Posters ScoVa16

Sequential Monte Carlo methods for estimation and prediction in longitudinal models

Danilo Alvares, Carmen Armero, and Anabel Forte

There are many studies where it becomes mandatory to update posterior distributions based on new data. In such situations using the posterior distribution as the prior and performing the usual Bayesian analysis is sensible. However, the analytic form of the posterior distribution is not always available and we only have an approximated sample of it, thus making the process "not-so-easy".

In this context, we present a dynamic approach using sequential Monte Carlo methods to obtain a posterior sample based on a prior sample and the new data. In particular, we focus on updating posterior distributions in the framework of longitudinal data analysis.

Results come from a simulation study. They show the savings in computational time and the equivalence of this approach with the inferential process based on the set that integrates the old and new data.

Assessment for plant disease through Bayesian models

Danilo Alvares, Carmen Armero, Anabel Forte, and Luis Rubio

Tiger nut tubers are the main ingredient in the production of *orxata* in València, a white soft sweet popular drink. In recent years, the appearance of black spots in the skin of tiger nuts has lead to important economical losses because diseased tubers are discarded. The disease is caused by a pathogenic organism which causal agent has not yet been identified.

In this work we discuss some statistical models to assess the transmission of the disease from seeds to tubers, and propose a measure of effectiveness for different treatments against the disease based on the probability of germination and the incidence of the disease. Statistical models for these studies include Bayesian Dirichlet-Multinomial inferential processes and logistic regression models.

Distribution species: a multivariate method using a conditional coregionalized linear model

Xavier Barber, David Conesa, Antonio López-Quílez, Iosu Paradinas, José M. Bellido, and Maria G. Pennino

The aim of this work is to present a Bayesian bivariate spatial modelling for the occurrence patterns of two cohabitants species. There are many situations of interest in which this methodology could be applied such as fisheries management, climate change or landscape management. In particular, in line with Banerjee et al (2014), we propose a conditional coregionalized Bayesian linear model that makes use of the reparametrization of the variance (Yan et al., 2007). We also stablish new default priors for the variances.
This modelling provides a better approach to the prediction problem of both species comparing to the separable approach of both species. Clearly it also provides a better approach that when analysing each species by itself. Two particular examples are presented to check the behaviour of this approach.

Monte Carlo error in the Bayesian estimation of risk ratios using log-binomial regression models

Juan Antonio Cano and Diego Salmerón

In cohort studies, binary outcomes are very often analyzed by logistic regression. However, it is well known that when the goal is to estimate a risk ratio, the logistic regression is inappropriate if the outcome is common. In these cases, a log-binomial regression model is preferable. On the other hand, the estimation of the regression coefficients of the log-binomial model is difficult owing to the constraints that must be imposed on these coefficients.

Bayesian methods allow a straightforward approach for log-binomial regression models and produce smaller mean squared errors in the estimation of risk ratios than the frequentist methods, and the posterior inferences can be obtained using the software WinBUGS. However, Markov chain Monte Carlo methods implemented in WinBUGS can lead to large Monte Carlo errors in the approximations to the posterior inferences because they produce correlated simulations, and the accuracy of the approximations are inversely related to this correlation.

To reduce correlation and to improve accuracy, we propose a reparameterization based on a Poisson model and a sampling algorithm coded in R.

Statistical modeling of proteome expression data in Manila clams, Ruditapes philipinarum, exposed to citrate capped gold nanoparticles (AuNP), as a model contaminant of environmental nanoparticle contamination

Raquel Gavidia Josa, Moritz Volland, Julian Blasco, Amparo Torreblanca, and Miriam Hampel

Among others, proteomic research aims to identify and quantify relative changes in protein abundance between individuals under different circumstances. In ecotoxicology, proteomics is applied to test organisms exposed to environmental contaminants, and is especially useful for the evaluation of chronic low-level exposure effects that manifest at molecular levels of organization. Obtained information can be used to better understand metabolic pathways and networks the contaminant interferes with, providing a better knowledge of the mode of action of the contaminant, as well as for contaminant specific biomarker development for environmental monitoring purposes. In a laboratory-based experiment, we exposed the non-target marine bivalve, the Manila clam *Ruditapes philippinarum* to an environmentally relevant concentration (0.75 μg L^{-1}) of weakly agglomerating citrate AuNPs ($\sim$20 nm). An 8-plex iTRAQ-based bottom-up proteomic approach followed by tandem mass spectrometry led to the identification and quantitation of 2200 expressed proteins. In order to identify statistically significant features, the obtained data was analysed using multiple contrast analysis corrected by FDR aimed at the control of the proportion of false positives among rejected hypotheses. However, the experimental design applied resulted in a higher number of variables than samples or observations leading to an over-conservative approach. Alternatively, the restricted model regressions Lasso and Elastic-Net were applied to provide better information on what proteins differentiate between control and treated clams through penalization of the coefficients of the variables that are not associated with the response variable zero. Elastic-Net revealed 105 identified proteins of which many were the same as those obtained using multiple contrasts. Ultimately, samples were classified after dimension reduction by PLS-DA. This provided a good classification of the samples, in which all variables are projected along the components creating the model. Several obtained proteins were related to oxidative stress, inflammatory response and cytoskeleton and correlated with several features identified in q-PCR gene expression analysis carried out on the same samples previously.

