# Interruptible Exact Sampling in the Passive Case

Keith Crank [1]

Division of Mathematical Sciences

National Science Foundation

kcrank@nsf.gov

and

James Allen Fill [2]

Department of Mathematical Sciences

The Johns Hopkins University

jimfill@jhu.edu and http://www.mts.jhu.edu/~fill/

*ABSTRACT*

We establish, for various scenarios, whether or not interruptible exact stationary sampling is possible when a finite-state Markov chain can only be viewed passively. In particular, we prove that such sampling is not possible using a single copy of the chain. Such sampling is possible when enough copies of the chain are available, and we provide an algorithm that terminates with probability one.

*AMS* 2000 *subject classifications.* Primary 60J10, 68U20; secondary 60G40, 62D05, 12D05.

*Key words and phrases.* Exact sampling, perfect simulation, passive case, Markov chain Monte Carlo, stopping time, interruptibility, reversibility, arborescence, tree distribution, Markov chain tree theorem, randomized algorithms, polynomial factorization.

*Date.* October 25, 2002.

[1]The views and findings presented here are those of the author and do not necessarily reflect the opinion of the National Science Foundation.

[2]This author's research was supported by NSF grants DMS–9803780 and DMS–0104167, and by The Johns Hopkins University's Acheson J. Duncan Fund for the Advancement of Research in Statistics; and was carried out in part while author was Visiting Researcher, Theory Group, Microsoft Research.

# 1 Introduction and summary

In recent years a large number of articles have been written about exact sampling (also called perfect sampling) using Markov chains. See [14] for an overview. The rough idea is as follows. One wishes to sample from the unique stationary distribution $\pi$ of an observed irreducible Markov chain. At each transition of the chain, a decision is made whether to continue observing the chain or to stop. When the observation is stopped, a value $S$ is output and it is desired that, for all states $i$, $P(S = i \,|\, \text{one ever stops observing the chain}) = \pi_i$. The decision about whether to stop at a particular time is made on the basis of the evolution of the chain up through that time, possibly together with some additional randomness independent of the chain.

The goal of our research leading to this paper was to determine whether or not it is possible to carry out interruptible exact sampling for finite-state chains in what Propp and Wilson [13] call the passive setting. (We will explain in Section 3.2 what is meant by "interruptible" and "the passive setting".) Our central result is the following:

*Interruptible exact sampling is not possible when one observes only a single trajectory.*

This result remains true even if we assume that the chain is aperiodic and reversible. [See Remark 6.2(b).] However, interruptible exact sampling *is* possible for an $N$-state chain when one is able to observe, simultaneously, $N$ trajectories. Here is a guide to our specific results.

(i) (positive:) We provide an algorithm (Algorithm 4.3) which, given an irreducible Markov chain on $N$ states as input, produces in (random) finite time an exact sample from the tree distribution, and hence also an exact sample from $\pi$. (The tree distribution is defined in Section 3.1.) The algorithm is interruptible, but requires $N$ independent synchronized trajectories from the chain. (See Theorem 4.4.)

(ii) (negative:) There is no algorithm in the passive setting for obtaining an observation from the stationary distribution of an irreducible aperiodic Markov chain on $N$ states which uses fewer than $N$ independent trajectories from the chain and which is both interruptible and exact. (See Theorem 5.1.)

(iii) (negative:) There is no algorithm in the passive setting for obtaining an observation from the common stationary distribution of any finite number of independent irreducible aperiodic Markov chains on $N$ states (with possibly different transition matrices) which is both interruptible and exact. (See Theorem 6.1.) This remains true even if we assume that all of the chains are reversible. [See Remark 6.2(a).]

# 2 Background

In 1992, Asmussen, Glynn, and Thorisson [3] demonstrated that exact sampling from a Markov chain is possible under certain circumstances. They also proved that it is not possible to obtain an exact sample from an arbitrary Markov chain without some prior knowledge about the chain; in particular, the size of the state space must be known.

Although their paper does provide a method for generating exact samples from an $N$-state Markov chain when $N$ is known, the paper is primarily of a theoretical nature, and the method is complicated and inefficient.

In 1995, Lovász and Winkler [11] provided a simpler and more efficient algorithm for obtaining an exact sample from an irreducible $N$-state Markov chain. Although not mentioned explicitly in their paper, the method described in Section 3 of Lovász and Winkler can in fact be used to obtain an exact sample from the tree distribution of the Markov chain (as defined in Section 3.1). Aldous [1], Broder [5], and Propp and Wilson [13] also describe algorithms for sampling from the tree distribution. Propp and Wilson [13] discuss and compare these and other methods of sampling from the tree distribution, and from the stationary distribution. Their discussion includes consideration of such issues as whether or not the sampling is exact or interruptible. To our knowledge, the question of whether interruptible exact sampling is possible in the passive case (as described in Section 3.2) has not previously been considered.

# 3 Preliminaries

## 3.1 The tree distribution

Throughout this paper we consider only finite-state irreducible Markov chains. We assume that the number of states, call it $N$, is known; in fact, it turns out that we may as well assume (and so we do) that the state space is known to be $[N] := \{1, \ldots, N\}$. We denote the transition matrix of such a chain generically by $\mathbf{P} = (p_{ij})$.

An irreducible Markov chain on $[N]$ can be viewed equivalently as a random walk on a connected weighted directed graph $G$. The vertex set of $G$ is $[N]$, and there is an edge from $i$ to $j$, with weight $p_{ij}$, if and only if $p_{ij} > 0$.

For the moment, let us consider an *undirected* graph $G$ with vertex set $[N]$. Then a subgraph $T$ of $G$ is called a *spanning tree* if it contains all $N$ vertices and is connected and acyclic. From any spanning tree, we obtain a *directed spanning tree* by assigning a direction to each edge. A directed spanning tree is called an *arborescence* rooted at a given vertex $r$ if all edges are directed towards $r$.

We define the *weight* $w(T)$ of an arborescence $T$ with edges $\{e_l\}$ as $w(T) := \prod_{l=1}^{N-1} p(e_l)$, where $p(e_l) := p_{ij}$ if $e_l$ is directed from $i$ to $j$. For the remainder of this paper, when we refer to a "tree" $T$ we mean an arborescence $T$ with $w(T) > 0$. The *tree distribution* of the Markov chain is the probability distribution on trees obtained by normalizing the weights $w(T)$ so as to sum to unity.

The Markov chain tree theorem is the well-known result (see, for example, [11] or [2]) that the stationary distribution $\pi$ of the chain can be expressed simply in terms of the tree distribution:

$$\pi_i = w_i / w, \quad i \in [N],$$

where, writing $\mathcal{T}_i$ for the set of trees rooted at $i$ and $\mathcal{T}$ for $\cup_{i \in [N]} \mathcal{T}_i$,

$$w_i := \sum_{T \in \mathcal{T}_i} w(T), \qquad w := \sum_{i \in [N]} w_i = \sum_{T \in \mathcal{T}} w(T).$$

In particular, any algorithm for sampling from the tree distribution provides a means of sampling from $\pi$: simply output the root of the tree.

### 3.2 The passive case; interruptible exact sampling

Propp and Wilson [13] distinguish between the active setting and the passive setting for sampling using a Markov chain. In the active setting, an algorithm is assumed to have access at all times to a transition generator, that is, to a routine which, given any input state $i$, generates an observation $j$ from the probability distribution $(p_{ij} : j \in [N])$, independent of all previously generated observations. In particular, a user can generate a trajectory from $\mathbf{P}$ with any desired initial state. In the passive setting, the algorithm has no control over the initial state and can only watch passively as the chain transitions from one state to the next.

