Exact Draws from the Stationary Distribution of Entry-Exit Models*

Takashi Kamihigashi[a] and John Stachurski[b]

[a]*Research Institute for Economics and Business Administration, Kobe University*

[b]*Research School of Economics, Australian National University*

September 21, 2012

Abstract

In entry and exit models of firm dynamics, the stationary equilibrium distribution of firms summarizes the predictions of the model for any given set of primitives. Typically this distribution has no analytical solution, and numerical methods are necessary to investigate its properties. To date, all of these methods have involved some form of approximation. Focusing on Hopenhayn's seminal model of firm dynamics with entry and exit (*Econometrica*, 60:5, 1992, p. 1127–1150), we provide a coupling algorithm that generates *exact* draws from the stationary distribution in finite time. We show that the technique is able to rapidly generate large numbers of exact and independent draws in applications.

Keywords: Simulation, stationary equilibrium, firm dynamics
Journal of Economic Literature Classifications: C61, C63

1 Introduction

In this paper we provide a method for generating exact, IID draws from the stationary distribution of Hopenhayn's (1992) model of firm dynamics with entry and exit. Hopenhayn's model now forms the foundations of a large and growing literature (see, for example, Hopenhayn and Rogerson, 1993; Cooley and Quadrini, 2001 or Melitz, 2003). The

*The second author acknowledges financial assistance from the ARC Discovery Outstanding Researcher Award DP120100321.

stationary distribution of the model represents a distribution of firms that is both consistent with the definition of equilibrium at any point in time and also invariant over time. For typical parameter values the stationary distribution is uniquely defined but has no analytical solution. To compare the model with data, it is necessary to either compute an approximation to the distribution or simulate from the distribution.

To compute the distribution itself typically involves some form of discretization, possibly with interpolation on finite grids. Alternatively, it is straightforward to set up an ergodic Markov process such that the invariant distribution of the Markov process coincides with the cross-sectional stationary distribution of firms from Hopenhayn's model. This permits approximate simulation from the stationary distribution simply by running the process from an arbitrary initial value until it is judged to have "nearly" converged. Simulating until the distribution of the state is approximately stationary is referred to as burn-in.

There are problems with both of these approaches. Regarding computation of the distribution by discretization, one difficulty is that even if one can show that a given method is globally convergent, the size of the error in the approximation at a given level of discretization and after a given number of iterations is hard to assess. The problem with the simulation method is somewhat similar: the length of burn-in required is the subject of guesswork and heuristics, and regardless of how much burn-in is performed, the resulting sample is never exactly stationary. Once again, the size of the error is largely unknown. Moreover, for a given method, the size of the error is likely to vary with the parameters since the parameters change the structure of the problem. This leads to uncertainty regarding accuracy of results and the potential for baised estimates.

In this paper we show that the simulation problem can be overcome for this class of models. By using a variation of Propp and Wilson's (1996) coupling from the past technique, we show that it is possible to generate *exact* independent draws form the stationary distribution of Hopenhayn's model. An additional and related benefit of the algorithm is that it terminates as soon as an exact draw has been generated. Hence there is no need for the heuristics used to judge burn-in time. By repeating the algorithm with independent seeds, it becomes possible to generate multiple independent draws from the stationary distribution. These processes can be run sequentially or in parallel. Even with sequential generation, the algorithm permits rapid generation of exact IID draws from the stationary distribution. Because the resulting sample is IID, the standard central limit theorem for independent observations can be used to give confidence intervals for moments or empirical distributions.

The coupling from the past method has been used successfully for models with large but discrete state spaces. Applications range from statistical mechanics to page ranking and the design of peer-to-peer file sharing systems (see chapter 2 of Levin *et al.*, 2008 for recent survey). In the case of continuous state spaces, applicability of the method is not auto-

matic, and depends on the particular structure of the model in question. Some ingenuity is usually required to exploit what structure is available. Murdoch and Green (1998) showed that coupling from the past can in theory be used in continuous state settings when the underlying process satisfies Doeblin's condition, although their method is not always efficent. Their work was extended by Foss and Tweedie (1998) and Athreya and Stenflo (2003). Corcoran and Tweedie (2001) showed how efficient perfect sampling could be achieved in the case where the Markov process is monotone increasing. Nishimura and Stachurski (2010) applied their idea to an incomplete markets model.

