Challenges and Advances in High Dimensional and High Complexity Monte Carlo Computation and Theory

David Ceperley (University of Illinois Urbana-Champaign),
Yuguo Chen (University of Illinois Urbana-Champaign),
Radu V. Craiu (University of Toronto),
Xiao-Li Meng (Harvard University),
Antonietta Mira (University of Lugano),
Jeffrey Rosenthal (University of Toronto)

March 18, 2012–March 23, 2012

1 Overview of the Field

It is commonly recognized in the literature that the only possible way to estimate many realistic highly structured and high dimensional statistical models that properly describe the real world and the complex interactions among the variables that come into play, is by using computational tools such as Monte Carlo methods. The development and application of Monte Carlo methods has been an active research area for the last two decades. Many useful Monte Carlo techniques have been proposed in the literature, including Markov chain Monte Carlo (MCMC), sequential Monte Carlo, adaptive MCMC, perfect sampling, and quantum Monte Carlo.

While these Monte Carlo algorithms have turned into standard tools over the past decade (with dedicated software developed to implement them, such as Winbugs, www.mrc-bsu.cam.ac.uk/bugs, and numerous packages within the R project, www.r-project.org), they still face difficulties in handling less regular problems such as those involved in deriving inference for high-dimensional models, data sets that contain severe collinearity, massively multimodal target distributions, hierarchical latent variable models, or rapidly evolving phenomena. Two of the main problems encountered when using MCMC in these challenging settings are: 1) it is difficult to design a Markov chain that efficiently samples the state space of interest; and 2) the resulting MCMC estimators have high variance thus producing less reliable inference and poor forecasts. As a result of these challenges, while the central keyword of the nineties was "convergence diagnostic" for MCMC algorithms, the research in this area now focuses on adaptive algorithms and variance reduction techniques.

On the contrary to the lack of new methods for complex problems, users of standard Monte Carlo methods sometimes suffer from an embarrassment of riches, in that there are many different related algorithms available which are all asymptotically valid. For example, the Metropolis requires a choice of proposal distribution, with some choices working much better than others. An important start on better understanding these choices was the pioneering paper of Roberts, Gelman, and Gilks [1], which theoretically determined the proposal distribution from among a fixed parametric family for random-walk Metropolis algorithms. This led to numerous follow-up papers. However, much work – both theoretical and practical – remains to be done regarding choice of proposal distribution and other MCMC tuning issues.

Another class of challenges in the MCMC world aims at reconciling the methodological advances in the computational tools with the applications where evolving and complex phenomena are overly present. Evolving systems clearly embrace a large variety of applied problems and it is thus useful to develop software and packages that allow the practitioners to implement in an almost automatic way the recent advances in the literature without having to write their own code. Currently there are many R routines to implement MCMC algorithms (such as MCMCpack, mcmc, MCMCglmm and mcclust), or to perform convergence diagnostic (such as BOA and CODA). There is a single routine to implement adaptive MCMC (AMCMC) and there is nothing available to reduce the variance of the resulting estimators mostly because the tools that have been so far proposed in the literature are not sufficiently general to motivate a dedicated software.

The area of exact Monte Carlo sampling with Markov chains, also known as perfect sampling, has had a meteoric rise due to the seminal concept of coupling from the past introduced by Propp and Wilson [2]. While the method is spectacularly successful in a number of applications involving finite state spaces, especially in statistical physics and point processes, its use remains problematic in statistical models where the parameter has a continuous (possibly unbounded) state space. It would be valuable to find a good trade-off between the classical MCMC methods whose implementability is very general but suffer from lack of reliable convergence diagnostics, and perfect sampling algorithms that eliminate the need for convergence diagnostics but require ingenious solutions for wider applicability.

2 Recent Developments and Open Problems

MCMC has had periods of rapid methodological development since its second birth in the early nineties as the main computational tool for performing Bayesian inference. Recently, one can notice that the incentive for further improvements and innovation comes mainly from three broad directions: the high dimensional challenge, the quest for adaptive procedures that can eliminate the cumbersome and often frustrating process of tuning "by hand" the simulation parameters for a complex MCMC design and the need for flexible theoretical support, arguably required by all recent developments as well as many of the traditional MCMC algorithms that are widely used in practice. Each of these directions is discussed below.

