

FDA-SEIO 2011

Libro de ACTAS

I Reunión de Trabajo del Grupo Análisis de Datos Funcionales

Sociedad Española de Estadística e I.O.

Universidad de Cantabria, Santander
15 de Junio de 2011





Libro de Actas

FDA-SEIO 2011

I Reunión de Trabajo del Grupo

Análisis de Datos Funcionales de la

Sociedad Española de Estadística e I.O.

Universidad de Cantabria, Santander, 15 de Junio de 2011

Editores

- Ana M. Aguilera del Pino
Dpto. Estadística e I.O.
Facultad de Ciencias
Universidad de Granada
18071-Granada. Spain
e-mail: aaguiler@ugr.es
- Juan A. Cuesta Albertos
Dpto. de Matemáticas
Facultad de Ciencias
Universidad de Cantabria
39071-Santander. Spain

© 2011. Los editores

Todos los derechos reservados. Esta publicación no puede ser reproducida, ni registrada, ni transmitida en ninguna forma ni por ningún medio, sin el permiso de los editores

Prólogo

La I Reunión de Trabajo del Grupo de Análisis de Datos Funcionales (FDA) de la Sociedad Española de Estadística I.O. (SEIO) se celebra el 15 de Junio de 2011 en Santander. La idea surge como iniciativa de la reunión del grupo de trabajo celebrada en A Coruña durante el XXXII Congreso Nacional de Estadística e I.O. Allí se decidió que podría ser interesante llevar a cabo una reunión del grupo el día anterior a la celebración del Second International Workshop on Functional and Operatorial Statistics (IWFOS 2011), facilitando de este modo la participación tanto de los miembros del grupo FDA-SEIO como de otros investigadores en FDA interesados.

El análisis de datos funcionales es actualmente un tema puntero de investigación estadística que está dando lugar a una amplia variedad de publicaciones en revistas de alto impacto desde un punto de vista tanto teórico como aplicado. Siguiendo la tendencia internacional, en España han aumentado en los últimos años tanto el número de investigadores como el de resultados importantes en este tema. Este hecho motivó la iniciativa de un grupo de miembros de la S.E.I.O. e investigadores en FDA de crear un grupo especializado en este tema cuyo objetivo es la realización de actividades que fomenten la comunicación y la colaboración entre sus miembros. La creación del grupo Análisis de Datos Funcionales de la SEIO (FDA-SEIO) fue aprobada por el Consejo Ejecutivo de la sociedad en febrero de 2009.

El objetivo de la I Reunión FDA-SEIO 2011 es la presentación y discusión de problemas abiertos y aplicaciones del análisis de datos funcionales que favorezcan tanto el intercambio de ideas y la colaboración entre los investigadores en FDA de la SEIO y de otros grupos externos interesados en este tema, como el desarrollo de herramientas computacionales diseñadas para el uso práctico de los métodos desarrollados.

A los organizadores les gustaría expresar su gratitud a todos los autores tanto de las conferencias invitadas como del taller con R y de los pósters, por su cooperación y entusiasmo. Muy especialmente, nos gustaría agradecer a los participantes del Instituto de Hidráulica Ambiental de la Universidad de Cantabria por aceptar nuestra propuesta y contribuir con sus presentaciones a la posibilidad de transferencia y aplicación de los resultados sobre FDA del grupo. También nos gustaría dar las gracias a todos los participantes que con su presencia han hecho posible que esta reunión pueda realizarse.

