Random generation of combinatorial objects

Sujet de stage M2

`anne.bouillard@ens.fr, ana.busic@ens.fr`

Context

Generating random objects according to a some distribution that is not easily computable is an important problem in various research fields, such as theoretical physics, communication networks, stochastic geometry and theoretical computer science. For example, it can be used to approximate $\#P$-complete problems: generating an object (let say the independent sets of a graph) uniformly can be used to estimate the number of independent sets of the graph. In many cases, this uniform distribution can easily be defined as the stationary distribution of a finite-state Markov chain.

The classical Markov Chain Monte Carlo simulation techniques suffer from two main problems:

- The convergence to the stationary distribution can be very slow, and it is in general difficult to estimate;
- Even if one has an effective convergence criterion, the sample obtained after any finite number of iterations is biased.

To overcome these issues, Propp and Wilson [7] have introduced a Perfect Sampling Algorithm (PSA) that has been later extended and applied in various contexts (see the annotated bibliography by Wilson [8]).

PSA uses coupling arguments to give an unbiased sample from the stationary distribution of an ergodic Markov chain on a finite state space $\mathcal{X}$. Assume the chain is given by an update function Φ and an i.i.d. sequence of innovations $(U_n)_{n\in\mathbb{Z}}$, so that

$$X_{n+1} = \Phi(X_n, U_{n+1}). \tag{1}$$

The algorithm is based on a backward coupling scheme: it computes the trajectories from all $x \in \mathcal{X}$ at some time in the past $t = -T$ until time $t = 0$, using the same innovations. If the final state is the same for all trajectories (i.e. $|\{\Phi(x, U_{-T+1}, \ldots, U_0) \,:\, x \in \mathcal{X}\}| = 1$, where $\Phi(x, U_{-T+1}, \ldots, U_0) := \Phi(\Phi(x, U_{-T+1}), U_{-T+2}, \ldots, U_0)$ is defined by induction on T), then we say that the chain has globally coupled and the final state has the stationary distribution of the Markov chain. Otherwise, the simulations are started further in the past.

Any ergodic Markov chain on a finite state space has a representation of type (1) that couples in finite time with probability 1, so Propp and Wilson's PSA gives a "perfect" algorithm in the sense that it provides a *unbiased* sample in *finite time*. Furthermore, the stopping criterion is given by the coupling from the past scheme and knowing the explicit bounds on the coupling time is not needed for the validity of the algorithm.

However, the applicability of PSA suffers from two drawbacks :

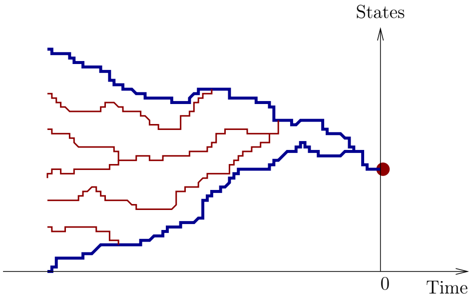

Figure 1: Coupling from the past

- The coupling time can be very large.
- One needs to run one simulation per state in $\mathcal{X}$, which limits its application only to chains with a state space of very small cardinality.

Various techniques have been developed to reduce the number of trajectories that need to be considered in the coupling from the past procedure. A first crucial observation already appears in [7]: for a monotone Markov chain, one has to consider only extremal initial conditions. This remark can be generalized in several ways:

- two bounding trajectories can sandwich the set of all the trajectories computed by the algorithm ([5, 2]).
- consider an enlarged state space so that one trajectory of the new state space include all the trajectories computed by the algorithm ([3]).

Objectives

There are two possible directions, depending on the background and the interests of the candidate:

Random graphs. The first direction is to generalize results in [3, 2]. One possibility is to study a specific problem from graph theory, or to define a class of problems where the technique in [3] can be transpose. An example of specific problem is to design an efficient perfect sampler for a maximum matching of a given bipartite graph. Some result exist in the case of dense regular bipartite graphs [4].

Coupling time. The second direction is to study the coupling time for some specific models, such as for instance the class of almost space-homogeneous Markov chains defined in [1], or problems from graph theory. The approach would combine the recent developments on mixing times for Markov chains [6]. Also, we wish to investigate the relations between the coupling time and the algorithmic complexity of problems. This could explain the impossibility of a good (efficient) sampling for some combinatorial problems.

Prerequisites. Basic knowledge of discrete probability theory is required. Some prior experience with programming (C, C++ or Matlab) is strongly recommended.

Funding. The internship is funded and there is a possibility to continue with a PhD.

References

[1] A. Busic, B. Gaujal, and F. Pin. Perfect sampling of markov chains with piecewise homogeneous events. *Performance Evaluation*, 2011. accepted for publication.

[2] A. Bušić, B. Gaujal, and J.-M. Vincent. Perfect simulation and non-monotone Markovian systems. In *3rd International Conference Valuetools'08*, Athens, Greece, Oct. 2008. ICST.

[3] M. Huber. Perfect sampling using bounding chains. *Ann. Appl. Probab.*, 14(2):734–753, 2004.

[4] M. Huber. Exact sampling from perfect matchings in dense regular graphs. *Algorithmica*, 44:183 193, 2006.

[5] W. S. Kendall and J. Møller. Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes. *Advances in Applied Probability*, 32(3):844–865, 2000.

[6] D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, 2006.

[7] J. G. Propp and D. B. Wilson. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms*, 9(1-2):223–252, 1996.

[8] D. B. Wilson. Perfectly random sampling with Markov chains. Annotated bibliograpy. `http://dimacs.rutgers.edu/ dbwilson/exact.html/`.