Bernoulli **18**(3), 2012, 1031–1041
DOI: 10.3150/10-BEJ350

Spatial birth–death swap chains

MARK HUBER

Department of Mathematics and Computer Science, 850 Columbia Ave., Claremont McKenna College, Claremont, CA 91711, USA. E-mail: mhuber@cmc.edu

Markov chains have long been used for generating random variates from spatial point processes. Broadly speaking, these chains fall into two categories: Metropolis–Hastings type chains running in discrete time and spatial birth–death chains running in continuous time. These birth–death chains only allow for removal of a point or addition of a point. In this paper it is shown that the addition of transitions where a point is moved from one location to the other can aid in shortening the mixing time of the chain. Here the mixing time of the chain is analyzed through coupling, and use of the swap moves allows for analysis of a broader class of chains. Furthermore, these swap moves can be employed in perfect sampling algorithms via the dominated coupling from the past procedure of Kendall and Møller. This method can be applied to any pairwise interaction model with repulsion. In particular, an application to the Strauss process is developed in detail, and the swap chains are shown to be much faster than standard birth–death chains.

Keywords: birth death process; coupling from the past; perfect simulation; spatial point processes; Strauss process; swap moves

1. Introduction

Spatial point processes are in wide use in statistical modeling (see [15] for an overview). Typically finite point processes are modeled as being absolutely continuous with respect to a Poisson point process. That is, they have a density $f(x)/c$ where $f(x)$ is an easily computable function but the normalizing constant c of the density is impractical to compute. A Monte Carlo algorithm gains information about $f(x)/c$ by studying random variates drawn from the distribution the density describes.

To obtain these variates, a Markov chain is built whose stationary distribution matches the target distribution. Metropolis–Hastings chains run in discrete time (see [9]), and the spatial birth–death chain approach of Preston [21] runs in continuous time. In [6] problems were given where the Metropolis–Hastings approach is faster than Preston's.

The drawback of these Markov chain Monte Carlo methods is that unless the mixing time of the Markov chain is known, the quality of the variates is suspect. Heuristics such as the autocorrelation test can prove that a chain has not mixed, but cannot establish the positive claim that a chain has mixed.

Perfect simulation algorithms solve this problem. They generate samples exactly from the desired distribution without the need to know the mixing time of a Markov chain. Kendall [18] showed how the coupling from the past (CFTP) idea of Propp and Wilson [22] could be used together with a spatial birth and death chain to obtain samples from area interaction processes. Kendall and Møller [19] showed how this method could be extended to any locally stable point process using a method they called *dominated CFTP*. They also considered perfect sampling

using Metropolis–Hastings chains, but restricted these chains to only adding or deleting a point at each step.

So [6] indicates that Metropolis–Hastings chains can beat continuous time chains, but [19] shows how to exactly sample using continuous time chains. The goal of this work is to introduce a new swap move to the continuous time chains that speeds up convergence, while still allowing for perfect simulation.

In Section 2 the theory behind spatial birth–death chains with the new swap move is developed, and an example of such a chain is given for the Strauss process. Section 3 reviews the use of dominated coupling from the past, and shows how the addition of swap moves fits into this protocol. Section 4 bounds the expected running time of the procedure for a restricted class of models.

2. Spatial point processes

Dyer and Greenhill [7] first introduced a swap move for hard core point processes in discrete spaces. In this section their method is extended to more general point processes.

For ease of exposition, we consider here point processes that do not contain multiple points. Let S be a separable measurable set, and λ be a diffuse measure on S (so $\lambda(\{v\}) = 0$ for all $v \in S$) such that $\lambda(S) < \infty$. (Typically S is a bounded Borel set of $\mathbf{R}^2$.) Then a Poisson point process is a finite subset of S chosen as follows. First, let N be a Poisson distributed random variable with parameter $\lambda(S)$. Next, let $X_1, \ldots, X_N$ be independently and identically distributed (i.i.d.) and drawn from the probability distribution $\lambda(\cdot)/\lambda(S)$. Then $\{X_1, \ldots, X_N\}$ (called a *configuration*) is a draw from a Poisson point process with intensity measure $\lambda(\cdot)$ over S. Let μ be the distribution of the configuration and Ω the set of all possible configurations. More details of μ and Ω can be found in [4,21].

As an example of data modeled using these types of processes, Harkness and Isham [12] studied locations of ant nests in a rectangular region R. With two types of ants, $S = R \times \{0, 1\}$ and λ is the product of Lebesgue and a measure on $\{0, 1\}$.

