BAYES 2017 LA MANCHA

22-23-24-25 May 2017

Faculty of Pharmacy
University of Castilla-La Mancha
Spain | Albacete Campus

http://www.bayes-pharma.org

info@bayes-pharma.org

The Adolphe Quetelet Society (IBS-Belgian Region)

ISBA INTERNATIONAL SOCIETY FOR BAYESIAN ANALYSIS

ROYAL STATISTICAL SOCIETY

UCLM CAMPUS DE EXCELENCIA INTERNACIONAL

Universidad de Castilla-La Mancha

FACULTAD DE FARMACIA UCLM

UCLM DPTO. DE MATEMÁTICAS

Welcome

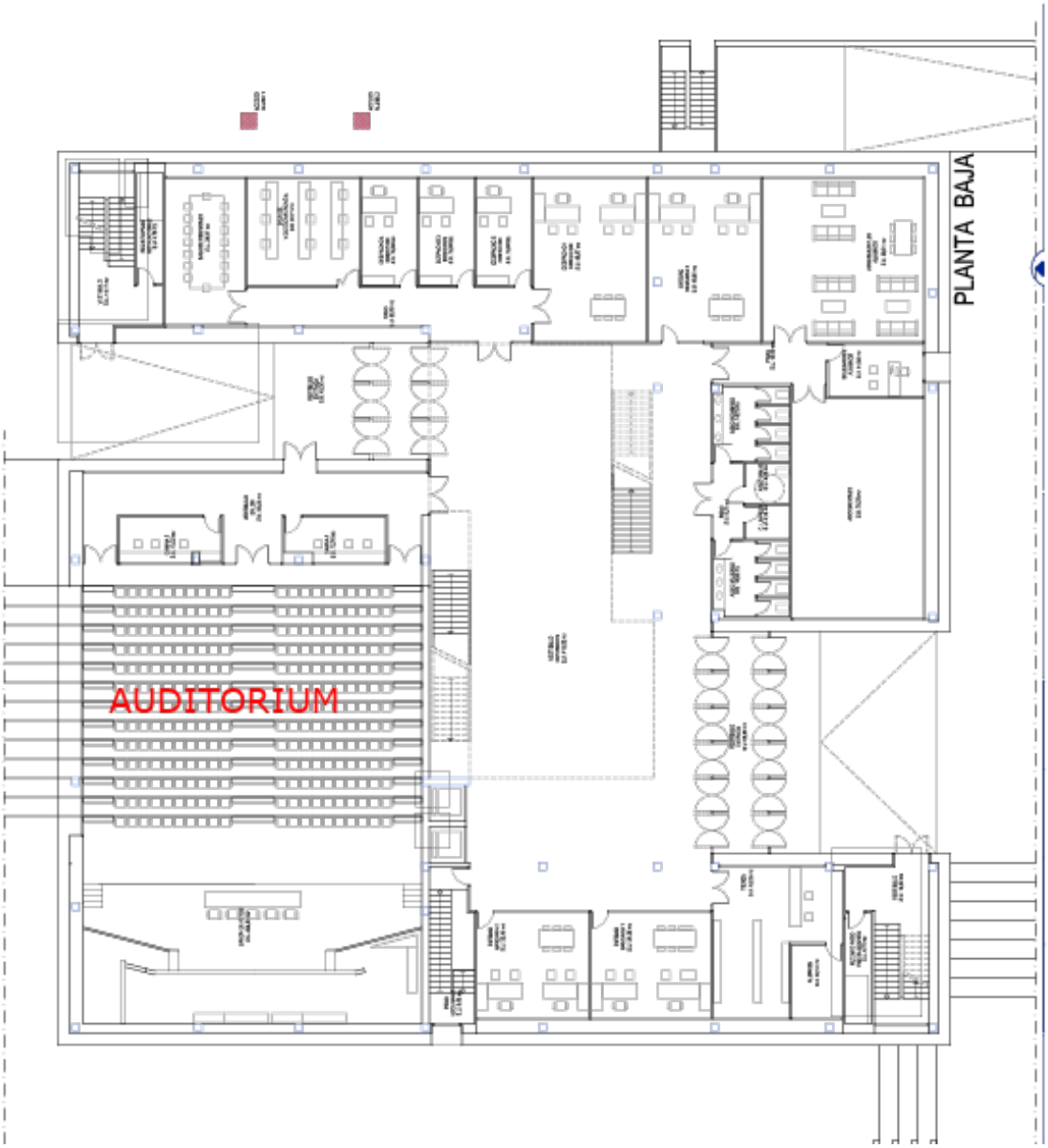
Somewhere in La Mancha, in a place whose name I do not care to remember...
we are organising the Bayes Pharma 2017 conference!!!

Dear participant,

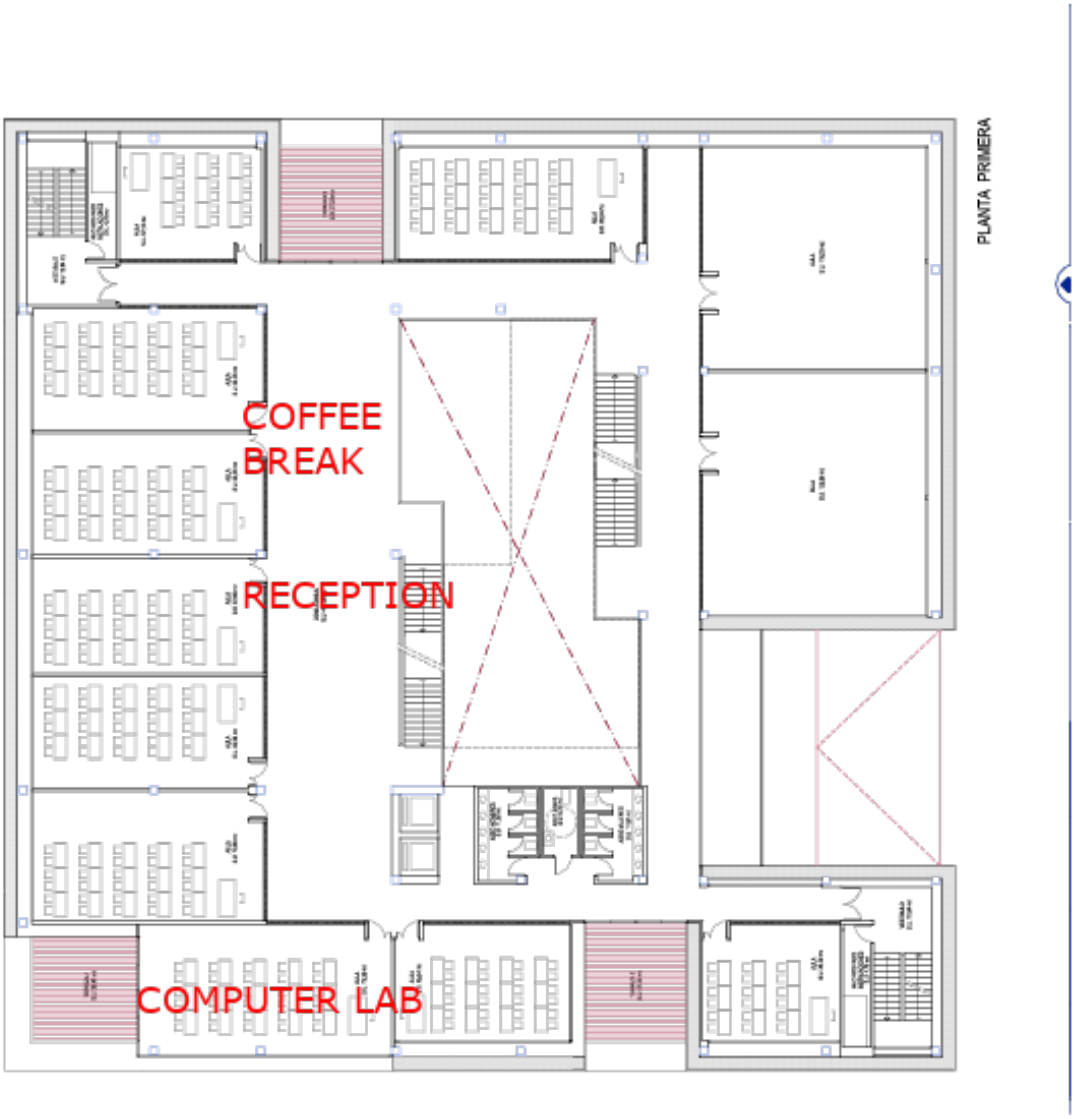
We welcome you to the 8th edition of the Bayes Pharma conference organised by the Faculty of Pharmacy of the University of Castilla-La Mancha. The conference continues its tradition of bringing together statisticians from the academy, the regulatory agencies and the pharmaceutical industry... with a Bayesian flavour! We wish you a fruitful conference and that you enjoy your time in Albacete.

