JAMES RENNIE BEQUEST

REPORT ON CONFERENCE

Conference Title: Sixth Workshop On Bayesian Inference In Stochastic Processes (BISP6)

Travel Dates: 14th June 2009 – 21st June 2009

Location: Bressanone, Italy

Group Member(s):   Simon Aeschbacher

Aims: a) To present a poster on "Joint parameter estimation in bottlenecked populations using Approximate Bayesian Computation"
b) To discuss open methodological issues in Approximate Bayesian Computation and meet with people working on them
c) To learn about recent advances in Bayesian Inference

OUTCOME:

The BISP workshop is held every second year, alternately in Spain and Italy. Its goal is to bring together people working on Bayesian inference in stochastic processes. Both theoretical and applied contributions are intended. The sixth edition in Bressanone from 18th to 20th of June 2009 was an outstanding meeting in terms of quality of the talks and in terms of the work that was presented.

With about 250 participants, BISP6 had two advantages typical to small meetings: It was informal, which encouraged discussions after the talks; and it had no parallel sessions, which avoided the problem of having to choose between alternative talks. There were nine sessions and though most of them had a particular theme, there was quite some thematic overlap. Among the dominant topics were spatio-temporal processes and modelling, state-space models, non-parametric Bayesian approaches, and parameter estimation in diffusion processes. Most talks started off from a particular real-world problem, extended previous approaches to solve them and than applied the new method to the original problem. There were some purely theoretical talks, and others that focussed on technical aspects of computation and optimisation of algorithms. In the following, I will pick out four talks and summarise them. Later, I will come to the poster session and discuss one contribution that is of importance for my work.

Paul G. Blackwell (University of Sheffield, UK) presented a way of predicting the age of ice layers according to their depth. There is no strictly linear relationship between the depth and the age of ice, but the settlement of dust from the atmosphere leaves a signal in the ice core. This process shows an annual periodicity and by scanning ice scores, one can extract the signal. However, the distance between two dust peaks is not equal across years. By modelling the stochastic building process of these 'age lines', Blackwell can now relate the depth of ice layers to their age and give a formal measure of uncertainty. The method is aimed at estimating the age of layers up to some 500 years old.

Peter Green (University of Bristol) gave a talk on Markov modelling in genome-wide association studies (GWAS). GWAS look for regions of high correlation between genetic markers (these days mostly Single Nucleotide Polymorphisms, SNPs) and a trait of interest (e.g. a disease). Currently used methods make some assumptions about the evolution of DNA that are not actually met. For example, demography, recombination

and mutations can create patterns of linkage disequilibrium (LD) such that the assumption of independence of SNPs along the chromosome is not valid. The large number of associations tested can lead to spurious findings if one does not control for multiple comparisons. Green proposed a more sophisticated DNA strand model that involves a Markov chain that switches between different seeds. This way, the LD between SNPs can be modelled and exploited. The relationship between the allele at the casual site and the disease was modelled by logistic regression. The strand and the disease model were combined to a full Bayesian model which allows inference about the position of the causal sites. Applying the method to two examples of causal effects, Green showed that his approach gives narrower peaks of correlation around the candidate polymorphism, and that it shows fewer spurious peaks further away (see Hosking et al. 2008, Genetic Epidemiology). In his talk, Green assumed correctly phased genotypes. Future work will have to be concerned with uncertainty introduced by phasing. Also, many traits are quantitative, which poses another challenge for inference.

Abel Rodriguez (University of California Santa Cruz, USA) talked on nonparametric inference in spatio-temporal processes. Abel focused on observations that arise as a mixture of different underlying distributions. The mixing as such again follows a distribution, the mixing distribution. Rodriguez described a novel class of nonparametric priors for such mixing distributions. They are based on a stick-breaking process and the weights of that process are constructed as probit transformations of Gaussian random variables. This approach is more flexible than previous ones, allowing for dependence of the mixing weights over time or space. Rodriguez applied his approach to extend multivariate stochastic volatility models. Here, the innovation is that one does not assume constant means of returns over time, and that covariances between different assets are allowed to change over time. The model can thus explain the correlation between negative mean return and periods of high volatility observed on the financial markets. In the discussion following the talk, the issue was raised whether one should invest into more complex models for volatility or rather into assessment of existing models and decision support for stock-brokers. Clearly, this question was motivated by recent developments on the financial markets.

Finally, Marc A. Suchard (University of California Los Angeles, USA) impressed the audience by speeding up time-demanding phylogenetic computations by a factor of 100. The ingenious idea consists of parallelising the algorithms to a high degree and bringing them onto the Graphical Processing Unit (GPU), rather than running them on the Central Processing Units (CPUs). The main challenge was to come up with an efficient way of transferring data to and from the GPU chips. Once this bottleneck was removed, the GPU's ability to process small functional steps many times in parallel could be efficiently exploited.

The poster session included about 30 posters, covering theoretical and applied Bayesian statistics. Most applied posters came from biostatistics (including health, behaviour and genetics) or econometrics. A growing interaction with machine learning and computer science was visible.

In my poster I presented the application of Approximate Bayesian Computation (ABC) to the joint estimation of parameters in population genetics. ABC is a likelihood-free approach to find approximations to posterior distributions. The approach depends on simulation of data sets under a stochastic model designed to capture the relevant processes that generate the observation in nature. When data are high dimensional, it has become a standard to use summary statistics to project onto a lower dimensional space. Simulated data points are retained if they fall 'close' to the observation and rejected otherwise. The posterior distribution is then estimated from the parameter values belonging to accepted points. I presented the ABC algorithm I had used, showed results from an application to Alpine ibex (Capra ibex) in the Swiss Alps, and gave an overview on the next steps and on open issues in ABC. Two important issues are the choice of summary statistics and the choice of a metric in the rejection step. Both are interrelated and there is growing activity in this field. One of my goals was to get input on how these issues should be addressed. It was therefore convenient that, next to my poster, Richard D. Wilkinson (University of Sheffield, UK) presented a poster claiming to provide guidance for the choice of metric, the measure of 'closeness' between simulated and observed data points. In particular, he showed that ABC gives exact inference under the assumption of a particular model with a particular distribution of the error. The approximation made in ABC can thus be interpreted in terms of the error and its putative distribution. However, since the argument assumes a given choice of the metric and is conditional on that choice, it is not clear to me how one should actually get insight into what metric should be chosen in practice. Richard and me discussed this issue, and it seems that the solution still waits to be uncovered.

Bressanone provided a perfect environment for the conference. It is a scenic town in South Tyrol, in the Northern Italian Alps, and it has an impressive history. Formerly, it belonged to the Austro-Hungarian

Empire, but was then assigned to Italy after the First World War. It is therefore officially bilingual, with ~70% of the people speaking German as their mother tongue and ~20% Italian. A minority speaks Ladin, a dialect of the Raetians who had settled there before being Romanised. The German influence goes back to the invasion of the Bavarians. Until 2008, Bressanone was home to the Bishop of the Diocese of Bressanone. Since then, the structure of the Dioceses has been changed and the seat of the Bishop moved to Bolzano. Bressanone is also known for good food and local specialities, which made the stay there a great experience.

The next BISP was announced for 2011 in Colmenarejo, near Madrid. It will be organised by Michael Wiper (University Carlos III, Madrid).