University of
BRISTOL

Intractable Likelihoods Workshop 2015

Conference Programme

20-22nd May 2015

Supported by SuSTaIn i-like.org.uk

Welcome to the Intractable Likelihoods Workshop 2015

This is the third edition of a yearly series of workshops dedicated to the topic of intractable likelihoods.

In most statistical contexts, it is recognised that inference methodology based on the likelihood function are usually methods of choice. However such methods are not always easy to implement. For instance, in complex problems often with massive data sets, it can sometimes be completely impossible to even evaluate the likelihood function. The computational statistics revolution of the 1990s provided powerful methodology for carrying out likelihood-based inference, including Markov chain Monte Carlo methods, the EM algorithm, many associated optimisation techniques for likelihoods, and Sequential Monte Carlo methods.

Although these methods have been and are highly successful in making likelihood-based inference accessible to a wide range of problems from virtually every area of science and technology, we now have a far better understanding of their limitations, for example in high-dimensional problems and for massive data sets. Thus many challenging statistical inference problems of the 21st century cannot be addressed using existing likelihood-based methods. Examples which motivate the current project come from genetics, genomics, infectious disease epidemiology, ecology, commerce, and bibliometrics.

However there have been various recent breakthroughs in computational and statistical approaches to intractable likelihood problems, including pseudo-marginal and particle MCMC, likelihood-free methods such as Approximate Bayesian Computation, composite and pseudo-likelihoods, new simulation methods for hitherto intractable stochastic models, and adaptive Monte Carlo methods. These advances coupled with developments in multi-core computational technologies such as GPUs, have enormous potential for extending likelihood methods to meet the most difficult challenges of modern scientific questions.

SuSTaIN

SuSTaIn is Statistics underpinning Science, Technology and Industry, a programme with the ambitious goal of strengthening the discipline of Statistics in the UK, by equipping it to face the challenges of future applications. Thus the focus is on rigorous and innovative new theory and methodology – core statistics for the 21st century – aimed at and stimulated by generic challenges raised by the ‘data revolution’, in areas as diverse as genomics, astronomy, telecommunications and finance.

It is funded principally by a 3.5million Science and Innovation award from EPSRC, and partly by the University of Bristol, and runs from 2006 to 2016.

Wednesday 20th May 2015

09.00-10.15 : Arrival and registration (including refreshments)

10.15-10.30 : Welcome talk by Guy Nason, Christophe Andrieu and Gareth Roberts

10.30-11.30 : Nicolas Chopin

11.30-12.30 Chris Jewell

12.30-13.30 Lunch

13.30-14.30 : Ioannis Kosmidis

14:30-15:30 : Samuel Kou

15:30-16:00 : Break

16:00-17:00 : Iain Murray

16.30-18.00 Poster presentation & wine reception

Thursday 21st May 2015

9:00-10:00 : Jonathan Gair

10:00-11:00 : Marc Suchard

11:00-11:30 Break

11:30-12:30 : Dawn Woodard

12:30-13:30 Lunch

13:30-14:30 : Sinan Yildirim

14:30-15:30 : Peter Glynn

15:30-16:00 Break

16:00-17:00 : Paul Jenkins

19.00 Conference dinner

Friday 22nd May 2015

9:00-10:00 : Max Welling

10:00-11:00 : Matthew Stephens

11.00-11.30 Break

11:30-12:30 : Alexandre Thiery

End of the meeting

ORAL PRESENTATIONS
(in order of appearance)

Nicolas Chopin, CREST, Paris

The Poisson transform for unnormalised statistical models

(joint work with Simon Barthelmé)

