Exact Sampling with Coupled Markov Chains and Applications to Statistical Mechanics *

James Gary Propp
propp@math.mit.edu

David Bruce Wilson
dbwilson@mit.edu

Department of Mathematics
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

July 16, 1996

A Primer on Perfect Simulation

Elke Thönnes
(Mathematical Statistics, Chalmers University of Technology)

Motivation

- Suppose we want to sample from a given distribution π on a finite state space S.
- Standard approach is to construct an ergodic markov chain which has π as its stationary distribution.
- After running the chain for M steps for some large M, the distribution over states tends to π, but hard to determine how large M needs to be to get within a given distance.
- The coupling-from-the-past algorithm determines M dynamically, in order to return exact samples from π.

Forward simulation

Suppose we start in state i^* at time $-M$ and we have access to a random subroutine *Markov*() which given state i produces state j according to the chain's transistion probabilities p_{ij}.
Fixed-time forward simulation:

$i_{-M} \leftarrow i^* \qquad (\text{start chain in state } i^* \text{ at time } -M)$

for $t = -M$ **to** -1

$\qquad i_{t+1} \leftarrow Markov(i_t)$

return i_0

Backward simulation

- Start by running the chain from time -1 to time 0. As we don't know the state of the chain at time -1, we must run the chain from -1 to 0 for each of the n states which might occur at time -1, i.e. we define a random function $f_{-1}(i) = Markov(i)$ for $i = 1, \ldots, n$, which we denote by $f_t = RandomMap()$.
- Define $F^0_{-M} = f_{-1} \circ f_{-2} \circ \cdots \circ f_{-M+1} \circ f_{-M}$
- Then $F^0_{-M}(i^*)$ has the same distribution as in forward simulation.

Coalescence

- F_t^0 can be updated through the equation $F_t^0 = F_{t+1}^0 \circ f_t$
- If F_t^0 becomes constant, i.e. $F_t^0(i) = F_t^0(j)$ for all i, j, then this remains true for all subsequent t' (i.e. $\forall t' < t$).
- When F_t^0 is constant say coalescence occurs from time t. When this occurs there is no need to continue the backward simulation to $-M$, as we must have $F_{-M}^0 = F_t^0$.
- This is true for any M, so in particular if coalescence occurs from time t, $F_{-\infty}^0 = F_t^0$ and the sample returned comes from the equilibrium distribution π.
- Coupling-from-the-past is the procedure of working backwards until t is large enough so that F_t^0 is constant, and then returning the unique value in the range of that map.

Coupling-from-the-past

$t \leftarrow 0$

$F_t^0 \leftarrow$ the identity map

repeat

$t \leftarrow t - 1$

$f_t \leftarrow RandomMap()$

$F_t^0 \leftarrow F_{t+1}^0 \circ f_t$

until $F_t^0(\cdot)$ is constant

return the unique value in the range of $F_t^0(\cdot)$

Theorem 1 *With probability 1 the coupling-from-the-past protocol returns a value, and this value is distributed according to the stationary distribution of the Markov chain.*

Proof: Since the chain is ergodic, there is an L such that for all states i and j, there is a positive chance of going from i to j in L steps. Hence for each t, $F_{t-L}^{t}(\cdot)$ has a positive chance of being constant. Since each of the maps $F_{-L}^{0}(\cdot), F_{-2L}^{-L}(\cdot), \ldots$ has some positive probability $\varepsilon > 0$ of being constant, and since these events are independent, it will happen with probability 1 that one of these maps is constant, in which case F_{-M}^{0} is constant for all sufficiently large M. When the algorithm reaches back M steps into the past, it will terminate and return a value that we will call $\overline{F}_{-\infty}^{0}$. Note that $\overline{F}_{-\infty}^{0}$ is obtained from $\overline{F}_{-\infty}^{-1}$ by running the Markov chain one step, and that $\overline{F}_{-\infty}^{0}$ and $\overline{F}_{-\infty}^{-1}$ have the same probability distribution. Together these last two assertions imply that the output $\overline{F}_{-\infty}^{0}$ is distributed according to the unique stationary distribution π. $\square$

Why from the past?

- If we run the chain from time 0 into the future, finding the smallest t such that $F_0^t(x)$ is constant and outputting that value, the samples obtained are biased.
- To see this imagine a chain in which some states have a unique predecessor, such states can't occur at the time of coalescence.
- So must have $F_t^0 = F_{t+1}^0 \circ f_t$ rather than $F_t^0 = f_t \circ F_{t+1}^0$.
- Also must keep all random number used so far the same, each time we decrement t.

Monotonicity

- For most models of interest, the number of states n is too large to check coalescence by simulating a chain starting in each state.
- Suppose S has a partial ordering $\leq$, and that there exist elements $\hat{0}$ and $\hat{1}$ such that $\hat{0} \leq x \leq \hat{1}$ for all $x \in S$.
- Set $f_t(x) = \phi_t(x, u_t)$ where ϕ_t is a deterministic function and u_t is a random variable. Suppose that ϕ_t has the property that $x \leq y \implies \phi_t(x, u_t) \leq \phi_t(y, u_t)$ almost surely w.r.t. u_t.
- Then rather than consider trajectories starting in all possible states, we can check coalescence by just looking at trajectories starting from $\hat{0}$ and $\hat{1}$.