The technique presented in this paper is relatively innovative. The state space is continuous, and, while the Markov process for productivity of incumbents is monotone increasing, the overall process that combines dynamics of incumbents and new entrants is not. We develop a coupling from the past type method that exploits both the monotonicity of productivity for incumbents and the renewal feature introduced by new entrants.

2 Preliminaries

In this section we briefly review an exact sampling method that gives intuition for our technique, and recall the major features of Hopenhayn's (1992) model.

2.1 Coupling From the Past

There are many versions of coupling from the past. Here we briefly review a version for Markov chains on finite state spaces that coincides conceptually with the algorithm for entry-exit models developed below. To begin, consider an irreducible aperiodic Markov chain $\{X_t\}$ on finite state space S with unique stationary distribution μ. Suppose that the dynamics of the model are given by $X_{t+1} = g(X_t, W_{t+1})$, where $\{W_t\}$ is IID. If we start at any X_0 and simulate $\{X_t\}$ by drawing $W_1, W_2, \ldots$ and successively applying the maps $g(\cdot, W_1), g(\cdot, W_2), \ldots$, then, in view of aperiodicity and irreducibility, the distribution of X_t will converge to (but in general never reach) the stationary distribution μ. A second immediate observation is that if $\{X_t\}$ and $\{X'_t\}$ are two chains simulated using the *same sequence* of shocks $\{W_t\}$, and if $X_\tau = X'_\tau$ at some point in time τ, then $X_t = X'_t$ for all $t \geq \tau$. We say that $\{X_t\}$ and $\{X'_t\}$ *couple* at time τ.

Now consider an IID sequence of shocks $\{W_t\}_{t=-\infty}^0$ indexed on the nonpositive integers, and let $\{X_t\}_{t=-\infty}^0$ be the process generated from these shocks, starting in the infinite past and terminating at time $t = 0$. Intuitively, since the burn-in is infinite, the distribution of X_0 will be exactly μ. Since this infinite chain cannot be simulated, it appears that X_0 is not observable, but it turns out that this is not always the case. To understand how X_0

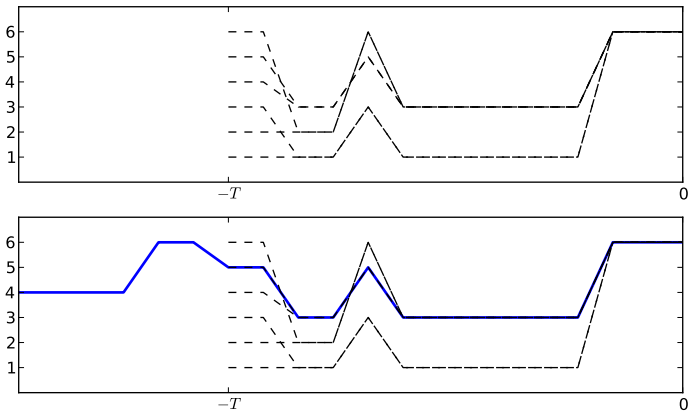

Figure 1: Coalescence of the tracking processes

can potentially be observed, suppose that we fix $T \in \mathbb{N}$, take the finite subset of shocks $W_{-T}, W_{-T+1}, \ldots, W_0$ and then, for each point x in the state space S, construct one version of the process $\{X_t\}$ that starts at this point x at time $-T$ and runs up until time zero. All of these processes are updated using the same shocks $W_{-T}, W_{-T+1}, \ldots, W_0$. We will call these processes *tracking processes*. Because the tracking processes are finite it is possible to simulate them. Two different visualizations of the tracking processes are shown in the top panel of figure 1 and figure 2 respectively. Here the state space is $\{1, \ldots, 6\}$. Notice that in the simulations, some of the processes couple and then run together up until time zero. In figure 1, all of the tracking processes have coupled by time zero, and only one final value is observed. If this is the case we will say that the tracking processes *coalesce*. In figure 2 multiple paths are still present at time zero, and the processes fail to coalesce.