We now explain what is meant by an (on-line, Markov-chain-based) interruptible exact sampling algorithm in the passive case; for simplicity, we will do this explicitly only in the case that a single trajectory from the chain is available and the desired output is an observation from the stationary distribution $\pi$ (rather than one from the tree distribution). Informally, an exact sampling algorithm must take as input a trajectory from the given Markov chain; possibly using external randomization to make its decisions, it watches the chain only until some finite time and then returns an observation distributed according to $\pi$. (Important note: The state returned is not necessarily the state of the chain at the stopping time.) More formally, we can define an exact sampling algorithm as a collection of functions $\phi_{k,i} : [N]^{k+1} \to [0,1]$ [with $\phi_{k,i}(x_0, \dots, x_k)$ to be interpreted informally as the conditional probability that the algorithm stops by time $k$ and outputs $i$, given that it sees the trajectory $(x_0, \dots, x_k)$ through time $k$] having the following properties, where (iii) and (iv) must hold for all $\pi$, for all $\rho$, and for all irreducible transition matrices $\mathbf{P} = (p_{ij})$ on $[N]$ with stationary distribution $\pi$:

(i) $\forall k \geq 0 \;\; \forall (x_0, \dots, x_k) \in [N]^{k+1} : \;\; \sum_j \phi_{k,j}(x_0, \dots, x_k) \leq 1$;

(ii) $\forall i \in [N] \;\; \forall k \geq 0 \;\; \forall (x_0, x_1, \dots) \in [N]^\infty : \;\; \phi_{k,i}(x_0, \dots, x_k) \uparrow$ as $k \uparrow$;

(iii) $\lim_{k\uparrow\infty} \sum_{j\in[N]} \sum_{x_0,x_1,\dots,x_k} \rho_{x_0} p_{x_0,x_1} \cdots p_{x_{k-1},x_k} \, \phi_{k,j}(x_0, x_1, \dots, x_k) > 0$;

(iv) $\forall i \in [N] : \;\; \lim_{k\uparrow\infty} \frac{\sum_{x_0,x_1,\dots,x_k} \rho_{x_0} p_{x_0,x_1} \cdots p_{x_{k-1},x_k} \, \phi_{k,i}(x_0,x_1,\dots,x_k)}{\sum_{j\in[N]} \sum_{x_0,x_1,\dots,x_k} \rho_{x_0} p_{x_0,x_1} \cdots p_{x_{k-1},x_k} \, \phi_{k,j}(x_0,x_1,\dots,x_k)} = \pi_i$.

In terms of the chain $X$ observed and the stopping time $\tau$ and output state $S$ for the algorithm, the properties can be interpreted informally as (i) $P(\tau \leq k \,|\, X) \leq 1$; (ii) $P(\tau \leq k \,|\, X) \uparrow$ as $k \uparrow$; (iii) $P(\tau < \infty) > 0$; and (iv) $P(S = i \,|\, \tau < \infty) \equiv \pi_i$. When the strengthening

(iii′) $\lim_{k\uparrow\infty} \sum_{j\in[N]} \sum_{x_0,x_1,\dots,x_k} \rho_{x_0} p_{x_0,x_1} \cdots p_{x_{k-1},x_k} \, \phi_{k,j}(x_0, x_1, \dots, x_k) = 1$

[interpreted as $P(\tau < \infty) = 1$] of (iii) holds, we will call the algorithm *terminating*. When (iv) can be strengthened to

(iv′) $\forall i \in [N] \;\; \forall k \geq 0$:

$$\sum_{x_0,x_1,\ldots,x_k} \rho_{x_0} p_{x_0,x_1} \cdots p_{x_{k-1},x_k} \, \phi_{k,i}(x_0, x_1, \ldots, x_k)$$
$$= \pi_i \times \sum_{j \in [N]} \sum_{x_0,x_1,\ldots,x_k} \rho_{x_0} p_{x_0,x_1} \cdots p_{x_{k-1},x_k} \, \phi_{k,j}(x_0, x_1, \ldots, x_k)$$

[interpreted as the independence $P(\tau \leq k, S = i) \equiv P(\tau \leq k)\pi_i$ of $\tau$ and $S \sim \pi$], we say that the algorithm is *interruptible*. An interruptible algorithm can be aborted without biasing output; see the discussion in [6]. For active-case algorithms, the leading example of a non-interruptible algorithm is coupling from the past [12], while interruptible algorithms include cycle popping [13], Fill's rejection-based algorithm [6] [8], and the Randomness Recycler [7]. The results of this paper, both positive and negative, are for interruptible algorithms.

# 4 A terminating algorithm for interruptible exact sampling in the passive case

In this section we present a terminating algorithm for interruptible exact stationary sampling in the passive case, assuming that one can watch $N$ synchronized copies $X_i = (X_i(t) : t = 0, 1, \ldots)$, $i \in [N]$, of a Markov chain with state space $[N]$ and irreducible transition matrix $\mathbf{P}$. We allow arbitrary initial distribution $\rho$ for the $N$-variate chain $\mathbf{X} := (X_1, \ldots, X_N)$, but we assume that $X_1, \ldots, X_N$ are conditionally independent given the initial state $(X_1(0), \ldots, X_N(0))$. The algorithm will produce an observation from the tree distribution corresponding to $\mathbf{P}$ (recall Section 3.1).

## 4.1 The algorithm in a restricted setting

In this subsection we present a terminating algorithm for interruptible exact tree-sampling in the passive case that works under the following additional restriction on $\mathbf{P}$:

**Assumption A:** $p_{1j} > 0$ for all $j \in [N]$.

While this assumption may seem unreasonably restrictive, we will show in Section 4.2 how a simple modification of the algorithm can handle the more general case.

To describe the algorithm we first define the following events for even $t \geq 2$. Informal interpretations in words are given parenthetically following the formal definitions; see also Figure 1.

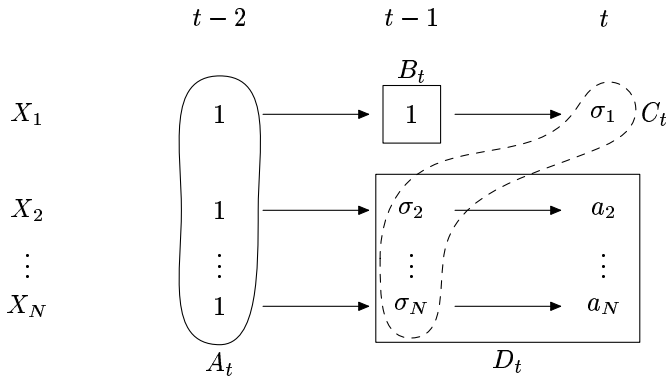


Figure 1: Illustration of Algorithm 4.1.

$$
\begin{aligned}
A_t &:= \cap_{l\in[N]}\{X_l(t-2)=1\} \quad \text{(Every copy of the chain is in state 1.)},\\
B_t &:= \{X_1(t-1)=1\} \quad \text{(The first copy remains in state 1 at the next step.)},\\
C_t &:= \{\{X_1(t), X_2(t-1), \dots, X_N(t-1)\} = [N]\}\\
&\qquad \text{(The states } (X_1(t), X_2(t-1), \dots, X_N(t-1)) \text{ form a permutation of } [N].),\\
D_t(T) &:= \{\text{the graph with directed edges from } X_l(t-1) \text{ to } X_l(t),\ 2\le l\le N,\\
&\qquad \text{is the arborescence } T\},\\
D_t &:= \cup_{T\in\mathcal{T}} D_t(T),\\
E_t(T) &:= A_t\cap B_t\cap C_t\cap D_t(T),\\
E_t &:= \cup_{T\in\mathcal{T}} E_t(T) = A_t\cap B_t\cap C_t\cap D_t.
\end{aligned}
$$

In Figure 1, $C_t$ is realized if and only if $(\sigma_1, \sigma_2, \dots, \sigma_N)$ is a permutation of $[N]$; and $D_t$ is realized if and only if the directed edges $\sigma_j \to a_j$, $j = 2, \dots, N$, form an arborescence with vertex set $[N]$.

