*Fourth Workshop on*

# *BAYESIAN INFERENCE IN STOCHASTIC PROCESSES*

*Villa Monastero, Varenna (LC), Italy*

*June, 2-4, 2005*

*TALKS …*

**Matthew J. Beal and Yee Whye Teh**

*Efficient Sampling Strategies for the Hierarchical Dirichlet Process: with Application to the Infinite Hidden Markov Model and its Variants*

We consider time series data modelling using Hidden Markov Models having an a priori unknown number of hidden states. We show that the Infinite Hidden Markov Model [1] can be recast in the framework of Hierarchical Dirichlet Processes (HDPs). The HDP framework [2] considers problems involving related groups of data: each (fixed) group of data is modelled by a DP mixture model, with the common base measure of the DPs being itself distributed according to a global DP. The base measure being discrete w.p.1 ensures that the group DPs share atoms (despite being countably infinite). [2] presents two sampling schemes for posterior inference in the HDP. We cast sequential data in the grouped data framework by assigning observations to groups, where the groups are indexed by the value of the previous state variable in the sequence; then the current state and its emission distributions define a group-specific mixture model. Thus the hidden state sequence implicitly defines a partition into groups, and induces constraints in the posterior that make the sampling methods proposed in [2] quite inefficient. We present novel effective MCMC methods to overcome this problem in the iHMM and it variants, and present results in text and bioinformatics domains.

[1] Beal, M.J., Ghahramani, Z. and Rasmussen, C.E. (2001). The Infinite Hidden Markov Model. In Advances in Neural Information Processing Systems 14:577-584, eds. T. Dietterich, S. Becker, Z. Ghahramani, MIT Press, 2002.

[2] Teh, Y.W., Jordan, M.I., Beal, M.J. and Blei, D.M. Hierarchical Dirichlet Processes. To appear in Advances in Neural Information Processing Systems 17, MIT Press, 2005. For more details see Technical Report 653, UC Berkeley Statistics, 2004. Also submitted to JASA.

**Alexandros Beskos**

*Exact inference for diffusion processes using MCMC*

We present a new methodology for parametric inference about diffusions given a set of discrete time observations. The method is based on an algorithm for the exact simulation of the stochastic model and involves imputation of the missing paths among the data. The exact simulation algorithm allows for carrying out a simple Gibbs step when sampling from the full conditional of the missing paths without the necessity of any kind of approximation.

**Paul G. Blackwell and K. J. Harris**

*Bayesian radio-tracking: inference for diffusions in heterogeneous environments*

In ecology and zoology, the radio tracking of animals is an important source of data on movement, behaviour and habitat use. Dependence between successive observations on

an individual needs to be allowed for in any statistical analysis; since observations are not necessarily made at equal time intervals, a natural approach is to regard the true underlying movement process as a diffusion. In this talk, we will discuss a range of fully-parametric diffusion models that capture key features of realistic patterns of animal movement. These include the two-dimensional Ornstein-Uhlenbeck process (Dunn and Gipson, 1977, Biometrics 33:85-101), diffusions driven by an underlying Markov chain representing behaviour (Blackwell 2003, Biometrika 90:613-627) and diffusions with different properties in different discrete regions of space (Blackwell and Harris, work in progress); the latter is an extension to two (or more) dimensions of continuous-time threshold auto-regressive processes. We will describe and illustrate techniques for carrying out fully Bayesian inference for these models, using a Markov Chain Monte Carlo approach.

**Paolo Bulla and Pietro Muliere**

*Bayesian Nonparametric Estimation for Reinforced Markov Renewal Processes*

Starting from the definitions and the properties of reinforced renewal processes and reinforced Markov renewal processes, we characterize, via exchangeability and de Finetti's representation theorem, a prior that consists of a family of Dirichlet distributions on the space of Markov transition matrices and beta-Stacy processes on distribution functions. Then, we show that this family is conjugate and give some estimate results.

**Pier Luigi Conti**

*Bayesian nonparametric inference on solutions of renewal equations*

Renewal process play an important role as models of stochastic systems. Generally speaking, many important characteristics associated to renewal processes (e.g., the "standard" renewal function, the distribution function of latest/first renewal before/after time $t$, etc.) are obtained as solution of a general integral equation, the so-called "renewal equation". To be concrete, let $(X_i \ i \geq 1)$ be a sequence of r.v.s ("inter-arrival times") i.i.d. conditionally on their d.f. $F$. A general renewal equation is an integral equation of the form

$$M(t) = m(t) + \int_{-\infty}^{+\infty} M(t-x)dF(x)$$

$m(\cdot)$ being an appropriate function depending on $F$. As a prior law for $F$, we assume first a Dirichlet process. Under appropriate regularity conditions, consistency of the posterior law of $M(\cdot)$ is shown. Then, the weak convergence of the posterior of $M(\cdot)$, when appropriately rescaled, to a Gaussian process is proved. Due to the possible unboundedness of $M(\cdot)$, this point is discussed in some detail. Approximations to the actual posterior law of $M(\cdot)$ are also discussed. Finally, some results are extended to more general prior laws.

**Lehel Csato and Manfred Opper**

*Efficient Gaussian Process Inference*

Gaussian processes (GP) have gained popularity among researchers in the machine learning community. Despite their theoretical simplicity and generality, the implementation of GP inference usually suffers from: (1) intractability of the Bayesian posterior or the predictive distribution for all but the simplest likelihood models and (2) the time required to find the solution is cubic in the size of the data set. In the presentation the proposed a solution to the problem of intractability is a sequential approximation of the non-tractable posterior with a GP. The same approximation allows for an estimation of the – not necessarily Gaussian – predictive distribution. With the approximation technique we estimate the marginal data likelihood or evidence to the data, which in turn can be used to adjust the parameters of the kernel function to the GP. The prohibitive scaling of the computation time is solved by projecting the GP into a low-dimensional subspace, this projection makes the computational time cubic with respect to the size of the basis or anchor set, making the GP inference applicable for large data-sets.

1. Cornford D, Csato L, Evans D, Opper M. (2004): Bayesian analysis of the Scatterometer Wind Retrieval Inverse Problem: Some New Approaches, Journal of the Royal Statistical Society B, Vol. 66, pp. 1–17.

2. Csato L, Opper M (2002): Sparse Gaussian Processes, Neural Computation, 14/3, The MIT Press.

3. Csato L (2002): Gaussian Processes – Iterative sparse approximations, Ph.D thesis, Neural Computing Research Group, Aston University.

4. Csato L (2004): Sparse Gaussian Processes Matlab Toolbox
www.tuebingen.mpg.de/∼csatol/OGP

**Alan Gelfand**

*Multivariate Spatial Process Modelling*

Models for the analysis of multivariate spatial data are receiving increased attention these days. In many applications it will be preferable to work with multivariate spatial processes to specify such models. A critical specification in providing these models is the cross covariance function. Constructive approaches for developing valid cross-covariance functions offer the most practical strategy for doing this. These approaches include separability, kernel convolution or moving average methods, and convolution of covariance functions. We review these approaches but take as our main focus the computationally manageable class referred to as the linear model of coregionalization (LMC). We introduce a fully Bayesian development of the LMC. We offer clarification of the connection between joint and conditional approaches to fitting such models including prior specifications. However, to substantially enhance the usefulness of such modelling we propose the notion of a spatially varying LMC (SVLMC) providing

a very rich class of multivariate nonstationary processes with simple interpretation. We illustrate the use of our proposed SVLMC with application to more than 600 commercial property transactions in three quite different real estate markets, Chicago, Dallas and San Diego. Bivariate nonstationary process models are developed for income from and selling price of the property.