High Dimensional Challenge for MCMC. We have reached a time when collecting large volume of data for increasingly complex models is becoming standard in a number of high-impact scientific explorations in genetics, medicine, astronomy, to name just a few. When the state space is very high dimensional, not only it is difficult to develop MCMC samplers that are able to move across the space, but once a reasonable candidate has been build, it is difficult to assess its performance. Possible approaches involve a "divide and conquer" strategy in which analyses based on smaller data subsamples are performed and must be combined in the last stage of the analysis. Alternatively, one may wish to replace costly and lengthy numerical calculations that are asymptotically unbiased with bias-controlled approximations that can be much faster to obtain.

Approximate Bayesian Computation. The challenges of dealing with high-volume data and complex models reflect on our ability to express the latter mathematically. For instance, in astronomy and genetics it is not unusual to encounter a "black-box" model (usually the result of an extremely complex computer model) for which one can obtain likelihood values corresponding to the parameters fed into the model, but no general mathematical formulation is available. This feature makes the traditional calculation of posterior probabilities impossible and thus Bayesian inference and, consequently, MCMC algorithms must be implemented in a very different form known as Approximate Bayesian Computation (ABC). A most challenging aspect of ABC-based inference is determining which model, among a number of competing ones, is most suitable for the data at hand.

Ergodicity of Adaptive MCMC. For many users of MCMC methods, it becomes quickly apparent that one must carefully tune the transition kernel parameters in order to produce an efficient sampling algorithm.

The class of Adaptive MCMC (AMCMC) allows the modification of the Markov chain's transition kernel automatically and "on the fly", i.e. it allows the simulation process to self-adjust at each iteration n based on the information provided by all the samples drawn by that time. Such an approach violates the Markovian property as the subsequent realizations of the chain depend not only on the current state but also on all past realizations. This implies that one can validate theoretically this approach only if one is able to prove from first principles that the adaptive algorithm is indeed sampling from the stationary distribution π.

Exploration and Exploitation for Adaptive MCMC. The designer of an AMCMC algorithm has to keep in mind two crucial tasks the adaptive scheme must achieve. First, the algorithm must exploit efficiently the information obtained each time it modifies its parameters, and second, it must continue to explore and search for new regions of the sample space that have significant importance with respect to the density (or probability mass function) of π. Striking the right balance between these two tasks is the main methodological challenge for any AMCMC algorithm.

The practical implementation of AMCMC samplers requires careful consideration in those cases when the target distribution is multimodal. For instance, a posterior distribution may be multimodal if the sampling distribution is represented as a mixture. The latter occurs when we assume that the population of interest is heterogeneous or when such a formulation is a convenient representation of a non-standard density. It is well known that MCMC sampling from multimodal distributions can be extremely difficult as the chain can get trapped in one region of the sample space due to areas of low probability (bottlenecks) between the modes. Therefore, a large amount of effort has been devoted to designing efficient MCMC sampling methods for multimodal target distributions.

Theoretical Developments. The theoretical developments follow closely the methodological advances described above. For practical implementation and interpretation, a statistical estimator must be provided with an estimator of its variance. The MCMC estimators are no exception, although the dependence between samples can pose challenges to proving a central limit-type theorem. The CLT is known to hold when the MCMC chain is geometrically ergodic. However, it is rather difficult to prove for a general and moderately complex sampling algorithm that its underlying Markov chain is geometrically ergodic. Recent developments strive to simplify the conditions under which geometric ergodicity holds.

3 Presentation Highlights

High Dimensional Data. Small sample data with large number of variables often occur in genetics, and **Chiara Sabatti** from Stanford University described Bayesian models that can be particularly fruitful in analyzing these data, as well as the computational challenges that they pose. **David van Dyk** from Imperial College London, has focused his talk on the analysis of massive data sets that quickly become standard in astronomy. In this context MCMC requires careful use of existent strategies and the design of new ones. For instance, in the case of high-energy spectral analysis the classical Gibbs sampling is inefficient so a partially collapsed Gibbs sampler is proposed instead. This strategy reduced conditioning in the Gibbs scheme and improves convergence. In the case of generalized linear models with a high number of predictors, Bayesian methods for variable selection pose a great challenge. **Faming Liang** from Texas A&M University introduced a novel Bayesian subset regression model which incorporates idea from regularized likelihood methods within the Bayesian paradigm.