Finalmente, estamos extremadamente agradecidos a las siguientes instituciones que con su colaboración y patrocinio han contribuido al éxito de la reunión FDA-SEIO 2011:

- Sociedad Española de Estadística e Investigación Operativa
(<http://www.seio.es/>)
- Instituto Cántabro de Estadística
(<http://www.icane.es/>)
- Universidad de Cantabria
(<http://www.unican.es/>)
- Universidad de Granada
(<http://www.ugr.es/>)

Granada-Santander, Junio de 2011

Ana M. Aguilera y Juan A. Cuesta

Comité organizador

- M. Carmen Aguilera Morillo (Universidad de Granada)
- Alicia Nieto Reyes (Universidad de Cantabria)
- Gema Quintana Portilla (Universidad de Cantabria)
- Paula Rodríguez Bouzas (Universidad de Granada)

Índice

1	Programa definitivo	1
2	Datos de interés	2
3	Lista de participantes	3
4	Comunicaciones invitadas	5
5	Taller de R con el paquete fda.usc	9
6	Comunicaciones en póster	10

1 Programa definitivo

9.00 h. Recepción de participantes y recogida de documentación

9.30 h. Inauguración

9.45 h. Encuentro bilateral con el Instituto de Hidráulica Ambiental de la Universidad de Cantabria (<http://www.fundacionih.unican.es/presentacion.asp>)

Moderador: Juan Antonio Cuesta

- **Eliminación de sesgos temporales en registros paleoclimáticos**

Melisa Menéndez*, Fernando J. Méndez, Iñigo J. Losada
IHCantabria, Universidad de Cantabria

- **Comparación entre series numéricas e instrumentales de parámetros de oleaje**

Antonio Tomás, Fernando J. Méndez*, Roberto Mínguez
IHCantabria, Universidad de Cantabria

11.15 h. Café- Descanso

12.00 h. Comunicaciones invitadas sobre FDA

Moderadora: Ana María Aguilera

- **Algunas ideas sobre selección de variables en problemas de clasificación con datos funcionales**

José Ramón Berrendero, Antonio Cuevas, José Luis Torrecilla*
(Universidad Autónoma de Madrid)

- **Análisis de datos funcionales basado en distancias: Presentación de la librería dbstats de R**

Eva Boj, Adrià Caballé*, Pedro Delicado, Josep Fortiana
(Universitat Politècnica de Catalunya)

- **Modelos de regresión funcional para predecir concentración de pólen**

Mariano J. Valderrama*, Ana M. Aguilera, Manuel Escabias,
Francisco A. Ocaña-Lara, Paula R. Bouzas, Nuria Ruiz-Fuentes,
Francisco M. Ocaña-Peinado, Juan E. Ruiz-Castro, M. Carmen Aguilera-Morillo (Grupo FQM-307. Universidad de Granada)

14:00 h. Almuerzo (Facultad de Ciencias)

16:00 h. **Taller de R con el paquete fda.usc**

Manuel Febrero-Bande*, Manuel Oviedo
(Universidad de Santiago de Compostela)

18.00 h. Sesión Póster

– **Análisis geoestadístico de datos funcionales**

M José Ginzo Villamayor*, Manuel Febrero-Bande
(Universidad de Santiago de Compostela)

– **Inferencia sobre la intensidad estocástica de partículas emitidas**

Grupo FQM-307. Modelización y Predicción con Datos Funcionales
(Universidad de Granada)

– **Suavización P-spline del análisis en componentes principales funcional**

M. Carmen Aguilera-Morillo*, Ana M. Aguilera
(Universidad de Granada)

18.30 h. Café-Descanso

19.00 h. Reunión del Grupo FDA-SEIO: conclusiones y direcciones futuras

21.30 h. Cena de clausura (Restaurante Deluz)

2 Datos de interés

- El lugar de celebración de las conferencias y reuniones es el Salón de Actos de la Facultad de Ciencias de la Universidad de Cantabria en Santander.
- Las comunicaciones orales tendrán una duración de 30 minutos de presentación con posterior discusión de hasta 10 minutos.
- Los pósters se colocarán durante el café-descanso de la mañana en el espacio habilitado para ellos y se retirarán durante el café-descanso de la tarde.