Example: random walk

Example 2.1 A random walk: *Suppose we have three balls which are distributed over two urns. With probability* $1/2$ *we pick a ball from the left urn and put it into the right urn. Alternatively, we take a ball from the right urn and put it into the left urn. If we find a chosen urn empty we do nothing. What is the long-run average number* M *of balls in the right urn?*

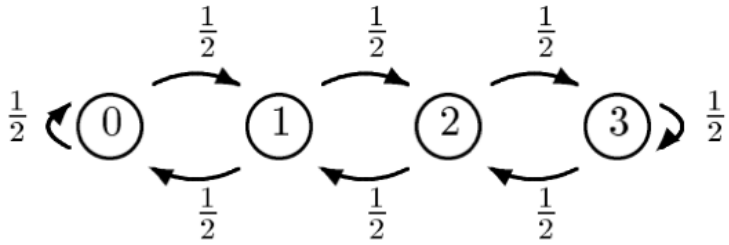


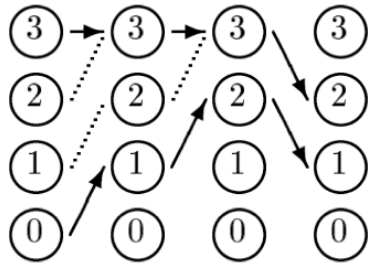
Figure 1: State-flow diagram for the Markov chain in Example 2.1.

- The update rule for this example is monotonic, and can be written as follows:

$$f(n, C) = \begin{cases} \min(n+1, 3) & \text{if } C = H \\ \max(n-1, 0) & \text{if } C = T \end{cases} \qquad n \in \{0, 1, 2, 3\} \qquad (3.1)$$

2.1.

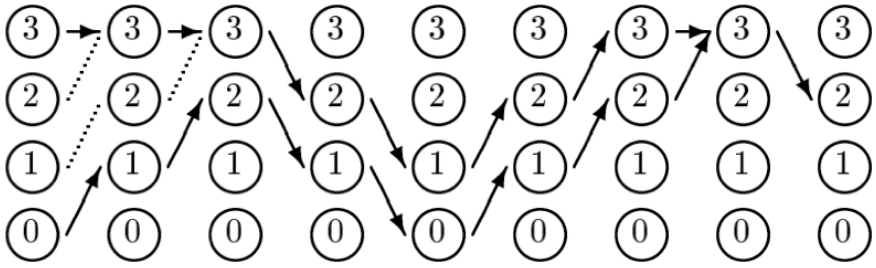
Iteration 1: $T_1 = 3$



Coin Flips: H H T

Time -3 -2 -1 0

Iteration 2: $T_2 = 8$



Coin Flips: H H T T H H H T

Time -8 -7 -6 -5 -4 -3 -2 -1 0

Figure 10: CFTP for Example 2.1. The paths started in state 0 and in state 3 are shown as solid lines. The dotted lines are the paths started from intermediate states. However we do not need to monitor these to determine complete coalescence. Note how the coin toss realisations of the previous iteration are reused! Complete coalescence occurs at time -1, however we continue till time 0 and sample state 2.

Example: Attractive Ising model

$$\pi(x) \quad = \quad \frac{1}{Z} \exp\Big(- H(x)\Big),$$

$$H(x) \quad = \quad -\frac{1}{KT}\Big[\, J\sum_{j\sim k} x_j x_k - Bm\sum_k x_k \,\Big]$$

$$\mathbb{P}\Big(x_n = +1 \,\Big|\, x_{-n}\Big) \quad = \quad \frac{\pi(x_n = +1,\ x_{-n})}{\pi(x_n = +1,\ x_{-n}) + \pi(x_n = -1,\ x_{-n})}.$$

More specifically, at each step k we independently draw a random number U_k which is uniform on the interval $(0,1)$ and a random number N_k which is uniform on the lattice Λ. We then assign an upward spin to the site N_k if

$$U_k \quad \leq \quad \mathbb{P}\Big(x_{N_k} = +1 \,\Big|\, x_{-N_k}\Big),$$

First consider the case when $J > 0$, that is the ferromagnetic Ising model. Then the probability $\mathbb{P}(x_N = +1 \mid x_{-N})$ is the greater the more neighbours of N have an upward spin. We may exploit this fact by equipping the state space of the Ising model with an appropriate partial order $\preceq$. We say the spin configuration x is smaller than y, that is $x \preceq y$ if

$$x_j \quad \leq \quad y_j \qquad \text{for all } j \in \Lambda.$$

Now, if $x \preceq y$ then

$$\mathbb{P}\Big(x_N = +1 \,\Big|\, x_{-N}\Big) \ \leq\ \mathbb{P}\Big(y_N = +1 \,\Big|\, y_{-N}\Big) \quad \text{for any } N \in \Lambda.$$