Sensitivity analysis in a Bayesian correlated model for assessing the prevalence of viruses in organic and non-organic agroecosystems

Elena Lázaro, Carmen Armero, and Luis Rubio

A Bayesian correlated binary model were formulated to assess the prevalence of three viruses in organic and non organic agriculture. Markov chain Monte Carlo (MCMC) methods have been used to approximate the posterior distribution of the uncertainties in the model. In random effects models the inappropriate choice of the prior distribution on the random effects scale parameter can concern posterior estimation. Several models were fitted regarding nine hyper-prior assumptions and inference discrepancies were examined in terms of changes in marginal posterior distributions both of the fixed effects and of the random effects. Hellinger distance and a derived measure of sensitivity were applied to assess differences. A calibration of results with respect to the standard normal distribution was done to facilite comparisons among the different models.

A Bayesian discrete-time model to predict Bronchiolitis incidence in the Valencia region.

Mónica López-Lacort, Francisco J. Santonja, and Ana Corberán-Vallet

Bronchiolitis is a disorder caused by viral lower respiratory tract infection. It is a common infection among young children that causes an elevated number of hospital admissions and a high demand for both primary care and emergency services. Consequently, the development of a predictive model that enables timely public health intervention is fundamental to reduce morbidity and mortality.

We present a Bayesian model in discrete time to study bronchiolitis dynamics in a population of children aged less than 2 years-old in Valencia that has been divided into four age groups. New infections are described taking into account that previous cases of bronchiolitis in all age groups act directly on the probability of infection. The seasonality that explains the cyclical pattern of infected cases is treated through indicator variables that account for monthly effects. The Bayesian analysis of the model allows us to calculate both the posterior distribution of the model parameters and the posterior predictive distribution, which facilitates the computation of point forecasts and prediction intervals for future observations.

Bayesian models in avian ecology

Marcial Marín, David Conesa, and Juan Monrós

Aspects such as observations made by multiple observers, missing data or the probability of detection of an individual can be sources of additional uncertainty in statistical modelling. Bayesian statistics allows for proper management of that uncertainty. Some data sets related with avian ecology are available to work with in a forthcoming Ph.D. thesis. These data sets have associated some of the uncertainty sources mentioned above, so we think that a Bayesian point of view would be useful to deal with it. We will propose a Ph.D. thesis based on examples of Bayesian approaches to comparison of biometric data collected by several observers, treatment of missing data, migration and survival studies using capture-recapture data and spatial distribution of bird species.

Bayesian hierarchical dynamic modeling of dengue cases

Daniel Martínez Bello and Antonio López-Quilez

Dengue is a viral disease, causing high morbidity in tropical countries around the world. Case report data is available from health surveillance institutions, and it is the primary source of epidemiological information for disease control. Dengue disease is vector transmitted and it is highly influenced by environmental factors. Rain, relative humidity, solar radiation, environmental temperature are factors influencing dengue case report, at local and regional levels. Establishing the association between these factors and dengue case report is important to develop early warning systems, helping the health institutions to decrease the disease burden. The objective of the present study is to establish the relation of dengue case report and meteorological variables through a hierarchical Bayesian dynamic modeling analysis. A Bayesian dynamic log Poisson model for the weekly count of dengue cases was fitted to data from a city in Colombia, South America, for the period 2008 to 2015, with two main structures in the predictors, accounting for the dynamic characteristics of the data. First main structure included autorregressive response of lag one, and time varying correlated errors of order one. The second main structure contained order one autorregressive time varying coefficients. For every main structure, non seasonal and seasonal models were considered. Seasonal models included trigonometric terms accounting for annual pattern in dengue report. Meteorological variables entered into every main structure as constant effects with all the possible interactions or as time varying effects, with order one autoregressive coefficients. Normal priors were assigned to the predictor coefficients. MCMC simulation was employed for parameter estimation and model selection proceeded using deviance information criteria and related statistics such as mean deviance, estimated deviance and number of effective parameters. Models were displayed using biplots. The model with the best selection criteria included time varying coefficients for relative humidity, solar radiation and temperature, time varying intercepts, and annual seasonal pattern. The main challenge of the hierarchical Bayesian modeling of dengue case report and meteorological variables, involves the interpretation of the coefficients, given the dynamic nature of the data.