The processes considered here are absolutely continuous with respect to μ with density f satisfying a *local stability condition* (as in [19]):

$$(\exists K > 0)(\forall x \in \Omega)(\forall v \in S \setminus x)\big(f(x \cup \{v\}) \leq K f(x)\big). \tag{1}$$

Many point processes of interest meet this condition, including the area interaction process [2, 25], the Strauss process [17,24] and the continuous random cluster model [11].

2.1. Spatial birth–death swap chains

The development of the swap move given here follows the framework of Preston [21], who introduced the use of spatial birth–death chains for these problems. These chains are examples of *jump processes*, where at a given state x, the chain stays in the state for an exponential length of time with expected value given by $1/\alpha(x)$. The state then jumps to a new state using kernel $\mathbf{K}$,

so the probability that the new state is in A is $\mathbf{K}(x, A)$), independent of the past history (see [8], Chapter X, for the details of jump processes).

In the Preston framework, the rate of births (addition of points to the configuration) and deaths (deletion of points from the configuration) depends only on the current state:

- There exists a non-negative measurable birth rate function b from $\Omega \times S$ equipped with the standard product σ-field to $\mathbf{R}$ with the Borel σ-field. Call $b(x, v)$ the *birth rate* at which point v is added to configuration x.
- There exists a non-negative measurable death rate function d from $\Omega \times S$ equipped with the standard product σ-field to $\mathbf{R}$ with the Borel σ-field. Furthermore, $w \in x \Rightarrow d(x, w) > 0$ and $w \notin x \Rightarrow d(x, w) = 0$. Then $d(x, w)$ is the *death rate* at which a point w is removed from configuration x.

To this birth–death framework we now add a swap rate:

- There exists a non-negative measurable swap rate function s from $\Omega \times S \times S$ equipped with the standard product σ-field to $\mathbf{R}$ with the Borel σ-field. Furthermore, $w \notin x \Rightarrow s(x, w, v) = 0$. So $s(x, w, v)$ is the *swap rate* at which point w is removed and point v is added.

The birth, death, and swap rates are used to build a kernel $\mathbf{K}$ for the Markov chain as follows. For all $A \in \mathcal{B}$, let $K_b(x, A) = \int_{v \in S} b(x, v)\mathbf{1}(x \cup \{v\} \in A)\lambda(\mathrm{d}v)$. When $K_b(x, \Omega) < \infty$ for all x in Ω, the birth kernel is $\mathbf{K}_b(x, A) = K_b(x, A)/K_b(x, \Omega)$. Similarly, $K_d(x, A) = \sum_{w \in x} d(x, w)\mathbf{1}(x \setminus \{w\} \in A)$, which always has a finite number of terms and so $\mathbf{K}_d(x, A) = K_d(x, A)/K_d(x, \Omega)$. The total rate of births is $r_b(x) = \int_{v \in S} b(x, v)\lambda(\mathrm{d}v)$, and the total rate of deaths is $r_d(x) = \sum_{v \in x} d(x, v)$.

For the swap kernel, set $K_s(x, A) = \sum_{w \in x} \int_{v \in S} s(x, w, v)\mathbf{1}(x \cup \{v\} \setminus \{w\} \in A)\lambda(\mathrm{d}v)$. When $K_s(x, \Omega) < \infty$ for all $x \in \Omega$, let

$$\mathbf{K}_s(x, A) = K_s(x, A)/K_s(x, \Omega), \qquad r_s(x) = \sum_{w \in x} \int_{v \in S} s(x, w, v)\lambda(\mathrm{d}v). \tag{2}$$

The overall rate at which the configuration changes is $\alpha(x) = r_b(x) + r_d(x) + r_s(x)$, and the overall kernel is:

$$\mathbf{K}(x, A) = \mathbf{K}_b(x, A)\frac{r_b(x)}{\alpha(x)} + \mathbf{K}_d(x, A)\frac{r_d(x)}{\alpha(x)} + \mathbf{K}_s(x, A)\frac{r_s(x)}{\alpha(x)}. \tag{3}$$

Harris recurrence guarantees that a Markov process has a unique invariant measure (see [1] for details of Harris recurrence in the continuous-time context). Kaspi and Mandelbaum [16] showed that a continuous-time chain is Harris recurrent if and only if there exists a non-zero σ-finite measure where X almost surely hits sets with positive measure.

In particular, for all the chains considered here, the death rate equals the number of points in the configuration, and the birth rate is bounded above by a constant. This forces the chain to visit the empty configuration infinitely often, making it Harris recurrent.

