dicho precio. Denotaremos como $f_{\Theta_i|S=s}$ a la función de densidad de la variable aleatoria $\Theta_i|S=s$, $i = 1, 2$, consideradas v.a.i.i.d..

D representará la demanda de un día y en uno de los períodos de ese día. D se supondrá inelástica al precio.

Cada empresa, independiente y simultáneamente, realiza una oferta (una acción) $b_i \in [0, b_{\max}]$, $i = 1, 2$, donde especifica el mínimo precio al que está dispuesto a vender la totalidad de su capacidad.

Una estrategia para la empresa i es, entonces, una función oferta de la forma $b_i(\cdot) : [a_1, a_2] \rightarrow [0, b_{\max}]$, $i = 1, 2$. Donde b_{\max} es el mayor precio aceptado por el Operador del Mercado. Supondremos que $b_i(\theta_i) = b(\theta_i)$ es utilizada por todos los jugadores¹, donde b es una función estrictamente creciente y diferenciable, satisfaciendo la condición² $b(a_2) = m$. Al conjunto de estas hipótesis lo llamaremos **Modelo Simétrico Correlado**.

En base a estas ofertas, el Operador del Mercado distribuye el despacho de electricidad. Ordena las ofertas de menor a mayor, de tal forma que el de menor

¹Es lógico suponer, por la simetría del problema, que en equilibrio se comporten de la misma manera estratégica. Por ello nos centraremos en la búsqueda de equilibrios simétricos.

²Supondremos además que $b_{\max} = m$, es decir el subastador ajusta el tope a la máxima oferta realizada en equilibrio por las empresas cuando esto sea posible.

oferta despacha primero. Si su capacidad no es suficiente para satisfacer la demanda, el de mayor oferta despacha la demanda residual, siempre y cuando ésta sea inferior a su propia capacidad. Luego la cantidad asignada para el despacho a la empresa i es:

$$Q_i(b_i, b_j) = \begin{cases} \min(1, D) & \text{si } b_i < b_j \\ 0 & \text{si } b_i > b_j \text{ y } \min(1, D) = D \\ \min(D - 1, 1) & \text{si } b_i > b_j \text{ y } \min(1, D) = 1 \end{cases}$$

Luego el beneficio obtenido por la empresa i es:

$$B_i(\theta_i, b_i, b_j) = Q_i(b_i, b_j)\rho_i(b_i, b_j) - c(Q_i(b_i, b_j))\theta_i$$

Donde $\rho_i(b_i, b_j)$ es el precio por unidad eléctrica que el Operador del Mercado paga a i si las ofertas pujadas son (b_i, b_j) (**Precio del Mercado**). Todos los aspectos de este juego Bayesiano, además del modelo de subasta a utilizar para las transacciones del Mercado Eléctrico, son de conocimiento público. El Precio del Mercado dependerá del modelo de subasta que se está empleando para realizar las transacciones. Hay tres modelos de subasta principales aplicables al Mercado Eléctrico: *Modelo de subasta Uniforme*, *Modelo de subasta Discriminatoria* y *Modelo de subasta Vickrey*.

En el modelo Uniforme, el precio que el Operador del Mercado paga a cada empresa, por unidad eléctrica suministrada, es igual a la mayor oferta aceptada. Todas las empresas que despachan en el mercado reciben el mismo precio por unidad. (e.d. $\rho_i(b_i, b_j) = \rho(b_i, b_j)$).

En el modelo Discriminatorio, el precio que el Operador del Mercado paga a cada empresa, por unidad eléctrica suministrada, es igual a su propia oferta. Las empresas que despachan en el mercado reciben precios distintos por unidad eléctrica. (e.d. $\rho_i(b_i, b_j) = b_i$).

En el modelo Vickrey, la regla mediante la cual el Operador del Mercado establece el precio es algo más complicada. A cada empresa le pagan, por cada unidad de electricidad despachada, el precio correspondiente a la unidad que desplaza si no ofertase dicha empresa. Por tanto a la empresa de mayor oferta le paga por unidad de demanda residual (si es mayor que cero) la mayor oferta aceptada, b_{\max} . Si la empresa i es la empresa de menor oferta y, por tanto, la que entra primero en el mercado, le paga la oferta contraria b_j si $D < 1$. Y si $D \geq 1$ esta empresa i , de menor oferta, recibe dos precios distintos por unidad: la oferta máxima permitida, b_{\max} , por $D - 1$ unidades y la oferta contraria, b_j , por las restantes.

En este trabajo se analiza una familia paramétrica de modelos de subasta. Esta familia se denomina Modelo de Subasta General. Se calculan los equilibrios Bayesianos de Nash, ingresos esperados por las empresas y el pago que espera hacer el Operador del Mercado. Este planteamiento permite comparar los modelos de subasta antes descritos en un contexto más amplio en el que aparecen nuevos modelos que tienen un cierto interés.

Modelos matemáticos para el problema de la ubicación de productos en estanterías

M. Palmer¹, J. Albarracín¹, J. P. García¹, M. Babiloni¹, A. Esparcia²

¹marpalga@omp.upv.es, Departamento de Organización de Empresas,
Universidad Politécnica de Valencia

²anna@iti.upv.es, Instituto Tecnológico de Informática, Universidad
Politécnica de Valencia

Abstract

El problema de la asignación de espacios físicos a los productos en una gran superficie comercial es cada vez más importante dada la rapidez con la que cambian las preferencias de los consumidores. En el presente trabajo se plantean y resuelven diversos modelos matemáticos para obtener soluciones acorde a los objetivos de la empresa.

Palabras Clave: programación matemática, asignación, logística

1. Introducción

El presente trabajo pretende aportar soluciones en el campo de la distribución de espacios en grandes superficies comerciales. El desarrollo de nuevos tipos de establecimientos, caracterizados por un mayor tamaño, no ha permitido resolver el problema, debido a que se ha producido un fuerte incremento en el número de productos ofrecidos en las tiendas, que supera el mencionado aumento de la superficie en los establecimientos.

El reparto de espacio que se asigna a cada uno de los productos, no es más que la consecuencia de la estrategia planteada por el detallista para conseguir sus objetivos. A continuación, se analizan con detalle los elementos que forman parte del problema descrito.

2. Definición del problema

El problema se puede caracterizar mediante la definición de los elementos que forman parte del problema:

- **Surtido:** Estará formado por aquellos artículos que los propietarios del establecimiento consideran oportuno para abastecer la demanda esperada. Dicho surtido estará subdividido en grupos homogéneos con funcionalidades distintas (secciones, categorías, familias y referencias).
- **Espacios:** La organización del espacio de venta es una de las tareas fundamentales a desarrollar, puesto que una distribución racional del espacio entre los clientes, los productos y el personal permitirá alcanzar los objetivos fijados en cuanto a la rentabilidad que el establecimiento espera alcanzar. Se pueden diferenciar pasillos, mobiliario, góndolas, estanterías y cabeceras.

- **Criterios de asignación:** Se puede identificar como objetivo el minimizar la dispersión de aquellas referencias que es interesante estén lo más cerca posible alejando aquellas que no lo son tanto, esto permitirá una compra agradable al consumidor así como una fácil gestión de los recursos implicados por parte del detallista.

3. Modelos matemáticos

Dada la complejidad del problema se procede a la enumeración de las diferentes situaciones que se pueden tipificar para su posterior modelización y resolución:

- Las necesidades de espacio cuantificadas en módulos, para ubicar las diferentes categorías que componen el surtido son iguales al número de módulos disponibles en las estanterías.
- Cuando las necesidades de espacio de las categorías y los módulos disponibles en el mobiliario no son coincidentes.

4. Experiencias computacionales

Los modelos matemáticos presentados en el apartado anterior se ha resuelto utilizando el software de programación matemática CPLEX presentando los resultados computacionales obtenidos.

5. Agradecimientos

El presente trabajo se enmarca dentro de las investigaciones realizadas conjuntamente entre el Departamento de Organización de Empresas y el Instituto Tecnológico de Informática de la Universidad Politécnica de Valencia dentro del proyecto ANALOG (Aplicación de Algoritmos Naturales a problemas de logística) subvencionado por el Instituto para la Pequeña y Mediana Empresa de la Generalitat Valenciana.

6. Bibliografía

- [1] Mouton, M. (1997). *Merchandising estratégico*. Gestión 2000.
- [2] Anderson, E.E., Amato, H.N. (1974): “A mathematical model for simultaneously determining the optimal brand collection and display area allocation”, Operations Research, vol. 22, january-february, pp. 13-21
- [3] Hansen, P., Heinsbroek, H. (1979), “Product selection and space allocation in supermarkets”. European Journal of Operations Research, vol. 3, pp. 474-484.
- [4] Corstjens, M., Doyle, P. (1981), “A model for optimizing retail space allocations” Management Science, vol. 27, nº 7, pp. 822-833.

MOM: Un Sistema Software para la optimización del tráfico ferroviario

P. Tormos¹, A. Lova¹, F. Barber², L. Ingolotti², M. Abril², M.A. Salido²

¹{ptormos,allova}@eio.upv.es, Departamento de Estadística e Investigación Operativa Aplicadas y Calidad, Universidad Politécnica de Valencia

²{fbarber,lingolotti,mabril,msalido}@dsic.upv.es, Departamento de Sistemas Informáticos y Computación, Universidad Politécnica de Valencia

Abstract

Las recientes regulaciones en el sector ferroviario han supuesto la separación entre los gestores de la infraestructura y los operadores que requieren su uso. Este hecho ha incrementado la necesidad de optimizar el uso de la infraestructura existente así como de ofrecer un servicio de calidad a los usuarios. En este trabajo presentamos MOM (Módulo de Optimizador de Mallas) un sistema de ayuda a la toma de decisiones para la optimización del tráfico ferroviario. La herramienta desarrollada permite obtener buenas soluciones a problemas complejos de tráfico ferroviario así como un alto nivel de interacción con el horarista. MOM está integrado de forma operativa en el sistema automático utilizado por el ADministrador de Infraestructuras Ferroviarias (ADIF).

Palabras Clave: Sistema de Ayuda a la toma de decisiones, Secuenciación, Horarios Ferroviarios

1. Introducción

Las recientes regulaciones del sector ferroviario han supuesto la separación efectiva entre los gestores de la infraestructura y los operadores que solicitarán el uso de la misma. Al mismo tiempo son un nuevo marco de competencia en el que se hace indispensable optimizar el uso de la infraestructura existente. Para ello uno de los objetivos del gestor de infraestructuras es lograr a partir de los requerimientos de los operadores, horarios optimizados que permitan maximizar el uso de la capacidad de la red y conocer los puntos críticos de la misma. En este proceso resulta indispensable disponer de sistemas automáticos de ayuda a la toma de decisiones.

En este trabajo presentamos un sistema software de ayuda a la toma de decisiones que proporciona al horarista las herramientas necesarias para llevar a cabo la secuenciación de trenes de modo que se verifiquen las restricciones existentes y se optimicen los criterios de calidad que se establezca.

2. Railway Scheduling Problem

El problema implica la programación (generación de horarios) de trenes que circulan en ambos sentidos en una línea de ferrocarril de vía única/doble por tramo (segmento de vía entre dos apeaderos o estaciones) de tal modo que se

verifiquen las restricciones existentes y se optimice uno o varios criterios de eficiencia.

Las restricciones se han definido conjuntamente con el personal de ADIF y se clasifican en: (i) **restricciones de tráfico**, que garantizan que no se producen conflictos de cruce y adelantamiento; (ii) **restricciones de usuario**, tales como tipo y número de trenes a secuenciar, frecuencia, intervalo de salida/llegada y (iii) **restricciones por la topología** de la red tales como número de vías en estación o tiempo de cierre en las estaciones. De acuerdo a las restricciones establecidas el objetivo es determinar los horarios de salida/llegada de cada tren a cada dependencia de la red de modo que se optimice un criterio de eficiencia seleccionado por el usuario: (FO1) minimizar el tiempo de recorrido (FO2) equilibrar los retrasos entre los trenes de ambos sentidos.

3. Arquitectura general del sistema

El sistema tiene tres componentes principales:

- El **interfaz**: a través del cual es posible establecer ciertos atributos correspondientes a las restricciones, infraestructura de la red y los trenes así como configurar el propio optimizador.
- El **optimizador**: es un procedimiento iterativo que aplica heurísticas guiadas por la función objetivo. Es posible escoger entre dos modos de optimización: *Normal* o *Flexible*. En el segundo caso es posible relajar determinadas restricciones.
- **Presentación de resultados**: informes gráficos interactivos, mediante diagramas tiempo-espacio en los que se representa la posición del tren en un tiempo dado; informes de texto, mediante tablas se proporciona la información del horario resultante.

La eficiencia del sistema desarrollado se ha contrastado utilizando mallas reales y actualmente MOM está siendo utilizado por ADIF.

4. Agradecimientos

Este trabajo ha sido parcialmente subvencionado por los proyectos TIN2004-06354-C02-01 (MEC-FEDER) y 20051019 de M.Fomento. Agradecemos a José Estrada Guijarro su apoyo y colaboración en el desarrollo de este trabajo.

5. Bibliografía

- [1] MOM: Módulo Optimizador de Mallas
<http://www.dsic.upv.es/users/ia/gps/MOM/>

Multidimensional Assignment Games

F. Perea¹

¹perea@us.es Department of Statistics and Operations Research.
University of Seville

Abstract

El problema de asignación multidimensional (MDA) es una generalización natural del problema de asignación clásico. En este trabajo analizamos la posible cooperación entre los agentes involucrados en un MDA. Debido a que los MDA son NP-duros, hacemos uso de algoritmos de aproximación para repartir beneficios entre los jugadores.

Keywords: Multidimensional assignment, cooperative game theory, approximation algorithms.

1. Introduction

Given a bipartite graph $G = (N, A)$ so that $N = N_1 \cup N_2$, $N_1 \cap N_2 = \emptyset$ and weights $b_a \geq 0 \forall a \in A$, the assignment problem consists of finding a matching from elements in N_1 to elements in N_2 so that the sum of weights is maximized [4]. Consider a w -partite graph $G = (N, A)$, that is, there exist N_1, \dots, N_w such that $N_1 \cup \dots \cup N_w = N$, $N_i \cap N_j = \emptyset \forall i \neq j$. Suppose that the association constituted by one element of each group gives rise to a benefit $c_a \geq 0$. The problem of finding the set of associations so that the sum of weights is maximized is called a Multidimensional Assignment problem, and was first introduced in [3]. MDA problems are combinatorial optimization problems known to be NP-hard. The case $w = 2$ leads us to the classical Assignment Problem.

In this paper it is considered the case in which elements of the graph that gives rise to a MDA problem have conflicting objectives. Nevertheless, the cooperation among them is allowed so the general benefit is maximized. It directly follows that cooperative game theory can be used to analyze such a situation [1].

2. MDA games: definition and properties

Consider the MDA problem defined by the graph $G = (N_1, \dots, N_w, A)$. The corresponding MDA game is defined as (N, v) where

1. $N = N_1 \cup \dots \cup N_w$
2. $v(S)$ is the optimal value of the MDA problem arising from graph $G = (N_1^S, \dots, N_w^S, A)$, where $N_j^S = N_j \cap S \forall j = 1, \dots, w, \forall S \subset N$.

Let (N, v) be a w -dimensional assignment game. One has that (N, v) is:

- w -normalized, that is, $v(S) = 0 \forall S : |S| < w$
- monotonic, that is, $v(S) \leq v(T) \forall S \subset T \subset N$
- superadditive, that is, $v(S) + v(T) \leq v(S \cup T) \forall S, T \subset N : S \cap T = \emptyset$.

2.1. The core

The core of MDA games is proven to be, in general, empty. The study of conditions for its non-emptiness is analyzed in this section.

3. Allocations and approximation algorithms

Due to the fact that MDA games are not balanced, the search for other allocations with good properties becomes more important. Note that, because of the NP-hardness of the MDA problem, the exact knowledge of the characteristic function is impracticable, since we should be able to solve $O(2^n)$ NP-hard LP problems. This is enough argumentation to look for other techniques to allocate the benefits generated. A class of approximation algorithms named *K-SGTS* are used in this paper [2].

Proposition 2 *Let (N, v) be a w -MDA game and x be an allocation obtained from the approximation algorithm K-SGTS for the case $K = 1$. One has that:*

1. x is a preimputation of (N, v) .
2. x satisfies the individual rationality principle.
3. $\frac{v(N)}{w} \leq x(N) \leq v(N)$.

4. Bibliography

- [1] Forgó, F., Szép, J. and Szidarovsky, F. "Introduction to the Theory of Games." Kluwer Academic Publishers, Dordrecht (1999).
- [2] Perea, F., de Waard, H.W. A new class of Heuristic Polynomial Time algorithms for solving the Multidimensional Assignment Problem. *Submitted to Information Fusion 2006*.
- [3] W. Pierskalla, (1968) The multidimensional assignment problem. *Operations Research* 16, 422-431.
- [4] Wosley, L.A. (1998) *Integer Programming*. John Wiley & Sons.

New Results on the Public Good Index

J. M. Alonso-Mejide¹, B. Casas-Méndez², M. J. Holler³, S. Lorenzo-Freire⁴

¹mejide@lugo.usc.es, Departamento de Estadística e Investigación Operativa,
Universidad de Santiago de Compostela

²eigradu3@usc.es, Departamento de Estadística e Investigación Operativa,
Universidad de Santiago de Compostela

³holler@econ.uni-hamburg.de, Institute of SocioEconomics, University of
Hamburg

⁴silvia_lorenzo@uvigo.es, Departamento de Estadística e Investigación
Operativa, Universidad de Vigo

Abstract

Holler (1982) proposed a power index for simple games: the Public Good Index. In this paper, we present a new characterization of this index and a method to compute it by means of the multilinear extension of the game. We apply our method to a real-world example taken from the political field.

Keywords: Power index, axiomatization, multilinear extension

1. Introduction

Simple games can be used to model decision-making processes and fruitful applications have been done in the field of political science. Power is one of the main concepts in this science and different power indices have been defined in order to assess the a priori distribution of power in the voting body. Power is a very difficult concept to measure and no consensus has been obtained in order to select a power index.

The main power indices that we can find in the literature are the Shapley-Shubik index (Shapley and Shubik, 1954), the Banzhaf index (Banzhaf, 1965), the Johnston index (Johnston, 1978), the Deegan-Packel index (Deegan and Packel, 1979), and the Public Good Index, also called Holler index (Holler, 1982). These measures of voting power are based on an evaluation of an actor's relative importance to each of the coalitions which might be formed.

In the context of simple games, an agent is considered critical when his elimination from a winning coalition turns this coalition into a losing coalition. A minimal winning coalition is one such that all its members are critical. The Public Good Index is determined by the number of minimal winning coalitions containing the voter divided by the sum of such numbers across all the voters.

The choice of an index can depend on the situation. To facilitate this choice, some desirable properties have been introduced in the context of power indices.

In Lorenzo-Freire et al. (2006) new characterizations of Deegan-Packel and Johnston power indices are obtained with monotonicity properties. In this paper, we provide a new characterization of the Public Good Index by means of a property of monotonicity with a similar flavour.

One of the main difficulties with these indices is that computation generally requires the sum of a very large number of terms. Owen (1972) defined the multilinear extension of a game. We introduce the procedure to calculate the Public Good Index by means of the multilinear extension and we apply it to some political examples.

2. Acknowledgement

Authors acknowledge the financial support of Spanish Ministry for Science and Technology and FEDER, through projects BEC2002-04102-C02-01, BEC2002-04102-C02-02, SEJ2005-07637-C02-01/ECON, and SEJ2005-07637-C02-02/ECON and *Xunta de Galicia* through project PGIDIT03PXIC20701PN.

3. Bibliography

- [1] Banzhaf, J.F. (1965). Weighted voting doesn't work: A mathematical analysis. *Rutgers Law Review* 19, 317-343.
- [2] Deegan, J. and Packel, E.W. (1979). A new index of power for simple n-person games. *International Journal of Game Theory* 7, 113-123.
- [3] Holler, M.J. (1982). Forming coalitions and measuring voting power. *Political Studies* 30, 262-271.
- [4] Johnston, R.J. (1978). On the measurement of power: some reactions to Laver. *Environment and Planning A* 10, 907-914.
- [5] Lorenzo-Freire, S., Alonso-Mejide, J.M., Casas-Méndez, B., and Fiestras-Janeiro, M.G. (2006). Characterizations of the Deegan-Packel and Johnston power indices. Forthcoming in *European Journal of Operational Research*.
- [6] Owen, G. (1972). Multilinear extensions of games. *Management Science* 18, 64-79.
- [7] Shapley, L.S. and Shubik, M. (1954). A method for evaluating the distribution of power in a committee system. *American Political Science Review* 48, 787-792.

Non Advantageous Merging in cost spanning tree problems. A new characterization of the Bird rule.

Maria Gómez - Rúa¹, Juan J. Vidal - Puga²

¹mariarua@uvigo.es, Dpto. Estatística e I.O., Universidade de Vigo

²vidalpuga@uvigo.es, Dpto. Estatística e I.O., Universidade de Vigo

Abstract

We introduce a new property in cost spanning tree problems called Non Advantageous Merging. This property means that no group of agents can be better off claiming to be a single node. We provide a new characterization of the Bird rule using Non Advantageous Merging.

Keywords: Cost spanning tree problems, Non Advantageous Merging, the Bird rule.

1. Introduction

Cost spanning tree problems (*cstp*) modelize situations where a group of agents, located at different geographical points, want some particular service which can only be provided by a common supplier, or *source*. For instance, some houses in a village may want to be connected to a common water source or to a power plant. Agents will be served through connections which entail some cost. The agents are not concerned with whether they are connected directly or indirectly to the source.

We assume that agents construct a minimal cost spanning tree (*mcst*). An algorithm for building a *mcst* is provided by Prim (1957). But building a *mcst* is only a part of the problem. Other important issue is how to divide the cost associated with the *mcst* between the agents. Different *rules* give different answers to this question.

One of the most common topics is the axiomatic characterization of rules. The idea is to propose "desirable" properties and to study which of them characterize each rule. Properties often help agents to compare different rules and to decide which rule is preferred in a particular situation.

In the literature there are several rules. The *Bird rule* was proposed by Bird (1976) and was recently characterized by Dutta and Kar (2004).

We introduce a new property in this class of problems called *Non Advantageous Merging*. The idea behind this property is the following: in *cstp* there exists a planner, who wants to construct a network to connect all the agents to a source. To do that, they ask them about which are interested to be connected.

In this kind of situations agents may have incentives to join in advance and to present themselves as a single node in order to get advantage. A rule satisfies Non Advantageous Merging if the agents have no incentives to do this.

We show that if the agents pay their own connection cost, Non Advantageous Merging is satisfied. This rule is the Bird rule. Moreover, we provide an new axiomatic characterization of the Bird rule using Non Advantageous Merging.

2. Acknowledgement

Financial support from the Spanish Ministerio de Ciencia y Tecnología and FEDER through grant SEJ2005-07637-C02-01/ECON is gratefully acknowledged.

3. Bibliography

- [1] Bird C.G. (1976). "On cost allocation for a spanning tree: A game theoretic approach". Networks 6: 335-350.
- [2] Dutta B. and Kar A. (2004) . "Cost monotonicity, consistency and minimum cost spanning tree games". Games and Economic Behavior 48: 223-248.

Non-cooperative cost spanning tree games with budget restrictions

Gustavo Bergantiños¹, Leticia Lorenzo²

¹gbergant@uvigo.es, Dpto. Estadística e I.O., Universidad de Vigo

²leticiap@uvigo.es, Dpto. Estadística e I.O., Universidad de Vigo

Abstract

We extend the non-cooperative game associated with the cost spanning tree problem introduced in Bergantiños and Lorenzo (2004) to situations where agents have budget restrictions. We study the Nash equilibria, subgame perfect Nash equilibria, and strong Nash equilibria of this game.

Keywords: Game Theory, cost spanning tree problem, equilibria.

1. Introduction

This paper studies cost spanning tree problems (*cst*). Consider a group of agents, located at different geographical places, who want some particular service, which can only be provided by a common supplier, called the source. The agents will be served through connections which entail some cost and they do not bother whether they are connected directly or indirectly to the source.

The first problem faced in *cst* problems is how to construct a minimal cost spanning tree (*mcst*), *i.e.*, the tree that serves all agents with the minimal total cost (Kruskal (1956), Prim (1957)). But another important issue is how to allocate the cost associated with the *mcst* among agents. Claus and Kleitman (1973) introduce this problem, but it is Bird (1976) the first to address this issue using Game Theory. Bird associates with any *cst* problem a cooperative game proposing a cost allocation rule called the Bird's rule. This paper generates further literature, for instance, Granot and Huberman (1984); Feltkamp, Tijs, and Muto (2000), and Kar (2002), but always within the cooperative Game Theory.

Bergantiños and Lorenzo (2004) show that in some cases agents do not cooperate to build the *mcst* and share the total cost in some specific way, but act independently and connect in several stages to the source or to a previously connected agent. They model this situation from the point of view of non-cooperative Game Theory using an extensive form game such that in each stage of the game unconnected agents decide, simultaneously, whether or not they want to connect to the source or some of the agents who are already connected. The game ends when everybody is connected or all unconnected agents decide simultaneously to remain unconnected. They study the Nash

equilibria (*NE*) and subgame perfect Nash equilibria (*SPNE*) of this game assuming, as in all the literature about *cst* problems, that agents want to be connected, even if they have to pay the highest connection cost in the network.

We consider this assumption may be unrealistic in some cases. For instance, the utility of being connected might not be enough to compensate agents if they have to pay high connection costs. Another simple reason is that agents often do not have enough money to pay high costs. Thus, in this paper we assume that, for each agent i there exists $\alpha_i \in \mathbb{R}$, that can be interpreted in two possible ways: firstly, the utility that agent i obtains when connected to the source and secondly, the money that agent i has available.

Following Bergantiños and Lorenzo (2004) we associate with each *cst* problem with budget restrictions a non-cooperative game. We study the *NE* and *SPNE* of this game and prove that the set of payoffs associated with *NE* coincides with the set of payoffs associated with *SPNE*.

Given an order among the agents, let us assume that they connect sequentially according to this order to one of the cheapest options available: a previously connected agent or the source. Thus, the set of payoffs associated with *SPNE*, and hence *NE*, coincides with the set of payoffs obtained when agents connect sequentially according to an order belonging to a subset of the set of permutations we call admissible permutations.

We also address the optimality issue studying which of the equilibria induce a (Pareto) undominated payoff in the set of all payoffs induced by strategies where no agent pays more than the amount he can afford. There are some *SPNE* (and hence *NE*), whose payoffs are dominated, but the set of undominated payoffs associated with *SPNE* (and hence *NE*) coincide with the set of payoffs associated with Strong Nash Equilibria (*SNE*). In non-cooperative games, *SNE* may not exist, but in this non-cooperative framework, *SNE* always exist and “coincide” with the set of undominated *SPNE*.

2. Contact

Address: Departamento de Estadística e Investigación Operativa

Facultad de Ciencias Económicas y Empresariales.

Lagoas - Marcosende s/n 36310 - Vigo (Pontevedra) Spain.

Telephone: (+34) 986812443 (+34) 988368773

Fax: (+34) 986812401

e-mail: leticiap@uvigo.es

www: <http://webs.uvigo.es/leticiap>

Nuevos criterios generalizados para modelar las preferencias del decisor en los métodos de relaciones de superación

Gabriela Fernández Barberis, María del Carmen Escribano Ródenas

Departamento de Métodos Cuantitativos, Facultad de Ciencias Económicas y Empresariales, Universidad San Pablo – CEU, Madrid
ferbar@ceu.es, escrod@ceu.es

Abstract

Los criterios generalizados son funciones de preferencia que se utilizan dentro del Proceso de Decisión, en la etapa de Modelación de las Preferencias. En el presente trabajo se proponen nuevos tipos de criterios generalizados en los Métodos PROMETHEE, generándose así, un amplio perfil de preferencias para el decisor.

Palabras Clave: Métodos de Relaciones de Superación, Criterios Generalizados, Funciones y Estructuras de Preferencia

AMS: 91B06

1. Introducción

Los Métodos de Relaciones de Superación constituyen una de las categorías más importantes, dentro del apasionante mundo de la Ayuda a la Decisión Multicriterio. Dentro de esa categoría, los métodos PROMETHEE han adquirido gran aceptación, principalmente debido a que son fácilmente comprensibles por el decisor y de sencilla aplicación.

Tales métodos, tal y como fueron originalmente propuestos, ofrecen la posibilidad de elegir entre seis tipos de criterios generalizados diferentes para representar la estructura de preferencias que el decisor tiene en mente. Los criterios generalizados son funciones de preferencia, que se utilizan dentro del Proceso de Decisión, en la etapa de Modelación de las Preferencias.

Dado que los problemas de decisión reales a los que se enfrenta un decisor son cada vez más variados y complejos, se proponen nuevos tipos de criterios generalizados, que ofrecen al analista la posibilidad de recurrir a otro tipo de funciones, distintas de las conocidas y definidas hasta ahora, en la ardua tarea de representar, de la forma más fiable y representativa posible, el esquema de preferencias que subyace en la mente del decisor. De esta forma, se enriquece notablemente la metodología de decisión, incorporando mayor realismo, fiabilidad y robustez a cualquier problema de Decisión Multicriterio Discreta.

2. Los criterios generalizados tradicionales

En los Métodos PROMETHEE una de las etapas más importantes es

la de Enriquecimiento de la Estructura de Preferencias. Para lograr tal enriquecimiento, a cada criterio inicial se le asocia una función de preferencia que recibe el nombre de criterio generalizado.

La elección del tipo de criterio generalizado no resulta una tarea sencilla, ya que deben definirse, según sea el tipo correspondiente, umbrales de indiferencia y/o preferencia. Por ello, es muy importante que se efectúe interactivamente entre el decisor y el analista, para reflejar de la manera más objetiva posible los matices o escalas de preferencia que el decisor tiene en mente. La labor del analista, reviste en esta etapa, una importancia trascendental.

3. Definición de nuevos criterios generalizados para modelar las preferencias del decisor

En el presente trabajo se proponen tipos de criterios adicionales a los tradicionales y que resultan de la combinación de dos o más de los seis tipos existentes. Sin lugar a dudas, la posibilidad de disponer de nuevas funciones de preferencia que permitan reflejar con mayor precisión las preferencias del decisor, ampliando la gama de preferencias y reflejando distintos matices en la intensidad de las mismas, representa una mejora extraordinaria de la metodología en su conjunto.

4. Conclusiones

Los nuevos criterios generalizados propuestos mejoran notablemente la etapa de Enriquecimiento de la Estructura de Preferencias, ampliando la gama de posibilidades que se presentan al decisor a la hora de reflejar la estructura que posee en su mente. Para que la labor concluya con éxito, es sumamente importante que el analista asista y asesore adecuadamente al decisor.

Podrían definirse otras funciones de preferencias y asociarlas a los criterios originales, pero es importante tener en cuenta que esta tarea debe efectuarse muy cuidadosamente para no distorsionar los resultados finales.

5. Bibliografía

- [1] Fernández Barberis, G. (1993): New Preference Structures for Multiple Criteria Decision Making: its extension to PROMETHEE Methods. *Central European Journal for Operational Research and Economics*, Vol 2, N° 1. PROFIS. Publishing House. Bratislava.
- [2] Roubens, M.; Vincke, P. (1985). *Preference Modelling*. Springer Verlag.
- [3] Roy, B.; Vincke, P. (1987). Pseudo-Orders: definition, properties and numerical representation. *Mathematical Social Sciences*, 14, pp. 263-274. North Holland.

Obtención de referentes más cercanos en la frontera eficiente DEA

Inmaculada Sirvent¹, Juan Aparicio², José Luis Ruiz³

¹isirvent@umh.es, ²j.aparicio@umh.es, ³jlruiz@umh.es

Centro de Investigación Operativa, Universidad Miguel Hernández

Abstract

En este trabajo proponemos un enfoque general para encontrar los referentes (targets) más cercanos a una unidad dada. Para las unidades ineficientes, esto supone determinar direcciones de mejora que permiten alcanzar la eficiencia con el menor esfuerzo, en contraposición con las proporcionadas habitualmente en DEA, que requieren el esfuerzo máximo.

Palabras Clave: Data Envelopment Analysis (DEA), Targets, Distancia mínima a la frontera

1. Introducción

El *Data Envelopment Analysis* (DEA) es una metodología utilizada para evaluar la eficiencia relativa de un conjunto de unidades (DMUs) que utilizan uno o varios *inputs* para producir uno o varios *outputs*. En general, para cada una de las unidades evaluadas, los modelos DEA proporcionan el valor de una medida de eficiencia relativa así como información acerca de los *peers* y *targets* que han sido utilizados en dicha evaluación. Para una unidad ineficiente, los peers son unidades eficientes observadas que dicha unidad ineficiente podría intentar imitar con el fin de alcanzar la eficiencia. Los targets son las coordenadas de la proyección de la DMU evaluada en la frontera eficiente y representan niveles de los inputs y los outputs con los que dicha unidad alcanzaría la eficiencia. Así, los peers y los targets desempeñan en ocasiones un interesante papel en la práctica, ya que pueden indicar a las unidades ineficientes claves para mejorar su comportamiento. En particular, los targets indican direcciones de mejora para las unidades ineficientes, especificando cambios concretos en los valores de sus inputs y sus outputs que la llevarían a operar de forma eficiente.

Independientemente del tipo de medida de eficiencia utilizada, los modelos DEA proporcionan targets que se corresponden con la proyección más lejana de la unidad evaluada sobre la frontera eficiente. Sin embargo, algunos autores creen que la proyección así obtenida puede no ser un referente representativo para la unidad evaluada y proponen alternativamente minimizar la distancia a esta proyección eficiente, de modo que los targets resultantes sean lo más parecidos posible a los valores de los inputs y los outputs de la unidad evaluada. La idea que sustenta esta propuesta es que una unidad puede aprender más fácilmente de otras que sean similares a ella y, además, cuanto más cercanos estén los

targets a los valores de las variables de la unidad evaluada, menor esfuerzo deberá realizar ésta para alcanzar la eficiencia.

En la literatura podemos encontrar distintos enfoques para interpretar esta idea de “similitud” o “cercanía” y, dependiendo de dicha interpretación, se han propuesto distintos procedimientos para encontrar los correspondientes targets más cercanos. En general, los procedimientos propuestos hasta la fecha consisten en minimizar alguna distancia a la frontera eficiente, lo que no constituye una tarea sencilla, puesto que este problema es equivalente a minimizar la distancia al complemento de un poliedro. De hecho, estos procedimientos presentan algunos inconvenientes cuando se utilizan en la práctica. O bien no garantizan siempre la obtención del punto más cercano perteneciente a la frontera eficiente, o bien requieren el uso de algoritmos que identifiquen todas las facetas de la frontera eficiente DEA, lo que conlleva una elevada carga computacional.

En este trabajo proponemos una metodología general que permite encontrar los targets más cercanos a una unidad dada, una vez se ha especificado el criterio de similitud que se quiere emplear. La clave de nuestro planteamiento radica en la caracterización del conjunto de puntos eficientes que dominan a la unidad que debe ser evaluada por medio de un conjunto de restricciones lineales que, en particular, involucran variables enteras. De este modo, una vez seleccionada una medida para evaluar la proximidad entre los inputs y los outputs de la unidad evaluada y su correspondiente proyección, se trata de formular un problema de programación matemática que consiste en minimizar dicha medida, sujeta a las restricciones obtenidas en la caracterización anterior.

2. Bibliografía

- [1] Cooper, W.W., Seiford, L.M. and Tone, K. (2000). *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*. Kluwer Academic Publishers.
- [2] González, E. and Álvarez, A. (2001). From efficiency measurement to efficiency improvement: the choice of a relevant benchmark, *European Journal of Operational Research*, 133, 512-520.
- [3] Pastor, J.T., Ruiz, J.L. and Sirvent, I. (1999). An Enhanced DEA Russell Graph Efficiency Measure, *European Journal of Operational Research*, 115 (3), 596-607.
- [4] Silva, M.C.A., Castro, P. and Thanassoulis, E. (2003). Finding Closest Targets in Non-Oriented DEA Models: The Case of Convex and Non-Convex Technologies. *Journal of Productivity Analysis*, 19, 251-269.

Obtención de targets globales e individuales mediante un nuevo modelo de asignación de recursos centralizado bajo retorno de escala constante

Jesús T. Pastor¹, Juan Aparicio²

¹jtpastor@umh.es, ²j.aparicio@umh.es

Centro de Investigación Operativa, Universidad Miguel Hernández de Elche

Abstract

En un gran número de organizaciones existe la figura del gestor capacitado para la toma de decisiones centralizadas, dedicado a la supervisión de las unidades formantes de la organización. Dichas unidades a pesar de actuar de manera independiente, siempre lo hacen bajo una dirección central. En dicho contexto, proponemos una nueva versión bajo retorno de escala constante del modelo centralizado de Lozano y Villa (2004).

Palabras Clave: Data Envelopment Analysis (DEA), Planificación Centralizada, Targets

1. Introducción

Existen diversas situaciones reales en las cuales las DMUs (Decision Making Units) se encuentran bajo la supervisión de un gestor central. Esta situación es bastante común en organizaciones de gran tamaño como, por ejemplo, redes bancarias, departamentos universitarios, etc. El gestor central, interesado en la eficiencia de operación de las DMUs, se encuentra, a su vez, preocupado por la producción global de los outputs y el consumo total de inputs. En este sentido Lozano y Villa (2004) recientemente han publicado un nuevo modelo DEA de asignación de recursos centralizado (modelo LV, desde este momento). Este modelo busca la expansión radial del vector de outputs total de la organización, mientras se impone la conservación del nivel global de inputs (en el caso, por supuesto, de orientación output). Además, el modelo LV proyecta el conjunto de DMUs sobre la frontera eficiente de manera simultánea, mientras que cuando analizamos la eficiencia relativa de una DMU a través de un modelo DEA clásico, esta unidad es proyectada de forma individualizada sobre la frontera eficiente.

El modelo LV tiene la ventaja de ser más sencillo que el resto de artículos publicados en DEA dedicados a la planificación centralizada de recursos (Golany et al. (1993), Golany y Tamir (1995), Athanassopoulos (1995), Athanassopoulos (1998), Färe et al. (1997), Kumar y Sinha (1999) y Beasley (2003)). Además, el modelo LV es el único de los anteriormente citados que

garantiza que los puntos proyección se encuentren localizados sobre la frontera eficiente del conjunto de posibilidades de producción (ver la introducción del artículo de Lozano y Villa (2004)).

En el artículo original de Lozano y Villa (2004) se asumió una tecnología bajo retorno de escala variable. Sin embargo, los autores argumentaban que simples modificaciones del modelo en su formulación, eran necesarias para incorporar la suposición de retorno de escala constante. No obstante, el modelo LV no funciona bien en el caso de trabajar con una tecnología bajo retorno de escala constante, como hemos podido constatar. En consecuencia, vamos a proponer un modelo de asignación de recursos centralizado alternativo, bajo retorno de escala constante, de forma tal que nos permita obtener targets globales (para la organización) e individuales (para cada una de las DMUs) más adecuados desde el punto de vista de la gestión.

2. Bibliografía

- [1] Lozano, S. y Villa, G. (2004) "Centralized resource allocation using Data Envelopment Analysis" Journal of Productivity Analysis 22, 143–161.

On the convenience of keeping components at a level greater than one half

Francesc Carreras¹, Josep Freixas², María Albina Puente²

¹francesc.carreras@upc.edu, Departamento de Matemática Aplicada 2,
Escuela Técnica Superior de Ingenieros Industriales y Aeronáuticos de
Terrassa.

²{josep.freixas;M.Albina.Puente}@upc.edu, Departamento de Matemática
Aplicada 3, Escuela Politécnica Superior de Ingeniería de Manresa.

Universidad Politécnica de Cataluña

Abstract

En este trabajo se considera una clase especial de sistemas decisivos para los cuales estudiamos la variación de su fiabilidad con respecto a diferentes suposiciones. Se demuestra que una forma de alcanzar un alto nivel de fiabilidad consiste en mantener la fiabilidad de sus componentes a un nivel superior a 1/2.

Keywords: semicoherent structure; self-dual k -out-of- n system

AMS: 06E30, 90B25

1. Summary

The class of k -out-of- n systems was introduced by Birnbaum, Esary and Saunders [1] (see Rushdi and Al.Hindi [3] for an extensive survey). These systems consist of n components of the same kind. The entire system is working if at least any k of its n components are operating and it fails if $n - k + 1$ or more components fail. Hence, a k -out-of- n system breaks down at the time of the $(n - k + 1)$ th component failure. Important particular cases of k -out-of- n systems are parallel and series systems corresponding to $k = 1$ and $k = n$, respectively. Applications of k -out-of- n systems can be found in the areas of target detection, communication, safety monitoring systems and, particularly, in the area of voting systems.

In this work we focus on the reliability of a subclass of these systems, namely the so-called *self-dual k -out-of- n systems*, specially when all components have the same reliability p . We also study the case with multiple self-dual k -out-of- n subsystems.

We especially study the variation of the system reliability under two assumptions: (i) if we slightly increase the common components reliability p for $p > 1/2$, and (ii) if we increase the number of components of the system. We also compare the system reliability with the components reliability.

According to Carreras et al. [2] self-dual k -out-of- n systems have an index of performance equal to $1/2$, this low level could one to expect a high degree of instability for the system when components work at a level of performance far from 1 .

The main results found in our study are asymptotic and reflect that the system reliability tends to 1 whenever the number of components in the system is large and the common component reliability is greater than $1/2$. This property permits to achieve a high level of performance with the only requirement of maintaining components with a performance above $1/2$. Although the number of path and cut sets for these systems is the same it is quite relevant the high effectiveness the system achieves by only maintaining components above $1/2$.

After studying the basic model we provide further extensions of the main result when the system is a combination of self-dual k -out-of- n subsystems. Parallel self-dual k -out-of- n systems and series self-dual k -out-of- n systems are conspicuous examples. The key-role that plays the value $1/2$ for the basic model is replaced by other appropriate values.

2. Acknowledgement

Research partially supported by Grant BFM 2003-01314 of the Science and Technology Spanish Ministry and the European Regional Development Fund.

3. Bibliography

- [1] Birnbaum, Z.W., Esary, J.D. and Saunders, S. (1961). Multi-component systems and structures and their reliability. *Technometrics* 3, 55–77.
- [2] Carreras, F., Freixas, J. and Puente, M.A. [2004]: “A performance index for semicoherent structures.” *Reliability Engineering and System Safety* 83, 323–332.
- [3] Rushdi, A. and Al.Hindi, K.A. (1993). Table of the Lower Boundary of the Region of Useful Redundancy for k -out-of- n systems. *Microelectronics and Reliability* 33, 979–992.

On the election of optimal persuadable voters

Josep Freixas, Montserrat Pons

{josep.freixas;montserrat.pons}@upc.edu, Departamento de Matemática Aplicada 3, Escuela Politécnica Superior de Ingeniería de Manresa, Universidad Politécnica de Cataluña

Abstract

Un observador pronostica la intención de voto de cada miembro de un comité con respecto a una propuesta, para ello establece una ordenación de los votantes según la predisposición a votar a favor de la propuesta. Su objetivo es determinar aquellos votantes que pueden tener un papel más decisivo. Los votantes elegidos serán los mejores candidatos a ser persuadidos antes de la votación.

Keywords: circumstantial and standard power indices; Owen's multilinear extension.

AMS: 91A12, 91B12.

1. Summary

Assume that a proposal P has to be submitted to the members of a certain committee, and that according to some prediction, each voter i has an independent a priori probability p_i of voting in favor of the proposal. It does not matter if the vector of probabilities is either "qualitative" or "quantitative," although the former possibility is more realistic when one intends to do a forecast on the voting intentions. In such a case the voters are ranked relatively to each other and classified in some distinguished categories, like for example "those almost in favor of P ," "those slightly in favor of P ," "those undecided," and so on.

Suppose, moreover, that an external influence might change the perception of some voter i with respect to P , and thus his/her probability of voting for the proposal. In this context, it is of special interest to know which particular voter is more decisive in order to get P approved. The ways to measure whether one voter is more decisive than other can be different, and depend on different particular factors, but anyone of them must take into account the voting rules and the independent a priori probability p_i of each player in relation to the proposal P .

In this work we are interested in finding optimal persuadable voters for a given voting system and for any vector of probabilities representing the willingness that players have to vote in favor of a certain proposal P according to an outsider prediction. Our results extend the ones given in [1]. Three preorderings defined in the set of voters play an important role.

The results we have found explain which one of two arbitrary voters has to be persuaded, in order to get the best possible global result, whenever they are related for some of the three preorderings. The combination of the three results associated to each preordering allow the construction of a step-by-step algorithm for finding a set of persuadable voters for any voting system and any vector of predictions. It is significant to note that for using the proposed algorithm it is only required the list of minimal winning coalitions and the ranking among voters according to their willingness to vote for the submitted proposal.

2. Acknowledgement

Research partially supported by Grant BFM 2003-01314 of the Science and Technology Spanish Ministry and the European Regional Development Fund.

3. Bibliography

- [1] Freixas, J. and Pons, M. (2005). Two measures of circumstantial power: influences and bribes. *Homo Oeconomicus* 22, 1–20.

On the Shapley value of a minimum cost spanning tree problem

Gustavo Bergantiños¹, Juan Vidal-Puga²

¹gbergant@uvigo.es ²vidalpuga@uvigo.es

Departamento de Estadística e I. O. – Universidade de Vigo

Abstract

Asociamos un juego con utilidad transferible "optimista" v^+ con cada problema del árbol de mínimo coste. Definimos $v^+(S)$ como el coste de conectar los agentes de S a la fuente, suponiendo que los agentes de $N \setminus S$ ya están conectados. Probamos que el valor de Shapley de v^+ coincide con una regla ya presente en la literatura.

Keywords: minimum cost spanning tree problems, Shapley value.

1. Introduction

In this paper we study minimum cost spanning tree problems (*mcstp*). Consider that a group of agents, located at different geographical places, want some particular service which can only be provided by a common supplier, called the source. Agents will be served through connections which entail some cost. However, they do not care whether they are connected directly or indirectly to the source.

There are many economic situations that can be modeled in this way. For instance, several towns may draw power from a common power plant, and hence have to share the cost of the distribution network. Other examples include communication networks, such as telephone, Internet, or cable television.

The literature on *mcstp* starts by defining algorithms for constructing minimal cost spanning trees (*mt*). Other important issue is how to allocate the cost associated with the *mt* among agents.

Bird (1976) associated with each *mcstp* a cooperative game with transferable utility (*TU* game) v . According with Bird, $v(S)$ denotes the cost of connecting agents in S to the source, assuming that agents in $N \setminus S$ are not present. Since the Shapley value is a suitable concept for *TU* games, we can use it in *mcstp*.

An *mcstp* is *irreducible* if reducing the cost of any arc, the cost of connecting agents to the source is also reduced. Given an *mcstp* C , Bird (1976) defines the irreducible form C^* associated with the *mcstp* C . Bergantiños and Vidal-Puga (2005) define the rule φ , of the *mcstp* C , as the Shapley value of the *TU* game v associated with C^* . They prove that φ coincides with the rule *ERO* introduced in Feltkamp *et al* (1994).

We associate with each *mcstp* another *TU* game v^+ . We define $v^+(S)$ as the cost of connecting agents in S to the source, assuming that agents in $N \setminus S$ are already connected, and agents in S can connect to the source through agents in $N \setminus S$.

Both *TU* games, v and v^+ , compute the cost of connecting agents to the source. The game v takes a pessimistic point of view because it assumes that agents in $N \setminus S$ are not connected to the source. The game v^+ takes an optimistic point of view because it assumes that agents in $N \setminus S$ are already connected to the source.

In general there is no relationship between v and v^+ . Nevertheless, it is possible to find a relationship between v and v^+ in an interesting class of problems. We prove that, in irreducible problems, v and v^+ are dual games (*i.e.* $v(S) + v^+(N \setminus S)$ is constant for all $S \subset N$).

We define two rules in *mcstp* using the optimistic *TU* game v^+ . The first rule is the Shapley value of the *TU* game v^+ associated with C . The second one is the Shapley value of the *TU* game v^+ associated with C^* .

We thus have four rules in *mcstp* based on the Shapley value of an associated *TU* game. We prove that, actually, we have two rules. The Shapley value of the characteristic function v^+ associated with C coincides with the Shapley value of the characteristic function v^+ associated with C^* , and with the Shapley value of the characteristic function v associated with C^* . The classical Shapley value (of the characteristic function v associated with C) is different.

Finally, we present a new characterization of φ using a property of equal contributions.

2. Bibliography

- [1] Bergantiños G and Vidal-Puga JJ (2005) A fair rule in cost spanning tree problems. Mimeo. Available at <http://webs.uvigo.es/gbergant/papers/cstrule.pdf>
- [2] Bird CG (1976) On cost allocation for a spanning tree: A game theoretic approach. Networks 6:335-350
- [3] Feltkamp V, Tijs S and Muto S (1994) On the irreducible core and the equal remaining obligation rule of minimum cost extension problems. Mimeo. Tilburg University.

On-line search algorithms for fault diagnostics in large-scale computer communication networks

David Alcaide¹, Eugene Levner², Ilya Levin³, Sergey Frenkel⁴,
Viktor Zakharov⁴

¹dalcaide@ull.es, Statistics, Operations Research and Computing Science
Department, University of La Laguna, Tenerife, Spain

²levner@hait.ac.il, Department of Computer Science, Holon Institute of
Technology, Holon, Israeli

³iliyal@post.tau.ac.il, School of Education, Tel-Aviv University, Tel-Aviv, Israel

⁴{sfrenkel, vzkharov}@ipiean, Institute of Informatics, Russian Academy of
Sciences, Moscow Russia

Abstract

Se considera el problema de diagnosticar fallos en redes de comunicación de ordenadores con el objetivo de minimizar las pérdidas esperadas causadas por dichos fallos. Puesto que un fallo pudiera ser inadvertido en una inspección, es posible inspeccionar varias veces el mismo subsistema (módulo). Se desarrollan algoritmos “on-line” para este problema.

Keywords: Fault diagnostics, testing, on-line algorithm

1. Introduction

Fault diagnostics is a crucial aspect of the design and exploitation of modern computer communication systems. Faults, in practice, are unavoidable. So, their detection and correction, under as less as possible consumptions, are essential for the reliability of any system. The approach to fault diagnostics in the present work is based on using test transactions or *probes* [1]. A *probe* is a test transaction whose outcome depends on the status of some components of the system. For example, a probe can be a program that is executed on a particular machine, a *probe station*, by sending a transaction to a network component and measuring the response. The well-known program *ping* is a common example of a probing tool useful to detect network availability. Also, the probes can be sent in the form of test e-mail messages, web-access requests, and so on. In this context, a system is usually represented by a “*dependency graph*” where nodes can be either hardware elements or software components or services, and links represent both physical, technological, and logical connections between them. We denote such *dependency graph* as $G=(N,E)$, where the set of nodes $N=1,\dots,n$ is a given set of system components, and the set of edges $E \subseteq N \times N$ describes the connections/interactions between them. There is also a set of probes P , corresponding to a testing-for-fault search procedure (diagnosis). The directed edge $(i, j) \in E$ denotes the fact that a fault in node i may cause an error in node j . Using a probabilistic model description, such directed edge is labeled with the probability π_{ij} of that j fails when i fails. Early estimation of the characteristics to be diagnosed of the target system could be obtained by

studying fault modules by various optimal search strategies [2], i.e., studying the problem of searching fault modules when the search effort resources are limited and only probabilities of the fault location and of the error of fault detection are given. Such approach may provide useful estimations and upper bounds for exploitation cost. This approach will be further developed in the present paper.

2. A brief description of the problem and a min-loss search algorithm

In its simplest form, the fault diagnostic and localization looks in this way: a set of input stimulus (tests) must first be selected and then a diagnosis is made from the test outcomes. A common test strategy aims at minimizing the costs of the sequence of tests required to reach a diagnostic conclusion and to locate faults in a system. The main question of fault localization is its cost which is often represented in terms of length of test (probe) sequence, i.e., the minimum necessary number of probes that yield zero (as ideal) some diagnosis error measure [2]. A large scale system is modelled as a collection of known components interacting with each other. Components are the basic granularity for localizing failures. Assume that there are n independent, stochastically failing modules. Several of them can fail in the system's lifetime. When a failure occurs, a series of sequential inspections must be performed in order to identify the failed module. The goal is to minimize expected losses incurred by the failure of the system before the latter is detected.

3. Conclusion

Our model permits to design a fast priority-based algorithm which efficiently estimates the number of search steps and time spent to detect a fault module in a computer system. The model also allows both to plan a strategy of optimal search of a fault module in a computer system and predicts time spent to the diagnostic, based on the information accessible on an early stage of the target system design. In fact, this real-time algorithm enable system managers to select their own trade-off between the computational cost and test sets needed for effective fault localization.

4. Acknowledgement

This work is partially supported by Spanish Government Research Projects DPI2001-2715-C02-02 and MTM2004-07550.

5. Bibliography

- [1] Brodie, M., Rish, I. and Ma, S. (2002) Intelligent probing: A cost-effective approach to fault diagnosis in computer networks, *IBM Syst. J.*, 41, 3.
- [2] Levner, E. (1994) Infinite-horizon scheduling algorithms for optimal search for hidden objects, *Int. Trans. in Oper. Res.*, vol. 1, n. 2, 241-250.

Operadores de recombinación para Algoritmos Evolutivos Multiobjetivo basados en ranking Pareto

P. Mateo Collazos¹, I. Alberto Moralejo²

¹mateo@unizar.es, ²isolina@unizar.es

Departamento de Métodos Estadísticos, Universidad de Zaragoza

Abstract

Una característica deseable de los operadores de los AEs es su capacidad de adaptarse a la evolución de la población. Se presenta un conjunto de operadores de recombinación para AEMOs que poseen un comportamiento dinámico y utilizan la definición de ranking Pareto. Los operadores se comparan sobre un conjunto de problemas test habitual en la literatura.

Palabras Clave: Optimización Multiobjetivo, Algoritmos Evolutivos

AMS: 90C29, 90C59

Optimal control of fed-batch processes with particle swarm optimization

A. Ismael F. Vaz¹, Eugénio C. Ferreira²

¹aivaz@dps.uminho.pt, Departamento de Produção e Sistemas, Escola de Engenharia, Universidade do Minho, Campus de Gualtar, 4710-057 Braga-P

²ecferreira@deb.uminho.pt, Centro de Engenharia Biológica, Universidade do Minho, Campus de Gualtar, 4710-057 Braga-P

Abstract

Los problemas de control óptimo son normalmente descritos a través de sets de ecuaciones algebraicas y diferenciales no lineales. Son revisados algunos problemas de control óptimo en el área de bioprocесos. Utilizamos un optimizador de soluciones de nube de partículas para resolver una aproximación no lineal para el problema de control óptimo.

Keywords: Optimal control, particle swarm optimization

1. Introduction

A great number of valuable products are produced using fermentation processes and thus optimizing such processes is of great economic importance. Fermentation modeling process involves, in general, highly nonlinear and complex differential equations. Often optimizing these processes results in control optimization problems for which an analytical solution is not possible. By reformulating the optimal control problem as a non-linear, non-differentiable, optimization problem we are able to solve it with a derivative free optimization technique. We also propose a numerical environment to address the optimization of a fermentation process using available software for non-linear optimization and a development of an external library to handle the dynamic equations resulting from the complex differentiable equations.

Particle swarm [4, 2] was the technique used to solve the resulting non-linear approximation problems and proves to be competitive in finding a problem solution.

2. The optimization problem

The optimal control problem (P) can be posed as

$$\max J(t_f) \quad (4.69)$$

$$s.t. \quad \dot{x} = f(x, u, t) \quad (4.70)$$

$$\underline{x} \leq x(t) \leq \bar{x}, \forall t \in [t_0, t_f] \quad (4.71)$$

$$\underline{u} \leq u(t) \leq \bar{u}, \forall t \in [t_0, t_f] \quad (4.72)$$

where (4.70) is a set of differential equations, x are the state variables, u are the control variables and $J(t_f) = \varphi(x(t_f), t_f) + \int_{t_0}^{t_f} \phi(x, u, t) dt$.

φ is the performance index of the state variables at final time t_f and ϕ is the integrated performance index during the operation.

Imposing an infinity penalty for the (4.71) constraints violation results in replacing the objective function $J(t_f)$ by

$$\hat{J}(t_f) = \begin{cases} J(t_f) & \text{if } x \leq x(t) \leq \bar{x}, \forall t \in [t_0, t_f] \\ +\infty & \text{otherwise.} \end{cases}$$

We address the functional constraints (4.72) by approximating functions $u(t)$ with linear splines. Let w_i , $i = 0, \dots, n$ be the linear spline values at n equally spaced point in $[t_0, t_f]$. Constraint (4.72) can then be approximated by $\underline{u} \leq w_i \leq \bar{u}$.