Maps

Faculty of Pharmacy. Ground Floor



Faculty of Pharmacy. First Floor



Bayes Pharma 2017

UCLM Campus

- Facultad de Ciencias Económicas y Empresariales
- Facultad de Farmacia

Conference Dinner

- Nuestro Bar

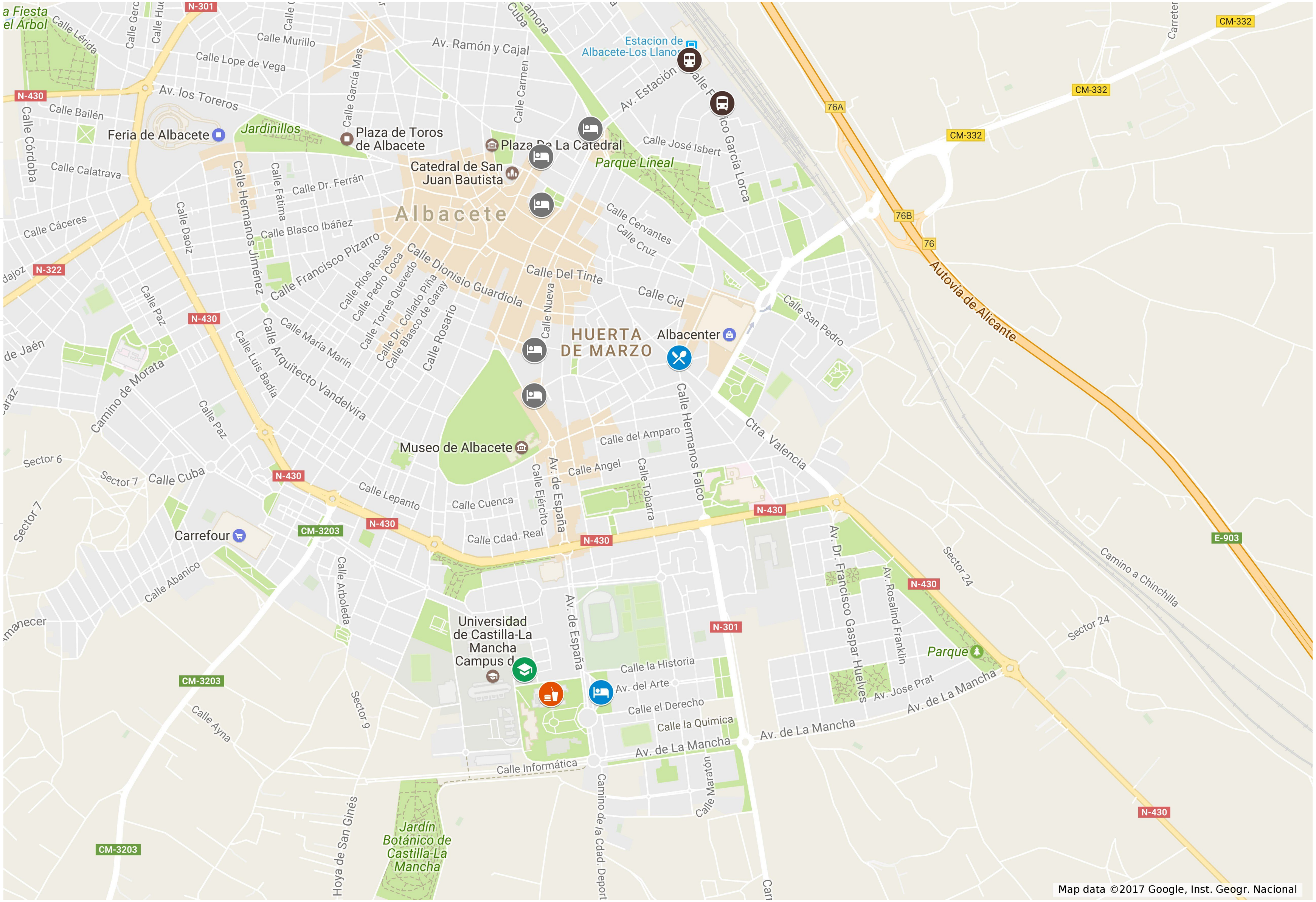
Public Transport

- Albacete-Los Llanos
- Bus Station

Hotels

- Hotel Universidad
- Hotel Sercotel Los Llanos
- Hotel San Jose Albacete
- Hotel Altozano
- Hotel San Antonio
- Hotel Castilla

Landmarks for Bayes Pharma 2017 conference



Conference Programme

Main locations

Pre-conference course	Fac. of Pharmacy	Computer lab, 1st floor
Reception	Fac. of Pharmacy	Hall, 1st floor
Coffe breaks	Fac. of Pharmacy	Hall, 1st floor
Lunch	Fac. of Law and Economics	Café, 1st floor
Main conference	Fac. of Pharmacy	Auditorium, ground floor
Conference dinner	Restaurante "Nuestro Bar"	c/ Alcalde Conangla, 102

DAY 1 - May 22nd, 2017

08:30 - 09:00	Registration
Fac. Pharmacy	
Pre-conference course (Computer Lab, 1st floor)	
09:00 - 10:30	Lecture 1
10:30 - 11:00	Coffee break
11:00 - 12:30	Practical 1
12:30 - 13:30	Lunch
Fac. Economics	
13:30 - 15:00	Lecture 2
15:00 - 15:30	Coffee break
15:30 - 17:00	Practical 2
18:00 - 20:00	Reception (Faculty of Pharmacy)