**Simon Godsill and Gary Yang**

*Inference for Gaussian and non-Gaussian continuous-time ARMA processes*

In many physical science and engineering applications a continuous-time stochastic process model is more appropriate than an approximate discrete time model. Here we present results for Bayesian inference in continuous-time ARMA models using Monte Carlo sampling procedures. The target distributions are challenging for MC exploration and we demonstrate effective performance using specially designed proposal functions and annealed samplers, which yield superior performance to simpler approximation methods from the literature. We discuss also how to extend from the Gaussian case to symmetric $\alpha$-stable Levy-driven ARMA processes through use of an augmented model based on the scale mixture representation of the symmetric stable law. This leads to a simulation-exact method for certain cases, while an approximate discretisation is required in the general case.

**Andrew Golightly and Darren J. Wilkinson**

*Bayesian Sequential Inference for Nonlinear Multivariate Diffusions*

We extend recently develop ed simulation-based sequential algorithms to the Bayesian analysis of partially and discretely observed diffusion processes. Typically, since observations arrive at discrete times, yet the model is formulated in continuous time, it is natural to work with the first order Euler approximation. As the interobservation times are usually too large to be used as a time step, it is necessary to augment the observed low-frequency data with the introduction of $m-1$ latent data points in between every pair of observations (Pedersen, 1995). Markov chain Monte Carlo (MCMC) methods can then be used to sample the posterior distribution of the latent data and model parameters. Unfortunately, if the amount of augmentation is large, high dependence between parameters and missing data results in arbitrarily slow rates of convergence of basic algorithms such as Gibbs samplers. We propose a simulation filter, exploiting the diffusion bridge construct of Durham & Gallant (2002), which allows on-line estimation of parameters and state (Liu & West, 2001) and doesn't break down as $m$ increases. We apply the method to the estimation of parameters governing some interesting nonlinear multivariate diffusions.

1. Durham, G. B. & Gallant, R. A. (2002). Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. Journal of Business and

Economic Statistics 20, 279316.
2. Liu, J. & West, M. (2001). Combined parameter and state estimation in simulation-based filtering. In Sequential Monte Carlo Methods in Practice, A. Doucet, N. de Freitas & N. Gordon, eds.

**Daniela Golinelli**

*Bayesian inference in a hidden stochastic two-compartment model for feline hematopoiesis*

In this paper we describe a hidden two-compartment process that has been adopted to model the kinetics of feline hematopoietic stem cells in continuous time. Because of the experimental design and the data collection scheme making inference in such model is extremely difficult. We introduce an RJMCMC algorithm that allows us to obtain an estimate of the posterior distribution of the parameters of interest. We show the performance of the algorithm on both simulated and real data. In particular we apply the introduced algorithm to the case of multiple cats, or multiple realizations.

**Mathieu Kessler, Omiros Papaspiliopoulos and Rui Paulo**

*Objective priors for spatial Gaussian processes*

We address the issue of deriving objective priors for inference based on the observation of a spatial Gaussian process on a finite grid. We consider the reference approach suggested by Bernardo, where an essential step consists in the replication of the experiment to reach perfect estimation asymptotically. We argue that replication of the experiment is not the natural way to reach perfect estimation in this stochastic process context and investigate the effect of considering infield asymptotics instead.

**Athanasios Kottas, Jason Duan and Alan E. Gelfand**

*Bayesian Nonparametric Spatio-temporal Modeling for Disease Incidence Data*

We propose a Bayesian nonparametric approach to modeling disease incidence data, which are typically available as rates or counts for specified regions, and are collected over time. We develop a hierarchical formulation using spatial random effects modeled with a Dirichlet process prior. The Dirichlet process is centered around a normal distribution, which is defined by first assuming a Gaussian process model for the underlying spatial surface, and then using block averaging of this Gaussian process to the areal units determined by the regions in the study. We employ a dynamic formulation for the spatial random effects to extend the model to spatio-temporal settings. Posterior inference is implemented with an efficient Gibbs sampler, which utilizes strategically introduced latent variables. We illustrate the methodology with data on lung cancer incidences for all 88 counties in the state of Ohio over an observation period of 21 years.

**Samuel Kou, Sunney Xie and Jun Liu**

*Bayesian Analysis of Stochastic Models in Single Molecule Biophysics*

Recent technological advances allow scientists for the first time to follow a biochemical process on a single molecule basis, which, unlike traditional macroscopic experiments, raises many challenging data-analysis problems and calls for a sophisticated statistical modeling and inference effort. This paper provides the first likelihood-based analysis of the single-molecule fluorescence lifetime experiment, in which the conformational dynamics of a single DNA hairpin molecule is of interest. The conformational change is modeled as a continuous-time two-state Markov chain, which is not directly observable and has to be inferred from changes in photon emissions from a dye attached to the DNA hairpin molecule. In addition to the hidden Markov structure, the presence of molecular Brownian diffusion further complicates the matter. We show that closed form likelihood function can be obtained and a Metropolis-Hastings algorithm can be applied to compute the posterior distribution of the parameters of interest. The data augmentation technique is utilized to handle both the Brownian diffusion and the issue of model discrimination. Our results increase the estimating resolution by several folds. The success of this analysis indicates there is an urgent need to bring modern statistical techniques to the analysis of data produced by modern technologies.

**Hedibert Lopes and Carlos Carvalho**

*Factor models with time-varying loadings and regime switiching*

In this article we use factor models to describe acertain class of covariance structure for financial time series models. More specifically, we concentrate on situations where the factor variances are modeled by a multivariate stochastic volatility structure. We build on previous work by allowing the factor loadings, in the factor model structure, to have a time-varying structure and to capture changes in asset weights over time motivated by applications with multiple time series of daily exchange rates. The factor loadings time-varying structure as well as the common factor and specific factor volatilities are modeled through Markov switching processes. Posterior inference is performed by designed Markov chain Monte Carlo and sequential Monte Carlo algorithm. We explore and discuss potential extensions to the models exposed here in the prediction area. This discussion leads to open issues on real time implementation and natural model comparisons.

**Ramses H. Mena and Stephen G. Walker**

*Representation of some Markov Models via Predictive Distributions*

Predictive distributions that arise from Bayesian settings constitute a useful tool for modelling dependency in time series analysis. This technique is of particular interest in non-linear and non-Gaussian cases. In this work we discuss an approach to model

stationary processes via predictive distributions. In particular, some developments within the continuous time framework are presented. The underlying approach provides alternative ways to simulate and estimate well-known continuous time processes.

