Perfect Simulation for Spatial Point Processes

by

W.S. Kendall

April 1997

Invited Paper, *Simulation of Stochastic Processes in Engineering* meeting
51st Session of ISI, Istanbul August 1997

This is `Paper.tex` (`RISC OS`).
Last edited: 19:20:14, 03/04/97.

Perfect Simulation for Spatial Point Processes.

Wilfrid S. Kendall, Statistics, University of Warwick, Coventry CV4 7AL, UK

1 Simulation, reversibility, and Markov chains

Several areas of mathematical science share a particular requirement for simulation techniques to estimate the normalising constant for a probability density known only up to proportionality. Thus, in statistical mechanics a fundamental technical challenge is presented by models specified as probability densities

$$p(x;\theta) \quad = \quad \exp\left(-E(x;\theta)\right)/K(\theta) \tag{1.1}$$

(where typically x ranges over a high-dimensional state space) for which the *energy* $E(x;\theta)$ is specified explicitly, but the *partition function* $K(\theta)$ is defined only implicitly by the requirement that it normalises the integral of the density. Asymptotic theory for $K(\theta)$ is available and important, but simulation also plays a fundamental rôle. (In fact historically the development of Markov point process theory owes much to the ideas developed in statistical mechanics to deal with this problem.)

In computer science the theory of randomised algorithms addresses problems such as the question of computing efficiently the volume of a high-dimensional bounded convex set K. In effect one requires the normalising constant of the *uniform* probability density over K. A standard argument shows how to do this if one can sample a point drawn uniformly from the convex set K.

Bayesian statistics deals with a likelihood $\ell(y|\psi)$ for the data y given the parameter ψ, and a prior density $p(\psi)$ for ψ. One requires the posterior density $p(\psi|y)$, and knows it is proportional to the joint density $\ell(y|\psi) \times p(\psi)$. Once again, the issue is to compute a normalising constant.

In point process theory important and flexible models for random point patterns arise from densities with respect to Poisson point processes. For example consider the Strauss point process (Strauss, 1975), with density with respect to a Poisson point process in a bounded window W given for $\gamma \in (0,1)$ by

$$p_s(\mathbf{x}) \quad = \quad \gamma^{s(\mathbf{x})}/\kappa_s(\gamma) \tag{1.2}$$

(where $s(\mathbf{x})$ counts the pairs of points in the pattern $\mathbf{x}$ which are closer than a specified radius r); or the area-interaction point process (Baddeley & van Lieshout, 1995), with density given for $\gamma \in (0,\infty)$ by

$$p_a(\mathbf{x}) \quad = \quad \gamma^{-A(U_{\mathbf{x}})}/\kappa_a(\gamma)\,, \tag{1.3}$$

where $U_{\mathbf{x}}$ is the union of balls of radius r centred at the points of $\mathbf{x}$, and $A(U_{\mathbf{x}})$ is the area of this union. In both cases the definition can be extended by standard techniques to define a point process over the whole plane $\mathbb{R}^2$ (though care must then be taken to consider whether phase-transitions occur). Practical application requires the computation of the normalising constants κ_s, κ_a.

Simulation-based solutions can be obtained in all these cases using Markov chains, typically exploiting special features of the problem in order to deliver realistic algorithms. In computer science one uses a random walk reflected at the boundary of K: its equilibrium distribution is uniform, so a sufficiently long-run simulation should deliver an approximately uniform sampling procedure – careful constructions and arguments establish that this delivers the normalising constant computed efficiently to a required accuracy.

In statistics one exploits graphical structure implicit in conditional independence relations between the parameter coordinates ψ_1, ..., ψ_k to build a reversible Markov chain whose state-space is parameter space, varying one parameter at a time. By reversibility (under lenient regularity conditions) this *Gibbs' sampler* has the posterior distribution as its equilibrium. In the statistical context this idea (sampling from a target distribution using the equilibrium distribution of a Markov chain) is referred to as *Markov Chain Monte Carlo* (MCMC).

Statistical mechanics substantially predates these other contexts, and uses these methods and further variations, including accept-reject methods known as *Metropolis-Hastings algorithms*. But once again the basic idea is to create a Markov chain, typically reversible, whose equilibrium distribution is the required distribution, and which can be simulated without knowledge of the normalising constant.