3 Lista de participantes

- Aguilera del Pino, Ana María. Universidad de Granada
(aaguiler@ugr.es)
- Aguilera-Morillo, María del Carmen. Universidad de Granada
(caguilera@ugr.es)
- Caballé Mestres, Adriá. Universidad Politécnica de Catalunya
(adriacaballe@yahoo.es)
- Carreño Madinabeitia, Sheila. Empresa TecnaliaR&I
(sheila101@hotmail.es)
- Crujeiras Casais, Rosa María. Universidad de Santiago de Compostela
(rosa.crujeiras@usc.es)
- Cuesta Albertos, Juan Antonio. Universidad de Cantabria
(juan.cuesta@unican.es)
- Cuevas, Antonio. Universidad Autónoma de Madrid
(antonio.cuevas@uam.es)
- Delicado Useros, Pedro. Universidad Politécnica de Catalunya
(pedro.delicado@upc.edu)
- Escabias Machuca, Manuel. Universidad de Granada
(escabias@ugr.es)
- Estébanez Gallo, David. Universidad de Cantabria
(david.eg86@yahoo.es)
- Febrero Bande, Manuel. Universidad de Santiago de Compostela
(manuel.febrero@usc.es)
- Ginzo Villamayor, María José. Universidad de Santiago de Compostela
(mariajose.ginzo@usc.es)
- González Gutiérrez, Pablo. Universidad de Cantabria
(pablogonzalezgutierrez@gmail.com)
- González Manteiga, Wenceslao. Universidad de Santiago de Compostela
(wenceslao.gonzalez@usc.es)
- Inacio, Vanda. Universidad de Lisboa
(vanda.kinets@gmail.com)

- Irigoien Garbizu, Itziar. Universidad del País Vasco
(itziar.irigoien@ehu.es)
- Martínez Calvo, Adela. Universidad de Santiago de Compostela
(amartinez@cesga.es)
- Méndez Incera, Fernando J. IHCantabria, Universidad de Cantabria
(mendezf@unican.es)
- Menéndez García, Melisa. IHCantabria, Universidad de Cantabria
(menendezm@unican.es)
- Mínguez, Roberto. IHCantabria, Universidad de Cantabria
(minguezsr@unican.es)
- Nieto Reyes, Alicia. Universidad de Cantabria
(alicia.nieto@unican.es)
- Oviedo de la Fuente, Manuel. Universidad de Santiago de Compostela
(manuel.oviedo@usc.es)
- Pateiro López, Beatriz. Universidad de Santiago de Compostela
(beatriz.pateiro@usc.es)
- Quintana Portilla, Gema. Universidad de Cantabria
(quintanagr@unican.es)
- Rodríguez Bouzas, Paula. Universidad de Granada
(paula@ugr.es)
- Rodríguez Poo, Juan Manuel. Instituto Cántabro de Estadística
(rodriguez_jm@gobcantabria.es)
- Ruiz Fuentes, Nuria. Universidad de Granada
(nfuentes@ujaen.es)
- Saumard, Matthieu. Universidad de Santiago de Compostela
(matthieu.saumard@gmail.com)
- Tomás Sampederro, Antonio. IHCantabria, Universidad de Cantabria
(antonio.tomas@unican.es)
- Torrecilla Nogueras, José Luis. Universidad Autónoma de Madrid
(joseluis.torrecilla@uam.es)
- Valderrama Bonnet, Mariano. Universidad de Granada
(valderra@ugr.es)