**Philip O'Neill and Nikolaos Demiris**

*Bayesian inference for stochastic epidemics using random graphs* We consider the problem of Bayesian inference for a class of stochastic epidemic models in which a population of individuals is partitioned into groups (such as households). Potentially infectious contacts occur both within groups, and between groups, as governed by Poisson processes of different rates. Given data on final outcome, namely the individuals who are ever infected during an outbreak, we are interested in inferring information about the infection rates. The main obstacle to this task is that the likelihood is intractable, which suggests that data augmentation may be profitable. Here we proceed by using a certain random graph construction that contains information about the eventual outcome, and use realisations of such graphs within a Markov chain Monte Carlo setting to perform the desired inference. The methods are illustrated with data on influenza outbreaks. The methods are very general and extensions will be discussed.

**Andrea Ongaro**

*A new random re-labelling scheme for posterior calculations of symmetric prior processes* In previous work (Ongaro 2003, JSPI) the connection between a certain random relabelling scheme (known as size-biased sampling) and a general class $P$ of processes to be used as nonparametric prior distributions was established. This led to representations of the class particularly suitable for inferential purposes and allowed the derivation of relatively simple results on the posterior and predictive distributions and on the structure of a sample of observations from the class. Here we aim at extending such a framework by considering a more general class of prior processes. In particular, the class $P$ is enlarged so that it essentially coincides with the class of so-called symmetrically distributed random probability measures (Kallenberg, 1986, Random measures). Consequently, we have constructed a new extension of the size-biased relabelling scheme which allows to generalise the results obtained for $P$ to the enlarged class.

**Jesus Palomo and David Dunson**

*Bayesian structural equation models with latent variables*

Structural equation models (SEMs) provide a general framework for modeling of multivariate data, particularly in settings in which measured variables are designed to measure one or more latent variables. In implementing SEM analyses, it is typically assumed that the model structure is known and that the latent variables have normal distributions. To relax these assumptions, this article proposes a semiparametric

Bayesian approach. Categorical latent variables with an unknown number of classes are accommodated using Dirichlet process (DP) priors, while DP mixtures of normals allow continuous latent variables to have unknown distributions. Robustness to the assumed SEM structure is accommodated by choosing mixture priors, which allow uncertainty in the occurrence of certain links within the path diagram. A Gibbs sampling algorithm is developed for posterior computation. The methods are illustrated using biomedical and social science examples.

**Sara Pasquali**

*Bayesian and Classical Inference for a Stochastic Predator-Prey System*

In this work we study the problem of parameter estimation in a discretely observed predator-prey model described by a system of stochastic differential equations. A classical approach based on maximum likelihood function and a Bayesian approach based on a MCMC algorithm are compared. The estimator based upon the ML function converges in mean square to the true value of the parameter, when the final time and the number of observations go to infinity and the interval between two consecutive observations goes to zero. In the Bayesian framework, considering a prior normal distribution for the parameter to be estimated, we obtain a normal posterior distribution whose mean is close to the classical estimator when an improper prior is chosen. Then we introduce latent data between two consecutive actual observations. A MCMC method based on the Metropolis-Hastings algorithm is applied to sample from the posterior distribution of the latent data. The Bayesian approach provides a reasonable posterior distribution for the parameter even when we dispose of a relatively small number of observations. Numerical results are illustrated.

**Havard Rue and Sara Martino**

*Approximative Deterministic Bayesian Inference for Hierarchical Gaussian Markov Random Field Models*

Hierarchical Gaussian Markov random field (GMRF) models are often used in statistical applications, and includes dynamic and temporal models, model-based geostatistical and spatial models, and spatio-temporal models. Most of these have a posterior of the following form,

$$\pi(x, \theta|y) \propto \pi(x|\theta)\pi(\theta) \prod_{i \in Y} \pi(y_i|x_i)$$

where $\theta$ are a low-dimensional hyperparameter of the GMRF $x$, and $y$ are pointwise observations of $\{x_i \ : \ i \in Y\}$. Although Bayesian inference about $(\theta, x)$ using MCMC techniques and preferable block-MCMC, we consider in this talk ways to do approximative deterministic inference for $\theta$ and $x_i$, $i = 1, \ldots, n$, conditioned on $y$. Such approximations can be derived from tractable Gaussian and non-Gaussian approximations of $\pi(x|y, \theta)$ and have several advantages to simulation based inference;

• Lack of (large) Monte Carlo error
• Extremely fast
• No burn-in/convergence problems
The disadvantage is a small bias.

**Alexandra M. Schmidt, Bruno Sansò and Aline A. Nobre**
*Spatio-temporal models based on spatial discrete convolutions*
We consider a class of models for spatio-temporal processes based on convolving spatial independent processes with a discrete kernel that is represented by a lower triangular matrix. This triangular matrix represents the Cholesky decompostion of an auto-regressive process of order $p$. We find that the proposed families of models provide a rich variety of covariance structures. These include covariance functions that are stationary and separable in space and time as well as time dependent non-separable and non-isotropic ones. We also address, under the proposed model, the problem of missing data and the problem of covariates which are measured at different locations from the response variable (spatial misalignment). A real data example, based on measurements of PM10 at the city of Rio de Janeiro, will be presented.

**David Stephens and Matthew Gander**
*Inference for Volatility Models driven by Lévy Processes*
We extend the currently most popular models for the volatility of financial time series, Ornstein-Uhlenbeck stochastic processes, to more general Non-Ornstein Uhlenbeck models. In particular, we investigate means of making the correlation structure in the volatility process more flexible. We implement a method for introducing quasi long-memory into the volatility model without recourse to superposition We demonstrate that the models can be fitted to real share price returns data, and that results indicate that for the series we study, the long-memory aspect of the model is not supported.

**Osnat Stramer and Jun Yan**
*Parametric inference for partially observed diffusion processes: a comparison study*
Diffusion models described by stochastic differential equations are used extensively in many areas of science-engineering, hydrology, financial economics and physics. While the models are formulated in continuous-time, the data are recorded at discrete points in time. For most diffusion models the likelihood function is unavailable. We describe four alternative approaches to inference of diffusion models.
• Bayesian inference using Gibbs sampling and data augmentation (Elerian at al., 2001; Roberts & Stramer, 2001).
• Simulated maximum likelihood estimation (Pedersen, 1995; Durham & Gallant, 2002).
• Analytical approximation of the likelihood (Ait Sahalia, 2002 and Ait Sahalia, 2005).

• Efficient method of moments (Gallant & Long, 1997).

All the above methods are computationally heavy. We review all four methods in terms of

• what models can each method handle,
• theoretical justification of the method,
• accuracy,
• speed.

We perform a set of Monte Carlo experiments to compare the performance of these approaches on two different type of models:

• Simple models like the square-root/Cox-Ingersoll-Ross models that can be solved explicitly and hence can be estimated via maximum likelihood.
• Models like the preferred model for short termi interest rates of Ait-Sahalia (1996) that cannot be solved explicitly and hence cannot be estimated via maximum likelihood.

In addition, we also provide a numerical technique to improve the computational efficiency of the simulated maximum likelihood method proposed by Durham & Gallant (2002).

**Jonathan Stroud, Michael Johannes and Nicholas Polson**

*Sequential Parameter Estimation in Stochastic Volatility Models with Jumps*

This paper analyzes the sequential learning problem for both parameters and states in stochastic volatility models with jumps. We describe the existing methods, the particle and practical filter, and then extend these algorithms to incorporate jumps. We analyze the performance of both approaches using both simulated and S&P 500 index return data. On both types of data, we find that both algorithms are effective in sequential learning of the jump parameters, although sensitivity analysis indicates that the practical filter performs marginally better. These conclusions are similar to those in Stroud, Polson and Muller (2004) regarding stochastic volatility models.