Spatio-temporal modeling of dengue cases in Bucaramanga, Colombia, for the period 2008 to 2013

Daniel Martínez Bello, Antonio López-Quilez, and Alexander Torres Prieto

Disease mapping is an important tool to understand the disease distribution in space and time, and to evaluate the public health measures impact to counteract the disease burden. Hierarchical Bayesian disease mapping is at the core of the geo-spatial representation of the disease. The objective of the study was to apply spatio-temporal hierarchical Bayesian disease mapping models to aggregated dengue cases from the city of Bucaramanga, Colombia, for the period 2008 to 2013, searching for a model which could be used as a representation of the disease spread and dynamics in Bucaramanga, supplying information in the search of a predictive model. Dengue cases for the period 2008 to 2013 were geocoded, and allocated to one of 295 city sections, extracted from the Census 2005 according to the national statistical office (DANE) by epidemiological periods of 4 epidemiological weeks. Spatio-temporal Bayesian hierarchical models type I, II, III and IV were fitted to model the relative risk of dengue in every census section. The full model contained spatial and temporal heterogeneity and clustering effects and interaction random effects. Parameter estimation was done using Monte Carlo Markov Chain simulations. Model convergence was established using Gelman and Brooks test and Geweke test and line plots, and model selection was accomplished using the Deviance Information Criteria (DIC). The results showed from several several fitted models, that the interaction type III models exhibited the smallest DIC, and from these, the model containing spatial heterogeneity and clustering effect in every epidemiological period was the selected model to produce relative risk maps for Dengue cases in Bucaramanga.

The Reverse Sigmoidal function, an alternative to Geostatistics?

Joaquín Martínez-Minaya, Antonio Vicent, Antonio López-Quílez, and David Conesa

Modelling spatial patterns of plant disease incidence using local environmental variables has been a growing area of research in last years. Geostatistical models are probably the most popular approach to estimate and predict the disease risk and relate it to the underlaying geographic and environmental factors. However, Geoestatistics may not be appropriate in situations which there exists a separability problem, because it can not work with the likelihood function. An alternative to Geostatistics is presented here based on the fact that the presence/absence of a disease can be expressed with a hierarchical Bayesian model that incorporates environmental variables and a geographic factor through a nonlinear function: the Reverse Sigmoidal. Bayesian inference on the parameters and prediction of disease presence/absence in new locations are performed using INLA, which is particularly adequate in the case of latent Gaussian models. The two spatial methods, Geostatistics and Reverse Sigmodial function, have been compared using two distinct case-studies: the presence/absence of the *Fasciola hepatica*, a parasite of cows in Galicia, in north-west Spain, and the presence/absence of citrus black spot, a disease caused by the fungus *Phyllosticta citricarpa*, in KwaZulu-Natal province in South Africa.

Modeling plant pathogen inoculum production using Bayesian growth curves

Joaquín Martínez-Minaya, Antonio Vicent, Antonio López-Quílez, and David Conesa

Growth curves are included in a class of nonlinear models widely used in biology to study different processes, such as inoculum and disease progress over time. However, in addition to the dynamic process, other factors related with the response variable may be required to improve model performance and applicability. Here, we use hierarchical Bayesian methods to analyze growth models including time and environmental factors as a nonlinear function. These models have been fitted using Markov chain Monte Carlo methods by means of the WinBUGS software. The methodology has been applied in the plant pathology context by the first time, in particular, to model the maturation process of the ascospore inoculum of the pathogenic fungus *Mycosphaerella nawae*, the causal agent of circular leaf spot of persimmon tree. Datasets have been assembled from different locations in Comunitat Valenciana, central-eastern Spain for the period 2010 to 2015. The main aim of the study has been to predict periods of potential inoculum availability based on environmental variables.