4 Comunicaciones invitadas

- **Título:** Eliminación de sesgos temporales en registros paleoclimáticos

Autores: Melisa Menéndez*, Fernando J. Méndez, Iñigo J. Losada

*IHCantabria, Universidad de Cantabria

Palabras clave: Series temporales, escalado, desfases temporales

Resumen: El estudio del sistema climático terrestre, el papel de los procesos físicos involucrados y el entendimiento de las inestabilidades climáticas requiere del conocimiento de las dinámicas actuantes en la Tierra desde hace muchos años, es decir, requiere de un estudio paleoclimático. Para ello, el enfoque temporal abarca una escala de tiempo geológica que se suele enmarcar en la última fase del Cuaternario (desde hace 1.7 millones de años hasta el presente). Uno de los aspectos más relevantes en la vida animal durante el Cuaternario es la evolución del género homo. Sin embargo, para reconstruir el clima existente antes de que los seres humanos fuéramos capaces de usar la tecnología es necesario el uso de diversas técnicas y aproximaciones científicas. Dado que no existen dataciones hechas por el hombre sobre variables directas del clima en el periodo pasado, para obtener registros paleoclimáticos se debe recurrir a indicadores indirectos, denominados proxies. Los proxies son registros paleoclimáticos que permiten inferir variaciones de variables climáticas directas (temperatura, precipitaciones, humedad, etc). Existen diversas fuentes de proxies, como las que aporta la dendrocronología (estudio del anillo de los árboles), los arrecifes de coral, el crecimiento de estalagmitas en cuevas (espeleotermas) o los que se obtienen a través de testigos de hielo y testigos sedimentológicos, lacustres o marinos. La inferencia de la información obtenida a partir de esta variedad de fuentes a variables climáticas directas conlleva una colección de técnicas físico-químicas variada. La técnica más común se basa en el análisis isotópico. Uno de las mayores aportaciones de proxies proviene de los registros paleoceanográficos a partir del estudio de organismos microscópicos marinos que resultan indicadores de variaciones climáticas. La reconstrucción mediante estas técnicas de registros paleoclimáticos permite obtener información muy valiosa desconocida hasta el momento. Sin embargo, se deben tener en cuenta sus inconvenientes, como que están geográficamente limitados, pueden tener sesgos en su variabilidad, inhomogeneidades en la resolución temporal, o la incertidumbre en la escala de baja frecuencia. A partir del estudio de los proxies se tiene constancia de una sucesión de situaciones climáticas que siguen un mismo patrón. Esta situación climática, a la que se suele referir como ciclo climático, está compuesto por un periodo Interglacial seguido de un periodo Glacial. Dentro de estos periodos, que duran varias decenas de miles de años, se tiene constancia de episodios milenarios de agudización de frío/calor denominados estúdiales/interstadiales respectivamente.

- **Título:** Comparación entre series numéricas e instrumentales de parámetros de oleaje

Autores: Antonio Tomás, Fernando J. Méndez*, Roberto Mínguez

*IHCantabria, Universidad de Cantabria

Palabras clave: Análisis de series temporales, bases de datos de oleaje, ajuste temporal, calibración de la fase

Resumen: El movimiento de la superficie del mar que denominamos oleaje es un fenómeno complejo, y no ha sido hasta la segunda mitad del siglo XX cuando los primeros modelos analíticos han empezado a describirlo. El oleaje se analiza como un proceso estocástico; para la simplificación de dicho análisis se suele requerir que el proceso cumpla una serie de propiedades, entre ellas que sea estacionario o al menos, débilmente estacionario, lo que implica que los estadísticos del proceso (por ejemplo, media y varianza) se mantengan invariables ante el paso del tiempo. Sin embargo, el oleaje es la respuesta de la superficie del mar a condiciones atmosféricas no estacionarias, luego para poder asumir la estacionariedad es necesario trabajar con "instantáneas" del proceso de una duración tal que, durante ese tiempo la variación de los estadísticos del proceso sea imperceptible. Estos intervalos de tiempo en que el proceso se puede considerar débilmente estacionario se denominan estados de mar. La duración de los estados de mar es un compromiso entre un tiempo lo suficientemente corto como para que el proceso sea débilmente estacionario y lo suficientemente largo como para que contenga la necesaria información estadística del proceso. En la actualidad, se utilizan estados de mar de una hora de duración, aunque el límite inferior por información estadística suele ser de unos 30 minutos. Los parámetros de estado de mar (por ejemplo, altura de ola del momento de orden cero espectral, H_{m0} , periodo de pico espectral, T_p , altura de ola significativa, H_s , periodo medio, T_z , etc.), a su vez, varían en cada estado de mar, constituyendo, a lo largo del tiempo, una serie temporal de estados de mar. Las características estadísticas de esta serie describen el clima marítimo a largo plazo en una determinada zona. Existen dos tipos de bases de datos de parámetros de estados de mar, las instrumentales y las numéricas. Las bases de datos instrumentales provienen de redes de observación de oleaje (generalmente constituidas por boyas) tienen los datos cuantitativamente mejores aunque están dispersas en el espacio y no suelen suministrar información de más de 15-20 años. Además no son continuas, presentan huecos (por avería o mantenimiento). Las bases de datos numéricas (o bases de datos de reanálisis) se generan a partir de modelos numéricos que son capaces de simular los procesos que dan lugar a la generación del oleaje a partir de forzamientos atmosféricos. De esta manera, se dispone de series históricas de los parámetros del oleaje de mucha más larga duración que los existentes en las bases instrumentales (20-60 años), homogéneas y de una mayor resolución espacial. Por otro lado, aunque estas bases de datos representan una aproximación a la realidad, al estar generadas numéricamente, tienen discrepancias respecto de los datos instrumentales. Estas diferencias suelen estar motivadas por las limitaciones de los modelos numéricos utilizados, aunque otras veces tan solo son debidas a las diferentes técnicas utilizadas para generar o registrar los datos comparados.