Now consider again the process $\{X_t\}_{t=-\infty}^0$ generated from the entire sequence of shocks $\{W_t\}_{t=-\infty}^0$. Since this process must pass through one point x in the state space at $-T$, and since it receives the same shocks as the tracking processes from that time forwards, it must follow the same path as the tracking process that started at x. Of course we do not know which of the tracking processes it is following. However, *if* the tracking processes coalesce, then there is only *one* final observation. This observation must be a realization of the time zero value X_0 of the process $\{X_t\}_{t=-\infty}^0$ that started in the infinite past, and is therefore a draw from μ. Such an outcome is illustrated in the bottom panel of figure 1. For comparison, an unsuccessful outcome is illustrated in the bottom panel of figure 2. Here there are two final values, and we do not know which is the time zero value of the process $\{X_t\}_{t=-\infty}^0$. In this case we can take one step further back in time, drawing the additional shock W_{-T-1} while preserving the existing shocks $W_{-T}, W_{-T+1}, \ldots, W_0$, recalculate the tracking processes, test for coalescence, and so on. This procedure will eventually terminate with an exact draw from μ.

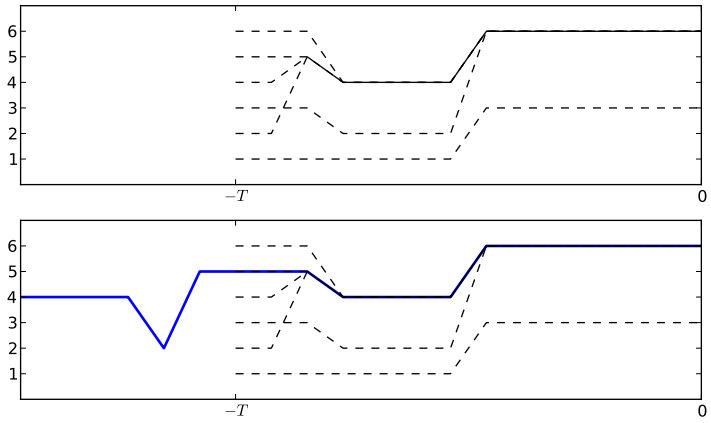

Figure 2: The tracking processes failing to coalesce

While this technique works when there are a finite number of states, it is clearly not possible to directly implement the same idea when the state space is infinite. However, variations on the idea can potentially be found by exploiting the structure of a given model. In the case of the entry-exit model studied below, we show how the monotonicity of productivity for incumbent firms combined with the renewal caused by new entrants can be combined to pin down a random but finite number of tracking processes. (Instead of considering processes generated from the past, we will talk about "reversed" processes. On a mathematical level, the idea is the same.)

2.2 The Model

Hopenhayn (1992) developed what is now a benchmark model for studying the evolution of a competitive industry where entry and exit is endogenously determined. In this section we briefly outline the model and review standard (non-exact) simulation. Most of our notation follows Hopenhayn (1992). If X is a random variable, then $\mathcal{D}(X)$ denotes the distribution of X. The symbol $\mathscr{B}$ represents the Borel sets on $[0, 1]$.

In the model there is a continuum of firms that produce a homogeneous good. The firms face idiosyncratic productivity shocks that follow a Markov process on $S := [0, 1]$. The conditional cumulative distribution function for the shock process is denoted by $F(\phi' \mid \phi)$. Hopenhayn assumes that F is decreasing in its second argument and that, for any $\epsilon > 0$ and any $\phi \in S$, there exists an integer n such that $F^n(\epsilon \mid \phi) > 0$ (cf., Hopenhayn, 1992, assumptions A.3 and A.4). We refer to these assumptions as the monotonicity and recurrence assumptions respectively. We also let P denote the corresponding stochastic kernel.

That is,

$$P(\phi, A) := \int_A F(d\phi' \mid \phi) \qquad (\phi \in S,\ A \in \mathscr{B}).$$

Incumbent firms exit the industry whenever their current productivity state falls below a reservation value denoted by x_t (cf., Hopenhayn, 1992, p. 1132). The value x_t is defined endogenously in Hopenhayn's model, but in what follows it suffices to treat x_t as a given. Letting M_t be the mass of entrants at time t and ν be the Borel probability measure from which the productivity of entrants is drawn, the sequence of firm distributions $\{\mu_t\}$ on S satisfies $\mu_{t+1}(A) = \int P(\phi, A)\mathbb{1}\{\phi \geq x_t\}\mu_t(d\phi) + M_{t+1}\nu(A)$ for all $A \in \mathscr{B}$, where $\mathbb{1}\{\cdot\}$ is an indicator function. At the stationary equilibrium, both x and M are constant, and a stationary distribution μ is a Borel probability[1] measure μ satisfying

$$\mu(A) = \int P(\phi, A)\mathbb{1}\{\phi \geq x\}\mu(d\phi) + M\nu(A) \qquad (A \in \mathscr{B}). \tag{1}$$

