Exact Simulation of Finite State Markov Chains

Professor Karl Sigman
Columbia University
New York City
USA
Here, we present a result/method introduced in 1996 by Propp and Wilson [3] for simulating, exactly, the stationary distribution of finite state Markov chains. The method is called **coupling from the past (CFTP)**. We present it in an “OR/Applied Probability" way. The purpose is to give a simulation algorithm that yields a random variable distributed exactly as the stationary distribution $\pi$ of a positive recurrent recurrent Markov chain, without a priori having solved for $\pi$. In essence, it is a different way of "finding" $\pi$. Such methods have been generalized in recent years to some Markov processes with a general state space in both discrete and continuous time, particularly in the context of queueing models.
Consider a discrete-time Markov chain (MC) \( \{X_n : n \geq 0\} \), with finite state space \( S = \{0, 1, \ldots, b\} \), and irreducible transition matrix \( P = (P_{ij}) \). Such a chain is positive recurrent with a unique stationary distribution \( \pi = (\pi_0, \pi_1, \ldots, \pi_b) \). For each \( i \in S \) let \( Y(i) \) denote a rv distributed as the \( i^{th} \) row of \( P \):

\[
P(Y(i) = j) = P(X_{n+1} = j | X_n = i) = P_{ij}, \ j \in S.
\]

\( Y(i) \) is distributed as “the next state visited if currently the chain is in state \( i \)".

Let \( Y = (Y(0), Y(1), \ldots, Y(b)) \), where for now we do not specify the joint distribution of this random vector, only the marginal distributions. Let \( \{Y_n : n \geq 1\} \) denote an iid sequence of such random vectors; \( Y_n = (Y_n(0), Y_n(1), \ldots, Y_n(b)) \).
Using the iid sequence $Y_n$, we can recursively construct a version of the Markov chain as follows: First take an initial value, $X_0$. Then define $X_1 = Y_1(X_0)$, $X_2 = Y_2(X_1) = Y_2(Y_1(X_0))$, and so on;

$$X_{n+1} = Y_{n+1}(X_n), \ n \geq 0.$$ 

This is actually how one normally simulates a Markov chain: Given $X_n = i$, independently generate a rv $X$ distributed as the $i^{th}$ row of $P$ and set $X_{n+1} = X$. 
Assume now that the matrix $P$ is \textit{strongly irreducible}, meaning that $P_{ij} > 0$ for all $i, j \in S$. In particular, $P_{i0} > 0$, $i \in S$ and thus \textit{if the coordinates $Y_i$ are chosen to be independent}, then

$$p \overset{\text{def}}{=} P(Y = 0) = P_{00}P_{10} \times \cdots \times P_{b0} > 0.$$

The key consequence of $p > 0$ is this:

\textit{If ever $Y_n = 0$ (which it will at some time $n$ since $p > 0$), then $X_n = 0$ with certainty, and this is so regardless of the value of $X_{n-1}$ (or any of the past).}
Stationary Regime

Let \( \{X_n : n \geq 0\} \) denote a stationary version of the MC, that is, we assume that \( X_0 \) is distributed as (a priori unknown) stationary distribution \( \pi \), and hence \( X_n \) is distributed as \( \pi \) for all \( n \geq 0 \). We can imagine that the chain was started initially at time \( n = -\infty \), and evolved step-by-step into the future until time \( n = 0 \) yielding a stationary sequence from the past:

\[
\{X_{-n} : n \geq 0\} = \cdots - X_{-3}, X_{-2}, X_{-1}, X_0.
\]

In particular, \( X_0 \sim \pi \).
(The rigorous justification for the existence of this “from the past" construction is ensured by Kolmogorov’s extension theorem from basic probability theory. We can also justify it by imagining using the time-reversed Markov chain.)
Our objective is to show that we can produce (simulate) a copy of $X_0$ that only requires using (generating) a finite (but random) number of our iid $\{Y_n : n \geq 1\}$, and nothing else.