- **Título:** Algunas ideas sobre selección de variables en problemas de clasificación con datos funcionales

Autores: José Ramón Berrendero, Antonio Cuevas, José Luis Torrecilla*

*Departamento de Matemáticas. Universidad Autónoma de Madrid

Palabras clave: Selección de variables, clasificación, correlación de distancias.

Resumen: Las técnicas de selección de variables son cada vez más utilizadas en el ámbito de la estadística con datos multivariados de gran dimensión. Dada la gran aplicabilidad práctica de estas técnicas, resulta natural estudiar su adaptación al contexto de los datos funcionales. En este trabajo en curso se está analizando el desarrollo de nuevos métodos de selección de variables para datos funcionales con vistas a su utilización en problemas de clasificación supervisada. Nuestra propuesta toma como punto de partida el método de "Máxima Relevancia y mínima Redundancia" (MRmR), combinado con una nueva medida de dependencia estadística basada en las funciones características. Estamos analizando las implicaciones teóricas y los resultados de la aplicación de este método a conjuntos de datos funcionales, tanto reales como simulados.

- **Título:** Análisis de datos funcionales basado en distancias: Presentación de la librería dbstats de R

Autores: Eva Boj, Adrià Caballé*, Pedro Delicado, Josep Fortiana

*Departamento de Estadística e I.O. Universidad Politécnica de Catalunya

Palabras clave: Distancias y semi-métricas entre funciones, mínimos cuadrados iterados y reponderados, modelo lineal generalizado, partial least squares, productos escalares entre funciones, regresión basada en distancias, regresión no paramétrica, regresión ponderada, verosimilitud local

Resumen: En 1989 Carles Cuadras introdujo la regresión basada en distancias. La contribución principal de esta metodología fue establecer que es posible ajustar un modelo de regresión lineal múltiple conociendo únicamente las distancias entre los diferentes individuos de la muestra, calculadas a partir de los valores que toman los regresores en cada uno de ellos. Posteriormente se fueron desarrollando versiones de otras técnicas estadísticas que comparten esa característica: su input es únicamente una matriz de distancias entre individuos. La aplicación de esta metodología al análisis de datos funcionales ha dado lugar a diversas publicaciones que muestran que el enfoque basado en el cálculo de distancias (o semi-métricas) entre funciones seguido de la aplicación de un método estadístico basado en distancias, es una alternativa competitiva. Recientemente hemos desarrollado una librería de R, llamada dbstats, que implementa la versión basada en distancias de la regresión lineal y del modelo lineal generalizado y de sus versiones no paramétricas (construidas mediante ajustes locales de los modelos anteriores), así como de la regresión por mínimos cuadrados parciales (PLS). Como muestra de la potencia de los métodos basados en distancias en el análisis de datos funcionales presentaremos varios ejemplos desarrollados con esta librería en combinación con librerías específicas de datos funcionales.