This is eq. (12) in Hopenhayn (1992). It follows from (1) and $\mu(S) = P(\phi, S) = \nu(S) = 1$ that $M = M(x, \mu) := \mu\{\phi \in S : \phi < x\}$. As a result, we can also write (1) as

$$\mu(A) = \int Q(\phi, A)\mu(d\phi) \tag{2}$$

where

$$Q(\phi, A) := P(\phi, A)\mathbb{1}\{\phi \geq x\} + \nu(A)\mathbb{1}\{\phi < x\}. \tag{3}$$

2.3 Simulation

Our initial task is to produce an ergodic Markov process the stationary distribution of which coincides with the distribution μ in (2). To do this, we need a method for sampling from the stochastic kernel Q. The first step is to simulate from the conditional distribution $P(\phi, \cdot) = F(\cdot \mid \phi)$. In particular, we seek a random variable U and a function g such that $\mathcal{D}(g(\phi, U)) = F(\cdot \mid \phi)$ for all $\phi \in S$. This can be achieved via the inverse transform method, where U is uniform on $[0, 1]$ and $g(\phi, u) = F^{-1}(u \mid \phi)$.[2] Now consider the process $\{\Phi_t\}$ defined by

$$\Phi_{t+1} = g(\Phi_t, U_{t+1})\mathbb{1}\{\Phi_t \geq x\} + Z_{t+1}\mathbb{1}\{\Phi_t < x\} \tag{4}$$

where $\{(U_t, Z_t)\}$ is IID with $\mathcal{D}(Z_t) = \nu$ and $\mathcal{D}(U_t) = \text{Uniform}[0, 1]$. Comparing (3) and (4), it can be seen that $\{\Phi_t\}$ is a Markov process with stochastic kernel Q. Indeed, fixing $A \in \mathscr{B}$ and applying the law of iterated expectations, we have

$$\begin{aligned}
\mathbb{P}\{\Phi_{t+1} \in A\} &= \mathbb{E}\left[\mathbb{E}_t\mathbb{1}\{\Phi_{t+1} \in A\}\right] \\
&= \mathbb{E}\left[\mathbb{E}_t\mathbb{1}\{\Phi_{t+1} \in A\}\mathbb{1}\{\Phi_t \geq x\} + \mathbb{E}_t\mathbb{1}\{\Phi_{t+1} \in A\}\mathbb{1}\{\Phi_t < x\}\right] \\
&= \mathbb{E}\left[P(\Phi_t, A)\mathbb{1}\{\Phi_t \geq x\} + \nu(A)\mathbb{1}\{\Phi_t < x\}\right].
\end{aligned}$$

[1] We focus only on normalized measures, since other cases are just scalar multiples.

[2] Here $F^{-1}(u \mid \phi)$ is the generalized inverse of $F(\cdot|\phi)$. That is, $F^{-1}(u \mid \phi) := \inf\{z : F(z \mid \phi) \geq u\}$.

Hence if $\mu_t := \mathcal{D}(\Phi_t)$ for all t, then $\mu_{t+1}(A) = \mathbb{E}\, Q(\Phi_t, A) = \int Q(\phi, A)\mu_t(d\phi)$. It now follows that if μ is the stationary distribution of the process $\{\Phi_t\}$, then (2) must hold and conversely.

3 Exact Sampling

Let $\{(U_t, Z_t)\}_{t\geq\mathbb{Z}}$ be an infinite sequence of IID shocks indexed on $\mathbb{Z}$ and with each pair (U_t, Z_t) having the product distribution Uniform$[0,1] \times \nu$. To simplify notation we will let $g_t := g(\cdot, U_t)$, so that, for example, $g_t \cdots g_1\, \phi := g_t \circ g_{t-1} \circ \cdots \circ g_1(\phi)$ is exogenous productivity after t periods, given initial productivity $\phi \in S$. By the monotonicity and recurrence assumptions, we know that g_t is increasing for all t and

$$\forall\, \epsilon > 0,\ \forall\, \phi \in S,\ \exists\, t \in \mathbb{N} \text{ with } \mathbb{P}\{g_t \cdots g_1\, \phi \leq \epsilon\} > 0. \tag{5}$$