Contrary to standard statistical models, unnormalised statistical models only specify the likelihood function up to a constant. While such models are natural and popular, the lack of normalisation makes inference much more difficult. Here we show that inferring the parameters of a unnormalised model on a space Ω can be mapped onto an equivalent problem of estimating the intensity of a Poisson point process on Ω. The unnormalised statistical model now specifies an intensity function that does not need to be normalised. Effectively, the normalisation constant may now be inferred as just another parameter, at no loss of information. The result can be extended to cover non-IID models, which includes for example unnormalised models for sequences of graphs (dynamical graphs), or for sequences of binary vectors. As a consequence, we prove that unnormalised parametric inference in non-IID models can be turned into a semi-parametric estimation problem. Moreover, we show that the noise-contrastive divergence of Gutmann & Hyvarinen (2012) can be understood as an approximation of the Poisson transform, and extended to non-IID settings. We use our results to fit spatial Markov chain models of eye movements, where the Poisson transform allows us to turn a highly non-standard model into vanilla semi-parametric logistic regression.
http://arxiv.org/abs/1406.2839

Christopher Jewell, Lancaster University

Rapid likelihood-based inference for epidemics — the hardware way.

Recent developments in MCMC for Bayesian inference on dynamical models have provided a tantalising vision for decision support in a wide range of applications. In a real-time setting, providing answers in a short time frame is critical to the utility of these methods, though this is often at odds with the computational cost of complex likelihood functions on large datasets. In the post-Moore's Law era, parallel computing has become the standard approach to achieve significant improvements in algorithmic runtime. However, the serial nature of MCMC presents challenges to its implementation on traditional CPU-based architectures. For example, inter-process communication limits scalability for within-chain parallelisation, whereas the length of the burn-in period imposes a lower bound on the runtime of a multiple-chain approach. In a surprising return to the early days of microcomputing, a promising solution lies in the use of co-processors, among which general-purpose graphics processing units (GPGPU) have gained populatiry. GPGPUs may acts a co-processors to accelerate the most computationally intensive stages of a MCMC algorithm, combining massive parallelisation with fast communication compared to CPU-based systems. This talk will describe such an approach to real-time Bayesian inference for epidemic models, using Nvidia's CUDA C++ library. Intensive likelihood calculations are delegated to the GPGPU, whilst the CPU directs an overall serial MCMC scheme. Using this technique, a highly scalable parallel algorithm is achieved at a fraction of the hardware cost of the equivalent CPU-based machine.

Ioannis Kosmidis, University College London

Reduced-bias inference for regression models with tractable and intractable likelihoods

This talk concerns work on a unified computational and conceptual framework for reducing bias in the estimation of statistical models from a practitioner's point of view. The talk will briefly discuss how shortcomings of classical estimators can be overcome via reduction of bias, and provide a few illustrations for well-used statistical models with tractable likelihoods, including regression models with categorical responses and Beta regression.

New results will then be presented on the use of bias reduction methods for linear mixed effects models by focusing on the, usually small-sample, setting of meta-regression in the presence of heterogeneity. The large effect that the bias of the nuisance variance components can have on inference motivates the application of the framework to deliver higher-order corrective methods to generalised linear mixed models. The challenges involved in this direction will be discussed and resolutions from current research will be presented.

Samuel Kou, Harvard

Fast analysis of dynamic systems via Gaussian emulator

Dynamic systems are used in modelling diverse behaviours in a wide variety of scientific areas. Current methods for estimating parameters in dynamic systems from noisy data are computationally intensive (for example, relying heavily on the numerical solutions of the underlying differential equations). We propose a new inference method by creating a system driven by a Gaussian process to mirror the dynamic system. Auxiliary variables are introduced to connect this Gaussian system to the real dynamic system; and a sampling scheme is introduced to minimize the 'distance' between these two systems iteratively. The new inference method also covers the partially observed case in which only some components of the dynamic system are observed. The method offers a substantial saving of computational time and fast convergence while still retaining high estimation accuracy. We will illustrate the method by numerical examples.