For point processes of the form described above one can simulate a *spatial birth-and-death process* (random patterns of points which are born and die as time passes in a Markov fashion) whose equilibrium distribution is the required point process law. Using the ideas of reversibility such a Markov chain can be constructed without knowledge of the normalising constant, so the simulation algorithm is feasible. A variation is the "add-delete" algorithm (alternating births and deaths systematically), and a wide-ranging generalization uses Metropolis-Hastings ideas. Møller (1997) presents a useful survey.

In every case it is vital that the Markov chain runs long enough to be approximately in equilibrium.

2 How long is a long-run simulation?

The single big implementation issue is to decide how long is "long enough"? This is more than a theoretical nicety. There are cases in the literature where hindsight has shown that simulations have clearly *not* been carried out for long enough for the chain to reach approximate equilibrium. This is a particularly vexed issue for point processes, where one is dealing with a probability measure over an infinite-dimensional space, and it would be foolish to ignore problems of existence, false convergence, and multi-modality.

An empirical response to this issue has been to devise diagnostics to warn when convergence has not yet been reached. For example, in the statistical context workers have proposed the use of multi-start algorithms: independent copies of the Markov chain in question are started off from several different starting points chosen randomly according to a distribution known to be over-dispersed with respect to the target equilibrium distribution. If there is no agreement between estimates of the normalising constant based on these different starts then the simulation needs to be continued for longer. This has been formalised in a number of ways (see various articles in Bernardo et al., 1992; Gilks, Richardson & Speigelhalter, 1995); Johnson (1996) couples the copies in a manner which is to some extent related to the work discussed below.

As is ever the case, none of these diagnostics will deliver us from the need for good judgement and mathematical insight. In particular for point processes the multi-start approach is badly hindered by the fact that it is hard to formalise the notion of "over-dispersion" in an infinite-dimensional context.

Ideally one would like more than a diagnostic. A perfect solution to the problem of how long is "long enough" would be to augment the simulation algorithm to produce a certificate when equilibrium has been attained. Unlikely as it seems, in modestly favourable circumstances this can be achieved!

3 Exact sampling for finite Markov chains

Conventional MCMC works by running a simulated Markov chain *forwards* in time. Propp & Wilson (1996) observe that by running it *backwards* in time one can construct a "Coupling From The Past" (CFTP) algorithm which produces a certificate of when equilibrium has been attained. We describe their construction in the case of a very simple example. Consider the (exactly computable!) question of determining the equilibrium distribution π of a simple random walk trapped between reflecting barriers at 0 and $N > 0$. Suppose that the potential jumps at various times are computed once-for-all as independent identically distributed innovations $\{Z_t : t \in \mathbb{Z}\}$. Implement the random walk $X_n^{\min}$ beginning at state 0 at time $-n$ by

$$X_n^{\min}(-n) \quad = \quad 0\,; \qquad X_n^{\min}(t+1) \quad = \quad \begin{cases} X_n^{\min}(t) + Z(t) & \text{if } X_n^{\min}(t) \in (0, N) \\ X_n^{\min}(t) + \max\{Z(t), 0\} & \text{if } X_n^{\min}(t) = 0 \\ X_n^{\min}(t) + \min\{Z(t), 0\} & \text{if } X_n^{\min}(t) = N \end{cases} \tag{3.1}$$

so that the barriers reflect simply by preventing moves across them (this non-standard reflection makes the chain aperiodic, but is also necessary for the CFTP construction, as it supplies a crucial monotonicity). We also implement $X_n^{\max}$ in the same way, *using the same innovations* $\{Z(t) : t \in \mathbb{Z}\}$, but with $X_n^{\max}(-n) = N$.

By standard Markov chain theory we have convergence in distribution to equilibrium: $\mathcal{L}(X_n^{\min}(0)) \Rightarrow \pi$. On the other hand, by the monotonic nature of the construction there is an *a.s.* finite *coalescence time*

$$T \quad = \quad \inf\{n : X_n^{\min}(0) = X_n^{\max}(0)\} \tag{3.2}$$

such that $X_n^{\min}(0) = X_n^{\max}(0) = X_T^{\min}(0)$ for all $n \geq T$. Hence $X_n^{\min}(0) \to X_T^{\min}(0)$ almost surely as $n \to \infty$ (even in the discrete metric!). It follows that $X_T^{\min}(0)$ has the equilibrium distribution π. Heuristically, we can view $X_T^{\min}(0)$ as the product of a "virtual simulation from time minus infinity" $X_\infty^{\min}(0)$.