Model selection from a Frequentist and Bayesian approach to solve an eating disorders problem

Elena Moreno, David Conesa, David Hervás, Rosa Baños, and Ausiàs Cebolla

Although eating styles assessed by the Dutch Eating Behavior Questionnaire (DEBQ) have been studied in both obesity and eating disorders according to several symptoms and potential risk factors, we can still go further. The fact that most of the studies have been done using a small number of predictors leads us to ask ourselves the next question: why not study the problem in all its complexity? The spectrum that involves eating disorders and its relationship with eating styles is too broad to reduce, but to consider a large number of predictors and relations among them might lead to a heavy and impractical analysis. So here's when variable selection turns up with methods like the Lasso or BayesVarSel, which bring us the opportunity of considering the full range of variables of interest, selecting between them and giving us, as a result, a more parsimonious and accurate model. Thus, for the three different eating styles analyzed and without ignoring the problem of missing data, Lasso and BayesVarSel have shown similar results which are consistent with previous studies, where both extreme weight conditions analyzed (obesity and anorexia) compared with healthy controls, present different patterns. So, again, results suggest that these differences can have implications for understanding the development and maintenance of obesity and eating disorders.

Selection of the bandwidth parameter in a Bayesian kernel regression model for genomic-enabled prediction

Sergio Pérez-Elizalde, Jaime Cuevas, Paulino Pérez-Rodríguez, and José Crossa

One of the most widely used kernel functions in genomic-enabled prediction is the Gaussian kernel. Usually selection of the bandwidth parameter for kernel regression is based on cross-validation. In this study, we propose a Bayesian method for selecting the bandwidth parameter h of a Gaussian kernel as the mode of its posterior distribution. We present a theory for the Bayesian selection of h in a Transformed Gaussian Kernel (TGK) model and its application in two genomic plant breeding data sets (maize and wheat) that were already predicted using the kernel averaging (KA) method within the context of the Reproducing Kernel Hilbert Spaces' (RKHS KA). We also compared the prediction accuracy of the proposed method (TGK) with a model that uses a Gaussian kernel (GK) and estimates the bandwidth parameter using restricted maximum likelihood method (GK REML). Results for the wheat data set show that the predictive ability of TGK was on average 3% higher than the predictive ability of model RKHS KA, with TGK showing a smaller Predictive Mean Squared Error (PMSE) than the other two approaches. The advantages of the TGK model over GK REML in terms of PMSE were clear for one trait in nine environments. For the maize data set, the TGK model had slightly better prediction accuracy than methods RKHA KA and GK REML.

Bayesian joint models applied to paediatric Chronic Kidney Disease

Hèctor Perpiñán, Carmen Armero, and Anabel Forte

Bayesian joint models are an extension of the marginal models commonly used for modelling longitudinal and survival data. One of the main benefits of those models is that both submodels can share common information. When dealing with models under the Bayesian paradigm we manage to properly estimate the data as well as predict the progression of the outcomes of interest. This is a very valuable information for understanding the epidemiology of chronic diseases.

Chronic Kidney Disease (CKD) is a progressive loss of renal function which is quantified by means of the Glomerular Filtration Rate (GFR). The evolution of CKD patients is followed by subsequent hospital visits. This followup provides different types of information: longitudinal data on the evolution of GFR and survival time to get dialysis/transplant or recovery (only possible in a child's case). Data contain information from 168 children with CKD living in the Comunitat Valenciana who were diagnosed from 1st January 2005 until 31st December 2010 or before.

We used a Linear Mixed-effect submodel for the longitudinal data and a left-truncated competing risk submodel for the time to an event of interest. The introduction of the submodel for survival data allows us to manage the censured problems (informative missing data) caused by events of interest and shared this information with the longitudinal submodel to obtain more accurate predictions per patient. To sum up, this model permit us to approach to personalized medicine from a statistical point of view.

A project on Bayesian survival and capture-recapture methods

Blanca Sarzo, Carmen Armero, David Conesa, Olof Olsson, Henrik Österbloom, and Jonas Hentati

This poster presents my thesis project, which main objective is to analyse the survival of a seabird species, the Common Guillemot (*Uria aalge*). In particular its juvenile survival.

The Guillemot colony is located in a small island in the Baltic Sea, Stora Karlsö (Sweden) and has been monitored from 1987 till present. A total amount of 35.642 chicks have been ringed in that period and, each year (along the reproductive period), re-sightings of the individuals are done at the island.

Most sea-birds, as guillemots, spend major part of their life at sea, arriving at breeding islands only in the reproductive period. Indeed, when chicks reach their sexual maturity, they come back to their natal colony to breed (recruitment). For the Guillemots, its sexual maturity is at four years.

This colony is 'special' because, in contrast with other colonies, the chicks come back to their natal colony before reaching sexual maturity. This aspect gives us the opportunity to study the juvenile survival, a difficult aim to study in seabirds.

We will present here different statistical models that are used to survival and population analysis, being our interest to analyse them from the Bayesian perspective. In particular, we will emphasize and summarize their main objectives (parameters estimation), statistical aspects (priors, likelihood), bias and advantages/disadvantages.