To further simplify notation, let

$$h_t(y) := g(y, U_t)\mathbb{1}\{y \geq x\} + Z_t\mathbb{1}\{y < x\} := g_t(y)\mathbb{1}\{y \geq x\} + Z_t\mathbb{1}\{y < x\}.$$

With this notation, (4) can be written as $\Phi_{t+1} = h_{t+1}\Phi_t := h_{t+1}(\Phi_t)$, and, more generally, the random variable Φ_t can be expressed as $\Phi_t = h_t \cdots h_1\, \Phi_0 := h_t \circ h_{t-1} \circ \cdots \circ h_1(\Phi_0)$. In all of what follows, Φ^* is a draw from μ that is independent of $\{(U_t, Z_t)\}_{t\geq\mathbb{Z}}$. Since Φ^* is a draw from the stationary distribution, it is a trivial observation that

Lemma 3.1. *$\mathcal{D}(h_{t+k} \cdots h_t\, \Phi^*) = \mu$ for all $t \in \mathbb{Z}$ and $k \in \mathbb{N}$.*

Now let $t > 1$ be given, and consider the set

$$\Sigma_{-t} := \{j \in \mathbb{N} : 1 < j \leq t \text{ and } g_{-j} \cdot g_{-j-1} \cdots g_{-t}\, 1 < x\} \quad \text{and} \quad \sigma := \max \Sigma_{-t}.$$

Clearly σ is only defined when Σ_{-t} is nonempty. However, the probability that Σ_{-t} is nonempty converges to one as $t \to \infty$ because of the recurrence property (5).

Lemma 3.2. *If the set Σ_{-t} is nonempty, then either $\Phi^* < x$ or $h_{-j} \cdots h_{-t}\Phi^* < x$ for some $j \in \{\sigma, \sigma+1, \ldots, t\}$.*

Proof. Suppose not. In this case, $\Phi^*, h_{-t}\, \Phi^*, h_{-t+1}h_t\, \Phi^*, \ldots, h_{-\sigma} \cdots h_{-t}\, \Phi^*$ are all greater than x. We know that if $\phi \geq x$, then $h_i\, \phi = g_i\, \phi$. It follows that $h_{-\sigma} \cdots h_{-t}\, \Phi^* = g_{-\sigma} \cdots g_{-t}\, \Phi^*$. By the definition of σ, we have $g_{-\sigma} \cdots g_{-t}\, 1 < x$. This leads to the contradiction

$$x \leq h_{-\sigma} \cdots h_{-t}\, \Phi^* = g_{-\sigma} \cdots g_{-t}\, \Phi^* \leq g_{-\sigma} \cdots g_{-t}\, 1 < x.$$

Here the second inequality is due to monotonicity of g_i and $\Phi^* \leq 1$. □

Lemma 3.3. *If Σ_t is nonempty, then*

$$h_1 \cdots h_t \Phi^* \in \bigcup_{j=\sigma-1}^{t} \{h_1 \cdots h_{j-1} Z_j\}. \tag{6}$$

Proof. Given that Σ_t is nonempty, we know that one of the events in lemma 3.2 has occurred. Suppose first that $h_j \cdots h_t \Phi^* < x$ for some $j \in \{\sigma, \ldots, t-1, t\}$. Observe that

$$h_j \cdots h_t \Phi^* < x \implies h_1 \cdots h_t \Phi^* = h_1 \cdots h_{j-2} h_{j-1} h_j \cdots h_t \Phi^* = h_1 \cdots h_{j-2} Z_{j-1}.$$

This leads to the set of possibilities

$$h_1 \cdots h_t \Phi^* = h_1 \cdots h_{j-2} Z_{j-1} \quad \text{for some} \quad j \in \{\sigma, \ldots, t-1, t\}.$$