The detailed balance conditions (that imply f is invariant) for jump processes are: $f(x)\alpha(x) \times \mathbf{K}(x, \mathrm{d}y)\,\mathrm{d}\mu(x) = f(y)\alpha(y)\mathbf{K}(y, \mathrm{d}x)\,\mathrm{d}\mu(y)$. For moves from configurations with n points to

those with $n+1$ (or vice versa), the detailed balance conditions are satisfied [21,23] when the rate of births balance the rate of deaths with respect to f. So

$$f(x)b(x,v) = f(x \cup \{v\})d(x \cup \{v\}, v). \tag{4}$$

Swap moves stay inside the same dimensional space, and it is straightforward to show that reversibility for swap moves holds when

$$f(x)s(x,w,v) = f(x \cup \{v\} \setminus \{w\})s(x \cup \{v\} \setminus \{w\}, v, w). \tag{5}$$

2.2. Locally stable repulsive point processes

Kendall and Møller [19] describe how to create a jump process with stationary density f for locally stable processes. Briefly, their method works as follows. Two coupled chains will be run: the *dominating chain* with state $D(t)$ at time t and the *target chain* with state $X(t)$ at time t. It will always be true that $X(t) \subseteq D(t)$. Each point $w \in D(t)$ has death rate $d(D(t), w) = 1$. If a point dies that is also in $X(t)$, it is removed from both $X(t)$ and $D(t)$. The rate of births for the dominating chain is $r_b = K\lambda(S)$, where K is the local stability constant in equation (1). If a birth occurs, a point v is chosen according to the probability measure $\lambda(\cdot)/\lambda(S)$. Then v is always added to $D(t)$ to get the next dominating state, but is only added to $X(t)$ with probability $f(X(t) \cup \{v\})/[Kf(X(t))]$. Assume that each point v born in $D(t)$ is marked with a uniform draw from $[0,1]$. Then the point is born in $X(t)$ if the mark falls below $f(X(t) \cup \{v\})/[Kf(X(t))]$.

Suppose $X(0) \subseteq D(0)$. Then since deaths are always accepted in both chains, but a birth in the dominating chain might not occur in the target chain, the dominating configuration will be a superset of the target configuration for all $t \geq 0$.

Adding a swap move to this birth death framework can be done automatically when the rejection of a birth v can be linked to a single point $w \in X(t)$. Consider an example.

Strauss model

In the Strauss model [17,24], the density has a factor that is exponential in the number of pairs of points that lie within distance R of each other. Let ρ be a metric on S (usually Euclidean distance), then the density can be written:

$$f_S(x) = Z^{-1}_{(\beta_1,\beta_2,R)}\beta_1^{\#x}\beta_2^{s(x)}, \qquad s(x) = \sum_{\{v,v'\}:v\in x, v'\in x\setminus\{v\}} \mathbf{1}\big(\rho(v,v') \leq R\big), \tag{6}$$

where $Z_{(\beta_1,\beta_2,R)}$ is the normalizing constant for the density. As noted in [17], in order for $Z_{(\beta_1,\beta_2,R)}$ to be finite (and hence for the density to exist) β_2 must be at most 1. In addition, [17] generalizes the Strauss process to the pairwise interaction process. All methods presented here are written for the Strauss process for simplicity, but work equally well for the pairwise interaction process.

Let x be the state of the target chain, and suppose point v is born in the dominating chain. Call point $w \in x$ a neighbor of v if $\rho(v,w) \leq R$. The Strauss process is locally stable with

$K = \beta_1$, so the chance of accepting v into x is $f(x \cup \{v\})/[Kf(x)] = \beta_2^{s(x,v)}$, where $s(x, v) = \sum_{v' \in x} \mathbf{1}(\rho(v', v) \leq R)$ is the number of neighbors of v in x.

Let $\mathsf{Bern}(p)$ denote the Bernoulli distribution with parameter p. One way to draw $B \sim \mathsf{Bern}(\beta_2^{s(x,v)})$ is to draw $B_1, \ldots, B_{s(x,v)} \overset{\text{i.i.d.}}{\sim} \mathsf{Bern}(\beta_2)$ and set $B = B_1 B_2 \cdots B_{s(x,v)}$. (Here $\overset{\text{i.i.d.}}{\sim}$ denotes that the draws are to be independent and identically distributed.)

When $B_i = 0$, say that the point indexed by i *blocks* the birth of v. Suppose that v is blocked by a single neighbor w. Then the swap move removes w, and allows the birth of v. Call this new configuration x'. The probability of swapping from x to x' (given birth v) is $\beta_2^{s(x,v)-1}(1 - \beta_2)$. This makes it straightforward to check that $f(x)s(x, w, v) = f(x')s(x', v, w)$, so (5) is satisfied. To implement this swap move, simply mark each point v born in $D(t)$ with an i.i.d. sequence of $\mathsf{Bern}(\beta_2)$ random variables.