The control optimization problems were coded in the AMPL [3] modeling language and an external dynamic library was made to compute the objective function values. The CVODE [1] library was used to solve the corresponding differentiable equations.

3. Acknowledgement

This work was partially supported by the Algoritmi research center and by FCT grant POCI/MAT/58957/2004.

4. Bibliography

- [1] S.D. Cohen and A.C. Hindmarsh. CVODE, a Stiff/Nonstiff ODE solver in C. *Computers in Physics*, 10(2):138–143, 1996.
- [2] R. Eberhart and J. Kennedy. New optimizers using particle swarm theory. In *Proceedings of the 1995 6th International Symposium on Micro Machine and Human Science*, pages 39–43.
- [3] R. Fourer, D.M. Gay, and B.W. Kernighan. A modeling language for mathematical programming. *Management Science*, 36(5):519–554, 1990.
- [4] J. Kennedy and R. Eberhart. Particle swarm optimization. In *Proceedings of the 1995 IEEE International Conference on Neural Networks*, pages 1942–1948, Perth, Australia. IEEE Service Center, Piscataway, NJ. <http://engr.iupui.edu/~shi/Coference/psopap4.html>.

Optimistic weighted Shapley rules in minimum cost spanning tree problems

Gustavo Bergantiños¹, Silvia Lorenzo-Freire²

¹gbergant@uvigo.es, Dpto. Estadística e I.O., Universidad de Vigo

²silvia_lorenzo@uvigo.es, Dpto. Estadística e I.O., Universidad de Vigo

Abstract

We introduce optimistic weighted Shapley rules in minimum cost spanning tree problems. We define them as the weighted Shapley values of the optimistic game v^+ introduced in Bergantiños and Vidal-Puga (2005b). We prove that they are obligation rules (Tijs, Branzei, Moretti, and Norde (2005)). Moreover, we present an axiomatic characterization using monotonicity properties.

Keywords: minimum cost spanning tree problems. Weighted Shapley values.

1. Introduction

A group of agents want some particular service which can only be provided by a common supplier, called the source. Agents will be served through connections which entail some cost. They do not care whether they are connected directly or indirectly to the source. This kind of situations are studied in minimum cost spanning tree problems, briefly *mcstp*.

Many real situations can be modeled in this way. For instance, Bergantiños and Lorenzo (2004) studied a real situation where villagers had to pay the cost of constructing pipes from their respective houses to a water supplier. Other examples are communication networks, such as telephone, Internet, wireless telecommunication, or cable television.

An *mcstp* is characterized by a set $N \cup \{0\}$ and a matrix C . N is the set of agents, 0 is the source, and for each $i, j \in N \cup \{0\}$, c_{ij} denotes the cost of connecting i and j .

A relevant issue of this literature is to define algorithms for constructing minimal cost spanning trees, briefly *mt*. Kruskal (1956) and Prim (1957) provide two algorithms for finding *mt*. Another important issue is how to allocate the cost associated with the *mt* among agents. Bird (1976), Kar (2002), and Dutta and Kar (2004) introduce several rules.

Bird (1976) associated with each *mcstp* a cooperative game with transferable utility, briefly *TU* game, v_C . According with Bird, $v_C(S)$ denotes the cost of connecting agents in S to the source, assuming that agents in $N \setminus S$ are not present. Kar (2002) studied the Shapley value of v_C whereas Granot and Huberman (1981, 1984) studied the core and the nucleolus.

An *mcstp* is *irreducible* if reducing the cost of any arc, the cost of connecting agents to the source is also reduced. Given an *mcstp*, Bird (1976) defined the irreducible form C^* associated with the matrix C . Bergantiños and Vidal-Puga (2005a) defined the rule φ in C as the Shapley value of v_{C^*} .

All the rules mentioned before allocate the cost among the agents taking into account only C . In some situations it could make sense to use another relevant information. For instance, in the case of Bergantiños and Lorenzo (2004) we can take into account also the income of each villager, which can be modelled by a vector $w = (w_i)_{i \in N}$ where $w_i > 0$ is the income of villager i .

The main objective of this paper is to find nice rules allocating the cost among agents using both, the cost matrix C and the weight system w . We will do it by “generalizing” the rule φ .

Bergantiños and Vidal-Puga (2005b) associated another *TU* game v_C^+ with each *mcstp*. Here, $v_C^+(S)$ is the minimal cost of connecting all the agents of S to the source assuming that agents in $N \setminus S$ are already connected, and agents in S can connect to the source through agents in $N \setminus S$. Thus, v_C is called the pessimistic game whereas v_C^+ is called the optimistic game.

Bergantiños and Vidal-Puga (2005b) proved that, for each *mcstp* C , the Shapley value of v_{C^*} , v_C^+ , and $v_{C^*}^+$ coincide whereas the Shapley value of v_C is different.

One possible way of defining rules using C and w is through the game v_C^+ . Thus, we define the *optimistic weighted Shapley rule* with respect to w as the weighted Shapley value of v_C^+ .

Tijs *et al* (2005) introduced the class of *obligation rules*, a family of rules which satisfy two appealing properties: population monotonicity (if a new agent joint the society nobody is worse off) and strong cost monotonicity (if the connection cost between any pair of agents increases, nobody is better off). We prove that optimistic weighted Shapley rules are obligation rules. This is a quite surprising result because they are defined in a completely different way.

Bergantiños and Vidal-Puga (2005a) proved that φ is the only rule satisfying population monotonicity, strong cost monotonicity, and equal share of extra cost. We prove that the optimistic weighted Shapley rule with respect to w is the only rule which satisfies population monotonicity, strong cost monotonicity, and weighted share of extra cost with respect to w .

Optimización conjunta del mantenimiento y los recursos considerando criterios RAMS+C

A. Sánchez¹, S. Martorell², J. Sanz³, G. Clemente⁴

^{1,3,4}aisanche@eio.upv.es, jsanz@eio.upv.es, gclemente@eio.upv.es

Departamento de Estadística, Investigación Operativa Aplicadas y Calidad,
Universidad Politécnica de Valencia

²smartore@iqn.upv.es, Departamento de Ingeniería Química y Nuclear,
Universidad Politécnica de Valencia

Abstract

La productividad es un factor clave en la competitividad de las empresas la cual depende de la fiabilidad, mantenibilidad y disponibilidad (RAM – Reliability, Availability, Maintainability) de los equipos. En este contexto, el objetivo de la presente ponencia se centra en la optimización conjunta de la frecuencia de mantenimiento y de los recursos, tanto materiales como humanos, considerando criterios RAM+C y utilizando como técnica de optimización un algoritmo genético multiobjetivo.

Palabras Clave: RAM, multiobjetivo, repuestos

1. Introducción

La necesidad de ser más competitivos ha obligado a que las empresas consideren la función de mantenimiento como un factor clave en el proceso de mejora de la eficacia técnica y de la eficiencia económica de sus procesos productivos. Este aspecto es particularmente crítico para aquellas empresas con un elevado grado de mecanización en sus procesos.

El concepto de la función de mantenimiento ha evolucionado de forma importante desde el mantenimiento correctivo (MC), típico de la década de los 40, el mantenimiento preventivo por tiempo (MPT) en los 60 hasta la introducción, de técnicas de inspección en servicio (IS) y mantenimiento predictivo (MD) en la década de los 80. Asimismo, el papel asignado a la función mantenimiento también ha cambiado notablemente, pasando de ser considerado como un mal necesario de la producción para transformarse en un factor clave para la competitividad donde la Gestión de atributos de fiabilidad, mantenibilidad y disponibilidad (RAM – Reliability, Availability, Maintainability) de los equipos es fundamental para alcanzar como metas una elevada fiabilidad, disponibilidad y seguridad del proceso que conlleve una mayor productividad y mayor calidad del producto, y que sea respetuosa con el medioambiente, además de ser efectiva al mínimo coste posible.

En la actualidad existe una gran variedad de estrategias disponibles para desarrollar la función de mantenimiento, MC complementado con MPT, MP-C

asociado con técnicas de MD e IS. No obstante, en general, no se observa que para ello se utilice una metodología sistemática capaz de asignar la proporción y frecuencia óptima de MC-MP-MD-IS necesaria en cada caso. Al contrario, se promueve de forma indiscriminada, y hasta cierto punto sin fundamento, la ventaja de utilizar dicha fórmula del MP-C asociado con técnicas de MD e IS, que puede llegar a conclusiones generalizadas que son erróneas,

2. Objetivos

El objetivo de la presente ponencia se centra en la optimización integral del mantenimiento. Así, el proceso de optimización propuesto integra los dos elementos básicos en el proceso de toma de decisiones: a) las estrategias de mantenimiento y b) los recursos (humanos y/o materiales) los cuales son habitualmente tratados de forma separada. Además los criterios de decisión se establecen sobre los atributos RAM+C de los equipos bajo consideración.

Este proceso de toma de decisiones se puede desarrollar a partir del planteamiento de un problema de optimización multi-objetivo donde las estrategias de mantenimiento-recursos actuarán como variables de decisión y donde los atributos RAM actuarán como criterios de decisión. Para ello resulta necesario establecer un modelo paramétrico que permita formular la dependencia de los indicadores RAM respecto de los parámetros que caracterizan al plan de mantenimiento existente (por ejemplo, tipos de estrategias de mantenimiento, periodicidad, duraciones, recursos humanos, costes, repuestos, etc.). Estos modelos son utilizados en el proceso de optimización multi-objetivo para simular la influencia que un cambio en dichos parámetros puede tener sobre los criterios de decisión.

La optimización multi-objetivo se realiza mediante un algoritmo genético multiobjetivo. Finalmente, se presenta un caso de aplicación a un equipo de seguridad de una planta nuclear.

3. Agradecimientos

El presente trabajo de investigación se encuentra financiado por la Universidad Politécnica de Valencia.

Optimización y estimación robusta de carteras financieras

Francisco J. Nogales¹, Victor deMiguel²

¹FcoJavier.Nogales@uc3m.es, Departamento de Estadística, Universidad Carlos III de Madrid

²avmiguel@london.edu, London Business School, UK

Abstract

En este trabajo se pretende avanzar en el estudio del problema de selección de carteras financieras, esto es, dado un conjunto de activos/posiciones donde invertir y un presupuesto limitado se debe decidir qué proporción del presupuesto se invertirá en cada uno de los activos, de acuerdo a algún criterio de optimalidad.

Palabras Clave: optimización de carteras, estimación robusta, riesgo.

1. Introducción

En este trabajo se pretende abordar problemas de toma de decisiones financieras en presencia de incertidumbre (beneficios, volatilidades, etc.). En particular, esta propuesta se va a centrar en el estudio del problema de selección de carteras financieras, esto es, dado un conjunto de activos/posiciones donde invertir y un presupuesto limitado se debe decidir qué proporción del presupuesto se invertirá en cada uno de los activos, de acuerdo a algún criterio de optimalidad.

Este problema fue el que analizó y estudió H. Markowitz en la década de los 50 y por el que recibió el Premio Nobel de Economía en 1990. Markowitz mostró que un inversor, preocupado solo por la media y la varianza de las rentabilidades de una cartera estática, debe escoger solo aquellas carteras que se encuentren en la denominada *frontera eficiente*. Esta frontera está formada por aquellas carteras que optimizan el balance entre la media y la varianza de las rentabilidades de la cartera. La implementación práctica de estas políticas requiere la estimación del vector de medias y de la matriz de varianzas del conjunto de rentabilidades.

Tradicionalmente, para estimar estos parámetros se han utilizado estimadores de máxima verosimilitud (MLE). Pero las carteras construidas con estos estimadores presentan pesos extremos que fluctúan substancialmente con el tiempo. Como consecuencia, la implementación de estas políticas de inversión requiere un alto volumen de transacciones y da lugar a un comportamiento malo en muestras correspondientes a periodos de validación (aquellas que no son usadas en el periodo de estimación).

La razón de la inestabilidad de las estimaciones basadas en MLE se debe a que la distribución empírica de las rentabilidades se aleja frecuentemente de las

hipótesis habituales (datos independientes e idénticamente distribuidos según una normal) requeridas en la estimación basada en MLE. Esto es debido a la presencia de volatilidad estocástica, colas pesadas, atípicos, etc. Por estos motivos, los estimadores basados en MLE, aún siendo muy usados en la práctica, no son los más adecuados en el problema de selección de carteras financieras.

Para solventar las dificultades asociadas a la inestabilidad que presentan los estimadores basados en MLE para optimización de carteras, en este trabajo planteamos el uso de estimadores dentro del campo de la *Estadística Robusta*. Esta rama de la Estadística se centra en el desarrollo y análisis de estimadores que son estables (robustos) frente a cambios en las distribuciones supuestas.

En particular, se propone mostrar cómo se puede calcular una política de gestión de carteras que minimice un estimador *robusto* del riesgo. Nos centraremos en carteras de mínimo riesgo ya que tradicionalmente presentan un mejor comportamiento *out-of-sample*, esto es, en muestras de validación, en comparación con aquellas políticas que optimizan un balance entre riesgo y rentabilidad. Esto es debido a que, en la práctica, los errores asociados a la estimación de las rentabilidades medias de los activos son muy superiores a los errores asociados a la estimación de la correspondiente matriz de varianzas. Por tanto, la potencial ventaja de utilizar carteras que balanceen rentabilidad y beneficio se ve descompensada por el impacto negativo del error de estimación.

Además, se pretende caracterizar las propiedades analíticas de las correspondientes políticas robustas de mínimo riesgo. En concreto, y como ocurre con los estimadores robustos habituales en estadística univariante, se presenta una cota analítica a la sensibilidad de los pesos de la cartera frente a cambios en la distribución de las rentabilidades de los activos considerados. Esta cota demuestra la estabilidad de las políticas propuestas frente a cambios en la distribución de las rentabilidades y frente a presencia de valores atípicos. A su vez, esta cota nos permitirá encontrar un límite asintótico al volumen de transacción asociado a las políticas propuestas. Este es un aspecto importante en finanzas debido a la presencia de costes ocasionados por el *rebalanceo* de la cartera.

Finalmente, se realizará una comparación exhaustiva entre el comportamiento *out-of-sample* de las políticas robustas propuestas y las políticas clásicas de mínima varianza y de otras usadas en la práctica.

Partnership formation and binomial semivalues

Francesc Carreras¹, Maria Dolors Llongueras², Maria Albina Puente³

¹francesc.carreras@upc.edu, Department of Applied Mathematics II and Industrial Engineering School of Terrassa, Technical University of Catalonia

²m.dolors.llongueras@upc.edu, Department of Applied Mathematics II and Industrial Engineering School of Terrassa, Technical University of Catalonia

³m.albina.puente@upc.edu, Department of Applied Mathematics III and School of Engineering of Manresa, Technical University of Catalonia

Abstract

Se estudia la formación de consorcios en juegos cooperativos y se aplican los semivalores binomiales para medir los efectos de dicha formación.

Keywords: cooperative game, partnership, binomial semivalue.

AMS: 91A12.

1. Introduction

The notion of *partnership* was introduced by Kalai and Samet [2]. Carreras [1] emphasized the significance and scope of this concept, first in cooperative games and later on for simple games. The aim of this work is the application of *binomial semivalues*, introduced by Puente in [4], to study the effects of the partnership formation. Our goal is to investigate how these values are modified if several players agree to form a partnership, thus complementing previous results on the Banzhaf value found by Llongueras and Magaña [3]. The fact that a parameter p ranging the interval $[0, 1]$ defines the family of binomial semivalues introduces new, interesting features in the evaluation of games, partnership formation and coalition formation.

2. Partnership and binomial semivalues

Definition 2.1 (Puente [4]) Let $p \in [0, 1]$ and take $p_k = p^k(1-p)^{n-k-1}$ for $k = 0, \dots, n-1$, with $0^0 = 1$ in the extreme cases ($p = 0$ and $p = 1$). Then $\{p_k\}_{k=0}^{n-1}$ is a weighting vector and defines a semivalue, which will be denoted as ψ^p and called the p -binomial semivalue. Then the p -binomial semivalue of player i in game v is $\psi_i^p[v] = \sum_{T \subseteq N: i \in T} p^{t-1}(1-p)^{n-t}[v(T) - v(T \setminus \{i\})]$, where $t = |T|$. Of course, $\psi^{1/2} = \beta$, the Banzhaf value.

Definition 2.2 (Kalai and Samet [2]) A coalition $S \subseteq N$ is a *partnership* in a game v if $v(R \cup T) = v(R)$ for all $T \subset S$ and all $R \subseteq N \setminus S$.

The formation of a partnership suggests for the involved players a way to obtain strategic advantages. Often, it may be viewed as a previous step to coalitional bargaining. The case where only one coalition S (with $|S| \geq 2$) turns into a partnership will be the only considered throughout this work. In this case, the partnership formation is stated as follows.

Definition 2.3 (Carreras [1]) Let v a game in N and $\emptyset \neq S \subseteq N$. The *partnership game* v^S in N (where S clearly becomes a partnership) is defined by

$$v^S(T) = \begin{cases} v(T) & \text{if } S \subseteq T, \\ v(T \setminus S) & \text{if } S \not\subseteq T. \end{cases}$$

The binomial semivalues may be used to measure the internal and external effects of the partnership formation. Jointly with the binomial semivalue of each player, it will be useful to consider the (*additive*) *p-binomial semivalue* of a nonempty coalition $S \subseteq N$ in a game v , defined in a natural way by $\psi_S^p[v] = \sum_{i \in S} \psi_i^p[v]$, and also the *partnership p-binomial semivalue of S in v*, defined as $\psi_S^p[v^S]$.

General expressions will be given for the variation of the binomial semivalues (particularly for the Banzhaf value), when a partnership is formed, respectively concerning (a) internal players, (b) external players, and (c) the partnership as a whole; next, we will determine the cases where the formation of a partnership is more interesting than the constitution of a coalition in the very sense of the term (alliance) and will specialize to simple games.

3. Acknowledgement

Research partially supported by Grant BFM 2003–01314 of the Science and Technology Spanish Ministry and the European Regional Development Fund.

4. Bibliography

- [1] Carreras, F. [1996]: On the existence and formation of partnerships in a game. *Games and Economic Behavior* 12, 54–67.
- [2] Kalai, E. and Samet, D. [1987]: On weighted Shapley values. *International Journal of Game Theory* 16, 205–222.
- [3] Llongueras, M.D. and Magaña, A. [2005]: Alliances, partnerships and the Banzhaf semivalue. Forthcoming in *Annals of Operations Research*.
- [4] Puente, M.A. [2000]: Contributions to the representability of simple games and to the calculus of solutions for this class of games. Ph.D. Thesis. Technical University of Catalonia, Spain.

Planificación de la producción en una acería. Análisis de un caso real

Fermín Mallor Giménez¹, Pedro M. Mateo Collazos²

¹mallor@unavarra.es, Depto. de Estadística, Investigación Operativa,
Universidad Pública de Navarra

²mateo@unizar.es, Depto. de Métodos Estadísticos, Universidad de Zaragoza

Abstract

En este trabajo presentamos un problema real de planificación de la producción en una acería, analizando una de sus líneas de producción, la denominada línea de silicatos. El modelo incorpora, entre otros, distintos problemas auxiliares de optimización multiobjetivo. El modelo planteado proporciona soluciones útiles para la toma de decisiones a corto y medio plazo.

Palabras Clave: Planificación. Programación multiobjetivo

1. Introducción

En este trabajo presentamos un problema real de planificación de la producción en una acería dedicada a la fundición de aceros inoxidables aleaciones de níquel, aceros resistentes al calor, aceros al carbono, etc. El catálogo de productos distintos es definido por las necesidades de los clientes y asciende a varios cientos. Estos productos no sólo se diferencian por su forma y tamaño sino también por la distinta calidad del acero y de su acabado, ya que algunos requieren tratamiento térmico, rayos X, mecanizado, etc.

Aunque el proyecto de colaboración en el que se enmarca este trabajo tiene por objetivo final la planificación global de toda la planta, en este trabajo nos centramos en la planificación de la cabecera de una de las cuatro grandes familias de productos: los silicatos. Esta parte incluye un horno de fundición de 4000kg. de capacidad, un equipo de trabajo para la preparación manual de moldes, otro para el moldeo automático y finalmente uno para el desmoldeo. Cada uno de estos elementos es una fuente de restricciones. Mencionamos algunas de las más importantes: el horno, además de las limitaciones impuestas por su capacidad, debe considerar una secuencia factible en las calidades de acero que funde, ya que hay calidades de acero que no pueden ser coladas después de determinadas calidades, exigiendo otra colada intermedia, llamada de limpieza, de calidad distinta. También existe un requerimiento de homogeneidad en relación al grosor de las piezas coladas. Tanto el moldeo manual como automático tienen una capacidad limitada de producción de moldes además de un número limitado de las cajas que se utilizan para la construcción de los moldes. Asimismo, existe un espacio limitado para realizar ambos tipos de moldeo.

El principal objetivo declarado por los gestores es que el horno trabaje de forma ininterrumpida y a máxima capacidad. Otros objetivos son el cumplimiento de las fechas de entrega pactadas con los clientes, la minimización de coladas de limpieza y la generación uniforme de producto que precisa de estaciones de trabajo posteriores que constituyen cuellos de botella. El algoritmo proporciona planificaciones a corto plazo seleccionando la secuencia de trabajos que deben ser procesados durante la semana, o semanas siguientes en función de la cartera actual de pedidos. También es capaz de generar planificaciones a medio plazo, definido como dos o tres meses, permitiendo de tal modo pactar fechas de entrega de productos con clientes que realizan nuevos pedidos. En este último caso se debe considerar la reserva de capacidad de producción para el reproceso de piezas defectuosas, por no cumplir unos requerimientos de calidad, detectadas en sucesivos controles de calidad.

El algoritmo diseñado utiliza distintas herramientas de la investigación operativa como son las funciones de utilidad, la programación lineal multiobjetivo entera y la programación de eventos discretos. En su diseño se consideró no solamente su capacidad para generar *buenas (óptimas)* planificaciones sino también su flexibilidad para poder introducir las numerosas modificaciones en la definición del problema que eran y son (y serán) introducidas por los gestores de la planta industrial.

Planificación de la producción ganadera mediante programación lineal

Sara Verónica Rodríguez Sánchez¹, Luis M. Plà Aragones²

¹sara090444@hotmail.com, ²limpla@matematica.udl.es
Departamento de Matemática, Universidad de Lleida

Abstract

Varias metodologías se han propuesto para la planificación de la producción de explotaciones ganaderas intensivas. En general todas ellas toman como referencia modelos estáticos rudimentarios. En este trabajo se propone un modelo de programación lineal para apoyar la toma de decisiones estratégicas relacionadas con la planificación de la producción ganadera.

Palabras Clave: Producción, ganadería, programación lineal

AMS: 90B90, 90C90

1. Introducción

La planificación de la producción de una granja dependerá directamente de los recursos económicos (ganado, instalaciones, maquinaria, mano de obra) y financieros con los que se cuente, así como de la meta productiva que se proyecte lograr en función de la demanda observada o de su pronóstico. Estableciendo para ello estrategias de manejo reproductivo. Los métodos clásicos de la planificación de la producción son sencillos pero se ha demostrado que, por ejemplo, no representan de forma precisa las necesidades de espacio para las diferentes etapas productivas en una explotación de ganadería intensiva (Plà et al., 2004).

La planificación de la producción debe considerar las diferentes variaciones naturales que se puedan presentar en el comportamiento reproductivo de los animales como incidencia de abortos, descenso de fertilidad, aumento de la mortalidad, etc. En este trabajo se propone un modelo de programación lineal que permite planificar la producción de una granja, permitiendo un diseño óptimo de sus instalaciones a fin de satisfacer las necesidades de espacio requeridas en cada etapa del proceso productivo en función de los parámetros considerados.

2. Formulación del modelo

El modelo que se plantea está basado en un modelo de decisión markoviano que representa la dinámica de explotaciones de reproductoras (Plà et al. 1998). El modelo asume que todos los animales tienen el mismo potencial productivo.

Un parámetro básico para el modelo consiste en la utilidad media para cada estado τ_i , siendo $i \in S$ el conjunto de estados dónde los principales por ciclo son: gestación, lactación, en espera de cubrición y en espera de reemplazamiento.

Bajo estas hipótesis y asumiendo la homogeneidad en el tiempo de los parámetros y para una política reproductiva determinada, el rebaño de animales reproductores de la explotación se representa por una cadena finita de Markov aperiódica e irreducible. La estructura del rebaño en el equilibrio, $\Pi = (\pi_1, \pi_2, \dots, \pi_{|S|})$, se puede obtener resolviendo el siguiente sistema de ecuaciones lineales:

$$\begin{aligned} \pi_j &= \sum_{k \in S} \pi_k p_{kj} & j \in S \\ \sum_{j \in S} \pi_j &= 1 \end{aligned} \quad (4.73)$$

donde p_{kj} representa las probabilidades de transición para pasar del estado k al j y $\{\pi_j, j \in S\}$ representa la distribución en el equilibrio. De (1) y con la utilidad media en cada estado τ_i , la utilidad media de la granja para la política preseleccionada se puede calcular:

$$B = \sum_{j \in S} \pi_j \tau_j \quad (4.74)$$

Utilizando (2) se puede planificar la producción de una granja y estudiar el impacto de diferentes alternativas sobre el producto final.

3. Agradecimientos

Este trabajo ha sido financiado con el proyecto de investigación del Ministerio de Educación y Ciencia MTM2005-09362-C03-02.

4. Bibliografía

- [1] Plá, L.M. ; Conde J. and Pomar, J., 1998. Sow model for decision aid at farm level. In: Giron, F.J. (Eds.), Applied Decision Analysis. Kluwer Academic Publishers, Boston, 47-62.
- [2] Plá, L.M.; Babot D. and Pomar, J., 2004. A mathematical model for designing and sizing sow farms. International Transactions in Operational Research 11: 485-494

Política óptima de un modelo EOQ con demanda parcialmente acumulable

L.A. San-José¹, J. Sicilia², J. García-Laguna³

¹Departamento de Matemática Aplicada, Universidad de Valladolid

²Departamento de Estadística, I.O. y Computación, Univ. de La Laguna

³Departamento de Estadística e Investigación Operativa, Univ. de Valladolid

¹augusto@mat.uva.es, ²jsicilia@ull.es, ³laguna@eio.uva.es

Abstract

Se estudia un sistema de inventario con retropedidos parciales en el que la demanda durante el periodo de escasez es parcialmente retropedida. La fracción de demanda retropedida es una mixtura de dos funciones exponenciales. Este modelo generaliza varios sistemas de inventario estudiados por otros autores. Se determina la política óptima para todos los posibles valores de los parámetros.

Claves: inventario, EOQ, retropedidos parciales.

AMS: 90B05.

1. Introducción

En los últimos años, los modelos de inventario en los que la demanda durante el periodo de escasez es parcialmente retropedida constituyen un foco de considerable atención para diversos investigadores.

Algunos autores modelan los retropedidos parciales utilizando el concepto de impaciencia. Abad [1] fue el primero en introducir la suposición de que la fracción de demanda retropedida durante el periodo de escasez es una función dependiente del tiempo de espera y propuso dos funciones para modelar la impaciencia de los clientes. Una de las funciones propuestas es la exponencial negativa, que posteriormente es utilizada en diversos trabajos (Papachristos y Skouri [6], Skouri y Papachristos [8], Abad [2], San José et al. [8]).

Sin embargo, Padmanabhan y Vrat [4] consideran que la fracción de demanda retropedida depende del número de pedidos pendientes de satisfacer en ese instante, es decir, del nivel neto de inventario. Esta situación también ha sido considerada en los trabajos de Padmanabhan y Vrat [5] y Chu et al. [3]. Esta hipótesis sobre la tasa de demanda retropedida es equivalente a suponer que la fracción de retropedidos parciales es una función exponencial negativa del tiempo transcurrido desde que se produjo la rotura.

En este trabajo se considera que el comportamiento de los clientes frente a la escasez depende tanto de la impaciencia (es decir, del tiempo que los clientes deben o deberían esperar hasta la llegada del siguiente pedido) como

del inventario acumulado (es decir, del tiempo que ha transcurrido desde que se produjo la rotura). Más concretamente, suponemos que la fracción de demanda que se satisface es una mixtura de dos funciones exponenciales negativas, una dependiente del tiempo de espera y la otra dependiente del tiempo transcurrido desde la rotura. Por tanto, el comportamiento de los clientes se representa de una forma más general que la contemplada en los enfoques anteriores de Abad y Padmanabhan y Vrat. Además, se considera que el coste de escasez tiene dos componentes significativos: el coste unitario de cada retropedido (que incluye un coste fijo y un coste proporcional al tiempo que se espera) y el coste unitario de pérdida de ventas (que incluye un coste constante por pérdida de confianza).

2. Bibliografía

- [1] Abad P.L. (1996). Optimal Pricing and Lot-sizing Under Conditions of Perishability and Partial Backordering. *Management Science* 42, 1093-1104.
- [2] Abad P.L. (2003). Optimal pricing and lot-sizing under conditions of perishability, finite production and partial backordering and lost sale. *European Journal of Operational Research* 144, 676-684.
- [3] Chu P., Yang K.L., Liang S.K. and Niu T. (2004). Note on inventory model with a mixture of back orders and lost sales. *European Journal of Operational Research* 159, 470-475.
- [4] Padmanabhan G. y Vrat P. (1990). Inventory model with a mixture of back orders and lost sales. *International Journal of Systems Sciences* 21, 1721-1726.
- [5] Padmanabhan G. y Vrat P. (1995). EOQ models for perishable items under stock dependent selling rate. *European Journal of Operational Research* 86, 281-292.
- [6] Papachristos S. and Skouri K. (2000). An optimal replenishment policy for deteriorating items with time-varying demand and partial -exponential type- backlogging. *Operations Research Letters* 27, 175-184.
- [7] San José L.A., Sicilia J. y García-Laguna J. (2006). Analysis of an inventory system with exponential partial backordering. *Int. J. Production Economics*, 100, 76-86.
- [8] Skouri K. and Papachristos S. (2002). A continuous review inventory model, with deteriorating items, time-varying demand, linear replenishment cost, partially time-varying backlogging. *Applied Mathematical Modelling* 26, 603-617.

Procedimiento constructivo por combinación de reglas para secuenciar modelos mixtos de vehículos en líneas de montaje

Joaquín Bautista Valhondo¹, Jaime Cano Belmán²

¹joaquin.bautista@upc.edu, Nissan Chair ETSEIB, Universitat Politècnica de Catalunya

²jaime.cano-belman@upc.edu, Departament d'Estadística e Investigació Operativ, Universitat Politècnica de Catalunya

Abstract

Se estudia el problema de secuenciación de modelos mixtos de vehículos en líneas de montaje minimizando la sobrecarga (MSP-W, Mixed-model Sequencing Problem, Work overload). Para resolver el problema se usa un procedimiento que combina soluciones representadas en el espacio de reglas heurísticas. Se realiza una experiencia computacional con el propósito de comprobar la eficiencia del procedimiento.

Palabras Clave: Secuenciación, exploración de entornos, heurísticas

1. Introducción

Yano y Rachamadugu [3] proponen una de las variantes para resolver el problema de secuenciación de modelos mixtos de vehículos en líneas de montaje. Los autores usan como medida de eficiencia la sobrecarga total que puede ocurrir en las estaciones de la línea. La sobrecarga es el trabajo que no puede ser completado dentro de una estación cerrada debido a que el tiempo de proceso de la unidad en curso es más grande que el tiempo del que dispone el operario para completarla.

Se han usado procedimientos heurísticos para abordar el problema, los cuales se basan en el uso de diferentes reglas de prioridad que condicionan la asignación de productos en cada posición de la secuencia; la aplicación de éstos ofrece resultados aceptables, sobre todo cuando consideran más de una regla de prioridad o cuando la regla combina diferentes aspectos del problema.

Otra clase de heurísticas empleadas en los últimos años son los métodos de búsqueda local. Se han aplicado, entre otros, Algoritmos Genéticos (GA, Genetic Algorithms), Recocido Simulado (SA, Simulated Annealing) o Búsqueda Tabú (TS, Tabu Search). Esta clase de heurísticas brinda alternativas para buscar soluciones en un espacio delimitado por la definición de un vecindario. Sin embargo, la definición general de vecindarios no guarda información sobre el problema como suele ocurrir con las heurísticas greedy.

En este trabajo proponemos aprovechar las características positivas de ambos tipos de heurísticas: las reglas de prioridad y la generación de nuevas soluciones

en el espacio de búsqueda. Para ello, usamos las bases del método llamado Búsqueda Dispersa (SS, Scatter Search). En la experiencia computacional se usan instancias generadas por [2].

2. Reglas de prioridad

Las reglas de prioridad se usan para establecer una lista ordenada de vehículos candidatos a ser secuenciados en cada etapa de decisión. En este trabajo se usan 20 reglas de prioridad. Las reglas consideran aspectos como tiempos de procesamiento, la demanda pendiente, la diferencia absoluta entre el tiempo de proceso de cada producto en cada estación y el tiempo de ciclo, los desplazamientos de los trabajadores, la estación cuello de botella, la sobrecarga, el tiempo improductivo, la regularidad en la producción y en la sobrecarga.

3. Búsqueda dispersa

Aunque nos basamos en SS, el procedimiento implementado difiere del original en algunos aspectos. Se prueban tres procedimientos de regeneración del *RefSet*. El tamaño del *RefSet* depende del número de reglas consideradas. La intensificación se realiza en función a la frecuencia con que cierta regla aparece en cierta posición de la secuencia en las cadenas de reglas que contiene el *RefSet*. También en función de esta tabla de frecuencias se realiza el proceso de Diversificación.

4. Agradecimientos

Agradecemos a Spanish Operation of Nissan y a la Cátedra Nissan UPC el apoyo dado al presente trabajo. Este trabajo ha sido parcialmente financiado por el proyecto PROTHIUS DPI2004-03475 del Gobierno de España y por la CONACYT del Gobierno de México.

5. Bibliografía

- [1] Laguna, M., Matí, R. (2003). Scatter Search. Methodology and implementations in C. Kluwer Academic Publishers. USA.
- [2] Scholl, A., Klein, R. y Domschke W. (1998). Pattern based vocabulary building for effectively sequencing mixed-model assembly lines. Journal of Heuristics 4, 359-381.
- [3] Yano, C.A. y R. Rachamadugu (1991). Sequencing to minimize work overload in assembly lines with product options. Management Science 37, 572-586.

Procedimientos de Flujo en Redes para Problemas de Planificación Open-Shop Bicriterio

Antonio Sedeño-Noda, David Alcaide, Carlos González-Martín

asedeno@ull.es, dalcaide@ull.es, cgonmar@ull.es

Departamento de Estadística, Investigación Operativa y Computación,
Universidad de La Laguna

Abstract

Se estudian problemas de planificación open – shop bicriterio con trabajos interrumpibles cuya realización debe respetar estrictamente determinadas ventanas temporales. Los criterios que se minimizan son el costo de ejecución y el instante final de completación de todos los trabajos. El problema se resuelve utilizando procedimientos de flujo en redes.

Palabras Clave: planif. open-shop, ventanas temporales, flujo en redes

AMS: 90B35, 90B10, 90C27

1. Introducción

El presente artículo trata problemas de planificación con máquinas especializadas. Los trabajos se dividen en operaciones y cada operación debe realizarse en la correspondiente máquina. Nos centramos en problemas de planificación open-shop donde los trabajos son interrumpibles y cada trabajo J_j , $j=1, \dots, n$ tiene asociada una ventana temporal $[r_j, d_j]$. En dicho intervalo deben realizarse y completarse todas las operaciones O_{ij} , $i=1, \dots, m$ que constituyen el trabajo J_j , siendo r_j su instante de disponibilidad y d_j la fecha límite en la cual dicho trabajo debe estar completado. La realización de los trabajos puede requerir diferente personal cualificado y también el empleo de diversos recursos (agua, energía, material, tecnología especializada, ...). Tanto el personal como el uso y consumo de dichos recursos generará costes de ejecución. Para computar estos costes se establecen unos turnos horarios que vendrán caracterizados por su propia tarifa. Este planteamiento de diferentes tarifas lo motiva la propia realidad ya que, por ejemplo, no se suelen pagar igual las horas de trabajo extras, nocturnas o de fin de semana, y tampoco son iguales los consumos energéticos necesarios para realizar las operaciones de día o de noche. En el presente artículo medimos este coste de ejecución a través de una función f . El otro criterio es el instante final de completación de todos los trabajos, conocido en la literatura anglosajona como “makespan”, y denotado por C_{max} . Así, siguiendo la terminología dada en [2], podemos denotar el problema considerado como $O|pmtn, r_j, d_j|(f, C_{max})$, donde f se define a continuación.

2. Planteamiento, formulación y resolución del problema

Para formular el problema se divide el tiempo en intervalos $I_k = [t_{k-1}, t_k]$ con $k = 1, \dots, K$ y $t_0 < t_1 < \dots < t_K$, siendo K el número total de intervalos.

Para definir los hitos o extremos t_k , $k = 0, 1, \dots, K$ se consideran los K_0 valores distintos entre los valores r_j , d_j , $j = 1, \dots, n$, con $K_0 \leq 2n$, que definen las ventanas temporales, y los $H+1$ valores distintos q_r , $r = 0, 1, \dots, H$ que determinan los puntos de cambio de tarifa, siendo $q_{r-1} < q_r \forall r$, y H el número de cambios de tarifa necesarios. La ordenación creciente de estos K valores distintos, con $K \leq K_0 + H$, da los hitos o extremos t_k . Así definimos las variables de decisión x_{ijk} como la cantidad procesada de la operación O_{ij} en el intervalo I_k . A partir de ellas, las variables Y_{jk} y Z_{ki} definen, respectivamente, la cantidad procesada del trabajo J_j y la cantidad que procesa la máquina M_i , en el intervalo I_k , entonces $f = \sum_{i,j,k} c_{ijk}^1 x_{ijk} + \sum_{i,j,k} c_{ijk}^2 x_{ijk}$, donde las constantes c_{ijk}^1 y c_{ijk}^2 representan, respectivamente, los costes, por unidad de tiempo, relativos al personal y al resto de recursos necesarios para la operación O_{ij} en el intervalo I_k .

La resolución del problema se hace mediante la elección adecuada de una secuencia de valores cota–parámetro C , y la resolución en secuencia de problemas de un solo criterio de la forma $O|pmtn, r_j, d_j| \sum_{i,j,k} c_{ijk}^1 x_{ijk} + \sum_{i,j,k} c_{ijk}^2 x_{ijk}$ con la restricción adicional $C_{max} \leq C$, donde C es el parámetro considerado. Supondremos factibilidad. Un algoritmo para determinar la factibilidad de este problema es dado en [3]. Resolver este problema de optimización unicriterio equivale a resolver un problema de flujo de coste mínimo sobre una red caracterizada por el número $k' = k'(C)$ de intervalos de tiempo necesarios. Con este esquema se determinan los valores óptimos para Y_{jk} y Z_{ki} relativos al parámetro C . A partir de ahí, se determinan los valores óptimos para x_{ijk} con un procedimiento dado en [3]. Finalmente, la variación adecuada de la cota – parámetro C permite resolver el problema bicriterio de partida.

3. Agradecimientos

Este trabajo está parcialmente subvencionado por los Proyectos de Investigación DPI2001-2715-C02-02 y MTM2004-07550, del Plan Nacional ayudados con fondos FEDER, y PI042004/078 del Gobierno de Canarias.

4. Bibliografía

- [1] Ahuja, R., Magnanti, T. and Orlin, J.B. (1993). *Network Flows*. Prentice-Hall.
- [2] Lawler, E.L, Lenstra, J.K., Rinnooy Kan, A.H.G. and Shmoys, D. (1993) Sequencing and scheduling: algorithms and complexity. In: Graves, S.C., Rinnooy Kan, A.H.G. and Zipkin, P.H. (Eds.), *Handbooks in Operations Research and Management Science*, vol. 4, North Holland, Chapter 9.
- [3] Sedeño-Noda, A., Alcaide, D. and González-Martín, C. (2005). Network flow approaches to pre-emptive open-shop scheduling problems with time-windows. Aparecerá en *European Journal of Operational Research*. Disponible en www.elsevier.com desde mayo de 2005.

Programación múltiple con funciones convexas generalizadas

A. Beato Moreno¹, M.B. Hernández Jiménez², Rafaela Osuna Gómez³

¹beato@us.es, Departamento de Estadística e Investigación Operativa,
Universidad de Sevilla

²beatriz.hernandez@dmat.uhu.es, Departamento de Matemáticas, Universidad
de Huelva

³rafaela@us.es, Departamento de Estadística e Investigación Operativa,
Universidad de Sevilla

Abstract

En este trabajo, se estudian problemas multiobjetivo con funciones invex generalizadas. El objetivo es deducir propiedades teóricas que permitan desarrollar algoritmos para obtener los puntos eficientes para los problemas analizados. También se analiza la aplicación de estas propiedades en programación cuadrática multiobjetivo.

Claves: Programación multiobjetivo, convexidad generalizada, eficiencia.

AMS: Clasificación AMS: 90C29, 90C25

1. Introducción

En las últimas décadas se han propuesto varias generalizaciones del concepto de convexidad. Uno de los objetivos de estas generalizaciones, es la aplicación de resultados, propiedades y algoritmos que se han obtenido para funciones convexas a conjuntos de funciones más amplios.

Por ejemplo, para funciones convexas diferenciables se cumplen ciertas propiedades como la relación entre punto estacionario y mínimo global. Si se estudia hasta dónde podemos aplicar estas propiedades, se obtienen las funciones convexas generalizadas, entre ellas las funciones invex (véase Ben Israel and Mond [2] y Craven and Glover [4]).

El concepto de ρ -convexidad fue introducido por Vial ([9]) y la noción de ρ -invexidad fue introducida por Jeyakumar ([5]). Ambos obtuvieron resultados sobre la geometría y las caracterizaciones de estas funciones, así como propiedades de optimidad en programación matemática.

El objetivo de esta comunicación es estudiar la obtención de puntos eficientes en el problema multiobjetivo con funciones convexas generalizadas. Se desarrollan procedimientos de búsqueda de puntos eficientes para este tipo de problemas. En particular se estudia la aplicación de estos procedimientos al problema multiobjetivo con funciones objetivo cuadráticas.

2. Agradecimientos

Este trabajo ha sido parcialmente financiado por los proyectos BFM2003-06579 del Ministerio de Ciencia y Tecnología y FQM-153 del Plan Andaluz de Investigación.

3. Bibliografía

- [1] A. Beato Moreno, P. Ruiz Canales, P. L. Luque Calvo y R. Blanquero Bravo (1998). Multiobjective Quadratic Problem: Characterization of the Efficient Points. In: J.P. Crouzeix, J.E. Martínez-Legaz and M. Volle (eds.), *Generalized Convexity, Generalized Monotonicity: Recent Results* Nonconvex Optimization and Its Applications, Volume 27. Kluwer Academic Publishers. 425-438.
- [2] A. Ben Israel and B. Mond (1986). What is invexity? *Journal of Australian Mathematical Society, Ser. B*, 28, 1-9.
- [3] M. Castellani and M. Pappalardo (1998). Characterizations of ρ -Convex Functions. In: J.P. Crouzeix, J.E. Martínez-Legaz and M. Volle (eds.), *Generalized Convexity, Generalized Monotonicity: Recent Results* Nonconvex Optimization and Its Applications, Volume 27. Kluwer Academic Publishers. 219-234.
- [4] B.D. Craven and B.M. Glover (1985). Invex functions and duality, *Journal of Australian Mathematical Society, (Series A)*, 39, 1-20.
- [5] V. Jeyakumar (1985). Strong and weak invexity in mathematical programming, *Mathematical Methods of Operations Research*, 55, 109-125, 1985.
- [6] R. Osuna-Gómez, A. Beato-Moreno and A. Rufian-Lizana (1999) Generalized convexity in multiobjective programming , *Journal of Mathematical Analysis and Applications* 233 , 205-220.
- [7] R. Osuna-Gómez, A. Rufian-Lizana and P. Ruiz-Canales (1998). Invex functions and generalized convexity in multiobjective programming. *Journal of Optimization Theory and Application*, 98(3), 651-661.
- [8] S. K. Suneja, C. R. Bector and C. Singh (1991). Duality in Multiobjective Fractional Programming Involving Strongly and Weakly Invex and Related Functions, *Opsearch*, 28(3), 153-164.
- [9] J.P. Vial (1983). Strong and Weak Convexity of Sets and Functions, *Mathematics of Operational Research*, 8(2), 231-259 1983.

Programación por metas con criterios fraccionales no lineales

Carmen Castrodeza¹, Teresa Peña²

¹ccch@eco.uva.es, Departamento de Economía Aplicada, Universidad de Valladolid

² maitepe@eco.uva.es, Departamento de Economía Aplicada, Universidad de Valladolid

Abstract

En este trabajo se analiza un problema de programación por metas con criterios fraccionales no lineales y se estudia como aprovechar la estructura fraccional de las funciones que aparecen en el mismo con el objeto de facilitar su resolución.

Palabras Clave: programación fraccional no lineal, programación por metas

AMS: 90C29, 90C32

1. Introducción

Dado el siguiente problema de programación por metas lexicográficas (Romero [3]):

$$\begin{aligned} \text{Lex} \quad & \min[h_1(\rho), h_2(\rho), \dots, h_l(\rho)] \\ \text{s.a.} \quad & \frac{n_i(x)}{d_i(x)} - \rho_i \leq b_i, \quad i = 1, \dots, k \\ & \rho_i \geq 0, \quad i = 1, \dots, k \\ & x \in S \end{aligned} \tag{4.75}$$

donde $x \in \mathbb{R}$ y $\rho = (\rho_1, \dots, \rho_k) \in \mathbb{R}^k$ son las variables y:

- $S = \{x \in D \subseteq \mathbb{R}^n \mid g_j(x) \leq 0, j = 1, \dots, m\}$ siendo D un conjunto convexo y abierto, $g_j : D \rightarrow \mathbb{R}, j = 1, \dots, m$ funciones continuas y convexas, S no vacío y compacto.
- $n_i : D \rightarrow \mathbb{R}, i = 1, \dots, k$ funciones continuas y convexas, $d_i : D \rightarrow \mathbb{R}, i = 1, \dots, k$ funciones continuas, cóncavas y positivas en S . Además, n_i es positiva en S si d_i no es lineal.
- $b_i, i = 1, \dots, k$ el nivel máximo de aspiración determinado por el centro decisor para cada uno de los objetivos $\frac{n_i(x)}{d_i(x)}, i = 1, \dots, k$ y l ($l \leq k$) el número de niveles de prioridad entre las metas establecidos por el mismo.
- $h_s(\rho), s = 1, \dots, l$ una función de las variables de desviación correspondientes al nivel de prioridad s , tal que $h_s(\rho) = \sum_{i \in Q_s} w_i \rho_i, s = 1, \dots, l$ (enfoque lexicográfico ponderado) o $h_s(\rho) = \max_{i \in Q_s} w_i \rho_i, s = 1, \dots, l$

(enfoque lexicográfico minimax) donde Q_s , $s = 1, \dots, l$ denota el nivel de prioridad s , $i \in Q_s$ si la meta $\frac{n_i(x)}{d_i(x)} \leq b_i$ está en el nivel s y w_i , $i \in Q_s$ son pesos positivos que representan la importancia relativa dada por el centro decisor a las variables de desviación, ρ_i , asociadas al nivel s .

Cuando se analiza la resolución de (1) bajo cualquiera de los dos enfoques considerados nos encontramos con la dificultad de tener que resolver para cada nivel de prioridad un problema no convexo.

2. Resolución del modelo

En este trabajo hemos demostrado que se puede plantear un problema convexo asociado al verdadero problema a resolver en cada nivel de prioridad de tal forma que la resolución de dicho problema asociado nos permita conocer si en el problema original existen o no soluciones factibles en las que se verifiquen todas las metas de dicho nivel y en el caso de que existan nos proporcione las mismas. Si no existen dichas soluciones, hemos comprobado que la solución del problema convexo asociado no tiene por qué coincidir con la del problema original, por lo que habría que resolver éste directamente. Para llevar a cabo esa tarea, hemos planteado diversos cambios en el mismo con el fin de obtener un problema equivalente con un conjunto factible convexo y en el que las funciones fraccionales aparezcan en la función objetivo. Para resolver esos problemas equivalentes proponemos distintos algoritmos (Benson [1], Gugat [2], Schaible [4]) que aprovechan la especial estructura de la función objetivo de cada uno de ellos y que permiten que sean resueltos de forma iterativa utilizando algoritmos de programación convexa.

3. Bibliografía

- [1] Benson, H.P. (2002). Global optimization algorithm for the nonlinear sum of ratios problem. *Journal of Optimization Theory and Applications*, 112, 1-29.
- [2] Gugat, M. (1996). A fast algorithm for a class of generalized fractional programs. *Management Science*, 42, 1493-1499.
- [3] Romero, C. (1991). *Handbook of Critical Issues in Goal Programming*. Pergamon Press, Oxford Wesley.
- [4] Schaible, S. (1976). Fractional programming II. On Dinkelbach's algorithm. *Management Science*, 22, 868-873.

Propiedades de los sistemas completos borrosos

Susana Díaz¹, Pedro Gil¹, Susana Montes¹

¹diazsusana,pedro,montes@uniovi.es, Departamento de Estadística,
Investigación Operativa y Didáctica de la Matemática, Universidad de Oviedo

Abstract

En este trabajo se estudia el comportamiento de los refinamiento y productos de sistemas completos de sucesos borrosos. Así, se han caracterizado aquellas t-normas para las cuales dichas operaciones permiten obtener de nuevo un sistema completo de sucesos borrosos.

Claves: sistema completo borroso, t-norma.

1. Introducción

Un concepto fundamental en la Teoría de la Información es el de medida de incertidumbre o entropía (ver, por ejemplo, [5]), la cual nos mide la incertidumbre o falta de información que tenemos sobre el suceso que ocurrirá al realizar un experimento dado, es decir, el suceso que se seleccionará de un sistema completo de sucesos. Por otro lado, algunos experimentos son modelados de forma más eficiente si se considera la posibilidad de que los resultados de los mismos no estén definidos con total precisión, es decir, si se considera que el resultado del experimento aleatorio puede ser un suceso borroso (ver, por ejemplo, [1]). En tales casos, los sistemas completos de sucesos, también podrían estar formados por sucesos borrosos, en cuyo caso todos los conceptos clásicos deben ser replanteados y estudiados con detenimiento. En esta dirección hemos ido dando algunos pasos ([3], [4]), hasta concluir en este trabajo, en el que se caracterizan totalmente aquellas t-normas para las que las operaciones entre sistemas cumplen los requerimientos naturales.

2. Conceptos previos y resultados

Asociado a un suceso borroso está siempre el conjunto borroso que representa los posibles resultados del espacio muestral Ω que suponen la realización de dicho suceso. Un conjunto borroso no es más que un conjunto al que los elementos de Ω pertenecen (valor 1), no pertenecen (valor 0), o pertenecen en cierto grado (valor en (0,1)), en contraposición a un conjunto clásico o nítido, al que los elementos simplemente pertenecen (1) o no pertenecen (0). En el caso de considerar conjuntos borrosos, las operaciones entre conjuntos se generalizan a través de una aplicación $T : [0, 1] \times [0, 1] \rightarrow [0, 1]$, denominada t-norma o norma triangular ([2]), que surgió inicialmente al generalizar la clásica desigualdad triangular en los espacio métricos a espacio métricos probabilísticos. Dicha aplicación T tiene que ser conmutativa, asociativa, creciente en cada componente y con elemento neutro en 1. Con ellas, la unión de dos conjuntos borrosos se define como el conjunto borroso con función

de pertenencia $A \cup B(\omega) = 1 - T(1 - A(\omega), 1 - B(\omega))$ y la intersección como $A \cap B(\omega) = T(A(\omega), B(\omega)), \forall \omega \in \Omega$. El concepto de ϵ -igualdad de conjuntos borrosos ([3]), denotado $=_\epsilon$, generaliza la igualdad de conjuntos clásicos. Así, una vez presentados los conceptos de unión, intersección e igualdad de conjuntos borrosos, definimos en ([3]) un sistema completo de sucesos borrosos como la familia de conjuntos borrosos cuya unión es ϵ -igual a Ω y cuyas intersecciones dos a dos son ϵ -iguales al conjunto vacío (conjuntos ϵ -disjuntos). Esta definición de sistema completo de sucesos borrosos generaliza otras muchas existentes en la literatura (véase [4]). Dado un sistema completo de sucesos borrosos, un refinamiento se define como una nueva familia de sucesos borrosos de forma que para cada elemento del sistema inicial, existen una serie de elementos del refinamiento que forman a su vez una partición de él. Se define el producto de dos sistemas como la familia formada por las intersecciones de los conjuntos de ambos sistemas. Con esto, dada una t-norma T el refinamiento de un sistema completo de sucesos borrosos es a su vez un sistema completo de sucesos borrosos si y sólo si $T(1 - \epsilon, x) = x$, para todo $x \in (\epsilon, 1 - \epsilon)$. Por otra parte, el producto de dos sistemas es un nuevo sistema completo de sucesos borrosos si y sólo si, además de la condición anterior, se cumplen las dos condiciones siguientes: (1) $\forall x \leq y \mid T(x, y) \leq \epsilon$, $T(S(x, x), S(x, y)) \leq \epsilon$; (2) $\forall x_1 \geq y_1 \geq y_2 \geq \dots \geq y_j > x_2 \mid T(x_1, x_2) \leq \epsilon$, $T(y_1, y_2) \leq \epsilon$ y $T(x_1, y_j) > \epsilon$, $T(x_1, S(y_1, \dots, y_j)) = S(T(x_1, y_1), \dots, T(x_1, y_j)) = T(S(x_1, x_2, m^1, x_2), S(y_1, \dots, y_j, x_2, m^2, x_2))$. Lejos de lo que pudiera parecer, estas condiciones no son demasiado restrictivas, de hecho, en el caso de trabajar con t-normas continuas, tanto la t-norma idempotente, como una amplia clase de t-normas sumas ordinarias las cumplen.

3. Agradecimientos

Este trabajo está subvencionado en parte por el Proyecto MTM2004-01269.

4. Bibliografía

- [1] Gil M.A. (1993). Análisis y tratamiento estadístico de elementos difusos en experimentos aleatorios. *Estadística Española* 35(134), 477-525.
- [2] Klement E.P., Mesiar R. y Pap E. (2000). *Triangular Norms*. Kluwer Academic Publishers.
- [3] Montes S., Couso I. y Gil P. (2001). One-to-one correspondences between epsilon-partitions, (1-epsilon)-equivalences and epsilon-pseudometrics. *Fuzzy Sets and Systems* 124, 87-95.
- [4] Montes S., Couso I. y Gil P. (2003). Fuzzy delta-epsilon-partitions. *Information Sciences* 152, 267-285.
- [5] Pardo, L. (1997). *Teoría de la Información estadística*. Ed. Hesperides.

Reglas hidráulicas para problemas de bancarrota con referencias

Mercedes Alemán Jiménez¹, Carlos González-Alcón²

¹ I.E.S. La Orotava-Manuel Glez. Pérez
malejim@gobiernodecanarias.org

² Dpto. de Estadística, Investigación Operativa y Computación
Universidad de La Laguna
cgalcon@ull.es

Abstract

Con ayuda de la representación hidráulica se proponen nuevas reglas y familias de reglas para problemas de bancarrota, tanto clásico como con referencias.

Claves: Bankruptcy, Claims Problems, Hydraulic Rules.

AMS: Clasificación AMS:

91A12, 91A40, 91B32.

1. Introducción

Siguiendo la línea comenzada por O'Neill (1982) han sido muchos los autores que han trabajado los problemas de reparto conocidos como de bancarrota. En estos problemas se ha de repartir un bien escaso entre diversos agentes cuyas demandas exceden el total disponible. Un buen resumen de lo publicado hasta ahora puede encontrarse en Thomson (2002).

Las reglas de reparto son propuestas para realizar la asignación a cada agente. Estas reglas deben satisfacer que todo agente reciba una asignación positiva y como mucho su demanda. Kaminski (2000) define un tipo de reglas que denomina "hidráulicas" porque admiten una representación mediante un sistema de vasos intercomunicados: el patrimonio total es ahora un líquido y a cada agente se le asigna la cantidad de ese líquido contenida en sus vasos.

2. Familias de reglas

Las familias de reglas de reparto son conjuntos de reglas que aplican un determinado principio y tienen una expresión común quedando indexadas por uno o varios parámetros. Moreno-Ternero y Villar (2002) definieron la familia Talmud, donde para determinado valor del parámetro se tiene la regla del Talmud.

Definimos otras familias en las que, al igual que en la familia Talmud, para los valores extremos del parámetro se obtienen las reglas de pagos y de pérdidas

iguales restringidas (Aumann (1985)). De esta manera se busca un equilibrio en las asignaciones obtenidas por los agentes con bajas demandas y los agentes con altas demandas.