**Algorithm 4.1 (Terminating interruptible tree-sampling, under Assumption A).** For even $t \ge 2$, let $E_t(T)$ and $E_t$ be defined as above, and let $E_0 := \emptyset$. The algorithm is:

$t \leftarrow 0$
`repeat`
    $t \leftarrow t + 2$
`until` $E_t$ holds
$S \leftarrow T$, for the unique $T \in \mathcal{T}$ such that $E_t(T)$ holds
`return` $S$

**Theorem 4.2.** *When Asumption A holds, Algorithm 4.1 is a terminating algorithm for interruptible exact tree-sampling.*

*Proof.* Let $\tau$ denote the supremum of the values of the variable $t$ during the operation of Algorithm 4.1. Now fix a candidate value $t$ of $\tau$. Let $T \in \mathcal{T}$ be an arborescence, say with edges $e_l$ directed from $i_l$ to $j_l$, which we choose to index (in some arbitrary

but fixed order) by $l \in \{2, \ldots, N\}$. The event $D_t(T)$ is a disjoint union of $(N-1)!$ subevents, with each subevent corresponding to a way of mapping the $N-1$ transitions $(X_l(t-1), X_l(t))$ to the $N-1$ edges $e_l$. These subevents will all enter symmetrically into the calculation below of $P(\tau = t,\, S = T)$. One such subevent is

$$D'_t(T) := \cap_{l=2}^{N} \{(X_l(t-1), X_l(t)) = (i_l, j_l)\}.$$

Let $\{i_1\}$ denote the singleton $[N] \setminus \{i_2, \ldots, i_N\}$.

Define

$$A'_t := \{\tau \geq t-2\} \cap A_t, \quad \text{even } t \geq 2.$$

Then, using the Markov property and independence of the trajectories,

$$\begin{aligned}
&P(\tau = t,\, S = T) \\
&\quad = \quad P(\{\tau \geq t-2\} \cap E_t(T)) = P(A'_t \cap B_t \cap C_t \cap D_t(T)) \\
&\quad = \quad (N-1)!\, P(A'_t \cap B_t \cap C_t \cap D'_t(T)) \\
&\quad = \quad (N-1)!\, P(A'_t)\, P(B_t \,|\, X_1(t-2) = 1) \left[\prod_{l=2}^{N} P(X_l(t-1) = i_l \,|\, X_l(t-2) = 1)\right] \\
&\qquad\qquad \times P(X_1(t) = i_1 \,|\, B_t) \left(\prod_{l=2}^{N} p_{i_l, j_l}\right) \\
&\quad = \quad (N-1)!\, P(A'_t)\, p_{11} \left[\prod_{l=2}^{N} p_{1,i_l}\right] p_{1,i_1} w(T) = (N-1)!\, P(A'_t)\, p_{11} \left(\prod_{l=1}^{N} p_{1l}\right) w(T).
\end{aligned}$$

Summing over $T \in \mathcal{T}$ we find

$$P(\tau = t) = (N-1)!\, P(A'_t)\, p_{11} \left(\prod_{l=1}^{N} p_{1l}\right) w$$

and therefore

$$P(\tau = t,\, S = T) = P(\tau = t) \frac{w(T)}{w},$$

which shows that Algorithm 4.1 is an interruptible exact tree-sampling algorithm. Using the fact that $\mathbf{X}$ visits $(1, \ldots, 1)$ at even times infinitely often (a.s.) together with the strong Markov property of $\mathbf{X}$, it is clear that termination occurs at the first success in an almost surely infinite sequence of Bernoulli trials with success probability $p_{11} \left(\prod_{l=1}^{N} p_{1l}\right) w > 0$ (note that this is where Assumption A is used). Thus $P(\tau < \infty) = 1$, that is, Algorithm 4.1 is terminating. □

## 4.2 The algorithm in the general setting

To avoid needing Assumption A, we can use the averaging technique of Lovász and Winkler [11]. Let $\mathbf{P}^k$ be the $k$-step transition matrix of the chain $\mathbf{X}$. Then $\overline{\mathbf{P}} := \frac{1}{N} \sum_{k=1}^{N} \mathbf{P}^k$ is a transition matrix with all entries positive. (Compare Problem 7(a) in Chapter 2 of [9].) Moreover, we can effectively use the original chain to sample from

this "averaged" chain. The resulting more general algorithm (Algorithm 4.3) obtains, interruptibly, an exact sample $T$ from the tree distribution of $\mathbf{P}$.

To describe Algorithm 4.3, which works in the general setting described at the outset of Section 4, for $t \geq 3N$ we define the following events to be used in the context of the algorithm. (See also Figure 2.) Informally speaking, the random variables (r.v.s) $U_{l0}$ (each uniform on $[N]$) provide the $N$ trajectories an opportunity to synchronize (in suitable fashion) their initializations at state 1; the r.v. $V$ ensures the possibility of a hold in state 1 for the first trajectory; and the r.v.s $U_{l1}$ $(l = 2, \ldots, N)$ are used in generating the $\overline{\mathbf{P}}$-analogues of the transitions from epoch $t-2$ to epoch $t-1$ in Algorithm 4.1, while $U_{11}$ is used in generating the $\overline{\mathbf{P}}$-analogue of the $X_1$-transition from epoch $t-1$ to epoch $t$ in Algorithm 4.1 (recall Figure 1).

$$
\begin{aligned}
A_t &:= \cap_{l\in[N]}\{X_l(t-3N-1+U_{l0})=1\},\\
B_t &:= \{X_1(t-3N-1+U_{10}+V)=1\},\\
C_t &:= \{\{X_1(t-3N-1+U_{10}+V+U_{11}), X_2(t-3N-1+U_{20}+U_{21}),\\
&\qquad \ldots, X_N(t-3N-1+U_{N0}+U_{N1})\} = [N]\},\\
D_t(T) &:= \{\text{the graph with directed edges from } X_l(t-3N-1+U_{l0}+U_{l1})\\
&\qquad \text{to } X_l(t-3N-1+U_{l0}+U_{l1}+1),\ 2\leq l\leq N, \text{ is the arborescence } T\},\\
D_t &:= \cup_{T\in\mathcal{T}} D_t(T),\\
E_t(T) &:= A_t\cap B_t\cap C_t\cap D_t(T),\\
E_t &:= \cup_{T\in\mathcal{T}} E_t(T) = A_t\cap B_t\cap C_t\cap D_t.
\end{aligned}
$$