DAY 2 - May 23rd, 2017

08:30 - 09:00	Registration	
Hall		
09:00 - 09:30	Welcome	
Auditorium		
09:30 - 10:30	Invited Talk	
Auditorium	Variable selection techniques for covariate dependent Dirichlet process mixture models	Prof. Maria de Iorio
10:30 - 11:00	Coffee Break	
11:00 - 12:20	Contributed Talks	
Auditorium	Chair: Gianluca Baio	
	Decision making in early phase clinical trials using multivariate endpoints of mixed type	Juan J. Abellán et al.
	Trial Design using Preposterior Analysis and the Expected Value of Sample Information	A. Heath et al.
	Incorporating pre-clinical information into phase I trials: a Bayesian decision-theoretic approach	H. Zheng et al.
	Assessing early signs of efficacy while dealing with unknown dose-response relationship using a Bayesian approach with informative priors	A. Kaiser et al.
12:30 - 13:30	Lunch	
Fac. Economics		
13:30 - 14:30	Invited Talk	
Auditorium	Comparing Consensus Monte Carlo Strategies for Distributed Bayesian Computation	Steve L. Scott
14:30 - 15:30	Contributed Talks	
Auditorium	Chair: Gonzalo García-Donato Layrón	
	Informed interim analysis decision-making – a Bayesian Markov modelling approach	B. Magnusson et al.
	Bayesian joint models with parametric and non-parametric specifications of the baseline hazard function	E. Lázaro et al.
	Bayesian joint spatio-temporal analysis of different diseases	F. Palmí-Perales et al.
	Fitting complex Bayesian models with the Integrated Laplace Approximation and Monte Carlo sampling	V. Gómez-Rubio
15:30 - 16:00	Coffee Break	
20:00 - 00:00	Conference Dinner	

DAY 3 - May 24th, 2017

09:00 - 10:00	Invited Talk	
Auditorium	The past is prologue: the use of prior information and Bayesian methodologies in designing medical device regulatory approval	Dr. Roseann White
10:00 - 10:30	Coffee Break	
10:30 - 12:30	Contributed Talks	
Auditorium	Chair: Emmanuel Lesaffre	
	Assessing treatment effect in rare diseases with a single arm non-randomized study: A patient historic Bayesian Poisson mixed model approach	A. Monseur et al.
	Using historical data to inform extrapolation decisions in children	I. Wadsworth
	Combination of evidence from different continuous outcomes with a Bayesian approach constructing a common categorical distribution	D. Prieto-Merino et al.
	Modeling agreement on continuous recordings	S. Vanbelle et al.
	A Bayesian Model for Drug Response Estimation and Biomarker Testing using Gaussian Processes	F. Dondelinger et al.
12:30 - 13:30	Lunch	
Fac. Economics		
13:30 - 14:30	Invited Talk	
Auditorium	Using Statistical Innovation to Influence Regulatory Thinking	Dr. Harry Yang
14:30 - 15:30	Contributed Talks	
Auditorium	Chair: V. Gómez-Rubio	
	A Bayesian framework to evaluate calibration frequency in an automated immunoassay platform based on stability studies	L. Natalis et al.
	Bayesian Subgroup Analysis with Regularization and Widely Applicable Information Criterion	J. Takeda
	PPQ Sampling plan determination accounting for uncertainty in the estimation of tolerance intervals	L. Diya
15:30 - 16:00	Coffee Break	
17:00 - 19:00	Guided Tour	

DAY 4 - May 25rd, 2017

09:00 - 10:00	Invited Talk	
Auditorium	Bayesian approach to Optimal Experimental Design	Prof. Jesús F. López-Fidalgo
10:00 - 10:30	Coffee Break	
10:30 - 12:30	Contributed Talks	
Auditorium	Chair: E. López-Cano	
	Joint Modeling of Zero-Inflated Counts and Delayed Entries with Time-Varying Effects	X. Piulachs et al.
	Bayesian variable selection for identifying the source of food-borne disease outbreaks	R. Jacobs et al.
	From the laboratory to the clinic: A Bayesian perspective to accelerate the validation of a diagnostic test	M. Hennion et al.
	QSAR model for the estimation of lipophilicity in antidepressant drugs. Comparison between a classical model and a Bayesian model	J. Fernandez-Sainz et al.
12:30 - 13:30	Lunch	
Fac. Economics		

Pre-conference Course: Introduction to Bayesian survival models

Carmen Armero, Universitat de València, Spain
Danilo Alvares, Universitat de València, Spain
Elena Lázaro, Universitat de València, Spain

Source: Les Éditions Albert René

Survival analysis is one of the most important areas of applied and theoretical research in Statistics, with many important contributions in life sciences. This small course focuses on Bayesian reasoning in survival models. It contains the most basic elements and procedures in the subject without a strong theoretical approach but a conceptual and applied perspective that enables comprehensive modeling. Topics will be illustrated by means of real studies that will be subsequently worked with more depth in practical sessions.

Programme

- Lecture 1. Basic statistical concepts: survival function, hazard rate, censoring and truncation. Time-to-event regression models: accelerated failure models and proportional hazards (Cox) models.
- Practical 1. WinBUGS for censored data. Exponential data: Estimation and Prediction.
- Lecture 2. Frailty models. Joint models of longitudinal and survival data.
- Practical 2. Estimation and Prediction in Frailty models. Joint models.

Bibliography

R. Christensen, W.O. Johnson, A. J. Branscum and T. E. Hanson (2011). *Bayesian Ideas and Data Analysis*. Boca Raton: Chapman and Hall.

J. G. Ibrahim, M.-H. Chen, and D. Sinha (2001). *Bayesian Survival Analysis.* New York: Springer

J. P. Klein and M. L. Moeschberger (2003). *Survival Analysis: Techniques for Censored and Truncated Data* Second Edition. New York: Springer-Verlag

Variable selection techniques for covariate dependent Dirichlet process mixture models

Maria de Iorio

University College London

Dirichlet Process Mixture (DPM) models have been increasingly employed to specify random partition models that take into account possible patterns within covariates. Furthermore with large numbers of covariates methods for selecting the most important covariates have been proposed. Commonly the covariates are chosen either for their importance in determining the clustering of the observations or for their effect on the level of a response variable (when a regression model is specified). Typically both strategies involve the specification of latent indicators that regulate the inclusion of the covariates in the model. Common examples involve the use of spike and slab prior distributions. In this work we review the most relevant DPM models that include covariate information in the induced partition of the observations and we focus on available variable selection techniques for these models. We highlight the main features of each model and demonstrate them using simulations and a real data application.

Comparing Consensus Monte Carlo Strategies for Distributed Bayesian Computation

Steve L. Scott

Google

Consensus Monte Carlo is an algorithm for conducting Monte Carlo based Bayesian inference on large data sets distributed across many worker machines in a data center. The algorithm operates by running a separate Monte Carlo algorithm on each worker machine, which only sees a portion of the full data set. The worker-level posterior samples are then combined to form a Monte Carlo approximation to the full posterior distribution based on the complete data set. We compare several methods of carrying out the combination, including a new method based on approximating worker-level simulations using a mixture of multivariate Gaussian distributions. We find that resampling and kernel density based methods break down after 10 or sometimes fewer dimensions, while the new mixture-based approach works well, but the necessary mixture models take too long to fit.