It is crucial that T is *observable*, in the sense that one can determine whether or not $T > n$ simply by comparing $X_n^{\min}(0)$ and $X_n^{\max}(0)$, so delivering the required certificate. Thus the above provides an effective simulation algorithm which samples exactly, or "perfectly", from the equilibrium distribution.

It is also informative to note that simulating as above but running forwards in time will lead to severe bias, since coalescence of the forwards simulation can happen only at states 0 or N.

Variations on this procedure work for any finite-state-space ergodic Markov chain. Of course it is necessary to deliver *effective* ways of determining whether or not coalescence has occurred. In general one needs some extra structure, typically monotonicity. Thus for example Propp & Wilson (1996) describe applications to the Ising model, and elsewhere to problems in computer science relating to Markov chains on partially ordered sets. Murdoch & Green (1997) avoid the monotonicity requirement, essentially by exploiting the Markov chain concept of "small sets", and so carry out CFTP for some continuum state space examples (this is nicely related to geometric ergodicity by Foss & Tweedie, 1997).

We note here that CFTP is not the only option for perfect simulation. Fill (1997) describes an algorithm using rejection sampling and based on the notion of strong uniform random times, which protects against possible bias caused by the user becoming impatient and not waiting for coalescence to happen.

4 The case of spatial birth-and-death processes

For spatial point process models, as given by (1.2) and (1.3), one needs to perform perfect simulation in a non-finite, non-uniformly ergodic situation. The difficulty may be seen by considering how to extend the random walk example above to the case when there is just one barrier at 0. There is in fact a limiting process $X^{\min}_\infty = \lim_{n\to\infty} X^{\min}_n$, and a coalescence time T determined by $T = \inf\{n : X^{\min}_n(0) = X^{\min}_\infty\}$, but T is now unobservable because there is no longer an observable maximal process $X^{\max}_n$. How to proceed?

In Kendall (1998), further illustrated using a (one-dimensional) birth-death process in Kendall (1997), it is shown how to proceed by constructing an observable dominating process D, guaranteed to lie above the virtual simulation from time minus infinity. For construction of spatial point patterns using spatial birth-and-death processes one censors births of a fixed constant-rate birth-and-death process, which can be viewed as a *space-time Boolean model* D of time-like segments of exponentially distributed length (the segments thus playing the part of the innovations process $\{Z(t) : t \in \mathbb{Z}\}$). In the case of area-interaction and the Strauss process it is conceptually convenient to fatten out the segments to be cylinders of radius r, corresponding to the radius of interaction.

In the attractive case (as for example (1.3) for $\gamma > 1$) minimal and maximal processes $X^{\min}_n$, $X^{\max}_n$ are formed by starting at time $-n$ at respectively the vacant pattern $\emptyset$ and the full pattern of the time-slice of D at $-n$. The simulation algorithm extends backwards in time till $X^{\min}_n(0) = X^{\max}_n(0)$, which is then the required sample.

The non-attractive case (as for example (1.2), or (1.3) for $\gamma < 1$) needs just a minor modification: points are added to $X^{\max}$ according to a rule computed using $X^{\min}$ and *vice versa*. This simple crossover suffices to ensure that $X^{\min}_n$, $X^{\max}_n$ sandwich all relevant realizations of the target process including the virtual simulation from time minus infinity. Thus CFTP is applicable to point processes, even though the state space is infinite and non-uniform, and even when the processes are not attractive.

Further details, including consideration of Metropolis-Hastings algorithms, will appear in Kendall & Møller (1997). It should also be noted that Häggström, Lieshout & Møller (1996) describe CFTP perfect simulation using a different approach, tied to the attractive case of area-interaction and related models, but exploiting special features to attain greater efficiency.

5 Perfect simulation in space

Perfect simulation can be used in favourable cases to deal with sampling from equilibrium in space as well as time. To be definite, consider the simplest case: how to sample from the area-interaction process (1.3) defined over the whole of the plane $\mathbb{R}^2$, in the attractive case $\gamma > 1$.