3. Problemas con referencias

En las situaciones de bancarrota con referencias además de repartir el patrimonio según las demandas de los agentes, se tienen en cuenta puntos de referencia externos para determinar una justa división del patrimonio (Pulido *et al.* (2002)). Para este tipo de problemas definimos nuevas reglas y familias de reglas hidráulicas, extensiones de otras para el problema de bancarrota clásico.

4. Bibliografía

- [1] Aumann, R. and Maschler, M. (1985). Game theoretic analysis of a bankruptcy problem from the Talmud. *Journal of Economic Theory* 36, 195-213.
- [2] Kaminski, M. (2000). “Hydraulic” Rationing. *Mathematical Social Sciences* 40, 131-155.
- [3] Moreno-Ternero, J.D. and Villar, A. (2002). The TAL-family of rules for bankruptcy problems. *Social Choice and Welfare* (En prensa).
- [4] Pulido, M., Sánchez-Soriano, J. and Llorca, N. (2002). Game Theory Techniques for University Management: An Extended Bankruptcy Model. *Annals of Operations Research* 109, 129-142.
- [5] Thomson, W. (2003). Axiomatic and game-theoretic analysis of bankruptcy and taxation problems: a survey. *Mathematical Social Sciences* 45, 249-297.

Regularidad métrica del conjunto de soluciones óptimas en Programación Semi-Infinita

M. J. Cánovas¹, D. Klatte², M. A. López³, J. Parra¹,

¹canovas@umh.es, parra@umh.es, Centro de Investigación Operativa,
Universidad Miguel Hernández de Elche,

²klatte@ior.unizh.ch, Operations Research Institute, University of Zurich,
CH-8044 Zurich, Switzerland

³ marco.antonio@ua.es, Departamento de Estadística e Investigación
Operativa, Universidad de Alicante

Abstract

Caracterizamos la regularidad métrica de la inversa de la multifunción conjunto óptimo asociada a una familia de problemas de optimización lineal parametrizada con respecto al miembro derecho de las restricciones y al vector de coeficientes de la función objetivo. Asumimos la existencia de un conjunto fijo de índices (posiblemente infinito), compacto Hausdorff, y la continuidad de los vectores de coeficientes de las restricciones con respecto al índice. También se analiza la posible extensión de los resultados al caso convexo.

Claves: Regularidad métrica, conjunto óptimo, programación semi-infinita.

1. Introducción

Consideremos los problemas de Programación Semi-Infinita, en \mathbb{R}^n ,

$$\begin{aligned} \pi(c, b) : \quad & \inf c'x \\ \text{s. t. } & a'_t x \geq b_t, \quad t \in T, \end{aligned} \quad \text{y} \quad \begin{aligned} P(c, b) : \quad & \inf f(x) + c'x \\ \text{s. t. } & a'_t x \geq b_t, \quad t \in T, \end{aligned}$$

donde $x \in \mathbb{R}^n$, T es un espacio compacto Hausdorff, $a \in C(T, \mathbb{R}^n)$, $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una función convexa, y $c \in \mathbb{R}^n$ y $b \in C(T, \mathbb{R})$ constituyen los parámetros del modelo. Denotaremos por $\sigma(b)$ al sistema de restricciones asociado a $\pi(c, b)$ y $P(c, b)$. En el espacio paramétrico $\mathbb{R}^n \times C(T, \mathbb{R})$ se considera la norma $\|(c, b)\| := \max\{\|c\|_2, \|b\|_\infty\}$. Sean $\mathcal{G} : \mathbb{R}^n \rightrightarrows C(T, \mathbb{R})$ y $\mathcal{G}^* : \mathbb{R}^n \rightrightarrows \mathbb{R}^n \times C(T, \mathbb{R})$:

$$\begin{aligned} \mathcal{G}(x) &:= \{b \in C(T, \mathbb{R}) \mid a'_t x \geq b_t \text{ para todo } t \in T\}, \\ \mathcal{G}^*(x) &:= \{(c, b) \in \mathbb{R}^n \times \mathcal{G}(x) \mid x \in \arg \min \{f(y) + c'y \mid b \in \mathcal{G}(y)\}\}. \end{aligned}$$

La regularidad métrica de \mathcal{G} (propiedad de Aubin, o pseudo-Lipschitz, de $\mathcal{F} := \mathcal{G}^{-1}$) ha sido analizada en [1].

2. Regularidad métrica de \mathcal{G}^*

\mathcal{G}^* es métricamente regular (m.r.) en \bar{x} para $(\bar{c}, \bar{b}) \in \mathcal{G}^*(\bar{x})$ si existen entornos U de \bar{x} y V de (\bar{c}, \bar{b}) , y una constante $\kappa \geq 0$, tal que $d(x, \mathcal{F}^*(c, b)) \leq \kappa d((c, b), \mathcal{G}^*(x))$ para todo $x \in U$ y todo $(c, b) \in V$. La siguiente propiedad, referida a $(\bar{x}, (\bar{c}, \bar{b})) \in gph\mathcal{G}^*$, desempeña un papel crucial en nuestra teoría (aquí $T_{\bar{b}}(\bar{x}) = \{t \in T \mid a'_t \bar{x} = \bar{b}_t\}$):

$$\begin{aligned} \sigma(\bar{b}) \text{ satisface la condición de Slater y no existe } D \subset T_{\bar{b}}(\bar{x}) \\ \text{con } |D| < n \text{ tal que } (\bar{c} + \partial f(\bar{x})) \cap \text{cone}(\{a_t, t \in D\}) \neq \emptyset. \end{aligned} \quad (4.76)$$

El siguiente teorema muestra que, para $\pi(c, b)$, la regularidad métrica de \mathcal{G}^* resulta ser equivalente a otros criterios de estabilidad de la literatura ([2], [3], [4]). $\mathcal{F}^* (= (\mathcal{G}^*)^{-1})$ representa a la multifunción conjunto óptimo, y $(\bar{x}, (\bar{c}, \bar{b})) \in gph\mathcal{G}^*$.

Theorem 3 Para $\pi(c, b)$, son equivalentes:

- (i) \mathcal{G}^* es m.r. en \bar{x} para (\bar{c}, \bar{b}) ;
- (ii) \mathcal{F}^* tiene la propiedad de strong Lipschitz stability en $((\bar{c}, \bar{b}), \bar{x})$;
- (iii) \mathcal{F}^* es una función punto a punto y continua en un entorno de (\bar{c}, \bar{b}) ;
- (iv) \mathcal{F}^* es una función punto a punto en algún entorno de (\bar{c}, \bar{b}) ;
- (v) Se cumple (4.76).

Theorem 4 Para $P(c, b)$, si (4.76) entonces \mathcal{G}^* es m.r. \bar{x} para (\bar{c}, \bar{b}) .

3. Bibliografía

- [1] M. J. CÁNOVAS, A. L. DONTCHEV, M. A. LÓPEZ AND J. PARRA (2005), *Metric regularity of semi-infinite constraint systems*, Math. Program., 104, 329-346.
- [2] D. KLATTE AND B. KUMMER (2002), *Nonsmooth Equations in Optimization-Regularity, Calculus, Methods and Applications*. Nonconvex Optimization and Its Applications. Kluwer, Dordrecht.
- [3] M. KOJIMA (1980), *Strongly stable stationary solutions in nonlinear programs*, in Analysis and Computation of Fixed points, S.M. Robinson, ed., Academic Press, New York, pp. 93-138.
- [4] J.-J. RÜCKMANN (1999), *On existence and uniqueness of stationary points in semi-infinite optimization*. Math. Program., 86, 387-415.

Reparto de holguras en una red PERT

Javier Castro¹, Daniel Gómez², Juan Tejada³

¹jcastroc@estad.ucm.es, Departamento de Estadística e Investigación operativa III, Universidad Complutense

²dagomez@estad.ucm.es, Departamento de Estadística e Investigación operativa III, Universidad Complutense

³jtejada@mat.ucm.es, Departamento de Estadística e Investigación operativa I, Universidad Complutense

Abstract

Se modeliza el problema del reparto de holguras (SPERT) en una red PERT mediante una nueva clase de problemas de reparto con restricciones y pesos (PCW), definiendo una familia de reglas de reparto que se caracterizada. Para el problema SPERT se analizan los pesos más adecuados y se define un algoritmo polinomial en media que calcula el reparto correspondiente.

Claves: PERT, Calendario, Reparto de holguras.

1. Introducción

Bergantiños y Sánchez (2002) introducen los problemas de reparto con restricciones y peticiones (PCC). Analizan como caso particular de este problema el problema SPERT, considerando que las peticiones de las actividades, vienen dadas por la holgura máxima que cada una de ellas puede disponer. Dicho planteamiento no proporciona soluciones consistentes ante la subdivisión de actividades, por lo que se introduce el PCW, donde los repartos factibles no están acotados por el valor del peso asignado a cada actividad.

Los pesos se interpretan como una medida de la importancia de las actividades en cualquier reparto, no como una aspiración a la que hay que intentar satisfacer. Por ello se propone una modificación de la regla de reparto introducida por Bergantiños y Sánchez (2002), que utiliza los mismos pesos en cada iteración y que se llamará regla proporcional a los pesos. Finalmente se define un algoritmo polinomial en media que calcula el reparto.

2. El problema

Un PCW se define como una 4-tupla (N, P, W, E) , donde:

- $N = \{1, \dots, n\}$, es el conjunto de agentes.
- $P = \{\pi_1, \dots, \pi_p\}$, es un conjunto de subconjuntos de N , que cubren N .
- $W = (w_1, \dots, w_n) >> 0$, es un vector de pesos de los agentes.

- $E = (E_1, \dots, E_p) > 0$, es el vector que determina las restricciones en P .

El PCW consiste en asignar una cantidad no negativa que depende de los pesos, de forma que la suma de las cantidades asignadas a los agentes de π_j no excedan de E_j , $\forall j = 1, \dots, p$.

Una vez definido el PCW, se modeliza un problema SPERT como una 4-tupla (N, P, W, E) , donde N , es el conjunto de actividades, P son los caminos completos de la red, W es el vector de pesos de las actividades y E son las holguras de cada uno de los caminos.

3. Regla para el problema SPERT

La regla proporcional a los pesos es una generalización de la regla proporcional a las duraciones definida en Castro et al. (2005), tomando como pesos las duraciones de las actividades. Dicha regla se caracterizó mediante cuatro propiedades (consistencia, mínimo valor por camino, optimalidad y no manipulabilidad) y se definía mediante un algoritmo de complejidad exponencial.

En el problema SPERT, se trata de definir pesos que aplicados a la regla proporcional a los pesos, garanticen un reparto tal que la probabilidad de retraso para actividades simétricas sea la misma. Se demuestra que para las distribución uniforme, triangular inferior y triangular superior, los únicos pesos que lo garantiza, son aquellos proporcionales a los rangos de las variables aleatorias que modelizan las duraciones de las variables.

Por último, se presenta un algoritmo polinomial en media para el cálculo de la regla proporcional a los pesos. Este algoritmo permite realizar un reparto de la holgura de forma proporcional a unos pesos o a unas peticiones en problemas SPERT con gran número de actividades.

4. Agradecimientos

Plan Nacional de I+D+i: MTM2005-09184-C02-01

5. Bibliografía

- [1] G. Bergantiños y E. Sánchez (2002). The proportional rule for problems with constraints and claims. *Mathematical Social Sciences*. 43, 225-249.
- [2] J. Castro, D. Gómez y J. Tejada (2005). A rule for slack allocation proportional to the durations in a PERT network. *Working paper*

Resolución del Response Time Variability Problem mediante programación matemática

Albert Corominas¹, Wieslaw Kubiak², Rafael Pastor¹

¹albert.corominas/rafael.pastor @upc.edu, Research Institute IOC,
Universidad Politécnica de Cataluña

²Faculty of Business Administration, Memorial University of Newfoundland

Abstract

El *Response Time Variability Problem* ha sido recientemente definido en la literatura, en la que se ha propuesto resolverlo óptimamente mediante programación matemática (PLEM), con un límite práctico de 25 unidades. En este trabajo se mejora el modelo de PLEM actual más eficiente hasta ahora mediante reformulación, restricciones redundantes y eliminación de simetrías, con lo que se llega a resolver de forma óptima ejemplares de hasta 40 unidades.

Palabras Clave: response time variability, optimización, secuenciación

1. Introducción

El Response Time Variability Problem (RTVP) es un problema de secuenciación que ha sido recientemente definido en la literatura –únicamente se conocen dos referencias sobre RTVP: León *et al.* (2003) y Corominas *et al.* (2004)–. RTVP es muy simple de definir, pero es un problema de secuenciación muy difícil de resolver de forma óptima.

Sea n el número de productos. Sea d_i el número de unidades de producto i , ($i = 1, \dots, n$) a secuenciar en una secuencia $s = s_1s_2\dots s_D$ de longitud D ($D = \sum_{i=1}^n d_i$), es decir, con D posiciones. Para todo tipo de producto con $d_i \geq 2$, sea t_k^i la distancia entre la posición de las unidades $k+1$ y k del producto i ; se asume que sigue de forma inmediata a s_D (es decir, es una secuencia circular), entonces $t_{d_i}^i$ es la distancia entre la primera unidad de producto i en un ciclo y la última unidad del mismo producto en el ciclo anterior. Sea $\bar{t}_i = D/d_i$ la distancia promedio entre dos unidades consecutivas de producto i . El objetivo consiste en minimizar $RTV = \sum_{i=1}^n \sum_{k=1}^{d_i} (t_k^i - \bar{t}_i)^2$, lo cual es una suma de variancias ponderadas con unos pesos iguales a las demandas: $RTV = \sum_{i=1}^n d_i \cdot \text{Var}_i$, donde $\text{Var}_i = \frac{1}{d_i} \sum_{k=1}^{d_i} (t_k^i - \bar{t}_i)^2$.

2. Resolución óptima de RTVP

RTVP es definido en Corominas *et al.* (2004), donde se prueba que es NP-duro. En dicho trabajo se propone un algoritmo óptimo para el caso de dos productos. Cuando se trabaja con más de dos productos se considera RTVP como un caso especial del *quadratic assignment problem*, el cual es formalizado

como un programa cuadrático entero y es linealizado aplicando tres técnicas diferentes. Con el mejor de los programas matemáticos obtenidos al linealizar (MILP), dichos autores consiguen resolver de forma óptima ejemplares de hasta unas 25 unidades.

Analizando las características especiales de RTVP, varias ideas son expuestas en este trabajo para mejorar la eficiencia del mejor programa matemático MILP presentado en la literatura. Las contribuciones del artículo son: i) resolver de forma optima problemas de mayor tamaño; y ii) mostrar el enorme impacto que tiene la formulación, el uso de restricciones redundantes y la eliminación de simetrías en la efectividad de los MILP.

Las ideas propuestas consisten en reformular el programa matemático original, incluir restricciones redundantes y eliminar diversas clases de simetrías que presenta el problema. Se realiza una experiencia computacional para probar la validez de las mejoras propuestas, con los siguientes resultados: i) se consigue resolver de forma óptima ejemplares de hasta unas 40 unidades; ii) para un conjunto de 20 ejemplares prueba, por un lado, con el modelo final propuesto se consigue resolver de forma óptima los 20 ejemplares (mientras que con el modelo original publicado tan solo se resuelve uno de dichos 20 ejemplares), y, por otro, el tiempo de cálculo se reduce desde más de 38.603 segundos (ya que el tiempo para resolver cada ejemplar está limitado a 2.000 segundos) a tanto solo 398.

3. Agradecimientos

Financiado por el proyecto del MCYT DPI2004-03472, cofinanciado por FEDER.

4. Bibliografía

- [1] Corominas A, Kubiak W and Moreno N. Response time variability. *Working paper IOC-DT-P-2004-08*, UPC, Barcelona, 2004.
- [2] León D, Corominas A and Lusa A. Resolución del problema PRV min-var. *Working paper IOC-DT-I-2003-03*, UPC, Barcelona, 2003.

Resolución óptima del Multi-Min-Max Assembly Line Balancing Problem (MMM-ALBP) mediante programación matemática

Rafael Pastor¹, Jordi Ojeda^{1,2}, Carles Sentinella³, Jordi Solsona³

¹rafael.pastor/jordi.ojeda@upc.edu, Research Institute IOC, Universidad Politécnica de Cataluña

²Fundación CIM

³ETS de Ing. Ind. de Barcelona, Universidad Politécnica de Cataluña

Abstract

Se define un nuevo problema de equilibrado de líneas de montaje, el *Multi-Min-Max Assembly Line Balancing Problem* (MMM-ALBP), que consiste en minimizar el tiempo de la estación más cargada, posteriormente el tiempos de la segunda estación más cargada, posteriormente el de la tercera, etc. y se resuelve de forma óptima mediante programación matemática.

Palabras Clave: equilibrado de líneas de montaje.

1. Introducción

Existen muchos problemas para los que se puede plantear un objetivo Mínmax, consistente en minimizar el máximo valor alcanzable: localización; discrepancia respecto a un valor ideal; carga de estaciones en líneas de montaje; ...

La resolución óptima de problemas con este tipo de objetivo consigue que el valor máximo sea lo menor posible, pero no se considera qué pasa con los demás valores, y éstos también pueden tener gran importancia (ya que, sin superar el valor máximo, su distribución puede ser muy irregular y, de esta forma, inadecuada).

El problema de equilibrado de líneas de montaje es un problema muy estudiado en la literatura y entre los últimos trabajos de síntesis cabe destacar los de Rekiek et al. 2002, Becker and Scholl 2006 y Scholl and Becker 2006.

En el problema de equilibrado de líneas de montaje el considerar el segundo valor máximo y el tercero, etc., tiene gran importancia ya que permite ganar en fiabilidad de la línea. No es lo mismo una distribución de la carga de trabajo entre 6 estaciones de, por ejemplo, 10, 10, 10, 4, 3, 3, que una distribución de 10, 6, 6, 6, 6, 6; el valor del tiempo ciclo es el mismo pero, indudablemente, la segunda distribución es más fiable que la primera. Por otro lado, y como se observa en el ejemplo anterior, el objetivo multi-min-max también repercute en conseguir una distribución homogénea de la carga de trabajo entre las diferentes estaciones de la línea.

Existen funciones objetivo que también tienden a homogeneizar la distribución de la carga de trabajo, como podría ser la suma, para todas las estaciones, del cuadrado de la discrepancia entre el tiempo medio de carga de las estaciones y la carga de dicha estación ($[MIN] Z = \sum_{j=1}^m (\bar{t} - t_j)^2$), siendo m el número de estaciones y t_j la carga de la estación j); pero cabe recordar que la función multi-min-max persigue un objetivo distinto.

2. El Multi-Min-Max Assembly Line Balancing Problem (MMM-ALBP)

En este trabajo se define un nuevo problema de equilibrado de líneas de montaje, el *Multi-Min-Max Assembly Line Balancing Problem* (MMM-ALBP), que consiste en minimizar el tiempo de la estación más cargada, posteriormente el tiempo de la segunda estación más cargada, posteriormente el de la tercera, etc.

3. Resolución óptima del MMM-ALBP

Para resolver el MMM-ALBP de forma óptima se han diseñado dos modelos de programación matemática independientes, denominados modelo jerarquizado global y jerarquizado sucesivo, los cuales han sido resueltos mediante *CPLEX*. La eficiencia de los modelos ha sido evaluada resolviendo ejemplares de prueba de la base de ejemplares accesible en www.assembly-line-balancing.de para el equilibrado de líneas de montaje.

4. Agradecimientos

Financiado por el proyecto del MCYT DPI2004-03472, cofinanciado por FEDER.

5. Bibliografía

- [1] Becker C and Scholl A. (2006) A survey on problems and methods in generalized assembly line balancing. *European Journal of Operational Research* 168, 694-715.
- [2] Rekiek B, Dolgui A, Delchambre A and Bratcu A. (2002) State of art of optimization methods for assembly line design. *Annual Reviews in Control* 26, 163-174.
- [3] Scholl A and Becker C. (2006) State-of-the-art exact and heuristic solution procedures for simple assembly line balancing. *European Journal of Operational Research* 168, 666-693.

Retrospectiva y desafíos del problema de planificación de la producción

J. Gutiérrez¹, A. Sedeño-Noda², M. Colebrook³, B. Abdul-Jalbar⁴,
J. Sicilia⁵

¹jmgrrrez@ull.es, Departamento de Estadística, Investigación Operativa y
Computación (DEIOC), Universidad de La Laguna

²asedeno@ull.es, DEIOC ³mcolebrook@ull.es, DEIOC ⁴babdul@ull.es, DEIOC

⁵jsicilia@ull.es, DEIOC

Abstract

Presentamos este trabajo como una recopilación esquematizada de las principales aportaciones que del problema de la planificación de la producción se han publicado en las últimas décadas. Además, propondremos líneas abiertas para futuros trabajos de investigación.

Claves: Planificación de la producción, Capacidades, Inventario.

AMS: 90B05,90B30

1. Introducción

Con este artículo pretendemos recoger las aportaciones más relevantes que se han ido desarrollando en las últimas décadas para el problema de la planificación de la producción con capacidades de almacenaje, así como plantear líneas abiertas para futuros trabajos de investigación. Formalmente, este problema considera un horizonte temporal dividido en T ($t = 1, 2, \dots, T$) períodos y, para cada uno de ellos, se asume una demanda d_t , una función de coste de producción $C_t(\cdot)$, una función de coste de mantenimiento-rotura $H_t(\cdot)$, y una capacidad de inventario S_t conocidas. El objetivo por lo tanto consiste en determinar un plan de producción $\mathbf{Q} = (Q_1, Q_2, \dots, Q_T)$ que satisfaga las demandas de cada periodo a mínimo coste. Matemáticamente, el modelo se puede plantear como sigue:

$$\begin{aligned} & \min \sum_{t=1}^T C_t(Q_t) + H_t(I_t) \\ \text{s.a. } & I_0 = I_T = 0 \\ & I_{t-1} + Q_t - I_t = d_t \quad t = 1, 2, \dots, T \\ & I_{t-1} + Q_t \leq S_t \quad t = 1, 2, \dots, T \\ & I_t \in \mathbb{Z}, Q_t \in \mathbb{N}_0 = \mathbb{N} \cup \{0\} \quad t = 1, 2, \dots, T \end{aligned}$$

Fue Love en 1973 quien, por primera vez, resuelve eficientemente este problema desarrollando una algoritmo de programación dinámica con complejidad temporal $O(T^3)$. Creemos que desde entonces el modelo no suscitó más interés

ya que había sido resuelto satisfactoriamente. Sin embargo, en los últimos años se han diseñado nuevos algoritmos para estructuras de coste más específicas cuya complejidad mejoran la del procedimiento propuesto por Love. A lo largo de la exposición nos centraremos en la discusión de cada una de estas variantes detallando la caracterización de sus soluciones óptimas y el método solución empleado. Concluiremos este trabajo proponiendo extensiones al modelo que se plantearán como líneas de investigación futura.

2. Bibliografía

- [1] Gutiérrez, J., A. Sedeño-Noda, M. Colebrook, J. Sicilia (2003). A New Characterization for the Dynamic Lot Size Problem with Bounded Inventory. *Computers and Operations Research* 30, 383-395.
- [2] Gutiérrez, J., A. Sedeño-Noda, M. Colebrook, J. Sicilia (2003). A Polynomial Algorithm for the Production/ordering Planning Problem with Limited Inventory. *Computers and Operations Research*. En prensa.
- [3] Gutiérrez, J., A. Sedeño-Noda, M. Colebrook, J. Sicilia. An Efficient Approach for Solving the Production/ordering Planning Problem with Time-varying Storage Capacities. *European Journal of Operational Research*. En revisión.
- [4] Love, S.F. (1973), Bounded Production and Inventory Models with Piecewise Concave Costs. *Management Science* 20, 313-318.
- [5] Sedeño-Noda, A., J. Gutiérrez, B. Abdul-Jalbar, J. Sicilia (2004). An O(TlogT) Algorithm for the Dynamic Lot Size Problem with Limited Storage and Linear Costs. *Computational Optimization and Applications* 28, 311-323.

Risk measures in MPSSP

A. Alonso-Ayuso¹, L.F. Escudero², C. Pizarro³

¹antonio.alonso@urjc.es, Departamento de Estadística e Investigación Operativa ,Universidad Rey Juan Carlos

²escudero@umh.es, Centro de Investigación Operativa, Universidad Miguel Hernández

³celeste.pizarro@urjc.es, Departamento de Estadística e Investigación Operativa ,Universidad Rey Juan Carlos

Abstract

En este trabajo se presenta un modelo de Programación Estocástica para el MPSSP: Multi-Period Single Source Problem con incertidumbre en la demanda y los costes. En el objetivo del problema se incluyen términos de medida del riesgo, que hacen poco eficientes las metodologías usadas en modelos de minimización del coste esperado. Como alternativa, se ha adaptado la heurística Fix and Relax Coordinado a esta nueva situación. Se presentan resultados computacionales para casos de grandes dimensiones.

Claves: programación estocástica, Fix-and-Relax Coordinado, mean-risk.

1. Introducción

El objeto de este trabajo es la presentación de un método de aproximación heurística para el MPSSP, usando medidas de riesgo, para problemas 0-1 mixtos bietápicos multiperiodo con incertidumbre en la demanda y en los coste. Este problema permite asignar fábricas a clientes a lo largo de un horizonte de tiempo, dependiendo de la demanda en cada período. Los modelos tradicionales sólo consideran la versión determinística del problema, donde todos los parámetros son conocidos. Esta simplificación difícilmente puede ser una representación precisa de la realidad. Se aborda el problema desde la óptica de la Programación Estocástica vía el análisis de escenarios. Los modelos estocásticos previos consideran como objetivo el valor mínimo esperado, el cual, en situaciones en las que se lleva a cabo la toma de decisiones bajo condiciones similares repetidas a lo largo del tiempo puede dar buenos resultados. Sin embargo, en casos en los que se presente variabilidad en los datos, el valor esperado puede dar malos resultados.

Por esta razón, es necesario la implementación de modelos de riesgo. En particular, en este trabajo se ha optado por el *mean-risk*. Se plantea una aproximación heurística basada en el Fix-and-Relax Coordinado, consistente en la asignación secuencial de clientes a fábricas. Se analizan diversas formas de

creación de los grupos de clientes que se van a fijar en cada etapa. Por último, se reportan resultados computacionales y se comparan con otros resultados referidos a métodos de resolución usados hasta el momento.

2. Bibliografía

- [1] A. Alonso-Ayuso, L.F. Escudero, and M.T. Ortúñoz. (2003). BFC, a Branch-and-Fix Coordination algorithmic framework for solving stochastic 0-1 programs. *European Journal of Operational Research* 151, 503-519.
- [2] A. Alonso-Ayuso, L.F. Escudero, H.E. Romeijn, C. Pizarro, D. Romero Morales (2006). On solving the multi-period single-sourcing problem under uncertainty. *Computational Management Science* 3, 29-53
- [3] A. Alonso-Ayuso, L.F. Escudero, M.T. Ortúñoz and C. Pizarro. (2005). On a Stochastic Sequencing and Scheduling problem. *Computers and Operations Research*, to appear.

Secuenciación mediante algoritmos evolutivos en complejos talleres flexibles

Rubén Ruiz¹, Thijs Urlings¹, Funda Sivrikaya Serifoğlu²

¹rruiz@eio.upv.es, Departamento de Estadística e Investigación Operativa Aplicadas y Calidad, Universidad Politécnica de Valencia

² serifoglu_f@ibu.edu.tr, Dept. of Management, Abant Izzet Baysal University, Bolu, Turkey

Abstract

Se muestra en este resumen la aplicación de un algoritmo genético a un complejo problema de programación de la producción que incluye, entre otras cosas, varias etapas de producción con máquinas no relacionadas, relaciones de precedencia y tiempos de cambio. Los resultados se comparan con un modelo previo de programación matemática.

Claves: algoritmos genéticos, secuenciación, taller de flujo híbrido.

AMS: 90B35

1. Introducción

La teoría de la secuenciación de la producción está todavía lejos, de manera general, de ser aplicable en la práctica ([1]). En este trabajo intentamos acercar un poco más la teoría a la práctica tratando un complejo problema de programación de la producción. Tenemos un conjunto de N trabajos ($N = 1, \dots, n$) que hay que procesar en un conjunto de M etapas ($M = 1, \dots, m$). En cada etapa tenemos un conjunto $M_i = 1, \dots, m_i$ de máquinas no relacionadas que pueden procesar los trabajos. Los trabajos pueden saltarse las etapas y para cada máquina tenemos un tiempo que indica cuándo está inicialmente disponible (r_{m_i}). En cada etapa y para cada trabajo tenemos un conjunto de máquinas que pueden procesar ese trabajo. Consideramos también tiempos de espera o tiempos de solape entre tareas consecutivas de un trabajo, lo que nos permite modelizar situaciones como secados o enfriados y solapamientos (proceso de lotes de productos). Por último, se consideran también tiempos de cambio de partida dependientes de la secuencia. Como se puede observar, una gran cantidad de problemas existentes en la realidad se asemejan al modelo expuesto. Como objetivo de optimización consideramos la minimización de la fecha máxima de finalización de los trabajos o *Makespan* (C_{max}). Se presentan los resultados de un algoritmo genético en un conjunto de instancias de prueba de pequeño y medio tamaño. Comparamos estos resultados con los de un modelo matemático existente. Todo indica que el algoritmo genético propuesto es capaz de encontrar la mayoría de las soluciones óptimas.

2. Algoritmo genético y Resultados

El algoritmo genético se basa en una permutación para la secuenciación. La población inicial es aleatoria pero factible desde el punto de vista de las relaciones de precedencia. Se usa selección por torneo, cruce de un punto y mutación por inserción. Aplicamos una serie de reglas de asignación de los trabajos en cada etapa y máquina, aprovechando las características del problema. Una de las reglas más eficaces consiste en asignar el trabajo a la máquina que pueda terminar antes el trabajo y permita que la siguiente tarea del mismo trabajo en la siguiente etapa pueda comenzarse lo antes posible. El algoritmo genético se prueba en un PC Pentium IV 3.0 GHz con 512Mbytes de Ram. Comparamos el rendimiento del algoritmo genético, contra un extenso conjunto de 7.680 instancias propuestas para este mismo problema en [2]. Estas instancias se resolvieron con un modelo de programación matemática MIP. Para la mayoría de las instancias se encontró la solución óptima mientras que para otras o bien se encontró una solución no óptima o no se encontró ninguna. En el Cuadro 1 vemos los resultados obtenidos. Como se puede observar, en las

n	m_i	m	2		3	
		1	3	1	3	
5		0,32	3,66	0,44	3,46	
7		0,05	2,78	-1,91	-1,03	
9		-2,31	-2,51	-8,16	-7,89	
11		-5,74	-9,88	-11,67	-13,51	
13		-7,55	-13,94	-16,08	-19,50	

Cuadro 1: Incremento porcentual medio sobre la mejor solución obtenida con el MIP

instancias más pequeñas, la desviación es muy pequeña, mientras que para las grandes de 11 y 13 trabajos, las desviaciones son negativas, lo que indica que el MIP no fue capaz de encontrar la solución óptima en muchos casos (únicamente una buena cota superior) y el algoritmo genético propuesto proporciona mejores resultados.

3. Bibliografía

- [1] McKay, K. N., Pinedo, M., and Webster, S. (2002). Practice-focused research issues for scheduling systems. *Production and Operations Management*, 11(2):249-258.
- [2] Ruiz, R. and Sivrikaya Şerifoğlu, F. (2005). On solving realistic scheduling problems, *7th Workshop on Models and Algorithms for Planning and Scheduling Problems MAPSP 2005*, 235-238, Universidad de Siena, Italia, 6-10 Junio, 2005.

SEDD, un sistema experto para el diagnóstico de desastres

Antonio Omaña Prieto¹, Begoña Vitoriano Villanueva², M. Teresa Ortúñoz Sánchez³

¹antonioomana@yahoo.es, Universidad Pontificia Comillas de Madrid

²bvitoriano@doi.icai.upcomillas.es, Departamento de Organización Industrial,
Universidad Pontificia Comillas de Madrid

³tortuno@mat.ucm.es, Departamento de Estadística e Investigación Operativa
I, Universidad Complutense de Madrid

Abstract

En este trabajo se presenta una herramienta para el análisis y toma de decisiones en el marco de las operaciones de emergencia de envío de ayuda humanitaria en casos de desastres naturales. La herramienta es un sistema experto con razonamiento basado en casos y lógica difusa dada la incertidumbre sobre los datos con la que se trabaja en estas situaciones.

Palabras Clave: Sistemas de Ayuda a la Decisión, Sistemas Expertos, Desastres y Emergencias

1. Introducción

Los desastres que azotan a diversas zonas del planeta sobrepasan en muchas ocasiones la capacidad de respuesta de los gobiernos y organizaciones de la zona, de modo que se requiere la intervención de distintas organizaciones de ayuda humanitaria para dar atención a las necesidades que se producen.

Gracias a las comunicaciones actuales, las noticias de estas tragedias llegan a la comunidad internacional en minutos, y la ayuda se puede movilizar en cuestión de horas. Este caudal de ayuda inmediata puede beneficiar considerablemente a un país devastado por un desastre si se corresponde con necesidades reales. Sin embargo, cuando la ayuda no se ha solicitado, o cuando las instituciones donantes tienen una visión equivocada de las necesidades reales, la ayuda inmediata puede convertirse rápidamente en una carga.

En este trabajo se desarrolla una herramienta cuya principal utilidad es ayudar a diseñar y/o tomar decisiones estratégicas en Operaciones de Ayuda Humanitaria en situaciones de emergencia debida a desastres naturales.

2. SEDD

SEDD es un sistema experto que trata de proporcionar información del sistema estudiado (desastre) imitando distintos tipos de razonamientos lógicos obtenidos del conocimiento de los expertos y de la experiencia. El fin para el que está diseñado es el de ayudar a una organización a estimar el efecto de un desastre en los primeros momentos en que se tiene noticia de que se ha

producido, con el fin de decidir si intervenir en él, y en caso de decidirse por la intervención, proporcionar los datos a tener en cuenta al llevarla a cabo.

Para realizar esta estimación, el sistema utiliza el razonamiento basado en casos. Por lo tanto, la base de conocimiento está formada fundamentalmente por casos. El sistema almacena la información obtenida de casos anteriores en la base de conocimiento, a la espera de que el usuario ejecute la consulta de diagnóstico, en tiempo real.

Las variables de entrada son proporcionadas en el momento del desastre a través de la BH (base de hechos). Mediante el motor de inferencia el sistema dará una información de salida comparando la BH con la base de conocimientos almacenada. Con la información de salida tratará de establecer un diagnóstico de la situación, en términos de “esfuerzo” estimado de realización de una serie de tareas genéricas. Estas tareas tendrán que ser acordes con las normas recogidas por el estándar internacional de la *Carta Humanitaria y Normas mínimas de respuesta humanitaria en casos de desastre* [1], editada por primera vez en el año 1999.

El motor de inferencia busca en la base de conocimiento el caso “más parecido” al que se proporciona, utilizando una cierta medida de distancia. Dado que en estos sistemas la información suele ser muy cambiante y sujeta a incertidumbre, se utiliza la lógica difusa para establecer grados de certeza sobre los datos, definiendo así filtros para los datos con los que se va a trabajar.

La base de conocimiento se amplía con los nuevos casos que se van incorporando, cuando éstos se dan ya por cerrados. Además el sistema incluye una utilidad que permite consultar vía web la base de datos EM-DAT [2] que recoge casos de desastres desde el año 1900, con menor detalle, pero que son una alternativa para casos no documentados en la propia base de conocimiento.

3. Agradecimientos

Este proyecto se ha desarrollado en colaboración con Cruz Roja Española (Dpto. de Cooperación Internacional).

4. Bibliografía

- [1] El Proyecto Esfera (2004). Carta Humanitaria y Normas mínimas de respuesta humanitaria en casos de desastre. <http://www.sphereproject.org>
- [2] EM-DAT <http://www.cred.be> The Centre for Research on the Epidemiology of Disasters, Université Catholique de Louvain, Belgique

Segmentación de imágenes digitales basada en problemas de coloración para grafos valorados

Daniel Gómez¹, Victoria López², Javier Montero², Javier Yáñez²

¹dagomez@estad.ucm.es, Departamento de Estadística e Investigación Operativa III, Universidad Complutense de Madrid

² {vlopez,jayage,monty}@mat.ucm.es, Departamento de Estadística e Investigación Operativa I, Universidad Complutense de Madrid

Abstract

En este trabajo se desarrolla un algoritmo de segmentación para imágenes obtenidas vía satélite. Para llevar a cabo esta segmentación, se modeliza la imagen mediante un grafo planar valorado. Se propone un algoritmo de coloración para grafos valorados que da lugar a una segmentación de la imagen. Este algoritmo de coloración se define por medio de una coloración binaria en grafos no valorados que van siendo determinados en cada iteración.

Claves: Segmentación de Imágenes, Teledetección, coloración de grafos.

AMS: Clasificación AMS (Opcional)

1. Introducción

La segmentación de imágenes digitales está relacionada con la búsqueda de objetos entre los que existen fronteras bien definidas. En estos casos, pueden observarse cambios espectrales relevantes al pasar de un pixel a otro que se encuentra en una zona diferente. El principal problema es que habitualmente, las figuras asociadas a imágenes tomadas vía satélite no muestran un borde muy claro, ni tampoco una forma estándar, sino un suave cambio gradual entre regiones.

Como se muestra en [3, 4] algunos problemas de segmentación y optimización pueden modelarse como problemas de coloración de grafos. En este trabajo se presenta un algoritmo de coloración sobre grafos **valorados** que dará lugar a una segmentación. Esta segmentación, puede considerarse un primer acercamiento a un análisis de clasificación especialmente en zonas de transición entre clases.

El principal argumento, al menos en el tipo de problemas donde se han realizado experiencias (ver [1, 2]), es que las regiones homogéneas tienen un núcleo de píxeles conectados fácilmente detectables, mostrando alrededor cierta degradación suavemente de una región a otra. Por lo tanto, el proceso en cada estado se verá condicionado por una fuerte influencia del comportamiento de los píxeles del entorno.

Para obtener una segmentación de la imagen digital, en primer lugar se modeliza la imagen mediante un grafo valorado planar. En este grafo los nodos representan los pixels de la imagen, mientras que las aristas representan el grado de disimilitud entre pixels adyacentes. Una vez modelizada la imagen vía satélite mediante un grafo valorado, la segmentación (coloración del grafo valorado) se obtiene mediante una sucesiva aplicación de un procedimiento de coloración binario para grafos no valorados que da lugar a una estructura de regiones anidadas.

Finalmente y para representar la segmentación final, cada región obtenida mediante la coloración del grafo se representará mediante un color RGB obtenido como agregación del espectro visible de aquellos pixels que pertenecen a dicha región. En este contexto, regiones diferentes pueden tener colores similares, esto sugeriría que dichas regiones pertenezcan a una misma clase. De todas formas, una comparación detallada entre regiones diferentes puede ser llevada a cabo con el objeto de conseguir una clasificación inicial de la imagen.

Es importante observar, que el mecanismo que aquí se desarrolla se plantea como parte de un proceso más complejo. Este proceso tendrá en cuenta la información dada por la segmentación para mejorar la precisión de algoritmos no supervisados que no tienen en cuenta, entre otras cosas, el entorno de cada pixel.

2. Agradecimientos

Este trabajo se ha realizado bajo la supervisión del proyecto MTM2005-08982-C04-01.

3. Bibliografía

- [1] A. Amo, D. Gómez, J. Montero and G. Biging: *Relevance and redundancy in fuzzy classification systems*. Mathware and Soft Computing 8, 203-216 (2001).
- [2] A. Amo, J. Montero and G. Biging: *Classifying pixels by means of fuzzy relations*. International Journal of General Systems 29, 605–621 (2000).
- [3] Muñoz S., Ortuño T., Ramírez J. and Yáñez J (2005). Coloring fuzzy graphs. *Omega* 33, 211-221.
- [4] Yáñez J. and Ramírez J. (2003). The robust coloring problem. *European Journal of Operational Research* 148, 546–558.

Selección de carteras utilizando funciones borrosas de riesgo lateral

Jose D. Bermúdez¹, José V. Segura², Enriqueta Vercher³

¹bermudez@uv.es, Departamento de Estadística e Investigación Operativa,
Universidad de Valencia

²jvsh@umh.es, Centro de Investigación Operativa, Universidad Miguel
Hernández de Elche

³vercher@uv.es, Departamento de Estadística e Investigación Operativa,
Universidad de Valencia

Abstract

Abordamos la construcción de nuevos modelos de selección de carteras en los que los rendimientos de los activos son considerados números borrosos y se utilizan intervalos de medias para el cálculo del rendimiento y del riesgo de la inversión. Estos modelos borrosos conducen a diferentes problemas de Programación Lineal.

Palabras Clave: rendimientos borrosos, función de riesgo downside, programación matemática fuzzy

AMS: 90C70, 91B28, 90C05

1. Introducción

El problema de la selección de carteras trata de determinar una cartera satisfactoria atendiendo al deseo del inversor, tanto en cuanto al beneficio que se quiere asegurar como al riesgo que está dispuesto a asumir. Supondremos que el inversor desea encontrar las proporciones óptimas del capital que ha de invertir en cada activo buscando un equilibrio entre la maximización del rendimiento y la minimización del riesgo de su inversión. La primera formulación matemática del problema de seleccionar una cartera buscando un cierto equilibrio entre riesgo y beneficio se debe a Markowitz [4].

En este trabajo la incertidumbre sobre el rendimiento de los activos se modeliza mediante números borrosos de tipo *LR* y se utilizan diferentes definiciones de la esperanza de estos números para evaluar el rendimiento y el riesgo esperados para una cartera dada [1, 2]. Además vamos a considerar dos órdenes para los números *LR*: el primero basado en el operador fuzzy-max, que es la extensión a números borrosos del operador máximo entre números reales y el segundo basado en la comparación de intervalos [5].

Para poder plantear adecuadamente el problema de seleccionar una cartera necesitamos determinar de manera explícita la función de riesgo. El riesgo de la inversión es aproximado mediante intervalos de medias que evalúan el riesgo lateral para una cartera borrosa dada y hemos calculado el riesgo borroso de manera explícita para diferentes familias de funciones de referencia y para las

dos definiciones de intervalos de medias. Dichas expresiones del riesgo borroso comparten dos elementos comunes: se trata de un intervalo cuyo extremo inferior es el cero y cuya amplitud es directamente proporcional a la amplitud del núcleo y a la dispersión del rendimiento de la cartera borrosa, lo que refuerza la idea de que esta representación recoge buena parte de la imprecisión que subyace en el problema de selección [6].

En cuanto al procedimiento para determinar una representación escalar del intervalo que modeliza el riesgo borroso esperado hemos seguido la propuesta dada en Leon *et al* [3], que consiste en minimizar el límite superior del intervalo que aproxima el riesgo lateral borroso. Esta escalarización tiene la ventaja adicional de que la función objetivo del problema borroso se convierte en una función lineal con coeficientes precisos, tanto para rendimientos borrosos de la misma forma como para aquellos cuyas funciones de referencia son de la misma familia.

Para números borrosos de tipo *LR* se realiza la comparación de los modelos de selección propuestos, que permiten determinar una cartera mediante la resolución de problemas de optimización lineales.

2. Agradecimientos

Parcialmente subvencionado por el Ministerio de Educación y Ciencia, TIN2005-08404-C04-04.

3. Bibliografía

- [1] Carlsson, C. and Fullér, R. (2001). On possibilistic mean value and variance of fuzzy numbers. *Fuzzy Sets and Systems* 122, 315-326.
- [2] Dubois, D. and Prade, H. (1987). The mean value of a fuzzy number. *Fuzzy Sets and Systems* 24, 279-300.
- [3] León, T., Liern, V., Marco, P., Segura, J.V. and Vercher, E. (2004). A downside risk approach for the portfolio selection problem with fuzzy returns. *Fuzzy Economic Review* 9, 61-77.
- [4] Markowitz, H. (1952). Portfolio Selection. *Journal of Finance* 7, 77-91.
- [5] Sengupta, A. and Pal, T.K. (2000). On comparing interval numbers, *European Journal of Operational Research* 127, 28-43
- [6] Vercher, E., Bermúdez, J.D. and Segura, J.V. (2005). Fuzzy portfolio optimization under downside risk measures. *Fuzzy Sets and Systems* (submitted).

Selección de variables y obtención de coeficientes de la función de discriminación lineal: Uso de Búsqueda Tabú

Silvia Casado¹, Joaquín Pacheco¹, Laura Nuñez²

¹scasado,jpacheco@ubu.es, Departamento de Economía Aplicada, Universidad de Burgos

²laura.nuñez@ie.edu, Departamento de Finanzas, Instituto de Empresa

Abstract

El problema tratado en este trabajo consiste en seleccionar las variables y determinar los coeficientes de las mismas que dan lugar a la mejor función de discriminación lineal, teniendo como objetivo obtener un porcentaje de aciertos elevado en la clasificación.

Palabras Clave: Selección de variables, clasificación, Búsqueda Tabú

1. Introducción

En los problemas de clasificación a través de un conjunto de ejemplos (de los que se conoce su clase) se pretende diseñar y generalizar el conjunto de reglas que permita clasificar, con la mayor precisión posible el conjunto de individuos. Existen varias metodologías para abordar este problema: Análisis discriminante clásico, Regresión Logística, Redes Neuronales, Árboles de decisión, Instance-Based Learning, etc. En la mayor parte de los métodos de discriminación lineal se trata de buscar funciones lineales a partir de las cuales realizar la clasificación (Wald, Fisher, etc). El uso de funciones lineales permite una mejor interpretación de los resultados (por ejemplo importancia y/o significación de cada variable en la clasificación) al analizar el valor de los coeficientes obtenidos.

El problema de determinación de coeficientes de la función discriminante lineal es un problema NP hard. De hecho los métodos clásicos como Regresión Logística Binaria (la función de transferencia logística) y el Análisis Discriminante Clásico (Wald, Fisher, etc) suelen dar lugar a buenas soluciones pero que no son óptimas, como después se va a comprobar

Independientemente del método usado para clasificar cuando se tienen muchas variables antes del diseño de cualquier método de clasificación es necesario seleccionar de las variables originales aquellas que son realmente necesarias. Es decir eliminar del análisis aquellas menos significativas. Este problema es conocido como problema de selección de variables.

Desde un punto de vista computacional la búsqueda del subconjunto de variables es un problema NP-Hard, es decir que no existe ningún método que garantice la solución óptima en un tiempo polinomial en el tamaño de problema,

(NP = *No polinomic*). Para este tipo de problemas se desarrollan dos tipos de algoritmos o métodos: los métodos exactos u óptimos, que encuentran la solución óptima pero que solo son aplicables cuando el tamaño del problema es pequeño; y los métodos aproximados o heurísticos que, aunque no garantizan el óptimo, encuentran buenas soluciones en un tiempo razonable.

En cuanto al problema de selección de variables para su uso posterior en análisis discriminante hay que señalar que en la literatura existe el conocido método *Stepwise* y variantes como [1] así como los métodos *Backward* y *Forward*. Estos son sencillos procedimientos de selección basados en criterios estadísticos (Landa de Wilks, F de Fisher, etc.) y que han sido incorporados en algunos de los paquetes estadísticos más conocidos como SPSS, BMDP, etc. Como destaca [2] estos métodos no son muy eficaces, y cuando las variables originales son muchas raramente alcanzan el óptimo.

El problema tratado en este trabajo consiste en simultáneamente seleccionar las variables y determinar los coeficientes de las mismas que dan lugar a la mejor función de discriminación lineal. Es decir, se trata de un problema a dos niveles: 1) encontrar un subconjunto de variables y 2) determinar los coeficientes de la función de discriminación lineal, en el sentido de obtener un clasificador eficiente. El objetivo consiste en obtener un porcentaje de aciertos elevado en la clasificación o en la determinación de la clase a la que pertenecen una serie de individuos caracterizados por atributos o variables.

En este trabajo, para resolver el problema planteado se diseña un algoritmo basado en la estrategia metaheurística Búsqueda Tabú [3] y se comparan las soluciones que obtiene con, quizás, algunos de los métodos más clásicos de obtención de funciones lineales discriminantes : Análisis Discriminante Clásico y Regresión Logística Binaria.

2. Bibliografía

- [1] O'Gorman, T.W.(2004). Using adaptive Methods to Select Variables in Case-Control Studies. *Biometrical Journal* 46, 5, pp.595-605.
- [2] Schulerud, H. and Albregtsen, F. (2004). Many are called, but Few are Chosen . Feature Selection and Error Estimation in High Dimensional Spaces. *Computers Methods and Programs in Biomedicine* 73, 91-99.
- [3] Glover, F. and Laguna, M.(1997). Tabu Search. *Kluwer Academic Publishers*, Boston.

Sharing costs in a proportionate flow shop situatio.

M. A. Mosquera¹, P. Borm², H. Hamers², A. Estévez-Fernández²

¹mamrguez@usc.es, Department of Statistics and Operations Research,
Santiago de Compostela University

²[P.E.M. Borm, H.J.M. Hamers, A.E. Fernandez]@uvt.nl, CentER and
Department of Econometrics and Operations Research, Tilburg University

Abstract

En una situación *flow shop proporcional* varios trabajos tienen que procesarse en una linea fija de máquinas y cada trabajo tiene el mismo tiempo de procesado en todas las máquinas. Partiendo de un orden inicial dado, al alcanzar un orden óptimo se obtienen unos ahorros. En este trabajo se estudia como repartir estos ahorros.

Keywords: Cooperative games, allocation rules, proportionate flow shop situations.

1. Introduction

In a flow shop situation several jobs have to be processed through a fixed sequence of machines. The current paper considers flow shops in which the processing time of each job is the same in all machines. This kind of flow shop situations is often referred to as *proportionate flow shop* (PFS) (cf. [2]). In [3] it is shown that for any number of machines and for any regular criterion one can restrict the search for an optimal schedule to schedules where the processing order of the jobs is the same in all machines, the so-called *orders*. Moreover, with total weighted completion time criterion this type of problems is solvable in $O(n^2)$ time.

By assuming that each job belongs to one agent or player and introducing cost functions for the players that are dependent on the completion time of their jobs, an associated cooperative game is defined in which the value of a coalition reflects the maximum cost savings this coalition can achieve reordering their jobs from the initial position to a feasible, optimal one. This kind of games are called *PFS games*. We show that PFS games are balanced. Moreover PFS games are convex if the initial order is the urgency order. We also propose and analyze two context-specific cost savings allocation rules. The first rule is based on the algorithm in [3] and on the idea of the equal gain splitting rule for sequencing situations introduced in [1]. It is shown that this rule leads to core elements. For the second rule we find the formula for some of the marginal vectors of the associated PFS game and we define a cost savings allocation rule as the average of these marginal vectors. Obviously, if the initial order is the

urgency order, this rule provides core elements of the associated PFS game. The advantage of these two rules is that we do not need to know the values of the game to compute them.

2. Contact

Address:

Manuel Alfredo Mosquera Rodríguez
 Departamento de Estadística e Investigación Operativa
 Facultad de Matemáticas - Universidade de Santiago de Compostela
 15782 - Santiago de Compostela (A Coruña)
 Spain

Telephone: (+34) 981563100 Ext. 13378

Fax: (+34) 981597054

e-mail: mamrguez@usc.es

3. Acknowledgement

M. A. Mosquera acknowledges the financial support of *Ministerio de Educación y Ciencia*, FEDER and *Xunta de Galicia* (projects BEC2002-04102-C02-02, SEJ2005-07637-C02-02 and PGIDIT03PXIC20701PN)

4. Bibliography

- [1] Curiel I., Potters J., Prasad R., Tijs S. and Veltman B. (1994). Sequencing and cooperation. *Operations Research*, 42:566–568.
- [2] Pinedo, M. (1995). *Scheduling: theory, algorithms and systems*. Englewood Cliffs, Prentice-Hall, New York.
- [3] Shakhlevich N., Hoogeveen H. and Pinedo M. (1998). Minimizing total weighted completion time in a proportionate flow shop. *Journal of Scheduling*, 1:157–168.

Simulación de una subasta en el sector de la construcción

Juan Momparler¹, Mario Hidalgo²

¹momparle@mat.uji.es, Dep. de Matemáticas, Universidad Jaume I

²marioseg29@yahoo.es, Dep. de Estadística, Univ. Tec. Metropolitana (Chile)

Abstract

El objeto de esta ponencia es presentar la simulación de una subasta en sobre cerrado al primer precio en el sector de la construcción. Partiendo del proceso seguido por una constructora, realizaremos una simulación de la subasta con dos métodos diferentes, analizando los resultados obtenidos.

Claves: Equilibrio Bayesiano Nash, Subastas, Simulación, Construcción.

1. Introducción

Es conocido que la mayoría de las empresas constructoras participan en concursos privados y públicos como procedimiento habitual para la adjudicación de obras. La participación en estos concursos con una estrategia óptima adquiere una importancia relevante para estas empresas. Para poder orientar a las empresas en una estrategia adecuada proponemos modelar estas subastas mediante dos técnicas diferentes. En este trabajo supondremos siempre que las empresas utilizan la misma función de puja, es decir, solo se diferenciará su oferta por la valoración del bien, lo que habitualmente se conoce en la literatura como subastas simétricas.

2. Planteamiento del problema

Supongamos que una empresa está interesada en adjudicarse un contrato de obra participando en una subasta organizada por la administración. Para ello estima los costes de la obra y calcula los beneficios que quiere obtener. Posteriormente realiza una estimación, tanto del número de participantes en la subasta, como una valoración de las posibles ofertas que realizarán sus competidores. Con todas estas consideraciones presentará su oferta con el objetivo de adjudicarse el contrato.

3. Simulación de la oferta presentada por la empresa

En primer lugar, analizaremos el método seguido por una empresa del sector de la construcción para calcular su puja, realizaremos una simulación de la subasta, suponiendo que la función de puja ($b_i = s^*(v_i)$) es la misma para todos los concursantes y que la valoración del bien (v_i) del resto de concursantes, se obtiene de una distribución de probabilidad conocida por todos, con función de densidad f y función de distribución F .

4. El modelo de referencia

En segundo lugar utilizaremos el **modelo de referencia**, es decir trataremos la subasta como un **juego estático de información incompleta**, como se puede ver en la literatura. En este modelo se intenta maximizar la utilidad esperada. En el caso de las licitaciones de obras se trata de **subastas a la baja**, si representamos por b_i la oferta de la empresa i se adjudicará el contrato si cumple que:

$$b_i < b_j \quad \forall j = 1, 2, \dots, i-1, i+1, \dots, n$$

La función de utilidad viene dada por:

$$u_i(b_1, \dots, b_n, v_i) = \begin{cases} b_i - v_i & \text{si } b_i < b_j \quad \forall j \neq i \\ 0 & \text{en otro caso} \end{cases}$$

En este caso, la función de puja s^* que maximiza la utilidad de todos los concursantes viene dada por:

$$s^*(v_i) = v_i + \frac{\int_{v_i}^1 [1 - F(w)]^{n-1} dw}{[1 - F(v_i)]^{n-1}}$$

5. Conclusiones

Por último compararemos los resultados obtenidos por los dos métodos y analizaremos las dificultades que se plantean en el caso de utilizar modelos más realistas.

6. Bibliografía

- [1] Krishna V. (2002). *Auction Theory*. Elsevier Science.
- [2] Maskin E., Ripley J. *Asimetric Auctions*. Review of Economic Studies. (2000)67, pp. 413-438
- [3] Menezes Flavio M. and Monteiro Paulo K. (2005). *An Introduction to Auction Theory*. OXFORD UNIVERSITY PRESS.
- [4] Perez J., Jimeno J., Cerdá E. 2004 *Teoría de Juegos*. PEARSON EDUCACIÓN S.A.

Sistema Geo^a/Geo^b/1

Maria Cruz Valsero Blanco, Ana Lorente Marín

mcruz@eio.uva.es

Departamento de Estadística e Investigación Operativa
Facultad de Ciencias, Universidad de Valladolid

Abstract

En este trabajo analizamos un sistema de colas discreto en el que los clientes entran y salen en masa. Modelamos el sistema como markoviano bidimensional. Vemos que el sistema tiene distribución estacionaria si el número de entradas es menor que el de salidas y calculamos la distribución estacionaria.

Palabras Clave: colas discretas

AMS: 60-XX, 90-XX

1. Introducción

Los sistemas de colas discretos actualmente suscitan gran interés. En esta comunicación presentamos el desarrollo de un sistema en el que el tiempo entre llegadas y el tiempo de servicio sigue una ley geométrica de parámetros α y β respectivamente. Tanto las llegadas como los servicios se producen en masa. Cada llegada produce a clientes y el servidor puede servir hasta b clientes.