The past is prologue: the use of prior information and Bayesian methodologies in designing medical device regulatory approval

Roseann White

Duke Clinical Research Institute

There have been many paper on the advantages and disadvantages of the Bayesian power prior in clinical trials. However, there have been several roadblocks to the acceptance of this methodology when planning medical device feasibility and approval clinical trials that are required regulatory approval in key geographies. Some of the challenges that MDIC has trying to address are the following:

1. Clarifying for the clinical community does it mean to design a clinical trial where one can combine the prior information about the device with the current trial data to improve the precison in the outcome estimates or potentially reduce the sample as compared to a frequentist approach
2. Clarifying what is expected by CDRH when putting together a protocol so that the review and approval process goes smoothly
3. Developing a transparent process that allows one to pre-specifyhow the prior information will be weighted based on type I error, assuring that the bias that will be introduced by the prior is clinically acceptable and the probability of success of the trial given the assumed true outcome is acceptable (the loss function).

This talk will highlight some of efforts that are currently under way or have already been completed for bullets 1 and 2 but will mostly focus on role of the statistician and methods that the statistician will employ to address bullet 3 when designing a clinical trial using Bayesian analysis with an informative prior trial.

Using Statistical Innovation to Influence Regulatory Thinking

Harry yang

MedImmune, USA

A looming deadline of submission, an unanswered regulatory inquiry, and a purification process seemingly short of regulatory expectation,... who would guess that an FDA-recommended acceptance criterion for residual host cell DNA would become the subject of statistical controversy. This presentation discusses how we resolved a critical CMC regulatory issue, by applying statistical innovation to extreme. It also speaks to the fact that close collaboration between statisticians and scientists often works wonders.

Bayesian approach to Optimal Experimental Design

Jesús F. López-Fidalgo

University of Navarra

A unified view of the topic is presented by putting experimental design in a decision theoretic framework. Experimental design is the only situation where it is meaningful within the Bayesian theory to average over the sample space. As the sample has not yet been observed the general principle of averaging over what is unknown applies. This framework justifies many optimality criteria and opens new possibilities. Various design criteria become part of a single coherent approach. Linear and nonlinear models will be considered as well as a particular application of an optimality criterion for discriminating between any two statistical models in the presence of prior information. If the rival models are not nested then, depending on which model is true, two different Kullback-Leibler distances may be defined. The Bayesian KL-optimality criterion is a convex combination of the expected values of these two possible Kullback-Leibler distances between the competing models. Concavity of the Bayesian KL-optimality criterion allows the use of the classical results of Optimal Design Theory. A standardized version of the proposed criterion is also given in order to take into account possible different magnitudes of the two Kullback-Leibler distances. Some illustrative examples will be provided.

Trial Design using Preposterior Analysis and the Expected Value of Sample Information

Anna Heath, Ioanna Manolopoulou and Gianluca Baio

University College London

Preposterior analysis is concerned with estimating the properties of a potential posterior distribution before relevant data have been collected. This idea can be embedded within a health economic framework to create a decision theoretic measure known as the Expected Value of Sample Information (EVSI), which gives the potential economic or monetary value of a future data collection exercise. Theoretically, this could be an important tool for trial design as researchers could compare the value of the trial directly with its cost and could also determine the optimal trial design in terms of monetary benefit by comparing value and cost across different trial designs.

Despite these useful features, the practical application of the EVSI in trial design has been restricted due to computational issues. However, methods have, recently, been developed to overcome these computational barriers using a large number of statistical tools. Our method focuses on determining the properties of the distribution of the preposterior mean incremental net benefit. While the posterior mean incremental net benefit is a constant after the data has been observed, the preposterior mean has a distribution. This distribution is induced because the data has not been observed and is therefore sampled from the prior predictive distribution. Our knowledge about the distribution of the preposterior mean can be combined with Bayesian non-linear regression to calculate the EVSI for different sample sizes with a significantly reduced computation cost compared to standard EVSI analysis.

Using historical data to inform extrapolation decisions in children

Ian Wadsworth, Lisa V. Hampson, Thomas Jaki and Graeme J. Sills

Lancaster University

When developing a new medicine for children, the potential to extrapolate from adult efficacy data is well recognised. However, significant assumptions about the similarity of adults and children are needed for extrapolations to be biologically plausible. One such assumption is that pharmacokinetic-pharmacodynamic (PK-PD) relationships are similar in these different age groups. In this presentation we consider how 'source' data available from historical trials completed in adults or adolescents treated with a test drug, can be used to quantify prior uncertainty about whether PK-PD relationships are similar in adults and younger children. A Bayesian multivariate meta-analytic model is used to synthesise the PK-PD data available from the historical trials which recruited adults and adolescents. The model adjusts for the biases that may arise since these existing data are not perfectly relevant to the comparison of interest, and we propose a strategy for eliciting expert prior opinion on the size of these external biases. From the fitted bias-adjusted meta-analytic model we derive prior distributions which quantify our uncertainty about the similarity of PK-PD relationships in adults and younger children. These prior distributions can then be used to calculate the probability of similar PK-PD relationships in adults and younger children which, in turn, may be used to inform decisions as to whether a complete extrapolation of efficacy data from adults to children is currently justified, or whether additional data in children are needed to reduce uncertainty. Properties of the proposed methods are assessed using simulation, and their application to epilepsy drug development is considered.

Bayesian variable selection for identifying the source of food-borne disease outbreaks

Rianne Jacobs, Emmanuel Lesaffre, Peter Teunis, Jan van de Kassteele

With food chains becoming increasingly complex and food products being transported across the globe with increasing ease, contaminated food products can rapidly cause food-borne disease outbreaks. Identification of contaminated food products is a cumbersome process. Analytic case-control studies are the main epidemiological tool in this process of identification. Once an outbreak has been detected, patients and controls fill out an extensive food consumption questionnaire. These are usually analyzed using classical logistic regression. One can, however, imagine the practical difficulties that subjects have trying to recall their dietary consumption and the resulting amount of missing values in such data. In addition, it is impossible to confirm whether controls are indeed true controls (i.e. not infected) or rather asymptomatic infections (i.e. infected but not ill), resulting in misclassification of the response. Due to the large number of different food products people may have consumed, one has a variable selection problem, where one attempts to identify relevant exposures. When searching for the cause of the outbreak we, therefore, need a far more sophisticated procedure. We argue that the Bayesian approach offers tools to deal with variable selection problems with missing covariates and misclassified responses. We present a Bayesian spike and slab variable selection method which extends scaled logistic regression for dealing with misclassified responses. To deal with missing covariates, we perform variable selection in each of the regression models of the covariate probability model as well. We apply our model to data of a large Salmonella outbreak in the Netherlands. We show that, using the posterior inclusion probabilities, we can correctly identify the contaminated food product. Further extensions of the model include dynamic updating, detailed prior construction and possibly principal component analysis to reduce the number of covariates in early stages of the outbreak analysis.

From the laboratory to the clinic: A Bayesian perspective to accelerate the validation of a diagnostic test

Maud Hennion and Bruno Boulanger

Arlenda S.A, Mont-saint-Guibert, Belgium

In the frame of a disease with a low prevalence –or unknown prevalence as for immunogenicity test- but with major consequences, the development and validation of diagnostic tests with high PPV –Positive Predictive Value (or NPV – Negative Predictive Value) is crucial for ensuring reliable decision making by the medical bodies. Not taking appropriate actions while needed could be deleterious for the patient and on the other side taking too often action could be unnecessarily expensive.