**Elisa Varini and Renata Rotondi**

*A sequential particle filter for a continuous-time state-space model*

A new state-space model for the analysis of earthquake sequences is proposed: a pure jump Markov process with observations from marked point processes. We considered three marked point processes for seismic sequences, each of them having a different physical foundation: the Poisson process, the stress release model and the Epidemic-Type Aftershock-Sequence model. We assumed that a time-magnitude seismic sequence is composed by a series of realizations of these three models (the observed process). The dynamic of their activation times is driven by an hidden pure jump Markov process (the state process). The inference of a so complex and rich model is carried out by

exploiting a Bayesian sequential Monte Carlo method in order to estimate the model parameters and to approximate the filtering distribution. We examined a simulated data set focusing our attention on the properties of the estimation methods proposed and on their sensitivity to the assessment of some particular parameters. When the model parameters are known, a solution of the filtering problem is proposed by exploiting the innovation method in the context of the martingale representation of a point process. This method can be also used to improve the approximation of the filtering distribution obtained by the sequential Monte Carlo procedure when the model parameters are unknown.

**Krisztina Vasas, Peter Elek and Laszlo Markus**

*A regime switching time series model of daily river flows*

We report on fitting a two-state regime switching autoregressive model for daily water discharge series registered at monitoring sites of River Tisza and of its tributaries. One peculiarity of the model is that the noise sequence switches distribution according to the exponential and normal law, the rising regime being governed by the exponential part. In the case of Markovian regime switching, the estimation can be carried out by a simple implementation of the MCMC algorithm. However, as change times of the regimes in hydrological series are known to deviate from the geometric distribution, Markov-modelling is not entirely satisfactory. When generalisations of the model are considered, more sophisticated estimation algorithms should be chosen. In some particular non-Markovian cases, either the efficient method of moments algorithm or reversible jump MCMC may be of help. In the talk we also analyse the joint behaviour of the main river and of its tributaries, by examining the pattern of regimes and the multivariate distribution of the noise sequences at different sites.

**Mike West**

*Stochastic Search & MCMC on "Big" Model Spaces*

I will present and discuss approaches to stochastic computation in "big" and "sparse" models - large, sparse graphical models, regression variable models with many candidate predictors, and large-scale, sparse factor models. Varieties of MCMC methods and non-MCMC evolutionary stochastic search methods will be discussed, compared and exemplified. Questions arise about convergence characteristics of the resulting stochastic processes - whether MCMC or stochastic optimisation is the goal - questions that challenge Bayesian probabilists and algorithm developers alike. Critical questions of Bayesian model/prior specification - issues that are simply central to scalability of Bayesian technologies as models/parameter spaces increase in dimension - will also be discussed.

**Darren J. Wilkinson**

*Bayesian Inference for Biochemical Network Dynamics*

This talk will give an overview of one of the key problems in the new science of Systems Biology - inference for the rate parameters underlying complex stochastic kinetic biochemical network models, using partial, discrete and noisy time course measurements of the system state. The basic problem will be introduced, highlighting the importance of stochastic modelling for effective estimation, and then a range of approaches to Bayesian inference will be reviewed and compared. Some approaches recognise the discrete nature of the underlying molecular dynamics, whilst others use a diffusion approximation to give a non-linear multivariate stochastic differential equation representation.

*and POSTERS*

**Carlos Almeida and Michel Mouchart**

*Bayesian encompassing test under partial observability: General Theory (with application to binary observability)*

In the framework of Partial observability, a Bayesian encompassing procedure is proposed in order to compare a parametric model against a (non-parametric) alternative; this paper considers the partial observability process completely known. The general procedure is illustrated by the case where only the sign is observable, and more generally when the available data come from a binary reduction of a vector of latent variables.

**Maria Concepcion Ausin and Hedibert Lopes**

*Large claims approximation in insurance risk processes*

Large claims can have a dramatic effect in the analysis of insurance risk processes. For example, the ruin probability for a given initial capital can be underestimated if an inappropriate model is considered for the claim size distribution. Frequently in the literature, large claims are fitted with long-tailed distribution (i.e. with tails that decay more slowly than exponentially) such as the Pareto and the Weibull. However, risk models with long-tailed claim sizes tend to be difficult to analyze and there are not explicit expressions for the ruin probabilities that are often approximated with simulation methods. Alternatively, in this work, we consider flexible models for the claim sizes based on mixtures of distributions. We develop Bayesian inference for a reparametrization of mixtures of Erlang distributions including a non informative prior, which allows us to approximate long-tailed distributions by increasing the number of components in the mixture. Explicit evaluations of the ruin probabilities are possible as the Erlang mixture belong to the set of distributions of phase type. We show how to estimate the ruin probability in different risk reserve processes such as the classical compound Poisson risk process and the Sparre Anderse process. We illustrate this approach with simulated and real data.

**Paloma Botella, M. Cambra and Antonio Lopez Quilez**

*Bayesian Inference on a Cox Model for Epidemiology of Plant Viruses*

Cox point process, or doubly stochastic, is a hierarchical model whose events appear by means of a random intensity function. It provides high flexibility in modelling complex phenomena, though inference can be difficult if probabilities in both levels are not parametrically related. In this work we carry out the spatial analysis of the infected trees by Sharka (Plum Pox Virus) in a parcel of peach trees. Geographic coordinates of the 53 trees that became infected in a 8900 trees parcel can be considered as a point pattern. The infection of these trees could come from a contiguous parcel, in which after detecting the virus the plantation was eliminated. This hypothesis is

included through an inhomogeneous point process modelling, based on the distance of the infected trees to the previously infected parcel. Wind direction is incorporated in the analysis implying a random intensity of the infection process. A Cox model is used to model this problem, with a first level that would gather the possible directions of infection and a second level that would relate the infection to the distance to the contiguous parcel in each direction. The analysis of the model has been made by means of the WinBUGS program.

**Javier Cano, David Rios Insua and Carla Tello**

*BRAMS. A system for Bayesian Reliability, Availability and Maintenance Support*

We describe BRAMS a system for Bayesian Reliability, Availability and Maintenance Support that we are developing with the software engineering industry in mind. The system includes modules referring to RBDs, CMTCs and Sofwtare reliability growth models, interlinked. Several examples will be shown.

**Roberto Casarin**

*Inference on diffusion processes by population Monte Carlo method*

Diffusion processes are widely used in many fields like physics, engineering, biology, economics and finance. In finance they allow to describe the evolution over time of many financial quantities, like interest rates, asset returns and volatility. The dynamic of a diffusion process can be described by means of a stochastic differential equation (SDE), which could depend on some usually unknown parameters and unobserved components.