Estudiamos el proceso X_n que cuenta el número de clientes. Modelamos el proceso como un proceso de Markov bidimensional con conjunto de estados (n, m) , siendo n el número de clientes en el sistema y m el número de clientes en servicio. $n = 0, 1, \dots; m \leq \min(b, n)$.

Para deducir la distribución en el equilibrio, planteamos las ecuaciones en el equilibrio y vemos que existe solución si se verifica $a x \alpha < b x \beta$.

La solución es una mixtura de distribuciones geométricas en función de los parámetros del sistema y de la única raíz en el intervalo $(0, 1)$ de la ecuación característica.

La demostración de que esta raíz es única en el caso $a x \alpha < b x \beta$ no se ha hecho utilizando el teorema de Rouché, si no calculando el punto de cruce de dos funciones convexas y positivas en el intervalo $(0, 1)$.

La solución general a este sistema nos servirá de base para encontrar soluciones a otros sistemas más generales en los que la distribución entre llegadas o del tiempo de servicio sea una binomial negativa.

Dirección:

Departamento de Estadística e Investigación Operativa

Facultad de Ciencias Prado de la Magdalena s/n

47005 – Valladolid Spain

e-mail: mcruz@eio.uva.es

Sistemas de ecuaciones e inecuaciones lineales: Estabilidad y distancia al mal planteamiento

M. J. Cánovas, F.J. Gómez-Senent, J. Parra

canovas@umh.es, paco.gomez@umh.es, parra@umh.es,

Centro de Investigación Operativa. Universidad Miguel Hernández de Elche

Abstract

Consideramos sistemas de restricciones lineales con una cantidad fija, posiblemente infinita, de desigualdades y una cantidad finita fija de igualdades. Proporcionamos una fórmula para determinar la distancia a la inconsistencia (o consistencia) en términos de los coeficientes del sistema, haciendo hincapié en las particularidades derivadas de incluir restricciones de igualdad.

Palabras Clave: Estabilidad, Programación Semi-infinita

AMS: 65F22, 90C34, 90C05, 15A39, 49J53

1. Breve descripción del trabajo

En este trabajo consideramos sistemas en \mathbb{R}^n del tipo

$$\sigma = a_t'x \geq b_t, \quad t \in T; \quad a_s'x = b_s, \quad s \in S,$$

donde $a_t, a_s \in \mathbb{R}^n$, $b_t, b_s \in \mathbb{R}$, los elementos de \mathbb{R}^n son considerados como martices columna (a' representa la traspuesta de $a \in \mathbb{R}^n$), T es un conjunto arbitrario, posiblemente infinito, y S es no vacío, disjunto de T , y tiene cardinal $m \leq n$. Cada sistema puede identificarse con una función (arbitraria) de $T \cup S$ en \mathbb{R}^{n+1} (dotado este último de una norma arbitraria), y la distancia entre sistemas viene dada por la distancia (en la recta real ampliada) del supremo.

El propósito primordial de este trabajo es determinar la distancia de un sistema dado a la frontera del conjunto de los sistemas consistentes (con conjunto factible no vacío). En una primera etapa caracterizamos el interior del conjunto de los sistemas consistentes, así como el de los inconsistentes, y en una segunda etapa calculamos la citada distancia. El problema de hallar la distancia a la mencionada frontera (distancia al mal planteamiento) en el espacio de los sistemas (infinito-dimensional si T es infinito) se traduce en el de hallar la distancia del origen a un subconjunto $G \subset \mathbb{R}^{n+1}$, que viene dado en términos de los coeficientes del sistema. En el trabajo se ponen de manifiesto las particularidades que se derivan de la presencia de igualdades en el modelo, distinguiendo claramente entre aquellas aquellos resultados que se obtienen desdoblando cada igualdad en dos desigualdades y aquellos en los que es preciso tratar las igualdades como tales.

Existen numerosas contribuciones a la teoría de la estabilidad de sistemas desigualdades lineales en diferentes contextos (finitos y semi-infinitos). Varios de los criterios de estabilidad apuntados en estas contribuciones se traducen en que el sistema considerado esté en el interior del conjunto de los consistentes (véase [3]). En [2] se aborda el cálculo de la distancia al mal planteamiento para sistemas de desigualdades lineales. En [1] se estudia la regularidad métrica de multifunciones que describen sistemas semi-infinitos lineales de igualdades y desigualdades, en el caso en el que T es compacto, los coeficientes dependen continuamente respecto del índice y sólo se permiten perturbaciones (continuas) del miembro derecho de las restricciones.

2. Agradecimientos

Este trabajo ha sido parcialmente financiado con los proyectos BFM2002-04114-C02-02 y MTM2005-08572-C03-02, del MEC y FEDER, y GV04B-648 and GRUPOS04/79 de la Generalitat Valenciana. F.J. Gómez-Senent agradece a la Consejería de Educación y Cultura de la Región de Murcia la concesión de una Licencia por Estudios.

3. Bibliografía

- [1] Cánovas, M.J., Dontchev, A.L., López, M.A. and Parra, J. (2005). Metric regularity of Semi-infinite constraint systems. *Mathematical Programming* 104, 329-346.
- [2] Cánovas, M.J., López, M.A., Parra, J. and Toledo, F.J. (2005). Distance to ill-posedness and the consistency value of linear semi-infinite inequality systems. *Mathematical Programming* 103, 95-126.
- [3] Goberna, M.A. and López, M.A. (1998). Linear Semi-Infinite Optimization. John Wiley and Sons. Chichester (UK).

Sistemas de Inventario/Distribución con demanda potencial y rotura

Beatriz Abdul-Jalbar¹, Joaquín Sicilia², Roberto Dorta³, José M. Gutiérrez⁴

¹babdul@ull.es, ²jsicilia@ull.es, ³rodorta@ull.es, ⁴jmgrrrez@ull.es,

Departamento de Estadística, Investigación Operativa y Computación
Universidad de La Laguna

Abstract

En este trabajo se estudia un sistema de Inventario/Distribución formado por un almacén central que suministra los artículos a varios minoristas. Se asume que la demanda de los clientes es conocida y que la distribución a lo largo del tiempo sigue un patrón potencial. Tanto en el almacén como en los minoristas se consideran costes de reposición, de mantenimiento y de rotura. El objetivo consiste en determinar políticas de reposición que minimicen los costes totales del sistema.

Claves: control de inventario, sistemas de inventario/distribución, demanda potencial.

AMS: 90B05, 90B30

1. Introducción

La alta competitividad en el ámbito empresarial y los ajustados márgenes de beneficios en muchos sectores económicos hacen que una adecuada gestión de los recursos disponibles sea esencial para aumentar los beneficios de las empresas. Teniendo en cuenta que los costes de inventario representan normalmente un gasto importante, la reducción de dichos costes es fundamental para conseguir una mejora en los resultados de explotación de las empresas.

En la mayoría de los sistemas de inventario están involucradas varias instalaciones. Por ejemplo, es muy frecuente que en la producción de un artículo éste no se fabrique en una sola localización sino que distintos componentes del artículo se produzcan en diferentes instalaciones para luego montar el producto final en otra instalación. Por otro lado, cuando una compañía distribuye productos en un área geográfica grande, normalmente, hace uso de un sistema de inventario formado por un almacén central, cercano a la fábrica que produce los artículos, y por un número determinado de minoristas cercanos a los clientes. Así, cada vez son más las empresas que dedican esfuerzos a conseguir un buen control de los inventarios a lo largo de todas las instalaciones que forman parte de la cadena de suministro. Es importante observar que las acciones llevadas a cabo por el responsable de un eslabón de la cadena pueden afectar a los

demás componentes de la cadena. Por lo tanto, para lograr un control efectivo de los inventarios es necesario una buena coordinación y cooperación entre las distintas instalaciones del sistema.

En particular, en este trabajo se estudia la situación en la que un almacén suministra los artículos a un conjunto de minoristas, los cuales a su vez satisfacen la demanda de los clientes. El problema 1-almacén N -minoristas, ha sido analizado en la literatura por diferentes autores como, por ejemplo, Roundy (1985), y Abdul-Jalbar et al. (2003, 2004). En esos trabajos la demanda de los clientes es constante y se distribuye uniformemente a lo largo del horizonte temporal considerado. Además, dicha demanda debe ser satisfecha sin caer en rotura. Mitchell (1987) y Chen (1999, 2000) también analizan este problema pero considerando la posibilidad de que se produzcan roturas.

En este trabajo se extiende el análisis al caso en el que la demanda es conocida pero el patrón de demanda es potencial. En este caso, la obtención de la expresión general del coste total del sistema es una tarea bastante compleja. Por ello, analizamos una clase especial de políticas que, debido a sus características, simplifican la modelización del problema. En concreto, el problema se formula haciendo uso de las políticas cíclicas.

2. Bibliografía

- [1] Abdul-Jalbar B., J. Gutierrez, J. Puerto y J. Sicilia (2003). Policies for Inventory/Distribution Systems: The effect of centralization vs. decentralization, *International Journal of Production Economics* 81-82, 281- 293.
- [2] Abdul-Jalbar B., J. Gutierrez, y J. Sicilia (2005). Integer-ratio policies for distribution/inventory systems, *International Journal of Production Economics*, 93-94, 407-415..
- [3] Chen F. (1999). Worst-case analysis of (R,Q) policies in a two-stage serial inventory system with deterministic demand and backlogging, *Operations Research Letters*, 25 (2), 51-58.
- [4] Chen F. (2000). Effectiveness of (R,Q) policies in one-warehouse multi-retailer systems with deterministic demand and backlogging, *Naval Research Logistics*, 47 (5), 422-439.
- [5] Mitchell J. (1987). 98 % Effective lot-sizing for one-warehouse multi-retailer inventory systems with backlogging, *Operations Research* 35, 399-404.
- [6] Roundy R.O. (1985). 98 % Effective integer-ratio lot sizing for one-warehouse multi-retailer systems, *Management Science* 31 (11), 1416-1430.

Sistemas Estratégicos de Rutas de Vehículos para Operaciones de Paz

¹*Eduardo Ramos Méndez,* ²*José A. Carrillo Ruiz*

^{1,2}eramos@ccia.uned.es, Departamento de Estadística, Investigación Operativa y Cálculo Numérico, UNED
²jacarrillo@ccia.uned.es, Ministerio de Defensa

Abstract

El objetivo de este trabajo es estudiar la posibilidad de que los modelos VRP puedan dar respuesta a los nuevos planteamientos surgidos de las necesidades logísticas propias de las misiones de paz y ayuda humanitaria que desarrollan las Fuerzas Armadas.

1. Introducción

Entre las nuevas misiones que las Fuerzas Armadas tienen encomendadas actualmente, resultan ser especialmente relevantes las misiones de paz y de ayuda humanitaria. En este ámbito, las capacidades logísticas, fundamentalmente en lo que se refiere a transporte y capacidad de movimiento, han de dar respuesta a las necesidades que surgen de la exigencia de proyectar fuerzas nacionales a cualquier lugar del mundo en el que estas misiones se desarrolle. El escenario en el que transcurren estas misiones se puede caracterizar, entre otros, por los siguientes aspectos:

- Son operaciones en las que intervienen medios personales y materiales de varios países. Igualmente, puede haber una mayor presencia de instituciones civiles, nacionales y extranjeras. Por lo tanto, es de esperar heterogeneidad en el ambiente en que se desarrollan. Además, la variedad de países que pueden verse involucrados en este tipo de operaciones, propicia la presencia de legislaciones potencialmente diferentes. Este hecho vuelve a incidir en el carácter heterogéneo en el desarrollo de dichas operaciones.
- En general, se puede hablar de un nivel alto de incertidumbre. Las operaciones tienen lugar en territorios que han sufrido conflictos bélicos o bien han sido víctimas de catástrofes naturales. Se trata de circunstancias que favorecen que el devenir de los acontecimientos no sea predecible. A estos factores, hay que unir en algunas ocasiones, la posibilidad de que las fuerzas que realizan estas misiones sean consideradas hostiles, y puedan ser objetivos de ataques.
- La ampliación geográfica del marco de actuación obliga a un alargamiento de las líneas de comunicación. Además, estas vías son muy variables tanto desde el punto de vista de su naturaleza como de su estado.

2. Conclusiones

Los modelos VRP se caracterizan por su capacidad de adaptación a las múltiples situaciones que presentan los problemas que pretende resolver. Por ello resultan especialmente útiles para enfrentarse al escenario descrito en la sección anterior. La aplicación de la metodología VRP a los sistemas

estratégicos de rutas de vehículos para operaciones de paz, permite extraer las siguientes conclusiones:

1. Las necesidades de transporte de las Fuerzas Armadas como consecuencia del cumplimiento de estas nuevas misiones pueden ser modeladas dentro del marco del problema de rutas de vehículos.
2. La adaptabilidad del modelo VRP permite contemplar los siguientes aspectos, todos ellos presentes en los problemas planteados:
 - *Aspectos temporales*: Las ventanas de tiempo son elementos que están presentes en modelos VRP. El alargamiento de las líneas de comunicación, y la considerable duración de estas misiones precisan horizontes de planeamiento muy amplios.
 - *Racionalización en el empleo de vehículos*: La utilización múltiple de vehículos, donde los vehículos representan un bien crítico.
 - *Flotas heterogéneas*: Característica de los modelos VRP ya tratados en la literatura.
3. Los modelos estocásticos dan respuesta rigurosa a los supuestos en los que el grado de incertidumbre es importante.
4. Los modelos VRP también son capaces de proporcionar respuestas adecuadas a los cambios no previstos de situación, considerando las versiones dinámicas de los problemas, sin que sea necesario reconsiderar todo el sistema de transporte y rutas.
5. El modelo combinado inventario-rutas permite optimizar de forma conjunta estos aspectos de la función logística, proporcionando respuestas óptimas a varios problemas de forma simultánea.
6. El modelo combinado localización-rutas permite optimizar decisiones en lo referente al despliegue de unidades, ya que además de aquellas consideraciones de tipo logístico permite modelar restricciones de tipo operativo.

3. Nota

Este trabajo ha sido financiado en parte por el Instituto Universitario General Gutiérrez Mellado de Investigación sobre la Paz, la Seguridad y la Defensa.

4. Bibliografía

- [1] M. Desrochers, Lenstra y M. P. Savelsbergh (1991). A classification scheme for vehicle routing and scheduling problem. *Eur. J. of Oper. Res.* 46, 322-332.
- [2] M. Dorigo y T. Stützle (2004). *Ant colony optimization*. MIT Press.
- [3] Hansen P. y Cheng R. (1999). Metaheuristics: Advances and trends in local search paradigms for optimization. In: Kluwer Academic Publishers. (eds.) *An introduction to variable neighborhood search*.
- [4] Ramos Méndez E. y Carrillo Ruiz J. (2005). *Sistemas estratégicos de rutas de vehículos para operaciones de paz*. Acuerdo de investigación 03/04. Instituto Universitario General Gutiérrez Mellado de Investigación sobre la Paz, la Seguridad y la Defensa

Sobre una clase de problemas inversos en control óptimo estocástico

Ricardo Josa Fombellida¹, Juan Pablo Rincón Zapatero²

¹ricar@eio.uva.es, Departamento de Estadística e Investigación Operativa,
Universidad de Valladolid

²jrincon@eco.uc3m.es, Departamento de Economía, Universidad Carlos III de
Madrid

Abstract

El objetivo del trabajo es identificar una clase interesante de problemas de control óptimo estocástico tales que el control óptimo sea también óptimo para el problema determinista asociado. Para ello utilizaremos el enfoque alternativo de resolución de este tipo de problemas mostrado en Josa–Fombellida y Rincón–Zapatero [2].

Claves: Control estocástico, ecuación de Hamilton–Jacobi–Bellman, problema de inversión y consumo.

AMS: 93E20, 90C39, 60H30, 91B28, 91B70.

1. Introducción

La teoría de control estocástico se aplica a numerosos campos en economía, ingeniería o medio ambiente. Algunos ámbitos de su aplicación son: extracción de recursos, crecimiento económico de sectores productivos, control de la población, inversión–consumo y gestión de planes de pensiones. La metodología usada en la resolución de estos problemas está basada generalmente en la ecuación de Hamilton–Jacobi–Bellman (HJB en lo sucesivo), que caracteriza a la función valor óptimo. Sin embargo, en general esta ecuación en derivadas parciales de segundo orden es no lineal. Este hecho puede dificultar el estudio de las propiedades analíticas y numéricas de sus soluciones.

Recientemente, en Josa–Fombellida y Rincón–Zapatero [2], se ha desarrollado un método alternativo para resolver estos problemas en algunos casos particulares que incluyen numerosos modelos económicos que aparecen en la literatura especializada. Este método considera, en lugar de la ecuación de HJB, un sistema de ecuaciones en derivadas parciales semilineales que caractériza directamente al vector de controles óptimos. Previamente este método fue desarrollado en Bourdache–Siguierdidjane y Fliess [1], en el caso determinista, y en Rincón–Zapatero *et al* [4], en juegos diferenciales.

2. Problema inverso

En este trabajo mostramos una nueva ventaja de este enfoque alternativo frente

a la ecuación de la programación dinámica: la resolución de un nuevo tipo de problema inverso.

Dado un control admisible de un problema de control estocástico, un problema inverso consiste en encontrar las funciones de pago instantánea y final tales que éste sea óptimo del problema de control.

El nuevo tipo de problema inverso que planteamos consiste en identificar las funciones de pago tales que un problema de control estocástico tenga el mismo control óptimo *feedback* que su problema determinista asociado. Mediante el sistema de ecuaciones semilineales mostramos cómo resolver en general este tipo de problema inverso.

3. Una aplicación económica

Una aplicación interesante, que se desarrolla en el trabajo, es la resolución del nuevo problema inverso de maximización de una utilidad del consumo con condiciones muy generales, donde toda la riqueza se invierte en un activo con riesgo. Este problema es un caso particular del modelo más general presentado en Merton [3], donde el agente puede diversificar su inversión, eligiendo el porcentaje de su riqueza que invierte en un conjunto de activos con riesgo, así como en un bono. En este contexto el problema inverso plantea encontrar la función de utilidad tal que un consumo admisible dado sea óptimo de dos problemas simultáneamente, el inicial (problema estocástico donde toda la riqueza se invierte en un activo con riesgo) y el perturbado (problema determinista donde ésta se invierte en un bono con idéntico tanto de retorno que el activo).

4. Bibliografía

- [1] Bourdache–Siguerdidjane, H. and Fliess, M. (1987). Optimal feedback control of nonlinear systems. *Automatica* 23, 365–372.
- [2] Josa–Fombellida, R. and Rincón–Zapatero, J.P. (2006). New approach to stochastic optimal control. Under revision in *Journal of Optimization Theory and Applications*. Presented in the XXVII Congress SEIO, 2003, Lleida.
- [3] Merton, R.C. (1971). Optimum consumption and portfolio rules in a continuous time model. *Journal Economic Theory* 3, 373–413.
- [4] Rincón–Zapatero, J.P., Martínez, J. and Martín–Herrán, G. (1998). New method to characterize subgame perfect Nash equilibria in Differential Games. *Journal of Optimization Theory and Applications* 96, 377–395.

Solving a multiple instance problem via separating concentric hyperspheres

José Gordillo¹, Emilio Carrizosa¹, Frank Plastria²

¹{jgordillo,ecarrizosa}@us.es, Dpto. Estadística e Investigación Operativa,
Universidad de Sevilla

²Frank.Plastria@vub.ac.be , MOSI, Vrije Universiteit Brussel

Abstract

Consideramos un problema de clasificación en el que tenemos que asignar una etiqueta $\{+1, -1\}$ a una serie de objetos formados por varios vectores de características. La regla de clasificación nos lleva a construir dos hiperesferas concéntricas. El problema subyacente es de optimización global entero mixto, y es resuelto heuristicamente con un algoritmo VNS.

Keywords: Classification, Global Optimization, Variable Neighborhood Search

1. Introduction

The multiple instance problem (see e.g. [1, 3]) is a classification problem where the objects to be classified are sets or bags of instances which are vectors measuring d different attributes.

According to these measurements, a label (+1 or -1) is assigned to each bag by means of the so-called MI assumption (see [3]), which states that a bag is positive (label +1) if at least one of its instances satisfies a determined condition and negative (label -1) otherwise.

In our case, the classification rule is defined in terms of a ball of center $x_0 \in \mathcal{R}^d$ and radius $R \in \mathcal{R}_+$ in the following way . Given a bag $X \subset \mathcal{R}^d$,

- classify in G_{+1} , if $\exists x \in X$ such as $\|x - x_0\|^2 < R^2$
 - classify in G_{-1} , otherwise, i.e., $\forall x \in X, \|x - x_0\|^2 \geq R^2$.
- (4.77)

Consider the database $\Omega \subset \mathcal{R}^d$, formed by objects $i = (X_i, Y_i) \in \Omega$, where each X_i is a set of feature vectors, $X_i = \{x_1^i, x_2^i, \dots, x_n^i\}$, with $x_k^i \in \mathcal{R}^d$, and where Y_i is the corresponding class, +1 or -1. If we consider a training sample $I \subset \Omega$, this set will be used to build the optimization problem which must be solved to obtain the optimal parameters $x_0 \in \mathcal{R}^d$ and $R \in \mathcal{R}_+$ for constructing the classification rule (4.77).

This optimization problem can be written as follows,

$$\max_{x_0, R} \min \left\{ \min_{i \in G_{+1}} \max_{x_k^i \in X_i} (R^2 - \|x_k^i - x_0\|^2), \min_{i \in G_{-1}} \min_{x_k^i \in X_i} (\|x_k^i - x_0\|^2 - R^2) \right\}. \quad (4.78)$$

The problem can be transformed into the question in which we have to build two concentric balls $B(x_0, r_{+1})$ and $B(x_0, r_{-1})$, where the first ball must contain at least one instance from each bag of G_{+1} and the second one cannot contain strictly any instance from the bags of G_{-1} .

This latter, this problem may be formulated as a mixed integer global optimization problem, and a heuristic procedure, based on the Variable Neighborhood Search algorithm (see [2] for a description), is proposed to solve it.

2. Acknowledgement

This work has been partially supported by projects BFM2002-04525, Ministerio de Ciencia y Tecnología, Spain, and FQM-329, Junta de Andalucía, Spain

3. Bibliography

- [1] Dietterich T.G., Lathrop R.H., Lozano-Pérez T. (1997). Solving the Multiple Instance Problem with Axis-Parallel Rectangles. *Artificial Intelligence* 89, 31-71.
- [2] Mladenovic N. and Hansen P. (1997). Variable Neighborhood Search. *Computers and Operations Research* 24, 1097-1100.
- [3] Xu X. (2003). *Statistical Learning in Multiple Instances Problems*. Master's thesis, University of Waikato.

Solving a School Bus Routing Problem

Jorge Riera-Ledesma, Juan-José Salazar-González

{jriera,jjsalaza}@ull.es

Departamento de Estadística, Investigación Operativa y Computación,
Universidad de La Laguna, 38271 La Laguna, Spain

Abstract

Este trabajo modeliza un problema de ruta de vehículos que subyace en el contexto del transporte escolar. En él se debe decidir qué parada debe utilizar cada estudiante para tomar el autobús, y qué ruta debe seguir cada vehículo hasta llegar al colegio. Se presentan dos algoritmos óptimos basados en sendos modelos. Asimismo se analiza una experiencia computacional llevada a cabo sobre varias familias de datos.

Keywords: Vehicle Routing Problem, Traveling Purchaser Problem

1. Introduction

Let us consider a set K of m vehicles in a depot 0. Each vehicle $k \in K$ has a capacity q_k . Let U be a set of students. Let I be the set of the potential stops for vehicles and J be the locations of a set of schools. Each student u must go to one school $j(u) \in J$ with one vehicle, and he/she can wait for this vehicle in a stop in a subset $I(u)$. For each school $j \in J$, the set of all its students will be denoted by $U(j)$. Let us consider an undirected graph $G = (V, A)$ where $V := \{0\} \cup I \cup J$ represents a set of all locations. Each arc $a \in A$ is associated with a cost c_a . For $S \subset V$, let us denote $A(S) := \{(i, j) \in A : i, j \in S\}$, $\delta^+(S) := \{(i, j) \in A : i \in S, j \in V \setminus S\}$ and $\delta^-(S) := \{(i, j) \in A : i \in V \setminus S, j \in S\}$. The problem is to decide the stop where each student must wait for a vehicle, and to find m routes such that each route is assigned to each vehicle, starts in 0, passes through some potential stops picking up some students going to the same school, and finishes in this school; each student must be collected by one vehicle; the number of students collected by vehicle k cannot exceed q_k and the sum of the lengths of the paths must be minimum.

2. Mathematical models

Let us consider the following decision variables:

$$x_a^k := \begin{cases} 1 & \text{if vehicle } k \in K \text{ uses the arc } a \\ 0 & \text{otherwise} \end{cases}$$

for all $a \in A$ and all $k \in K$;

$$y_i^k := \begin{cases} 1 & \text{if vehicle } k \in K \text{ visits } i \in I \cup J \\ 0 & \text{otherwise} \end{cases}$$

for all $i \in I \cup J$ and all $k \in K$;

$$z_{ui}^k := \begin{cases} 1 & \text{if student } u \text{ is picked up at } i \text{ with } k \in K \\ 0 & \text{otherwise} \end{cases}$$

for all $k \in K$, all $u \in U$ and all $i \in I(u)$. A 3-index formulation is:

$$\min \sum_{a \in A} c_a \sum_{k \in K} x_a^k$$

$$\begin{aligned}
x^k(\delta^+(0)) &= 1 & \forall k \in K \\
x^k(\delta^+(i)) = x^k(\delta^-(i)) &= y_i^k & \forall k \in K, \forall i \in I \\
x^k(\delta^-(j)) &= y_j^k & \forall k \in K, \forall j \in J \\
x^k(\delta^+(S)) \geq y_i^k & & \forall k \in K, \forall i \in S \subseteq I \\
\sum_{k \in K} \sum_{i \in I(u)} z_{ui}^k &= 1 & \forall u \in U \\
z_{ui}^k \leq y_i^k & & \forall k \in K, \forall u \in U, \forall i \in I(u) \\
\sum_{i \in I(u)} z_{ui}^k \leq y_{j(u)}^k & & \forall k \in K, \forall u \in U \\
\sum_{u \in U} \sum_{i \in I(u)} z_{ui}^k \leq q_k y_j^k & & \forall k \in K, \forall j \in J \\
x_a^k &\in \{0, 1\} & \forall k \in K, \forall a \in A \\
y_i^k &\in \{0, 1\} & \forall k \in K, \forall i \in I \cup J \\
z_{ui}^k &\in \{0, 1\} & \forall k \in K, \forall u \in U, \forall i \in I(u).
\end{aligned}$$

The above model has a polynomial number of variables and an exponential number of constraints. Therefore, it is suitable for a cutting-plane approach. An alternative model based on a polynomial number of constraints and an exponential number of variables is also given. This second model is suitable for a column generation approach. Both approached are analyzed on different families of instances.

3. Acknowledgement

Supported by “Ministerio de Educación y Ciencia” TIC2003-05982-C05-02.

Solving the Ordered Median Spanning Tree Problem with a greedy algorithm

Elena Fernández¹, Justo Puerto², Antonio M. Rodríguez Chía³

¹ Dpt. Estadística i Investigació Operativa, Univ. Politècnica de Catalunya.
e-mail: e.fernandez@upc.edu

² Dept. Estadística e Investigación Operativa, Universidad de Sevilla.
e-mail:puerto@us.es

³ Dpt. Estadística e Investigación Operativa. Universidad de Cádiz.
e-mail:antonio.rodriguezchia@uca.es

Abstract

Consideramos el problema del árbol de expansión de coste mínimo con la función objetivo de mediana ordenada. Utilizamos un modelo de programación lineal entera para formular el problema y estudiamos la descripción del correspondiente poliedro. Demostramos que el problema puede resolverse óptimamente mediante un algoritmo de tipo greedy.

Keywords: Árbol de expansión, Mediana ordenada, algoritmo greedy.

1. Introduction

In this work we study the Ordered Median Spanning Tree Problem (OMSTP). It is a generalization of the well-known Minimum Cost Spanning Tree Problem, where we consider an ordered median objective function. Roughly speaking, with ordered median objective functions, the contribution to the objective function of a variable, say x_i , depends not only on its cost coefficient c_i , but also on the “order” of the value $c_i x_i$ relative to the other cost values $c_j x_j$. That is, there is another vector of weights, λ , so that if $c_i x_i$ is the $k - \text{th}$ ordered cost value, its contribution to the objective function is $\lambda_k c_i x_i$.

Ordered median objective functions have as particular cases classical objective functions like the sum, minmax, cent-dian, etc., and ordered median optimization problems have been studied for continuous and discrete location problems (see, for instance, [1, 2, 3]). To the best of our knowledge, these objective functions have not been considered so far for other combinatorial problems.

2. The Ordered Median Spanning Tree Problem

Let $G = (V, E)$, $|V| = m$ and $|E| = n$, be a graph and a cost function $c : E \rightarrow \mathbf{R}$. Let also $N = \{1, 2, \dots, n\}$ denote the set of indices for the edges. For any subset of edges $F = \{e_j : j \in K_F\} \subseteq E$, $K_F \subseteq N$ let $o_F : \{1, 2, \dots, |F|\} \rightarrow K_F$ be such that $c_{e_{o_F(1)}} \geq c_{e_{o_F(2)}} \geq \dots \geq c_{e_{o_F(r)}}$.

Then, the OMSTP can be formally defined as follows:

$$MSTOC(G, K, d, c) \quad \max \quad \sum_{k \in K_F} \lambda_k c_{e_{o_F(k)}} \quad (4.79)$$

$$\text{s.t.} \quad F = \{e_j : j \in K_F\} \text{ is a tree in } G \quad (4.80)$$

$$c_{e_{o_F(1)}} \geq c_{e_{o_F(2)}} \geq \cdots \geq c_{e_{o_F(r)}} \quad (4.81)$$

We give a linear integer formulation of the problem and we study the structure of the associated polyhedron. Finally prove that the problem can be optimally solved using a greedy algorithm.

3. Acknowledgement

This work has been partially supported through grants MTM2004-0909 and TIC2003-05982-C05/4. These supports are gratefully acknowledged

4. Bibliography

- [1] Boland, N., Dominguez-Marin, P., Nickel, S. and Puerto, J. (2005). Exact Procedures for Solving the Discrete Ordered Median Problem, *Computers and Operations Research*, to appear.
- [2] Nickel, S., and Puerto, J. (2005). *Location Theory. A Unified approach*. Springer.
- [3] Dominguez-Marin, P., Nickel, S., Mladenović, N. and Hansen, P. (2005). Heuristic Procedures for Solving the Discrete Ordered Median Problem, *Annals of Operations Research* 136, 145-173.

Técnicas Bayesianas aplicadas a la fusión de datos en un Robot Modular

A. Moreno Díaz¹, M.I. Rodríguez Galiano¹, M.A. Virtó García²

¹{amoreno, irodriguez}@fi.upm.es, Dpto. de Inteligencia Artificial, Facultad de Informática, Universidad Politécnica de Madrid

²mvirto@fi.upm.es, Área de Ayuda a la Decisión, Ministerio de Defensa

Abstract

La utilización de robots modulares o articulados es necesaria en muchos entornos no estructurados, como son los astilleros navales, entornos de construcción o en operaciones militares. En este trabajo nos centramos en los modelos bayesianos aplicados a la fusión probabilística de la información procedente de varios sensores o de varios módulos.

Claves: Fusión Probabilística, Análisis Bayesiano, Robot Modular.

AMS: 62C10, 93E35, 93C85

1. Introducción

El concepto clásico de robot industrial tiene aplicación cuando la tarea a realizar se desarrolla en un entorno estático o con dinamismo controlado y completamente estructurado. Para entornos altamente dinámicos, poco estructurados y con tareas que rara vez se realizan de la misma forma, es necesario el uso de sistemas robóticos modulares o articulados (serpentoides u otros). Nos estamos refiriendo a entornos tales como astilleros, plantas de fabricación de estructuras singulares o de grandes dimensiones, mantenimiento de estructuras complejas o extensas como puentes, operaciones militares como detección de minas, localización de víctimas en construcciones derruidas, etc. En estos ámbitos se requeriría disponer de un gran número de robots, normalmente muy complejos, cada uno adaptado a una tarea particular. Como esto no resulta viable económicamente, es necesario buscar estructuras que puedan adaptarse de forma simple a la tarea a realizar. Para ello, se busca una arquitectura modular.

La construcción de este tipo de robots requiere la participación de varios equipos de investigación en un entorno multidisciplinar. Dentro de ese entorno nuestra labor consistirá en el estudio de técnicas bayesianas para combinar la información recibida de varios sensores dentro de un mismo módulo o unidad y/o de varios módulos del robot, con el fin de desarrollar un controlador bayesiano que guíe al robot y coordine la acción conjunta de los diversos módulos que lo integran.

2. Objetivos

Estamos interesados en organizar la sensorización de forma que se integren los diversos sensores para la obtención de la información deseada. Algunos trabajos ya incluyen análisis bayesiano para la realización de esta tarea. Por ejemplo, [3] aplica teoría de decisión bayesiana para la detección de puertas. Más recientemente, en [4] se introduce un algoritmo bayesiano descentralizado de decisión que puede ser aplicado a la fusión de información sensorial en redes de sensores. Por otra parte, [2] utiliza un programa basado en inferencia bayesiana para todas las tareas de control de un robot, entre las que se encuentra la fusión de sensores. [1] también se encuadra dentro de la teoría bayesiana en especial para detectar fallos del sistema sensorial del robot a partir de las lecturas de los sensores.

A partir del estudio de estos trabajos y de la realización de una revisión bibliográfica profunda de lo publicado al respecto, veremos cómo ampliar y mejorar estos métodos aplicados a la fusión probabilística de datos, con el fin de reducir y tratar la incertidumbre presente en el problema. Además debemos adaptar estas técnicas para conseguir posicionar correctamente cada módulo del robot y conseguir su coordinación.

Como objetivo final pretendemos desarrollar un controlador bayesiano para la navegación en el entorno considerado, integrando la información previamente obtenida.

3. Bibliografía

- [1] Aznar F., Pujol M. and Rizo R. (2005). Obtaining a Bayesian map for data fusion and failure detection under uncertainty. *Lecture Notes in Artificial Intelligence* 3533, 342-352.
- [2] Lebeltel O., Bessière P., Diard J. and Mazer E. (2004). Bayesian Robot Programming. *Autonomous Robots* 16, 49-79.
- [3] Kristensen S. (1997). Sensor planning with Bayesian Decision Theory. *Robotics ans Autonomous Systems* 19 (3-4), 273-286.
- [4] Makarenko A. and Durrant-Whyte H. (2005). Decentralized Bayesian algorithms for active sensor networks. *Information Fusion*, in press.

Técnicas para la programación de tráfico ferroviario heterogéneo

A. Lova¹, P. Tormos¹, F. Barber², L. Ingolotti², M. Abril², M.A. Salido²

¹{allova,ptormos}@eio.upv.es, Departamento de Estadística e Investigación Operativa A. y C., Universidad Politécnica de Valencia

²{fbarber,lingolotti,mabril,msalido}@dsic.upv.es, Depto. de Sistemas Informáticos y Computación, Universidad Politécnica de Valencia

Abstract

¹ El problema de programación ferroviaria considerado en este trabajo implica la optimización de nuevos trenes en una línea ferroviaria que está ocupada (o no) por otros trenes con horario fijo. Presentamos un procedimiento constructivo de soluciones factibles que está guiado por heurísticas basadas en el objetivo a minimizar: el tiempo de trayecto de cada nuevo tren. La técnica de resolución desarrollada está siendo utilizada por el ADministrador de la Infraestructura Ferroviaria Española (ADIF).

Palabras Clave: programación de horarios ferroviarios, técnicas heurísticas, satisfacción de restricciones

1. Introducción

La optimización del tráfico ferroviario en una línea con tramos de vía única y tramos de vía doble es conocido como un problema NP_Hard con respecto al número de conflictos a resolver en la programación. Por tanto su resolución requiere el uso de técnicas heurística que permitan obtener buenas soluciones en un tiempo computacional aceptable. La asignación eficiente de la infraestructura ferroviaria a la demanda de pasajeros y mercancías en una red ferroviaria es un problema de alto interés práctico y social. El objetivo que se plantea en este trabajo es la generación de horarios factibles de nuevos trenes de diferente tipo (grandes líneas, regionales, cercanías o de mercancías) que tienen que circular sobre una infraestructura que está siendo utilizada (o no) por trenes de diferente tipo con horarios ya fijados. La programación de los nuevos trenes no puede modificar el horario de los ya existentes puesto que se encuentran en circulación (es decir sus horarios son públicos).

2. Descripción del problema. Técnicas de resolución

La técnica desarrollada consiste en un procedimiento iterativo que en cada iteración obtiene una solución factible en el contexto del siguiente escenario: (i) Dos conjuntos de dependencias (estaciones, apeaderos, bifurcaciones...) ordenadas según orden de visita. Una por cada sentido del trayecto ferroviario.

¹Este trabajo ha sido parcialmente subvencionado por los proyectos TIN2004-06354-C02-01 (MEC-FEDER) y 20051019 del Ministerio de Fomento.

- (ii) Un conjunto de trenes en cada dirección indicando el sub-trayecto a recorrer.
- (iii) Tiempo de recorrido de cada tramo del trayecto de cada tren y paradas comerciales en las estaciones.

Para definir el conjunto de restricciones que debe cumplir toda solución al problema se ha trabajado estrechamente con ADIF con el fin de estudiar la problemática real en el sistema ferroviario español. Las restricciones que se tienen en cuenta para la programación de nuevos trenes son las siguientes:

- Intervalo de salida de la estación origen y de llegada a la estación destino.
- Tiempo mínimo de sucesión entre trenes que circulan en el mismo sentido.
- Capacidad de cada estación (número de vías con/sin andén).
- Tiempos de Cierre de cada estación en los que los trenes pueden circular pero no parar o bien no pueden circular.
- Dos trenes circulando en sentido contrario no pueden utilizar simultáneamente un tramo de vía única.
- Trenes de diferente tipo sólo pueden adelantarse en estaciones y no en tramos, aunque éstos sean de vía doble.
- Control de paradas no solicitadas.
- Tiempos de recepción y expedición entre trenes de distinto sentido que coinciden en una estación.

La técnica desarrollada explora un árbol de búsqueda cuyo recorrido está guiado por la función objetivo: minimizar el tiempo de trayecto de cada tren. Como resultado de cada iteración del algoritmo, se obtiene una solución que cumple con todas las restricciones mencionadas. El proceso termina cuando se cumple un criterio de parada dando la mejor solución obtenida hasta el momento.

3. Resultados

La técnica desarrollada permite al usuario obtener una rápida respuesta sobre el horario de un conjunto de nuevos trenes cumpliendo con todas las restricciones del usuario, de la infraestructura de la red ferroviaria y del tráfico existente. La técnica de resolución desarrollada está siendo utilizada por el **ADministrador de la Infraestructura Ferroviaria Española (ADIF)**. De hecho ADIF ha proporcionado un conjunto de problemas reales con los que evaluar la técnica propuesta.

4. Bibliografía

- [1] Lova, A., Tormos, P., Barber, F., Ingolotti, L., Salido, M.A., Abril, M. Intelligent Train Scheduling on a High-Loaded Railway Network. ATMOS. LNCS (Springer Verlag). To appear. 2006

Teoría de Juegos y Cohesividad en Redes de Comunicación Probabilísticas

Daniel Gómez¹, Enrique González Arangüena¹, Conrado Manuel¹,
Guillermo Owen², Mónica del Pozo¹

¹dagomez,egaran,conrado,mpozo@estad.ucm.es, Departamento de Estadística e Investigación Operativa III, Universidad Complutense de Madrid

²Gowen@nps.edu, Naval Posgraduate School, Monterrey, CA.

Abstract

En esta comunicación se introduce una medida, basada en conceptos juego-teóricos, de la cohesividad de un conjunto de individuos cuyas relaciones vienen dadas por una red social o económica, que se representará mediante un grafo. Se obtienen propiedades de esta medida que los analistas de redes sociales consideran relevantes.

Palabras Clave: juegos cooperativos, cohesividad, valor de Myerson

1. Introducción

Uno de los problemas mas relevantes en el análisis de redes sociales es la identificación de *subgrupos cohesivos* dentro de una red. Un grupo se considera cohesivo en la medida en que sus miembros poseen conexiones con otros miembros del grupo que contribuyen a mantenerlo unido. Así, un subgrupo cohesivo en una red social en un subconjunto de individuos entre los que existen conexiones relativamente fuertes, directas o frecuentes ver [3,4] para mas detalles.

Para representar las relaciones entre los individuos así como los intereses que motivan sus interacciones, utilizaremos el concepto juego-teórico de situación de comunicación [2]. Una situación de comunicación es una terna (N, v, Γ) , en la que (N, v) es un juego n-personal en forma de función característica y (N, Γ) un grafo no dirigido.

Si se asume que las relaciones entre los miembros del grupo no son deterministas, se tiene una situación de comunicación probabilística (N, v, Γ, p) , donde p es una distribución de probabilidad, en principio arbitraria, sobre el conjunto de los subgrafos de Γ .

En este contexto, definimos la cohesividad de un subgrupo de individuos A de N como la proporción del valor de A que se mantiene al pasar de la situación de comunicación determinista a la probabilística, es decir, si se considera que las relaciones no son seguras y pueden romperse.

En este trabajo se generalizan los resultados obtenidos en [1]. En particular, se prueba que la inclusión de una nueva conexión entre los individuos del grupo

incrementa la cohesividad del mismo, independientemente de la distribución de probabilidad subyacente (siempre que la nueva distribución sea consistente con la anterior). Esta propiedad es considerada por los analistas de redes sociales como imprescindible para cualquier medida de cohesividad.

2. Bibliografía

- [1] Gomez D., González-Arangüena E., Manuel C., Owen G., Pozo M., Saboyá M. (2006). *Annals of Operation Research. (Pendiente de publicación)* (2006)
- [2] Slikker M., and Nouweland A. (2001). Social and economic networks in cooperative game theory. *Kluwer Academic Publishers, Norwell, MA.*
- [3] White D.R. and Harary F. (2001). The cohesiveness of blocks in Social Networks. Node connectivity and conditional density'. *Social Methodology* 31, 305-359.
- [4] Wasserman, S. and Faust, K. (1994). Social Networks Analysis. *Cambridge University Press.*

The Harsanyi paradox and the 'right to talk' in bargaining among coalitions

Juan J. Vidal-Puga¹

¹vidalpuga@uvigo.es, Statistics and Operations Research Department,
University of Vigo

Abstract

I describe a new coalitional value from a non-cooperative point of view, assuming coalitions are formed for the purpose of bargaining. The idea is that all the players have the same chances to make proposals. This means that players maintain their own "right to talk" when joining a coalition.

Keywords: weighted Shapley value, non-cooperative bargaining, coalition structures, Harsanyi paradox

AMS: 91A10

1. Introduction

Many real-life situations can be modelled as a set of agents or players with independent interests who may benefit from cooperation. Moreover, it is not infrequent that these agents have partitioned themselves into coalitions (such as unions, cartels, or syndicates) for the purpose of bargaining.

Assuming that cooperation is carried out, in this paper I address the question of how to share the benefit between the coalitions and between the members inside each coalition, i.e. which "value" best represents the expectation of each individual.

I describe a new coalitional value from a non-cooperative point of view, assuming coalitions are formed for the purpose of bargaining. The idea is that all the players have the same chances to make proposals. This means that players maintain their own "right to talk" when joining a coalition. The resulting value coincides with the weighted Shapley value in the game between coalitions, with weights given by the size of the coalitions. I apply this value to an intriguing example presented by Krasa, Temimi and Yannelis (2003) and show that the Harsanyi paradox (forming a coalition may be disadvantageous) disappears. These results throw certain doubts on the reasonability of the Carrier axiom as presented by Hart and Kurz (1983).

The latest version of this paper can be found at <http://webs.uvigo.es/vidalpuga/>

2. Contact

Address:

Juan José Vidal Puga
Facultad de Ciencias Sociales - Campus A Xunqueira s/n
36005 - Pontevedra (Spain)

Telephone: (+34) 986 802014 (+34)986 813511

Fax: (+34) 986812401

e-mail: vidalpuga@uvigo.es

www: <http://webs.uvigo.es/vidalpuga/>

3. Acknowledgement

Financial support by the Spanish Ministerio de Ciencia y Tecnología and FEDER through grant SEJ2005-07637-C02-01/ECON is gratefully acknowledged.

4. Bibliography

- [1] Hart S. and Kurz M. (1983) Endogenous formation of coalitions. *Econometrica* 51(4), 1047-1064.
- [2] Krasa, S., Temimi A. and Yannelis N.C. (2003). Coalition structure values in differential information economies: is unity a strength? *Journal of Mathematical Economics* 39, 51-62.

The use of relative priorities in optimizing the performance of a queueing system

Justo Puerto¹, Francisco R. Fernández², Rafael Hassin³

^{1,2}Department of Statistics and Operations Research, University of Seville.
41012 Sevilla, Spain.

³Department of Statistics and Operations Research, Tel Aviv University, Tel Aviv 69987, Israel.

Abstract

Relative priorities can be used in a n -class queueing system to reduce server and customer costs. This property is demonstrated in a model where the goal is to minimize a non-linear cost function of the expected waiting times. Special attention is given to the case where the expected waiting time of each class is restricted and the goal is to minimize the server's costs.

Keywords: Queues, Dynamic processor sharing.

1. Introduction

Control of queueing systems to maximize profits or welfare has been the subject of numerous papers. The common methods are by setting prices or by setting an adequate priority regime. (See [2] for several models in this context.) An alternative approach is that of *relative priorities* where the priority given to a class also depends on state variables associated with other classes. In this paper we demonstrate several new possible uses of such regimes.

We consider a single server and several customers. Each of the customers submits jobs to be processed by the server. The demand generated by customer i is Poisson with rate λ_i and the service rate is exponential with mean $1/\mu$. A function $f(W_1, \dots, W_n)$ gives the cost incurred by the system when i -jobs have expected waiting time of W_i , $i = 1, \dots, n$. We also consider a variation of this model where the service rate is a decision variable variable, and the cost function is extended to include the cost associated with the chosen service rate. In both cases we give conditions under which relative priorities reduce costs.

We elaborate on a special case of the above model, where customer i requires that the expected time his jobs stay in the system (waiting and being processed) is bounded by a constant t_i . The server is free to choose the service rate m and a priority rule. The server incurs a cost $C(m)$ per unit of time if the chosen service rate is m . The function C is monotone non-decreasing. We investigate the optimal choices to be made by the server. In particular, we show that the server can profit by using a regime of relative priorities.

We consider the priority scheme called *discriminatory processor sharing* (DPS). Under this model there exist nonnegative parameters $x_i \in [0, 1]$, $\sum_{i=1}^n x_i = 1$ representing *relative priority* of customers of the classes. If n_i customers are

present in the system, $i = 1, \dots, n$, an i -customer receives $\frac{x_i}{\sum_{i=1}^n n_i x_i}$ of the service capacity of the server. In particular, the total capacity dedicated to class- i is $\frac{n_i x_i}{\sum_{i=1}^n n_i x_i}$. Of course, $x_i = 1$ means that the class i obtains absolute priority.

The DPS discipline is used in several queueing models in the computer science and communication literature, see [1, 2, 5, 7]. Clearly, DPS gives more options than can be achieved by absolute priorities, and one may claim that it is expected that by applying DPS a server should be able to achieve better performance or profit than otherwise. However, at least in one notable case this assumption turns to be false. Hassin and Haviv [4] considered two customer classes and a single server who sets both prices and relative priority. They observed that it follows from Mendelson and Whang [6] that when the server is not restricted in choosing these variables, there exists an optimal solution with absolute priorities and thus the application of DPS doesn't improve the welfare achieved by the system. However, they also showed that if the server is restricted to a given set of prices, or if the server must set a common price to both classes, then relative priorities may be used to increase profits. Thus, it is a question of interest to identify other settings where the use of relative priorities can be helpful.

The paper is organized as follows. In Section 2 we analyze how to reduce system costs by using DPS as opposed to the use of absolute priorities. Specifically, we give conditions that ensure, for a given cost function, when DPS outperforms FCFS. Section 3 considers a different model where each class fixes its aspiration level on the waiting time and the problem is to ensure those levels at a minimum service rate.

2. Bibliografía

- [1] Altman, E., T. Jimenez and D. Kofman (2004), "DPS queues with stationary ergodic service times and performance of TCD in overloads," in *Proceedings of INFOCOM*, Hong-Kong, 2004.
- [2] Boland, T. and L. Massoulie (2001), "Impact of fairness in Internet performance," in *Sigmetrics*, Cambridge, MA.
- [3] Hassin, R. and M. Haviv (2003), *To Queue or not to Queue: Equilibrium Behavior in Queueing Systems*, Kluwer International Series.
- [4] Hassin, R. and M. Haviv (2006), "Who should be given priority," *Operations Research Letters*, **34** 191-198.
- [5] Kleinrock, L. (1976), *Queueing Systems, Vol. 2: Computer Applications*
- [6] Mendelson, H. and S. Whang (1990), "Optimal incentive-compatible priority pricing for the $M/M/1$ queue," *Operations Research* **38** 870-883.
- [7] Rege, K.M. (1994), "A decomposition theorem and related results for the discriminatory processor sharing queue," *Queueing Systems* **18** 333-351.

The virtually subgame perfect equilibrium for extensive games

Ignacio García-Jurado¹, Julio González-Díaz²

¹igjurado@usc.es, Departamento de Estadística e Investigación Operativa,
Universidad de Santiago de Compostela

²julkin@usc.es, Departamento de Economía e Historia Económica,
Universidad Autónoma de Barcelona

Abstract

En este trabajo analizamos el concepto de *virtually subgame perfect equilibrium* en el contexto de los juegos en forma extensiva. En particular, presentamos algunos resultados que estudian su existencia y lo relacionan con el equilibrio perfecto en subjuegos y con otros refinamientos del equilibrio de Nash.

Keywords: Extensive games, subgame perfect equilibrium

AMS: 91A18

Subgame perfect equilibrium, shortly SPE, is probably the most important equilibrium concept for extensive games. It was introduced in Selten (1965). Its main target is to disregard those Nash equilibria which are only possible if some players give credit to irrational plans of others. More formally, a SPE is a Nash equilibrium which, moreover, induces a Nash equilibrium in every subgame. Van Damme (1991) provides a deep survey on equilibria in extensive games.

Here we deal with a new equilibrium concept for extensive games: the *virtually subgame perfect equilibrium*, shortly VSPE, which was introduced in García-Jurado and González-Díaz (2006). This concept has the same effect as subgame perfection, but it only concentrates on those subgames which are relevant for a given strategy profile; relevant in the sense that they are reachable if exactly one player deviates from the strategy profile in any subgame which has already been classified as relevant. Despite of being based on the same idea, SPE and VSPE are different concepts, the latter existing in many games which do not have SPE. There are many extensive games without SPE, but still, they can have sensible equilibria. This is the case when the non-existence of SPE is because some subgames which are irrelevant for a certain strategy profile do not have Nash equilibria. VSPE solves this problem; it is especially useful when dealing with extensive games having large trees.

In this paper we show some interesting properties of VSPE, and we provide some results on its existence, on its relation with SPE, and on its connections with other refinements.

1. Acknowledgement

This research was partially supported by the Spanish Ministry for Education and Science and FEDER through project SEJ2005-07637-C02-02, and by Xunta de Galicia through project PGIDT03PXIC20701PN.

2. Bibliography

- [1] García-Jurado, I. and González-Díaz, J. (2006). The Role of Commitment in Repeated Games. To appear in *Optimization*.
- [2] Selten, R. (1965). Spieltheoretische Behandlung eines Oligopolmodells mit Nachfrageträgheit. *Zeitschrift für die gesamte Staatswissenschaft* 12, 301-324.
- [3] Van Damme, E. (1991). *Stability and Perfection of Nash Equilibria*. Springer-Verlag.

Ubicación de productos en estanterías de grandes superficies: una revisión bibliográfica

J. M. Albarracín¹, M. Palmer¹, J. P. García¹, M. Babiloni¹, A. Esparcia²

¹jmalbarr@omp.upv.es, Departamento de Organización de Empresas,
Universidad Politécnica de Valencia

²anna@iti.upv.es, Instituto Tecnológico de Informática, Universidad
Politécnica de Valencia

Abstract

Ubicar correctamente productos en estanterías de grandes superficies es de vital importancia para el adecuado funcionamiento en pro de la consecución de los objetivos definidos por el detallista. En el presente trabajo se analizan las diferentes soluciones planteadas por diversos autores identificando objetivos, restricciones y soluciones comentadas.

Palabras Clave: programación matemática, asignación, logística

1. Introducción

La asignación de espacio, recurso escaso y finito, entre los distintos productos y referencias que conforman el surtido de una superficie comercial que, por otra parte, son cada vez más numerosos, es hoy día, una de las preocupaciones esenciales de la dirección detallista. El desarrollo de nuevos tipos de establecimientos, caracterizados por un mayor tamaño, no ha permitido solventar el problema, ya que, paralelamente, se ha producido un fuerte incremento en el número de productos ofrecidos en las tiendas, que supera el mencionado aumento de superficie en los establecimientos.

La importancia de tal decisión tiene que ver de su relación directa con la rentabilidad detallista, en cuanto que afecta tanto a los costos como a los ingresos. La razón fundamental se deriva de la ausencia de personal de ventas que caracteriza al mismo y que exige al producto auto argumentarse. Para ello, además de otras consideraciones, es necesaria una presencia suficiente del producto, que permita al cliente su percepción, ya que de otro modo, si éste no ve un artículo no podrá comprarlo.

2. Modelos clásicos

Los modelos más habituales utilizados en el sector han sido:

- **Planificación del stock:** el reparto de espacio se basa en el plan de stock desarrollado para cada línea de productos. Dicho plan consiste en determinar el surtido ideal, considerando los distintos estilos, tamaños, marcas, variedades y colores, entre otras variables, de acuerdo con las preferencias y necesidades del grupo objetivo de clientes.

- **Método del ratio de productividad del espacio:** este método se fundamenta en asignar espacio a las distintas categorías de productos que conforman cada sección o departamento, de acuerdo con el nivel de ventas o beneficio que generen por unidad de espacio.

3. Modelos informáticos comerciales

Los programas informáticos de gestión de espacio proporcionan un soporte en la toma de decisiones sobre la disposición de los productos en el mobiliario de presentación y el espacio a asignar a cada uno de ellos. Estos programas generan detallados informes que analizan y comparan el comportamiento de los productos y categorías de productos.

En el mercado existen varios programas informáticos de gestión de espacio, si bien los más conocidos y analizados son Spaceman, Apollo y Gold'n Vision.

4. Modelos de optimización

Los modelos matemáticos de optimización buscan encontrar la solución idónea introduciendo en el modelo todas las variables y objetivos que el decisor quiere tener en cuenta a la hora de repartir el espacio disponible.

Estos modelos, a diferencia de los diseños que pretenden medir el impacto sobre el cliente de la relación espacio asignado-ventas producidas, tienen en cuenta otros factores, además de las ventas, en las consideraciones sobre asignación de espacio. Se analizan y comentan los modelos de: Anderson y Amato, Hansen y Heinsbroek, Corstjens y Doyle, Zufryden, Yang y Chen, Lim, Quian y Rodrigues

5. Agradecimientos

El trabajo se enmarca en las investigaciones realizadas por el Departamento de Organización de Empresas y el Instituto Tecnológico de Informática de la Universidad Politécnica de Valencia en el proyecto “Aplicación de Algoritmos Naturales a problemas de logística” subvencionado por el Instituto para la Pequeña y Mediana Empresa de la Generalitat Valenciana.

6. Bibliografía

- [1] Anderson, E.E., Amato, H.N. (1974): “A mathematical model for simultaneously determining the optimal brand collection and display area allocation”, *Operations Research*, vol. 22, january-february, pp. 13-21
- [2] Yang, M., Chen, W. (1999), “A study of shelf space allocation and management”. *International Journal of Production Economics*, 60-61, pp. 309-317.
- [3] Lim, A., Rodrigues, B., Zhang, X. (2004), “Metaheuristics with local search techniques for retail shelf-space optimization” *Management Science*, vol. 50, n° 1, pp. 117-131.

Un algoritmo basado en la entropía para la construcción de redes Bayesanas

Ana María Martínez Rodríguez¹, Luis G. Vargas², Luis Parras Guijosa³

¹ammartin@ujaen.es, Departamento de Estadística, Investigación Operativa,
Universidad de Jaén

²Katz Graduate School of Business, University of Pittsburgh

³Departamento de Estadística, Investigación Operativa, Universidad de Jaén

Claves: Redes Bayesianas, distancia de Kullback-Leibler.