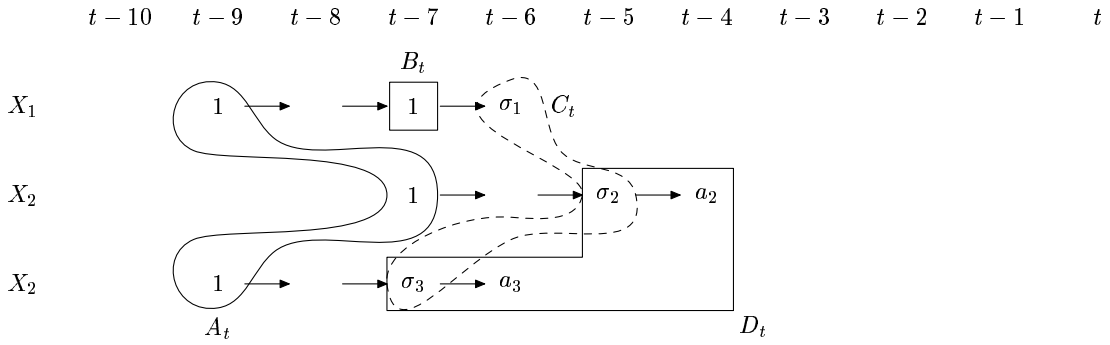


Figure 2: Illustration of Algorithm 4.3. Here $N = 3$, and $U_{10} = 1$, $U_{20} = 3$, $U_{30} = 1$, $V = 2$, $U_{11} = 1$, $U_{21} = 2$, $U_{31} = 2$.

In the following algorithm, successive calls to `Random()`are assumed to generate independent random numbers, each uniformly distributed over $[N]$.

**Algorithm 4.3 (Terminating interruptible stationary sampling).** For $t \geq 3N$, let $E_t(T)$ and $E_t$ be defined as directly above, and let $E_0 := \emptyset$. The algorithm is:

$t \leftarrow 0$

```
repeat
    t ← t + 3N
    V ← Random()
    for l ← 1 to N
        U_{l0} ← Random()
        U_{l1} ← Random()
until E_t holds
S ← T, for the unique T ∈ 𝒯 such that E_t(T) holds
return S
```

By modifying slightly the proof of Theorem 4.2, we obtain the following result.

**Theorem 4.4.** *Algorithm 4.3 is a terminating algorithm for interruptible exact tree-sampling.* □

**Remark 4.5.** Our interest in providing Algorithm 4.3 is more of a theoretical nature (to establish the possibility of terminating interruptible exact sampling, given enough copies of a chain) than of a practical nature (to provide an efficient algorithm). Thus we have not fine-tuned Algorithm 4.3 to improve its performance, and we will not analyze its running time here.

**Remark 4.6.** If we make no assumption regarding the independence of the trajectories, then interruptible sampling becomes impossible for $N \geq 2$ states, no matter how many trajectories are available. Indeed, it is then possible that we are in the extreme case that all the trajectories are identical, i.e., that there is "really" only one trajectory, in which case Theorem 5.1 applies.

# 5 Impossibility of interruptible exact sampling (I)

Algorithm 4.3 requires $N$ independent synchronized Markov chain trajectories. This may seem excessive, especially since for interesting chains $N$ is often enormously large. But our next main result, Theorem 5.1, shows that this is best possible. Note that to prove Theorem 5.1, we need only show that interruptible exact sampling is impossible using $N-1$ independent trajectories. Indeed, if interruptible exact sampling is possible with $m$ independent trajectories, then for any $m' \geq m$ it is possible with $m'$ independent trajectories, since extra trajectories can always be ignored.

**Theorem 5.1.** *There is no algorithm in the passive setting for obtaining an observation from the stationary distribution of an irreducible aperiodic Markov chain on $N$ states which uses fewer than $N$ independent trajectories from the chain and which is both interruptible and exact.*

*Proof.* We first establish an equation [(5.3)] that must hold if there exists an interruptible exact sampling algorithm for $N$-state chains (for given $N \geq 2$) that uses only a *single* trajectory; in that case the discussion of Section 3.2 applies verbatim. A similar

equation, namely (5.4), must hold if interruptible exact sampling is possible using $N-1$ trajectories. But (5.4) will lead to a contradiction via a transition-balancing argument.

So we begin with the case of a single trajectory. Suppose that functions $\phi_{k,i}$ satisfying (i)–(iii) and (iv′) of Section 3.2 exist. We remind the reader that (iii) and (iv′) were required to hold for all initial distributions $\rho$; throughout the present proof it will suffice to consider trajectories starting deterministically at 1. Taking $\rho$ to be unit mass $\delta_1$ at 1 and $p_{ij}$ to be identically $1/N$, we find from (iii) that, for some $0 \le k < \infty$,

$$\sum_{j\in[N]} \sum_{x_1,\dots,x_k} \phi_{k,j}(1, x_1, \dots, x_k) > 0 \tag{5.1}$$

Let $k_0$ be the minimum such $k$, and define $\phi_j(\mathbf{x}) \equiv \phi_j(x_1, \dots, x_{k_0}) := \phi_{k_0,j}(1, x_1, \dots, x_{k_0})$ and $\mathcal{X}_j := \{\mathbf{x} = (x_1, \dots, x_{k_0}) : \phi_j(\mathbf{x}) > 0\}$ for $j \in [N]$. Again taking $\rho$ to be $\delta_1$ and $p_{ij}$ to be identically $1/N$, we find from (iv′) and (5.1) that $\mathcal{X}_i \neq \emptyset$ for $i \in [N]$. Using (iv′) again, we find that for any transition matrix $\mathbf{P}$ with positive entries and stationary distribution $\pi$,

$$\forall i \in [N]: \quad \sum_{\mathbf{x}\in\mathcal{X}_i} \phi_i(\mathbf{x})\, p_{1,x_1} \cdots p_{x_{k_0-1},x_{k_0}} = \pi_i \times \sum_{j\in[N]} \sum_{\mathbf{x}\in\mathcal{X}_j} \phi_j(\mathbf{x})\, p_{1,x_1} \cdots p_{x_{k_0-1},x_{k_0}}, \tag{5.2}$$

and all terms on both sides of (5.2) are positive. Recalling the notation of Section 3.1, it now follows in particular that

$$w_2 \sum_{\mathbf{x}\in\mathcal{X}_1} \phi_1(\mathbf{x}) \prod_{i,j} p_{ij}^{n_{ij}(\mathbf{x})} = w_1 \sum_{\mathbf{x}\in\mathcal{X}_2} \phi_2(\mathbf{x}) \prod_{i,j} p_{ij}^{n_{ij}(\mathbf{x})}, \tag{5.3}$$

where we write $n_{ij}(\mathbf{x})$ for the number of $i \to j$ transitions in the trajectory $(1, x_1, \dots, x_{k_0})$ and again all terms on both sides of the equation are positive.