In the literature many estimation approaches have been proposed and simulation based methods result particularly useful when dealing with nonlinear SDE. Elerian, Chib and Shephard [5] apply MCMC method to likelihood inference for nonlinear diffusion processes following the high frequency augmentation (HFA) method which has been independently proposed by Jones [13] and Eraker [6]. Billio, Monfort and Robert [1], through an example, apply simulated likelihood ratio method to diffusion processes. Johannes and Polson [11] give an updated review of the MCMC methods for inference on diffusion processes also with partially observed components, with application to finance. Within the simulation based inference framework the Bayesian approach has been widely applied in many recent studies, due the natural way the Monte Carlo approximation can enter in the inference procedure. Moreover the Bayesian framework allows to account also for nonstationarity of the model and for prior information about the parameters. The need for on-line data processing and for more efficient simulation techniques gave rise in the last years to the use of sequential importance sampling methods and of adaptive MCMC methods. See Doucet, de Freitas and Gordon [4] and Robert and Casella [17] for an introduction to general simulation methods and see for

example Golightly and Wilkinson [7] and Johannes, Polson and Stroud [12] for recent advances on diffusion process inference.

This work deals with Bayesian simulation based inference for SDE, when observations from the continuous time process are available at regularly spaced discrete time points. In order to make inference on the diffusion process we apply the Euler-Maruyama method and obtain a discretisation of the SDE. Furthermore, in the Bayesian approach parameters are random quantities and once the prior distribution has been chosen, the Bayesian estimator is a function of the posterior distribution. For complex models the estimator is not available in a closed-form, thus a numerical approximation is needed.

The first aim of the work is to propose the iterated importance sampling method (see Guillin, Marin and Robert [8]) as efficient alternative to MCMC method for SDE parameter estimation. In particular we show how the iterative importance sampling method, also called Population Monte Carlo (PMC) method, due to Celeux et al. [3], Cappe et al. [2], can improve Bayesian parameter estimation and reduce the computational effort.

The second aim is to compare basic and weighted PMC estimators and to study the optimal size of the particle sample at each iteration of the importance sampling algorithm.

Finally we combine the PMC with the HFA method and provide some simulation results for inference on the geometrical Brownian motion and constant elasticity volatility process (see Hull and White [9] for a financial application), with particular attention to the Ornstein-Uhlenbeck and square root mean reverting processes.

[1] M. Billio, A. Monfort and C.P. Robert. The Simulated Likelihood Ratio Method, Doc. du Travail Crest, INSEE, Paris, 1998.

[2] O. Cappe, A. Guillin, J.M. Marin and C.P. Robert. Population Monte Carlo, J. Computational and Graphical Statistics, to appear, 2002.

[3] G. Celeux and J.M. Marin and C.P. Robert. Iterated importance sampling in missing data problems, Cahiers du Ceremade, N. 0326, Univ. Paris Dauphine, 2003.

[4] A. Doucet, N. de Freitas and N.J. Gordon. Sequential Monte Carlo Methods in Practice, Springer-Verlag, New York, 2001.

[5] O. Elerian, S. Chib and N. Shephard. Likelihood inference for discretely observed nonlinear diffusions, Econometrika, 69(4), 959-993, 2001.

[6] B. Eraker. Markov Chain Monte Carlo Analsyis of Diffusion Models with Applications to Finance, Journal of Business and Economics Statistics, 19, 177-191, 2001.

[7] A. Golightly and D.J. Wilkinson. Bayesian Sequential Inference for Nonlinear Multivariate Diffusions, Working paper, University of Newcastle, 2004.

[8] A. Guillin, J.M. Marin and C.P. Robert. Estimation bayesienne approximative par

echantillonnage preferentiel, Revue Statist. Appliquée, to appear, 2004.
[9] J. Hull and A. White. The pricing of Options on Assets with Stochastic Volatility, Journal of Finance, Vol. 42, pp. 281-300, 1987.
[10] Y. Iba. Population-based Monte Carlo algorithms, Trans. Japanese Soc. Artificial Intel l., 16(2):279-286, 2002.
[11] M. Johannes and N. Polson. MCMC Methods for Financial Econometrics. To appear in Handbook of Financial Econometrics, Y. Ait-Sahalia and L. Hansen eds, 2004.
[12] M. Johannes, N. Polson and J.R. Stroud. Nonlinear Filtering of Stochastic Differential Equations with Jumps, Working Paper, 2004.
[13] C.S. Jones. Bayesian Estimation of Continuous Time Finance Models, Working Paper, Rochester University, 1998.
[14] I. Karatzas and S.E. Shreve. Brownian Motion and Stochatic Calculus, 2nd Ed., Springer Verlag, New York, 1991.
[15] P.E. Kloeden and E. Platen. Numerical Solution of Stochastic Differential Equations, Springer Verlag, Berlin Heidelberg, 1992.
[16] D. Revuz and M. Yor. Continuous Martingales and Brownian Motion, 3rd Ed., Springer Verlag, Berlin Heidelberg, 1999.
[17] C.P. Robert and G. Casella. Monte Carlo Statistical Methods, 2nd Ed., Springer Verlag, New York, 2004.
[18] M. Tanner and W. Wong. The calculation of posterior distributions by data augmentation. Journal of the American Statistical Association, 82, 528-550, 1987.

**Maria Eugenia Castellanos, Javier Morales, Roland Fried and Carmen Armero**

*A Bayesian walk through the Machine Interference Problem*

A The machine interference problem is the most popular queueing model with a finite population of customers. Because most applications of this model occur in management and industrial processes, its description is usually presented as a system with working machines (the customers) and operators (the servers). When a machine breaks down and there is at least one operator idle the machine is repaired immediately. The machine needs to wait for service otherwise. Basically, repair times and machine life times are assumed to be independent and exponentially distributed. The Bayesian analysis of the model is oriented towards evaluation of staffing levels and system efficiency. This includes prediction of general congestion measures of the system in equilibrium, such as the number of machines which are down or the time necessary to put an out of order machine into operation again, and more specific measures like the net production rate and the machine availability.

**Aart F. de Vos and Marc K. Francke**

*Marginal Likelihood, Jeffreys' Rule and Unit Root Tests*

In inference on the covariance matrix of the general linear model, the regression and scale parameters are nuisance. Classical marginal likelihood is known as a good way to get rid of these parameters. We argue that the same likelihood can also be used in Bayesian inference, directly or by application of the independence Jeffreys' prior. The latter differs from Jeffreys' rule, while reference priors are ambiguous. In case only one parameter describes the covariance structure, classical tests based on marginal likelihood ratio are uniformly most powerful invariant (UMPI) if this ratio is a monotone function of some statistic. For this case we show that model choice based upon Bayes Factors corresponds to a classical test with a size determined by prior considerations. The data are used in the same way. Our example is the much discussed unit root model where we obtained tests that are almost UMPI using classical marginal likelihood, while Bayesian inference based upon Jeffreys' rule fails. Essentially the same test may be obtained from a Bayesian analysis using independence priors. The correspondence between its size and the prior however is complex, specifically in relation to the sample size; it is studied for the AR(1)-model.

**Vanja Dukic, B. Elderd and G. Dwyer**

*A Bayesian SEIR approach to modeling smallpox epidemics*

Much of the recent US public-policy debate regarding smallpox vaccination has focused on mass versus trace vaccination strategies; namely, whether the public can be better protected by vaccination of the entire population or of only those who have been in contact with infected individuals. Much of the previous work on smallpox epidemics has generally employed relatively complex deterministic models, with many biological parameters fixed, and focusing mostly on a single point estimate of the disease reproductive rate (the number of newly infected individuals arising from a single infected individual). We present a Bayesian analysis of the multiple historical smallpox epidemics that yields an estimate of the distribution of the disease reproductive rate, taking into account the uncertainty of all other parameters in the model, as well as population and geographical heterogeneity. We then present a comparison of the two vaccination strategies based on the posterior predictive distribution of the outcomes under several scenarios.