Iain Murray, University of Edinburgh

Deep conditional density estimation

Given a model with parameters t of data D, we may be able to simulate a large number of data-parameter pairs. From these, we might try to fit the conditional distribution p(t|D). _If_ we're able to do that, then given observed data D, the posterior is directly available, without running iterative inference algorithms, likelihood-based or otherwise. I'll talk about recent work on practical learning algorithms that adapt deep neural networks from machine learning to conditional density estimation.
http://homepages.inf.ed.ac.uk/imurray2/pub/14dnade/
http://arxiv.org/abs/1502.03509

Jonathan Gair, University of Cambridge

Title and abstract are appended to this programme.

Marc Suchard, UCLA

Sex, lies and self-reported counts: almost intractable general birth-death processes to model self-reported count of sexual behaviour

Abstract: Surveys often ask respondents to report non-negative counts, but respondents may misremember or round to a nearby multiple of 5 or 10. The error inherent in this heaping can bias estimation. To avoid bias, we propose a novel reporting distribution arising from a general birth-death process whose underlying parameters are readily interpretable as rates of misremembering and rounding. The process accommodates a variety of heaping grids and allows for quasi-heaping to values nearly but not equal to heaping multiples. Inference using this stochastic process requires novel, efficient techniques to compute finite-time transition probabilities for arbitrary birth-death processes that we provide through Laplace transforms and a continued fraction representation. We present a Bayesian hierarchical model for longitudinal samples with covariates to infer both the unobserved true distribution of counts and the parameters that control the heaping process. Finally, we apply our methods to longitudinal self-reported counts of sex partners in a study of high-risk behaviour in HIV-positive youth.

Dawn Woodard

Efficiency of ABC-MCMC / Driving Time Reliability Prediction

I will present two topics. The first consists of our recent results on the efficiency of approximate Bayesian computation (ABC). We address Markov chain Monte Carlo versions of ABC, presenting the surprising result that multiple pseudo-samples typically do not improve the efficiency of the algorithm as compared to employing a high-variance estimate computed using a single pseudo-sample. This means that it is unnecessary to tune the number of pseudo-samples, and is in contrast to particle MCMC methods, in which many particles are often required to provide sufficient accuracy.
The second topic is prediction of the reliability of driving time on a road network, for use in mapping services like Bing Maps or Google Maps. Such mapping services provide recommended routes between a specified origin and destination, along with predictions of driving time on those routes. However, there can be considerable uncertainty in those predictions, due for example to incomplete knowledge of traffic conditions. Accurate probabilistic forecasts of driving time account for this uncertainty, and can be used to report driving time reliability to a user, or as a component of fleet vehicle decision support systems. I present methods for probabilistic prediction of driving time on arbitrary routes in a road network at arbitrary times, and apply those methods to large volumes of mobile phone GPS data from the Seattle metropolitan region.

Sinan Yildirim, University of Bristol

On an alternative class of exact-approximate MCMC algorithms

(Joint work with Christophe Andrieu and Arnaud Doucet.)

Consider the standard Metropolis-Hastings (MH) algorithm for a given distribution P on x. This talk is on exact-approximate algorithms that expand the scope of MH to situations where its acceptance ratio r(x, x') is intractable.
We present a novel class of exact-approximate MH algorithms. The motivation is the desire to benefit averaging of multiple noisy estimates of r(x, x') and still preserving detailed balance w.r.t. P. We show that this is indeed possible with the use of a pair of proposal kernels and asymmetric acceptance ratios. Moreover, the steps within one iteration that increase statistical efficiency with the cost of extra computation are parallelizable.
One interesting application of the methodology that will be discussed in the talk is a simple extension of the exchange algorithm for doubly intractable distributions. Use of the methodology for general latent variable models will also be demonstrated with a toy example.

Peter Glynn, Stanford University

Dealing with the Initial Transient in Markov Chain Simulations

The "initial transient" refers to the initial segment of a Markov chain's simulation that is most impacted by bias related to a non-equilibrium initialization. Theoretical analysis of this problem typically involves computing bounds on the "second eigenvalue/spectral gap" of the process. In this talk, we discuss sample-based methods for mitigating the effect of the initial transient and measuring its magnitude. Such sample-based methods, derived from a simulation of the Markov chain, are potentially applicable in settings where good theoretical estimates of rates of convergence to equilibrium are unavailable.