By the same reasoning, if there exists an interruptible exact sampling algorithm for $N$-state chains that uses $N-1$ independent trajectories, then there exist integer $k \ge 0$ and nonempty sets $\mathcal{X}_1$ and $\mathcal{X}_2$ of $(N-1)$-tuples

$$\mathbf{x} = (x_1(1), \dots, x_1(k); x_2(1), \dots, x_2(k); \dots; x_{N-1}(1), \dots, x_{N-1}(k))$$

of $k$-tuples from $[N]$ such that, for any transition matrix $\mathbf{P}$ with positive entries,

$$w_2 \sum_{\mathbf{x}\in\mathcal{X}_1} \phi_1(\mathbf{x}) \prod_{i,j} p_{ij}^{n_{ij}(\mathbf{x})} = w_1 \sum_{\mathbf{x}\in\mathcal{X}_2} \phi_2(\mathbf{x}) \prod_{i,j} p_{ij}^{n_{ij}(\mathbf{x})}, \tag{5.4}$$

where, for $l = 1, 2$ and and every $\mathbf{x} \in \mathcal{X}_l$, we have $\phi_l(\mathbf{x}) > 0$, and where $n_{ij}(\mathbf{x})$ is the sum over $1 \le m \le N-1$ of the numbers of $i \to j$ transitions within the trajectories $(1, x_m(1), \dots, x_m(k))$. To complete the proof, we will show that (5.4) cannot possibly hold. We will make key use of the observation that, for any $\mathbf{x} \in \mathcal{X}_1 \cup \mathcal{X}_2$,

$$0 \le n_{1+}(\mathbf{x}) - n_{+1}(\mathbf{x}) \le N-1, \tag{5.5}$$

where we have introduced the notation

$$n_{1+}(\mathbf{x}) := \sum_{j=2}^{N} n_{1j}(\mathbf{x}), \qquad n_{+1}(\mathbf{x}) := \sum_{i=2}^{N} n_{i1}(\mathbf{x}) \tag{5.6}$$

for the total numbers of transitions out of and into state 1, respectively. Indeed, since each trajectory $(1, x_m(1), \ldots, x_m(k))$ starts in state 1, the number of transitions out of state 1 within such a trajectory either equals or exceeds by one the number of transitions into state 1.

To obtain the desired contradiction, we begin by observing that (5.4) can be written in the form (eliminating the diagonal variables $p_{ii}$) that

$$w_2 f_1 = w_1 f_2 \tag{5.7}$$

for all $(p_{ij} > 0 : 1 \le i \ne j \le N)$ such that $\sum_{j:j\ne i} p_{ij} < 1$ for every $i \in [N]$, where

$$f_l := \sum_{\mathbf{x}\in\mathcal{X}_l} \phi_l(\mathbf{x}) \left[\prod_{i,j:i\ne j} p_{ij}^{n_{ij}(\mathbf{x})}\right] \left[\prod_{i} \Big(1 - \sum_{j:j\ne i} p_{ij}\Big)^{n_{ii}(\mathbf{x})}\right], \qquad l = 1, 2. \tag{5.8}$$

Using continuity it follows that (5.7) holds for all $(p_{ij} \ge 0 : 1 \le i \ne j \le N)$ such that $\sum_{j:j\ne i} p_{ij} \le 1$ for every $i \in [N]$.

For $l = 1, 2$, note that $f_l$ and $w_l$ are both polynomial expressions in the variables $p_{ij}$, $1 \le i \ne j \le N$ (we will denote this entire collection of $N(N-1)$ variables by $\mathbf{p}$); in fact, $w_1$ is a polynomial expression in the $(N-1)^2$ variables $p_{ij}$ with $i, j \in [N]$ and $i \notin \{1, j\}$ (with a similar reduction in number of variables possible for $w_2$). Applying Proposition A.1 (see the Appendix) to $F := w_2 f_1 - w_1 f_2$, we conclude that (5.7) holds as an equality in the ring of polynomials in the variables $\mathbf{p}$ over the complex field. Henceforth we shall write $G_1 \equiv G_2$ to indicate such an identity of polynomials $G_1, G_2$.

According to Lemma A.2 in the Appendix, the polynomial $w_1$ (again, over the complex field) is irreducible; likewise, so is $w_2$. From the polynomial identity $w_2 f_1 \equiv w_1 f_2$ at (5.7) it then follows that we can write

$$f_l \equiv w_l f, \qquad l = 1, 2, \tag{5.9}$$

for some polynomial $f$ in $\mathbf{p}$. Of course, the polynomial identities (5.9) remain true as we now reduce the number of variables to three by setting $p_{1j}$ to $\alpha$ for $j \ne 1$, $p_{i1}$ to $\beta$ for $i \ne 1$, and $p_{ij}$ to $\gamma$ if $i \ne 1$, $j \ne 1$, and $i \ne j$. Observe that now

$$\begin{aligned} f_l(\alpha, \beta, \gamma) \;\equiv\; & \sum_{\mathbf{x}\in\mathcal{X}_l} \phi_l(\mathbf{x}) \alpha^{n_{1+}(\mathbf{x})} \beta^{n_{+1}(\mathbf{x})} \gamma^{n_{++}(\mathbf{x})} \\ & \times [1 - (N-1)\alpha]^{n_{11}(\mathbf{x})} [1 - \beta - (N-2)\gamma]^{H(\mathbf{x})}, \quad l = 1, 2, \end{aligned} \tag{5.10}$$

recalling (5.6) and defining

$$n_{++}(\mathbf{x}) := \sum_{i,j\in\{2,\ldots,N\}:\, i\ne j} n_{ij}(\mathbf{x}), \qquad H(\mathbf{x}) := \sum_{i=2}^{N} n_{ii}(\mathbf{x}). \tag{5.11}$$

Also now, by a simple generalization of the bijection argument ([10], Section 2.3.4.4, p. 390) showing that the number of arborescences rooted at 1 is $N^{N-2}$,

$$w_1(\beta,\gamma) \equiv \beta[\beta + (N-1)\gamma]^{N-2}; \tag{5.12}$$

and

$$w_2(\alpha,\beta,\gamma) \equiv \alpha\, v_2(\beta,\gamma) \tag{5.13}$$

for some polynomial $v_2(\beta,\gamma)$ which is not divisible by $\beta$ [the explanation for (5.13) being that any $T \in \mathcal{T}_2$ has precisely one directed edge leaving vertex 1 and that there exists $T \in \mathcal{T}_2$ for which 1 is a leaf. In fact, it can be shown that $v_2(\beta,\gamma) \equiv [\beta+(N-1)\gamma]^{N-2}$, but we won't need this.]

The idea for the remainder of the proof is to derive from the identities (5.9)–(5.10) a polynomial identity in the single variable $\beta$, namely (5.14), and then show that (5.14) leads to a contradiction. We will produce (5.14) by eliminating (using suitable divisibility arguments) first $\alpha$ and then $\gamma$. These arguments are carried out in the next two lemmas. □

**Lemma 5.2.** *Suppose that there exists an interruptible exact algorithm in the passive setting for sampling from the stationary distribution of an irreducible aperiodic Markov chain on $N$ states which uses fewer than $N$ independent trajectories from the chain. Then there exist nonempty sets $\mathcal{X}_1'$ and $\mathcal{X}_1''$ and a polynomial $r$ such that*

$$\sum_{\mathbf{x}\in\mathcal{X}_1''} \phi_1(\mathbf{x})\beta^{n_{+1}(\mathbf{x})-m_1(\beta)}(1-\beta)^{H(\mathbf{x})} \equiv \beta^{N-2} r(\beta), \tag{5.14}$$

*where*

$$m_1(\beta) = \min_{\mathbf{x}\in\mathcal{X}_1'} n_{+1}(\mathbf{x}).$$

*Proof.* Let $m_l(\alpha)$ denote the highest power of $\alpha$ that divides $f_l$ at (5.10) and define $\tilde{m}_l(\alpha) := \min_{\mathbf{x}\in\mathcal{X}_l} n_{1+}(\mathbf{x})$. We claim that $m_l(\alpha) = \tilde{m}_l(\alpha)$, and note that this sort of highest-power observation will be used frequently—and without accompanying proof—in the sequel. [Indeed, $m_l(\alpha) \geq \tilde{m}_l(\alpha)$ is clear. To see the reverse inequality, divide $f_l$ by $\alpha^{\tilde{m}_l(\alpha)}$ and set $\alpha$ to 0 to obtain the expression

$$\sum_{\mathbf{x}\in\mathcal{X}_l:\, n_{1+}(\mathbf{x})=\tilde{m}_l(\alpha)} \phi_l(\mathbf{x})\beta^{n_{+1}(\mathbf{x})}\gamma^{n_{++}(\mathbf{x})}[1-\beta-(N-2)\gamma]^{H(\mathbf{x})} =: g_l(\beta,\gamma), \tag{5.15}$$

which is not the zero polynomial since it has a positive value when $\beta = 1/N = \gamma$.]