**Marc K. Francke and Aart F. de Vos**

*Marginal Likelihood in State Space Models* In this paper we examine different formulations of the likelihood in the general linear model from Bayesian and classical viewpoints. Starting point is the diffuse Kalman filter. This is an essentially Bayesian "trick" to compute estimates of parameters and the likelihood function running a fil-

ter. It is specifically useful to cope with diffuse initial conditions, or as Bayesians call it, improper priors. There is a problem however in the resulting formula, caused by the fact that a likelihood is degenerate in such a case. This makes one question the validity of the derivation. Indeed there appears to be a snag, but for many purposes not a serious one. But there are examples where the diffuse likelihood differs among reformulations of the same model. The solution is to use the marginal likelihood. The classical concept goes back to Harville, 1974, the Bayesian equivalent results by using Jeffreys Independence priors. We give the simple adjustments needed in the diffuse Kalman filter to get the marginal likelihood and discuss applications and problems with nonlinear models.

**Pilar Gargallo and Manuel Salvador**
*Monitoring for Models of the Exponential Family. A Bayesian Decision Approach*
This paper proposes an algorithm for monitoring two-sided changes in mean of series from the exponential family. The adopted approach is bayesian and uses a loss function which take into account the run lengh extending the monitoring scheme proposed in Harrison and Veerapen (1994). The algorithm is illustrated with several examples in which the power of the monitor is analized.

**Alessandra Guglielmi, Ilenia Epifani and Eugenio Melilli**
*Distributional results for functionals of Dirichlet processes*
A fundamental problem in a nonparametric Bayesian framework is the computation of the laws of functionals of random probability measures. For instance, in this context, inference about the variance of the frequency distribution of a characteristic in a population requires the knowledge of its posterior distribution. The aim of this work is to show some new results concerning the law of the functional variance $V$ of a Dirichlet process $P$, which can be considered quite general to some extent. For instance, we will see that the law of the variance of Dirichlet mixtures of location-invariant densities is a mixture of distributions of random Dirichlet variances. We establish a simple distributional relationship between $V$ and the random variable $(X-Y)^2$, where $X$ and $Y$ are independent copies of the random mean of the Dirichlet process $P$. Useful expressions for some integral transforms of $V$ are also obtained and illustrative examples are given. Finally, some classes of moment-based distribution approximations will be reviewed and an application to the posterior law of $V$ will be illustrated.

**Konstantinos Kalogeropoulos, Gareth O. Roberts and Petros Dellaportas**
*Bayesian Inference for Stochastic Volatility Diffusion Models Using Time Change Transformations*
This paper presents a new Markov chain Monte Carlo approach for diffusion driven stochastic volatility models. The algorithm is based on a data augmentation scheme

where we treat both the paths of the unobserved volatility and the paths between any two observed points as missing data. We use a time change transformation to break down the dependence between the diffusion parameters and the transformed paths. We then present an appropriate Markov chain Monte Carlo algorithm to sample from the posterior of the diffusion paths and the parameters of the stochastic volatility model.

**Theodore Kypraios and Gareth O. Roberts**

*A semi-Parametric Time Series Model Based on Latent Branching Trees*

A branching process is used to introduce a class of semi-parametric time series models. We describe in detail how one can construct such a branching process and we list the model's properties. Markov Chain Monte Carlo methods are adopted to make inference for the parameters. Furthermore, Non-Centered and partially Non-Centered parameterisations are also presented in order to improve the mixing of the Markov chain. This is work is based on Neal (2002, Bayesian Statistics, 7).

**Ales Linka and Petr Volf**

*Segmentation of texture with local MRF models*

The contribution deals with the analysis of textures and their MRF models. It is assumed that in different parts of texture the parameters of model may differ. The objective is to find these areas (and model parameters), i.e. to segment the texture with respect to local models. In the Bayes scheme, for given number of different models ($K$), we combine the priors of parameters, the Potts MRF of local values of parameters (it supports the connectivity of areas with the same models) and pseudo-likelihood of data, given the configuration of models. Solution is then obtained from the MH algorithm, parameter of Potts model is used as a tuning parameter. However, as we search for an optimal configuration of models (not for representation of posterior), the search may be simplified and solved with the aid of simulated annealing. The same concerns the case with unknown $K$: Instead of full RJ MH procedure, it suffices to compare optimal results for different $K$, penalized by corresponding priors. The application deals with the detection of nonhomogeneous spots (e.g. of poor quality) in digitized image of non-woven textile material.

**Antonio Lopez-Quilez, Carmen Armero and Rut Lopez-Sanchez**

*Bayesian Analysis of the Time to Diagnosis after an Abnormal Screening Mammography*

Breast cancer is one of the diseases with more impact on health in women. Its prediction basically depends on the extension of the damage at the detection time. Consequently, an early diagnosis is the better way to improve the possibilities of a recovery. In the Pais Valencia, one of the Spanish autonomous regions that has taken on the management of all health services, the Breast Screening Programme has been designed to

detect unsuspected cancer in healthy women. This study focus attention on women with an abnormal screening mammography taking part in this programme during year 2002. Specifically, we are interested in comparing the length of the times to diagnosis in connection with the different medical procedures performed, main women risk factors and hospitals taking part in the program. Generalized linear mixed models are used in order to incorporate random effects characterizing heterogeneity among hospitals. Bayesian reasoning and Markov Chain Monte Carlo simulation are considered for estimation and inference.

**László Márkus and Miklos Arato**

*A Inhomogeneous Spatial Poisson Process for Modelling Car Insurance Data*

The results of the analysis of spatial dependence structure for claims occuring in third party liability car insurance in Hungary is going to be reported. An inhomogeneous spatial Poisson process is fitted to the data. Presently the following model fit is obtained for 168 subregions of the country by an MCMC approach. The conditional intensity of the Poisson process is $\lambda \exp(\theta_i t_i)$ with $\lambda$ being the common claim frequency, $t_i$ the overall time-at-risk of all policyholders at the $i$-th region, and $\theta_i$ the local risk factor. The vector of all local risk factors $\Theta = (\theta_i,\ i = 1, \ldots, 168)$ is supposed to be distributed normally, with 0 mean and a variance matrix $\Sigma$ of two parameters $\tau^2$, $\rho$:

$$\Sigma = \frac{1}{\tau^2}(I - \rho A)^{-1}$$

where A is the neighbourhood matrix of zeros and ones, with 1 at the $(i, j)$ position indicating the $i$-th and $j$-th subregions to be neighbours. Supposing $\tau^2$ and $\rho/\rho_{max}$ to be distributed respectively as $\Gamma(\alpha, \beta)$, and $Beta(p, q)$ sufficiently adequate fit can be achieved even by a simple Metropolis-Hastings algorithm. However, when risks for all the 3112 settlements of Hungary are considered separately the algorithm needs unacceptedly long time to stationarise. Also, the fit is intended to be carried out simultaneously for other risk parameters such as sex, age and car type. The dependence of the solution on priors is also a problem to be addressed.