Paul Jenkins, University of Warwick

A tractable approach to intractable likelihoods in population genetics models with recombination

Population genetics models provide a rich source of intractable likelihoods; only under very restrictive assumptions about the evolution of a population can a likelihood be written down in closed form. However, advances in DNA sequencing are providing a wealth of data on genetic variation, and we should like to perform inference on the numerous biological and demographic processes shaping this variation: mutation, natural selection, historical migrations, population structure, and so on. In practice we must usually resort to computationally intensive Monte Carlo approaches, summary statistics, heuristic model simplifications, or a combination of these. In this talk I will describe a new analytic method for the purposes of inference about the process of recombination. Recombination is a fundamental aspect of reproduction which causes the shuffling of genetic variants, or alleles, along a chromosome so that the genetic makeup of an offspring differs from that of its parent. Quantifying recombination is vital in for example locating genes associated with complex diseases. I will show how an application of the martingale central limit theorem can be used to derive an accurate model of recombination with a key property: its likelihood is entirely tractable. The result is illustrated by embedding the likelihood in a reversible jump Markov chain Monte Carlo algorithm, and applying this to genomic data from the model fruit fly Drosophila melanogaster. We constructed the first genome-wide maps of fine-scale recombination rate variation in this organism.

Max Welling, University of Amsterdam

Approximate Bayesian Computation with Noisy Gradients: From Big Data to Complex Simulations

Bayesian inference is a natural and elegant tool to handle uncertainty. Unfortunately, in all but trivial cases, exact Bayesian inference is computationally highly demanding. In the face of very large datasets, even approximate inference schemes, such as MCMC may become impractical. In this talk I will present progress for a class of approximate Bayesian inference methods that scale to very large datasets. These “stochastic gradient MCMC” methods use only small subsets of the data for every update and computation can be conveniently distributed across many machines. We will discuss some applications of this technique to Bayesian Topic Modelling, Collaborative Filtering and Network Models. Bayesian inference is also the method of choice in likelihood free settings, where the likelihood can only be assessed through, often highly complex, simulations. We show how the same stochastic gradient MCMC methods can also be successfully applied to high dimensional likelihood free inference problems.

Matthew Stephens, University of Chicago

Dynamic Statistical Comparisons

When several statistical methods exist for the same task, it is common, and useful, to compare their performance on empirical benchmark data - both real and simulated. However, the way this benchmarking is usually done in practice - by publishing comparisons in journal papers - is far from ideal. For example, comparisons are usually performed by the research group that developed one of the methods, which almost inevitably favours that method. Furthermore, performing these kinds of comparisons is incredibly time-consuming, requiring careful familiarization with software implementing the methods, and the creation of pipelines and scripts for running and comparing them. And in fast-moving fields new methods or software updates appear so frequently that comparisons are out of date before they even appear. In summary, the current system results in a large amount of wasted effort, with multiple groups performing redundant and sub-optimal comparisons. In this talk I will describe our (recently started and ongoing) work to try to help improve this situation. Our goal is to enrol your help to make this kind of benchmarking more dynamic - by developing (somewhat) standardized formats and structures for making statistical comparisons, and, most importantly, by publishing code, data and pipeline scripts in easily-extendible public internet repositories.

Alexandre Thiery, National University

Asymptotic Analysis of Random-Walk Metropolis on Ridged Densities

The asymptotic behaviour of local-move MCMC algorithms in high-dimensions is by now well understood and the emergence of diffusion processes as trajectory limits has been proved in many such contexts. The results obtained so far involve mainly i.i.d scenarios, though there are now a number of generalizations in high/infinite dimensions. We adopt a different point of view and look at cases when the target distributions tend to exhibit `ridges' along directions of the state space. Such contexts could arise for instance in classes of models when data arrive with small noises or when there are non-identifiable subsets of parameters. In an asymptotic context all probability mass will concentrate on a manifold. We show that diffusion limits (on a manifold) abound also in this set-up of ridged densities and are particularly useful for identifying computational costs or providing optimality criteria.