Abstract

Para que la información que está almacenada en las bases de datos sea útil a los investigadores, esta ha de ser interpretada. Una red Bayesiana es un modelo probabilístico de las relaciones entre variables. En este trabajo se propone un nuevo algoritmo para la construcción de redes Bayesianas a partir de los datos almacenados en grandes bases de datos.

1. Introducción

Desde un punto de vista no formal una Red Bayesiana constituye la representación gráfica de una distribución de probabilidad n-dimensional mediante un grafo dirigido acíclico, en el cual cada nodo representa una variable de interés mientras que los arcos representan las probabilidades condicionales de relaciones de dependencia entre las variables.

El problema de encontrar la estructura de red Bayesiana que mejor se ajuste a los datos se conoce con el nombre de *aprendizaje estructural* y ha sido tratado ampliamente por los investigadores en los últimos años [2]. En general en la literatura aparecen dos metodologías para afrontar el problema del aprendizaje estructural. Por una parte están los métodos de métrica más búsqueda que consisten en definir una métrica de la bondad de una estructura de red dados los datos y un algoritmo de búsqueda (heurístico) que trata de optimizar dicha métrica [3]. Por otra parte están los métodos de aprendizaje restringido que tratan de recuperar la estructura de red Bayesiana que mejor representen las relaciones de (in)dependencia condicional que hay entre las variables [5].

En este trabajo proponemos un nuevo método de medida más búsqueda para el aprendizaje estructural. Para ello definiremos una nueva medida de calidad a partir de la *distancia de Kullback-Leibler* [4] y a continuación se propondrá un algoritmo para optimizar dicha medida.

2. Medida de calidad

Dado un grafo $G = (V, \Gamma)$ consideramos los elementos de la matriz de

adyacencia. A partir de los ellos obtenemos una expresión de la distribución estimada a partir de los datos $P_a(V_1, \dots, V_n)$ que, junto a la definición de la *distancia de Kullback-Leibler* [4] permite definir la siguiente medida asociada a una estructura B :

$$Q(B) = \sum_{k=1}^n \left(\sum_{v_k, \pi_k} p(v_k, \pi_k) \log p(v_k | \pi_k) \right) \quad (4.82)$$

donde π_k es una realización de los padres de k . El objetivo es maximizar de dicha medida, esto es, obtener la estructura de red que mejor valor proporcione de $Q(\cdot)$, para ello, puesto que todos los sumandos $Q(B)$ tienen el mismo signo, vamos a buscar el máximo de cada uno de esos sumandos.

3. Algoritmo de búsqueda

Se presenta un nuevo algoritmo para optimizar la medida (4.82). Para ello vamos a proponer un algoritmo que calcule el máximo de cada uno de dichos sumandos. El algoritmo partirá del grafo vacío y posteriormente irá determinando aquellos nodos que serán padres de V_k .

4. Aplicación

Aplicamos el algoritmo a una base de datos de la red ALARM [1].

5. Bibliografía

- [1] Beinlich, I. A., Suermondt, H. J, Chavez, R. M. & Cooper, G. F. (1989). *The ALARM monitoring system: A case study with two probabilistic inference techniques for belief networks*. Proceeding of the second European Conference on Artificial Intelligence in Medicine, 247-256.
- [2] Buntine, W. (1996). *A guide to the literature on learning probabilistic networks from data* IEEE Transactions on Knowledge and data Engineering, 8, 195-210.
- [3] Cooper, G.F. & Herskovits, E. (1992). A Bayesian method for the induction of probabilistic networks from data. *Machine Learning* 9, 309-347.
- [4] Kullback, S. (1968). *Information Theory and Statistics*. Dover Publication Inc.
- [5] Geiger, D., Paz, A. & Pearl, J. (1990). *Learning causal trees from dependence information*, Technical Report R-149, Computer Science Department, University of California.

Un algoritmo de Generación de Columnas para Máquinas de Vector de Apoyo

Emilio Carrizosa¹, Belén Martín-Barragán², Dolores Romero Morales³

¹ecarrizosa@us.es, Universidad de Sevilla (Spain)

²belmart@us.es, Universidad de Sevilla (Spain)

³dolores.romero-morales@sbs.ox.ac.uk, University of Oxford (U.K.)

Abstract

En este trabajo proponemos un modelo que, basado en Máquinas de Vector de Apoyo selecciona automáticamente qué variables juegan un papel importante en la clasificación, y qué valores de dichas variables son críticos.

Claves: Máquinas de Vector de Apoyo, Generación de Columnas, Minería de datos.

1. Introduction

En este trabajo, nos centramos en el problema de la Clasificación Supervisada, donde tenemos un conjunto de objetos Ω y el objetivo es construir una regla de clasificación que prediga la clase c^u a la que pertenece un objeto u por medio del vector predictor x^u . El vector predictor x^u toma valores en un conjunto X , normalmente un subconjunto de \mathbb{R}^p , como por ejemplo $\{0, 1\}^p$. La componente del objeto u , llamada la clase, c^u , toma valores en un conjunto predefinido de posibles clases \mathcal{C} . Para la construcción de la regla de clasificación disponemos de un conjunto de objetos $I \subset \Omega$, llamado muestra de aprendizaje, en cuyos objetos, tanto el vector predictor x^u , como la clase c^u , son conocidos. Nos restringimos al caso en el que hay dos clases $\mathcal{C} = \{-1, 1\}$.

La herramienta conocida como Máquinas de Vector de Apoyo, del inglés Support Vector Machines (SVM), ha demostrado dar buenos resultados en problemas de Clasificación Supervisada. Sin embargo, otros métodos como los Árboles de Clasificación, se han vuelto más populares en las áreas de aplicación, debido a que las reglas de clasificación obtenidas son fácilmente interpretables.

El modelo de propuesto en este trabajo, usa una regla de clasificación basada en preguntas del tipo:

$$\boxed{\text{¿es la variable predictora } \ell \text{ mayor o igual que } b?} \quad (4.83)$$

donde las variables ℓ y los valores de b usados son elegidos automáticamente, siguiendo las ideas de las Máquinas de Vector de Apoyo[1, 2], mediante la maximización del margen.

Consideramos teóricamente todas las posibles preguntas del tipo (4.83), formalizadas matemáticamente por la función:

$$\phi_{\ell b}(x) = \begin{cases} 1 & \text{si } x_\ell \geq b \\ 0 & \text{en otro caso.} \end{cases} \quad (4.84)$$

para $b \in B_\ell = \{x_\ell^u : u \in I\}$ y $\ell = 1, 2, \dots, p$. En lo que sigue, a cada función del tipo $\phi_{\ell b}$ la llamaremos *característica*. Dichas características se usan para clasificar de la siguiente manera: cada característica $\phi_{\ell b}$ tiene asociado un peso $\omega_{\ell b}$, que mide su contribución a la clasificación en la clase -1 ó 1 . La suma ponderada de las características y un umbral β constituyen la *función de puntuación*,

$$f(x) = \omega^\top \Phi(x) + \beta = \sum_{\ell=1}^p \sum_{\{b \in B_\ell | x_\ell \geq b\}} \omega_{\ell b} + \beta, \quad (4.85)$$

donde $\Phi(x) = (\phi_{\ell b}(x))_{b \in B_\ell, \ell=1,2,\dots,p}$ y $\omega^\top \Phi(x)$ denota el producto escalar de los vectores ω y $\Phi(x)$. Un objeto se clasificará en la clase -1 si $f(x) < 0$, y en la clase 1 si $f(x) \geq 0$.

Para elegir los coeficientes $\omega_{\ell b}$ y el umbral β seguimos el enfoque del SVM, basado en encontrar el hiperplano que maximiza el margen en el espacio definido por las características. En este trabajo, dicho problema se formula como un problema de Programación Lineal donde el número de variables de decisión es muy alto, por lo que se propone un algoritmo de generación de columnas para resolverlo.

Los resultados empíricos muestran que el método propuesto se comporta, en términos de clasificación correcta, de manera comparable al SVM clásico y claramente mejor que los árboles de clasificación.

2. Agradecimientos

Este trabajo está financiado parcialmente por los proyectos MTM2005-09362-C03-01 del Ministerio de Ciencia y Tecnología, España, y FQM-329 del Plan Andaluz de Investigación, Andalucía (España).

3. Bibliografía

- [1] J. Shawe-Taylor, P.L. Bartlett, R.C. Williamson, and M. Anthony. Structural risk minimization over data-dependent hierarchies. *IEEE Transactions on Information Theory*, 44(5):1926–1940, 1998.
- [2] V. Vapnik. *Statistical Learning Theory*. Wiley, 1998.

Un algoritmo descendente local para generar la región eficiente de un problema de programación lineal multiobjetivo

Jesús M. Jorge Santiso

jjorge@ull.es, Departamento de Estadística, Investigación Operativa y Computación, Universidad de La Laguna

Abstract

En este trabajo proponemos un nuevo método para generar el conjunto de soluciones eficientes de un problema de programación lineal multiobjetivo. El algoritmo que presentamos incorpora tests de eficiencia apropiados para determinar, para cada vértice eficiente, todas las caras eficientes maximales incidentes en el mismo, empleando para ello una búsqueda descendente (en el sentido de la inclusión) de las caras exploradas.

Palabras Clave: Programación lineal multiobjetivo, región eficiente, caras eficientes maximales

AMS: 90C29, 90C99

1. Introducción

La programación lineal multiobjetivo (MOLP) se encarga del estudio y resolución de aquellos problemas de decisión multicriterio en los que tanto los objetivos como las restricciones del modelo son lineales.

Como es bien sabido, la región eficiente de un MOLP es un conjunto no convexo, aunque si conexo, que puede ser representado como la unión de todas las caras eficientes maximales del problema ([9]). Dada la dificultad, tanto teórica como computacional, que comporta la generación del conjunto de soluciones eficientes de un MOLP, a lo largo del tiempo se han propuesto en la literatura diferentes estrategias y numerosos algoritmos.

Atendiendo a si la exploración de las caras de la región factible se realiza de menor a mayor dimensión o viceversa, podemos clasificar a tales algoritmos en *ascendentes* o *descendentes*, respectivamente. Por otra parte, si las búsquedas de caras eficientes se practican a medida que se detecta cada vértice eficiente o no, hablamos de técnicas *locales* o *globales*. Esto nos da una clasificación de métodos en cuatro categorías con propiedades bien diferenciadas ([7]). Entre los ascendentes locales se encuentran los procedimientos de Armand ([1]), Ecker-Hegner-Kouada ([3]), Gal ([4]) o Isermann ([5]). De tipo ascendente global destacamos el de Armand-Malivert ([2]) y, como ejemplos de procedimientos descendentes globales, podemos citar los de Sayin ([8]) y Yu-Zeleny([10]). Hasta donde sabemos, y a pesar de sus evidentes ventajas, todavía no ha sido publicado en la literatura especializada ningún procedimiento de tipo descendente local.

El algoritmo que presentamos en este trabajo se engloba en esta última categoría y emplea sencillos tests de eficiencia para las caras derivados a partir de sus descriptores ([6]). Estas características hacen que la herramienta desarrollada sea práctica e implementable, resultando especialmente adecuada cuando se den situaciones de alta dimensionalidad relativa en las caras eficientes maximales del problema MOLP a resolver.

2. Agradecimientos

Este trabajo ha sido parcialmente subvencionado a través del proyecto de la Dirección General de Universidades e Investigación del Gobierno de Canarias número PI042004/071.

3. Bibliografía

- [1] Armand, P. (1993). Finding All Maximal Efficient Faces in Multiobjective Linear Programming. *Mathematical Programming* 61, 357-375.
- [2] Armand, P. and Malivert, C. (1991). Determination of the Efficient Set in Multiobjective Linear Programming. *Journal of Optimization Theory and Applications* 70, 467-489.
- [3] Ecker, J., Hegner, N. and Kouada, I. (1980). Generating All Maximal Efficient Faces for Multiple Objective Linear Programs. *Journal of Optimization Theory and Applications* 30, 353-381.
- [4] Gal, T. (1977). A General Method for Determining the Set of All Efficient Solutions to a Linear Vectormaximum Problem. *European Journal of Operational Research* 1, 307-322.
- [5] Isermann, H. (1977). The Enumeration of the Set of All Efficient Solutions for a Linear Multiple Objective Program. *Operational Research Quarterly* 28, 711-725.
- [6] Jorge, J. (2003). Maximal Descriptor Set Characterizations of Efficient Faces in Multiple Objective Linear Programming. *Operations Research Letters* 31, 124-128.
- [7] Jorge, J. (2004). Programación Vectorial Lineal y Entera. Tesis doctoral. Servicio de Publicaciones de la Universidad de La Laguna.
- [8] Sayin, S. (1996). An Algorithm based on Facial Decomposition for Finding the Efficient Set in Multiple Objective Linear Programming. *Operations Research Letters* 19, 87-94.
- [9] Steuer, R. (1986). *Multiple Criteria Optimization: Theory, Computation and Application*. John Wiley & Sons.
- [10] Yu, P. and Zeleny, M. (1975). The Set of All Nondominated Solutions in Linear Cases and a Multicriteria Simplex Method. *Journal of Mathematical Analysis and Applications* 49, 430-468.

Un algoritmo genético híbrido basado en una codificación innovadora para el RCPSP

Javier Alcaraz Soria¹, Concepción Maroto Álvarez²

¹jalcaraz@eio.upv.es, Departamento de Estadística e Investigación Operativa Aplicadas y Calidad, Universidad Politécnica de Valencia

²maroto_con@gva.es, Conselleria de Territori i Habitatge, Generalitat Valenciana

Abstract

En este trabajo se presenta un nuevo algoritmo genético para resolver el problema de la programación de proyectos con recursos limitados (RCPSP). Éste hace uso de mecanismos de mejora y de un nuevo tipo de codificación para las soluciones, incorporando información específica del problema que es utilizada en el proceso genético así como en el proceso de construcción del calendario.

Palabras Clave: Metaheurísticas, Algoritmos Genéticos, Programación de Proyectos con Recursos Limitados.

1. Introducción

En los últimos años se han desarrollado diversos algoritmos para resolver el problema de la programación de proyectos con recursos limitados (RCPSP). La mayoría de éstos hacen uso de la representación estándar por lista de actividades (ALR), debido a que diversos estudios indicaban que ésta era la mejor representación posible para las soluciones al problema, independientemente del paradigma utilizado para resolverlo: algoritmos genéticos, tabu search, simulated annealing, etc. Estudios posteriores demuestran que otras representaciones, basadas en ALR que incorporan conocimiento específico del problema, permiten obtener un mejor comportamiento. La nueva representación que se describe a continuación combina las dos anteriores y utiliza la información contenida de forma eficiente tanto en el proceso genético como en el procedimiento utilizado para construir el calendario de ejecución.

2. Algoritmo genético y nueva representación

Hartmann [2] extiende la representación estándar por lista de actividades incorporando un bit adicional, que permite que se utilice tanto el esquema serie como el paralelo para construir el calendario de ejecución. Por otro lado, Alcaraz y Maroto [1] incorporan a ALR un bit, al que llaman gen forward/backward que indica si el calendario de ejecución debe construirse hacia delante o hacia atrás. Ambas extensiones de ALR demuestran su eficiencia, mejorando el comportamiento de la representación estándar.

La nueva codificación que proponemos combina las dos anteriores e incorpora ambos bits a la representación estándar, de manera que permite combinar los cuatro tipos de métodos para construir el calendario de ejecución a partir

de la lista de actividades: serie/adelante, serie/atrás, paralelo/adelante y paralelo/atrás. Esta información adicional es utilizada además en el proceso genético para generar de forma más eficiente a los hijos a partir de los padres, adaptando los operadores genéticos definidos por Alcaraz y Maroto [1] para operar sobre esta nueva representación. Además, para evitar problemas de convergencia y evitar quedar atrapados en óptimos locales, se ha incorporado al algoritmo un procedimiento que permite incorporar nuevo material genético en el proceso en determinados instantes del mismo. El algoritmo también utiliza un procedimiento de mejora de soluciones, basado en justificaciones hacia delante y hacia atrás, que también utiliza la nueva información que incorporan las soluciones con el fin de incrementar su eficiencia.

3. Resultados

En el cuadro siguiente se muestran los resultados obtenidos por este algoritmo utilizando los proyectos medianos (J60) de la librería estándar PSPLIB. Por restricciones de espacio en este resumen, no se han incluido los resultados para otros proyectos. Estos resultados se comparan con los obtenidos con los mejores algoritmos publicados, pudiéndose observar el excelente comportamiento de nuestro algoritmo.

Autor (Año)	Tipo Alg.	Max. nº calendarios	
		1.000	5.000
Alcaraz and Maroto (este trab.)	Genético	11.67	11.05
Kochetov and Stoylar (2003)	Gen., Tabu	11.71	11.17
Debels et al. (2004)	Scatter	11.73	11.10
Valls et al. (2005)	Genético	12.21	11.27
Hartmann (2002)	Genético	12.21	11.70
Valls et al. (2005)	Sampling	12.73	12.35

Cuadro 2: Medio de incremento con respecto a la duración dada por el camino crítico. J60 PSPLIB

4. Agradecimientos

Este trabajo ha sido financiado con cargo al Proyecto de Investigación DPI2004-06366-C03-01, concedido por el Ministerio de Educación y Ciencia.

5. Bibliografía

- [1] Alcaraz, J. and Maroto C. (2001). A robust genetic algorithm for resource allocation in project scheduling. *Annals of Oper. Res.* 102, 83-109.
- [2] Hartmann, S. (1999). Project Scheduling under Limited Resources. Springer.

Un algoritmo heurístico para Clasificación Semisupervisada

Rafael Blanquero¹, Emilio Carrizosa¹, Norberto Díaz²

¹{rblanquero,ecarrizosa}@us.es, Dpto. de Estadística e Investigación Operativa, Universidad de Sevilla

²ndiaz@lsi.us.es, Dpto. de Lenguajes y Sistemas Informáticos, Universidad de Sevilla

Abstract

En este trabajo se aborda el problema de clasificación semisupervisada. La búsqueda del clasificador se basa en la resolución de un problema combinatorio, para lo cual se propone una metodología metaheurística de resolución. Presentamos nuestra experiencia computacional en diferentes bases de datos.

Claves: Clasificación semisupervisada, Variable Neighborhood Search, Minería de Datos.

AMS: 90C27, 90C10, 90C59

Un algoritmo para optimizar una función lineal sobre la región eficiente de un problema multiobjetivo lineal entero

Jesús M. Jorge Santiso¹, Marianela Carrillo Fernández²

¹jjorge@ull.es, Departamento de Estadística, Investigación Operativa y Computación, Universidad de La Laguna

²mcarrif@ull.es, Departamento de Economía Aplicada, Universidad de La Laguna

Abstract

En este trabajo presentamos un algoritmo implementable para optimizar una función lineal sobre la región eficiente de un problema multiobjetivo lineal entero. El método propuesto se basa en resolver una sucesión de problemas escalares lineales enteros progresivamente más restringidos que evitan el inconveniente de generar de forma explícita la región eficiente del problema.

Palabras Clave: Optimización sobre la región eficiente, programación multiobjetivo, programación entera

AMS: 90C29, 90C99

1. Introducción

La programación multiobjetivo lineal entera (MOILP) trata de resolver una importante clase de problemas que surgen en la toma de decisiones multicriterio cuando todas o algunas de las variables del modelo son discretas.

Tradicionalmente la resolución de este tipo de problemas ha estado orientada a la generación de todo el conjunto de soluciones eficientes o, al menos, de partes significativas del mismo ([8]). Para ello se han empleado un amplio abanico de técnicas, mereciendo ser destacadas entre otras, programación dinámica ([8]), ramificación y acotación ([6]), métodos iterativos ([5]), retículos completamente eficientes ([3]), etc.

Sin embargo, y a pesar de tales esfuerzos, el cálculo explícito de la región eficiente de un MOILP suele estar desaconsejado por razones prácticas para problemas medios o grandes. Efectivamente, aparte de los enormes recursos computacionales que consumen tales algoritmos, el desmesurado tamaño habitual de la región eficiente suele saturar al decisor, hasta el punto de que le puede resultar imposible seleccionar su solución más preferida de entre el conjunto generado.

Un enfoque más realista, que ha sido tratado con cierto detenimiento en el caso continuo por numerosos autores a lo largo del tiempo (ver, por ejemplo, [2,4,9]), pero que apenas ha sido explorado en el caso entero ([1]), consiste en optimizar sobre la región eficiente una función lineal arbitraria que permita

medir la importancia de o discriminar entre las soluciones eficientes disponibles ([2]).

Siguiendo esta línea, en este trabajo presentamos un método finito e implementable (utilizando, por ejemplo, paquetes de software convencionales de programación entera tales como Cplex) que permite optimizar de forma exacta una función lineal arbitraria, no necesariamente combinación lineal de los objetivos del MOILP subyacente, sobre la región eficiente del mismo. El algoritmo no necesita generar de forma explícita la región eficiente del MOILP y se basa en resolver una sucesión de problemas escalares lineales enteros progresivamente más restringidos inspirados en las ideas dadas en [5] y [7].

2. Agradecimientos

Este trabajo ha sido parcialmente subvencionado a través del proyecto de la Dirección General de Universidades e Investigación del Gobierno de Canarias número PI042004/071.

3. Bibliografía

- [1] Abbas, M. and Chaabane, D. (2005). Optimizing a Linear Function over an Integer Efficient Set. To appear in *European Journal of Operational Research*.
- [2] Benson, H. (1984). Optimization over the Efficient Set. *Journal of Mathematical Analysis and Applications* 98, 562–580.
- [3] Jorge, J (2004). Un Algoritmo para Determinar la Región Eficiente de un Problema de Programación Multiobjetivo Lineal Entero. *Actas del XXVIII Congreso Nacional de Estadística e Investigación Operativa*.
- [4] Jorge, J. (2005). A Bilinear Algorithm for Optimizing a Linear Function over the Efficient Set of a Multiple Objective Linear Programming Problem. *Journal of Global Optimization* 31, 1-16.
- [5] Klein, D. and Hannan, E. (1982). An Algorithm for the Multiple Objective Integer Linear Programming Problem. *European Journal of Operational Research* 9, 378-385.
- [6] Marcotte, O. and Soland, R. (1986). An Interactive Branch-and-Bound Algorithm for Multiple Criteria Optimization. *Management Science* 32, 61-75.
- [7] Sylva, J. and Crema, A. (2004). A Method for Finding the Set of Non-Dominated Vectors for Multiple Objective Integer Linear Programming. *European Journal of Operational Research* 158, 46-55.
- [8] Teghem, J. and Kunsch, P. (1986). A Survey of Techniques for Finding Efficient Solutions to Multi-Objective Integer Linear Programming. *Asia-Pacific Journal of operational Research* 3, 95-108.
- [9] Yamamoto, Y. (2002). Optimization over the Efficient Set: Overview. *Journal of Global Optimization* 22, 285-317.

Un estudio sobre la coherencia de preferencias lingüísticas

Susana Díaz¹, José Luis García Lapresta², Susana Montes¹

¹ {diazsusana,montes}@uniovi.es, Departamento de Estadística, Investigación Operativa y Didáctica de la Matemática, Universidad de Oviedo

²lapresta@eco.uva.es, Departamento de Economía Aplicada (Matemáticas), Universidad de Valladolid

Abstract

En este trabajo se lleva a cabo un estudio sobre la consistencia de las preferencias lingüísticas. En él se analiza la relación existente entre la transitividad de las preferencias lingüísticas y la de sus correspondientes relaciones de preferencia estricta y de indiferencia nítidas.

Claves: preferencia lingüística, transitividad.

1. Introducción

Una parte fundamental del proceso de toma de decisiones consiste en modelizar las preferencias del decisor. De esta manera surgen las relaciones de preferencia e indiferencia nítidas y, con ellas, las estructuras de preferencia nítidas. Sin embargo, las relaciones nítidas no permiten representar con precisión las decisiones de un individuo, ya que no admiten grados de cumplimiento: o se prefiere o no se prefiere, sin que existan situaciones intermedias. Con objeto de modelizar de una forma más precisa las respuestas de un decisor sobre sus preferencias surgen las relaciones de preferencia e indiferencia borrosas o difusas, que además de los valores 1 y 0 (presencia y ausencia de relación, respectivamente), admiten cualquier valor intermedio ([2, 4]). Aunque las relaciones borrosas son ampliamente utilizadas en la práctica, las preferencias lingüísticas pueden ser más adecuadas en algunas ocasiones para capturar la falta de precisión del comportamiento humano ([3, 5]). Sin embargo, para decantarnos por estas relaciones, debemos analizar su consistencia mediante el estudio de propiedades como la transitividad, como se ha hecho con las relaciones borrosas [1].

2. Preferencias lingüísticas y transitividad

Consideremos un conjunto finito de alternativas A y un conjunto de etiquetas lingüísticas $\mathcal{L}_n = \{l_0, l_1, \dots, l_{2n}\}$, con $n \geq 1$, ordenadas linealmente: $l_0 < l_1 < \dots < l_{2n}$. Una relación de preferencia lingüística sobre A basada en \mathcal{L}_n es una aplicación $R : A \times A \longrightarrow \mathcal{L}_n$ que satisface la condición de reciprocidad: $R(a, b) = l_k \Rightarrow R(b, a) = l_{2n-k}$, para cualesquiera $a, b \in A$ y $k \in \{0, 1, \dots, 2n\}$; en tal caso, se denotará por $R \in \mathcal{L}_n(A)$. El significado de las etiquetas es el siguiente: $R(a, b) = l_{2n}$, si a es absolutamente preferida a b ; $l_n < R(a, b) < l_{2n}$, si a es preferida a b en cierto grado; $R(a, b) = l_n$, si a es indiferente a b ; $l_0 < R(a, b) < l_n$, si b es preferida a a en cierto grado; y $R(a, b) = l_0$, si b es absolutamente preferida a a .

Toda relación de preferencia lingüística tiene asociadas dos relaciones nítidas: la de preferencia (fuerte) P , que se satisface cuando $R(a, b) \geq l_{n+1}$; y la de indiferencia I , que se satisface cuando $R(a, b) = l_n$. En este trabajo estudiamos la racionalidad en el contexto de las preferencias lingüísticas a través del concepto de transitividad. Es bien conocido que en el ámbito de las preferencias borrosas existe una amplia gama de propiedades de transitividad, generalmente basadas en t-normas. En nuestro marco lingüístico se utilizarán *operadores monótonos* sobre $\mathcal{L}_n^+ = \{l_n, \dots, l_{2n}\}$: aplicaciones $F : \mathcal{L}_n^+ \times \mathcal{L}_n^+ \rightarrow \mathcal{L}_n$ que satisfagan $F(l_j, l_i) \leq F(l_k, l_i)$ y $F(l_i, l_j) \leq F(l_i, l_k)$, para cualesquiera $l_i, l_j, l_k \in \mathcal{L}_n^+$ tales que $l_j \leq l_k$. Dados un operador monótono F y $R \in \mathcal{L}_n(A)$, se dirá que R es *F-transitiva moderada fuerte* si $R(a, b) \geq F(R(a, c), R(c, b))$ para cualesquiera $a, b, c \in A$, tales que $R(a, c), R(c, b) \in \mathcal{L}_n^+$. Por otra parte, se dirá que R es *F-transitiva moderada débil* si $R(a, b) \geq F(R(a, c), R(c, b))$ para cualesquiera $a, b, c \in A$, tales que $R(a, c), R(c, b) \in \mathcal{L}_n^{++} = \{l_{n+1}, \dots, l_{2n}\}$.

En nuestro trabajo consideramos estos dos tipos de transitividad lingüística y caracterizamos aquellos operadores monótonos F para los que la F -transitividad moderada fuerte de una relación de preferencia lingüística R permiten asegurar la transitividad de la relación de indiferencia nítida asociada I . También caracterizamos los operadores monótonos que permiten garantizar la transitividad de la relación de preferencia nítida asociada P a partir de las F -transitividades moderadas, tanto fuerte como débil, de cualquier relación de preferencia lingüística. Además demostramos las relaciones existentes entre las condiciones que garantizan las transitividades de P y de I , así como las que aseguran que las composiciones de P e I están contenidas en P .

3. Agradecimientos

Este trabajo está subvencionado en parte por el Proyecto MTM2004-01269.

4. Bibliografía

- [1] Díaz, S., De Baets, B., Montes, S. (en prensa): "Transitivity Bounds in Additive Fuzzy Preference Structures". *IEEE Transactions on Fuzzy Systems*.
- [2] Fodor, J., Roubens, M. (1994): *Fuzzy Preference Modelling and Multicriteria Decision Support*. Kluwer Academic Publishers, Dordrecht.
- [3] García-Lapresta, J.L. (2006): "A general class of simple majority decision rules based on linguistic opinions". *Information Sciences* 176, pp. 352-365.
- [4] García-Lapresta, J.L., Llamazares, B. (2000): "Aggregation of fuzzy preferences: Some rules of the mean". *Social Choice and Welfare* 17, pp. 673-690.
- [5] Zadeh, L.A. (1975): "The concept of a linguistic variable and its applications to approximate reasoning". *Information Sciences*. Part I: 8, pp. 199-249. Part II: 8, pp. 301-357. Part III: 9, pp. 43-80.

Un método simplex para problemas de caminos mínimos con pivotaje múltiple

C. González Martín, A. Sedeño Noda, S. Alonso Rodríguez

cgonmar@ull.es, asedeño@ull.es, salonso@ull.es

Departamento de Estadística, Investigación Operativa y Computación.
Universidad de La Laguna

Abstract

Utilizando una regla de pivotaje múltiple, proponemos un algoritmo simplex para resolver problemas de caminos mínimos sobre un red dirigida, con n vértices y m arcos, que se ejecuta en $O(nm)$. La aplicación del algoritmo se ilustra con el correspondiente estudio computacional

Palabras Clave: Problemas de caminos mínimos, regla de pivotaje múltiple, método simplex para redes

AMS: 90B10, 90B06, 90C27

1. Introducción

El problema de camino mínimo (*CM*), sobre una red dirigida, consiste en encontrar caminos dirigidos de longitud mínima desde un vértice fuente a cualquier otro vértice o , en su caso, detectar ciclos dirigidos (circuitos) de longitud negativa. Este problema aparece en un gran número de contextos (en general, sobre redes de transportes o comunicaciones) y ha sido intensivamente estudiado.

Para resolver un problema *CM* sobre una red dirigida con n vértices y m arcos, es conocido que el método fuertemente polinomial de Bellmam-Ford-Moore (Bellman [1], Ford [3], Moore [6]) es el más eficiente desde el punto de vista computacional teórico. Existen, también, otros procedimientos que son variantes del método simplex para problemas de caminos mínimos (ver, Dantzig [2], Orlin [7], Goldfarb *et al.* [4],...). Cabe destacar que Goldfarb and Jin [5] introducen el primer método simplex para problemas de caminos mínimos que se ejecuta en $O(nm)$ al igual que el método de Bellmam-Ford-Moore.

Debido a la estructura especial del problema *CM*, es posible ejecutar ejecutar simultáneamente varias operaciones de pivotaje (pivotaje multiple). Esta particularidad pretende ser aprovechada en este trabajo.

2. Formulación del problema y algoritmo

Dada una red dirigida con n vértices y m arcos, destacamos un vértice particular, que denominamos *fuente*, y asociamos a cada arco la correspondiente longitud.

Como es conocido, el problema *CM* se puede modelizar como un problema de flujos de coste mínimo y, por tanto, su resolución se puede afrontar utilizando las propiedades y herramientas propias de los problemas de flujos en redes. En este trabajo, el problema *CM* se resuelve con un método Simplex para redes en el que, aprovechando las propiedades específicas, se utiliza una nueva regla de selección simultánea de varias variables no básicas que, por ello, supone la realización de pivotaje múltiple. Este algoritmo es eficiente desde la perspectiva computacional y se ejecuta en $O(nm)$.

3. Bibliografía

- [1] Bellman, R., "On a Route Problem", *Quart. Of Appl. Math.* 16 (1958), 87-90.
- [2] Dantzig, G. B., "Discrete-variable extremum principles", *Operations Research* 5 (1957), 266-277.
- [3] Ford, L. R. 1956. *Network Flow Theory*. The Rand Corporation Report P-923, Santa Monica, Calif.
- [4] Goldfarb D., Hao, J. and Kai, S.R., "Efficient Shortest Path Simplex Algorithms", *Operations Research* 38 (4) (1990), 624-628.
- [5] Goldfarb D. and Jin, Z., "An $O(nm)$ -Time Network Simplex Algorithm for the Shortest Path Problem", *Operations Research* 47 (3) (1999), 445-448.
- [6] Moore, Z. F., "The Shortest Path Through a Maze", *In Proceedings of the International Symposium on Theory of Switching*, Part II (1957), 285-292.
- [7] Orlin, B."On the Simplex Algorithm for Networks and Generalized Networks", *Mathematical Programming Study* 25 (1985), 166-178.

Un modelo de asignación en el mercado inmobiliario con intervención del estado

Antonio Armada Estévez¹, Joaquín Sicilia Rodríguez²

antarm@tiscali.es, jsicilia@ull.es

Departamento de Estadística, Investigación Operativa y Computación,
Universidad de La Laguna

Abstract

En este trabajo se estudia un modelo de juego de asignación para los agentes que intervienen en un mercado inmobiliario (compradores-vendedores o propietarios-inquilinos). El objetivo consiste en obtener el máximo número de contratos, intentando satisfacer las preferencias de los agentes implicados.

Palabras Clave: Juegos cooperativos, problema de asignación

AMS: 91A12, 91A80

1. Introducción

Los juegos de asignación fueron introducidos por Shapley y Shubik [4] y consisten en un tipo particular de juegos cooperativos donde el conjunto de todos los jugadores se divide en dos subconjuntos, de forma que sólo deben relacionarse o emparejarse jugadores de subconjuntos distintos. Dichos autores comprobaron que el Core de un juego de asignación corresponde al conjunto de soluciones óptimas del problema dual relacionado con el problema de programación entera que caracteriza a dichos juegos.

Tijs y otros [6] introdujeron los juegos de permutación como juegos de coste. Probaron que cada juego de asignación es un juego de permutación y que los juegos de permutación y de asignación son totalmente balanceados. Balinski y Gale [1] probaron que el Core de un juego de asignación tiene como mucho $\binom{2m}{m}$ puntos extremos, donde m es el mínimo de los cardinales de los dos subconjuntos en que se divide el conjunto de jugadores. Posteriormente, Solymosi y Raghavan [5] dieron un algoritmo para encontrar el nucleolo de un juego de asignación.

Por otro lado, Rochford [3] introdujo las asignaciones balanceadas y emparejadas simétricamente para juegos de asignación como una forma de alcanzar un reparto del beneficio generado por cada emparejamiento. Por último, Curiel [2] dedica un capítulo de su libro a revisar y presentar los principales resultados sobre juegos de asignación y permutación.

En este trabajo nos centraremos en un juego de asignación planteado en el mercado inmobiliario. La situación que planteamos es la de una agencia estatal cuya misión es la de facilitar el acceso a la vivienda a personas con determinadas dificultades (jóvenes, empleados con salarios inferiores a cierta cantidad, jubilados, etc.). Para llevar a cabo su labor dispone de una lista de personas que desean comprar (o alquilar) una vivienda y otra lista de personas dispuestas a vender (o alquilar) viviendas a través de la agencia (gracias a rebajas de impuestos, subsidios o determinadas compensaciones ofrecidas por la administración). Para conseguir llegar a acuerdos, la agencia se compromete a subvencionar la diferencia entre lo que está dispuesto a pagar el comprador y el precio que desea el vendedor.

El objetivo consiste en obtener el máximo número de contratos, minimizando la perdida total que se produce en los mismos e intentando satisfacer las preferencias de los agentes.

2. Agradecimientos

Este trabajo ha sido parcialmente financiado por el Ministerio de Educación y Ciencia, Proyecto de Investigación MTM2004-07550 del Plan Nacional de Investigaciones Científicas, Desarrollo Tecnológico e Innovación.

3. Bibliografía

- [1] Ballinski, M.L. y D. Gale (1990). On the Core of the assignment game, en: *Functional Analisis, Optimization and Matchmatics Economics*. Oxford University Press, 274-289.
- [2] Curiel, I. (1997). *Cooperative Game Theory and Applications*. Kluwer Academic Publishers.
- [3] Rochford, S.C. (1984). Symetrically Pairwise-bargained allocations in an assignment market. *Journal of Economics Theory* 34, 262-281.
- [4] Shapley, L. S. y M. Shubik (1972). The assignment game I: The Core. *International Journal of Game Theory* 1, 111-130
- [5] Solymosi, T. y T.E.S. Raghavan (1994). An algorithm for finding the nucleoulus of assignment games. *International Journal of Game Theory* 23, 119-143.
- [6] Tijs, S.H., T. Parthasarathy, J.A.M. Potters y V. Rajendra Prasad (1984). Permutation games: Another class of totally balanced games. *OR Spektrum* 6, 119-123.

Un modelo de simulación para planificación de tráfico ferroviario

Iosu Mirem Martínez¹, Begoña Vitoriano², Antonio Fernández-Cardador³, Paloma Cucala⁴

¹imartinez@recuerdo.net, Universidad Pontificia Comillas de Madrid

²bvitoriano@upcomillas.es, Instituto de Investigación Tecnológica,
Universidad Pontificia Comillas de Madrid

³antonio.fernandez@iit.upcomillas.es, Instituto de Investigación Tecnológica,
Universidad Pontificia Comillas de Madrid

⁴paloma.cucala@upcomillas.es, Instituto de Investigación Tecnológica,
Universidad Pontificia Comillas de Madrid

Abstract

La planificación y regulación de sistemas ferroviarios son tareas que requieren de herramientas que permitan simular el tráfico ferroviario para evaluar y comparar estrategias. Se presenta un modelo de simulación, fundamentalmente el modelado de los parámetros básicos (en concreto, el tiempo de parada), para una línea metropolitana regulada por intervalo.

Palabras Clave: Planificación, simulación, tráfico ferroviario

1. Introducción

En la gestión y operación de los sistemas ferroviarios se plantean distintos niveles de decisión. En un primer nivel se encuentran las decisiones estratégicas, que son decisiones a medio y largo plazo, y que engloban todo lo que es la infraestructura del proyecto. En un segundo nivel, definido para medio y corto plazo, se encuentra la planificación, y en un tercer nivel se encuentra la operación/regulación que sucede en tiempo real.

En planificación se toman decisiones sobre cómo van a funcionar los trenes que van a circular por la línea y que van a dar servicio a un determinado número de viajeros, sujetos a restricciones técnicas. La planificación es a medio y corto plazo, mientras que la operación / regulación son decisiones que se toman en tiempo real para lograr los objetivos de la planificación recibiendo de ésta las reglas de control que hay que aplicar.

Como herramientas de planificación, las más útiles son los simuladores que evalúan las decisiones y permiten prever la respuesta del sistema ante diversas decisiones. Un simulador es un evaluador de consecuencias de decisiones antes de implantarlas sobre el sistema. Funcionan off-line y ayudan a la planificación en la toma de decisiones.

En este artículo se presenta un simulador para líneas ferroviarias metropolitanas reguladas por intervalo. Para realizar este simulador, se tienen que modelar las

variables que intervienen en el proceso. El modelado debe ser un compromiso entre un modelo exhaustivo, que pueda llevar a no poder resolver ni utilizar el modelo, y un modelo simplista que no permita evaluar lo que sería la respuesta del sistema real.

La mayor parte del estudio se centra en el modelado de los tiempos de parada del tren en una estación y de los tiempos de intervalo entre trenes consecutivos. Los tiempos de parada del tren son parámetros básicos del simulador, y son variables aleatorias muy relacionadas con el número de viajeros que se encuentran en el andén en el momento en que llega el tren. En la literatura sobre el tema ([1], [2] y [3]) siempre se ha considerado que el tiempo de parada está íntimamente relacionado con el tiempo que transcurre entre dos trenes consecutivos, afirmándose que existe una relación lineal entre este tiempo y el tiempo de parada, que resulta al fin fundamental para las reglas de control propuestas. En el estudio se analizan diversas hipótesis buscando la distribución que mejor se ajuste a estos tiempos, así como la posible relación entre los tiempos de parada y los tiempos de intervalo entre dos trenes consecutivos. El resultado es contrario a lo que se ha venido manteniendo en la literatura tanto para la distribución de los tiempos, siendo la lognormal la que mejor se ajusta a los valores, como para la relación entre las dos variables, que se ha mostrado prácticamente nula.

Un tema de estudio también importante es el tema de las incidencias. Estas incidencias se presentan en la línea de metro y afectan al normal funcionamiento de la misma. Por lo tanto, es importante lograr identificar y separar estas incidencias, es decir, qué tiempos de parada se van a considerar que son anormalmente largos y en consecuencia son incidencias, teniendo un estudio y tratamiento aparte de los tiempos de parada “normales”.

Finalmente se presentan las conclusiones generales del estudio, conclusiones que van a permitir definir qué características va a tener el simulador, y se presenta un modelo simplificado del simulador.

2. Bibliografía

- [1] Breusegem V. Van, Campion G. and Bastin G. (1991). Traffic Modeling and State Feedback Control for Metro Lines. *IEEE Transactions on Automatic Control*, 36, 7, 770-784.
- [2] Ceder A. and Marguier P. (1985). Passenger waiting time at transit stops. Supplied by the British Library- “The world’s knowledge”.
- [3] Silvino C. and Milani B. (2001). Regulação Robusta de Tráfego em Linhas de Metro. *Controle & Automação*, 12, 2, 118-130.

Un modelo EOQ con demanda exponencialmente dependiente del precio

L.A. San-José¹, J. Sicilia², J. García-Laguna³

¹Departamento de Matemática Aplicada, Universidad de Valladolid

²Departamento de Estadística, I.O. y Computación, Univ. de La Laguna

³Departamento de Estadística e Investigación Operativa, Univ. de Valladolid

¹augusto@mat.uva.es, ²jsicilia@ull.es, ³laguna@eio.uva.es

Abstract

Se considera un modelo de inventario de tamaño del lote económico donde el precio de venta es una variable a determinar. Se admite que la demanda y el precio de venta están inversamente relacionados a través de una función exponencial negativa. Se determinan el tamaño óptimo de pedido y el precio óptimo. Para ilustrar los resultados teóricos obtenidos, se dan ejemplos numéricos y se realiza un análisis de sensibilidad.

Claves: modelos EOQ, demanda dependiente, precio.

AMS: 90B05.

1. Introducción

Dos de las hipótesis que con mayor frecuencia se consideran en los estudios de modelos EOQ son que la tasa de demanda es constante y determinística, y que el precio unitario de venta es constante e independiente de la demanda (Hadley y Whitin (1963), Silver et al. (1998), Axsäter S. (2000), Zipkin (2000)). Sin embargo, en muchas situaciones, la tasa de demanda y el precio unitario de venta pueden estar interrelacionadas. Por ejemplo, en un entorno fabril con demanda alta y amplia capacidad productiva puede ser razonable incrementar la tasa de producción para reducir los costes unitarios de fabricación, y con ello el precio del producto. Análogamente, en un entorno comercial, si se tiene una demanda suficientemente alta puede ser interesante hacer pedidos grandes si con ellos se consiguen descuentos y se posibilita la bajada del precio de venta. En consecuencia, en tales circunstancias parece razonable considerar que la tasa de demanda y el precio unitario están inversamente relacionados.

El problema de determinar tanto el precio óptimo de venta como el tamaño óptimo de pedido ha sido estudiado en Ray et al. (2005), bajo la hipótesis de que la tasa de demanda y el precio unitario están relacionados isoelásticamente y, así mismo, cuando la demanda depende linealmente del precio de venta.

En este trabajo se considera un modelo de inventario de un solo producto con demanda independiente, horizonte de planificación infinito y patrón de demanda uniforme. Se admite que la demanda y el precio están inversamente

relacionados a través de una curva exponencial negativa. El reabastecimiento del inventario se realiza en lotes del mismo tamaño que se reciben de una sola vez y el tiempo de reabastecimiento es nulo. El coste unitario de compra es constante e independiente del tamaño del lote, y el coste de realizar un pedido es constante e independiente de la cantidad solicitada. El coste de almacenamiento es una función lineal basada en el inventario medio. El control del inventario se realiza mediante un sistema de revisión continua y no se permite escasez. Las variables de decisión son el tamaño del lote económico y el precio de venta. Bajo las hipótesis anteriormente citadas se determinan el tamaño óptimo de pedido y el precio óptimo.

2. Agradecimientos

Este trabajo ha sido parcialmente subvencionado por el *Ministerio de Educación y Ciencia*, Proyecto de Investigación MTM2004-07550, del Plan National de Investigación Científica, Desarrollo Tecnológico e Innovación. También, ha sido parcialmente subvencionado por el proyecto OTP/04/03 del Gobierno Regional de la *Comunidad Autónoma de Castilla y León*. Una parte de esta investigación ha sido realizado durante una estancia del segundo autor en la Universidad de Valladolid (Beca de Estancia de la Dirección General de Universidades e Investigación del Gobierno de la *Comunidad Autónoma de Canarias*).

3. Bibliografía

- [1] Axsäter S. (2000). *Inventory Control*. Kluwer Academic Press.
- [2] Hadley G. y Whitin T.M. (1963). *Analysis of Inventory Systems*. Prentice-Hall.
- [3] Ray S., Gerchak Y. y Jewkes E.M. (2005). Joint pricing and inventory policies for make-to-stock products with deterministic price-sensitive demand. *International Journal of Production Economics* 97, 143-158.
- [4] Silver E.A. Pyke D.F. y Peterson R. (1998). *Inventory Management and Production Planning and Scheduling*. John Wiley.
- [5] Zipkin H.P. (2000). *Foundations of Inventory Management*. McGraw Hill.

Un nuevo algoritmo GRASP para el problema del empaquetamiento bidimensional

R. Álvarez-Valdés¹, F. Parreño², J. M. Tamarit³

¹ramon.alvarez@uv.es, ³jose.tamarit@uv.es Departamento de Estadística e
Investigación Operativa, Universitat de València

²fparreno@infor-ab.uclm.es, Departamento de Sistemas Informáticos,
Universidad de Castilla-La Mancha

Abstract

En este artículo se propone un nuevo algoritmo heurístico basado en la metodología GRASP para el problema de empaquetar un conjunto de piezas rectangulares en un rectángulo de anchura fija y altura ilimitada, de manera que se minimice la altura total requerida.

Palabras Clave: Corte y Empaquetamiento, Heurísticos, GRASP

1. El problema de empaquetamiento

El problema de empaquetamiento bidimensional en un rectángulo de altura ilimitada (*Strip Packing Problem*) consiste en hallar la mejor ubicación de un conjunto de piezas rectangulares de dimensiones conocidas (h_i, w_i) , $i=1..n$, en un rectángulo de anchura W y altura infinita, de forma que se minimice la altura requerida H . El problema tiene muchas aplicaciones en las industrias del papel y cartón, vidrio o metal, cuando el material se recibe en bobinas grandes, cuya longitud puede considerarse infinita, y de las que se han de cortar piezas de los tamaños requeridos por los clientes.

En los últimos años se han desarrollado muchos algoritmos heurísticos para este problema. Los heurísticos constructivos sencillos se han integrado en procedimientos metaheurísticos más complejos, sobre todo algoritmos genéticos (Bortfeld, 2005), aunque también existen algoritmos Tabu Search (Iori et al., 2003) y un algoritmo GRASP desarrollado por Beltrán et al. (2002). Un estudio completo de este tipo de problemas puede encontrarse en Hopper y Turton (2001).

Nosotros hemos desarrollado un nuevo algoritmo basado en GRASP y lo hemos probado en la mayor parte de los problemas test existentes en la literatura. Los resultados indican que nuestro procedimiento iguala y frecuentemente mejora los mejores resultados conocidos hasta el momento.

2. El algoritmo GRASP

Los tres elementos básicos de un algoritmo GRASP son el procedimiento constructivo *greedy*, la aleatorización y la fase de mejora. Para la primera fase

hemos desarrollado un nuevo procedimiento que elige en cada paso la pieza a cortar de acuerdo con un criterio que mide la adaptación de la pieza al perfil de la solución parcial construida. Esta elección de la pieza no es determinista, sino que se elige al azar dentro de un conjunto restringido de buenos candidatos, siguiendo la estrategia de GRASP reactivo.

Para la fase de mejora, después de diseñar y probar cuatro métodos alternativos de búsqueda local, utilizamos el procedimiento propuesto por Beltrán et al. (2002) en el que se eliminan las últimas piezas de la solución y se vuelven a colocar con un procedimiento determinista.

3. Resultados computacionales

Para cada elemento del algoritmo se diseñaron y probaron diversas alternativas y se eligieron las más eficientes. El algoritmo completo se ha probado sobre los conjuntos de problemas test más utilizados en la literatura. En total, se han estudiado 662 instancias, incluyendo instancias de pérdida cero, generados cortando repetidamente un rectángulo grande, e instancias adaptadas de otros problemas de empaquetamiento y corte, para los que la solución óptima contiene zonas vacías que se consideran pérdidas.

El estudio computacional muestra que en cada uno de los conjuntos de instancias que se han considerado nuestro algoritmo obtiene los mejores resultados, en tiempos de computación relativamente cortos.

La combinación de la metodología GRASP, que se adapta especialmente bien a estos problemas, y las estrategias específicas desarrolladas en cada fase del proceso han dado como resultado un algoritmo muy eficiente para el problema del empaquetamiento.

4. Bibliografía

- [1] Beltran J., Calderón J.E., Cabrera R.J., Moreno J.M. (2002): Procedimientos constructivos adaptados (GRASP) para el problema del empaquetado bidimensional, *Revista Iberoamericana de Inteligencia Artificial* 15, 26-33..
- [2] Bortfeldt A. (2005): A genetic algorithm for the two-dimensional strip packing problem with rectangular pieces, *European Journal of Operational Research*, en prensa.
- [3] Hopper E., Turton B.C.H. (2001): An empirical investigation of metaheuristic and heuristic algorithms for a 2D packing problem, *European Journal of Operational Research* 128, 34-57..
- [4] Iori M., Martello S., Monaci M. (2003): Metaheuristic algorithms for the strip packing problem, En: Pardalos, Panos y Korotkikh (eds.): *Optimization and Industry: New frontiers*, Kluwer, Boston, pp. 159-179.

Un nuevo método simplex eficiente para resolver problemas de caminos mínimos

A. Sedeño Noda, C. González Martín

asedeñ@ull.es, cgonmar@ull.es

Departamento de Estadística, Investigación Operativa y Computación.
Universidad de La Laguna

Abstract

Utilizando una regla de pivotaje multiple, proponemos un algoritmo simplex para resolver problemas de caminos mínimos sobre un red dirigida, con n nodos y m arcos, que se ejecuta en. $O(nm)$. La aplicación del algoritmo se ilustra con el correspondiente estudio computacional

Palabras Clave: Problemas de caminos mínimos, método simplex, eficiencia computacional

AMS: 90B10, 90B06, 90C27

1. Introducción

El problema de camino mínimo (CM), sobre una red dirigida, consiste en encontrar caminos dirigidos de longitud mínima desde un nodo fuente a cualquier otro nodo o, en su caso, detectar ciclos dirigidos (circuitos) de longitud negativa. Este problema aparece en un gran número de contextos (en general, sobre redes de transportes o comunicaciones) y ha sido intensivamente estudiado (ver, por ejemplo, Ahuja *et al.* [1]).

Entre los diversos métodos propuestos en la literatura, cabe señalar que Goldfarb and Jin [2] introducen el primer método simplex para problemas de caminos mínimos que se ejecuta en $O(nm)$ y que utiliza la idea de nodo etiquetado permanentemente. Nuestra propuesta incluye un procedimiento con igual eficiencia computación tomando como base el nuevo concepto de nodo pseudo permanentemente etiquetado.

2. Formulación del problema y algoritmo

Dada la red $G = (V, A)$, sea $V = \{1, \dots, n\}$ el conjunto de nodos y A el conjunto de m arcos. Distinguimos en V un nodo s que denominamos fuente. Si, para cada $(i, j) \in A$, $c_{ij} \in \mathbb{R}$ es la longitud asociada, la correspondiente longitud de un camino dirigido es la suma de las longitudes de los arcos que contiene. El problema CM consiste en determinar los caminos de longitud mínima entre s y cualquier otro nodo $i \in V \setminus \{s\}$, o, en su caso, detectar un ciclo dirigido de longitud negativa.

Si asociamos un flujo x_{ij} con cada arco (i, j) , una disponibilidad $b_s = n - 1$

con el nodo s y disponibilidades $b_i = -1, \forall i \in V \setminus \{s\}$, obtenemos la siguiente formulación para el problema CM (ver Ahuja et al [1]):

$$\begin{aligned} \text{min. } & c(x) = \sum_{(i,j) \in A} c_{ij} x_{ij} \\ \text{s.a. } & \sum_{j : (j,i) \in A} c_{ij} x_{ji} - \sum_{j : (i,j) \in A} c_{ij} x_{ij} = b_i, \quad \forall i \in V \\ & x_{ij} \geq 0, \quad \forall (i,j) \in A \end{aligned}$$

Los potenciales de los nodos (variables duales) que corresponden a un árbol básico T se obtienen haciendo $\pi_i(T) = 0$ a través de la ecuación $c_{ij} - \pi_i(T) + \pi_j(T) = 0 \quad \forall (i,j) \in T$. Definimos $d_i(T) = -\pi_i(T)$ como la etiqueta distancias del nodo i , $\forall i \in V$.

Definición 1. Un nodo $j \in V$ está pseudo permanentemente etiquetado (pseudo permanently labeled) en T , si y sólo si, $\bar{c}_{ki}(T) \geq 0 \quad \forall k \in \Gamma_i^- = \{l \in V \mid (l,i) \in A\}$ y $\forall i \in B_j(T)$, donde $B_j(T)$ es el conjunto de nodos en el camino dirigido de T entre s y j .

La anterior definición permite obtener un algoritmo del tipo simplex que tiene una complejidad $O(nm)$.

3. Agradecimientos

Este trabajo está parcialmente subvencionado por el Proyecto de Investigación PI042004/078 del Gobierno de Canarias.

4. Bibliografía

- [1] Ahuja, R., T. Magnanti, J.B. Orlin. 1993. *Network Flows*. Prentice-Hall, inc.
- [2] Goldfarb D., Z. Jin. 1999. An $O(nm)$ -Time Network Simplex Algorithm for the Shortest Path

Un sistema multicomponentes reparable en tiempo discreto

Juan Eloy Ruiz Castro¹, Rafael Pérez Ocón², Gemma Fernández-Calvillo Villodre

¹jeloy@ugr.es, ²rperezo@ugr.es

Facultad de Ciencias. Campus Fuentenueva s/n. Departamento de Estadística e Investigación Operativa, Universidad de Granada

Abstract

En este trabajo se desarrolla un sistema de fiabilidad reparable en tiempo discreto con múltiples unidades dispuestas en reserva pasiva. Los tiempos de funcionamiento y de reparación son generales considerando su representación tipo fase. Se calculan las medidas de interés del modelo de forma algorítmica, implementándose computacionalmente con *Matlab*.

Claves:Fiabilidad, Distribución tipo fase, Sistema multicomponentes.

AMS: 90B25, 60J10, 65C40

1. Introducción

El mantenimiento de sistemas reparables en fiabilidad ha sido estudiado por muchos autores. Casi todos los trabajos han sido realizados considerando tiempo continuo en el funcionamiento del sistema. Entre los más recientes está el de Pérez-Ocón and Ruiz-Castro [3] en el que desarrollaron un sistema reparable con dos unidades con distintos tipos de fallo. En cambio, el análisis de sistemas en tiempo discreto ha sido poco tratado. En este trabajo se modeliza un sistema discreto multicomponentes con tiempos de funcionamiento que siguen distribuciones generales. Se construye la matriz de probabilidades de transición, se obtiene explícitamente la distribución estacionaria del modelo y se calculan distintas medidas de efectividad en el caso transitorio y estacionario, todo ello de forma algorítmica, algebraica y matricial. El trabajo ha sido implementado computacionalmente con *Matlab* y se ha aplicado a un ejemplo simulado.

2. El modelo

Se considera un sistema con n componentes, siendo una la principal y las restantes dispuestas en reserva pasiva. Las hipótesis del modelo son similares a las dadas en [2] pero en el caso discreto. El tiempo de funcionamiento de la unidad principal, así como el tiempo de reparación, siguen distribuciones de probabilidad generales discretas e independientes. Es sabido que cualquier distribución de probabilidad discreta tomando valores en N es una distribución tipo fase ([1]). Por ello tomaremos los tiempos de funcionamiento y reparación del sistema distribuidos tipo fase, (α, \mathbf{T}) y (β, \mathbf{S}) respectivamente. El sistema es

modelizado por un proceso vectorial cuyo espacio de macro-estados es similar al dado en [2]. Se ha construido la matriz de probabilidades de transición que tiene estructura de QBD discreto. Se ha calculado explícitamente la distribución estacionaria del proceso, π_k (macro-estado k), obteniéndola de forma matricial y algebraica desde la condición de normalización y las ecuaciones de equilibrio.