3. Perfect simulation by dominated CFTP

In the previous section it was shown how to couple a dominating chain and target chain using standard birth–death chains and the new birth–death swap chain. Here a further coupling is built that allows exact draws to be taken from the stationary distribution of the target chain using the dominating CFTP (dCFTP) method of Kendall and Møller [19].

Both $X(t)$ and $D(t)$ are time-reversible, so they can be run backwards in time as easily as forwards while maintaining the property that if $X(0) \subseteq D(0)$, then $X(t) \subseteq D(t)$ for all $t \in (-\infty, 0]$. (A more detailed introduction to dCFTP can be found in [19].)

So far two chains (the dominating and target) have been coupled, but now consider two more chains, called the *lower chain* and *upper chain*, denoted $L(t)$ and $U(t)$, respectively. Suppose that these four chains have the *sandwiching property* that

$$L(t) \subseteq X(t) \subseteq U(t) \subseteq D(t) \qquad \text{for all } t \in (-\infty, 0]. \tag{7}$$

The process $(L(t), U(t))$ can also be thought of as a bounding process for $X(t)$ (see [14]).

Suppose $X(0)$ is drawn from the stationary distribution. Then if $L(0) = U(0)$, $X(0)$ also equals the lower and upper chain, and the state they all equal is a draw from the stationary distribution. This is the idea behind CFTP.

For each positive integer N, a lower and upper chain can be created. Consider $D(t)$ moving backward through time, and let τ_N denote the time where the Nth backward event occurs. Set $L_N(\tau_N)$ to the empty configuration, and $U_N(\tau_N) = D(\tau_N)$.

Every time there is an event at time t (either a birth or death in the dominating process moving forwards in time) it is important to ensure that $U_N(t)$ and $L_N(t)$ continue to bound $X(t)$ once the event updates the chain. That is, if a point v is added to the target chain state, it must also be added to the upper chain. If a point w is removed from the target chain state, it must also be removed from the lower chain. Such a coupling has the *funneling property* (see [3]). All the couplings used here have this important property.

An induction argument shows that the funneling property implies $L_N(0) \subseteq X(0) \subseteq U_N(0)$. Note if $L_N(0) = U_N(0)$, then $X(0)$ is trapped between them and also equals this common value. This is the coupling part of CFTP.

The "from the past" part of CFTP works as follows. Suppose $L_N(0) \neq U_N(0)$. Then increase the value of N and try again. Let $N' > N$. The first N events for the dominating process (looking backward in time from time 0) have already been generated, these same events must be used in subsequent evaluations of the bounding process. Therefore, only $N' - N$ additional events need to be generated. Once these events have been generated, run $L_{N'}$ and $U_{N'}$ forward until $L_{N'}(0)$ and $U_{N'}(0)$ can be compared.

If $L_N(0) = U_N(0)$ for some N, then $L_{N'}(0) = U_{N'}(0)$ for all $N' > N$ as well, so it is not necessary to try every value of N. Propp and Wilson [22] noted that by doubling N at each step, the total number of checked events is at most twice the minimum number. The choice of N_{initial} is arbitrary, but $L_N(0)$ cannot equal $U_N(0)$ unless every point in $D(\tau_N)$ has died by time 0. For simplicity, here N_{initial} is set equal to the expected number of points in the dominating process at time 0, which is $K\lambda(S)$ (see [3] for a more advanced approach to choosing N_{initial}).

Kendall and Møller showed (Theorem 2.1 of [19]) that as long as the probability that $D(t)$ visits the empty configuration in $[0, t]$ goes to 1 as t goes to infinity, this procedure will terminate in finite time with probability 1. The resulting configuration $L_N(0) = U_N(0)$ is a draw exactly from the target distribution.

Now consider the question: How should the lower and upper chains be updated for each event in the dominating process so the funneling property holds for the swap move?

3.1. Updating the bounding process

For a jump process $A(t)$, let $A(t-)$ denote the limit as ε goes to 0 of $A(t - \varepsilon)$, that is, the state of the process right before time t. The bounding process needs to be updated if a point is born or dies at time t. The procedure followed is the same as given in [14].

If a point $w \in X(t-)$ dies, it is removed from $X(t)$, and so can be removed from both $L_N(t)$ and $U_N(t)$. Now suppose point v is born into the dominating chain at time t.

Case 1: Point v is blocked by at most one point w in $U_N(t-)$. Then $X(t-) \subseteq U_N(t-)$ and so if $w \in X(t)$, then w is swapped away by v, and if $w \notin X(t-)$, then v can be born. So either way $X(t) = X(t-) \setminus \{w\} \cup \{v\}$, w is removed from $U_N(t)$ (and $L_N(t)$ if it is there also) and v is added to both $L_N(t)$ and $U_N(t)$.