In this presentation we'll present a global Bayesian perspective to develop and validate in a clinical setting a diagnostic test from the data published to the final assessment of the performances. First we'll reformulate the overall objective of sensitivity, specificity, prevalence and PPV (NPV) in a Bayesian setting. We'll point out the benefit, in this context, of assessing the test's performance based on the NPV (or PPV) and not on the usual assay characteristics as Sensitivity and Specificity. Second, we'll recall the methodology to derive quantitatively priors based on the Meta-analytic Predictive Prior (MAP) methodology adapted to binary outcomes using the Beta-Binomial distribution. Third, we'll show how to theoretically leverage available laboratory and published data to inform the design. Finally, we'll also show the ways and simulations to perform to optimize and justify a clinical trial to obtain the best posterior information about the PPV (NPV), bringing the appropriate level of confidence.

PPQ Sampling plan determination accounting for uncertainty in the estimation of tolerance intervals

Luwis Diya, Martin Otava, Hans Coppenolle and Helena Geys

Janssen

Process Validation is all about the collection and analysis of data from the design to the manufacturing phases of a product with the aim of confirming that the process is capable of outputting quality products. Process Validation consists of three steps, namely process design, process qualification and continued process verification. In process qualification there are two phases, that is equipment qualification and process performance qualification (PPQ).

PPQ can further be subdivided into two sub-steps: (i) PPQ sampling plan and (ii) PPQ validation. At the stage of the design of the PPQ sampling plan, the number of within-batch samples is determined in order to demonstrate whether the process is generating quality product. To that end, process tolerance intervals and the probability to fail specifications (POOS) for a future random batch are used.

However, at the stage of validation, batch specific tolerance intervals are often used (by comparison with specifications) to confirm that the process is capable of reproducible commercial manufacturing. The determination of the sampling plan is not completely aligned with the expectations of the validation step since (1) the POOS is a statement on batch rejection and not validation, and (2) the process tolerance interval is a "point estimate" of the distribution of tolerance intervals (neglecting the uncertainty in the estimation).

We propose combining the two metrics (tolerance intervals and POOS) so that the design of the PPQ sampling plan is aligned with the expectations of the PPQ validation stage. That is, we propose using the calculation of the probability that a tolerance interval for a future (validation) batch will be within specifications using a double simulation approach. In doing so, we take the uncertainty in the estimation of tolerance interval into account and we quantify the risk of the tolerance interval not falling within specifications.

Bayesian joint models with parametric and non-parametric specifications of the baseline hazard function

E. Lázaro[1], C. Armero[1], D. Alvares[1] and M. Rué[2,3]

[1] Universitat de València, Department of Statistics and Operational Research, Doctor Moliner, 50, 46100-Burjassot, Spain

[2] Universitat de Lleida-IRBLLEIDA, Department of Basic Medical Sciences, Avda. Rovira Roure, 80, 25198-Lleida, Spain

[3] Health Services Research Network in Chronic Diseases (REDISSEC), Spain

Joint models for longitudinal and survival data allow to incorporate endogenous temporal covariates in survival studies. The simplest structure considers a longitudinal mixed effects model for the trajectory of the time-dependent predictor and a Cox proportional hazard model for the relevant time-to-event. The frequentist paradigm has provided little prominence to the hazard baseline function. In the Bayesian paradigm, however, this function has to be specified.

We analyse the impact of different parametric and non-parametric proposals for the baseline hazard function in the framework of Bayesian joint models. Parametric approaches provide strictly monotone baseline hazard estimations. On the contrary, non-parametric choices allow flexible patterns. We considered a Weibull distribution as a parametric choice, and piecewise constant and B-splines basis functions as non-parametric proposals. Within the non-parametric approaches, we consider prior scenarios that impose different smooth shape conditions: from prior independence to correlated structures such as discrete-time martingale processes, Gamma process priors, and random-walk priors.

The proposals were discussed in a real study devoted to assess the relationship between the risk of death or be discharged alive in intensive care units and a longitudinal severity index. The joint model was specified by means of a longitudinal mixed effects and a survival competing risk submodels. Both submodels were connected through a vector of shared random effects. Markov Chain Monte Carlo methods were adopted for parameter estimation by means of the JAGS software. Model selection was based on the deviance information criterion (DIC) and the Log Pseudo-Marginal Likelihood criterion (LPML).

References

C. Armero, C. Forné, M. Rué, A. Forte, H. Perpiñán, G. Gómez, and M. Baré (2016). Bayesian joint ordinal and survival modelling for breast cancer risk assessment. Statistics in Medicine, 35(28), 5267-5282.

J. G. Ibrahim, M. H. Chen, and D. Sinha, D. (2005). Bayesian survival analysis. John Wiley & Sons, Ltd.

Informed interim analysis decision-making – a Bayesian Markov modelling approach

Baldur Magnusson, Marina Savelieva, Dieter Häring, Ekkehard Glimm, Wenping Wang

Novartis Pharma AG

A common approach for phase 3 studies in many disease areas is to define a primary endpoint in terms of time to a specific event of interest (e.g. time to disease progression). These studies are typically large and expensive, and use of interim analyses for early futility stopping is therefore an appealing option. However, if events are defined retrospectively (e.g. after a confirmation period) then applying standard time-to-event models to inform interim decision making is suboptimal.

We develop a model for disease progression that allows long-term prediction of the primary endpoint given short-term interim data. Longitudinal data from historical studies is modelled with a Bayesian discrete time Markov model where, at every visit, patients may transition between progressive and non-progressive states. Considerations are given to a proper handling of dropouts, model validation and development of a simulation engine to compute interim predictive power in a future phase 3 study.

Several versions of the Markov model have been explored, including joint models that explicitly take into account the time to dropout. Posterior predictive diagnostics indicate that these models can reproduce historical disease progression and Kaplan-Meier curves.

The use of a Markov model allows tracking of patient transitions from stable to progressing disease over time. Compared to standard time-to-event models, the Markov model is more efficient as it allows the use of information from patients who have an onset of progression but insufficient follow-up time to confirm the event, and may therefore allow earlier and more reliable decision making than time-to-event methods.

Assessing early signs of efficacy while dealing with unknown dose-response relationship using a Bayesian approach with informative priors

Andreas Kaiser and Richardus Vonk

Research & Clinical science Statistics, Bayer AG

To support decision making about larger investments in drug development e.g. for indication expansions, often small clinical trials are conducted to generate early signs of efficacy. If the shape of the underlying dose-response relationship is unknown, efficacy evaluation becomes more difficult due to the small number of patients included in such trials. To tackle this challenge, a four step approach is proposed, allowing a rough assessment of the underlying dose-response shape and the quantification of efficacy signals. In the first step, the potential underlying treatment responses at each dose are analyzed using Bayesian approach allowing the incorporation of prior information. Next, potential underlying dose-response relationships are compared with pre-specified models. By this, probabilities for a model being the best approximation of underlying dose-response are derived. In the third step, posterior probabilities for maximum treatment effect are calculated for each model. Finally, these posterior probabilities are averaged using the probabilities for model selection as weights resulting in a probabilistic quantification of early signs of efficacy. Results from simulation studies about the operational characteristics of this approach will be presented.

Bayesian Subgroup Analysis with Regularization and Widely Applicable Information Criterion

Jun Takeda

Astellas Pharma Inc., Tokyo, Japan

The Bayesian subgroup analysis is a methodology where the posterior distribution of the subgroups is interpreted considering the research purpose. Typically, the posterior distribution has some shrinkage: the configuration of the subgroups is moved toward the center from the original configuration. The degree of the shrinkage depends on the probabilistic model and the choice of the prior distribution. The most well-known models are 1) complete-pooling model, 2) partial-pooling model (or hierarchical model), and 3) no-pooling model. These models have strongest shrinkage, moderate shrinkage, and no shrinkage respectively. The exchangeability-nonexchangeability (EXNEX) model was also proposed to bridge the two models 2) and 3) as a mixture.