POSTER PRESENTATIONS

Jim Barrett, University of Birmingham

CARMA models for the efficient characterisation of noise in the time domain

Noise is an inevitable part of any experimental measurement. It is therefore of great importance to develop techniques of efficiently characterising and understanding the properties of all types of noise. We present a general Continuous Auto-Regressive Moving Average (CARMA) model for correlated noise in the time domain, together with an algorithm to efficiently calculate its likelihood function.
The model can be used to create a 'whitening filter' that removes correlations from noise and renders it ~N(0,1) white noise. We expect this technique to have a wide range of applications, especially in the field of astronomy, where data-sets are typically very large and unevenly sampled in time.

Murali Haran, Penn State University

Computationally Efficient Emulation and Calibration for an Ice Sheet Model

This is joint work with Won Chang (U. of Chicago), Klaus Keller (Penn State Geosciences), Roman Olson, Dave Pollard and Patrick Applegate (Penn State Earth and Environmental Systems Institute).

Predicting the future behaviour of the Greenland and West Antarctic ice sheets is a challenging problem that is of great interest from both science and policy perspectives. A sound approach to prediction requires combining output from ice sheet models with observational data regarding the past and present state of the ice sheet. A statistical model that "emulates" the ice sheet model's behaviour at various parameter settings acts as a surrogate likelihood function. This function connects the ice sheet model output to the parameters. Inference for these parameters based on observational data is challenging due to the nature and size of the data and model output. I will describe computationally expedient Gaussian process-based methods that account for various sources of uncertainty when performing inference for parameters of an ice sheet model.

Luke Kelly, University of Oxford

Lateral transfer on phylogenetic trees

Bayesian phylogenetic methods for inferring the phylogeny of homologous traits based on vertical inheritance, while promising, are prone to errors and inconsistencies when the data-generating mechanism includes lateral transfer. While the individual trait histories are tree-like, they conflict with the overall phylogeny. To address this model misspecification, we describe and fit a phylogenetic model for the diversification of homologous traits which explicitly incorporates lateral transfer.
We take the binary stochastic Dollo model (Nicholls and Gray, 2008; Ryder and Nicholls, 2011) as our starting block and extend it to allow for the lateral transfer of traits between lineages of a phylogenetic tree. The model differs from those of Nakhleh et al. (2005) and Szöllősi et al. (2012) in that it does not induce a graph structure or require modifications to the tree as we integrate over all possible trait-borrowing times and locations.

The likelihood is intractable as the integration step requires us to integrate over all possible states of the process as time progresses, and the number of states grows exponentially with the number of taxa. The dimension of the system poses a number of computational problems so we present a tractable approximation based on Green's Functions solutions to the systems of differential equations.

References:

L. Nakhleh, D. Ringe, and T. Warnow. Perfect Phylogenetic Networks: A New Methodology for Reconstructing the Evolutionary History of Natural Languages. Language, 81(2):pp. 382–420, 2005.

G.K. Nicholls and R.D. Gray. Dated ancestral trees from binary trait data and their application to the diversification of languages. Journal of the Royal Statistical Society: Series B, 70(3):545–566, 2008.

R.J. Ryder and G.K. Nicholls. Missing data in a stochastic Dollo model for binary trait data, and its application to the dating of Proto-Indo-European. Journal of the Royal Statistical Society: Series C, 60(1):71–92, 2011.

G.J. Szöllősi, B. Boussau, S.S. Abby, E. Tannier, and V. Daubin. Phylogenetic modeling of lateral gene transfer reconstructs the pattern and relative timing of speciations. Proc. Natl. Acad. Sci. U.S.A., 109(43):17513–17518, 2012.