Case 2: The point v is blocked by at least two points in $L_N(t-)$. Then there are at least two blocking points in $X(t-)$, so the birth does not occur in $L_N(t)$, $X(t)$ or $U_N(t)$.

Case 3: the point v is blocked by at most one point in $L_N(t-)$, and at least two points in $U_N(t-)$. Then if $X(t-)$ contains the two blocking points in $U_N(t-)$, the swap does not occur, but if it only contains the single blocking point in $L_N(t-)$, the swap does occur. The result is that the birth v must be added to $U_N(t)$ (but not to $L_N(t)$) to ensure $X(t) \subseteq U_N(t)$, and any blocking point in $L_N(t-)$ must be removed from $L_N(t)$.

Figure 1 shows the running time advantage gained by using the swap move. The times are measured in number of events generated by dominated CFTP (dCFTP). On the left are the raw number of times for the chain without the swap move and with the swap move. The plot on the right shows the ratio of these two times. Note that as β_1 gets larger, the speedup gained by using the swap move also increases.

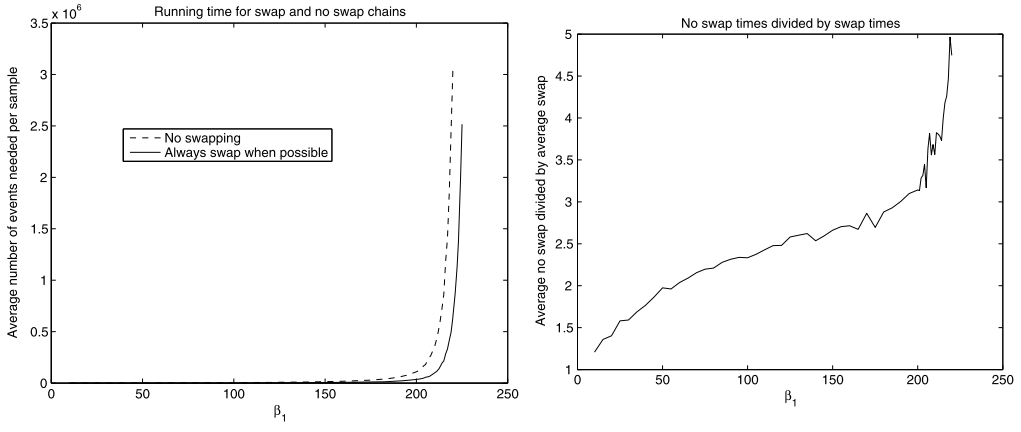

Figure 1. Running time of dCFTP for Strauss model on $S = [0, 1]^2$, $\beta_2 = 0.5$, $R = 0.05$, λ is Lebesgue measure.

4. Analyzing the running time

Consider how many events must be generated before the dominated coupling from the past procedure terminates, that is, before $U_N(0) = L_N(0)$. Deaths in $U_N(t) \setminus L_N(t)$ cause the bounding process to move together, while births can add a point to $U_N(t)$ but not to $L_N(t)$, and the swap move sometimes removes a point from $L_N(t)$ but not $U_N(t)$. Therefore, it is reasonable that the perfect simulation algorithm will run faster in situations where the birth rate is low.

In this section it is shown that, for perfect simulation of the Strauss process, the original no swap chain takes (with high probability) a small number of steps per perfect sample when β_1 and R are not too large, and β_2 is not too small. By creating a mixture of the swap chain and no swap chain, it is possible to improve this result to where it applies for values of β_1 that are twice as large as for the no swap chain.

The mixture works as follows: At each step, with probability p_{swap}, the swap move chain is used, while with probability $1 - p_{\text{swap}}$, the original no swap chain is used. The best theoretical bound is achieved when $p_{\text{swap}} = 1/4$.

Theorem 4.1. *Suppose that N events are generated backwards in time and then run forward to get $U_N(0)$ and $L_N(0)$. Let $B(v, R)$ be the area within distance R of $v \in S$, let $r = \sup_{v \in S} \lambda(B(v, R))$.*

If $\beta_1(1 - \beta_2)r < 1$, then for the chain without the swap move

$$\mathbf{P}\big(U_N(0) \neq L_N(0)\big) \leq 2\exp(-0.09N) + \beta_1\lambda(S)\exp\big(-N\big(1 - \beta_1(1 - \beta_2)r\big)/(4\beta_1\lambda(S))\big). \quad (8)$$