**Miguel Ángel Martínez Beneito, Oscar Zurriaga Llorens, Hermelinda Vanaclocha Luna, Antonio López Quílez, Carmen Armero, David Conesa, Juan José Abellán Andrés and Jordi Pérez Panadés**

*Parametric and Semi-Parametric Approaches to Bayesian Survival Analysis with Spatial Term*

Bayesian inference allows to incorporate a great variety of information and structures when modelling survival times of a given specific population. In particular, recent applications in this topic introduce spatial information about patients at different levels of aggregation. In general, parametric and semi-parametric approaches are the

two main ways to include spatial information in a survival study. In this work, both techniques are used to evaluate sojourn times of patients with renal diseases from the Comunidad Valenciana Renal Registry. Covariates as age, sex, treatment or period are included, and geographical residence is also incorporated in the analysis. Discussion about advantages and limitations of each one of these approaches are presented. In particular, semi-parametric procedure has the advantage of considering dynamic covariates properly.

**William J. McCausland**

*Time Reversibility of Stationary Regular Finite State Markov Chains*

We propose an alternate parameterization of stationary regular finite-state Markov chains, and a decomposition of the parameter into time reversible and time irreversible parts. We demonstrate some useful properties of the decomposition, and propose an index for a certain type of time irreversibility. Two empirical examples illustrate the use of the proposed parameter, decomposition and index. One involves observed states; the other, latent states.

**Loukia Meligkotsidou and Paul Fearnhead**

*Ancestral inference in population genetics via importance sampling*

Population genetics is concerned with analysing a sample of DNA sequences taken from individuals of a population. A genealogical tree can be constructed to represent the ancestry of the sample. Coalescence or branching events in the tree correspond to individuals in the genealogy sharing a common ancestor. The mutation model assumes that mutations occur as events in a Poisson process, with each new mutation occurring at a site that hasn't experienced mutation before. Ancestral inference in population genetics involves estimating the branching times in genealogical trees. Under the natural coalescent prior on the branch lengths the posterior distribution of the coalescence times is difficult to handle. We propose using a non-informative phylogenetic prior on the branch lengths and following the approach of Fearnhead and Meligkotsidou (2004) to approximate the posterior distribution of coalescence times. Then we correct for the true posterior distribution via Importance Sampling. We show how this approach can be applied to deal with a class of coalescent models.

**Alicia Quiros Carretero, Raquel Montes and Juan Antonio Fernandez**

*Bayesian Inference in functional Magnetic Resonance Imaging*

Functional magnetic resonance imaging (FMRI) is a technique that creates images of a subject\222s brain that are sensitive to changes in blood oxygenation, caused by neural activation. We are interested in identifying those brain activity regions by observing the differences in blood magnetisation due to the haemodinamic response. The data sets produced by an fMRI experiment are typically very large and therefore

computationally expensive. A typical output from an fMRI experiment might consist of voxel data points, modelled by a general linear model. t-test are usually performed to look for significantly activated areas, adjusted to allow for multiple comparisons. Using a Bayesian framework provide us with the ability to probabilistically incorporate prior information by modelling the pattern as a Markov Random Field (MRF). To correctly infer on parameters of interest, for example, the haemodynamic response function parameters is not an easy task giving the complexity of the model. Monte Carlo Markov Chain (MCMC) methods enable sampling from the posterior distributions of interest to give estimates of the summary statistics.

**Diego Salmeron and Juan Antonio Cano**

*Regularized particle filter for Bayesian estimation in unobserved diffusion processes*

In this work we study the goodness of an approximation based on Euler schemes for Bayesian estimation of the parameters appearing in a stochastic differential equation. We consider statistical models in which the observations are related with an unobserved diffusion in a nonlinear way. Although the Euler approximation jointly with a Gibbs sampler and a Metropolis-Hastings algorithm ia a good methodology for Bayesian inference in this model, it is known that the rate of convergence of the Gibbs sampler can be arbitrarily slow if the amount of the augmentation obtained by the Euler discretization, is large. In this paper a simple method for sequential estimation is proposed, based on simulation and regularization, without the use of the Gibbs sampler.

**Simo Särkkä, Aki Vehtari and Jouko Lampinen**

*Rao-Blackwellized Particle Filter for Tracking Unknown Number of Targets in Clutter*

We present a new Rao-Blackwellized particle filtering based algorithm for tracking an unknown number of targets in clutter. The algorithm is based on formulating probabilistic stochastic process models for target states, data associations, and birth and death processes. The tracking of these stochastic processes is implemented using sequential Monte Carlo sampling or particle filtering, and the efficiency of the Monte Carlo sampling is improved by using Rao-Blackwellization

**Chris Sherlock and Paul Fearnhead**

*A Gibbs Sampler for the Markov Modulated Poisson Process*

We consider the general Markov Modulated Poisson Process (MMPP) where only the Poisson Process is observed, and describe a Gibbs sampler that first samples from the exact conditional distribution of the (hidden) Markov chain. Bayes Factors can be calculated from the Gibbs sampler output, which enables model choice to be performed. The technique is applied to occurences of the Chi site in the E.coli genome. We also examine the mixing properties of various MCMC algorithms over a selection of 2-dimensional MMPP's.

**Fabio Spizzichino and Rachele Foschi**

*An Invariance Property of Spatial Mixed Poisson Processes*

It is a very well known result in queueing theory that the output of an infinite server queue $M_t/G/\infty$ (with a Poisson arrival process, then) is Poisson as well. Such a closure-type result can be extended in a number of different directions.

Most natural types of extensions can be obtained as follows:

i) by replacing Poisson arrival processes with Mixed Poisson arrival processes

ii) by replacing Poisson arrival processes with M-Poisson processes on the line

iii) by replacing the $M_t/G/\infty$ model with a more general transformation, where each jump $S_i$ in the departure process is obtained by combining an arrival $T_i$ with a random variable $Z_i$:

$$Z_i = \psi(T_i, Z_i)$$

(being $\psi(T_i, Z_i) = T_i + Z_i$, for the specific scheme of an infinite server queue). $Z_1, Z_2, \ldots$ are assumed to be i.i.d.

The extension in ii) is preliminary, and turns out to be basilar, for the extension in iii) (see [1]). The latter case can, in its turn, be extended to more general models with spatial Poisson arrival processes.

Substantially such closure-type results are based on the "Order Statistics Property" (OSP) of the arrival process (see [2] and [3], for the case in i) and [1] for ii) and iii)).

Also the "$p$-thinning Property" of the arrival process (see e.g. [4]) has a fundamental role in this context (see also [5] for some remarks in this concern).

In the talk, we consider models characterized by generic transformations $\psi$ of "spatial-mixed Poisson" arrival processes and present a more general (closure-type) result, that can be obtained in terms of the Order Statistics and p-thinning properties.

Possibly, we also aim to discuss some aspects in the Bayesian estimation of the unobservable parameter $M$ of the spatial-mixed Poisson process $N$ ($N$ being a conditionally spatial Poisson process, given $M$).

[1] Brown M. (1969). J. Appl. Probability, 6, 453-458.

[2] Huang, W. and Puri, P. (1990). Sankhya, Ser. A, 52, 232-243.