Sam Livingstone, University College London

Geometric ergodicity of the Random Walk Metropolis with position-dependent proposal covariance.

We consider a Metropolis-Hastings algorithm with proposals x' \sim N(x, hV(x)). In words, a Gaussian random walk in which the proposal variance changes dependent on the current position. The algorithm has been proposed a few times in the literature, in one instance by Roberts & Stramer (2002) as a discretisation of a tempered Langevin diffusion, with V(x) = \pi^{-1}(x)I, and more recently through the Kernel Adaptive Metropolis-Hastings method (Sejdinovic et al., 2014), in which V(x) is chosen by projecting past samples from the chain into some feature space (chosen through a reproducing kernel) and performing principal component analysis there. The latter in particular has generated some interest as a method which does not require derivatives of \log\pi(x) (so can be used in conjunction with pseudo-marginal approaches) but can still adapt to the local curvature of the target density.

Our interest lies in how different choices of V(x) affect the ergodic properties of the method, under different tail assumptions on \pi(x). We present results in one dimension outlining some necessary and sufficient growth conditions on V(x) and tail behaviour of \pi(x) for the algorithm to produce a geometrically ergodic Markov chain, and give an illustrative example in the multi-dimensional case.

Felipe Medina Aguayo

Stability of Noisy Metropolis-Hastings

This is joint work with Anthony Lee and Gareth Roberts.

Pseudo-marginal Markov chain Monte Carlo methods for sampling from intractable distributions have gained recent interest and have been theoretically studied in considerable depth. Their main appeal is that they are exact, in the sense that they target marginally the correct invariant distribution. However, the pseudo-marginal

Markov chain can exhibit poor mixing and slow convergence towards its target. As an alternative, a subtly different Markov chain can be simulated, where better mixing is possible but the exactness property is sacrificed. This is the noisy algorithm, initially conceptualised as Monte Carlo within Metropolis (MCWM), which has also been studied but to a lesser extent. In this poster we provide a further characterisation of the noisy algorithm, with a focus on fundamental stability properties like positive recurrence and geometric ergodicity. Sufficient conditions for inheriting geometric ergodicity from a standard Metropolis-Hastings chain are given, as well as convergence of the invariant distribution towards the true target distribution.

Christian Naesseth, Linköping University

Nested Sequential Monte Carlo Methods

We propose nested sequential Monte Carlo (NSMC), a methodology to sample from sequences of probability distributions, even where the random variables are high-dimensional. NSMC generalises the SMC framework by requiring only approximate, so called properly weighted, samples from the SMC proposal distribution, while still resulting in a correct SMC algorithm. Furthermore, NSMC can in itself be used to produce such properly weighted samples. Consequently, one NSMC sampler can be used to construct an efficient high-dimensional proposal distribution for another NSMC sampler, and this nesting of the algorithm can be done to an arbitrary degree. These properties allow us to consider complex and high-dimensional models using SMC. We illustrate the merits of the method on e.g. a climatological model for drought prediction, with a 1056-dimensional state space.

Dennis Prangle, University of Reading

Approximate Bayesian computation (ABC) performs approximate inference when the likelihood function is expensive or impossible to compute. Instead simulations from the model are performed for various parameter values and accepted if they are close enough to the observations. There has been much progress on deciding which data features to use to judge closeness, but less work on how to weight them. This poster presents a scheme to adaptively update these weights within the ABC-SMC algorithm of Toni et al. (2009).

Nick Tawn, University of Warwick

Improving the Efficiency of the Parallel Tempering Algorithm

Bayesian inference typically requires MCMC methods to evaluate samples from the posterior, however it is important that the MCMC procedure employed samples 'correctly' from the distribution for the sample estimates to be valid. For instance, if the posterior distribution was multi-modal then by running an MCMC procedure for only a finite number of runs it is possible that the chain can become trapped and not explore the entire state space. Well known algorithms to aid mixing in multimodal settings are the Parallel and Simulated tempering algorithms. The poster will introduce these and demonstrate their powers and weaknesses for sampling in such situations, and then describe the way in which these algorithms can be setup to achieve optimal efficiency when sampling.