To this end we shall first assign $Y_n$ to the random variable $X_{-n}$, $n \geq 1$, to be used recursively from the past to reconstruct (in distribution) the MC up to time $n = 0$:

\[
\begin{align*}
X_0 &= Y_1(X_{-1}) \\
X_{-1} &= Y_2(X_{-2}) \\
X_{-2} &= Y_3(X_{-3}) \\
&\vdots \\
X_{-n+1} &= Y_n(X_{-n}).
\end{align*}
\]
The point is that (in theory) for any fixed \( n \geq 1 \) we could, by the Markov property, replace \( X_0, X_{-1}, \ldots, X_{-n+1} \) by \( Y_1(X_{-1}), Y_2(X_{-2}), \ldots, Y_n(X_{-n}) \) and end up (in distribution) with the same stationary Markov chain. Thus, if we could do this, then we could use \( Y_1(X_{-1}) \) as our desired \( X_0 \) since it would be distributed as \( \pi \) which is our objective.

*The problem is that we do not know the value of \( X_{-n} \), so we do not actually seem to be able to do the replacement—at first glance.*
But at second glance we can do it: To make this possible let

\[ T = \min\{n \geq 1 : Y_n = 0\}. \]

\((T \text{ has a geometric distribution, } P(T = n) = (1 - p)^{n-1}p, \ n \geq 1.)\)

At time \(T\), we have \(Y_T = 0\) and hence \(X_{-T+1} = 0\) regardless of the value of \(X_T\) or any of the past before \(X_T\).

By the Strong Markov property, the past is independent of the future given the present state!!
Coupling from the past (6)

\[
\begin{align*}
X_0 &= Y_1(X_{-1}) \\
X_{-1} &= Y_2(X_{-2}) \\
X_{-2} &= Y_3(X_{-3}) \\
&\vdots \\
X_{-T+2} &= Y_{T-1}(0) \\
X_{-T+1} &= Y_T(X_{-T}) = 0.
\end{align*}
\]

For example, if \( T = 1 \), then \( Y_1 = 0 \), and thus \( X_0 = 0 \).
If \( T = 2 \), then \( X_{-1} = 0 \) and so \( X_0 = Y_1(0) \), where now \( Y_1 \neq 0 \): it has the conditional distribution of a \( Y \) given that not all its coordinates are 0.
Algorithm for generating a rv $X$ distributed as $\pi$:

1. Sequentially generate iid copies $Y_n$, of $Y$ until time $T = \min\{n \geq 1 : Y_n = 0\}$. (Keep a record of all the values $\{Y_n : 1 \leq n \leq T\}$ to use in Step (2) below.)

2. If $T = 1$, then set $X = 0$; if $T = 2$, then set $X = Y_1(0)$. If $T = 3$, then set $X = Y_1(Y_2(0))$. In general, if $T = n \geq 3$, then set $X = Y_1(Y_2(\cdots(Y_{n-1}(0))))$.

Note that we could also accomplish this via first generating a copy of a geometric rv $T$, and then generating $T - 1$ iid copies of $Y$ having the conditional distribution of $Y$ given that not all its coordinates are 0.
We can make this algorithm more efficient in several ways. First, instead of using \( T = \min\{n \geq 1 : Y_n = 0\} \), use

\[
T = \min\{n \geq 1 : Y_n(0) = Y_n(1) = \cdots = Y_n(b)\},
\]

the first time that all coordinates are identical, and replacing state 0 by (random) state \( J = Y_T(0) \) (the common value of the \( Y_T(i) \)). For then the value of \( p = P(Y_n(0) = Y_n(1) = \cdots = Y_n(b)) \) is larger than before so that \( T \) will be smaller than before; less simulation time is needed to get \( X_0 \).
Another improvement results by generating the vectors \( Y \) purposely making its coordinates correlated so as to increase the value of \( p \). For example: take a uniform \( U \) over \((0, 1)\) and use the inverse-transform method to construct the \( Y(i) \) using this same \( U \) for each coordinate. This would, for example, yield

\[
p = P(Y = \mathbf{0}) = \min\{P_{j,0} : 0 \leq j \leq b\}
\]

instead of the much smaller \( p = P(Y = \mathbf{0}) = P_{00}P_{10} \times \cdots \times P_{b0} \) (from the independent coordinates case).
References