Adaptive MCMC. Jeffrey Rosenthal from the University of Toronto presented the fundamental theoretical results one can use to prove ergodicity for adaptive MCMC and he introduced an interesting counterexample where an intuitively appealing adaptive Gibbs sampler fails to converge, thus clearly proving the need for rigorous theoretical backing of any adaptive MCMC algorithm. **Éric Moulines** and **Gersende Fort** from Telecom Paris-Tech shared a talk on two papers, one on adaptive tempering and the other one on equi-energy sampling. **Nando de Freitas** from University of British Columbia has shown some interesting connections between the field of Bayesian optimization and AMCMC.

Approximate Bayesian Computation. The challenges of performing model selection within the ABC framework have been clearly stated by **Christian Robert** from Universite Paris-Dauphine. His discussion has focused on theoretical results related to the importance of the summary statistics in computing the ABC Bayes factor for two competing models. The difficulties of ABC computation in high dimensions have been discussed by **Scott Sisson** from University of New South Wales. Possible remedies based on regression adjustment and a novel marginal-adjustment strategy were proposed.

Sequential Monte Carlo. **Jun Liu** from Harvard University presented two ideas to improve the efficiency of sequential Monte Carlo methods. One is to allow lookahead in the sequential sampling so that future information can be used in generating samples and computing the weights. The other is to design a sequential rejection control sampler on lower resolution spaces in order to effectively sample the target distribution on a high resolution space. **Gareth Roberts** from University of Warwick presented recent work on a sequential importance sampler which provides online unbiased estimation for irreducible diffusions (that is ones for which the reduction to the unit diffusion coefficient case by the Lamperti transform is not possible).

Theoretical Developments. Novel convergence results for commonly-used Gibbs samplers were reported by **James Hobert** from the University of Florida. He has shown conditions that guarantee the geometrical ergodicity for a Gibbs sampler that is used for performing Bayesian quantile regression. Similarly, geometric ergodicity is shown for the Gibbs sampler used in Bayesian linear mixed models. A novel method for proving geometric ergodicity results for hierarchical models has been introduced by **Krzysztof Latuszynski** from University of Warwick.

New MCMC algorithms. While perfect sampling may be difficult to implement in Bayesian analyses, **Duncan Murdoch** from University of Western Ontario proposed to use near perfect samplers to gain insights on the performance of MCMC samplers. The talk by **Zhiqiang Tan** from Rutgers University compared the efficiency of resampling and subsampling strategies in MCMC.

Topics in Bayesian Inference. **Helene Massam** from York University considered the calculation of Bayes factors for hierarchical loglinear models for discrete data given under the form of a contingency table with multinomial sampling. Motivated by the invariance of copulas to monotone transformations, **Francois Perron** from University of Montreal proposed Bayesian inference for the copula based on ranks. The ABC method is applied to estimate an Archimedean copula using a nonparametric Bayesian approach.

4 Interactions Between Scientists and Statisticians

Over the past half century, physicists and chemists have made significant contributions to the Monte Carlo literature. Several key Monte Carlo algorithms, such as the Metropolis Algorithm, Gibbs sampler, and the Wang-Landau algorithm, are first invented by physicists and chemists, but are made more general by statisticians years later. The flow of knowledge is mainly unidirectional from physics to statistics, and usually it may take approximately 10 to 15 years before the cross-fertilization process bears fruits. One of the precise aims of this workshop is to shorten this time by having statisticians and physicists together exchanging ideas. Moreover, we believe that statisticians have something to offer to physicists so that the flow of knowledge can be bidirectional.

Lines of research originated in the physics literature that have been recently explored by some of the participants are the zero-variance principle which is a control variates variance reduction technique and the well-tempered ensemble which is the biased ensemble sampled by well-tempered metadynamics. The seminal papers by Assaraf and Caffarel [3] and Per, Snook and Russo [4] were published in the physics literature and propose a very interesting and effective variance reduction method for Monte Carlo simulation. Their original idea, named “Zero Variance”, could be useful also to statisticians who make wide use of Monte Carlo and MCMC simulation. Similarly, the papers by Parrinello and his collaborators on the so called “Well Tempered Metadynamics” [5] [6] could offer to statisticians a new way of designing Metropolis-Hastings type algorithms and move between very separated modes in a quite effective way.