By (5.9), (5.12), and (5.13),

$$m_2(\alpha) = m_1(\alpha) + 1 \qquad \text{and} \tag{5.16}$$

$$g_l \equiv v_l g, \qquad l = 1, 2, \tag{5.17}$$

where $g_l$ is the polynomial defined at (5.15) [recalling $\tilde{m}_l(\alpha) = m_l(\alpha)$], $v_1 := w_1$, $v_2$ is defined at (5.13), and $g$ is obtained from $f$ by dividing by $\alpha^{m_1(\alpha)}$ and then setting $\alpha = 0$.

Define $\mathcal{X}_l' := \{\mathbf{x} \in \mathcal{X}_l : n_{1+}(\mathbf{x}) = m_l(\alpha)\} \neq \emptyset$ for $l = 1, 2$. Then, similarly, the highest power $m_l(\beta)$ of $\beta$ dividing $g_l$ is $\min_{\mathbf{x}\in\mathcal{X}_l'} n_{+1}(\mathbf{x})$;

$$m_2(\beta) = m_1(\beta) - 1; \tag{5.18}$$

and, with

$$h_1(\beta, \gamma) := \sum_{\mathbf{x}\in\mathcal{X}_1'} \phi_1(\mathbf{x})\beta^{n_{+1}(\mathbf{x})-m_1(\beta)}\gamma^{n_{++}(\mathbf{x})}[1 - \beta - (N-2)\gamma]^{H(\mathbf{x})},$$

we have

$$h_1(\beta, \gamma) \equiv [\beta + (N-1)\gamma]^{N-2} h(\beta, \gamma) \tag{5.19}$$

for some polynomial $h$.

The highest power $m_1(\gamma)$ of $\gamma$ dividing $h_1$ is $\min_{\mathbf{x}\in\mathcal{X}_1'} n_{++}(\mathbf{x})$. Divide both sides of (5.19) by $\gamma^{m_1(\gamma)}$ and set $\gamma$ to 0 to find that (5.14) holds for some polynomial $r$, where $\mathcal{X}_1'' := \{\mathbf{x} \in \mathcal{X}_1' : n_{++}(\mathbf{x}) = m_1(\gamma)\} \neq \emptyset$. □

**Lemma 5.3.** *The identity* (5.14) *cannot hold.*

*Proof.* It follows from (5.14) that $n_{+1}(\mathbf{x}) \geq m_1(\beta) + N - 2$ for all $\mathbf{x} \in \mathcal{X}_1''$. But then, for any such $\mathbf{x}$ and some $\mathbf{x}' \in \mathcal{X}_2'$,

$$\begin{aligned}
n_{+1}(\mathbf{x}) &\geq m_1(\beta) + N - 2 \\
&= m_2(\beta) + N - 1 \quad \text{by (5.18)} \\
&= n_{+1}(\mathbf{x}') + N - 1 \\
&\geq n_{1+}(\mathbf{x}') \quad \text{by the second inequality in (5.5)} \\
&= m_2(\alpha) \\
&= m_1(\alpha) + 1 \quad \text{by (5.16)} \\
&= n_{1+}(\mathbf{x}) + 1,
\end{aligned}$$

contradicting the first inequality in (5.5). □

# 6 Impossibility of interruptible exact sampling (II)

Algorithm 4.3 succeeds in using $N$ independent synchronized Markov chain trajectories to carry out interruptible exact sampling. But the algorithm assumes that each of the trajectories has not only (i) the same stationary distribution, but also (ii) the same transition matrix. In this section we show (Theorem 6.1) that interruptible exact sampling becomes impossible when assumption (ii) is dropped, no matter how (finitely) many trajectories are available.

**Theorem 6.1.** *There is no interruptible algorithm in the passive setting for obtaining an observation exactly from the common stationary distribution of any finite number of independent irreducible aperiodic Markov chains on $N$ states.*

*Proof.* Let $M$ denote the number of trajectories available. We first prove the impossibility of interruptible exact sampling when $M = N = 2$, then more generally when $N = 2$ (regardless of $M$), and finally for general $N$.

For $M = N = 2$, we note that if

$$0 < p_{12}, p_{21} < 1 \quad \text{and} \quad 0 < \rho < 1/\max\{p_{12}, p_{21}\}, \tag{6.1}$$

then

$$\mathbf{P} := \begin{pmatrix} 1 - p_{12} & p_{12} \\ p_{21} & 1 - p_{21} \end{pmatrix} \qquad \text{and} \qquad \mathbf{Q} := \begin{pmatrix} 1 - \rho p_{12} & \rho p_{12} \\ \rho p_{21} & 1 - \rho p_{21} \end{pmatrix}$$

are irreducible aperiodic transition matrices with common stationary distribution

$$\pi = \left( \frac{p_{21}}{p_{21} + p_{12}}, \ \frac{p_{12}}{p_{21} + p_{12}} \right).$$

Arguing as in the proof of Theorem 5.1, if there exists an interruptible exact sampling algorithm in the present setting, then there exist integer $k \geq 0$, nonempty sets $\mathcal{Z}_1$ and $\mathcal{Z}_2$ of pairs

$$\mathbf{z} = (\mathbf{x}, \mathbf{y}) = (x_1, \ldots, x_k; y_1, \ldots, y_k)$$

of $k$-tuples from $\{1, 2\}$, and positive numbers $\psi_l(\mathbf{z})$ ($\mathbf{z} \in \mathcal{Z}_l$, $l = 1, 2$) such that, whenever (6.1) holds,

$$p_{12} f_1 = p_{21} f_2 \tag{6.2}$$

where, using transition-count notation $n_{ij}$ like that in the proof of Theorem 5.1,

$$\begin{aligned} f_l \ = \ & \sum_{\mathbf{z} \in \mathcal{Z}_l} \psi_l(\mathbf{z}) p_{12}^{n_{12}(\mathbf{z})} p_{21}^{n_{21}(\mathbf{z})} \rho^{n_{12}(\mathbf{y}) + n_{21}(\mathbf{y})} (1 - p_{12})^{n_{11}(\mathbf{x})} (1 - p_{21})^{n_{22}(\mathbf{x})} \\ & \times (1 - \rho p_{12})^{n_{11}(\mathbf{y})} (1 - \rho p_{21})^{n_{22}(\mathbf{y})}, \qquad l = 1, 2. \end{aligned} \tag{6.3}$$

Using induction on the $\rho$-degree of the polynomial $p_{12} f_1 - p_{21} f_2$ and Proposition A.1, it is easy to show that (6.2) holds as an equality in the ring of polynomials in the variables $p_{12}, p_{21}, \rho$ over the complex field.