3. Medidas de efectividad

Para el modelo desarrollado se han calculado distintas medidas de efectividad de interés en el campo de la fiabilidad, tanto en el caso transitorio como en el caso estacionario. Entre ellas, en el caso estacionario, tenemos,

Disponibilidad: $A = 1 - \pi_n \mathbf{e}$.

Probabilidad Condicional de Fallo:

$$\nu = \pi_1 \left[(\mathbf{T} \otimes \mathbf{S}^0) (\mathbf{I} - \mathbf{T})^{-1} \mathbf{T}^0 + \left(\sum_{i=1}^{n-1} \prod_{k=1}^{i-1} \mathbf{R}_k \right) (\mathbf{T}^0 \otimes \mathbf{e}) \right],$$

siendo \mathbf{e} un vector columna de unos de dimensión apropiada y \mathbf{S}^0 y \mathbf{T}^0 vectores columna cuyos elementos resultan de restar a la unidad la suma de la correspondiente fila de las matrices \mathbf{S} y \mathbf{T} respectivamente.

4. Aplicación

Se ha aplicado el modelo considerando una unidad principal con distribución binomial negativa de parámetros 3 y 0,05, para la reparación una distribución tipo fase de media 48,3333 y tres unidades en reserva.

5. Agradecimientos

El primer y segundo autor agradecen al Ministerio de Ciencia y Tecnología de España la concesión del Proyecto MTM2004-03672, soporte económico del trabajo.

6. Bibliografía

- [1] Alfa, A.S. and Castro, I.T. (2002). Discrete Time Analysis of a Repairable Machine. *Journal of Applied Probability*, **39**, 3, 503-516.
- [2] Pérez-Ocón, Rafael and Montoro-Cazorla, Delia (2004). A multiple system governed by a quasi-birth-and-death process. *Reliability Engineering & System Safety*, **84**, 2, 187-196.
- [3] Pérez-Ocón, Rafael and Ruiz-Castro, Juan Eloy (2004). Two models for a repairable two-system with phase-type sojourn time distributions . *Reliability Engineering & System Safety*, **84**, 3, 253-260.

Una aproximación de programación matemática para la depuración de datos categóricos

Sergio Delgado Quintero¹, Juan José Salazar González²

{¹sdelquin,²jjsalaza}@ull.es, Departamento de Estadística, Investigación Operativa y Computación, Universidad de La Laguna

Abstract

El problema de edición e imputación de una encuesta consiste en corregir con el menor número de cambios posibles incoherencias originadas por unas reglas predefinidas. Presentamos aquí un modelo matemático de tipo “problema de partición de conjuntos” para su resolución. Analizamos la relajación lineal y presentamos resultados computacionales.

Claves: Programación Matemática, Edición e Imputación.

AMS: 90C57, 90C90

1. Introducción

Los datos que provienen de una encuesta estadística son potencialmente erróneos debido a una gran cantidad de fuentes de error (errores de comprensión, errores de escritura, errores de grabado, etc.). Antes de procesar estos datos es necesario llevar a cabo un procedimiento de depuración de los mismos, para eliminar las posibles incoherencias que existan en ellos. Este procedimiento se encarga de resolver el problema de *edición e imputación de datos* que es una fase muy importante dentro del ciclo de vida de una encuesta.

2. Descripción del problema

El conjunto de datos sobre el que se trabaja se puede dividir en unidades de información llamadas *registros*. Un registro corresponde a una instancia de un cuestionario y está formado por unidades más pequeñas de información denominadas *campos*. Un campo corresponde a una instancia de una pregunta y se caracteriza por una *variable*. Cada variable posee un *dominio* (en este caso discreto) con los posibles *valores* que puede asumir. Por ejemplo, la variable *sexo* tiene un dominio con cardinal 2 y puede asumir los valores *varón* ó *hembra*.

Además, en el problema se necesitan una serie de reglas de coherencia entre variables. Estas reglas restringen los posibles valores de unas variables en función de otras. En la terminología inglesa se utiliza el término *edit* para referirse a estas reglas. Por ejemplo, un edit podría ser “*Si (estado_civil = casado) Entonces (edad ≥ 18)*”.

Una vez que evaluamos todos los edits definidos sobre el conjunto de datos

tendremos una serie de incoherencias que deben ser corregidas utilizando el menor número posible de cambios. Este problema se divide en dos fases: en primer lugar detectar qué variables son las que hemos de modificar (*problema de edición*) y en segundo lugar detectar qué valores hemos de dar a esas variables (*problema de imputación*).

3. Modelo matemático

Lo que perseguimos es una transformación del problema original en un modelo matemático que nos permita hacer uso de las herramientas de programación matemática para resolverlo de forma óptima. Partimos de un registro cualquiera r del conjunto de datos de entrada. Este registro está formado por n variables x^1, x^2, \dots, x^n discretas.

Supongamos que tenemos una variable x^i cuyo dominio es $D^i = \{v_1^i, v_2^i, \dots, v_{s_i}^i\}$. Dado que el dominio es discreto podemos asignarle un orden a cada valor de manera consecutiva desde 1 hasta s_i . Esto nos permite obviar los valores y trabajar directamente sobre índices.

Creamos una variable entera $\{0, 1\}$ por cada valor asociado a la variable. Es decir, tendríamos s_i variables tales como $Y^i = \{y_1^i, y_2^i, \dots, y_{s_i}^i\}$, donde y_j^i corresponde al j -ésimo valor de la variable x^i . Además, creamos un vector de pesos $W^i = \{w_1^i, w_2^i, \dots, w_{s_i}^i\}$ asociado a la variable x^i . El modelo final, que no presentamos por motivos de espacio, es de tipo "problema de partición de conjuntos"(y por tanto de programación entera 0-1).

4. Resultados computacionales

El modelo anterior se resuelve con ayuda del código comercial de optimización CPLEX 9. Analizamos las cotas inferiores procedentes de la relajación lineal, y damos desigualdades para mejorarla. Se presentarán y analizarán tablas numéricas que recogen nuestras experiencias computacionales. Los ejemplos resueltos han sido generados aleatoriamente, pero también usamos algunos procedentes de bases de datos reales (no sujetas a secreto estadístico).

5. Agradecimientos

Investigación financiada por el Instituto de Estadística de Canarias.

6. Bibliografía

- [1] Fellegi, I.P., Holt, D., (1976). A systematic approach to automatic edit and imputation. *Journal of the American Statistical Association* 71, 17-35.

Una generalización del modelo con retropedidos parciales de Lee-Nahmias

L.A. San-José¹, J. Sicilia², J. García-Laguna³

¹Departamento de Matemática Aplicada, Universidad de Valladolid

²Departamento de Estadística, I.O. y Computación, Univ. de La Laguna

³Departamento de Estadística e Investigación Operativa, Univ. de Valladolid

¹augusto@mat.uva.es, ²jsicilia@ull.es, ³laguna@eio.uva.es

Abstract

En un trabajo anterior (San-José et. al [8]) se ha estudiado un modelo EOQ con demanda parcialmente acumulable en el que, cuando hay escasez, el comportamiento de los clientes se modeliza por medio de una función de impaciencia escalonada que depende del tiempo de espera. En este trabajo se generaliza tal situación considerando, además, el tiempo transcurrido desde el instante de ruptura del inventario.

Claves: modelos EOQ, retropedidos parciales, función de impaciencia.

AMS: 90B05.

1. Introducción

Los primeros modelos de Teoría de Inventarios en los que se considera demanda parcialmente acumulable fueron los desarrollados por Montgomery et al. [3], Rosenberg [7] y Park [6]. Estos modelos tienen en común que solamente una fracción fija β de la demanda es retropedida y la fracción restante ($1 - \beta$) constituye las ventas perdidas. En posteriores trabajos se ha introducido la modelización de la tasa de demanda retropedida utilizando diferentes enfoques.

Padmanabhan y Vrat [4] fueron los primeros en incorporar el concepto de tasa variable dependiente del tiempo durante el periodo de escasez. Así, ellos consideran que, en un instante dado del periodo de escasez, la tasa de demanda retropedida es una función de la cantidad de órdenes acumuladas en ese instante. En esta misma línea, Padmanabhan y Vrat [5] estudian un modelo de inventario para artículos perecederos con stock dependiente del precio de venta, donde la tasa de ventas es una función del nivel actual de inventario y la tasa de deterioro es constante.

Sin embargo, Abad [1] introduce una nueva perspectiva para la modelización de la tasa de retropedidos parciales teniendo en cuenta el comportamiento de los clientes. Él considera que los clientes son impacientes y supone que la tasa de retropedidos es una función decreciente del tiempo de espera hasta la llegada del siguiente pedido. Además, propone diferentes funciones para modelizar la tasa de retropedidos cuando los clientes son impacientes.

Lee y Nahmias [2] consideran la situación en la cual los clientes permanecen en el sistema si el tiempo que deben esperar es inferior a un valor fijo y, en caso contrario, abandonan el sistema. En San-José et al. [8] se presenta un estudio detallado de este sistema utilizando el concepto de función de impaciencia.

En este trabajo desarrollamos un modelo EOQ con retropedidos parciales en el que la fracción de demanda retropedida se describe a través de una función de comportamiento de clientes que depende de la cantidad de tiempo que el cliente debe o debería esperar hasta la llegada del siguiente pedido y del tiempo transcurrido desde que se produjo la rotura del stock. Se introduce, por tanto, una nueva perspectiva que generaliza los enfoques anteriormente citados. Más concretamente, la función de comportamiento que se considera es una función escalonada que generaliza la citada en Lee y Nahmias [2].

2. Bibliografía

- [1] Abad P.L. (1996). Optimal Pricing and Lot-sizing Under Conditions of Perishability and Partial Backordering. *Management Science* 42, 1093-1104.
- [2] Lee H.L. y Nahmias S. (1993). Single-Product Single-Location Models, in Graves S.C., Rinnoy A.H.C. & Zipkin P.H.(eds). *Handbooks in Operations Research and Management Science*, Vol. 4, Logistics of Production and Inventory. North-Holland, 3-55.
- [3] Montgomery D.C., Bazaraa M.S. y Keswani A.K. (1973). Inventory models with a mixture of backorders and lost sales. *Naval Research Logistics Quaterly* 20, 255-263.
- [4] Padmanabhan G. y Vrat P. (1990). Inventory model with a mixture of backorders and lost sales. *International Journal of Systems Sciences* 21, 1721-1726.
- [5] Padmanabhan G. y Vrat P. (1995). EOQ models for perishable items under stock dependent selling rate. *European Journal of Operational Research* 86, 281-292.
- [6] Park K.S. (1982). Inventory model with partial backorders. *International Journal of Systems Sciences* 13, 1313-1317.
- [7] Rosenberg D. (1979). A new analysis of a lot size model with partial backlogging. *Naval Research Logistics Quaterly* 26, 346-353.
- [8] San-José L.A., Sicilia J. y García-Laguna J. (2005). The lot size-reorder level inventory system with customers impatience functions. *Computers and Industrial Engineering*, 49, 349-362.

Una nueva herramienta software para resolver problemas de localización multicriterio sobre redes

Marcos Colebrook, Sergio Alonso, Joaquín Sicilia

[mcolesan, salonso, jsicilia]@ull.es, Departamento de Estadística, Investigación Operativa y Computación, Universidad de La Laguna

Abstract

Presentamos una nueva aplicación de software que recopila varios algoritmos para resolver problemas de localización de servicios deseados y no deseados sobre redes multicriterio. Por medio de una interfaz gráfica, el programa permite modelar y editar los datos de entrada, así como mostrar la solución óptima sobre la red.

Palabras Clave: Localización, Decisión Multicriterio, Grafos y redes

1. Introducción

En los últimos años, nuestro grupo de investigación ha desarrollado varios algoritmos que resuelven problemas de localización multicriterio/multiobjetivo sobre redes considerando varios pesos en los nodos y varias longitudes en las aristas ([1]-[4]). En este trabajo presentamos una aplicación software que nos permite de forma sencilla e interactiva diseñar una red dibujando sus nodos y aristas, ejecutar los algoritmos directamente, mostrar la solución gráfica sobre la red y los valores numéricos en la terminal, y poder realizar una traza mostrando las diferentes funciones objetivo.

2. Problemas implementados

Sea $N = (V, E)$ una red no dirigida, simple y conexa, con $V = \{v_1, \dots, v_n\}$ y $E = \{(v_s, v_t); v_s, v_t \in V\}$, con $m = |E|$.^o Sobre cada nodo $v_i \in V$ se define un vector de p pesos $w(v_i) = (w_i^1, \dots, w_i^p)$ y sobre cada arista $e = (v_s, v_t) \in E$ definimos un vector de q longitudes $l(e) = \{l_e^1, \dots, l_e^q\}$, con $k = p \times q$. Se define $d^r(v_a, v_b)$ como la longitud del camino mínimo entre los nodos v_a y v_b considerando la longitud l_e^r , con $1 \leq r \leq q$. Dada una arista $e = (v_s, v_t) \in E$, un punto interno a ella $x \in e$, y un nodo $v_i \in V$, se define $d^r(x, v_i) = \{c_e^r(x, v_s) + d^r(v_s, v_i), l_e^r - c_e^r(x, v_s) + d^r(v_t, v_i)\}$, donde $c_e^r(x, v_s)$ denota la longitud del segmento de línea en la arista e entre el punto x y el nodo v_s considerando la longitud l_e^r . Dado un punto $x \in N$, un índice de peso $1 \leq s \leq p$, y un índice de longitud $1 \leq r \leq q$, definimos las siguientes funciones objetivo: $f_{\min}^{sr}(x) = \min_{v_i \in V} w_i^s d^r(x, v_i)$, $f_{\sum}^{sr}(x) = \sum_{v_i \in V} w_i^s d^r(x, v_i)$, $f_{\max}^{sr}(x) = \max_{v_i \in V} w_i^s d^r(x, v_i)$, $f_{cd}^{sr}(\lambda, x) = \lambda \max_{v_i \in V} d^r(x, v_i) + \frac{(1-\lambda)}{\sum_{v_i \in V} w_i^s} \sum_{v_i \in V} w_i^s d^r(x, v_i)$ y $f_{acd}^{sr}(\lambda, x) = \lambda f_{\min}^{sr}(x) + (1-\lambda) f_{\sum}^{sr}(x)$.

$(1 - \lambda)f_{\text{sum}}^{sr}(x)$. Los problemas implementados son, entre otros:

Problema	Tipo
$\min_{x \in N} (f_{\max}^{11}(x), f_{\max}^{12}(x), \star, f_{\max}^{pq}(x)) \in \S^k$	Centro Multiobjetivo (<i>minimax</i>)
$\min_{x \in N} (f_{\text{sum}}^{11}(x), f_{\text{sum}}^{12}(x), \star, f_{\text{sum}}^{pq}(x)) \in \S^k$	Mediana Multiobjetivo (<i>minisum</i>)
$\min_{x \in N} (f_{\text{cd}}^{11}(\lambda, x), f_{\text{cd}}^{12}(\lambda, x), \star, f_{\text{cd}}^{pq}(\lambda, x)) \in \S^k$	Cent-Dian Multiobjetivo
$\max_{x \in N} (f_{\min}^{11}(x), f_{\min}^{12}(x), \star, f_{\min}^{pq}(x)) \in \S^k$	Anti-Centro Multiobjetivo (<i>maxmin</i>)
$\max_{x \in N} (f_{\text{sum}}^{11}(x), f_{\text{sum}}^{12}(x), \star, f_{\text{sum}}^{pq}(x)) \in \S^k$	Anti-Mediana Multiobjetivo (<i>maxisum</i>)
$\max_{x \in N} (f_{\text{acd}}^{11}(\lambda, x), f_{\text{acd}}^{12}(\lambda, x), \star, f_{\text{acd}}^{pq}(\lambda, x)) \in \S^k$	Anti-Cent-Dian Multiobjetivo

3. Agradecimientos

Este trabajo ha sido financiado parcialmente por el Ministerio de Educación y Ciencia con el proyecto de investigación MTM2004-07550.

4. Bibliografía

- [1] Ramos, R.M., Ramos, M.T., Colebrook, M. y Sicilia, J. (1999). Locating a facility on a network with multiple median-type objectives. *Annals of Operations Research* 86, 221-235.
- [2] Colebrook, M. y Sicilia, J. (2004). A polynomial algorithm for the multicriteria cent-dian location problem. *European Journal of Operational Research* (en prensa).
- [3] Colebrook, M. y Sicilia, J. (2003). Undesirable facility location problems on multicriteria networks. *Computers & Operations Research* (en prensa).
- [4] Colebrook, M. (2003). *Localización simple de servicios deseados y no deseados en redes con múltiples criterios*. Tesis Doctoral. Editado por el Servicio de Publicaciones de la Universidad de La Laguna, ISBN: 84-688-5038-1.

Unequal Probability Selection in multivariate finite population sampling. A Vector-Optimization approach

Emilio Carrizosa¹

¹ecarrizosa@us.es, Departamento de Estadística e Investigación Operativa,
Universidad de Sevilla

Abstract

The Hansen-Hurwitz estimator for estimating a population mean of a positive variable yields minimal variance when individuals are selected with probabilities proportional to the variable under study.

In this note we address the problem of determining selection probabilities in a multivariate setting, when the aim is the simultaneous minimization of variances for each variable under study, and a superpopulation model is considered. As an application, we describe the problem encountered in Auditing, where both the fraction of misstatements, and the average amount of such misstatements are of interest.

Keywords: Finite population sampling. Pareto-optimality.

1. Introduction

Let $\mathcal{U} = \{u_1, \dots, u_N\}$ be a population. Associated with each u_i we have an r -valued vector $Y_i = (Y_{i1}, Y_{i2}, \dots, Y_{ir})$. The vector ϑ of means,

$$\vartheta = \left(\frac{1}{N} \sum_{i=1}^N Y_{i1}, \frac{1}{N} \sum_{i=1}^N Y_{i2}, \dots, \frac{1}{N} \sum_{i=1}^N Y_{ir} \right),$$

is estimated by drawing from \mathcal{U} a sample with replacement of size n , and considering as estimator $\hat{\vartheta}$ the r -dimensional Hansen-Hurwitz estimator, [2],

$$\hat{\vartheta} = \left(\frac{1}{N} \sum_{i=1}^N \frac{Y_{i1} f_i}{n \alpha_i}, \frac{1}{N} \sum_{i=1}^N \frac{Y_{i2} f_i}{n \alpha_i}, \dots, \frac{1}{N} \sum_{i=1}^N \frac{Y_{ir} f_i}{n \alpha_i} \right),$$

where f_i denotes the frequency of u_i in the sample, and α_i denotes the probability of selection of u_i at each draw.

Assume a superpopulation model, in which Y_{ij} are independent random variables, with expected value $E(Y_{ij}) = \mu_{ij} < +\infty$ and variance $\text{var}(Y_{ij}) = \sigma_{ij}^2 \geq 0$. For simplicity we assume in what follows that each Y_{ij} is not degenerate to zero, thus

$$\mu_{ij}^2 + \sigma_{ij}^2 > 0 \quad \forall i, j.$$

For a given vector $(\alpha_1, \dots, \alpha_N)$ of selection probabilities, the expected squared error $\varepsilon_j(\alpha_1, \dots, \alpha_N)$ in the j -th component is given by $\varepsilon_j(\alpha_1, \dots, \alpha_N)$,

$$\varepsilon_j(\alpha_1, \dots, \alpha_N) = E \left(\left(\frac{1}{N} \sum_{i=1}^N Y_{ij} - \frac{1}{N} \sum_{i=1}^N \frac{Y_{ij} f_i}{n \alpha_i} \right)^2 \right)$$

Different possible strategies to determine $(\alpha_1, \dots, \alpha_N)$ taking into account all errors ε_j exist. We follow a multicriteria approach, and seek selection probabilities $\alpha_1, \dots, \alpha_N$ minimizing simultaneously the r criteria. In other words, we consider the nonlinear multicriteria optimization problem

$$\begin{aligned} \min & \quad (\varepsilon_1(\alpha_1, \dots, \alpha_N), \varepsilon_2(\alpha_1, \dots, \alpha_N), \dots, \varepsilon_r(\alpha_1, \dots, \alpha_N)) \\ \text{s.t.} & \quad \sum_{j=1}^N \alpha_j = 1 \\ & \quad \alpha_j \geq 0, \quad j = 1, \dots, r, \end{aligned} \tag{4.86}$$

and we seek the set \mathcal{P} of *Pareto-optimal* solutions to (4.86). See e.g. [1, 3, 4, 6] for other multiple-objective design problems.

2. Acknowledgement

Research supported by Ministerio de Educación y Ciencia (MTM2005-09362-C03-01) and Junta de Andalucía (FQM-329).

3. Bibliography

- [1] J.E. GENTLE, S.C. NARULA, AND R.L. VALLIANT "Multicriteria Optimization in Sampling Design". *Statistics of Quality* (S. Ghosh, W. Schucany, and T. Smith, Eds.) pp 411–425. Marcel-Dekker, New York, 1997.
- [2] HANSEN, M.M., AND W.N. HURWITZ "On the theory of sampling from finite populations". *Annals of Mathematical Statistics* **14** (1943) 333–362.
- [3] IMHOF,L., AND W. K. WONG "A Graphical Method for Finding Maximin Efficiency Designs". *Biometrics* **56** (2000) 113–117.
- [4] MALEC, D. "Selecting multiple-objective fixed-cost sample designs using an admissibility criterion". *Journal of Statistical Planning and Inference* **48** (1995) 229–240.
- [5] MAYOR, J.A. "Optimal cluster selection probabilities to estimate the finite population distribution function under PPS cluster sampling" *Test* **11** (2002) 73–88.
- [6] WONG, W.K. "Recent advances in multiple-objective design strategies" *Statistica Neerlandica* **53** (1999) 257–276.

Who gains and how much by cooperation in decentralized supply chains

Ana Meca

ana.meca@umh.es, Operations Research Center, Miguel Hernández University of Elche

Abstract

Estudiamos la coordinación de acciones y el reparto de beneficios en cadenas de distribución bajo control descentralizado en las que un solo distribuidor abastece a varios comerciantes al por menor, los cuales operan en sus propios mercados separados. Puesto que dichos sistemas de distribución resultan ser ineficientes, la cooperación entre las empresas mediante la coordinación de sus acciones permite mejorar los beneficios individuales.

Keywords: Decentralized supply chain, cooperation, gain.

AMS: 91A12, 90B99

1. Introduction

We study single period distribution chains with a single product. In such a supply chain, retailers place one-time orders for the product at the supplier. After production, the supplier delivers the goods to the retailers via a warehouse. This warehouse acts as an intermediary without costs or revenues. When the goods arrive, the non-competing retailers sell these on their own separate markets. The larger the quantity that is put on the market, the lower the expected revenue per unit for the retailer. Each retailer chooses its order quantity such that its profit is maximized.

The retailer pays the supplier a wholesale price per unit product ordered and delivered. This price is a decreasing function of the quantity ordered. Hence, incentives for cooperation among retailers exist. If the retailers combine their orders into one large order then they enjoy a lower wholesale price per unit. They can do so because the warehouse only informs the supplier about the quantities ordered and not about which retailer orders how much. Besides, retailers may want to cooperate with the supplier which implies a further reduction in cost inefficiency due to the absence of the intermediate wholesale prices. Obviously, the total profit under full cooperation is larger than the sum of the individual profits.

Because of the incentives for cooperation, we use cooperative game theory to study these distribution chains. For each chain we define a corresponding cooperative game in which the supplier and the retailers are the players. The

value of a coalition of players equals the optimal joint profit they can achieve. We show that the core of such a game is never empty, that is, all companies in the chain are willing to cooperate because there exists stable distributions of the total profit among the companies upon which no coalition can improve. Any distribution of profits that belongs to the core has a nice interpretation in terms of the underlying distribution chain. Further, we introduce a specific allocation of the total profit for distribution chains, the so-called minimal-gain-per-capita (mgpc) solution. This solution is a stable distribution of the profits, that is, it always belongs to the core of the game, and it possesses several nice properties. In particular, it takes into account the importance of the supplier to achieve full cooperation. Finally, a characterization of the mgpc-solution is provided.

Our contribution to the literature is twofold. First, we include the supplier in our analysis and study cooperation among retailers and the supplier. Second, we introduce a tailor-made allocation of the joint profit that always belongs to the core of the game. The existing literature only considers cooperation among retailers and hardly pays attention to suitable allocations of the joint benefits.

Sección 5

Probabilidad

A closed-form expression for the quantile function of the $\Gamma(2, \lambda)$ distribution

José A. Adell, P. Jodrá

Dpto. de Métodos Estadísticos, Universidad de Zaragoza, Spain
 {adell,pjodra}@unizar.es

Abstract

We provide an explicit formula for the quantile function of the $\Gamma(2, \lambda)$ distribution, $\lambda > 0$, which is related to the Lambert W function.

Keywords: Erlang distribution, Lambert W function, quantiles.

AMS: 60E05, 33B30

1. Introduction

For any $n = 1, 2, \dots$, let X_n be a $\Gamma(n, 1)$ distribution. Denote by $F_n(x)$ its distribution function, $x > 0$, and by $Q_n(u)$ its quantile function, $0 < u < 1$. As far as we know, in the literature only explicit solutions to the equation $F_n(x) = u$ are given for $n = 1$, thus obtaining $Q_1(u) = -\log(1 - u)$ for the standard exponential distribution. For $n > 1$, a number of numerical algorithms are available for computing the quantiles of the $\Gamma(n, 1)$ distribution (see Phien [4] for a survey).

The aim of this note is to provide a closed-form expression for the quantile function of the $\Gamma(2, \lambda)$ distribution. It is interesting to observe that such expression involves the Lambert W function. The Lambert W function is defined as the solution to the equation

$$W(x)e^{W(x)} = x, \quad x \geq -1/e. \quad (5.87)$$

For real values of $W(x)$, the range of x is limited to $x \geq -1/e$ (see Figure 1). For $-1/e \leq x < 0$, there are two possible real branches of $W(x)$. We shall only be interested in the branch taking on values in $(-\infty, -1]$, denoted by $W_{-1}(x)$. It is known that $W_{-1}(-1/e) = -1$, $W_{-1}(x)$ is decreasing and that $W_{-1}(x) \rightarrow -\infty$ as $x \rightarrow 0$. A review of this function and its applications in applied and pure mathematics may be found in Corless et al. [2], Barry et al. [1], Hayes [3], and the references therein.

With the preceding notations, we state the following.

Proposition. The $\Gamma(2, 1)$ distribution has quantile function

$$Q_2(u) = -W_{-1}(-(1 - u) e^{-1}) - 1, \quad 0 < u < 1. \quad (5.88)$$

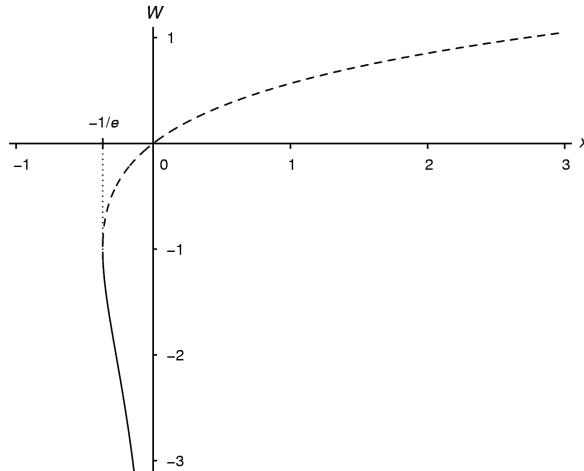


Figure 1: The two real branches of $W(x)$. ———, $W_{-1}(x)$; ——, $W_0(x)$.

Moreover, it is well known that $\lambda Q_{n;\lambda}(u) = Q_{n;1}(u)$, where $Q_{n;\lambda}(u)$ denotes the quantile function of the $\Gamma(n, \lambda)$ distribution. As a consequence, $Q_{2;\lambda}(u)$ can be directly obtained from (5.88).

Remark. Formula (5.88) gives us an explicit form for $Q_2(u)$ in terms of the Lambert W function. From a computational point of view, since this function is implemented in various computer algebra systems, $Q_2(u)$ can be evaluated in a straightforward manner. In addition, some basic properties of $Q_2(u)$ can be established from the analytical expression (5.88). However, we do not know any function implemented in some computer algebra system in terms of which $Q_n(u)$ could be expressed whenever $n > 2$.

2. Bibliography

- [1] D.A. Barry, J.-Y. Parlange, L. Li, H. Prommer, C.J. Cunningham and F. Stagnitti (2000). Analytical approximations for real values of the Lambert W -function. *Math. Comput. Simulation* **53**, 95–103.
- [2] R.M. Corless, G.H. Gonnet, D.E.G. Hare, D.J. Jeffrey and D.E. Knuth (1996). On the Lambert W function. *Adv. Comput. Math.* **5**, 329–359.
- [3] B. Hayes (2005). Why W ? *Amer. Sci.* **93**, 104–108.
- [4] H.N. Phien (1991). On the computation of gamma quantiles. *The Frontiers of Statistical Computation, Simulation and Modeling* (Vol. 1 of the Proceedings of ICOSCO-I), 253–269.

Berry–Esseen Bounds for Standardized Lévy Processes via Moduli of Smoothness

José A. Adell¹, Alberto Lekuona²

¹adell@unizar.es, Departamento de Métodos Estadísticos, Universidad de Zaragoza

²lekuona@unizar.es, Departamento de Métodos Estadísticos, Universidad de Zaragoza

Abstract

We introduce moduli of smoothness techniques to deal with Berry–Esseen bounds, and illustrate them by considering standardized Lévy processes with finite variance. Instead of the classical Berry–Esseen smoothing inequality, we give an easy inequality involving the second modulus of smoothness.

Keywords: Berry–Esseen bounds; Lévy process; moduli of smoothness

AMS: AMS 60F05

1. Introduction

We obtain Berry–Esseen bounds for standardized Lévy processes with finite variance by means of an easy inequality involving the second modulus. Under finite third moment assumptions, such an inequality provides the main term of the approximation with small constants, even asymptotically sharp constants in the lattice case. Under infinite third moment assumptions, we show that the optimal rate of convergence can be simply written in terms of the first modulus of smoothness of an appropriate function, depending on the characteristic random variable of the Lévy process.

We believe that the use of moduli of smoothness in dealing with Berry–Esseen bounds may clarify the proofs, on the one hand, and may allow us to obtain explicit small constants, on the other, without paying an extra effort.

2. Acknowledgements

This work has been supported by research grants MTM2005–08376–C02–01 and DGA E–64, and by FEDER funds.

Campos de difusión biparamétricos no constantes en los ejes. El caso lognormal

R. Gutiérrez¹, C. Roldán²

¹rgjaimez@ugr.es, Departamento de Estadística e Investigación Operativa,
Universidad de Granada

²iroldan@ujaen.es, Departamento de Estadística e Investigación Operativa,
Universidad de Jaén

Abstract

En este trabajo se considera una extensión de la definición de campo de difusión introducida por D. Nualart (1983) que nos lleva a un modelo que podría considerarse más realista en la representación de distintas situaciones y problemas de Ciencias Medioambientales, tales como la difusión de contaminantes en la atmósfera. Vamos a considerar campos aleatorios que son difusiones lognormales en los ejes del espacio paramétrico y que toman un valor constante en el origen. Fijadas las condiciones iniciales, los campos de difusión lognormales se introducen a partir la familia de densidades de transición, se estudian distintas propiedades, se calculan los momentos y se obtienen las ecuaciones de Kolmogorov.

Claves: Proceso de difusión lognormal, Campo aleatorio de difusión, Ecuaciones de Kolmogorov.

AMS: 60J60

1. Introducción

Las técnicas de predicción y simulación para difusiones lognormales biparamétricas que desarrollamos en Gutierrez, Roldán, Gutiérrez-Sánchez and Angulo (2005) se basan en la definición de difusión que introdujo Nualart (1983). El mencionado trabajo describe técnicas para la estimación y predicción de campos de difusión que son constantes en los ejes del espacio paramétrico. Sin embargo, al buscar situaciones reales en las que poder aplicar los resultados obtenidos nos pareció extraña una situación real en la que la difusión se comporta como un valor constante en los ejes del espacio.

En este trabajo vamos a considerar una extensión de la definición de campo de difusión al considerar campos aleatorios que son difusiones lognormales en los ejes del espacio paramétrico y que toman un valor constante en el origen. Técnicamente, vamos a considerar campos de Markov biparamétricos que se comportan como procesos de Markov (uniparamétricos) en los ejes del espacio paramétrico. Partiendo de estos campos y considerando trayectorias muestralmente continuas vamos a introducir un concepto de

difusión biparamétrica orientado al estudio de difusiones Gaussianas y lognormales.

Para establecer la definición de difusión biparamétrica nos hemos basado en las condiciones que hay que exigir para concluir que el campo sea Gaussiano (equivalentemente lognormal) y en la definición que dieron Nualart y Sanz (1979). No impondremos que el campo se comporte como una difusión uniparamétrica en cada coordenada (como hicieron Nualart y Sanz) y sólo necesitaremos que el campo se modelice como una difusión uniparamétrica en los ejes del espacio paramétrico con un valor constante en el origen. Además, sólo trabajaremos con los coeficientes de tendencia y difusión biparamétricos, ya que, las densidades de transición de los campos de difusión Gaussianos y lognormales están caracterizadas por dichos coeficientes. No se consideran los coeficientes de difusión mixtos ya que, para estos campos, éstos se podrían obtener a partir de los coeficientes de tendencia y difusión biparamétricos.

A partir de la definición se estudian distintas propiedades de los campos lognormales, se obtienen sus momentos condicionados y no condicionados y utilizando la metodología de Ricciardi (1977), se obtienen las ecuaciones adelantada y atrasada de Kolmogorov.

2. Agradecimientos

Este trabajo ha sido parcialmente financiado por el Ministerio de Educación y Ciencia, Proyecto de investigación MTM2005-09209.

3. Bibliografía

- [1] Gutierrez R. and Roldán C. (2006) Some analytical and statistical aspects related to 2D lognormal diffusion random fields. *Journal of Applied Probability*. Sometido para publicación.
- [2] Gutierrez R., Roldán C., Gutiérrez-Sánchez R. and Angulo J.M. (2005) Estimation and prediction of a 2D lognormal diffusion random field. *Stochastic Environmental Research and Risk Assessment* 19(4), 258-265.
- [3] Nualart D. (1983) Two-parameter diffusion processes and martingales. *Stochastic Processes and their Applications* 15, 31-57.
- [4] Nualart D. and Sanz M. (1979) A Markov property for two-parameter Gaussian processes. *Stochastica* III(1), 1-16.

Caracterización de distribuciones de vida mediante la ordenación estocástica de variables doblemente truncadas

Miguel A. Sordo¹

¹mangel.sordo@uca.es, Departamento de Estadística e Investigación Operativa, Universidad de Cádiz

Abstract

En este trabajo se caracterizan varias clases de distribuciones de vida (incluyendo las clases IFR, DFR y DRHR), de interés en fiabilidad, en términos de la ordenación estocástica de variables aleatorias doblemente truncadas. Estas caracterizaciones formalizan algunas propiedades intuitivas de las clases consideradas.

Palabras Clave: Ordenes estocásticos, variables truncadas, distribuciones de vida

AMS: 60E10, 60E15

1. Introducción

Sea X una variable aleatoria con función de distribución $F(t)$. En fiabilidad, cuando X describe la vida de una unidad, la variable aleatoria $X_t = [X - t \mid X > t]$, $t > 0$, representa la vida residual de la unidad en el instante t , dado que la unidad que ha sobrevivido a dicho instante t . En la literatura, la ordenación estocástica de las vidas residuales en diferentes instantes ha permitido caracterizar diversas clases de distribuciones de vida mediante resultados de la forma:

$$X \in C \iff X_s \prec X_t, \forall s < t \quad (5.89)$$

siendo C la clase considerada y \prec un cierto orden estocástico. Una variable relacionada con la vida residual X_t es la variable truncada por la izquierda en el instante t , denotada por $X_{[t,\infty)}$ y dada por $X_{[t,\infty)} = X_t - t$. Algunas caracterizaciones del tipo (1) pueden formularse en términos de variables truncadas por la izquierda, dando lugar a resultados de la forma

$$X \in C \iff X_{[s,\infty)} \prec X_{[t,\infty)}, \forall s < t. \quad (5.90)$$

Numerosos resultados del tipo (1) y (2) pueden encontrarse en el libro de Shaked y Shanthikumar (1994) y en los trabajos de Belzunce, Candel y Ruiz (1996), Pellerey y Shaked (1997), Belzunce (1999) y Fernández-Ponce y Suárez-Llorens (2003), entre otros.

Otra variable de interés en fiabilidad es la variable doblemente truncada en los instantes s y t , dada por $X_{[s,t]} = [X \mid s \leq X \leq t]$, que representa la vida de

una unidad que falla en el intervalo $[s, t]$. Esta variable ha sido considerada en este contexto por Navarro, Belzunce y Ruiz (1997). Hay que destacar que no se encuentran en la literatura caracterizaciones de distribuciones de vida en términos de variables doblemente truncadas.

2. Resultados

En este trabajo se caracterizan varias clases de distribuciones de vida mediante la ordenación estocástica de variables doblemente truncadas. Los resultados son de la forma:

$$X \in C \iff X_{[s,t]} \prec X_{[s',t']} , \forall [s,t] \text{ y } [s',t'] \in D,$$

siendo C una clase de distribuciones de vida, \prec un cierto orden estocástico y D un conjunto de restricciones que deben satisfacer los intervalos $[s, t]$ y $[s', t']$. Estas caracterizaciones formalizan ciertas propiedades intuitivas de las clases de distribuciones consideradas, entre las que se encuentran las clases IFR (increasing failure rate), DFR (decreasing failure rate) y DRHR (decreasing reversed hazard rate).

3. Agradecimientos

Este trabajo ha sido parcialmente subvencionado por el Proyecto SEJ2005-06678 del MEC.

4. Bibliografía

- [1] Belzunce, F., Candel, J and Ruiz (1996). Dispersive orderings and characterizations of ageing classes. *Statistics and Probability Letters* 28, 321-327.
- [2] Belzunce, F. (1999). On a characterization of right spread order by the increasing convex order. *Statistics and Probability Letters* 45, 103-110.
- [3] Fernández-Ponce, J. M. and Suárez-Llorens, A. (2003). An aging concept based on majorization. *Probability in the Engineering and the Informational Sciences* 17, 107-117.
- [4] Navarro, J., Belzunce, F and Ruiz, J. M. (1997). New stochastic orders based on double truncation. *Probability in the Engineering and the Informational Sciences* 11, 395-402.
- [5] Pellerey, F and Shaked, M. (1997). Characterizations of the IFR and DFR aging notions by means of the dispersive order. *Statistics and Probability Letters* 33, 389-393.
- [6] Shaked, M. Shanthikumar, J. G. (1994). Stochastic Orders and their Applications. Academic Press, San Diego, CA.

Cercanía de la distribución de un vector aleatorio a la de un conjunto aleatorio

I. Cascos Fernández¹, M. López-Díaz²

¹Dpto. de Estadística. Universidad Carlos III de Madrid,
e-mail:ignacio.cascos@uc3m.es

²Dpto. de Estadística e I.O. y D.M. Universidad de Oviedo,
e-mail:mld@uniovi.es

Abstract

El objetivo de esta comunicación es introducir el concepto de función de proximidad asociada a un funcional de capacidad.

Claves: Función de proximidad, Funcional de capacidad.

1. Introducción

Mediante este trabajo tratamos de cuantificar la proximidad de la distribución de probabilidad inducida por un vector aleatorio en \mathbf{R}^d a la distribución de probabilidad inducida por un conjunto aleatorio también en \mathbf{R}^d . La distribución del conjunto aleatorio viene determinada por un funcional de capacidad y por tanto, se introduce la noción de función de proximidad asociada a un funcional de capacidad como una aplicación definida sobre el espacio de probabilidades sobre la σ -álgebra de Borel. Dicha aplicación deberá satisfacer ciertas propiedades que la hagan apropiada para cuantificar la proximidad de una probabilidad al citado funcional de capacidad.

En concreto, las propiedades que se exigen a una función de proximidad son las siguientes:

- *invarianza por isometría:* dado el funcional de capacidad de un conjunto aleatorio, la probabilidad inducida por un vector aleatorio y una transformación ortogonal, la proximidad de la probabilidad a la capacidad debe coincidir con la proximidad de la probabilidad inducida por la transformación ortogonal del vector aleatorio al funcional de capacidad de la transformación ortogonal del conjunto aleatorio,
- *maximalidad:* dada una probabilidad con soporte en aquellos puntos que son “fijos” para el conjunto aleatorio con dicho funcional de capacidad, su proximidad ha de ser máxima,
- *acotación sobre mixturas:* dada una mezcla de dos probabilidades, su proximidad a un funcional de capacidad no puede ser inferior a la mínima de las proximidades de tales probabilidades a dicho funcional,

- *minimalidad*: dado un funcional de capacidad “acotado” en algún sentido y una sucesión de probabilidades que se “aleja del origen”, entonces la sucesión de proximidades debe converger a cero.

Utilizando una función de proximidad, podemos ordenar distribuciones de probabilidad según su proximidad respecto de un funcional de capacidad. Esto nos permitirá desarrollar diversas aplicaciones en problemas reales.

En la comunicación se plantearán dos funciones de proximidad particulares, se estudiarán sus propiedades principales y se analizará su comportamiento empírico. Debe indicarse que dicho comportamiento puede analizarse desde una doble perspectiva, por un lado desde el punto de vista de la distribución de probabilidad y por otra parte desde el funcional de capacidad.

2. Agradecimientos

Los autores desean agradecer la ayuda económica recibida de los proyectos del Ministerio de Ciencia y Tecnología MCT-BFM2002-01057, MCT-BFM2002-03263 y MTM2005-02254.

Characterization of Distributions through Martingales

R. Gouet¹, F.J. López², G. Sanz³

¹rgouet@dim.uchile.cl, Departamento de Ingeniería Matemática. Centro de Modelamiento Matemático. Universidad de Chile.

²javier.lopez@unizar.es, ³gerardo@unizar.es, Departamento de Métodos Estadísticos, Universidad de Zaragoza

Abstract

We consider the problem of finding all distribution functions F such that $N_n - cM_n$ is a discrete time martingale, where N_n is the counting process of upper records, $M_n = \max\{X_1, \dots, X_n\}$ is the process of partial maxima and c is a positive constant.

Keywords: Characterization of distributions, martingales.

1. Introduction

Let (X_n) be a sequence of nonnegative, integrable, independent and identically distributed random variables, with distribution function F . For $n \geq 1$, let $M_n = \max\{X_1, \dots, X_n\}$ be the process of partial maxima, $I_n = \mathbb{1}_{\{X_n > M_{n-1}\}}$ the indicator of an upper record and $N_n = \sum_{k=1}^n I_k$, $n \geq 1$, the counting process of records, with $M_0 = -1$. It is easy to check that $N_n - cM_n$ is a martingale with respect to the natural filtration when the X_n are exponential or geometric, with $c > 0$ depending on the distribution. One interesting application of the martingale $N_n - pM_n$ is the derivation of a central limit theorem for the counting process of records of geometric random variables. Details and extensions to other discrete models can be found in Gouet, López and Sanz (2005). Our aim in this paper is to solve the following problem:

P_c : Determine all distributions F such that $N_n - cM_n$ is a martingale, for a given constant $c > 0$.

In contrast to most characterization results, we do not impose any restriction on the class of distribution functions under consideration aside from integrability, necessary to define conditional expectations.

For a closed set $A \subseteq \mathbb{R}^+ = [0, \infty)$, let \mathcal{I} be the family of open disjoint intervals such that $\cup_{I \in \mathcal{I}} I = \mathbb{R} \setminus A$; the contiguous intervals of A are the bounded intervals in \mathcal{I} . Also, let $0 \leq \alpha_A = \min(A)$ and $\omega_A = \sup(A) \leq \infty$.

The support of F , given by $\{x \in \mathbb{R} \mid F(x + \epsilon) - F(x - \epsilon) > 0, \forall \epsilon > 0\}$ (which coincides with the smallest closed set A such that $F(A) = 1$) is denoted by $\text{supp}(F)$. The α, ω notation introduced above for closed sets A is simplified

when $A = \text{supp}(F)$, writing α_F or ω_F in place of $\alpha_{\text{supp}(F)}$ or $\omega_{\text{supp}(F)}$. When F is a solution of \mathbf{P}_c we write $F \in \mathbf{P}_c$.

2. Main results

In our main result (Theorem 5) we completely solve problem \mathbf{P}_c . The trivial solution with support $A = \{\alpha_A\}$ is excluded from Theorem 5 below, considering only closed sets A such that $\alpha_A < \omega_A \leq \infty$.

Theorem 5 *Let $c > 0$. Consider the family \mathcal{A} of closed subsets $A \subseteq [0, \infty)$ such that either*

(a) *A is unbounded and the contiguous intervals have length $< 1/c$*

or

(b) *A is bounded, $(\omega_A - 1/c, \omega_A)$ is a contiguous interval and all other contiguous intervals have length $< 1/c$.*

Then $(p, A) \rightarrow F_{p,A}$ is a parameterization of \mathbf{P}_c by $[0, 1) \times \mathcal{A}$, where $\text{supp}(F_{p,A}) = A$ and $F_{p,A}(\alpha_A) = p$.

Sketch of the proof.- We prove the existence of solutions starting from the case where A is a set of isolated points. The next step towards the general case is the construction of a sequence of discrete distributions solving \mathbf{P}_c , defined on adequate subsets of A . Finally, we show that the previously constructed sequence converges weakly to a distribution function which is proved to be the unique solution of \mathbf{P}_c .

Theorem 6 *Let $c > 0$. If $F \in \mathbf{P}_c$ has no atoms then F is the α -shifted exponential distribution with parameter c , for some $\alpha \geq 0$. That is,*

$$F(x) = F_{0,[\alpha,\infty)}(x) = (1 - e^{-c(x-\alpha)}) \mathbf{1}_{\{x \geq \alpha\}}.$$

3. Acknowledgements

We gratefully acknowledge financial support from the FONDAP Project in Applied Mathematics, FONDECYT grants 1020836, 7020836 (CHILE) and project MTM2004-01175 of MEC (SPAIN). The authors are members of the research group Modelos Estocásticos (DGA).

4. References

GOUET, R.; LÓPEZ, F.J. AND SANZ, G. (2005). Central limit theorem for the number of records in discrete models. *Adv. Appl. Prob.* **37**, 781–800.

GOUET, R.; LÓPEZ, F.J. AND SANZ, G. (2006). A characteristic martingale related to the counting process of records. To appear in *J. Theor. Probab.*

Comparación Dispersiva de Distribuciones Wishart

J.M. Fernández Ponce¹, M.T. Gómez Gómez¹, J.L. Pino Mejías¹,
Rosario Rodríguez Griñolo¹

¹Departamento de Estadística e Investigación Operativa ,

Universidad de Sevilla

ferpon@us.es, terego@us.es, jlpino@us.es, grinolo@us.es

Abstract

El orden dispersivo univariante ha sido extensamente caracterizado en las últimas décadas. Sin embargo, la versión multivariante carece de un estudio tan amplio. En este trabajo, presentamos condiciones necesarias y suficientes para conservar el orden multivariante bajo transformaciones. Como aplicación, estudiamos "*el esparcimiento de la dispersion*" mediante la comparación en dispersión de distribuciones Wishart.

Claves: Contracción, orden dispersivo, distribuciones Wishart.

1. Introducción

El orden dispersivo univariante fue inicialmente definido por Bickel and Lehmann [2] y posteriormente desarrollado por Shaked and Shanthikumar [7]. La definición es la siguiente. Sean F y G las funciones de distribución de X e Y , respectivamente. Se dice que X está menos dispersa que Y , denotado por $X \leq_{disp} Y$, si

$$F^{-1}(\beta) - F^{-1}(\alpha) \leq G^{-1}(\beta) - G^{-1}(\alpha), \text{ para todo } 0 < \alpha < \beta < 1,$$

donde F^{-1} y G^{-1} son las funciones inversas de F y G , respectivamente. Posteriormente, Bartoszewicz [1] y Rojo and He [6] estudian diferentes condiciones suficientes para conservar el orden dispersivo bajo transformaciones. Sin embargo, el orden dispersivo multivariante no ha sido tan extensamente desarrollado. El primer autor que presentó nociones sobre la dispersión multivariante fué Oja [5]. Giovagnoli and Wynn [4] dan dos definiciones de la ordenación dispersiva multivariante, la versión fuerte y la versión débil. Por otra parte, Fernández-Ponce and Suárez-Lloréns [3], definen la dispersión multivariante mediante el concepto de cuantiles condicionados. Estos autores ordenan dos distribuciones Wishart no singulares con el mismo grado de libertad en el sentido dispersivo multivariante. Sin embargo, no es fácil comparar dos distribuciones Wishart multivariante usando los resultados dados en Fernández-Ponce and Suárez-Lloréns [3] ni los resultados dados en Giovagnoli and Wynn [4].

En este trabajo, el orden dispersivo multivariante fuerte, denotado por \leq_{SD} , se estudiará para distribuciones Wishart debido a su aplicación en análisis

de regresión Bayesiano. Nótese que una medida de influencia Bayesiana está basada en la comparación de las correspondientes densidades predictivas. Por tanto, una primera etapa para la definición de una nueva medida de influencia, es la comparación de distribuciones Wishart en sentido del orden \leq_{SD} . De aquí, nuestro interés por establecer condiciones suficientes que conserven el orden dispersivo multivariante bajo transformaciones.

2. Contactar

Dirección:

Departamento de Estadística e Investigación Operativa.
 Facultad de Matemáticas - c/ Tarfia s/n
 41012 - Sevilla
 España

Teléfono: (+34) 954557927 **Fax:** (+34)954622800

www: <http://destio.us.es/>

3. Bibliografía

- [1] Bartoszewicz, J. (1985). Moment inequalities for order statistics from ordered families of distributions. *Metrika*, **32**, 383-389.
- [2] Bickel, P.J. and Lehmann, E.L. (1979). *Descriptive statistics for non-parametric models IV. Spread*, in: Jureckova, ed., Contributions to Statistics, (Reidel, Dordrecht) pp. 33-40
- [3] Fernández-Ponce, J.M. and Suárez-Lloréns, A. (2003). A multivariate dispersion ordering based on quantiles more widely separated. *Journal of Multivariate Analysis*, **85**, 40-53.
- [4] Giovagnoli, A. and Wynn, H.P. (1995). Multivariate dispersion orderings. *Statistics and Probability Letters*, **22**, 325-332.
- [5] Oja, H. (1983). Descriptive statistics for multivariate distributions. *Statistics and Probability Letters*, **1**, 327-332.
- [6] Rojo, J. and He, G.Z. (1991) New properties and characterizations of the dispersive ordering. *Statistics and Probability Letters*, **11**, 365-372.
- [7] Shaked, M. and Shanthikumar, J. G. (1994). *Stochastic Orders and Their Applications*. Academic Press. New York.

Contrastes para órdenes convexos basados en L-estadísticos

Miguel A. Sordo¹, Héctor M. Ramos², Jorge Ollero³

¹mangel.sordo@uca.es, ²hector.ramos@uca.es, ³jorge.ollero@uca.es
Departamento de Estadística e Investigación Operativa, Universidad de Cádiz

Abstract

Dadas dos distribuciones de probabilidad, proponemos una familia de tests basados en L-estadísticos para contrastar la hipótesis nula de igualdad estocástica frente a la hipótesis alternativa de que ambas distribuciones están estrictamente ordenadas en el orden convexo (respectivamente: orden creciente convexo, orden en dilatación y orden de Lorenz).

Palabras Clave: L-estadísticos, orden convexo, orden de Lorenz

AMS: 60E10, 60E15

1. Introducción

Dadas dos variables aleatorias X e Y , se dice que:

- (a) X está dominada por Y en el orden convexo (denotado $X \leq_{cx} Y$) si

$$E[\Phi(X)] \leq E[\Phi(Y)] \quad (5.91)$$

para toda función Φ convexa tal que las esperanzas en (1) existen.

- (b) X está dominada por Y en el orden creciente convexo (denotado $X \leq_{icx} Y$) si (1) se verifica para toda función creciente y convexa.

- (c) Siendo μ_X y μ_Y las medias respectivas de X e Y , X está dominada por Y en el orden en dilatación (denotado $X \leq_{dil} Y$) si

$$E[\Phi(X - \mu_X)] \leq E[\Phi(Y - \mu_Y)] \quad (5.92)$$

para toda función Φ convexa tal que las esperanzas en (2) existen.

- (d) Siendo X e Y no negativas, X está dominada por Y en el orden de Lorenz (denotado $X \leq_L Y$) si

$$E[\Phi(X/\mu_X)] \leq E[\Phi(Y/\mu_Y)] \quad (5.93)$$

para toda función Φ convexa tal que las esperanzas en (3) existen.

2. Resultados

Dadas dos variables aleatorias X e Y tales que $X \leq_{cx} Y$ (respectivamente, \leq_{icx} , \leq_{dil} , \leq_L), proponemos una familia de tests basados en L-estadísticos (combinaciones lineales de estadísticos ordenados) para contrastar la hipótesis nula $H_0 : X \equiv_{st} Y$ (respectivamente $X \equiv_{st} Y$, $X \equiv_{st} Y + k_1$, $X \equiv_{st} k_2 Y$, siendo k_1 y $k_2 > 0$ ciertas constantes) frente a la hipótesis alternativa $H_1 : X \not\leq_{cx} Y$ (respectivamente, $\not\leq_{icx}$, $\not\leq_{dil}$, $\not\leq_L$), donde “ \equiv_{st} ” denota igualdad en distribución y “ $\not\leq$ ” denota orden estricto. Los contrastes se derivan de las caracterizaciones de órdenes estocásticos obtenidas por Ramos y Sordo (2003) y Sordo y Ramos (2006).

3. Agradecimientos

Este trabajo ha sido parcialmente subvencionado por el Proyecto SEJ2005-06678 del MEC.

4. Bibliografía

- [1] Ramos, H. M., Sordo M. A. (2003). Dispersion measures and dispersive orderings. *Statistics & Probability Letters* 61, 123-131.
- [2] Sordo, M. A. and Ramos, H. M (2006). Characterization of stochastic orders by L-functionals. Por aparecer en *Statistical Papers*.

Cuasi probabilidad en el anillo \mathbf{B}^*

Joaquín Ferrer Sánchez¹, Juan Cuadra Carreño¹, Francisco Herrera Cuadra¹

¹ fherrer@ual.es, ² jferrer@ual.es, ³ jcudra@ual.es, Departamento de Estadística y Matemática Aplicada, Universidad de Almería

Abstract

Se realiza la construcción del anillo \mathbf{B}^* , anillo que nos sirve como soporte para realizar la definición de cuasi variables aleatorias sobre anillos, es decir, viene a sustituir a la σ -álgebra de Borel del caso clásico. Tras esto se define la cuasi probabilidad inducida, la función de distribución asociada, etc.

Palabras Clave: Cuasi probabilidad sobre anillos. Cuasi variables aleatorias.

AMS: Clasificación AMS (60A05)

1. Introducción

Este trabajo desarrolla los conceptos tratados en las comunicaciones *Cuasi variables aleatorias*, *Cuasi variables aleatorias sobre el σ -anillo \mathbf{T}^** y *Cuasi probabilidad sobre el anillo \mathbf{F}^** presentadas en los XXVII y XXVIII Congreso Nacional de Estadística e I.O., respectivamente.

En el caso que nos ocupa, trabajaremos sobre un anillo A que no sea un álgebra, al que exigiremos que sea un espacio medible, es decir $\bigcup_{E \in A} E = \Omega$,

por lo que Ω es infinito puesto que de lo contrario estaríamos en un álgebra.

Sobre este anillo A se define una medida finita μ con la condición de que tenga supremo acotado, esto es: $\sup_{E \in A} \mu(E) = 1$. Se trata de generalizar la idea del anillo \mathbf{F}^* , añadiéndole los subconjuntos numerables. A Dicho anillo lo denominaremos \mathbf{B}^* .

2. Cuasi variable aleatoria sobre \mathbf{B}^* .

Tomaremos la clase $\mathbf{E} = \{] -\infty, x], \mathbf{N} \}$, como punto de partida (siendo \mathbf{N} el σ -anillo de la numerabilidad), y $\mathbf{B}^* = < \mathbf{E} > = < \{] -\infty, x], \mathbf{N} \} >$, el anillo generado por ella. Se denota \mathbf{B}^* por ser casi la σ -álgebra de Borel. El álgebra generada por este anillo vendrá dada por, $\mathbf{A}_{\mathbf{B}^*} = < \mathbf{B}^* > = \{ E \subset \mathbf{R} / E \in \mathbf{B}^* \vee E^c \in \mathbf{B}^* \}$ en la que cualquier tipo de intervalo de \mathbf{R} esté o no acotado pertenecería a la misma.