- **Título:** Functional regression models to forecast pollen concentration

Autores: Mariano J. Valderrama*, Ana M. Aguilera, Manuel Escabias, Francisco A. Ocaña-Lara, Paula R. Bouzas, Nuria Ruiz-Fuentes, Francisco M. Ocaña-Peinado, Juan E. Ruiz-Castro, M. Carmen Aguilera-Morillo

*Departamento de Estadística e I.O. Universidad de Granada

Resumen: Several approaches have been developed in order to forecast pollen concentration from environmental data. In the first one, given time interval a functional regression model is derived by means of a two-step procedure, considering the air temperature in a previous interval as the input. This estimation is carried out by functional principal component (FPC) analysis and the residual noise is also modelled by FPC regression, taking as the explicative process the pollen concentration during the earlier interval. The prediction performance is then tested on cypress pollen data series recorded in Granada (Spain) over a period of 10 years. The second approach considers a binary response regression model selecting environmental explicative variables by stepwise regression and testing the performance with a set of olive pollen data.

References

Aguilera AM, Escabias M and Valderrama MJ (2007): Forecasting binary longitudinal data by a functional PC-ARIMA model. *Computational Statistics and Data Analysis*, 52 (6), 3187-3197

Ocaña-Peinado FM, Valderrama MJ and Aguilera AM (2008): A transfer function-principal components model to forecast air pollen concentration. *Stochastic Environmental Research and Risk Assessment* 22 (1), 59-63

Valderrama MJ, Ocaña-Lara FA, Aguilera AM and Ocaña-Peinado FM (2010). Forecasting pollen concentration by a two-step functional model. *Biometrics* 66 (2), 497-504

5 Taller de R con el paquete *fda.usc*

Autores: Manuel Febrero-Bande*, Manuel Oviedo de la Fuente

*Departamento de Estadística e I.O. Universidad de Santiago de Compostela

Este taller práctico tiene como objetivo introducir a los participantes en el manejo del paquete *fda.usc*, creado para R (versión 2.10 o posterior) por Manuel Febrero Bande y Manuel Oviedo de la Fuente. El paquete *fda.usc* implementa algunas herramientas útiles para el análisis de datos funcionales. Dicho paquete (no restringido a L_2) integra y complementa dos de las principales referencias del análisis de datos funcionales, como son el paquete *fda* del software R y las funciones implementadas por *Ferraty y Vieu (2006)*. Permite llevar a cabo un análisis exploratorio y descriptivo de datos funcionales, explorando sus características más importantes: mediciones de profundidad o detección de outliers funcionales, entre otros. También ayuda a explicar y modelar la relación entre una variable dependiente e independiente (modelos de regresión) y hacer predicciones. Además, incluye métodos para la clasificación supervisada o no supervisada de un conjunto de datos funcionales, en relación a una característica de los datos, así como un ANOVA para datos funcionales. En esta línea, los temas desarrollados en el taller son los siguientes:

1. Análisis de datos funcionales

- Definición de datos funcionales
- Exploración de datos funcionales
- Réplicas bootstrap como medidas de dispersión
- Detección de outliers funcionales

2. Modelos de regresión funcional (MRF)

- MRF con covariable funcional
- MRF con representación básica
- Modelo de regresión funcional no paramétrico
- MRF con covariable funcional o no funcional

3. Clasificación funcional

- Clasificación supervisada
- Clasificación no supervisada

4. ANOVA funcional

6 Comunicaciones en póster

- **Título:** Análisis geostadístico de datos funcionales

Autores: M José Ginzo Villamayor*, Manuel Febrero-Bande

*Departamento de Estadística e I.O. Universidad de Santiago de Compostela

Palabras clave: Base de funciones; cokriging; validación cruzada; modelo lineal funcional; Kriging; predicción espacial multivariante