For $l = 1, 2$, let

$$m_l = \min\{[n_{12}(\mathbf{y}) + n_{21}(\mathbf{y})] : \mathbf{z} = (\mathbf{x}, \mathbf{y}) \in \mathcal{Z}_l \text{ for some } \mathbf{x}\}$$

denote the highest power of $\rho$ that divides $f_l$. Then, by (6.2), $m_2 = m_1$. Divide both sides of (6.2) by $\rho^{m_1}$ and then set $\rho$ to 0 to obtain

$$p_{12} g_1 \equiv p_{21} g_2, \tag{6.4}$$

where

$$g_l := \sum_{\mathbf{z} \in \mathcal{Z}_l'} \psi_l(\mathbf{z}) p_{12}^{n_{12}(\mathbf{z})} p_{21}^{n_{21}(\mathbf{z})} (1 - p_{12})^{n_{11}(\mathbf{x})} (1 - p_{21})^{n_{22}(\mathbf{x})}, \qquad l = 1, 2,$$

with $\mathcal{Z}_l' := \{\mathbf{z} = (\mathbf{x}, \mathbf{y}) \in \mathcal{Z}_l : n_{12}(\mathbf{y}) + n_{21}(\mathbf{y}) = m_1\} \neq \emptyset$. But [cf. (5.5) with $N = 2$], if $\mathbf{z} = (\mathbf{x}, \mathbf{y}) \in \mathcal{Z}_1 \cup \mathcal{Z}_2$, then $n_{12}(\mathbf{y}) = \lceil m_1/2 \rceil$ and $n_{21}(\mathbf{y}) = \lfloor m_1/2 \rfloor$. Dividing both sides of (6.4) by $p_{12}^{\lceil m_1/2 \rceil} p_{21}^{\lfloor m_1/2 \rfloor}$ we obtain the polynomial identity

$$p_{12} h_1 \equiv p_{21} h_2, \tag{6.5}$$

where

$$h_l := \sum_{\mathbf{x} \in \mathcal{X}_l} \phi_l(\mathbf{x}) p_{12}^{n_{12}(\mathbf{x})} p_{21}^{n_{21}(\mathbf{x})} (1 - p_{12})^{n_{11}(\mathbf{x})} (1 - p_{21})^{n_{22}(\mathbf{x})}, \qquad l = 1, 2, \tag{6.6}$$

with

$$\mathcal{X}_l := \{\mathbf{x} : \text{there exists } \mathbf{y} \text{ such that } (\mathbf{x}, \mathbf{y}) \in \mathcal{Z}_l'\} \neq \emptyset, \qquad l = 1, 2$$

and, for $\mathbf{x} \in \mathcal{X}_l$,

$$\phi_l(\mathbf{x}) := \sum_{\mathbf{y} \colon (\mathbf{x}, \mathbf{y}) \in \mathcal{Z}_l'} \psi_l(\mathbf{x}, \mathbf{y}) > 0.$$

But (6.5) is the case $N = 2$ of (5.4), which, as shown in the proof of Theorem 5.1, cannot hold. This contradiction establishes the theorem in the case $M = N = 2$.

We leave to the reader the routine extension of the above proof to the case of arbitrary $M$ and $N = 2$. A sketch is that now there are $M - 1$ parameters $\rho_j$, but by using the same sort of argument for each $\rho_j$ in succession that we used above for $\rho$, one again obtains a contradiction of the form (6.5) [with $\phi_l(\mathbf{x}) > 0$ for all $\mathbf{x} \in \mathcal{X}_l \neq \emptyset$, $l = 1, 2$].

We complete the proof of the theorem by showing that an algorithm for interruptible exact sampling using $M$ independent trajectories from chains with $N \geq 3$ states could be converted into one for two-state chains.

Indeed, while watching independent trajectories of $M$ generic irreducible aperiodic two-state chains $X_1, \ldots, X_M$ with common (unknown) stationary distribution $\pi = (\pi_1, \pi_2)$, contemporaneously construct $M$ independent irreducible aperiodic $N$-state chains $Y_1, \ldots, Y_M$ by letting $Y_i(t) = 1$ whenever $X_i(t) = 1$ and selecting an independent uniform random value from $\{2, \ldots, N\}$ as the value of $Y_i(t)$ at each time $t$ such that $X_i(t) = 2$. The stationary distribution for each $Y_i$ is $(\pi_1, \pi_2/(N-1), \ldots, \pi_2/(N-1))$. Applying the size-$N$ algorithm to $Y_1, \ldots, Y_M$, suppose the output state is $S'$. To finish the construction of the two-state algorithm, output $S := \min\{S', 2\}$. □

**Remark 6.2.** (a) Any two-state chain is reversible, as are the chains $Y_i$ constructed in the preceding paragraph. Thus Theorem 6.1 remains true even if we assume that the chains are all reversible.

(b) Similarly, as mentioned in Section 1, interruptible exact sampling from the stationary distribution is not possible when one observes only a single trajectory from an irreducible aperiodic reversible finite-state chain.

(c) For $N \geq 3$ we do not know whether Theorem 5.1 remains true if one assumes that the chain is reversible.

**Acknowledgment.** We thank Dan Naiman for helpful discussions related to the Appendix and Nevin Kapur for help with Figures 1 and 2.

## References


[1] Aldous, D. J. On simulating a Markov chain stationary distribution when transition probabilities are unknown. In D. J. Aldous, P. Diaconis, J. Spencer, and J. M. Steele, editors, *Discrete Probability and Algorithms*, volume 72 of *IMA Volumes in Mathematics and its Applications*, pages 1–9, Springer–Verlag, 1995.

[2] Aldous, D. J. and Fill, J. A. *Reversible Markov Chains and Random Walks on Graphs.* Book in preparation, 200x. Draft available from `http://www.stat.berkeley.edu/users/aldous/`.

[3] Asmussen, S., Glynn, P. W., and Thorisson, H. Stationary detection in the initial transient problem. *ACM Transactions on Modeling and Computer Simulation* **2** (1992), 130–157.

[4] Barshay, J. *Topics in Ring Theory.* W. A. Benjamin, New York, 1969.

[5] Broder, A. Generating random spanning trees. In *Proceedings of the 30th Annual Symposium on Foundations of Computer Science*, pages 442–447, 1989.

[6] Fill, J. A. An interruptible algorithm for perfect sampling via Markov chains. *Annals of Applied Probability* **8** (1998), 131–162.

[7] Fill, J. A. and Huber, M. The Randomness Recycler: a new technique for perfect sampling. In *Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, pages 503–511, 2000.

[8] Fill, J. A., Machida, M., Murdoch, D. J., and Rosenthal, J. S. Extension of Fill's perfect rejection sampling algorithm to general chains. *Random Structures & Algorithms* **17** (2000), 290–316.

[9] Karlin, S. and Taylor, H. M. *A First Course in Stochastic Processes.* Second edition. Academic Press, New York, 1975.

[10] Knuth, D. E. *The Art of Computer Programming. Volume 1* (second edition). *Fundamental Algorithms.* Addison–Wesley, Reading, Mass., 1973.

[11] Lovász, L. and Winkler, P. Exact mixing in an unknown Markov chain. *Electronic Journal of Combinatorics* **2** (1995). Paper #R15.

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

[13] Propp, J. G. and Wilson, D. B. How to get a perfectly random sample from a generic Markov chain and generate a random spanning tree of a directed graph. *Journal of Algorithms* **27** (1998), 170–217.