In this study we propose another bridge with regularization and with Widely Applicable Information Criterion (WAIC). The L1 regularization (Bayesian Lasso) and L2 regularization (Bayesian Ridge) correspond to the Laplace prior distribution and the normal prior distribution respectively. The tuning parameter is compared with WAIC on how well the model is fitted to the data.

The proposed model was applied to various data sets with subgroups, including data sets from global clinical trials, oncology trials, and meta-analysis settings. In many cases, the models with regularization have some tuning values that have less WAIC than the models 1), 2), 3), and the EXNEX model with equal prior mixture weights. This means that the proposed model may be one way to improve modeling in Bayesian subgroup analysis from the existing models. In addition, the proposed model may give some insight to understand the nature of the data because the model continuously connects the models 1) and 3) with the fixed tuning parameter and allow people to observe gradual change of the relation between the tuning parameter, WAIC, and the configuration of the subgroups.

Decision making in early phase clinical trials using multivariate endpoints of mixed type

Juan J Abellan, Yansong Cheng, Rosemary Schroyer and Nicky Best

GlaxoSmithKline

In the early phases of the clinical development of a new drug, study teams sometimes have multiple endpoints of interest that can potentially be used for defining study success. When there is no clear preference of one of the endpoints over the others, decision rules for study success are typically constructed using the results of the endpoints analysed independently connected with Boolean operators (e.g. obtain a relevant result in endpoint 1 OR endpoint 2). However, this type of decision rules poses multiplicity issues, thus decreasing the chances of making the correct decision at the end of the study. Within the Bayesian framework, one way to overcome multiplicity issues is to jointly model the endpoints of interest and build decision rules based on posterior probabilities for the regions of the joint parameter space of interest linked to success.

When all endpoints are continuous, a multivariate normal can be used to jointly model the endpoints of interest. When there are endpoints of different nature (e.g. continuous, binary and time-to-event), one could resort to other approaches such as copula models, which may not be so straightforward. Here we present an approach based on Bayesian hierarchical models whereby endpoints are correlated in the parameter space (as opposed to the endpoint space). We illustrate the use of such models to build multivariate decision criteria for study success and demonstrate the benefits over independent analysis of each endpoint using simulations to assess the operating characteristics. A phase 2a study with binary, time-to-event and continuous endpoints provides the motivating example.

Modeling agreement on continuous recordings

Sophie Vanbelle[1] and Emmanuel Lesaffre[2]

[1] Department of Methodology and Statistics, Maastricht University, Maastricht, The Netherlands
[2] Leuven Biostatistics and Statistical Bioinformatics Centre, KU Leuven, Leuven, Belgium

Continuous recordings are defined as second by second or even closer records in time and belong to intensive longitudinal data. They are increasingly frequent in medical and behavioral sciences. They can result, for example from the observation of complex behaviors on video recordings or from the collection of data through handled portable electronic data-entry and storage devices (e.g. e-Health). In both cases, the reliability and the validity of the coding scheme/measurement instrument have to be assessed before using the continuous recordings in daily practice.

For example, the CAM study was designed to validate a new single-unit activity monitor (CAM) in patients with chronic organ failure (Annegarn et al., 2011). In the study, the activity (non-weight bearing posture (NWBP), weight-bearing posture (WBP) or dynamic activity (DA)) of 10 patients in rehabilitation was recorded during one hour of daily routine by the CAM, worn simultaneously on the leg and on the trunk for comparative purposes. The patients were also videotaped by a researcher during the same time period. The video, considered as criterion standard, was then analyzed second by second by a researcher blinded to the values obtained by the CAM. The aim of the study was to determine the body place providing the highest agreement level between the CAM and the video.

The existing methods to determine agreement in the presence of a covariate structure and repeated measurements have to be ruled out to analyze the present study because of the large amount of repeated measurements (i.e., 3600) and the small number of subjects involved (i.e., 10). We therefore developed a partial-Bayesian approach, based on previous work (Vanbelle and Lesaffre, 2015), to model agreement levels obtained on continuous recordings over time in the presence of a covariate structure. In particular, we consider that the time spent under a position over a small time period follows a betabinomial distribution.

References

- Annegarn J., Spruit M.A., Uszko-Lencer N.H.M.K., Vanbelle S., Savelberg H.H.C.M., Schols A.M.W.J., Wouters E.F.M. and Meijer, K. (2011). Objective Physical Activity Assessment in Patients with Chronic Organ Failure: A Validation Study of a New Single-Unit Activity Monitor, Archives of Physical Medicine and Rehabilitation, 92, 1852–1857.e1, DOI: 10.1016/j.apmr.2011.06.021

- Vanbelle S. and Lesaffre E. (2015). Modeling agreement on categorical scales in the presence of random scorers. Biostatistics, 17, 79–93. DOI: 10.1093/biostatistics/kxv036

Incorporating pre-clinical information into phase I trials: a Bayesian decision-theoretic approach

Haiyan Zheng, Lisa Hampson

Department of Mathematics and Statistics, Lancaster University, UK

Bayesian model-based procedures become increasingly important to design phase I dose-escalation trials, where the principal aim is to estimate the maximum tolerated dose of a novel compound. Conventionally, vague prior distributions have been used for model parameters. However, information on the dose-toxicity relationship is commonly available from pre-clinical toxicology studies by the time first-in-man trials are conducted. Should the pre-clinical experimental findings be commensurate with the toxicity in humans, incorporating them will lead to more efficient decision process and enhanced precision of the resulting recommended dose in the phase I trial. Such advantages, however, must be balanced against the risk that more patients may be treated with excessively toxic doses when prior-data conflict commences.

We propose a Bayesian decision-theoretic approach, with which the degree of commensurability can be dynamically measured. Central to this approach is a utility function that formalises the benefits of incorporating pre-clinical information under various potential risks. In particular, pre-clinical information is used to predict whether the incoming patients would experience dose-limiting toxicity (DLT) or not. These predictions are optimal in the sense of maximising the prior expected utility. Correct predictions will be assigned with a utility of 1, incorrect predictions of no-DLT with a utility of 0, and incorrect predictions of DLT with a utility between 0 and 1. At each interim analysis these prior predictions are compared with the actual human outcomes. The attained predictive utility, expressed as a fraction of the maximum utility achieved when all prior predictions are correct, is then helping quantify the weight to be attributed to the pre-clinical information. Simulations demonstrate our approach is competitive and robust. First, our approach can essentially borrow strength from pre-clinical information according to the assessed commensurability. Second, pre-clinical information will be downweighted quickly if incorporating it may undermine the safety of patients

Bayesian joint spatio-temporal analysis of different diseases

V. Gómez-Rubio[1], **F. Palmí-Perales**[1], G. López Abente[2], R. Ramis Prieto[2], P. Fernández Navarro[2]

[1] University of Casilla-La Mancha , Spain
[2] Instituto de Salud Carlos III - ISCIII, Spain

In recent years, different Bayesian hierarchical models have been proposed to analyze public health aggregated data. Many of this proposals are focused on extending or adapting the approach of Besag et. al. (1991), such as Abellan et. al. (2008) or Guangquan et. al. (2012). Disease mapping and spatial epidemiology have mainly been focused on modelling a single disease, but there are diseases that share risk factors and may have similar behaviours so a joint approach can be usefut, such as Downing et. al. (2008) present.