Resumen: Este trabajo trata la predicción de curvas, cuando se dispone de una muestra de las curvas de una región con continuidad espacial. La motivación de este trabajo es ofrecer una solución al problema de predecir curvas en aquellas zonas no muestreadas de una región, basándonos en estas las ramas estadísticas, el análisis de datos funcionales y la estadística espacial. Se revisan tres métodos para la predicción espacial de los datos funcionales. Inicialmente, se propone un predictor que tiene la misma forma que el predictor kriging clásico, pero teniendo en cuenta las curvas en lugar de datos de una sola dimensión. Los otros predictores surgen de adaptaciones de modelos lineales funcionales con respuesta funcional en el caso de datos funcionales espacialmente correlacionados. Por un lado, se define un predictor que es una combinación de kriging y del modelo funcional lineal point-wise (concurrente). Por otra parte, se utiliza el modelo funcional lineal total para extender dos métodos clásicos geostatísticos multivariantes para el contexto funcional. El primer predictor se define en términos de parámetros escalares. En el resto de los casos, los predictores implican parámetros funcionales. Se adapta un criterio de optimización, criterio utilizado en predicción espacial multivariante para estimar los parámetros escalares y funcionales que intervienen en los predictores propuestos. En todos los casos se da un enfoque no paramétrico basado en la expansión en términos de bases de funciones que se usa para obtener las curvas a partir de datos discretos. De la misma manera que los métodos estadísticos estándares han sido generalizados para ser utilizados en FDA, es posible pensar que los métodos geostatísticos pueden ser adaptados a este tipo de datos. Las metodologías propuestas se ilustran mediante el análisis de un conjunto de datos real correspondiente a la curva de temperatura que es función del tiempo. En resumen, lo que contiene este proyecto fin de máster es una revisión crítica de los métodos que se han considerado previamente, en estadística espacial con datos funcionales y aplicados a un conjunto de datos real como es el de la temperatura en Galicia. Este conjunto de datos tiene tanto componente espacial y funcional.

- **Título:** Inferencia sobre la intensidad estocástica de partículas emitidas
Autores: Grupo FQM-307. Modelización y Predicción con Datos Funcionales*

*Departamento de Estadística e I.O. Universidad de Granada

Palabras clave: Proceso de Cox, análisis funcional en componentes principales, inferencia simultánea

Resumen: Este trabajo se inspira en el problema que surge en el laboratorio de comprobar el adecuado funcionamiento de un espectrómetro u otro tipo de detector de partículas, es decir, en la calibración de aparatos de detección y recuento de partículas. Para ello, se modeliza el recuento de partículas y se propone una herramienta para comprobar si el dispositivo de medición está bajo control o no. El proceso puntual de emisiones de partículas radiactivas se modeliza como un proceso de Cox es decir, permitiendo la aleatoriedad de su intensidad. La intensidad estocástica del proceso se estima mediante análisis funcional en componentes principales utilizando un método ad hoc para este proceso (Ramsay and Silverman, 1997; Bouzas et al., 2010a; Bouzas et al, 2010b). Se observa que la estimación de la intensidad tiene una estructura y distribución que permite construir un intervalo de confianza para su media en cada instante del intervalo total de observación (Bouzas et al, 2010b). Como consecuencia, se construye también el correspondiente contraste de hipótesis para testar si la media de la intensidad de nuevas trayectorias observadas es igual a la estimada o no en cada instante de tiempo. Se debe estimar de forma precisa la intensidad de un recuento de partículas cuando el dispositivo se sabe en perfectas condiciones para que haga el papel de patrón y a partir de entonces, para nuevas mediciones, comparar los resultados con los considerados como patrón. Ya que se ha construido un test para contrastar la media de la intensidad de nuevas trayectorias en cada instante del intervalo de observación, que las observaciones se hacen en un conjunto discreto de instantes de tiempo y teniendo en cuenta que no debemos considerar cada instante por separado, proponemos el siguiente contraste múltiple para comparar nuevas trayectorias observadas con el patrón: la media estimada de las nuevas trayectorias observadas es igual a la patrón para todos los instantes de observación en contra de que no sea así. Por tanto, tenemos un test de bondad de ajuste que consiste en un contraste múltiple que consta de tantos individuales como instantes de observación y por sus características, se ha escogido el criterio stepdown Benjamini-Hochberg como criterio de decisión (Benjamini and Hochberg, 1995; Bouzas et al, 2010b). Finalmente, se decide aceptar la hipótesis nula si se aceptan todos los individuales con sus p-valores adaptados, lo que significaría que las mediciones son acordes con el modelo patrón y por tanto el detector funciona correctamente. En caso contrario, se interpretaría que el detector no funciona correctamente, no está bajo control. No siempre es posible observar el fenómeno durante mucho tiempo por multitud de razones, por lo que este trabajo sugiere la utilización de un estimador puntual para las nuevas trayectorias más simple que el utilizado para la estimación de la intensidad patrón (Bouzas et al, 2006). Los isótopos radiactivos ^{226}Ra y ^{137}Cs son isótopos de larga vida y por tanto sus emisiones se pueden suponer no dependientes del pasado del proceso por eso se considera su modelización