Unfortunately the physics literature is in general not easy to read for a statistician since it contains notation and definitions that might not be standard (at least to the statistician) and with which he/she is typically not familiar. This is a serious problem that researchers encounter when trying to transfer knowledge from one scientific area to a different one. Indeed, the first issue in cross fertilization attempts, is to be aware of some interesting principle or idea that can be successfully transferred from one area to another. The second difficulty is the lack of a common background and "vocabulary" that allows the transfer flow of knowledge to be fast and effective.

Two of the speakers invited were Prof. Michele Parrinello (Italian physicist, 2009 Dirac medal, well know for his works in molecular dynamic, affiliated both with the ETH, Zurich, and the USI, Lugano) and Prof. Roland Assaraf (chemist by background, Université Pierre et Marie Curie, Paris VI, expert in quantum and fermion Monte Carlo and efficient MCMC). Besides their brilliant presentations, many conference participants reported that interacting with them was very enlightening and interesting to have a different prospective and to reach a valuable cross-fertilization among disciplines and fields. We are confident that, having these "foreign" experts available for research discussions during the conference, will foster fast scientific progress not only among the statistics and probability community, but also among the chemists and physicists since cross-fertilization is a bi-directional pathway.

5 Outcome of the Meeting

By bringing together an international group composed of statisticians, physicists, chemists, mathematicians, and computer scientists working in the Monte Carlo field, this meeting provided a platform for exchanging ideas and developing collaborations that will have a positive impact for years to come. Such exchanges have been at the origin of several directions in Monte Carlo and their benefit is still reverberating within scientific communities. This workshop gave participants an update on the current state of the field and identified new challenges in Monte Carlo computation and theory.

The feedback we received from workshop participants is overwhelmingly positive. Some participants feel that this is the "best structured workshop" that they have ever attended, "large enough to have significant content, but intimate such that it was possible to chat with everyone." The participants like the congenial atmosphere at the workshop which is "very conducive to discussions and planting the seeds for new research directions and projects." The balance of structured and unstructured time "allowed for intense exchanges".

The participants are happy to see a mix of experts with different interests (from theory to applications) in Monte Carlo. They enjoyed the presentations and found that the topics are interesting and stimulating. Some participants feel that "it was especially good to see how some serious applications were making substantial use of the theory for guidance."

Many participants said that they learned interesting ideas from the talks, and from informal discussions with the participants. Several participants brought back some ideas from the workshop that they plan to investigate in the near future. Some learned ideas of specific directions for the project they are currently working on. Some participants collaborated with each other at the workshop to write research papers or develop ideas for future publications.

All participants are thankful for superb facilities that the Banff Centre provides, from the accommodations, cafeterias, to the conference space with its video-recording system. BIRS has also been very helpful through the organization of the workshop. We would like to conclude the report by thanking BIRS for making the workshop such a wonderful experience for every participant.

References

[1] G. O. Roberts, A. Gelman and W. R. Gilks, Weak convergence and optimal scaling of random walk Metropolis algorithms, *The Annals of Applied Probability* **7** (1997), 110–120.

[2] J. G. Propp and D. B. Wilson, Exact sampling with coupled Markov chains and applications to statistical mechanics, *Random Structures and Algorithms* **9** (1996), 232–252.

[3] R. Assaraf and M. Caffarel, Zero-variance principle for Monte Carlo algorithms, *Physical Review Letters* **83** (1999), 4682–4685.

[4] M. C. Per, I. K. Snook and S. P. Russo, Zero-variance zero-bias quantum Monte Carlo estimators for the electron density at a nucleus, *Journal of Chemical Physics* **135** (2011), 134112.

[5] A. Barducci, M. Bonomi and M. Parrinello, Metadynamics, *Computational Molecular Science* **1** (2011), 826–843.

[6] G. A. Tribelloa, J. Cuny, H. Eshet and M. Parrinello, Exploring the free energy surfaces of clusters using reconnaissance metadynamics, *Journal of Chemical Physics* **135** (2011), 114109.