Definición Llamaremos cuasi variables aleatorias a aquellas aplicaciones

$$\xi : (\Omega, A) \longrightarrow (\mathbf{R}, \mathbf{B}^*)$$

de forma que

$$1) \xi^{-1}(E) \in A \forall E \in \mathbf{B}^* \quad 2) \sup_{E \in \mathbf{B}^*} \mu(\xi^{-1}(E)) = 1$$

obsérvese que la segunda condición no tiene paralelismo en el caso clásico.

Teorema Es condición necesaria y suficiente para que una aplicación ξ sea una cuasi v.a. que se verifique:

- 1) $\xi^{-1}(]-\infty, x]) \in A \forall x \in \mathbf{R}$
- 2) $\xi^{-1}(E) \in A \forall E \in \mathbf{N}$
- 3) $\sup_{x \in \mathbf{R}} \{\mu(\xi^{-1}(]-\infty, x]))\} = 1$

A partir de dicha definición y de su teorema de caracterización se realiza un estudio paralelo al caso tradicional, obteniéndose:

Definición Dada una cuasi v.a. $\xi : (\Omega, A) \longrightarrow (\mathbf{R}, \mathbf{B}^*)$ se define la cuasi probabilidad inducida como

$$\mu_\xi(E) = \mu(\xi^{-1}(E))$$

lo que nos conduce a la definición de la correspondiente función de distribución

$$F_\xi(x) = \mu_\xi(]-\infty, x])$$

y por lo tanto la función de masa-densidad correspondiente será:

$$f_\xi(x) = \begin{cases} \alpha \cdot f_{\xi_d}(x) & x \in S_\xi \\ \beta \cdot f_{\xi_{ac}}(x) & x \in S_\xi^c \end{cases}$$

siendo

$$\alpha = \mu_\xi(S_\xi); \quad \beta = \sup_{x \in \mathbf{R}} [\mu_\xi(]-\infty, x] - S_\xi)]$$

3. Bibliografía

- [1] Herrera, F. y Cuadra J. (1998). *Probabilidad sobre σ -anillos*. XXIV Congreso SEIO.
- [2] Herrera, F., Ferrer, J. y Cuadra J. (2003). *Cuasi variables aleatorias* XXVII Congreso SEIO.
- [3] Ferrer, J. y Herrera, F. (2003). *Cuasi variables aleatorias sobre el σ -anillo \mathbf{T}^** . XXVII Congreso SEIO.
- [4] Herrera, F., Ferrer, J. y Cuadra J. (2004). *Cuasi probabilidad sobre el anillo \mathbf{F}^** . XXVIII Congreso SEIO.

De la geometría composicional al espacio de Hilbert de densidades

J.J. Egozcue¹, V. Pawlowsky-Glahn², J.L. Díaz-Barrero¹

¹juan.jose.egozcue@upc.edu,jose.luis.diaz@upc.edu, Dep. de Matemática Aplicada III, Universidad Politécnica de Cataluña. España.

²vera.pawlowsky@udg.es, Dep. de Informática y Matemática Aplicada, Universidad de Girona, España.

Abstract

Dos décadas de análisis de datos composicionales han llevado a estructurar su espacio muestral, el simplex, como espacio Euclídeo. Una densidad de probabilidad se puede aproximar por una composición, lo que inspira el estudio de un espacio de Hilbert de densidades: la suma es el teorema de Bayes y la métrica corresponde a la del simplex.

Claves: Aitchison geometry, compositional data, Bayes, perturbation.

1. Introducción

En la actualidad se conciben los datos composicionales como aquellos vectores de componentes positivas que sólo aportan información por el valor relativo o cocientes de sus componentes. De aquí que pueda elegirse un representante de una composición de forma que sus componentes sumen 1, ó 100 para porcentajes ó 10^6 para partes por millón. La consecuencia de este tipo de normalización es que los datos composicionales pueden representarse en el simplex (de tantas partes como componentes). Las dificultades para el análisis de este tipo de datos llevaron a J. Aitchison [1] a definir unas operaciones (perturbación y potenciación) y una distancia en el simplex. Estas operaciones y métrica han resultado ser la clave para estructurar el simplex como espacio Euclídeo. Tras estudiar las bases ortogonales ha sido posible representar las composiciones por sus coordenadas. Estas coordenadas ya están libres de las restricciones del simplex y se puede trabajar con ellas como vectores reales con la geometría ordinaria. J. Aitchison [1] señaló que la operación que hace de suma en el simplex, la perturbación, es equivalente al teorema de Bayes cuando se utilizan parámetros discretos. Resultaba natural, por tanto, que en el desarrollo de la geometría del simplex mencionada se intentara generalizar a composiciones de número infinito de partes, es decir, a las densidades de probabilidad. Los resultados que siguen son el inicio de este estudio.

2. Un espacio de Hilbert de densidades

Se considera un intervalo finito I de longitud η y densidades de medidas positivas soportados en él. Como en el caso de las composiciones se consideran

equivalentes dos densidades proporcionales. Las densidades de medidas finitas podrán ser normalizadas como una de probabilidad. Las correspondientes a medidas infinitas no precisan de una normalización específica, como sucede a los *priori* o funciones de verosimilitud impropios que aparecen en la estadística bayesiana. Se define perturbación de dos densidades, f_1 , f_2 , sobre I , y la potenciación por un escalar α , como

$$(f_1 \oplus f_2)(x) = f_1(x) \cdot f_2(x), \quad (\alpha \odot f_1)(x) = f_1(x)^\alpha,$$

donde la igualdad se entiende como proporcionalidad (salvo en conjuntos de medida nula). La operación \oplus se identifica con el teorema de Bayes. Puede demostrarse que con estas operaciones las (clases de) densidades sobre I constituyen un espacio vectorial de dimensión infinita. Se ha demostrado [2] que el subespacio de densidades, $A^2(I)$ cuyo logaritmo es de cuadrado integrable sobre I tiene estructura de espacio de Hilbert con el producto escalar, inspirado en el de Aitchison en el simplex,

$$\langle f_1, f_2 \rangle_a = \frac{1}{2\eta} \int_{I \times I} \ln \frac{f_1(x)}{f_1(y)} \cdot \ln \frac{f_2(x)}{f_2(y)} dx dy.$$

Este espacio contiene tanto densidades de probabilidad como otras impropias. La norma y la distancia entre densidades se deducen de esta definición. También es posible trasladar las bases de Hilbert del espacio $L^2(\mathbf{R})$ a $A^2(I)$. Una vez seleccionada una base, una densidad puede representarse por sus coeficientes de Fourier. Los coeficientes permiten reducir las operaciones a la suma y producto por escalar de sucesiones. La distancia de densidades a la distancia ℓ^2 de las sucesiones de coeficientes y el truncamiento de la sucesión de coeficientes es una proyección ortogonal. Las aplicaciones más inmediatas corresponden a la utilización de la distancia de Aitchison entre densidades para sustituir a otras distancias o divergencias (Hellinger, Kullback-Leibler, L^2 , Kolmogorov-Smirnov, etc.). Desde el punto de vista teórico, destaca la conservación de la distancia de Aitchison entre dos *priori* (proprios o improprios) después de obtener sus respectivos *posteriori* mediante una función de verosimilitud.

3. Agradecimientos

Ministerio de Ciencia y Tecnología (es): proyecto BFM2003-05640/MATE.

4. Bibliografía

- [1] Aitchison, J. (1986). *The Statistical Analysis of Compositional Data*. Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London, (UK), 416 p.
- [2] Egozcue, J. J., Díaz-Barrero, J.L. and Pawlowsky-Glahn, V. (2006). Hilbert Space of Probability Density Functions Based on Aitchison Geometry, *Acta Mathematica Sinica*, 22, 1, 1-8.

Densidades de cópulas considerando la estructura de su espacio soporte

M.I. Ortego¹, G. Mateu-Figueras²

¹ma.isabel.ortego@upc.edu, Departament de Matemàtica Aplicada III,
Universitat Politècnica de Catalunya

²gloria.mateu@udg.es, Departament d'Informàtica i Matemàtica Aplicada,
Universitat de Girona

Abstract

El espacio soporte de las funciones cópula es el cuadrado unidad. Éste puede ser considerado como subespacio de R^2 , o como un espacio vectorial euclíadiano con una estructura diferente a la de R^2 que induce una medida diferente a la habitual del espacio real. Mediante una aproximación gráfica se estudia cómo la elección de la estructura afecta a las densidades de algunas familias paramétricas de cópulas.

Claves: derivada Radon-Nikodým, medida, funciones cópula.

1. Introducción

Las aplicaciones estadísticas utilizan cada vez con más frecuencia el concepto de cópula entre dos o más variables. El teorema de Sklar, [1], asegura que cualquier función de distribución conjunta puede expresarse mediante la correspondiente cópula y las funciones de distribución marginales. En el caso bidimensional, esto corresponde a que

$$F_{XY}(x, y) = C_{XY}[F_X(x), F_Y(y)] \quad ,$$

para todo punto (x, y) del soporte, donde C_{XY} denota la función cópula . Este enfoque resulta de mucho interés, puesto que permite considerar de forma separada la estructura de dependencia entre las variables y las marginales. La dependencia entre las variables queda totalmente representada mediante su cópula. Además, la cópula es invariante bajo transformaciones estrictamente monótonas de las variables aleatorias.

Las funciones cópula son esencialmente funciones de distribución con marginales uniformes. Son pues funciones definidas sobre el espacio $[0, 1] \times [0, 1]$. Si $C_{UV}[u, v]$ es una cópula 2-dimensional, entonces,

- a) $C_{UV}[u, 1] = u, \quad C_{UV}[1, v] = v$, para todo u, v en $[0, 1]$.
- b) $C_{UV}[u, 0] = C_{UV}[0, v] = 0$ para todo u, v en $[0, 1]$.
- c) $C_{UV}[a_2, b_2] - C_{UV}[a_1, b_2] - C_{UV}[a_2, b_1] + C[a_1, b_1] \geq 0$, cuando $a_1 \leq a_2, \quad a_1 \leq a_2 \in [0, 1]$.

Por tanto, la cópula posee un conjunto de propiedades de gran interés. No obstante, en algunas ocasiones surge la necesidad de utilizar la función de densidad de la cópula en lugar de su función de distribución.

2. Funciones de densidad de cópulas

La definición general de función de densidad como derivada de Radon-Nikodým de una probabilidad relativa a una medida es válida en cualquier espacio. Habitualmente trabajamos con variables o vectores aleatorios reales y por lo tanto utilizamos densidades respecto a la medida de Lebesgue del espacio real. Sin embargo, los problemas aparecen cuando trabajamos con espacios soporte donde no es coherente considerar la medida de Lebesgue. Pensemos por ejemplo en un espacio discreto; en estos casos se da la función de densidad respecto de la medida consistente: la medida cardinal.

Nuestro estudio se centrará en funciones de densidad de las cópulas, definidas sobre el espacio $[0, 1] \times [0, 1]$. Podemos considerar este espacio como un subconjunto de R^2 y consecuentemente podemos utilizar densidades respecto a la medida de Lebesgue. No obstante, Pawlowsky-Glahn [2] introduce una estructura de espacio vectorial euclíadiano diferente a la de R^2 que induce una medida diferente a la habitual del espacio real. En este contexto no tendrá sentido definir una densidad de probabilidad relativa a la medida de Lebesgue puesto que ésta no será coherente con la estructura del espacio.

En este trabajo nos proponemos revisar las expresiones correspondientes a las densidades de algunas familias paramétricas de cópulas de uso común y expresarlas con respecto a la medida consistente con la estructura del espacio soporte definida en [2]. Posteriormente, mediante un estudio comparativo gráfico describiremos las principales diferencias entre las densidades clásicas, expresadas respecto a la medida de Lebesgue, y las densidades que se proponen, expresadas respecto a la medida introducida en [2].

3. Agradecimientos

Este trabajo ha recibido financiación mediante el proyecto BFM2003-05640/MATE (Dir. General de Investigación; Minis. de Ciencia y Tecnología).

4. Bibliografía

- [1] Nelsen, R.B. (1999). *An introduction to copulas* Springer-Verlag.
- [2] Pawlowsky-Glahn, V. (2003) Statistical modelling on coordinates. In: *Proceedings of CoDaWork'03 - Compositional Data Analysis Workshop-*.

Duality for Multi-State Probabilistic Cellular Automata

F.J. López¹, G. Sanz², M. Sobottka³

¹javier.lopez@unizar.es, ²gerardo@unizar.es, ³sobottka@unizar.es,
Departamento de Métodos Estadísticos, Universidad de Zaragoza

Abstract

The present work treats dualities for probabilistic cellular automata (PCA). A general result of duality is presented and it is used to study two models of PCA: the multi-opinion noisy voter model; and the multi-opinion noisy biased voter model.

Keywords: Probabilistic cellular automata; duality; voter model.

1. Introduction

In this work we study one-dimensional probabilistic cellular automata (PCA), that are discrete time stochastic processes with state space $X := W^{\mathbb{Z}}$, where $W = \{0, 1, \dots, M-1\}$, and with finite range interactions on the integer lattice \mathbb{Z} . More precisely, suppose $\ell, r \geq 0$ and $f : W \times W^{\ell+r+1} \rightarrow [0, 1]$, and consider the stochastic process $\eta_t = \{\eta_t(x) \in W : x \in \mathbb{Z}\}$ which evolution satisfies for all $t \in \mathbb{N}$, $x \in \mathbb{Z}$, and $w, w_{-\ell}, \dots, w_r \in W$:

$$\mathbb{P}\{\eta_{t+1}(x) = w \mid \eta_t(x-\ell) = w_{-\ell}, \dots, \eta_t(x+r) = w_r\} = f(w; w_{-\ell}, \dots, w_r).$$

In other words, η_t is an interacting particle system which evolves in discrete time. Let η_t and ξ_t be two interacting particle system with state spaces X and Y respectively, and let $H : X \times Y \rightarrow \mathbb{R}$ be a bounded measurable function. Then, we say η_t and ξ_t are dual to one another with respect to H if

$$\mathbb{E}_{\eta}[H(\eta_t, \xi)] = \mathbb{E}_{\xi}[H(\eta, \xi_t)], \quad (5.94)$$

for all $\eta \in X$ and $\xi \in Y$. If η_t and ξ_t are discrete time Markov chains with transition matrix P and Π respectively, then the previous expression for duality can be written as $PH = H\Pi^T$.

For the case of η_t evolving in continuous time, López and Sanz [4] studied a different type of duality, which is given by the equation:

$$\mathbb{E}_{\eta}[H(\eta_t, \xi)] = \mathbb{E}_{\xi}\left[H(\eta, \xi_t)e^{\int_0^t V(\xi_s)ds}\right], \quad (5.95)$$

for all $\eta \in X$ and $\xi \in Y$, and for some function $V : Y \rightarrow [0, \infty)$. Such equation of duality was used in [4] to obtain the limit behavior and conditions of ergodicity for some continuous time multi-opinion voter models.

On the other hand, for discrete time interacting particle systems, Katori et al. [1] study the duality for particular cases of the Domany-Kinzel model, which is the PCA with state space $\{0, 1\}^{\mathbb{Z}}$, and evolution defined by the parameters $p_0, p_1, p_2 \in [0, 1]$, as follows:

$$\mathbb{P}\{\eta_{t+1}(x) = 1 | \eta_t(x-1) = w_{-1}, \eta_t(x) = w_0, \eta_t(x+1) = w_1\} = p_{w_{-1}+w_1}.$$

When $p_0 = 0$ and $p_2 \leq p_1$, Katori et al. present a dual for η_t . The technique used in [1] was to represent η_t by a transition matrix P , and so to find matrixes H and Π which satisfy the duality equation. In [2] and [3], Konno presents results of self-duality for a class of multi-state finite range PCA, which include that studied in [1].

In the present work we propose an alternative type of duality for general PCA, which is the discrete time version of (5.95). Given a PCA η_t on X we are interested in finding a discrete time Markov chain ξ_t on Y , and a duality function $H : X \times Y \rightarrow \mathbb{R}$, such that

$$\mathbb{E}_{\eta}[H(\eta_t, \xi)] = d(\xi)\mathbb{E}_{\xi}[H(\eta, \xi_t)], \quad (5.96)$$

for some function $d : Y \rightarrow [0, \infty)$. Describing the evolution of η_t through a transition matrix P , the equation (5.96) can be written as $PH = H(D\Pi)^T$, where D is the diagonal matrix with $D_{\xi\xi} = d(\xi)$. We obtain conditions on D and Π for the ergodicity of the process, and as examples, two PCA models are studied: the multi-opinion noisy voter model; and the multi-opinion noisy biased voter model. Such models include the Domany-Kinzel model for arbitrary parameters $0 \leq p_0 \leq p_1 \leq p_2 \leq 1$.

2. Acknowledgements

Research supported by project MTM2004-01175 of MEC, and research group “Modelos Estocásticos” (DGA).

3. Bibliography

- [1] Katori, M. Konno, N. Sudbury, A., and Tanemura, H. (2004). Dualities for the Domany-Kinzel model. *J. of Theoretical Probability*, **17**, 1, 131-144.
- [2] Konno, N. (2002). Dualities for a class of finite range probabilistic cellular automata in one dimension. *J. of Statistical Physics*, **106**, 516, 915-922.
- [3] Konno, N. (2002). Self-duality for multi-state probabilistic cellular automata with finite range interactions. *J. of Statistical Physics*, **106**, 516, 923-930.
- [4] López, F.J. and Sanz, G. (2000). Duality for general interacting particle systems. *Markov Process and Relat. Fields*, **6**, 305-328.

Estimación cuadrática a partir de observaciones con retraso aleatorio correlado

R. Caballero¹, A. Hermoso², J. Jiménez¹, J. Linares², S. Nakamori³

¹raguila@ujaen.es, jdomingo@ujaen.es, Departamento de Estadística e Investigación Operativa, Universidad de Jaén

²ahermoso@ugr.es, jlinares@ugr.es, Departamento de Estadística e Investigación Operativa, Universidad de Granada

³nakamori@edu.kagoshima-u.ac.jp, Department of Technology, Faculty of Education, Kagoshima University

Abstract

Se considera el problema de estimación cuadrática de mínimos cuadrados de una señal aleatoria discreta, usando observaciones perturbadas por ruido blanco aditivo, que pueden estar aleatoriamente retrasadas una unidad de tiempo. El retraso aleatorio se modeliza mediante una sucesión de variables de Bernoulli correladas en instantes de tiempo consecutivos.

Claves: Estimación cuadrática, retraso aleatorio.

AMS: 62M20, 60G35

1. Introducción

Existen diversos problemas de estimación de señales dinámicas en los que la llegada de datos no es instantánea; en tales casos, los algoritmos de estimación convencionales no son aplicables, y es preciso modificarlos para incorporar el efecto de los posibles retrasos que serán, en general, función del tiempo. Aunque en ciertos casos el retraso es una función determinística, muchas situaciones reales se modelizan de manera más adecuada suponiendo que los retrasos son aleatorios. Algoritmos de estimación a partir de observaciones aleatoriamente retrasadas, obtenidos bajo diferentes aproximaciones, han sido propuestos por diversos autores para tratar algunas de tales situaciones ([1]-[2]).

En este trabajo se trata el problema de estimación cuadrática de mínimos cuadrados de señales discretas a partir de observaciones afectadas por ruido blanco aditivo, cuando la medida disponible en cualquier instante de tiempo puede no ser la observación de la señal en dicho instante, sino la correspondiente al instante anterior. Se supone que el retraso en cada una de las observaciones es aleatorio, y que la probabilidad de retraso es conocida. Concretamente, el retraso en cada instante se describe mediante una variable de Bernoulli que toma el valor uno si la observación correspondiente está retrasada y el valor cero si está actualizada.

Suponiendo que las variables de Bernoulli que modelizan el retraso en las observaciones son mutuamente independientes, en [3] y [4] se proponen algoritmos de estimación de mínimos cuadrados (lineal y cuadrática, respectivamente) para los problemas de filtrado y suavizamiento punto fijo. Tales algoritmos se obtienen a partir de las funciones de covarianzas de los procesos involucrados en la ecuación de observación.

En el presente trabajo se generaliza el estudio realizado en [4], caracterizando el retraso en las medidas por variables de Bernoulli correladas en instantes consecutivos. Esta forma de correlación permite considerar modelos en los que dos observaciones consecutivas no pueden estar retrasadas; concretamente, los algoritmos propuestos pueden aplicarse a modelos de transmisión con sensores *stand-by*, que detectan cualquier fallo de transmisión, procediéndose al reemplazamiento inmediato del sensor antiguo.

Además de las distribuciones marginales de las variables de Bernoulli, especificadas por las probabilidades de retraso, y la función de correlación de tales variables, los algoritmos de filtrado y suavizamiento punto fijo propuestos requieren exclusivamente el conocimiento de los momentos (hasta el cuarto orden) de la señal y el ruido blanco aditivo de las observaciones. La técnica empleada para abordar el problema de estimación cuadrática consiste en aumentar los vectores señal y observación agregando a los originales sus potencias de segundo grado; así, obtenido el estimador lineal de la señal aumentada basado en las observaciones aumentadas, sus primeras componentes proporcionan directamente los estimadores cuadráticos de la señal original.

2. Agradecimientos

Este trabajo ha sido parcialmente financiado por el Ministerio de Educación y Ciencia, a través del proyecto MTM2005-03601.

3. Bibliografía

- [1] C.L. Su and C.N. Lu, "Interconnected network state estimation using randomly delayed measurements", *IEEE Transactions on Power Systems*, vol.16, pp.870-878, 2001.
- [2] A.S. Matveev and A.V. Savkin, "The problem of state estimation via asynchronous communication channels with irregular transmission times", *IEEE Transactions on Automatic Control*, vol.48, pp.670-676, 2003.
- [3] S. Nakamori, R. Caballero, A. Hermoso and J. Linares (2005). Recursive estimators of signals from measurements with stochastic delays using covariance information. *Applied Mathematics and Computation* 162, 65-79.
- [4] S. Nakamori, A. Hermoso and J. Linares (2005). Quadratic estimation of multivariate signals from randomly delayed measurements. *Multidimensional Systems and Signal Processing* 16, 417-438.

Estimación de señales mediante aproximación de mixturas

R. Caballero¹, A. Hermoso², J. Jiménez¹, J. Linares², S. Nakamori³

¹raguila@ujaen.es, jdomingo@ujaen.es, Departamento de Estadística e Investigación Operativa, Universidad de Jaén

²ahermoso@ugr.es, jlinares@ugr.es, Departamento de Estadística e Investigación Operativa, Universidad de Granada

³nakamori@edu.kagoshima-u.ac.jp, Department of Technology, Faculty of Education, Kagoshima University

Abstract

En este trabajo se propone un algoritmo de filtrado no lineal para señales gaussianas a partir de observaciones inciertas perturbadas por un ruido blanco correlado con la señal. La técnica empleada se basa en aproximaciones de mixturas de distribuciones normales que proporcionan, a su vez, una aproximación del estimador óptimo de la señal.

Claves: Filtrado, observaciones inciertas, correlación señal-ruido

AMS: 62M20, 60G35

1. Introducción

Usualmente, para abordar el problema de estimación de señales aleatorias se supone que las observaciones disponibles contienen siempre información sobre dicha señal. Sin embargo, en diversos campos relacionados con las telecomunicaciones o el procesamiento de imágenes, pueden existir interrupciones aleatorias en el mecanismo de medida, de forma que la probabilidad de que una observación sea únicamente ruido puede ser positiva. Los sistemas que modelizan estas situaciones se denominan *sistemas con observaciones inciertas* y están caracterizados por incluir en la ecuación de observación no sólo ruidos aditivos, sino también un ruido multiplicativo modelizado por una sucesión de variables aleatorias de Bernoulli. Debido a esta componente multiplicativa, incluso bajo la hipótesis de gaussianidad de la señal y los ruidos del sistema, la distribución conjunta de la señal y las observaciones no es gaussiana y, en consecuencia, el estimador de mínimos cuadrados no puede calcularse, generalmente, de forma simple. Por este motivo, el interés en este tipo de sistemas se ha centrado en la búsqueda de estimadores subóptimos que puedan obtenerse recursivamente. En particular, el problema de estimación lineal a partir de observaciones inciertas ha sido ampliamente estudiado, tanto bajo la hipótesis de que el modelo de espacio de estados de la señal es conocido [1], como cuando dicho modelo no es

completamente conocido, pero se dispone de información alternativa referida, por ejemplo, a las funciones de covarianza de los procesos que intervienen [2]. En los trabajos mencionados, se supone que la señal y los ruidos del sistema son mutuamente independientes. Sin embargo, existen numerosas situaciones en las que esta hipótesis no es realista. Suponiendo correlación entre la señal y el ruido aditivo de las observaciones, el problema de estimación lineal ha sido también abordado por diversos autores cuando se conoce el modelo de espacio de estados de la señal [3] y cuando dicho modelo es desconocido [4]. En este artículo, proponemos un algoritmo recursivo de filtrado para el problema de estimación de señales gaussianas a partir de observaciones inciertas, que no requiere el conocimiento completo del modelo de espacio de estados de la señal, sino únicamente las funciones de covarianza de los procesos implicados. Como ya se ha indicado, la incertidumbre de las observaciones hace que, en este tipo de sistemas, el cálculo del estimador óptimo no sea sencillo, por lo que proponemos un método que, mediante aproximaciones sucesivas de mixturas de distribuciones normales, permite obtener de forma aproximada la función de densidad condicionada de la señal dadas las observaciones. La esperanza de dicha distribución condicionada aproximada proporciona un estimador subóptimo no lineal de la señal, que puede obtenerse de forma recursiva. Mediante un ejemplo de simulación se pone de manifiesto que el filtro propuesto mejora el comportamiento del filtro lineal obtenido en [4].

2. Agradecimientos

Este trabajo ha sido parcialmente financiado por el Ministerio de Educación y Ciencia, a través del proyecto MTM2005-03601.

3. Bibliografía

- [1] W. NaNacara and E. E. Yaz (1997). Recursive estimator for linear and nonlinear systems with uncertain observations. *Signal Processing* 62, 215-228.
- [2] S. Nakamori, R. Caballero, A. Hermoso and J. Linares (2003). New design of estimators using covariance information with uncertain observations in linear discrete-time systems. *Applied Mathematics and Computation* 135, 429-441.
- [3] A. Hermoso and J. Linares (1994). Linear estimation for discrete-time systems in the presence of time-correlated disturbances and uncertain observations. *IEEE Transactions on Automatic Control* 39(8), 1636-1638.
- [4] S. Nakamori, R. Caballero, A. Hermoso, J. D. Jiménez and J. Linares (2005). An innovation approach to the smoothing problem from uncertain observations with correlated signal and noise. *Mathematical Methods in the Applied Sciences* 28, 1569-1584.

Estudio de variables temporales asociadas a procesos estocásticos

R. Gutiérrez Jáimez, P. Román Román, D. Romero Molina, J.J. Serrano Pérez, F. Torres Ruiz

{rgjaimez, proman, deromero, jjserra, fdeasis}@ugr.es

Departamento de Estadística e Investigación Operativa

Universidad de Granada

Abstract

Se propone un método para estimar la función de densidad de variables temporales generales asociadas a procesos estocásticos en tiempo continuo. Se compara el método propuesto con la aproximación numérica de la solución de la ecuación integral de Volterra que verifica la densidad del tiempo de primer paso a través de barreras en procesos de difusión.

Claves: Tiempo de primer paso, proceso estocástico, proceso de difusión.

AMS: 60G07

1. Introducción

En el estudio de fenómenos dinámicos, modelizados mediante procesos estocásticos, se pueden plantear, entre otros, los siguientes problemas:

- Tiempo necesario para alcanzar un determinado valor o, en general, atravesar una determinada función del tiempo.
- Tiempo en que se produce la inflexión en modelos asociados a curvas de tipo sigmoidal, por ejemplo, en procesos de crecimiento.
- Tiempo en que se produce un cambio en la evolución de la tendencia.

Este tipo de problemas, asociados a un proceso estocástico en tiempo continuo $\{X(t); t \geq t_0\}$, se puede formular en la forma general

$$T = \inf_{t \geq t_0} \{t : X(t) \text{ verifica cierta propiedad}\} \quad (5.97)$$

y su solución consiste en la obtención de la distribución de tal variable temporal.

Este tipo de variables incluye el tiempo de primer paso por barreras variables en el tiempo

$$T = \inf_{t \geq t_0} \{t : X(t) > S(t)\},$$

(y, en particular, por barreras constantes, $S(t) = S$), cuya función de densidad, en el caso de un proceso de difusión, entre otros, verifica una ecuación integral de Volterra de segunda especie [1]. Dicha ecuación puede ser resuelta de forma explícita para cierto tipo de barreras, mientras que, en general, hay que recurrir a aproximaciones numéricas. En este caso, se plantean problemas computacionales debido a que un paso excesivamente pequeño en la aplicación del algoritmo que garantice la convergencia supone un tiempo de computación excesivo en algunas ocasiones (sobre todo cuando el tiempo de primer paso ocurre lejos del instante inicial).

En este trabajo se propone un procedimiento para la estimación de la función de densidad de la variable temporal (5.97) basado en la simulación de trayectorias del proceso y estimación no paramétrica de la densidad a partir de la información proporcionada por la evaluación de la variable temporal en cada una de ellas.

El procedimiento propuesto se aplica en el estudio del tiempo en que se produce la inflexión en el proceso de difusión Gompertz propuesto por Romero [2]. Asimismo, se considera el problema de tiempo de primer paso mostrando, en este caso, las ventajas que, en general, presenta este método en cuanto a la disminución del tiempo de computación frente a la aproximación numérica, cuando ésta es aplicable.

2. Agradecimientos

Este trabajo ha sido parcialmente financiado por el Ministerio de Educación y Ciencia, Proyecto de Investigación MTM2005-09209.

3. Bibliografía

- [1] Gutiérrez, R., Ricciardi, L. Román, P., Torres, F. (1997). First-passage-time densities for time-non-homogeneous diffusion process *Journal of Applied Probability* 34, 623-631.
- [2] Romero, D. (2005) *Aportaciones al estudio de modelos estocásticos asociados a curvas de crecimiento: un nuevo proceso de difusión tipo Gompertz*. Tesis Doctoral, Universidad de Granada.

Interpretación geométrica de λ para la familia de distribuciones gaussianas discretas

Miguel Ángel Tíscar Soria¹, José Rodríguez Avi², Antonio J. Sáez Castillo³, Antonio Conde Sánchez⁴

¹mats@terra.es, Departamento de Matemáticas, IES Castillo de la Yedra.
Cazorla- Jaén

² jravi@ujaen.es, ; ³ ajsaez@ujaen.es; ⁴aconde@ujaen.es

Departamento de Estadística e Investigación Operativa, Universidad de Jaén

Abstract

El parámetro λ de una familia de distribuciones hipergeométricas gaussianas $GHD(\alpha, \beta, \gamma, \lambda)$ de rango no finito tiene un fuerte componente geométrico, que le confiere un carácter de medida de forma, en cierto modo similar al concepto de pendiente. Por otro lado, la relación entre los parámetros determina la localización de modas locales.

Claves: Distribuciones discretas, Distribucion hipergeométrica de Gauss. Estimación.

AMS: 60E05

1. Introducción

En la familia de distribuciones hipergeométricas de Gauss [2], la función de masa de probabilidad viene dada por:

$$P[X = r] = f_r = \frac{(\alpha)_r (\beta)_r}{(\gamma)_r} \frac{\lambda^r}{r!} f_0 \quad \text{con} \quad f_0 = \left[\sum_{k=0}^{\infty} \frac{(\alpha)_k (\beta)_k}{(\gamma)_k} \frac{\lambda^k}{k!} \right]^{-1},$$

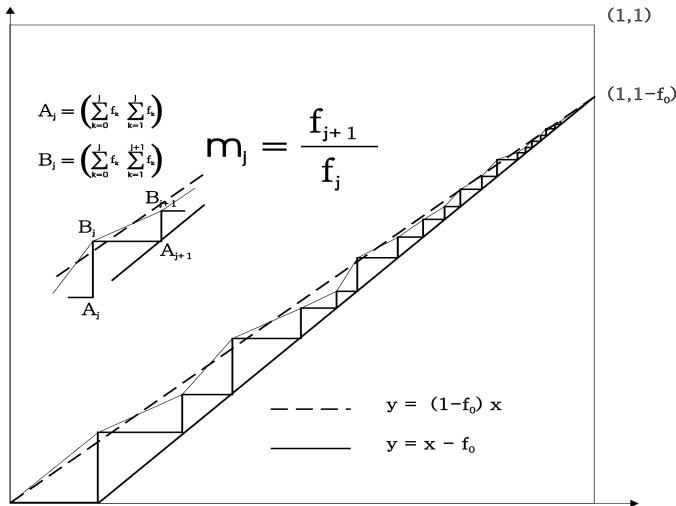
siendo $(\alpha)_k$ el símbolo de Pochhammer de α de orden k . En el caso no finito se puede considerar el límite de la siguiente razón de frecuencias consecutivas:

$$\lim_{r \rightarrow \infty} \frac{f_{r+1}}{f_r} = \lim_{r \rightarrow \infty} \frac{\lambda (\alpha + r) (\beta + r)}{(\gamma + r) (r + 1)} = \lambda. \quad (5.98)$$

2. Interpretación

A partir de (5.98) podemos obtener una primera interpretación del parámetro λ , como *factor de forma*: se trata de un medidor de la cantidad de información que aporta cada dato respecto a la aportada por su antecesor inmediato.

Sin embargo, la representación gráfica de la Figura 1 arroja una interpretación más próxima al concepto de pendiente. En dicha gráfica el eje X de es la función

Figura 1: Interpretación gráfica de λ .

de distribución $F(x)$ y $F^*(x) = F(x) - f_0$ el eje Y. Representamos en esas coordenadas la recta de puntos $A_j = (F(j), F^*(j))$ y la función escalonada dada por los puntos $B_j = (F(j), F^*(j+1))$. De esta forma, la pendiente de cada incremento de la función escalonada es el cociente $\frac{f_{r+1}}{f_r}$, siendo la pendiente de la recta de puntos A_j , $1 - f_0$.

Por otra parte, las modas locales de las distribuciones vienen dadas por las condiciones $\frac{f_r}{f_{r-1}} > 1$ y $\frac{f_{r+1}}{f_r} < 1$. Si se utiliza la reparametrización

$$b_1 = \gamma - 1 \quad a_2 = \lambda \quad a_1 = \lambda(\alpha + \beta) \quad a_0 = \lambda\alpha\beta$$

ambas inecuaciones resultan ser las inecuaciones lineales siguientes:

$$\begin{aligned} &\text{Si } r = 0, \quad a_0 < b_1 + 1 \\ &\text{Si } r \geq 1, \quad \begin{cases} a_2 r^2 + a_1 r + a_0 < b_1 (r+1) + (r+1)^2 \\ b_1 r + r^2 < a_2 (r-1)^2 + a_1 (r-1) + a_0 \end{cases} \end{aligned}$$

3. Aplicaciones

La representación de la Figura 1 pone de manifiesto que $1 - f_0$ puede ser utilizado como semilla u origen de métodos iterativos de resolución de la estimación del valor de λ . Por otro lado, la caracterización de modas locales, permite introducir nuevas restricciones sobre los parámetros a la hora de computar los diferentes métodos de estimación.

Observaciones imprecisas: variables aleatorias difusas

Laura Garrido Blanco¹, Teófilo Brezmes Brezmes², Pedro Gil Alvarez³

¹garridolaura@uniovi.es, becaria del Departamento de Estadística, Investigación Operativa y Didáctica de la Matemática, Universidad de Oviedo

²teofilo@uniovi.es, Departamento de Estadística, Investigación Operativa y Didáctica de la Matemática, Universidad de Oviedo

³pedro@uniovi.es, Departamento de Estadística, Investigación Operativa y Didáctica de la Matemática, Universidad de Oviedo

Abstract

En el trabajo realizado trasladamos resultados existentes en el tratamiento de la información imprecisa por medio de conjuntos aleatorios al caso en que dicha información sea modelada a través de una variable aleatoria difusa, teniendo en cuenta que una variable aleatoria difusa es una generalización de un conjunto aleatorio.

Palabras Clave: variable aleatoria difusa, envolvente probabilística, conjunto gradual

1. Introducción

En múltiples ocasiones, la medida de cierta característica de los individuos de una población no puede realizarse de forma exacta, sino que únicamente podemos asegurar que pertenece a cierto conjunto de valores. A la hora de modelar matemáticamente esta situación, no podemos considerar una variable aleatoria ordinaria. Para solucionar esto, la información imprecisa se modela bien a través de un conjunto aleatorio o bien a través de una variable aleatoria difusa.

En estudios anteriores se ha tratado principalmente el primer caso, definiendo la envolvente probabilística de un conjunto aleatorio y obteniendo diversos resultados para dicha envolvente. Así mismo, se ha definido la envolvente probabilística de una variable aleatoria difusa, pero los resultados obtenidos para ella son muy escasos.

En nuestro trabajo extendemos algunos de los resultados existentes para la envolvente probabilística de un conjunto aleatorio a la de una variable aleatoria difusa, teniendo en cuenta que los α -cortes de una variable aleatoria difusa son conjuntos aleatorios, así como el hecho de que las envolventes probabilísticas de los α -cortes son una representación por conjuntos graduales de la envolvente probabilística de la variable aleatoria difusa.

Entre otros resultados, nos planteamos la siguiente cuestión: si el espacio sobre el que está definida la variable aleatoria difusa no es completo, podemos considerar su compleción, ¿qué sucede entonces con la envolvente probabilística de la variable aleatoria difusa en ambos casos? Probamos, gracias a la teoría de Conjuntos Graduales [3], que *si las envolventes probabilísticas de los α -cortes coinciden al considerar el espacio inicial y su completado, también coincidirán las asociadas a la variable aleatoria difusa.*

Otra cuestión importante la siguiente: dado que no podemos obtener de forma exacta la probabilidad inducida por la variable aleatoria que define el experimento (de la cual tenemos una observación imprecisa) es importante tratar de acotar lo mejor posible el intervalo de valores en los que se encuentra. En el caso en el que la información imprecisa se modela a través de un conjunto aleatorio, tenemos asociadas a él las probabilidades superior e inferior de Dempster. En [1] se estudia para qué casos dichas probabilidades nos dan la mejor cota superior e inferior posible de la envolvente probabilística, y por tanto las mejores cotas posibles para el valor de la probabilidad inducida que queremos conocer. Basándonos en la estrecha relación existente entre variables aleatorias difusas y conjuntos aleatorios, *definimos unas nuevas cotas para los posibles valores de la probabilidad inducida buscada y determinamos en qué casos los valores definidos nos dan las mejores cotas posibles.*

2. Bibliografía

- [1] I. Couso (1999) *Teoría de la Probabilidad para datos imprecisos. Algunos aspectos.* Tesis doctoral. Departamento de Estadística e I.O. y Didáctica de la Matemática. Universidad de Oviedo.
- [2] I.R. Goodman y H.T. Nguyen (1985) *Uncertainty Models for Knowledge-Based Systems.* Elsevier Science Publishers. Amsterdam.
- [3] J.A. Herencia (1995) *Origen y Uso de los Conjuntos Graduales.* Tesis Doctoral. Departamento de Matemáticas. Universidad de Córdoba.

Ordenación stop-loss multivariante de vectores de sumas aleatorias en ambientes dependientes

Félix Belzunce², Eva-María Ortega¹, Franco Pellerey³, José M. Ruiz²

¹evamaria@umh.es, Dept. Estadística, Matemática e Informática, Centro de Investigación Operativa, Universidad Miguel Hernández

²belzunce@um.es, jmruiizo@um.es, Departamento de Estadística e Investigación Operativa, Universidad de Murcia

³pellerey@calvino.polito.it, Dipartimento di Matematica, Politecnico di Torino

Abstract

En el modelo colectivo de riesgo, la cuantía total de reclamaciones se define mediante sumas aleatorias y en la literatura actuarial, se han estudiado condiciones para el orden stop-loss de sumas aleatorias. En este trabajo consideramos sumas aleatorias dependientes mediante un vector de parámetros aleatorios, y obtenemos resultados de ordenación stop-loss multivariante de vectores de sumas aleatorias, cotas para la distribución de la cuantía total de reclamaciones y aplicaciones en cálculo de primas.

Claves: stop-loss order, total claim amounts, dependence notions.

1. Introducción y motivación

Las sumas aleatorias se han utilizado para describir los tiempos de vida de determinados sistemas en fiabilidad y supervivencia, y como modelo estocástico en otras áreas de investigación. Por ejemplo, la cuantía total de reclamaciones en el modelo colectivo de riesgo en ciencias actuariales, definida en términos del número aleatorio de reclamaciones y las cuantías de las reclamaciones, y que constituye uno de los tópicos importantes de la matemática actuarial. Por otro lado, en la literatura se han estudiado condiciones para la ordenación stop-loss (denotada sl) de sumas aleatorias de variables aleatorias en el modelo colectivo de riesgo, que asume independencia entre el número de reclamaciones y las cuantías, donde (ver Shaked y Shanthikumar (1994) o Müller and Stoyan (2002)) dadas dos variables aleatorias X e Y con funciones de supervivencia \bar{F} y \bar{G} , entonces X es menor que Y en el orden stop-loss (creciente convexo), y se escribe $X \leq_{sl} Y$ si

$$\int_x^{+\infty} \bar{F}(t)dt \leq \int_x^{+\infty} \bar{G}(t)dt, \forall x \in \mathbb{R}, \quad (5.99)$$

o equivalentemente, $X \leq_{sl} Y \Leftrightarrow E[\phi(X)] \leq E[\phi(Y)]$, para toda función ϕ creciente y convexa, para la que existan las esperanzas anteriores.

El esquema general de estos resultados es el siguiente: dadas dos sucesiones de variables aleatorias no-negativas $\mathbf{X} = (X_1, X_2, \dots)$ e $\mathbf{Y} = (Y_1, Y_2, \dots)$ y dos variables aleatorias no-negativas N_1, N_2 ,

$$\mathbf{X} \leq_{ord} \mathbf{Y} \Rightarrow \sum_{i=1}^{N_1} X_i \leq_{sl} \sum_{i=1}^{N_2} Y_i. \quad (5.100)$$

donde ord denota alguna ordenación multivariante y $N_1 \leq_{ord*} N_2$, donde $ord*$ es algún orden estocástico.

La propiedad (5.100) permite comparar las primas netas de la cuantía total de reclamaciones en el principio de reaseguro de limitación de pérdida y se verifica para los órdenes estocástico multivariante, stop-loss multivariante y supermodular (Müller and Stoyan (2002)). Esta propiedad se ha estudiado también introduciendo dependencia entre el número aleatorio de sumandos y los sumandos (ver Hu and Pan (1999), Cossette and Marceau (2000), Denuit et al. (2002) o Lillo et al. (2003)). En problemas reales del ámbito financiero y actuaria, las variables aleatorias dependen de condiciones económicas, geográficas o ambientales que tienen naturaleza estocástica. El objetivo de nuestro trabajo es estudiar la propiedad (5.100) considerando vectores de sumas aleatorias dependientes de parámetros aleatorios, que constituye un contexto de dependencia que describe situaciones reales del ámbito actuaria.

2. Ordenaciones y nociones de dependencia

En esta sección definimos las ordenaciones y nociones de dependencia que utilizaremos para establecer las condiciones suficientes de los resultados (ver Joe (1997)).

3. Resultados principales

Las principales aportaciones de este trabajo son resultados de ordenación de vectores de sumas aleatorias dependientes de parámetros aleatorios: obtenemos condiciones para la ordenación $iccx, sl$ (icx) multivariante de vectores de sumas aleatorias y de valores esperados de sumas aleatorias. Como el orden sl multivariante se preserva por marginales, entonces las componentes (sumas aleatorias) están también ordenadas sl . También obtenemos cotas en orden $iccx, sl$ (icx) multivariante y de valores esperados a partir de las sumas aleatorias de variables independientes con idénticas marginales. Una generalización de la propiedad supermodular multivariante, nos permite obtener otras condiciones suficientes de nuestros resultados y de otros en la literatura. También obtenemos aplicaciones y ejemplos. Todos estos resultados pueden verse en el trabajo *Belzunce, F., Ortega, E. M., Pellerey, F. and Ruiz, J.M. (2006). "Variability of total claim amounts under dependence between claims severity and number of events" en prensa en Insurance: Mathematics and Economics.*

Paralelización embarazosa distribuida sin modificación del código fuente

C. Carleos¹, J. Á. Baro², N. Corral¹

¹carleos@uniovi.es, Departamento de Estadística e Investigación Operativa y Didáctica de la Matemática, Universidad de Oviedo

²baro@agro.uva.es, Departamento de Ciencias Agroforestales, Universidad de Valladolid

Abstract

Muchos algoritmos estadísticos son susceptibles de paralelización embarazosa. Presentamos un programa informático que permite la distribución de código fuente entre varias máquinas. Se comentan las diferentes estrategias para decidir la conclusión del cálculo.

Claves: paralelización, MCMC.

1. Introducción

Se han popularizado los métodos estadísticos que requieren uso intensivo del ordenador (EM, MCMC, recocido, ...). A pesar del rápido avance de las capacidades de los nuevos ordenadores (MHz del procesador, GB de RAM) a menudo resultan insuficientes. Si la tarea por resolver puede dividirse en subtareas independientes (*paralelización embarazosa*), es sencillo crear la infraestructura necesaria para dividir el trabajo entre distintas máquinas.

2. Métodos

El trabajo simultáneo de varios ordenadores puede inscribirse en alguna de estas dos categorías:

Clúster: Varios ordenadores fuertemente integrados en un área local con conexiones rápidas. Ejemplos: OpenMosix, Beowulf; Condor; PVM, MPI.

Grid: Ordenadores heterogéneos en redes grandes. Ejemplo: BOINC.

Para la paralelización embarazosa es suficiente con una infraestructura de grid.

Tradicionalmente, el reparto de tareas se integra con el código original mediante el uso de una biblioteca. Por ejemplo, BOINC requiere modificar las entradas y salidas del programa (C, C++ o Fortran) y luego reparte binarios a los clientes.

Nuestra propuesta consiste en repartir directamente el código fuente. Requiere la creación de un guión que se ejecutará en los clientes y un guión en el servidor que decide cómo construir las tareas y cuándo parar.

3. Criterios de finalización

Los criterios para la determinación del número de tareas paralelas por realizar varían según el estudio. En algunos casos se trata de un valor constante, como en una validación cruzada. En otros, hay que comprobar una condición de convergencia.

En problemas de optimización (algoritmo EM, recocido simulado, etc.) se fija un número mínimo de veces que se debe alcanzar el mismo óptimo.

En simulación de Montecarlo de cadenas markovianas (MCMC), el método más práctico para comprobar convergencia de forma automática es el estudio de $\hat{R} = \sqrt{1 + (B/W - 1)/n}$, donde B y W representan variación *entre* y *dentro* de cadenas [2]. Un valor $\hat{R} < 1,2$ (en cada parámetro) es suficiente para la mayoría de los casos.

Una estrategia más elaborada estudia la variabilidad de Montecarlo para vigilar el grado de convergencia a través de intervalos de confianza sobre los parámetros [1]. (Nota: este método no está completamente implementado en nuestra plataforma.)

4. Discusión

Hemos aplicado la distribución de fuente en tres proyectos: cálculo de parámetros genéticos mediante Loki; validación cruzada sobre Fuzzy-ARTMAP; optimización combinatoria con recocido.

Otros enfoques para el aseguramiento de la convergencia no son posibles con el planteamiento presentado. Por ejemplo, el muestreo perfecto requiere cadenas emparejadas, susceptibles de paralelizarse en un clúster, pero no en un grid [3].

5. Bibliografía

- [1] Jones G.L., Haran M., Caffo B.S. y Neath R. (2005). Fixed-width output analysis for MCMC. *Johns Hopkins University, Dept. of Biostatistics Working Papers*. Working paper 72.
- [2] Kass R.E. (1998). MCMC in practice. *The American Statistician* 52, 93–100.
- [3] Röhrl A.S.A. (1998). Fast, portable, parallel and scalable exact simulation using Markov chains. *Computing Science and Statistics: Proceedings of the 30th Symposium on the Interface*, <http://citeseer.ist.psu.edu/247003.html>.
- [4] Sorensen D. y Gianola D. (2002). *Likelihood, Bayesian, and MCMC methods in quantitative genetics*. Springer.

Probability Inequalities for the Ruin Probability in a Markov Controlled Risk Model

Maikol Diasparra¹, Rosario Romera²

¹mdiaspar@est-econ.uc3m.es, Departamento de Estadística,
Universidad Carlos III de Madrid

²mrromera@est-econ.uc3m.es, Departamento de Estadística,
Universidad Carlos III de Madrid

Abstract

We consider a discrete risk process with proportional reinsurance and investment in a financial market modeled by a Markov Decision Process. Recursive and integral equations for the ruin probability are given. Generalized Lundberg inequalities for the ruin probabilities are derived. Numerical results are presented.

Keywords: risk process, ruin probability, Markov decision processes.

AMS: 60K10, 62N05, 91A90.

1. Resumé

A general X -valued discrete-time stochastic process $\{X_n\}_{n \geq 0}$ is considered which can be observed and controlled at the beginning of each period. The stochastic development is determined by a sequence of random variables $\{W_n\}_{n \geq 1}$ on some probability space (Ω, \mathcal{F}, P) with $W_n = (Y_n, Z_n, R_n)$. Let $\{Y_n\}_{n \geq 1}$ and $\{Z_n\}_{n \geq 1}$ be two independent sequences i.i.d. nonnegative random variables. We assume also independency with respect to the investment return process $\{R_n\}_{n \geq 1}$. In risk theory (see [1]), $X_n \in X$ (usually $X = \mathbb{R}$) describes the *surplus* (size of the fund of reserves) of an insurance company after n periods, Y_n denotes the total claims during the n -th period, i.e. from time $n-1$ to time n , Z_n represents the random period n , and R_n denotes the total retribution of the financial market during the n -th period.

A policy is a sequence $a = \{a_n\}_{n \geq 1}$ of decision functions $\varphi_n : X \rightarrow A$, with $A = [b, 1] \times [0, 1]^d$. Then $\varphi_n(X_n) = a_n = (b_n, \delta_n)$ will represent the action chosen at the beginning of the period $n+1$, where b_n is the retention level and δ_n is the proportional investment in the financial market. We consider Markovian control policies depending only on the current state, and stationary. The measurable function $h(b, y)$ specifies the part of the claim y paid by the insurer, and the prime's function $c(b)$ may be calculated according to the expected value principle (see [3]).

For an initial state $X_0 = x$ and a control policy a , the value of the risk process at the beginning of the period $n+1$, say $X_n^{x,a} = X_n$, is given according to

$$X_n = X_{n-1}(1 + \langle \delta_{n-1}, R_n \rangle) + c(b_{n-1})Z_n - h(b_{n-1}, Y_n) \quad (5.101)$$

We adopt a realistic point of view and we let the return process $\{R_n = (R_n^1, \dots, R_n^d)\}_{n \geq 1}$ to be statistically dependent. Thus, the price process given by $I_n = \langle \delta_{n-1}, R_n \rangle$ follows a Markov chain (see [2]). We define for all $a \in A$

$$\Pr\{I_{n+1}^a = i_t | I_n^a = i_s, a\} = p_{st}^a \geq 0$$

where $\sum_{t=0}^N p_{st}^a = 1$ for $s = 0, 1, \dots, N$.

The ruin probabilities with infinite and finite horizon and with return model (5.101), initial surplus x , action a and given $I_0 = i_s$ are

$$\psi^a(x, i_s) = \Pr\left\{\bigcup_{k=1}^{\infty} (X_k^{x,a} < 0)\right\} \text{ and } \psi_n^a(x, i_s) = \Pr\left\{\bigcup_{k=1}^n (X_k^{x,a} < 0)\right\} \quad (5.102)$$

In this paper we consider a problem of minimization of the ruin probability in a discrete time risk process. Nevertheless obtaining explicit optimal solutions is a difficult task in a general setting. An analytic method commonly used in ruin theory is to derive inequalities for ruin probabilities.

The equation (5.101) can be rewritten in a recursive form and then recursive equations for the ruin probabilities (5.102) are obtained. Using inductive approach and martingale approach (see [4]), probability inequalities Lundberg-type are obtained for the ruin probabilities. To illustrate these results numerical examples for PH-type claims distributions are presented.

2. Bibliography

- [1] Asmussen, S. (2000). *Ruin Probabilities*, World Scientific.
- [2] Cai J. and Dickson D. (2004), *Ruin probabilities with a Markov chain interest model*, Insurance: Mathematics and Economics, 35, 513-525.
- [3] Schäl M. (2004). *On Discrete-Time Dynamic Programming in Insurance: Exponential Utility and Minimizing the Ruin Probability*, Scandinavian Actuarial Journal, 3, 189-210.
- [4] Willmot G. and Lin X. (2000). *Lundberg Approximations for Compound Distributions with Insurance Applications*, Springer-Verlag.

Proceso estocástico lognormal triparamétrico: Simulación y aplicación

R. Gutiérrez¹, R. Gutiérrez-Sánchez, A. Nafidi, E. Ramos

¹rgjamez@ugr.es, Departamento de Estadística e Investigación Operativa,
Universidad de Granada

Abstract

Se estudia una extensión del modelo lognormal biparamétrico unidimensional, que se obtiene añadiendo un parámetro umbral. Se muestra la simulación del proceso y su aplicación a un caso real. Se estiman los parámetros en base los datos simulados y a los datos reales.

Claves: Difusión lognormal triparamétrica; Estimación máximo-verosímil;
Simulación y aplicación.

AMS: 60J60; 62M05

1. Introducción

En las últimas décadas, se han desarrollado varios modelos estocásticos de tipo difusión definidos desde dos puntos de vista, el primero a partir de las ecuaciones diferenciales estocásticas (E.D.E.) y el segundo a partir de las ecuaciones de Kolmogorov. Dichos modelos se han aplicado con éxito en ajustes y predicciones de varios fenómenos y en diferentes campos como Física, Biología, Economía y Sociología.

El proceso de difusión lognormal es uno de los procesos más relevantes en el campo de la aplicación. Los trabajos pioneros de este último se deben a Tintner et al. [4] y Gutiérrez et al. [1], que lo aplicaron sobre todo en temas relacionados con la Economía. Se han considerado varias extensiones de dicho proceso y entre ellas se cita por ejemplo la extensión al caso no homogéneo basada en el hecho de que el proceso está afectado por factores exógenos, siendo estudiado y aplicado por los autores citados anteriormente. Otra extensión de dicho modelo se obtiene al añadir un parámetro umbral al proceso lognormal biparamétrico, como se muestra en un contexto general en Ramos, E. [3] y Gutiérrez et al. [2].

2. Objetivos

En el presente trabajo se define un nuevo proceso lognormal triparamétrico unidimensional como extensión del proceso lognormal biparamétrico, añadiendo un parámetro umbral. Este modelo se obtiene a partir de las ecuaciones de Kolmogorov. Se establece la densidad de transición y los momentos del proceso. A partir de dicha transición se estiman los parámetros por máxima verosimilitud en base a un esquema de muestreo discreto. Posteriormente, a

través de la E.D.E. de Itô, se simula la trayectoria del proceso y se estiman los parámetros de la muestra simulada. Finalmente, se aplica la metodología del presente trabajo a un caso real, se trata de la modelización de la evolución de un indicador demográfico básico como es la nupcialidad, concretamente la variable "*Edad media al primer matrimonio en varones en España*".

3. Agradecimientos

Este trabajo está parcialmente subvencionado por el Proyecto MTM2005-09209 (Ministerio de Educación y Ciencia, España).

4. Bibliografía

- [1] Gutiérrez, R., Angulo, J.M., González, A. y Pérez, R. (1991). Inference in lognormal multidimensional diffusion processes with exogenous factors: application to modelling in economics. *Applied Stochastic Models and Data Analysis*, **7**, 295-316.
- [2] Gutiérrez, R., Gutiérrez-Sánchez, R., Ramos, E. (2004). Inference in tri-parameter lognormal multidimensional diffusion processes with exogenous factors. *Proceedings of the 17 European Meeting on Cybernetics and Systems Research*, **1**, 157-162. Viena.
- [3] Ramos, E., (2005). *Difusiones lognormales triparamétricas con factores exógenos*. PhD thesis, Universidad de Granada, España.
- [4] Tintner, G. y Sengupta, J.K. (1972). *Stochastics Economics*. Academic Press.