Suppose that the required parameter $\gamma > 1$ is not too large (in a sense to be made precise below). One can simulate perfectly that part of such a process which lies in a bounded square window $W_0 = [-1, 1] \times [-1, 1]$, as follows. First perfectly simulate the process as observed within $W_1 = [-2, 2] \times [-2, 2]$ (a) subject to the (minimal) boundary conditions that no point of the process ever lies in $W_1 \oplus \text{ball}(0, r) \setminus W_1 = \{x + y : x \in W_1, y \in \text{ball}(0, r)\} \setminus W_1$, (b) subject to the (impossible, but maximal) boundary conditions that every point on the boundary ∂W_1 belongs to the process for all time, but in both cases using the same space-time Boolean model of cylinders. If these two simulations agree on W_0 then return the agreed pattern of points in W_0. Otherwise extend until maximal and minimal simulations agree on W_0: for step $n > 1$ suppose one has already simulated over $W_{n-1} = [-2^{n-1}, 2^{n-1}] \times [-2^{n-1}, 2^{n-1}]$, then one extends the common space-time Boolean model to the space $W_n = [-2^n, 2^n] \times [-2^n, 2^n]$ and uses this to perfectly simulate X (a) subject to minimal boundary conditions, (b) subject to maximal boundary conditions. One returns the agreed pattern of points on W_0 if such agreement exists, otherwise proceeds to step $n + 1$.

Agreement will certainly be achieved for some (possibly large) n if the underlying space-time Boolean model of cylinders does not percolate to infinity, since in that case for large enough n the boundary conditions at ∂W_n cannot affect the point pattern at W_0. If parameter $\gamma > 1$ is small enough then almost surely this percolation will not occur, and then the algorithm will work.

In fact the doubling approach sketched here (W_n of semi-side-length 2^n) is not the most efficient choice; a heuristic argument based on extreme-value theory shows that the amount of work for a perfect simulation in a square of side-length ℓ is proportional to around $\ell^2(\log \ell)^2$. Consequently binary search arguments suggest $\tilde{W}_n = [-2^{n/2}, 2^{n/2}] \times [-2^{n/2}, 2^{n/2}]$ will be better.

This particular topic, as well as its application to edge-effect correction using the EM algorithm/missing data approach of Geyer (1997), will be investigated further in subsequent work.

Bibliography

Baddeley, A.J. & M.N.M. van Lieshout (1995), "Area-interaction point processes," *Ann. Inst. Statist. Math.* **47**, 601–619.

Bernardo, J.M., J.O. Berger, A.P. Dawid & A.F.M. Smith (1992), *BAYESIAN STATISTICS 4*, Oxford Univ. Press, New York, NY.

Fill, J. (1997), *An algorithm, unbiased for user impatience, for exact sampling via Markov Chains*, Johns Hopkins University Dept Math Sciences, WWW reference: `http://www.mts.jhu.edu/~fill/`.

Foss, S.G. & R.L. Tweedie (1997), "Perfect Simulation and Backward Coupling," *Preprint*.

Geyer, C.J. (1997), "Markov Chain Monte Carlo," in *Proceedings Séminaire Européen de Statistique, "Stochastic geometry, computation and likelihood"*, O.E. Barndorff-Nielsen, W.S. Kendall & M.N.M. van Lieshout, eds., Chapman and Hall, New York, NY.

Gilks, W.R., S. Richardson & D.J. Speigelhalter (1995), *Markov Chain Monte Carlo in Practice*, Chapman and Hall, New York, NY.

Häggström, O., M.N.M. van Lieshout & J. Møller (1996), "Characterisation results and Markov chain Monte Carlo algorithms including exact simulation for some spatial point processes," *Aalborg Mathematics Department Research Report* **R-96-2040**.

Johnson, V. (1996), "Studying convergence of Markov Chain Monte Carlo Algorithms using coupled sample paths," *J. Amer. Statist. Assoc.* **91**, 154–166.

Kendall, W.S. (1997), "On some weighted Boolean models," in *Advances in Theory and Applications of Random Sets*, D. Jeulin, ed., World Scientific Publishing Company, Singapore, 105–120.

Kendall, W.S. (1998), "Perfect simulation for the area-interaction point process," in *Probability Perspective*, C.C. Heyde & L. Accardi, eds., Springer-Verlag, New York–Heidelberg–Berlin, 218–234, Also Warwick Statistics Research Report 292 (1996).

Kendall, W.S. & J. Møller (1997), "Perfect simulation of point processes using Metropolis-Hastings algorithms," *In preparation*.