[14] Wilson, D. B. Annotated bibliography of perfectly random sampling with Markov chains. In *Microsurveys in discrete probability (Princeton, NJ, 1997)*, pages 209–220, Amer. Math. Soc., Providence, RI, 1998. Latest updated version is posted at `http://dbwilson.com/exact/`.

# A Appendix: Polynomials

In this Appendix we establish two basic facts about polynomials; these were used in the proof of Theorem 5.1. Throughout the Appendix, we write $F \equiv G$ to indicate that $F$ and $G$ are the same element in the ring of polynomials (in some specified finite collection of variables) over the complex field.

The first fact is quite simple. For completeness, we include an elementary proof.

**Proposition A.1.** *Let*

$$\mathbf{x} = (x_{ij} : 1 \le i \le n,\ 1 \le j \le k_i)$$

*be a double array of variables, where $n \ge 0$ and $k_i \ge 1$ for $1 \le i \le n$. If $F(\mathbf{x})$ is a polynomial expression that vanishes whenever $x_{ij} \ge 0$ for all $i, j$ and $\sum_{j=1}^{k_i} x_{ij} \le 1$ for all $i$, then $F(\mathbf{x}) \equiv 0$.*

*Proof.* Let $K := \sum_{i=1}^{n} k_i$. The proof is by (strong) induction on $\kappa := K + \deg F$, for which (if $F$ is not the zero polynomial) the smallest possible value is $\sum_{1 \le i \le 0} 1 + 0 = 0$. The base case $\kappa = 0$ of the induction is trivial.

For the induction step we may assume $n \ge 1$ and $k_n \ge 1$. Dividing the polynomial $F(\mathbf{x})$ by $x_{n,k_n}$, we can write

$$F(\mathbf{x}) \equiv x_{n,k_n} F_1(\mathbf{x}) + F_2(\mathbf{x}') \tag{A.1}$$

for polynomials $F_1$ and $F_2$, where the variables collection $\mathbf{x}'$ excludes the single variable $x_{n,k_n}$. Setting $x_{n,k_n}$ to 0 in (A.1), we see that $F_2(\mathbf{x}')$ is a polynomial satisfying the hypothesis of the proposition; and (in obvious notation) $K_2 = K - 1$ and $\deg F_2 \le \deg F$, so that $\kappa_2 < \kappa$. By induction, $F_2(\mathbf{x}') \equiv 0$, and so from (A.1) we now have $F(\mathbf{x}) \equiv x_{n,k_n} F_1(\mathbf{x})$. But now $K_1 = K$ and $\deg F_1 = \deg F - 1$, so that $\kappa_1 = \kappa - 1$, and one sees that $F_1(\mathbf{x})$ satisfies the hypothesis of the proposition. By induction, $F_1(\mathbf{x}) \equiv 0$; we conclude that $F(\mathbf{x}) \equiv 0$, as desired. □

As is well known (e.g., [4], Chapter 4), for any $n \ge 1$ the ring $\mathbf{C}[x_1, \ldots, x_n]$ of polynomials in the variables $x_1, \ldots, x_n$ over the complex field $\mathbf{C}$ is a unique factorization domain. This means that every nonzero polynomial in $\mathbf{C}[x_1, \ldots, x_n]$ can be written uniquely (up to complex scalar multiples) as a (possibly empty) finite product of irreducible polynomials. (A polynomial is said to be irreducible if it cannot be factored as the product of two nonconstant polynomials.)

**Lemma A.2.** *The polynomial $w_1$ [i.e., the polynomial in the $(N-1)^2$ variables $p_{ij}$ with $i, j \in [N]$ and $i \notin \{1, j\}$ defined in Section 3.1] is irreducible over the complex field.*

*Proof.* The proof is by induction on $N$. For $N = 1$, the polynomial $w_1 \equiv 1$ (in no variables) is certainly irreducible. For $N = 2$, the polynomial $w_1 \equiv p_{21}$ in the single variable $p_{21}$ is irreducible. To carry out the induction step for $N \geq 3$, we will use another induction, on $l$, to prove the following claim.

CLAIM. For $3 \leq l \leq N + 1$, let $y_l$ denote the polynomial in $(N-1)^2 - (N+1-l)$ variables obtained from $w_1$ by setting $p_{m1}$ to 0 for $l \leq m \leq N$. Then $y_l$ is irreducible for $4 \leq l \leq N + 1$.

To prove the claim, we begin by noting that $y_3$ has the factorization

$$y_3 \equiv p_{21}\omega_2, \tag{A.2}$$

where the polynomial

$$\omega_2 = \omega_2((p_{ij} : 2 \leq i, j \leq N \text{ and } i \notin \{2, j\}))$$

is obtained from the polynomial $w_1$ for the state space $[N-1]$ by changing each variable name from $p_{ij}$ to $p_{i+1,j+1}$. By the induction hypothesis for our $N$-induction, $\omega_2$ is irreducible. Since $p_{21}$ is clearly irreducible, we conclude that (A.2) is a prime factorization of $y_3$.

We now treat the base case $l = 4$ of our $l$-induction. Observe that $y_4 \not\equiv 0$ (consider, e. g., the tree $N \to N-1 \to \cdots 2 \to 1$) and that $y_4$ is linear in $p_{31}$. If $y_4$ is reducible, then we can write

$$y_4 \equiv (g_1 p_{31} + g_2)g_3, \tag{A.3}$$

where $g_i$ is a polynomial free of the variable $p_{31}$ $(i = 1, 2, 3)$ and $g_3$ is nonconstant. If we now set $p_{31}$ to 0 in (A.3), the result is $y_3 \equiv g_2 g_3$. From the prime factorization (A.2) we conclude that either $p_{21}$ or $\omega_2$ divides $g_3$. But this is wrong: (i) $p_{21}$ does not divide $g_3$ because it clearly does not divide $y_4$ (consider, e. g., the tree $N \to N-1 \to \cdots 4 \to 2 \to 3 \to 1$), and (ii) $\omega_2$ does not divide $g_3$ because (we claim) it, too, fails to divide $y_4$. (Indeed, setting $p_{m2}$ to 0 for $3 \leq m \leq N$ causes $\omega_2$—but clearly not $y_4$—to vanish.) From this contradiction we conclude that $y_4$ is irreducible, establishing the $l$-induction base case.

For the $l$-induction step, let $l \geq 5$. If $y_l$ is reducible, then we can write

$$y_l \equiv (h_1 p_{l-1,1} + h_2)h_3, \tag{A.4}$$

where $h_i$ is a polynomial free of the variable $p_{l-1,1}$ $(i = 1, 2, 3)$ and $h_3$ is nonconstant. If we now set $p_{l-1,1}$ to 0 in (A.4), the result is $y_{l-1} \equiv h_2 h_3$. By the $l$-induction hypothesis, it must be that $h_3$ is a nonzero complex scalar multiple of $y_{l-1}$; from (A.4) we then deduce that $y_{l-1}$ divides $y_l$. But this is wrong, because setting $p_{m1}$ to 0 for $2 \leq m \leq l-2$ causes $y_{l-1}$—but clearly not $y_l$—to vanish. From this contradiction we conclude that $y_l$ is irreducible, completing the $l$-induction.

Finally, set $l$ to $N+1$ in the claim to find that $w_1 \equiv y_{N+1}$ is irreducible, completing the $N$-induction and the proof of the lemma. □