If $\beta_1(1 - \beta_2)r < 2$, then for the chain where a swap is executed with probability $1/4$,

$$\begin{aligned}\mathbf{P}\big(U_N(0) \neq L_N(0)\big) &\leq 2\exp(-0.09N) \\ &\quad + \beta_1\lambda(S)\exp\big(-N\big(1 - 0.5\beta_1(1 - \beta_2)r\big)/(4\beta_1\lambda(S))\big).\end{aligned} \quad (9)$$

Why the value of $1/4$ for the probability? This is an artifact of the proof technique. The theorem only gives sufficient, not necessary, conditions for the algorithm to be fast, and simulation experiments indicate that the algorithm actually takes the fewest steps when the swap moves are used as often as possible (reasons why this could be true are noted below in the proof of the theorem).

Theorem 4.1 has immediate consequences for the expected running time of dominated CFTP. Recall that in dCFTP the number of events was doubled each time. Say $\mathbf{P}(U_N(0) \neq L_N(0)) \leq a\exp(-bN)$, and let T be the number of events generated in a call of dCFTP. Then for $T \geq t$, dCFTP must have failed on a run of length at least $t/2$. So

$$\mathbf{E}[T] = \sum_{N=1}^{\infty} \mathbf{P}(T \geq N) \leq \left[\sum_{N=1}^{\lceil (2/b)\ln a \rceil} 1 \right] + \sum_{N=\lceil (2/b)\ln a \rceil}^{\infty} a\exp(-bN/2), \tag{10}$$

which makes $\mathbf{E}[T] = \mathrm{O}(\ln a/b)$, and the mean running time $\mathrm{O}(\beta_1\lambda(S)(\ln \beta_1\lambda(S)))$ for the no swap chain when $\beta_1(1-\beta_2)r < 1$ and in the $1/4$-swap chain when $\beta_1(1-\beta_2)r < 2$.

Proof of Theorem 4.1. Recall $U_N(\tau_N) = D(\tau_N)$, a Poisson spatial point process with parameter $\beta_1\lambda(S)$. $L_N(\tau_N)$ is the empty configuration, and the bounding processes are run forward in time. Let $Q(t) = U_N(t) \setminus L_N(t)$. Then the chains have come together if and only if $\#Q(0) = 0$. Begin by considering the no swap chain.

Strauss no swap move. All individual death rates are 1, so the total rate of deaths of points in $Q(t)$ is just $\#Q(t)$. Call a death a *good event* since it reduces $\#Q(t)$ by 1.

For $\#Q(t)$ to increase by 1 (call this a *bad event*), a birth must occur at v and be added to $U_N(t)$ but not $L_N(t)$. Let w be any point in $Q(t)$. Then for $Q(t)$ to give rise to another point in $Q(t)$, a point v must be born within distance R of w and the $\mathsf{Bern}(\beta_2)$ draw must be 0. The area surrounding w is at most r, and the Bernoulli draw acts as a thinning procedure in a Poisson process (see Appendix G of [20].) So the rate at which w creates new points in $Q(t)$ is at most $\beta_1(1-\beta_2)r$, and the overall rate of bad events is at most $\beta_1(1-\beta_2)r\#Q(t)$.

Suppose the rate of bad events is smaller than the rate of good events. The probability that one event occurs in the time interval from t to $t+h$ is proportional to h, the probability that n events occurs is $O(h^n)$. Hence

$$\mathbf{E}\big[\mathbf{E}[\#Q(t+h)|U(t), L(t)] - \#Q(t)\big] \leq \mathbf{E}\left[\big(\#Q(t)\beta_1(1-\beta_2)r - \#Q(t)\big)h + \sum_{i=2}^{\infty} i\mathrm{O}(h^i) \right],$$

which means

$$\lim_{h\to 0} \frac{\mathbf{E}[\mathbf{E}[\#Q(t+h)|U(t), L(t)] - \#Q(t)]}{h} \leq -\mathbf{E}\big[\#Q(t)\big(1-\beta_1(1-\beta_2)r\big)\big].$$

Let $q(t) = \mathbf{E}[\#Q(t)]$, and let τ_N be the time of the Nth event moving backwards in time. Then $q(\tau_N) \leq \mathbf{E}[\#D(\tau_N)] = \beta_1\lambda(S)$, so together with $q'(t) \leq -q(t)(1-\beta_1(1-\beta_2)r)$:

$$q(t) \leq \beta_1\lambda(S)\exp\big(-t\big(1-\beta_1(1-\beta_2)r\big)\big).$$

By Markov's inequality, $\mathbf{P}(Q(0) \neq \varnothing) = \mathbf{P}(\#Q(0) \geq 1) \leq q(0)$.