The other remaining possibility is that $\Phi^* < x$. In this case, $h_t \Phi^* = Z_t$, and, as a consequence, $h_1 \cdots h_t \Phi^* = h_1 \cdots h_{t-1} Z_t$. Thus, combining all possible outcomes,

$$h_1 \cdots h_t \Phi^* = h_1 \cdots h_{j-2} Z_{j-1} \quad \text{for some} \quad j \in \{\sigma, \ldots, t-1, t, t+1\},$$

which is equivalent to the statement in the lemma. □

The importance of lemma 3.3 is as follows: The values in the set on the right-hand side of (6) are analogous to the final values of the tracking processes in the top panels of figures 1 and 2. If, for some t, we find that Σ_t is nonempty and that this set of values in (6) is a singleton, we will say that the process has *coalesced at* t. In view of (6), the realization $h_1 \cdots h_t \Phi^*$ is equal to the value of the singleton. Since the value in question is observed in the simulation, the implication is that $h_1 \cdots h_t \Phi^*$ has also been observed.

Some caution is required here. While lemma 3.1 shows that the distribution of $h_1 \cdots h_t \Phi^*$ is precisely μ, if we fix t, run the process, test for coalescence, and keep the value of the singleton only when coalescence occurs, we are effectively conditioning on coalescence at $j \leq t$. To avoid such conditioning, we need to obtain coalescence for every random seed. This can be done by fixing the seed and taking t larger and larger until coalescence occurs. Algorithm 1 shows how this is done. The algorithm terminates with an exact draw from μ. Replication with independent shocks will generate independent draws.

An straightforward implementation of algorithm 1 written in C can be obtained from `http://johnstachurski.net/papers/hh_sampling.html`. We found the algorithm to be relatively fast. For example, when $x = 0.25$, the conditional distribution $F(\cdot \mid \phi)$ is equal to the distribution of ϕB for $B \sim \text{Beta}(5, 1)$, and the distribution of new entrants is $\text{Beta}(5, 1)$, the program produces about 16,500 independent draws from μ per second (about 1 million per minute).

Algorithm 1: Generates an exact draw from μ

fix t to be an integer strictly greater than 1;
draw $(U_1, Z_1), \ldots, (U_t, Z_t)$ independently from their correct distributions;
let $\mathbb{S}$ be the list of shocks $(U_1, Z_1), \ldots, (U_t, Z_t)$;
repeat
    compute the set Σ_t using the list of shocks $\mathbb{S}$;
    if Σ_t *is nonempty* **then**
        compute the set in (6) using the list of shocks $\mathbb{S}$;
        if *the set is a singleton* **then**
            set r to be the value of that singleton;
            break;
        end
    end
    draw (U_{t+1}, Z_{t+1}) independently and append to the end of list $\mathbb{S}$;
    set $t = t + 1$;
end
return r;

References

[1] Athreya, K. B., Stenflo, O., 2003. Perfect sampling for Doeblin chains. Sankhya 65, 763–777.

[2] Cooley, T.F. and Quadrini, V., 2001. Financial markets and firm dynamics. American Economic Review, 91 (5), 1286–1310.

[3] Corcoran, J. N., Tweedie, R. L., 2001. Perfect sampling of ergodic Harris chains. Annals of Applied Probability 11, 438–451.

[4] Foss, S., Tweedie, R. L., 1998. Perfect simulation and backward coupling. Stochastic Models 14, 187–204.

[5] Hopenhayn, H. A., 1992. Entry, exit and firm dynamics in long run equilibrium. Econometrica, 60 (5), 1127–1150.

[6] Hopenhayn, H. and Rogerson, R., 1993. Job turnover and policy evaluation: A general equilibrium analysis. Journal of Political Economy, 101 (5), 915–938.

[7] Levin, D. A., Y. Peres, E. L. Wilmer, 2008. Markov Chains and Mixing Times. American Mathematical Society.

[8] Melitz, M.J., 2003. The impact of trade on intra-industry reallocations and aggregate industry productivity. Econometrica, 71 (6), 1695–1725.

[9] Meyn, S., Tweedie, R. L., 1993. Markov Chains and Stochastic Stability. Springer-Verlag, New York.

[10] Murdoch, D. J., Green, P. J., 1998. Exact sampling from a continuous state space. Scandinavian Journal of Statistics 25, 483–502.

[11] Nishimura, K. and J. Stachurski, 2010. Perfect Simulation of Stationary Equilibria. Journal of Economic Dynamics and Control, 34, 577–584.

[12] Propp, J. G., Wilson, D. B., 1996. Exact sampling with coupled Markov chains and applications to statistical mechanics. Random Structures and Algorithms 9, 223–252.