In this study a novel Bayesian hierarchical spatio-temporal model is proposed for the joint analysis of several diseases. This model allows for specific and joint spatial and temporal patterns, so that different effects can be disentangled. Our proposal extends other spatio-temporal models in a number of ways. First of all, it allows areas to show specific patterns that depart from joint spatial and temporal patterns. Secondly, the model also includes shared spatial and temporal patterns common to all diseases, as well as specific spatial and temporal patterns to each disease. Dependence on all shared spatial and temporal patterns is modulated by disease-specific weights that can be used to assess whether different diseases have similar patterns.

The suggested model has been used to study three different cancers (oral cavity, esophagus and stomach cancer) in Spain at province level. Shared and specific spatial, temporal and spatio-temporal effects have been estimated and mapped in order to study similarities and di erences among these causes. Results have been compared with studies of the same diseases such as López-Abente et. al. (2007) or López-Abente (2014).

QSAR model for the estimation of lipophilicity in antidepressant drugs. Comparison between a classical model and a Bayesian model

Jesús Fernández-Sainz[1], Pablo Atiénzar-Martínez[1], Carla Sáez-García[1], Silvia López-González[1], Pedro J. Pacheco-Liñán[1], Iván Bravo[1], Andrés Garzón[1], José Albaladejo[1], Matías Gámez[2], Emilio L. Cano[2,*]

[1] Department of Physical Chemistry, Faculty of Farmacia, University of Castilla-La Mancha (UCLM), Paseo de los Estudiantes, s/n, 02071, Albacete, Spain

[2] Quantitative Methods and Socio-economic development Group, Institute for Regional Development (IDR), University of Castilla-La Mancha (UCLM), Albacete, Spain

[*] Corresponding author: emilio.lcano@uclm.es

The drug's mechanism does not only depend on the interaction between the drug molecule and the biological target, but also on different physical properties as lipophilicity, solubility in water, acidity/basicity, and stability, among others [1]. Lipophilicity plays a main role in governing kinetic and dynamic aspects of drug action. Optimal lipophilicity values lead to good absorptions of drugs in the intestine by passive diffusion [2]. The common quantitative descriptor of lipophilicity is the octanol-water partition coefficient, log P, which is defined as

Log P = log ($[\text{drug}]_{oct}/[\text{drug}]_{wt}$),

where $[\text{drug}]_{oct}$ and $[\text{drug}]_{wt}$ are the solubility of the drug in n-octanol and water. Log P is an important parameter monitored by medical chemist in drug discovery. In this work, we present a Quantitative Structure-Activity Relationship (QSAR) model to estimate the log P in a set of antidepressant drugs based on a series of molecular properties (independent variables) such as dipole moment, solvation energy, polarizability, molecular volume, solvent accessible surface area, etc. These properties were calculated by means of different Quantum Methods. The resulting QSAR classical prediction model under the linear regression approach is compared with a Bayesian regression model incorporating expert knowledge about the physical understanding of the compounds.

References

[1] C. Quiñones-Torrelo et al., Journal of Chromatography B, 2004, 801, 185-198.

[2] R. Mannhold et al., Journal of Pharmaceutical Sciences, 2009, 98, 861-893

Combination of evidence from different continuous outcomes with a Bayesian approach constructing a common categorical distribution

David Prieto-Merino[1,2], Jana Vandrovcova[1], Juan Pablo Casas[1], Luis Prieto-Valiente[2]

[1] University College London, United Kingdom
[2] Universidad Católica San Antonio de Murcia, Spain

When studying the association of an exposure (X) with a disease (D) the clinician is often interested in whether the exposure is beneficial, neutral or harmful. However, we might find studies associating that exposure with different signs or symptoms of the disease (Y1, Y2... Yk) measured in different scales: Odds Ratios (OR), Risk Ratios (RR), Mean Differences (MD). The question is how to formally combine all these associations to give a final believe on the beneficial, null or harmful effect of X on D. For example, suppose that, a genetic trait (X) is negatively associated with cognitive function (OR=0.95 [0.85, 1.05]), with volume of the hippocampus (MD= -10 % [-25 %, +5 %]), and with motor function (RR= 0.90 [0.80, 1.10]). None of these associations are "significant" but they all point to a negative effect of trait X on some sort of neurodegenerative disease (D). Can we quantify a posterior probability for this?

We have developed a Bayesian method with the following steps: 1) Calculate the posterior of each outcome in a suitable continuous scale (log(OR), log(RR)...), 2) Convert each posterior to a 3-category posterior (beneficial / null / harmful) using medically suitable boundaries in the continuous scale of each outcome, 3) Combine categorical posteriors of all outcomes by multiplying probabilities for each category and normalising across categories (at this point the posteriors could be weighted by their clinical relevance). Further, for step-1 we have developed an algorithm to take prior probabilities in the 3-category scale [beneficial / null / harmful] which we believe is more user-friendly. We present an example with real data for the association of 1016 genes with Alzheimer's Disease combining information from GWAS meta-analyses on 7 different outcomes. We have designed specific graphs and we show some sensitivity analysis of the categorical priors.

Joint Modeling of Zero-Inflated Counts and Delayed Entries with Time-Varying Effects

Xavier Piulachs[1], Dimitris Rizopoulos[2], Eleni-Rosalina Andrinopoulou[2], Montserrat Guillen[1]

[1] Department of Econometrics, Riskcenter-IREA, University of Barcelona, Barcelona, Spain

[2] Department of Biostatistics, Erasmus Medical Center, Rotterdam, The Netherlands

In the context of health care studies, repeated measurements of medical claims are intermittently collected on each subject as a usage rate indicator of medical services. This outcome is usually restricted to a small range of non-negative integer values, affected by some level of overdispersion if the observed variance exceeds the observed mean. Additionally, counts with a large number of zeros become highly usual owing to the nature of insurance data, where a lack of information from some subjects happens regarding the behavior of interest. The aging process taking place in developed countries leads to a natural interest about assessing the relationship between critical care demand and survival. Hence, elderly policyholders need to be informed about adequately annuities according to its individualized health status, whereas insurance companies want to plan the potential costs of tackling lifetimes above the mean expectations. Building in such a scheme, joint modeling techniques provide useful tools to properly address the relationship between historical records and the hazard for death event.

Our work is motivated by a health insurance dataset of Spanish policyholders aged 65 and over, since they request critical medical care and hospitalization more often than younger individuals. The response of interest in the longitudinal analysis is the annual count of emergency claims, recorded over the 8-year study period. The observation of some subjects started after the age of 65 years, defined here as time zero, so we have to account for delayed entries in the survival approach in addition to the usual censoring mechanism. We propose a joint model to analyze the degree of dependence between two approaches. The counting sequence is undertaken by a hierarchical zero-inflated response, and the time-to-event approach is handled by a proportional hazards survival model. Because an increasing frequent demand is well-documented as policyholder ages, a time-varying association parameter between both outcomes is proposed. The estimation of parameters is performed under a Bayesian perspective by Markov chain Monte Carlo simulation methods.