Now fix N, the number of events to run back in time, and set $t = N/[4\beta_1\lambda(S)]$. The chance $Q(0)$ does not equal 0 starting at $-t$ is at most $\exp(-N/[4\beta_1\lambda(S)](1-\beta_1(1-\beta_2)r))$.

Using Chernoff bounds [5], it can be shown that for $A \sim \mathsf{Pois}(\alpha)$, $\mathbf{P}(A > 2\alpha) \leq \exp(-\alpha(2\ln 2 - 2 + 1))$. So after t time, the probability that more than $N/2$ events were generated in a Poisson process with rate $\beta_1\lambda(S)$ is at most $\exp(-(N/4)(2\ln 2 - 2 + 1))$. Both the times of the births and times of deaths (viewed individually) are Poisson processes with rate $\beta_1\lambda(S)$, therefore the probability that either uses more than $N/2$ events (by the union bound) is at most $2\exp(-0.09N)$. But if at this time each process used at most $N/2$ events, then moving back in time N events puts the user even farther back in time, and if coalescence occurs at $-t$, it will also occur starting at τ_N. Again using the union bound, the probability of failure is at most

$$2\exp(-0.09N) + \exp\bigl(-N\bigl(1-\beta_1(1-\beta_2)r\bigr)\bigr).$$

Strauss with swap move. Now consider what happens when $p_{\text{swap}} > 0$. The rate of good events (deaths) remains unchanged, but the rate of bad events changes. In Section 3.1, Case 1 leaves $\#Q(t)$ unchanged or reduces it by 1, Case 2 leaves $\#Q(t)$ unchanged, and Case 3 increases $\#Q(t)$ by 1 or 2. To be precise, let A_L be the set of blocking points in $L_N(t-)$, and A_U be the set of blocking points in $U_N(t-)$. Then the situations that change $\#Q(t)$ are:

Type	$\#A_U$	$\#A_L$	$\#Q(t) - \#Q(t-)$ no swap	$\#Q(t) - \#Q(t-)$ with swap
1	1	0	1	−1
2	at least 2	1	0	2
3	at least 2	0	1	1

Let b_1 denote the area of the region where a birth is Type 1, with b_2 and b_3 defined similarly. Together, the rate of change from births is:

$$b_1[(1-p_{\text{swap}}) - p_{\text{swap}}] + b_2[2p_{\text{swap}}] + b_3[(1-p_{\text{swap}}) + p_{\text{swap}}].$$

Any point in b_3 neighbors at least two points in $\#Q(t-)$, and points in b_1 or b_2 neighbor at least one. Each point in $Q(t-)$ has r area adjacent to it, so $b_1 + b_2 + 2b_3 \leq \#Q(t)r$.

The variable p_{swap} can be set to any number from 0 to 1: letting $p_{\text{swap}} = 1/4$ gives an upper bound on the bad event rate of $(1/2)b_1 + (1/2)b_2 + b_3 \leq (1/2)\#Q(t)r$.

Recall the bad event rate when $p_{\text{swap}} = 0$ was bounded above by $\#Q(t)r$. With $p_{\text{swap}} = 1/4$, the bad event rate is bounded above by $\#Q(t)r/2$, and this factor of two carries throughout the remainder of the proof to give (9). □

Häggström and Steif gave a result similar to the previous theorem for finitary codings for high noise Markov random fields [10], but their analysis involves moving backwards rather than forwards in time, and their result does not employ the swap move.

Figure 2 illustrates the mean run time for a fixed value of λ as the probability of a swap varies from $p = 0$ up to $p = 1$. The running time (as measured by generated iterations) decreases as the chance of swapping increases. This same phenomenon was noted for hard core gas models on graphs [13], and at present is unexplained by theory.

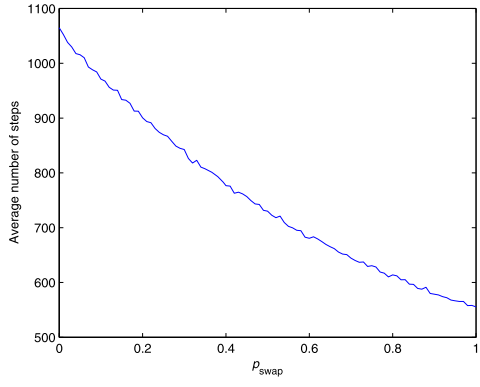

Figure 2. Running time of dCFTP for Strauss model on $S = [0, 1]^2$, $\beta_1 = 50$, $\beta_2 = 0.5$, $R = 0.05$, λ is Lebesgue measure, as p_{swap} runs from 0 to 1.