mediante un proceso de Cox. Aplicando lo explicado anteriormente, se estima la intensidad estocástica de las emisiones para ambos isótopos. Así mismo, se comprueba la eficacia del test de bondad de ajuste comparando nuevas trayectorias de emisiones con y sin ligeras distorsiones. Los resultados obtenidos prueban que la detección de trayectorias con distorsiones se consigue en un altísimo porcentaje por tanto se considera que el test es muy sensible detectando incluso observaciones ligeramente erráticas.

Referencia

P.R. Bouzas, N. Ruiz-Fuentes, A. Matilla, A. M. Aguilera, M. J. Valderama, A Cox model for radioactive counting measure: inference on the intensity process, Chemometrics and Intelligent Laboratory Systems 103 (2010) 116-121

- **Título:** Análisis en Componentes Principales Funcional Suavizado vía P-splines

Autores: M. Carmen Aguilera-Morillo*, Ana M. Aguilera

*Departamento de Estadística e I.O. Universidad de Granada

Palabras clave: Análisis en componentes principales funcional, P-splines, expansión básica

Resumen: En general, las observaciones muestrales de una variable funcional son funciones que proceden de la observación temporal de una variable estadística. En la práctica, las funciones muestrales se observan en un conjunto finito de puntos. Con objeto de reducir la dimensión y explicar la estructura de dependencia de un conjunto de datos funcionales en términos de variables incorreladas, es usual utilizar el Análisis en Componentes Principales Funcional (ACPF). Cuando las curvas muestrales son observadas con ruido, las autofunciones o curvas de componentes principales presentan mucha variabilidad y son difíciles de interpretar. En este trabajo proponemos dos versiones diferentes sobre el ACPF suavizado, ambas basadas en el suavizado con splines penalizados (P-splines), usando bases de B-splines. En primer lugar, penalizamos la aproximación de las curvas muestrales y posteriormente, se lleva a cabo un ACPF no suavizado. Por otro lado, introducimos la penalización de la rugosidad en la construcción de las componentes principales. Con objeto de ilustrar el buen comportamiento de estos métodos, se ha llevado a cabo un estudio de simulación en el que se compara el ACPF no suavizado con las nuevas versiones propuestas sobre ACPF suavizado via P-splines.

Notas

Grupo Análisis de Datos Funcionales de la SEIO
<http://www.seio.es/grupos/FDA/>



Organizadores
Ana María Aguilera del Pino (Univ. Granada)
Juan Antonio Cuesta Albertos (Univ. Cantabria)