Møller, J. (1997), "Markov Chain Monte Carlo and Spatial Point Processes," in *Proceedings Séminaire Européen de Statistique, "Stochastic geometry, computation and likelihood"*, O.E. Barndorff-Nielsen, W.S. Kendall & M.N.M. van Lieshout, eds., Chapman and Hall, New York, NY.

Murdoch, D.J & P.J. Green (1997), *Exact Sampling from a Continuous State Space*, University of Bristol Dept Maths, WWW reference: `http://www.stats.bris.ac.uk/MCMC/`.

Propp, J.G. & D.B. Wilson (1996), "Exact sampling with coupled Markov chains and applications to statistical mechanics," *Random Structures and Algorithms* **9**, 223–252.

Strauss, D.J. (1975), "A model for clustering," *Biometrika* **63**, 467–475.

Summary: A survey is given of the ideas of perfect simulation involving Coupling From The Past (CFTP) as introduced by Propp & Wilson (1996), and its application to the perfect simulation of spatial point processes. A sketch is given of an extension to the perfect simulation of point processes defined over all $\mathbb{R}^2$.

Sommaire: Voici un exposé des idées de la simulation parfaite au moyen de "Coupling From The Past" (CFTP) selon Propp & Wilson (1996), et son application à la simulation des processus ponctuels en espace. Une indication est fournie du moyen de conduire la simulation parfaite des processus ponctuels étendus à travers tout $\mathbb{R}^2$.

Other University of Warwick Department of Statistics Research Reports
authored or co–authored by W.S. Kendall.

161: The Euclidean diffusion of shape.
162: Probability, convexity, and harmonic maps with small image I: Uniqueness and fine existence.
172: A spatial Markov property for nearest–neighbour Markov point processes.
181: Convexity and the hemisphere.
202: A remark on the proof of Itô's formula for C^2 functions of continuous semimartingales.
203: Computer algebra and stochastic calculus.
212: Convex geometry and nonconfluent Γ-martingales I: Tightness and strict convexity.
213: The Propeller: a counterexample to a conjectured criterion for the existence of certain convex functions.
214: Convex Geometry and nonconfluent Γ-martingales II: Well–posedness and Γ-martingale convergence.
216: *(with E. Hsu)* Limiting angle of Brownian motion in certain two–dimensional Cartan–Hadamard manifolds.
217: Symbolic Itô calculus: an introduction.
218: *(with H. Huang)* Correction note to "Martingales on manifolds and harmonic maps."
222: *(with O.E. Barndorff-Nielsen and P.E. Jupp)* Stochastic calculus, statistical asymptotics, Taylor strings and phyla.
223: Symbolic Itô calculus: an overview.
231: The radial part of a Γ-martingale and a non-implosion theorem.
236: Computer algebra in probability and statistics.
237: Computer algebra and yoke geometry I: When is an expression a tensor?
238: *Itovsn3*: doing stochastic calculus with *Mathematica*.
239: On the empty cells of Poisson histograms.
244: *(with M. Cranston and P. March)* The radial part of Brownian motion II: Its life and times on the cut locus.
247: Brownian motion and computer algebra (Text of talk to BAAS Science Festival '92, Southampton Wednesday 26 August 1992, with screenshots of illustrative animations).
257: Brownian motion and partial differential equations: from the heat equation to harmonic maps (Special invited lecture, 49th session of the ISI, Firenze).
260: Probability, convexity, and harmonic maps II: Smoothness via probabilistic gradient inequalities.
261: *(with G. Ben Arous and M. Cranston)* Coupling constructions for hypoelliptic diffusions: Two examples.
280: *(with M. Cranston and Yu. Kifer)* Gromov's hyperbolicity and Picard's little theorem for harmonic maps.
292: Perfect Simulation for the Area-Interaction Point Process.
295: On some weighted Boolean models.
296: A diffusion model for Bookstein triangle shape.
301: COMPUTER ALGEBRA: an encyclopaedia article.
308: Perfect Simulation for Spatial Point Processes (Invited lecture, 51st session of the ISI, Istanbul).

If you want copies of any of these reports then please mail your requests to the secretary of the Department of Statistics, University of Warwick, Coventry CV4 7AL, UK, or use email (`L.K.Butterwick@warwick.ac.uk`).