Fitting complex Bayesian models with the Integrated Laplace Approximation and Monte Carlo sampling

Virgilio Gomez-Rubio

[1] Universidad de Castilla-La Mancha, Albacete (Spain)

The Integrated Nested Laplace Approximation (INLA, Rue et al., 2009) is now a popular method for approximate Bayesian inference on models that can be expressed as a latent Gaussian Markov random field (GMRF). In addition to providing accurate estimates of the posterior marginals, INLA relies on computational methods that reduce computing time as compared to other estimation methods based on Monte Carlo sampling, such as MCMC.

Recently, Gomez-Rubio and Rue (2017) have shown how to combine INLA and MCMC to extend the class of models that can be fitted with INLA provided that the model can be written as a conditional latent GMRF model. They develop a Metropolis-Hastings algorithm with INLA to estimate the joint posterior distribution of a small ensemble of parameters. The posterior marginals of the other parameters are obtained by Bayesian model averaging the conditional marginals obtained with INLA at each step of the Metropolis-Hastings algorithm.

The idea of obtaining an estimate of the posterior distribution for a small ensemble of parameters by combining Monte Carlo methods and INLA can be applied to many other situations. For example, mixture models can be expressed as conditional latent GMRF on a set of auxiliary variables that assign observations to mixture components. Furthermore, other Monte Carlo methods can be used together with INLA to fit complex Bayesian models. In particular, importance sampling can be efficiently combined with INLA to obtain a sample from the posterior marginal of a small ensemble of parameters and a set of conditional marginals for all the other parameters. These conditional marginals can be combined to obtain the posterior marginals. As importance sampling is not a sequential method, this allows model fitting to be done in parallel.

In my talk I will describe how inference on complex models can be tackled with INLA and other Monte Carlo methods in different contexts. In particular, I will discuss the Bayesian lasso for variable selection, fitting models with missing values in the covariates in a study on cholesterol and obesity and fitting complex multivariate spatial models for disease mapping. I will also give some hints about how these models can be fitted with INLA using parallel computing.

References

- V Gomez-Rubio and H. Rue (2017) Markov Chain Monte Carlo with the Integrated Nested Laplace Approximation. arXiv preprint https://arxiv.org/abs/1701.07844

- Rue H., Martino S. and Chopin N. (2009) Approximate Bayesian Inference for Latent Gaussian Models Using Integrated Nested Laplace Approximations (with discussion). Journal of the Royal Statistical Society, Series B, 71, 319-392

A Bayesian Model for Drug Response Estimation and Biomarker Testing using Gaussian Processes

Dennis Wang, James Hensman, Jonathan Dry, Michael P. Menden and **Frank Dondelinger**

Lancaster University

Large-scale drug response assays on cancer cell lines have in recent years produced a wealth of data that benefits cancer treatment, drug development and biomarker discovery (see e.g. Barretina et al. 2012, Garnett et al. 2012). The raw data from these experiments usually takes the form of a doseresponse curve. The default approach for analysing this data is to fit a sigmoidal curve to the measurements, and extract summary values that reflect the drug response, such as the IC50 and area under the curve.

In this work, we present a novel Bayesian approach for modelling dose-response curves using Gaussian processes. Our model has several advantages over the sigmoid fit: 1) it is non-parametric and allows us to fit a variety of responses; 2) it allows for a hierarchical Bayesian setup with information sharing across different curves; and 3) we automatically obtain a measure of the uncertainty of our curve fits from the variance of the Gaussian process. We extend the model with a Bayesian biomarker testing framework that allows us to test for a difference in the proportion of responsive curves in mutated versus wild type cell lines.

We test the model on cell line drug response data from the Cancer Genome Project (Garnett et al. 2012, Iorio et al. 2016). The data consists of 256 anti-cancer drugs assayed on 1,000 cancer cell lines. While our curve fits are in agreement with sigmoidal fits on the majority of cell lines when comparing summary measures, we demonstrate that the Gaussian process model shows greater robustness to outliers and to unusual response patterns. The Bayesian testing model successfully identifies known biomarkers, and is able to leverage information about the complete dose-response curve, rather than relying on summary measures.

A Bayesian framework to evaluate calibration frequency in an automated immunoassay platform based on stability studies

Laurent Natalis[1], Didier Poirault[2] and Bruno Boulanger[1]

[1] Arlenda SA, Belgium
[2] Biomérieux, France

BioMérieux, a world leader in the field of In Vitro Diagnostics (IVD) solutions, is the owner and vendor of the VIDAS® automated immunoassay platform. Currently, more than 3000 VIDAS systems are in use all over the world to diagnose more than 100 different assays.

Calibration of the VIDAS assay is done by one calibrator sample provided in the kit. Calibration must be performed each time a new lot of reagents is opened and every 14 or 28 days, depending on the assay. This operation provides instrument-specific calibration curves and compensates for possible minor variations in assay signal throughout the shelf-life of the kit. For each new assay under development, the frequency of calibration (i.e. 14 or 28 days), must be verified through a specific procedure.

Knowing that the required frequency of calibration is directly linked to the stability/instability of the reagents used in the assay, there is an opportunity to predict from stability studies the frequency of calibration that will guarantee the performance of the assay throughout the shelf-life of the kit.

Using Bayesian methods, the process to compute doses has been simulated, making it possible to evaluate the optimal frequency of calibrations and to answer the question which is in fact a very "Bayesian" one; what is the calibration frequency required to keep the probability of being outside the specifications below a given threshold, given the signal profiles observed during stability studies? This approach constitutes a new way of thinking for the evaluation of calibration frequencies. Furthermore, Bayesian solutions also offer the possibility of handling degradations following non-linear trends, a problem that can hardly be solved using frequentist approaches.

Assessing treatment effect in rare diseases with a single arm non-randomized study: A patient historic Bayesian Poisson mixed model approach

Arnaud Monseur[1], Bruno Boulanger[1], Marco Munda[1], Joe Parker[2]

[1] Arlenda SA

[2] Prometic Biotherapeutics.

Orphan drugs often cannot resort to randomized and controlled trials for ethical reasons (for example compassionate usage). As such, new techniques need to be developed in order to assess the performance of a new drug without possible comparisons. In this context, the observed diminution of hypoplasminogenic lesions thanks to injections of Plasminogen Intravenous (Human) needs to be assed with innovative techniques.

A Bayesian modeling of each patients' (n=12) historical number of lesions across time is implemented based on a Mixed Poisson Model. This model is than used to simulate predictive intervals at the future assessment dates of treated patients. A comparison of the observed number of lesions and the predictive interval of a non-treated case is realized. The probability to observe the true values based on these intervals are than computed. This will allow to show that the probability that the observed reduction in lesions cannot be due to a normal behavior of the disease. Bayesian designs allows us to computed intervals for these probabilities which a frequentist design would not.

Furthermore, we can assess the expected date at which as certain amount of predicted values will insure a total amount of zero lesions. This date can be compared to the model where a covariate representing treatment and the observed data is added to the model.

This strategy allows an assessment of the treatment effect using a various strategy. Indeed, a 95 % interval on the probability of observing the data is given. If the Upper bound of the latter is sufficiently small, it can be concluded that the observed values after treatment are not due to a normal behavior but rather to an effect of an outside event. In the current context, it is shown that Plasminogen Intravenous proves to be efficient in reducing the number of lesions. Indeed, the predicted probabilities of having the observed values is lower than the selected threshold.