[3] Berg M. and Spizzichino F. (2000). Math. Methods of Oper. Res., 51, 301-314.

[4] Grandell J. (1997). Chapman & Hall, London.

[5] Foschi R. (2005). Tesi di Laurea in Matematica. Universita' "La Sapienza".

**Osnat Stramer and Jun Yan**

*Parametric inference for partially observed diffusion processes: a comparison study*

Diffusion models described by stochastic differential equations are used extensively in many areas of science-engineering, hydrology, financial economics and physics. While the models are formulated in continuous-time, the data are recorded at discrete points

in time. For most diffusion models the likelihood function is unavailable. We describe four alternative approaches to inference of diffusion models.
• Bayesian inference using Gibbs sampling and data augmentation (Elerian at al., 2001; Roberts & Stramer, 2001).
• Simulated maximum likelihood estimation (Pedersen, 1995; Durham & Gallant, 2002).
• Analytical approximation of the likelihood (Ait Sahalia, 2002 and Ait Sahalia, 2005).
• Efficient method of moments (Gallant & Long, 1997)
All the above methods are computationally heavy. We review all four methods in terms of
• what models can each method handle,
• theoretical justification of the method,
• accuracy,
• speed.
We perform a set of Monte Carlo experiments to compare the performance of these approaches on two different type of models:
• Simple models like the square-root/Cox-Ingersoll-Ross models that can be solved explicitly and hence can be estimated via maximum likelihood.
• Models like the preferred model for short term interest rates of Ait-Sahalia (1996) that cannot be solved explicitly and hence cannot be estimated via maximum likelihood.
In addition, we also provide a numerical technique to improve the computational efficiency of the simulated maximum likelihood method proposed by Durham & Gallant (2002).

**Paola Tardelli**
*Dynamical estimate for lifetimes of particles belonging to a hetereogeneous population*
A heterogeneous population of identical particles, divided into a finite number of classes, according to their level of health, is considered. The partition can change during the time and is not observed. We just observe the cardinality of a particular class. Our aim is to find the conditional distribution of particles lifetimes, given such observation. To this end, we use a Bayesian dynamical approach that is stochastic filtering techniques.

**Stefano F. Tonellato**
*Random field priors for spectral density functions*
In this paper we discuss how a Gaussian random field with Matern covariance function can represent prior uncertainty about the log-spectral density, $g(\omega)$, of a stationary time series. Hyperparameters can be suitably tuned in order to determine the mean square

differentiability and the range of autocorrelation of the random field $g(\omega)$. However, Bayesian computations cannot be easily performed under such prior elicitation. We suggest therefore to approximate the Gaussian random field priors with a class of Gaussian Markov random fields which preserve most of the smoothness properties of the genuine prior distributions. Such approximation allows us to implement MCMC methods efficiently. Applications to simulated and real data will be shown.

**Matilde Trevisani and Nicola Torelli**

*Spatial misalignment modeling for small area estimation problems*

In this paper we specify, within a hierarchical Bayesian setting, appropriate atom-based models to solve the following small area estimation (SAE) questions: (i) combining auxiliary covariates which are available on non nested areal partitions (misaligned areal regression problem); (ii) providing small area estimates by using planned domains data (misaligned areal interpolation problem). To illustrate our approach we consider the problem of estimating the number of unemployed at Local Labour Market area (small area or target zone) level by using two misaligned source data: auxiliary information available on different administrative partitions; reliable estimates of unemployed on Labour Force Survey planned domains. Thus we explore the close connection that typical SAE issues show to have with spatial misalignment problems. Ob ject of SAE is, in fact, inference on survey non-planned "minor domains" (the so called small areas): based on direct domain data (when available), it leads to estimates of poor quality. Thereby models are set up for borrowing strength from indirectly related data sources. Similarly, spatial misalignment models are set up whenever "target zones" for which data are needed are different from source zones on which data are available.

**Mark Van Lokeren and F. Thomas Bruss**

*Optimal Stopping for Sequences of Random Variables with Unknown Success Parameter*

Let $X_1, X_2, \ldots, X_n$ be a finite sequence of identically distributed random variables taking values in $\{0, 1\}$. If $X_k = 1$ for some $1 \leq k \leq n$, we say a success is observed at time $k$. We assume the distribution of $X_1$ to be a Bernoulli law with unknown success parameter p. Observing the random variables $X_1, X_2, ..., X_n$ sequentially, we want to find a strategy which maximizes the probability of stopping at the last success (if any) in the sequence. No recall on preceeding successes is allowed. We study the case where the prior distribution of $p$ is taken as a Beta law. Using a Bayesian approach to update the prior distribution each time a random variable is observed, we determine an explicit form of the optimal strategy. The strategy relies on a recurrence relation, which also gives the optimal value of the maximum probability of stopping at the last success. The result is a generalisation of [1]. Finally we discuss some applications to clinical trials and sequential search problems.

[1] F.T. Bruss, Sum the odds to one and stop, Ann. Prob., 28(3) 1384-1391, 2000.

**Petr Volf**

*Cluster analysis of time series of Poisson counts*

The paper deals with the statistical analysis of aggregated unemployment data modeled as a time series of Poisson counts. Cluster analysis is employed for selection of sub-populations with similar development of unemployment in Czech Republic in recent years of economic transformation. It is shown how the clusters are connected with certain characteristics (covariates) of these sub-populations. We start from the model-based clustering yielding simultaneously the set of models for distinct clusters and detecting outlied sequences. This methodology still suffers one problem, namely the assessing the number of clusters, though a set of more-less ad-hoc criteria have been proposed. Therefore, we use an alternative based on the consistent Bayes approach, assuming sequential priors for models and taking the number of clusters as a random variable, too. The problem is then solved with the aid of an RJ-MCMC procedure. Further, inside the clusters, an additional analysis of heterogeneity is performed.

**Simon P. Wilson and M. J. Costello**

*Predicting Future Discoveries of European Marine Species using a Non-homogeneous Renewal Process*

Predicting future rates of species discovery and the number of species remaining are important in efforts to preserve biodiversity, discussions on rate of species extinction and comparisons on the state of knowledge of animals and plants of different taxa. In this paper, data on discovery dates of species in 32 European marine taxa are analysed using a class of thinned temporal renewal process models. These models allow for both under and over-dispersion with respect to the non-homogeneous Poisson process. An approach for implementing Bayesian inference for these models is described that uses Markov chain Monte Carlo simulation and that is applicable to other types of thinned process. Predictions are made on the number of species remaining to be discovered in each taxon.

**Michael P. Wiper, Maria Concepcion Ausin and Rosa Lillo**

*Bayesian control of the number of servers in a $GI/M/c$ queueing system*

In this paper we consider the problem of designing a $GI/M/c$ queueing system. Given arrival and service data, our objective is to choose the optimal number of servers so as to minimize an expected cost function which depends on quantities, such as the number of customers in the queue. A semiparametric approach based on Erlang mixture distributions is used to model the general interarrival time distribution. Given the sample data, Bayesian Markov chain Monte Carlo methods are used to estimate the system parameters and the predictive distributions of the usual performance measures.

We can then use these estimates to minimize the steady-state expected total cost rate as a function of the control parameter c. We provide a numerical example based on real data obtained from a bank in Madrid.