Procesos Puntuales de Incrementos Condicionalmente Independientes

Miguel Martín Díaz¹, Tomás Prieto Rumeau², Ricardo Vélez Ibarrola³

¹Dpto. de Estadística e Investigación Operativa, Fac. de CC Matemáticas,
Universidad Complutense de Madrid. miguel_martin@mat.ucm.es

²Dpto. de Estadística, Investigación Operativa y Cálculo Numérico, Fac. de Ciencias, UNED. tprieto@ccia.uned.es

³Dpto. de Estadística, Investigación Operativa y Cálculo Numérico, Fac. de Ciencias, UNED. rvelez@ccia.uned.es

Abstract

En este trabajo, se introducen los procesos puntuales de incrementos condicionalmente independientes (que son una generalización de los procesos puntuales de incrementos independientes, como, por ejemplo, el proceso de Poisson), y los caracterizamos mediante sus distribuciones marginales y condicionadas.

Claves: procesos puntuales, procesos de Poisson, procesos de incrementos condicionalmente independientes

AMS: 60G55, 60J27

1. Introducción

Se considera un espacio métrico completo y separable E , de sigma-álgebra de Borel \mathcal{E} , y un proceso puntual $\mathbf{N} = \{N(A)\}_{A \in \mathcal{E}}$ definido sobre un espacio de probabilidad (Ω, \mathcal{F}, P) . Se supone que \mathbf{N} cumple las propiedades estándar de los procesos puntuales; ver [1].

Habitualmente, se estudian los procesos puntuales de incrementos independientes, es decir, aquéllos para los que, dados conjuntos disjuntos $A_i \in \mathcal{E}$, $i \in I$, las variables aleatorias $N(A_i)$, $i \in I$, son independientes. Los procesos de incrementos independientes incluyen a los procesos de Poisson.

En este trabajo, debilitamos la hipótesis de incrementos independientes, que es muy restrictiva, sustituyéndola por una de incrementos condicionalmente independientes. Concretamente, diremos que un proceso puntual es de incrementos condicionalmente independientes (y lo escribiremos PPICI) si, para todo $A \in \mathcal{E}$, el proceso en A es condicionalmente independiente del proceso en $N(A^c)$ dada la variable aleatoria $N(A)$. Ésta es una propiedad de tipo markoviano (cf. [2]).

Estudiaremos los PPICI que verifican, además, una propiedad de simetría respecto de una medida difusa μ y una propiedad de homogeneidad.

Nuestro objetivo es caracterizar los PPICI mediante: (i) sus distribuciones marginales $N(A) \stackrel{d}{=} \{p_j(a) : j \in \mathbb{N}\}$, donde $\mu(A) = a$, y (ii) sus distribuciones condicionadas $\{P_{ij}(x) : i \geq j \geq 0, 0 \leq x \leq 1\}$, que son las probabilidades de que, siendo $N(A) = i$, sea $N(B) = j$, donde $B \subseteq A$ y $\mu(B)/\mu(A) = x$.

2. PPICI simples

En primer lugar, estudiamos los procesos simples. En este caso, la matriz $P(x) \equiv [P_{ij}(x)]_{i,j}$ tiene filas que corresponden a distribuciones binomiales de parámetros i y x , y las funciones generatrices de las marginales, i.e. $G(a, z) = \sum_{j=0}^{\infty} p_j(a)z^j$ son de la forma

$$G(a, z) = \psi(a(1 - z)), \quad \text{para } a > 0 \text{ y } |z| \leq 1, \quad (5.103)$$

donde ψ es la transformada de Laplace de una variable aleatoria positiva.

Así, la construcción de un PPICI simple es como sigue: dado un conjunto $A \in \mathcal{E}$, se escoge el número de puntos en A a partir de (5.103), y éstos se distribuyen independiente y uniformemente (respecto de μ) en A .

Como casos particulares se obtienen los procesos de Poisson, pero también procesos geométricos (i.e. sus marginales tienen distribución geométrica) y procesos binomiales negativos, entre otros.

3. PPICI generales

Si el PPICI no es simple, logramos caracterizar sus distribuciones condicionadas a través de las características de la matriz infinitesimal $P'(1)$.

El análisis de las distribuciones marginales es sensiblemente más complicado que para los PPICI simples. En el caso de que exista una solución de la ecuación de invariancia

$$\mathbf{w}P'(1) = \mathbf{0}, \quad (5.104)$$

se puede probar que el PPICI está totalmente determinado por ψ (ver (5.103)), \mathbf{w} (ver (5.104)) y la matriz infinitesimal $P'(1)$, construyéndose las trayectorias del proceso según una pauta parecida a la descrita en la Sección 2.

4. Bibliografía

- [1] Daley, D.J. and Vere-Jones, D. (1988). *An Introduction to the Theory of Point Processes*. Springer.
- [2] Van Lieshout, M.N.M. (2000). *Markov Point Processes and Their Applications*. Imperial College Press.

Redundancia activa y en espera para sistemas con componentes dependientes

Félix Belzunce, Helena Martínez, José María Ruiz

Departamento de Estadística e Investigación Operativa, Universidad de Murcia

belzunce@um.es, helenamp@um.es, jmruizgo@um.es

Abstract

El problema de donde colocar una componente redundante es un problema importante en teoría de fiabilidad. Este problema ha sido considerado para el caso de componentes independientes por distintos autores. El objetivo de este trabajo es obtener resultados en el caso de componentes dependientes.

Claves: Ordenaciones conjuntas, redundancia activa y en espera.

AMS: 62E10, 60K10

1. Introducción

El problema de dónde colocar una componente redundante en un sistema con el fin de optimizar el tiempo de vida del mismo es uno de los problemas importantes en teoría de fiabilidad. Las formas más comunes de colocar una componente en redundancia son dos, la redundancia activa (la cual supone considerar el máximo de dos variables) y la redundancia en espera (en este caso hemos de considerar la convolución de dos variables). Para ilustrar este problema consideramos el caso sencillo en que tenemos dos componentes con tiempos de vida aleatorios T_1 y T_2 , que forman un sistema en serie. Consideremos ahora una componente adicional con tiempo de vida S que puede ser puesta en redundancia activa con cualquiera de las dos componentes anteriores. Esto daría lugar a dos sistemas $U_1 = \min\{\max\{T_1, S, T_2\}\}$ y $U_2 = \min\{T_1, \max\{T_2, S\}\}$, y el problema está en decidir cuál de estos sistemas es mejor que el otro en algún sentido probabilístico, es decir con cuál de las dos componentes sería preferible hacer la redundancia activa.

Este problema viene estudiándose de forma intensa en la década de los 90, como puede verse, por ejemplo en los trabajos de Boland, El-Newehi y Proschan (1992), El-Newehi y Sethuraman (1993), Singh y Misra (1994), Singh y Singh (1997), Mi (1999), y ha recobrado nuevo interés con los trabajos de Valdés y Zequeira (2003), Romera, Valdés y Zequeira (2004) y Valdés y Zequeira (2006). Todos estos trabajos han considerado el caso en que las componentes del sistema son independientes y el objetivo de este trabajo es considerar el problema anteriormente citado en el caso de componentes dependientes. La herramienta a utilizar en estos resultados son los órdenes conjuntos introducidos

por Shanthikumar, Yamazaki y Sakasegawa (1991) y Shanthikumar y Yao (1991).

2. Agradecimientos

Este trabajo esta financiado por el proyecto BFM2003-02947/MATE del Ministerio de Ciencia y Tecnología y CARM 00698/PI/04 de la Fundación Séneca.

3. Bibliografía

- [1] Boland, P.J., El-Newehi, E. and Proschan, F. (1992). Stochastic order for redundancy allocations in series and parallel systems. *Adv. Appl. Prob.*, **24**, 161–171.
- [2] El-Newehi, E. and Sethuraman, J. (1993). Optimal allocation under partial ordering of lifetimes of components. *J. Appl. Prob.*, **25**, 914–925.
- [3] Mi, J. (1999). Optimal active redundancy allocation in k-out-of-n system, *J. Appl. Prob.*, **36**, 927–933.
- [4] Romera, R., Valdés, J.E. and Zequeira, R.I. (2004). Active-redundancy allocations in systems. *IEEE Transactions on reliability*, **53**, 313–318.
- [5] Shanthikumar, J.G., Yamazaki, G. and Sakasegawa, H. (1991). Characterization of optimal order of servers in a tandem queue with blocking. *Oper. Res. Lett.*, **10**, 17–22.
- [6] Shanthikumar, J.G. and Yao, D.D. (1991). Bivariate characterization of some stochastic order relations. *Adv. Appl. Prob.*, **23**, 642–659.
- [7] Singh, H. and Misra, N. (1994). On redundancy allocations in systems. *J. Appl. Prob.*, **31**, 1004–1014.
- [8] Singh, H. and Singh, R.S. (1997). Note: Optimal allocation of resources to nodes of series systems with respect to failure rate ordering, *Nav. Res. Log.*, **44**, 147–152.
- [9] Valdés, J.E. and Zequeira, R.I. (2003). On the optimal allocation of an active redundancy in a two-component series system, *Stat. Prob. Lett.*, **63**, 325–332.
- [10] Valdés, J.E. and Zequeira, R.I. (2006). On the optimal allocation of two active redundancies in a two-component series system, *Oper. Res. Lett.*, **34**, 49–52.

Relaciones entre las esperanzas de Aumann y Herer de un conjunto aleatorio

Pedro Terán¹

¹teran@unizar.es, Grupo Decisión Multicriterio Zaragoza, Facultad de Económicas, Universidad de Zaragoza

Abstract

El propósito de este trabajo es analizar exhaustivamente las relaciones entre las dos nociones de esperanza de un conjunto aleatorio cerrado acotado convexo X debidas a Aumann y Herer.

Este trabajo ha sido presentado al Premio Ramiro Meléndreras 2006.

Claves: Conjunto aleatorio. Esperanza. Geometría Estocástica.

AMS: 60D05

1. Planteamiento del problema

La esperanza de Aumann E_AX se define como la clausura del conjunto de esperanzas de sus selecciones. La esperanza de Herer E_HX es el lugar geométrico de los puntos x tales que todo punto está al menos tan cerca de x como lo está en media del punto más lejano de X .

Se conoce que si X es unipuntual, entonces ambas coinciden con la esperanza habitual, que $E_AX \subset E_HX$ y que ambas cumplen una ley fuerte de los grandes números. Hess [2] demostró que en un espacio de Hilbert ambas esperanzas son iguales siempre que el conjunto aleatorio cumpla una condición de integrabilidad cuadrática: $E\|X\|^2 < \infty$.

Pero en general no se conoce *para ningún espacio de Banach, ni siquiera \mathbf{R}* si ambas esperanzas son en general iguales o distintas, y en ese caso qué relación puede determinarse entre ellas. Uno de los problemas abiertos planteados en el libro "*Theory of Random Sets*"[4] de Ilya Molchanov es el de dar condiciones minimales sobre un espacio de Banach que garanticen que las dos definiciones son equivalentes.

2. Resultados

El problema de la relación entre las esperanzas de Aumann y Herer resulta estar íntimamente relacionado con las propiedades de la familia \mathcal{M} de todas las intersecciones de bolas cerradas. Intuitivamente, las dos esperanzas estarán más cercanas cuantos más conjuntos puedan ponerse como intersección de bolas.

Por su parte, las propiedades de \mathcal{M} pueden estudiarse a través de las propiedades geométricas de la bola unidad del espacio dual, específicamente de

sus puntos extremos o sus diversas generalizaciones en espacios de dimensión infinita. Esta labor comenzó a llevarse a cabo en la literatura analítica sólo muy recientemente [1] y por tanto nos hemos visto en la necesidad de desarrollar diversos resultados en esta línea para afrontar nuestro problema y en los que se fundamentan parcialmente nuestros resultados principales.

Entre estos últimos se cuentan:

1. La esperanza de Herer siempre está acotada superior e inferiormente por funciones de la de Aumann.
2. Bajo condiciones no muy restrictivas sobre el espacio (en particular si el dual es separable), se tiene que $E_H X$ es la intersección de todas las bolas que contienen a E_AX .
3. Si X es un conjunto aleatorio compacto (alternativamente, es de Mazur), entonces $E_H X$ es siempre la intersección de todas las bolas que contienen a E_AX .
4. La solución al problema planteado por Molchanov es que las esperanzas son iguales si y sólo si todo conjunto cerrado acotado convexo puede ponerse como intersección de bolas (Propiedad de la Intersección de Mazur [3]).

Los resultados obtenidos y las dos nociones de esperanza son coherentes con el enfoque axiomático de la esperanza en espacios métricos que hemos propuesto en [5].

3. Bibliografía

- [1] A. S. Granero, J. P. Moreno, R. R. Phelps (2004). Convex sets which are intersections of closed balls. *Adv. in Math.* **183**, 183–208.
- [2] C. Hess (2000). The Doss integral for random sets. Comparison with the Aumann integral. *Proc. 8th Int. Conf. on Information Processing and the Management of Uncertainty in Knowledge-Based Systems*, 515–520.
- [3] S. Mazur (1933). Über schwache Konvergenz in den Räumen L^p [*Sobre la convergencia débil en espacios L^p* , en alemán]. *Studia Math.* **4**, 128–133.
- [4] I. S. Molchanov (2005). *Theory of random sets*. Springer, Berlin.
- [5] P. Terán, I. Molchanov (200x). The law of large numbers in a metric space with a convex combination operation. *J. Theoret. Probab.*, en prensa.

Restauración de imágenes ruidosas usando información de covarianza

M. J. García-Ligero¹, A. Hermoso¹, J. Linares¹, S. Nakamori²

¹mjgarcia@ugr.es, ahermoso@ugr.es, jlinares@ugr.es, Departamento de Estadística e Investigación Operativa, Universidad de Granada

²nakamori@edu.kagoshima-u.ac.jp, Department of Technology, Faculty of Education Kagoshima University

Abstract

Se considera el problema de restauración de imágenes afectadas por un ruido blanco. Bajo la suposición de que el modelo de espacio de estados de la imagen original es desconocido y que el ruido blanco y el proceso imagen son correlados, proponemos un algoritmo recursivo de filtrado de mínimos cuadrados obtenido mediante la técnica de innovaciones.

Claves: Filtrado de imágenes, ecuación de Wiener-Hopf.

AMS: 62M20, 60G35.

1. Introducción

En muchas situaciones prácticas, las imágenes observadas son versiones degradadas de la imagen original. Estas degradaciones pueden ser debidas a diferentes causas como, por ejemplo, a errores en la transmisión, que se modelizan mediante componentes aleatorias denominadas ruido. Así, la imagen observada se modeliza como la imagen original más la componente aleatoria o ruido. El problema de restauración de imágenes afectadas por un ruido aditivo ha sido abordado utilizando técnicas de filtrado clásico. Concretamente, las técnicas de estimación recursiva han merecido una atención considerable en la literatura ya que el esfuerzo computacional para desarrollar filtros recursivos es menor que para métodos iterativos. En este sentido, la aplicación del filtrado de Kalman para restaurar imágenes se ha utilizado ampliamente, desarrollándose diferentes versiones del filtro de Kalman en dos dimensiones que han sido útiles para restaurar imágenes perturbadas por un ruido blanco gaussiano, debido a la eficiencia computacional que proporcionan.

El problema de restauración de imágenes utilizando algoritmos de filtrado tipo Kalman requiere el conocimiento completo del modelo de espacio de estados. Sin embargo, no siempre se dispone de dicho modelo y, por tanto, no es posible utilizar técnicas recursivas como el filtro de Kalman. En esta situación, el problema de estimación puede tratarse usando la información proporcionada por la función de covarianzas de la imagen y la del ruido aditivo de la observación. En particular, Nakamori [1] y Nakamori et al. [2] han obtenido algoritmos recursivos para problemas de filtrado y suavizamiento intervalo fijo

y sus resultados se han aplicado para restaurar imágenes perturbadas por un ruido blanco gaussiano.

En este trabajo, consideramos el problema de restauración de imágenes monocromáticas suponiendo que no se dispone de la ecuación que describe el proceso imagen, bajo las hipótesis de que la imagen observada está perturbada por un ruido blanco y que los procesos imagen y ruido son correlados. Para estimar el nivel de gris de la imagen en cada coordenada tendremos en cuenta la información proporcionada por algunos niveles de gris localizados en la misma fila del pixel que se quiere estimar. La obtención del algoritmo recursivo de filtrado sólo requiere para su aplicación el conocimiento de los momentos de primer y segundo orden del proceso imagen y del ruido de la observación, así como la función de correlación entre ambos. El algoritmo de filtrado que proponemos se obtiene utilizando un enfoque por innovaciones el cual, como es conocido, proporciona una deducción simple de los algoritmos de estimación de menor error cuadrático medio. Finalmente, para estudiar el comportamiento de los estimadores se obtienen fórmulas recursivas para las matrices de covarianzas de los errores de estimación.

2. Agradecimientos

Este trabajo ha sido parcialmente subvencionado por el "Ministerio de Educación y Ciencia". Proyecto MTM2005-03601.

3. Bibliografía

- [1] Nakamori S. (1993). Optimal Filtering Algorithm using Covariance Information in Linear Discrete-time Distributed Parameter Systems. *Signal Processing* 33, 45-56.
- [2] Nakamori S., García-Ligero, M.J., Hermoso-Carazo A. y Linares-Pérez J. Derivation of Fixed-Interval Smoothing using Covariance Information in Distributed Parameter Systems. *Applied Mathematics and Computation*. (doi:10.1016/j.amc.2005.012).

Una generalización de la distribución Beta-Binomial. Aplicaciones y propiedades

Antonio Conde Sánchez¹, M^a José Olmo Jiménez², José Rodríguez Aví³, Antonio José Sáez Castillo⁴

¹aconde@ujaen.es, ²mjolmo@ujaen.es, ³jravi@ujaen.es, ⁴ajsaez@ujaen.es,
Departamento de Estadística e I.O., Universidad de Jaén

Claves: Distribución Beta-Binomial, mixturas, sobredispersión.

AMS: 60E05

Abstract

En este trabajo se describen algunas propiedades y aplicaciones de una nueva distribución generada por la función hipergeométrica de Gauss y que puede verse como una extensión de la distribución Beta-Binomial. Al igual que ésta, se trata de una mixtura de la distribución Binomial con una generalización de la distribución Beta.

1. Introducción

La distribución Beta-Binomial se ha utilizado para modelizar datos binarios de recuento en presencia de sobredispersión (Chatfield y Goodhart [4], Williams [6]). Esta distribución puede obtenerse como mixtura de una distribución Binomial y una distribución Beta. Rodríguez et al. [5] han presentado una generalización de dicha distribución introduciendo un nuevo parámetro, que también surge como mixtura de la distribución Binomial, en este caso, con una distribución Beta generalizada. Una de las ventajas de esta nueva distribución es su mayor flexibilidad, que permite ajustar conjuntos de datos que no pueden modelizarse adecuadamente mediante la distribución Beta-Binomial.

En este trabajo se continúa el estudio de esta generalización aportando nuevas propiedades y aplicaciones. Para mostrar las mismas se utilizan los datos presentados en Alanko y Lemmens [2] sobre el número de días a la semana en los que se consume alcohol. Entre otras, se tratarán las siguientes cuestiones:

- Predicción
- Distribución condicional de la propensión.

2. Predicción

En Alanko y Duffy [1] se señala que hay dos clases de predicciones que son de interés:

1. Se pueden estimar las distribuciones relacionadas con la frecuencia de un fenómeno particular en un periodo de tiempo mayor basándose en los datos de encuesta relativos a un periodo de referencia más corto. Por ejemplo, se puede predecir la proporción de consumidores que beben menos (o más) de un número particular de días en un periodo de tiempo mayor que el periodo de referencia observado.
2. Si se consideran dos periodos de tiempo que no se solapan y en cada uno de ellos se define la variable de interés "número de intervalos de tiempo (días) en los que, al menos, un suceso ocurre", puede estimarse la esperanza condicionada de una variable dada la otra. Es decir, se puede estimar el número esperado de días al año que consume alcohol una persona según el número de días a la semana que ha bebido alcohol.

3. Distribución condicional de la propensión

Es conocido que para la Beta-Binomial la distribución condicionada de $P|_{X=x}$ es Beta. Esto mismo sucede con la generalización aquí propuesta. Así, aunque la propensión a beber, P , no es directamente observable, se puede obtener el valor esperado para P dado un número observado de días en los que se bebe.

4. Bibliografía

- [1] Alanko, T. and Duffy, J.C. (1996). Compound Binomial distributions for modelling consumption data. *The Statistician*, 45, 269-286.
- [2] Alanko, T. and Lemmens, P.H. (1996). Response effects in consumption surveys: an application of the Beta-Binomial model to self-reported drinking frequencies. *Journal of Official Statistics*, 12, 253-273.
- [3] Brooks, S.P., Morgan, B.J.T., Ridout, M.S. and Pack, S.E. (1997). Finite mixture models for proportions. *Biometrics*, 57, 1097-1115.
- [4] Chatfield, C. and Goodhardt, G.J. (1970). The Beta-Binomial model for consumer purchasing behaviour. *Applied Statistics*, 19, 240-250.
- [5] Rodríguez-Avi, J., Conde-Sánchez, A., Sáez-Castillo, A.J. and Olmo-Jiménez, M.J. (2006). A generalization of the Beta-Binomial distribution. *Submitted to Applied Statistics*.
- [6] Williams, D. (1975). The analysis of binary responses from toxicological experiments involving reproduction and teratogen. *Biometrics*, 31, 949-952.

Una solución al problema de estimación no lineal basada en los desarrollos en serie aproximados

R. M. Fernández Alcalá¹, J. Navarro Moreno², J. C. Ruiz Molina³

¹rmfernand@ujaen.es, Departamento de Estadística, Investigación Operativa,
Universidad de Jaén

²jnavarro@ujaen.es, Departamento de Estadística, Investigación Operativa,
Universidad de Jaén

³jcruiz@ujaen.es, Departamento de Estadística, Investigación Operativa,
Universidad de Jaén

Abstract

Se presenta un nuevo algoritmo de estimación no lineal basado en especificaciones de la covarianza de la señal. El método propuesto utiliza los desarrollos de Karhunen-Loève aproximados para la identificación de un modelo de estado aproximado para la señal de interés y entonces, se aplican las técnicas de estimación estándar sobre el modelo identificado. La principal ventaja de nuestra solución es que puede ser aplicada sin necesidad de que la señal verifique un modelo de espacio de estados¹.

Palabras Clave: Estimación no lineal, representaciones en serie de procesos estocásticos.

AMS: 60G35, 93E10

¹Este trabajo ha sido parcialmente financiado por el Ministerio de Ciencia y Tecnología, proyecto MTM2004-04230 del Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica. Este proyecto está cofinanciado por el FEDER.

Uso del Método de Rayleigh-Ritz para la Simulación de Campos Aleatorios

A. Oya Lechuga¹, J. Navarro Moreno², J.C. Ruiz Molina³

¹aoya@ujaen.es, Departamento de Estadística e I.O, Universidad de Jaén

²jnavarro@ujaen.es, Departamento de Estadística e I.O, Universidad de Jaén

³jcrui@ujaen.es, Departamento de Estadística e I.O, Universidad de Jaén

Abstract

El objetivo de esta comunicación es proponer un método de simulación para campos aleatorios utilizando el método de Rayleigh-Ritz. La metodología que se presenta únicamente requiere conocer la función de correlación del campo aleatorio pero no exige que éste sea estacionario o Gaussiano.

Palabras Claves: Campos aleatorios, Método de Rayleigh-Ritz, Representación de Karhunen-Loève, Simulación.

AMS: 60G12

1. Introducción

La teoría de campos aleatorios proporciona modelos matemáticos apropiados utilizados frecuentemente para modelizar entidades dependientes del entorno, esto es, características espaciales correladas que se pueden encontrar en muchas aplicaciones incluyendo ecología, ingeniería, hidrología, etc. En particular, la simulación de campos aleatorios es una potente herramienta para generar realizaciones numéricas que tienen en cuenta la dependencia estructural en el espacio, del proceso natural [1]. De tal forma, que la investigación de las propiedades del campo generado nos proporcionará información útil sobre las variaciones espaciales del proceso real. Una gran variedad de métodos han sido desarrollados para simular campos aleatorios [1]-[3]. Sin embargo, es usual que en muchos de estos procedimientos se requiera que el campo aleatorio sea estacionario y/o Gaussiano, hipótesis esta última que en muchos casos se asume por conveniencia en el análisis. Una de las metodologías de simulación más generales, aplicable incluso en aquellos casos en los que no es apropiado considerar que el campo aleatorio sea Gaussiano, es la basada en la representación de Karhunen-Loève del campo aleatorio [1].

Desde el punto de vista de la aplicación práctica, la técnica de simulación de Karhunen-Loève no es completamente satisfactoria. En efecto, su implementación requiere que los autovalores y las autofunciones de la función de correlación implicada sean calculados explícitamente. El hecho de la no existencia de procedimientos generales de resolución de la ecuación del operador asociado hace que

la obtención de los autovalores y autofunciones sea una tarea complicada. La propuesta alternativa consiste en utilizar un procedimiento numérico, el método de Rayleigh-Ritz, para la resolución aproximada de la ecuación del operador [4]. El uso de este método proporciona una base para construir un desarrollo aproximado para campos aleatorios que nos permite generar realizaciones a partir únicamente del conocimiento de la función de correlación.

2. Agradecimientos

Este trabajo ha sido parcialmente financiado por el Ministerio de Educación y Ciencia, Proyecto MTM2004-04230 del Plan Nacional de I+D+I. Este proyecto está financiado conjuntamente por FEDER.

3. Bibliografía

- [1] Christakos G. (1992). *Random field models in earth sciences*. Academic Press.
- [2] Yamazaki F. and Shinouzuka M. (1988). Digital generation of non-Gaussian stochastic fields. *J Eng Mech (ASCE)* 114(7), 1183-1197.
- [3] Deodatis G., Popescu R. and Prevost JH. (1998). Simulation of homogenous non-Gaussian stochastic vector fields. *Probabilistic Eng Mech* 13(1), 1-13.
- [4] Chen M., Chen Z. and Chen G. (1997). *Approximate solutions of operator equations*. World Scientific Pub. Co.

Índice de autores

- Abad Montes, Francisco, 53
Abdul-Jalbar Betancor, Beatriz, 597, 619
Adell Pascual, José Antonio, 273, 685, 687
Aguilera del Pino, Ana María, 361
Álamo Santana, Fayna, 385
Albarracín Guillem, José Miguel, 533, 643
Alberto Moralejo, I. Olina, 559
Albiach Vicent, José, 523
Alcaide López de Pablo, David, 557, 579
Alcalá Nalváiz, José Tomás, 83, 319
Alcaraz Soria, Javier, 651
Alemán Jiménez, Mercedes, 587
Alfaro Navarro, Jose Luis, 177, 329
Allahverdi, Ali, 503
Almenara Barrios, Jose, 199
Almiñana Alemany, Marc, 417, 419
Alonso Ayuso, Antonio, 599
Alonso Fernandez, Andres M., 269
Alonso Meijide, José M^a., 539
Alonso Morales, Francisco Javier, 175
Alonso Pérez, Estrella, 531
Alonso Rodríguez, Sergio, 365, 371, 405, 659, 677
Alonso, Andrés M., 91, 243
Alvarez Esteban, Pedro César, 337
Álvarez Esteban, Ramón, 349
Alvarez Gonzalez, Francisco, 231
Alvarez Valdés Olaguíbel, Ramón, 403, 505, 667
Amer Ramon, Rafael, 391, 397
Amo Salas, Mariano, 171
Andrés Romano, Carlos, 469, 475
Aneiros Pérez, Germán, 145
Angel García-Escudero, Luis, 347
Angulo Ibáñez, José Miguel, 93, 175, 179, 201
Aparicio Baeza, Juan, 547, 549
Aparicio Izquierdo, Francisco, 249
Aparisi, Francisco, 117, 141
Arana Jiménez, Manuel, 457
Arbelo Álvarez, Antonio, 63
Arcas Pons, Antoni, 305
Arcos Cebrián, Antonio, 353
Arenas Ramírez, Blanca, 249
Armada Estévez, Antonio, 661
Armengol, Josep María, 191
Arroyo Pérez, Andrés, 227
Artés Rodríguez, Eva M., 181
Artalejo, Jesús, 425
Arteaga Moreno, Francisco Javier, 15, 253
Artiles Romero, Juan, 97
Ausín Olivera, María Concepción, 75, 315
Avila Zarza, Carmelo, 19
Ayala Gallego, Guillermo, 335
Azcárate Camio, Cristina, 477
Babiloni, M., 533, 643
Bacala, Nora, 47
Balasch Parisi, Sebastiá, 239
Ballestín, Francisco, 407
Barceló Cerdá, Susana, 293

- Barea Barrera, Rosario del Carmen, 51
- Bargueño Díaz-Villarejo, Félix, 411, 413
- Baro, Jesús Ángel, 721
- Barranco Chamorro, Inmaculada, 213
- Barreiro, Juan José, 111
- Barrios, Agustín, 407
- Bautista Valhondo, Joaquín, 445, 465, 497, 577
- Beamonte San Agustín, M^a Asunción, 285 285
- Beato Moreno, Antonio, 511, 581
- Bécue-Bertaut, Mónica-María, 349
- Belzunce Torregrosa, Felix, 217, 719, 729
- Bergantiños Cid, Gustavo, 543, 555, 563
- Berihuete Macias, Angel, 101
- Bermúdez Edo, José Domingo, 217, 287, 607
- Bermúdez Parrado, Silvia, 227
- Berné, Carmen, 103
- Bernardos Rodríguez, Eva, 249
- Bielza Lozoya, Concha, 491
- Blázquez Zaballos, Antonio, 37, 49, 79
- Blanco Izquierdo, Víctor, 459
- Blanquero Bravo, Rafael, 227, 653
- Boj del Val, Eva, 279
- Borm, Peter, 611
- Brezmes Brezmes, Teófilo, 717
- Caballero Águila, Raquel, 709, 711
- Cabrera, Enrique, 525
- Calduch Losa, Ángeles, 325, 333, 343
- Calvo, Miquel, 305
- Cancela Carollo, Alberto, 327
- Cano Alsúa, Santiago, 393
- Cano Belmán, Jaime, 497, 577
- Cano Sevilla, Francisco, 393
- Cánovas Cánovas, M^a Josefa, 517, 589, 617
- Cao Abad, Ricardo, 257
- Capacho, Liliana, 467
- Capilla Romá, Carmen, 239
- Cárdenas, Olesia, 81
- Carleos Artíme, Carlos Enrique, 721
- Carlos Alberola, Sofía, 409, 565
- Carollo Limeres, M^o Carmen, 327
- Carot Sierra, José Miguel, 57, 209, 151, 241
- Carpente Rodríguez, María Luisa, 509
- Carreras Escobar, Francesc, 427, 551, 569
- Carrión García, Andrés, 341, 525
- Carrillo Fernández, Marianela, 655
- Carrillo Ruiz, Jose Antonio, 621
- Carrizosa Priego, Emilio, 625, 647, 653, 679
- Casado Yusta, Silvia, 609
- Casado, David, 91
- Casas Méndez, Balbina, 509, 539
- Cascos Fernández, Ignacio, 221, 693
- Castaño Martínez, Antonia, 233, 271
- Castellano Méndez, María, 119, 121
- Castellote Varona, Antonio, 411, 413
- Castro Cantalejo, Javier, 591
- Castrodeza Chamorro, Carmen, 583
- Chacón Durán, José Enrique, 301
- Chapinal, Nuria, 67
- Claramunt, M. Mercè, 279
- Clemente Marín, Gonzalo, 565
- Coladas Uría, Luis, 439
- Colebrook Santamaría, Marcos, 405, 597, 677
- Colubi Cervero, Ana, 219, 345
- Conchado Peiró, Andrea, 151, 241
- Conde Sánchez, Antonio, 715, 735
- Coppi, Renato, 219

- Corberán Vallet, Ana, 287
Corominas, Albert, 451, 499, 593
Corpas Nogales, Elena, 183
Corral Blanco, Norberto, 721
Cortés, Angela, 33
Couso Blanco, Inés, 483
Coves Moreno, Anna María, 401, 481, 527
Crespo i Escobar, Enric, 505
Cristóbal Cristóbal, José Antonio, 83, 319
Crujeiras-Casais, Rosa María, 99
Cruz, João Pedro, 195
Cuadra Carreño, Juan, 701
Cubiles de La Vega, María Dolores, 237
Cucala García, Asunción Paloma, 663

Daunis i Estadella, Josep, 67, 153, 311
De La Fuente, David, 495
De La Horra Navarro, Julián, 283
De Las Rivas Sanz, Javier, 27
De Padura y Dieguez, Juan Antonio, 411, 413
De Waard, Huub W., 395
Debón Aucejo, Ana, 307, 525
Del Águila del Águila, Yolanda, 131
Del Pozo, Mónica, 635
Delgado Alvarez, Francisco Javier, 229
Delgado Quintero, Sergio, 385, 673
Delicado Useros, Pedro, 255, 261
Demiguel, Victor, 567
Diasparra Ramos, Maikol Alejandra, 723
Díaz Barrero, José Luis, 703
Díaz Díaz, Norberto, 653
Díaz de Sarralde, Santiago, 43, 435
Díaz Vázquez, Susana, 585, 657
Dolors Llongueras, Maria, 427, 569
Domingo Esteve, Juan, 335
Domingo Ferrer, Josep, 381

Dominguez Rios, Miguel Angel, 405
Dominguez, Elena, 61
Dorta González, María Isabel, 471
Dorta González, Pablo, 431, 515, 529
Dorta Guerra, Roberto, 133, 143, 371, 619
Duque-Gómez, Carlos, 365

Edo Aparicio, Miguel Ángel, 325
Egozcue Rubí, Juan José, 39, 355, 703
Enguix González, Alicia, 41, 213
Eraso Goicoechea, M^a Luisa, 477
Escabias Machuca, Manuel, 361
Escribano Ródenas, María del Carmen, 513
Escribano, Alvaro, 109
Escudero Bueno, Laureano F., 417, 419, 599
Esparcia, A., 533, 643
Espejo Miranda, M^a Inmaculada, 461, 463
Espinel Febles, María Candelaria, 223
Esteban Lefler, María Dolores, 155, 367
Estévez Fernández, Arantza, 611

Faraldo Roca, P., 99
Färber, Gerrit, 401, 481
Febles Acosta, Jaime, 455
Febrero Bande, Manuel, 121
Felipe Ortega, Angel, 393
Femia Marzo, Pedro, 317
Fermín P., Wilmer J., 363
Fernández Alcalá, Rosa María, 737
Fernández Aréizaga, Elena, 445, 493, 629
Fernández Barberis, Gabriela Mónica, 513, 545
Fernández Cardador, Antonio, 663
Fernández de Castro, B. M., 99

- Fernández del Pozo, Juan Antonio, 491
 Fernández García, Francisco R., 639
 Fernández Gómez, María José, 37, 47, 49
 Fernández Hernández, Pascual, 423
 Fernández, Isabel, 495
 Fernández, M^a José, 33
 Fernández Pascual, Rosaura, 201
 Fernández Ponce, José María, 51, 321, 697
 Fernández Rodríguez, Arturo Javier, 127, 263
 Fernández Temprano, Miguel, 277
 Fernández-Calvillo Villodre, Gemma, 671
 Ferreira, Eugénio C., 561
 Ferrer Martí, Laia, 453
 Ferrer Riquelme, Alberto, 253, 293
 Ferrer Sánchez, Joaquín, 701
 Fortiana Gregori, Josep, 23, 279
 Framiñán Torres, José Manuel, 501
 Freitas Lopes, Hedibert, 315
 Freixas Bosch, Josep, 551, 553
 Frenkel, Sergey, 557
 Frías Bustamante, María del Pilar, 175
 Gáfaro Rojas, Aurora Inés, 477
 Galeano San Miguel, Pedro, 113
 Galindo Villardón, M^a Purificación, 33, 37, 47, 49, 55, 79, 81, 87, 229, 247, 309, 363
 Gámez Cámaras, Manuel, 105
 Garay, József, 105
 Garcet Rodríguez, Sandra Adelaida, 129
 García Díaz, Juan Carlos, 89, 117, 141
 García de Las Heras, Joaquín, 237
 García García, Gloria M., 295
 García, Gastón, 35
 García, Inmaculada, 423
 García, J. P., 533, 643
 García Jurado, Ignacio, 509, 641
 García Laguna, Juan, 443, 473, 575, 665, 675
 García Lapresta, José Luis, 657
 García Leal, Julia, 95, 183, 207
 García Luengo, Amelia V., 181
 García Nogales, Agustín, 301
 García Pérez, Alfonso, 73
 García Román, Manuel, 365
 García Sipols, Ana E., 109
 García-Ligero Ramírez, M^a Jesús, 107, 733 733
 García-Martos, Carolina, 61, 243, 251
 Gargallo Valero, Pilar, 103, 285
 Garrido Blanco, Laura, 717
 Gibert Oliveras, Karina, 147, 35
 Gil Álvarez, María Ángeles, 345
 Gil Álvarez, Pedro, 585, 717
 Gil Saura, Irene, 15
 Giménez Cánovas, Domingo, 383
 Giménez Pradales, José Miguel, 391, 397
 Ginebra Molins, Josep, 123, 133, 137, 143, 193
 Giner Bosch, Vicent, 325, 333, 343
 Gispert, Rosa, 137
 Goberna, Miguel Ángel, 479
 Gómez Gasquet, Pedro, 475
 Gómez Gómez, Alberto, 495
 Gómez Gómez, M^a Teresa, 51, 321, 697
 Gómez González, Daniel, 591, 605, 635
 Gómez Rúa, María, 541
 Gómez Senent, Francisco José, 617
 Gómez Villegas, Miguel Ángel, 139, 291
 González Alcón, Carlos M., 519, 587
 González Arangüena, Enrique, 635
 González Caballero, Juan Luis, 199

- González, Camino, 323
González, Carlos, 111
González Carmona, Andrés, 187, 189
González Dávila, Enrique, 63, 133, 143, 371
González Díaz, Julio, 641
González Farias, Isabel, 59, 289
González Fernández, Camino, 249
González-Gallarza Granizo, Martina, 15
González Manteiga, Wenceslao, 167, 187, 377
González Martín, Carlos, 579, 659, 669
González Pérez, Beatriz, 139
González Rodríguez, Gil, 219, 345
González Scotto, Manuel, 185, 195
González Sierra, Miguel A., 371
González Yanes, J. Alberto, 371
Gordaliza Ramos, Alfonso, 347
Gordillo Santofimia, José F., 625
Gouet Bañares, Raúl, 695
Gozálvez, Javier, 415
Grané Chávez, Aurea, 23
Guardiola Alcalá, Luis Antonio, 437
Gutiérrez Expósito, José Miguel, 597, 619
Gutiérrez Jáimez, Ramón, 689, 713, 725
Gutiérrez Pérez, José María, 101, 231
Gutiérrez Sánchez, Ramón, 725
Gutiérrez Vaquero, César, 433
- Hall, Andreia, 185, 195
Hammers, Herbert, 421, 611
Hassin, Rafael, 639
Hermoso Carazo, Aurora, 107, 709, 711, 733
Hernández Flores, Carmen Nieves, 97, 173
- Hernández Jiménez, Beatriz, 511, 581
Hernandez, Leonor, 241
Hernández Martínez, Pedro J., 357
Herrador, Montserrat, 367
Herrera Cuadra, Francisco, 701
Hidalgo Ruiz, Eva, 207
Hidalgo, Mario, 613
Hobza, Tomas, 155
Holler, Manfred J., 539
Huete Morales, María Dolores, 53
- Iglesias Pérez, María del Carmen, 265
Iglesias Patiño, C. L., 99
Iradi Arrieta, Josu, 369
- Jabaloyes Vivas, José Manuel, 57, 151, 209, 241, 341
Jácome Pumar, María Amalia, 265
Jiménez Fernández, Javier, 323
Jiménez Jiménez, Andres, 101, 231
Jiménez López, José Domingo, 709, 711
Jiménez Martín, Bienvenido, 433
Jiménez Meana, Jorge, 449
Jodra Esteban, Pedro, 273, 685
Jones, M. C., 13
Jorge Santiso, Jesús Manuel, 649, 655
Jornet Pla, Valentín, 479, 489
Josa Fombellida, Ricardo, 623
Judice, Joaquim, 5
- Kalcsics, J., 487
Klatte, Diethard, 589
Kontoghiorghes, Erricos, 3
Kubiak, Wieslaw, 593
- Lacruz Casauau, Beatriz, 71, 169
Lagares Franco, Carolina, 199
Landete Ruiz, Mercedes, 417, 419
Lara Porras, Ana María, 95, 183, 207

- León Hernández, Coromoto, 471
 León Mendoza, María Teresa, 335
 Lehtonen, Risto, 9
 Lekuona Amiano, Alberto, 687
 Letón Molina, Emilio, 297
 Levin, Ilya, 557
 Levner, Eugene, 557
 Lillo Rodríguez, Rosa Elvira, 31
 Linares Pérez, Josefa, 709, 711, 733
 Llamazares Rodríguez, Bonifacio, 399
 Llorca Pascual, Natividad, 415
 Lombardía Cortiña, María José, 99, 373, 377
 López Blázquez, F., 233
 López Cerdá, Marco Antonio, 517, 589
 López Díaz, Miguel, 693
 López Espín, Jose Juan, 383
 López Fidalgo, Jesús F., 27, 129, 131, 135
 López García, Inmaculada, 105
 López Herrero, María Jesús, 447
 López López, Angel, 313
 López López, Victoria, 605
 López Lorente, F. Javier, 695, 707
 López Pintado, Sara, 91
 Lopez Redondo, Juana, 423
 López Vizcaíno, Esther, 211, 375
 Lorente Marín, Ana, 615
 Lorenzo Freire, Silvia, 539, 563
 Lorenzo Picado, Leticia, 543
 Lova Ruiz, Antonio, 633
 Lucas, M. del Carmen, 415
 Luengo Merino, Inmaculada, 97, 173
 Luis Ruiz, José, 547
 Luna del Castillo, Juan de Dios, 157, 159, 317
 Macías Páez, Rodrigo, 93, 189
 Magña Nieto, Antonio, 397
 Maín Yaque, Paloma, 291
 Maldonado Jurado, Juan Antonio, 107
 Mallor Giménez, Fermín, 571
 Manjarres, Liney A., 525
 Manteca, Xavier, 67
 Manuel, Conrado, 635
 Manzoni, Pietro, 523
 Marco Almagro, Lluís, 149
 Marhuenda García, Yolanda, 155, 367
 Marín-Sanguino, A., 519
 Marmolejo Martín, Juan Antonio, 53
 Maroto Álvarez, Concepción, 651
 Martín Apaolaza, Nirian, 17, 303
 Martín-Barragán, Belén, 647
 Martín Bernia, Encarnación, 379
 Martín Díaz, Miguel, 727
 Martín Fernández, Josep Antoni, 311, 351
 Martín González, Germán, 403
 Martín Moreno, R. M., 19
 Martín Vallejo, Javier, 37, 49, 363
 Martínez Alzamora, Nieves, 161
 Martínez Gómez, Mónica, 57, 151, 209, 241
 Martínez Lopez, Ignacio, 125
 Martínez Martínez, Iosu Mirem, 663
 Martínez Miranda, M. Dolores, 187, 353
 Martínez Molada, Eulalia, 523
 Martínez Ortigosa, Pilar, 423
 Martínez Rodríguez, Ana María, 645
 Martínez Ruiz, Francisco, 307
 Martínez Talaván, Pedro, 29, 429
 Martínez, Helena, 729
 Martínez, Noemi, 103
 Martorell Alsina, Sebastián, 409, 565
 Mateo Collazos, Pedro, 559, 571
 Mateo Sanz, Josep Maria, 267, 381

- Mateu-Figueras, Gloria, 67, 311,, 705
Meca Martínez, Ana, 681
Minella, Gerardo, 521
Miñarro, Antoni, 305
Miquel Fernández, Silvia, 421
Mira Mcwilliams, José, 61, 323
Miralles Insa, Cristóbal, 469
Miró i Martínez, Pau, 57, 209
Moguerza, Javier M., 507
Molero Gómez, Manuel, 411, 413
Molina Peralta, Isabel, 373, 377
Molina Vila, Mariola, 479
Momparler Pechuán, Juan, 613
Moncada Herrera, Juan, 55
Mondéjar Jiménez, José, 329
Monge Ivars, Juan Francisco, 417, 419
Montanero Fernández, Jesús, 301
Montenegro, Manuel, 219
Montero Alonso, Miguel Ángel, 93, 165, 189
Montero de Juan, Javier, 605
Montes Diez, Raquel, 245
Montes Rodríguez, Susana, 449, 483, 585, 657
Montes Suay, Francisco, 307
Montesinos Guillot, Amparo, 239
Montoro Cazorla, Delia, 485
Morales González, Domingo, 25, 155, 367, 377
Moreno Díaz, Arminda, 631
Moreno García, Toni, 275
Moreno Rebollo, Juan Luis, 41, 213
Morgado Nunes, Sara, 55, 247
Mosquera Rodríguez, Manuel Alfredo, 611
Muñoz Fernández, Ovidio, 411, 413
Muñoz Orellana, Juan Antonio, 165, 189
Muñoz Pichardo, Juan Manuel, 41, 213, 237
Mullor Ibáñez, Rubén, 161
Nafidi, Ahmed, 725
Nakamori, Seiichi, 709, 711, 733
Navarro, Jorge, 357
Navarro Moreno, Jesús, 737, 739
Nickel, S., 487
Nogales Martín, Francisco Javier, 567
Norde, Henk, 421
Novella, Ricardo, 151
Novo Sanjurjo, Vicente, 433
Núñez Antón, Vicente, 197
Núñez Letamendia, Laura, 609
Oña Casado, Inmaculada, 181
Ocaña, Jordi, 85
Ojeda Cabrera, Jorge Luis, 83
Ojeda, Jordi, 595
Olaeta Goiriena, Haritz, 369
Olave Rubio, Pilar, 319
Olivares González, Alberto, 507
Oller Sala, Josep M., 295
Ollero Hinojosa, Jorge, 699
Olmo Jiménez, María José, 735
Olvera-Cervantes, A., 271
Omaña Prieto, Antonio, 603
Orbe Lizundia, Jesus, 197
Oreja Rodríguez, Juan Ramón, 455
Ortega Dato, Juan Francisco, 177
Ortega Pastor, Eva María, 719
Ortego, Maribel, 355, 705
Ortiz Rodríguez, Isabel María, 131
Ortiz-Busto, A., 51
Ortuño Sánchez, M. Teresa, 603
Osuna Gómez, Rafaela, 457, 511, 581
Owen, Guillermo, 635
Oya Lechuga, Antonia, 739
Oyáñez Ugidos, Cristina, 215
Pacheco Bonrostro, Joaquin, 609
Pagès, Jérôme, 349
Palarea Albaladejo, Javier, 351
Palegrín Pelegrín, Blas, 431
Palmer, M., 533, 643

- Pardo Llorente, Julio Angel, 281
 Pardo Llorente, Leandro, 17, 303
 Pardo Llorente, María del Carmen, 203
 Parra López, Juan, 517, 589, 617
 Parras Guijosa, Luis, 645
 Parreño, Francisco, 667
 Parreño, José, 495
 Pastor Ciurana, Jesús T., 549
 Pastor Moreno, Rafael, 451, 453, 467, 469, 499, 593, 595
 Patino Alonso, M^a Carmen, 87
 Pawlowsky Glahn, Vera, 39, 703
 Peña, Daniel, 113, 269
 Peña García, Teresa, 399, 583
 Pelegrin Pelegrin, Blas, 423
 Pellerey, Franco, 719
 Peralta Saez, Juan Luis, 199
 Perea Rojas-Marcos, Federico, 395, 537
 Pereira Gude, Jordi, 445, 465
 Pérez Bernabeu, Elena, 65, 341
 Pérez Bonilla, Alejandra, 147
 Pérez Casany, Marta, 193
 Pérez Fernández, Paloma, 301
 Pérez González, Ana, 167
 Pérez González, Carlos, 127, 263, 365
 Pérez González, Paz, 501
 Pérez López, César, 43, 435
 Pérez Martín, Agustín, 25, 367
 Pérez Ocón, Rafael, 485, 671
 Pérez Palomares, Ana, 169
 Pewsey, Arthur, 13
 Pino Mejías, José Luis, 51, 321, 697
 Pino Mejías, Rafael, 237
 Pizarro Romero, Celeste, 599
 Plà Aragonès, Lluís M., 573
 Planells del Pozo, Antonio, 411, 413
 Plastria, Frank, 625
 Plaza Martínez, Gema, 389
 Poch García, Jordi, 153
 Pons Vallès, Montserrat, 553
 Porcu, Emilio, 299
 Portela García-Miguel, Javier, 77, 235
 Pozo Chía, Antonio, 227
 Prado Valle, Cristina, 369
 Prados Carrasco, Ferran, 153
 Prieto Rumeau, Tomás, 727
 Prieto Sánchez, Carlos, 27
 Pruneda González, Rosa Eva, 71
 Puente del Campo, María Albina, 551
 Puente, M^a Albina, 427, 569
 Puerto Albandoz, Justo, 459, 487, 629, 639
 Puig Oriol, Xavier, 123, 137, 193
 Pujol Rodriguez, Rogelio, 419
 Pulido, Manuel A., 415
 Quesada Rubio, José Manuel, 183, 207
 Quirós Carretero, Alicia, 245
 Rabasa Dolado, Alejandro, 417, 419
 Rafael Castro, Claudio, 309
 Ramírez Cobo, Pepa, 31
 Ramos Ábalos, Eva M^a, 207, 725
 Ramos Méndez, Eduardo, 621
 Ramos Romero, Héctor M., 699
 Ramos, Carmen, 223
 Ramos, Rosa, 223
 Raya Miranda, Rocío, 187, 207
 Reig Pérez, Miguel Jorge, 65
 Riba Civil, Alexandre, 193
 Riera Ledesma, Jorge, 627
 Rincón Zapatero, Juan Pablo, 623
 Ríos Mercado, Roger Z., 493
 Rivas López, María Jesús, 27, 125
 Rodero de Lamo, Lourdes, 123, 137
 Rodríguez Álvarez, Margarita, 489
 Rodríguez Avi, José, 715, 735
 Rodríguez Calvo, Ericka Zulema, 527
 Rodríguez Chía, Antonio Manuel, 461, 463, 487, 629

- Rodríguez Díaz, Juan Manuel, 115, 171
Rodríguez Flores, Beatriz, 45
Rodríguez Galiano, María Isabel, 631
Rodríguez González, S., 19
Rodríguez Gríñolo, Rosario, 697
Rodríguez, Gustavo, 35
Rodríguez Pérez, Patricia, 215
Rodríguez Puerta, Julio, 243, 251, 269, 323
Rodríguez Sánchez, Sara Verónica, 573
Rodríguez Torreblanca, Carmelo, 115
Rojas-Medar, Marko A., 511
Roldán López de Hierro, Concepción, 689
Roldán Nofuentes, José Antonio, 157, 159
Román Montoya, Yolanda, 95
Román Román, Patricia, 107, 713
Romera Ayllón, Rosario, 723
Romero Molina, Desirée, 713
Romero Morales, Dolores, 647
Romo Urroz, Juan José, 91, 313, 331
Rosa González, Felipe, 63
Rueda García, M. Mar, 353
Rueda Sabater, Cristina, 277, 337
Rufián Lizana, Antonio, 457
Ruiz Castro, Juan Eloy, 207, 671
Ruiz de La Torre, José Luís, 67
Ruiz Gándara, Africa, 321
Ruiz García, Rubén, 503, 521, 601
Ruiz Garzón, Gabriel, 457
Ruiz Medina, María Dolores, 175, 179, 201
Ruiz Molina, Juan Carlos, 737, 739
Ruiz, José María, 719, 729
Saavedra Santana, Pedro, 97, 173
Saei, Ayoub, 373
Sáez Aguado, Jesús, 441
Sáez Castillo, Antonio José, 715, 735
Salamanca-Miño, B., 233
Salazar González, Juan José, 385, 627, 673
Salicrú Pagés, Miquel, 111, 191
Salvador Figueiras, Manuel, 103, 285
Salvador González, Bonifacio, 277, 473
San José Nieto, Luis Augusto, 443, 575, 665, 675
San Matías Izquierdo, Susana, 325, 333, 343
San Miguel, Pedro Galeano, 75
Sánchez, Ana, 451, 499
Sánchez Barba, Mercedes, 205
Sánchez Barbié, Ángel, 367
Sánchez Bello, Patricio, 69
Sánchez Borrego, Ismael Ramón, 207
Sánchez-Espigares, José Antonio, 85
Sánchez Fernández, Patricio, 69, 211
Sánchez Galdón, Ana Isabel, 161, 409, 565
Sánchez León, Guillermo, 171
Sánchez Naranjo, María Jesús, 61, 243, 249, 251
Sánchez Rodríguez-Morcillo, Ismael, 21, 59, 289
Sánchez Santos, José Manuel, 27
Sánchez Soriano, Joaquín, 415, 417, 419
Sanchis Saez, Javier, 293
Santamaría Arana, Laureano, 367, 377
Santana del Pino, Angelo, 97
Santos Martín, María Teresa, 109, 125
Santos Peñate, Dolores Rosa, 431, 515, 529

- Sanz Juan, José, 565
 Sanz Sáiz, Gerardo, 695, 707
 Saralegui, Jorge, 367
 Sedeño Noda, Antonio Alberto, 405, 579, 597, 659, 669
 Segovia, M^a Carmen, 485
 Seguí Llinares, Vicente Jesús, 65
 Segura González, Carlos, 471
 Segura, José V., 217, 607
 Sellés Cantó, Miguel Ángel, 65
 Sentinella, Carles, 595
 Serrano Pérez, Juan José, 713
 Sicilia Rodríguez, Joaquín, 443, 575, 597, 619, 661, 665, 675, 677
 Simón de Blas, Clara, 393
 Simeone, Bruno, 7
 Sirvent, Inmaculada, 547
 Sivaramakrishnan, Kartik K., 389
 Sivrikaya Serifoglu, Funda, 601
 Smrekar, Marcelo, 255
 Sobottka, Marcelo, 707
 Solanas, Agustí, 225, 381
 Soler Fernández, David, 523
 Solsona, Jordi, 595
 Sordo Díaz, Miguel Angel, 691, 699
 Stania, Larissa, 475
 Suárez Vega, Rafael, 431, 515, 529
 Suárez-Llorens, A., 271
 Susi García, Rosario, 291
 Tamarit Goerlich, José Manuel, 403, 505, 667
 Tapia García, Jesus Alberto, 473
 Tarifa Blanco, José, 207
 Tejada Cazorla, Juan Antonio, 531, 591
 Terán, Pedro, 259, 731
 Terlaky, Tamás, 389
 Thió-Henestrosa, Santiago, 153
 Tiscar Soria, Miguel Ángel, 715
 Toledo Casado, Andrés, 215
 Toledo Melero, Francisco Javier, 517
 Tommasi, Chiara, 135
 Tormos Juan, Pilar, 535
 Torrente Orihuela, Ester Aurora, 331
 Torres Darias, N., 519
 Torres Ruiz, Francisco de Asís, 107, 713
 Tort-Martorell Llabrés, Xavier, 149
 Trandafir, Paula-Camelia, 135
 Trujillo Carmona, José, 163
 Ureña Ureña, Carmen, 379
 Urlings, Thijs, 601
 Vázquez Burguete, José Luis, 349
 Vélez Ibarrola, Ricardo, 727
 Vaamonde Liste, Antonio, 69, 211, 375
 Valderrama Bonnet, Mariano José, 361
 Valero Franco, Concepción, 461, 463
 Vallada Regalado, Eva, 521
 Valls, Vicente, 407
 Valsero Blanco, María Cruz, 615
 Van Velzen, Bas, 421
 Varela, M., 79
 Varga, Zoltán, 105
 Vargas, Luis G., 645
 Vargas Vargas, Manuel, 329
 Vaz, Ismael, 561
 Vera Vera, José Fernando, 93, 165, 189
 Vercher González, Enriqueta, 217, 287, 607
 Ves Cuenca, Esther De, 335
 Vicente Galindo, Purificación, 37, 47, 49, 87, 205
 Vicente Villardón, José Luis, 33, 79, 81, 229
 Vidal Puga, Juan José, 541, 555, 637
 Vieu, Philippe, 145

- Vilar Fernández, Juan Manuel, 167,
257
- Villa Juliá, M^a Fulgencia, 505
- Villaroya, Angel, 305
- Virto García, Miguel Ángel, 631
- Vitoriano Villanueva, Begoña, 603,
663
- Vives, Sergi, 305
- Wiper, Michael P., 31
- Wondmagegnehu, Eshetu T., 357
- Yáñez, Cotos, 339
- Yáñez Gestoso, Javier, 429, 605
- Zakharov, Viktor, 557
- Zapata Babío, Carlos, 327
- Zucarello, Pedro, 335
- Zuluaga Arias, Pilar, 297