The key feature of these algorithms is the ability to share information from the mixing in the hotter states to aid the mixing of the chain in the hotter states. This poster also presents a new approach based on reparameterisation that could potentially enhance the algorithms' efficiency. Empirical evidence is illustrated to show that this new algorithm appears to vastly enhance the trade of mixing information between temperature levels when targeting certain posterior distributions.

Will Vousden, University of Birmingham

Dynamic temperature selection for parallel tempered MCMC samplers

The emerging field of gravitational wave astronomy – and many other areas of physics – requires efficient methods for characterising complex, high-dimensional, and often multi-modal probability distributions. Most popular methods, such as Markov chain Monte Carlo sampling, perform poorly on strongly multi-modal distributions, being liable to settle on one mode without finding others. Parallel tempering addresses this problem by sampling simultaneously with separate Markov chains from tempered versions of the target distributions in such a way that the modes are flattened out, and therefore easier for the sampler to migrate between.

An open problem in the application of parallel tempering is the selection of the temperatures at which to sample from the target distribution. We propose an adaptive scheme that selects temperatures in order to optimise the communication between samplers of different temperatures, and therefore maximise the efficiency of the sampler. In our tests, scheme reduces the autocorrelation time of the sampler by a factor of 1.5 or more.

Sophie Watson, University of Bristol

Adaptive Summary Statistic Selection within a Sequential Monte Carlo algorithm for Approximate Bayesian Computation

Sequential Monte Carlo (SMC) algorithms for Approximate Bayesian Computation (ABC) have been shown to improve the accuracy of inference, compared to standard rejection ABC. The reason for this improvement is that standard rejection ABC proposes particles from the prior, whereas the proposal distribution in SMC approaches the posterior. However, as we demonstrate here, standard SMC-ABC methods are still very sensitive to the dimensionality of summary statistics. In addition, such methods can be expensive to implement as they require many simulations from the model. Often a large proportion of simulations are quickly `rejected' and subsequently ignored. SMC-ABC methods are also affected by the choice of summary statistics: Summarising the data poorly results in worse inference. We present a novel algorithm which uses knowledge gained from all historical simulations from the model to select summary statistics to use for the next iteration of SMC. We show that our algorithm performs favourably on a toy model in which summary statistics have been deliberately poorly chosen and obtains estimates which are remarkably close to those which could be obtained using sufficient statistics.

Giacomo Zanella, University of Warwick

Bayesian complementary clustering, MCMC and Anglo-Saxon placenames

Common cluster models for multi-type point processes model the aggregation of points of the same type. In complete contrast, in the study of Anglo-Saxon settlements it is hypothesized that administrative clusters involving complementary names tend to appear. We investigate the evidence for such a hypothesis by developing a Bayesian Random Partition Model based on clusters formed by points of different types (complementary clustering).

As a result we obtain an intractable posterior distribution on the space of matchings contained in a k-partite hypergraph. We apply the Metropolis-Hastings (MH) algorithm to sample from this posterior. We consider the problem of choosing an efficient MH proposal distribution and we obtain consistent mixing improvements compared to the choices found in the literature. Simulated Tempering techniques can be used to overcome multimodality and a multiple proposal scheme is developed to allow for parallel programming. Finally, we discuss results arising from the careful use of convergence diagnostic techniques.

This allows us to study a dataset including locations and placenames of 1316 Anglo-Saxon settlements dated approximately around 750-850 AD. Without strong prior knowledge, the model allows for explicit estimation of the number of clusters, the average intra-cluster dispersion and the level of interaction among placenames. The results support the hypothesis of organization of settlements into administrative clusters based on complementary names.