5. Conclusions

The regular birth–death chains only move when no point blocks the birth of a point in the dominating process. The birth–death swap chains move when at most one point blocks the birth of a point in the dominating process. This alone means that more moves are being taken, and helps to explain the improved analysis and improved performance when used for perfect sampling with dominated coupling from the past.

References

[1] Azéma, J., Kaplan-Duflo, M. and Revuz, D. (1967). Mesure invariante sur les classes récurrentes des processus de Markov. *Z. Wahrsch. Verw. Gebiete* **8** 157–181. MR0222955

[2] Baddeley, A.J. and van Lieshout, M.N.M. (1995). Area-interaction point processes. *Ann. Inst. Statist. Math.* **47** 601–619. MR1370279

[3] Berthelsen, K.K. and Møller, J. (2002). A primer on perfect simulation for spatial point processes. *Bull. Braz. Math. Soc. (N.S.)* **33** 351–367. MR1978833

[4] Carter, D.S. and Prenter, P.M. (1972). Exponential spaces and counting processes. *Z. Wahrsch. Verw. Gebiete* **21** 1–19. MR0310967

[5] Chernoff, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statistics* **23** 493–507. MR0057518

[6] Clifford, P. and Nicholls, G. (1994). Comparison of birth-and-death and Metropolis–Hastings Markov chain Monte Carlo for the Strauss process. Unpublished manuscript.

[7] Dyer, M. and Greenhill, C. (2000). On Markov chains for independent sets. *J. Algorithms* **35** 17–49. MR1747721

[8] Feller, W. (1966). *An Introduction to Probability Theory and Its Applications. Vol. II*. New York: Wiley. MR0210154

[9] Geyer, C. (1999). Likelihood inference for spatial point processes. In *Stochastic Geometry* (*Toulouse*, 1996). *Monogr. Statist. Appl. Probab.* **80** 79–140. Boca Raton, FL: Chapman & Hall/CRC. MR1673118

[10] Häggström, O. and Steif, J.E. (2000). Propp-Wilson algorithms and finitary codings for high noise Markov random fields. *Combin. Probab. Comput.* **9** 425–439. MR1810150
[11] Häggström, O., van Lieshout, M.C.N.M. and Møller, J. (1999). Characterization results and Markov chain Monte Carlo algorithms including exact simulation for some spatial point processes. *Bernoulli* **5** 641–658. MR1704559
[12] Harkness, R.D. and Isham, V. (1983). A bivariate spatial point pattern of ants' nests. *Appl. Stat.* **32** 293–303.
[13] Huber, M. (2000). A faster method for sampling independent sets. In *Proceedings of the Eleventh Annual ACM-SIAM Symposium on Discrete Algorithms* (*San Francisco*, *CA*, 2000) 625–626. New York: ACM. MR1755521
[14] Huber, M. (2004). Perfect sampling using bounding chains. *Ann. Appl. Probab.* **14** 734–753. MR2052900
[15] Illian, J., Penttinen, A., Stoyan, H. and Stoyan, D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*. Chichester: Wiley. MR2384630
[16] Kaspi, H. and Mandelbaum, A. (1994). On Harris recurrence in continuous time. *Math. Oper. Res.* **19** 211–222. MR1290020
[17] Kelly, F.P. and Ripley, B.D. (1976). A note on Strauss's model for clustering. *Biometrika* **63** 357–360. MR0431375
[18] Kendall, W.S. (1998). Perfect simulation for the area-interaction point process. In *Probability Towards the Year* 2000 (L. Accardi and C. C. Heyde, eds.) 218–234. Springer, New York.
[19] Kendall, W.S. and Møller, J. (2000). Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes. *Adv. in Appl. Probab.* **32** 844–865. MR1788098
[20] Møller, J. and Waagepetersen, R.P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. *Monographs on Statistics and Applied Probability* **100**. Boca Raton, FL: Chapman & Hall/CRC. MR2004226
[21] Preston, C.J. (1977). Spatial birth-and-death processes. *Bull. Inst. Int. Stat.* **46** 371–391. MR0474532
[22] Propp, J.G. and Wilson, D.B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures Algorithms* **9** 223–252. MR1611693
[23] Ripley, B.D. (1977). Modelling spatial patterns (with discussion). *J. Roy. Statist. Soc. Ser. B* **39** 172–212. MR0488279
[24] Strauss, D.J. (1975). A model for clustering. *Biometrika* **62** 467–475. MR0383493
[25] Widom, B. and Rowlinson, J.S. (1970). A new model for the study of liquid-vapor phase transition. *J. Chem. Phys.* **52** 1670–1684.

Received June 2008 and revised November 2010