Sankhyā : The Indian Journal of Statistics
2003, Volume 65, Part 4, pp 763-777

PERFECT SAMPLING FOR DOEBLIN CHAINS

By KRISHNA B. ATHREYA
Cornell University, Ithaca, USA and Iowa State University, Ames, USA
and
ÖRJAN STENFLO
Stockholm University, Stockholm, Sweden

SUMMARY. For Markov chains that can be generated by iteration of i.i.d. random maps from the state space X into itself (this holds if X is Polish) it is shown that the Doeblin minorization condition is necessary and sufficient for the method by Propp and Wilson for "perfect" sampling from the stationary distribution π to be successful. Using only the transition probability $\mathbf{P}$ we produce in a geometrically distributed random number of steps N a "perfect" sample from π of size $N!$.

1. Introduction

The problem of sampling exactly from the stationary distribution of an ergodic Markov chain has received much attention in the Markov Chain Monte Carlo literature after the pioneering work of Propp and Wilson (1996). The present work explores this problem in some detail for Markov chains on general state spaces.

Let $(X, \mathcal{B})$ be a measurable space, and $\mathbf{P} : X \times \mathcal{B} \to [0, 1]$ be a transition probability. That is, for each $x \in X$, $\mathbf{P}(x, \cdot)$ is a probability measure on $(X, \mathcal{B})$ and for each $A \in \mathcal{B}$, $\mathbf{P}(\cdot, A)$ is $\mathcal{B}$-measurable. Let $\mathbf{P}$ satisfy the Doeblin hypothesis:

There exist a probability measure ν on $(X, \mathcal{B})$, and constant $0 < \alpha < 1$, such that

$$\mathbf{P}(x, \cdot) \geq \alpha\nu(\cdot), \text{ for all } x \in X. \tag{1}$$

It is known, see e.g. Orey (1971), Athreya and Ney (1978) or Nummelin (1978), that for such a $\mathbf{P}$:

Paper received April 2001; revised February 2002.
AMS (2000) *subject classification.* 65C05.
Keywords and phrases. Exact sampling, perfect sampling, iterated function systems, Markov chains.

i) there exists a unique invariant probability measure π on $(X, \mathcal{B})$ i.e. a probability measure satisfying,

$$\pi(A) = \int_X \mathbf{P}(x, A) d\pi(x), \text{ for all } A \in \mathcal{B}, \tag{2}$$

and

(ii) if $\{Z_n^{\mu_0}\}_{n=0}^{\infty}$ denotes a Markov chain with transition probability $\mathbf{P}$ and initial probability distribution μ_0, then for any μ_0

$$\|P(Z_n^{\mu_0} \in \cdot) - \pi(\cdot)\| \leq (1-\alpha)^n \text{ for } n \geq 0, \tag{3}$$

where $\|\cdot\|$ denotes the total variation norm. Thus $Z_n^{\mu_0}$ may be regarded as a sample from a distribution that is close to π.

The goal of this paper is to show that if the Markov chain with transition probability $\mathbf{P}$ satisfies Doeblin's condition and can be generated by iteration of i.i.d. random maps then it is possible to produce in a finite number of steps using only $\mathbf{P}$ a sample of π-distributed random variables. In fact, our scheme produces a random sample of random size M, say $S := \{x_1, \ldots, x_M\}$ such that each x_i is marginally distributed as π and conditional on M, they are identically distributed and $M = N!$, where N is a random variable with geometric(α) distribution. Such a sample S, has been referred to as an "exact" or "perfect" sample in the Markov Chain Monte Carlo literature (See Wilson (2002)).

It is also shown here that for Markov chains that can be generated by iteration of i.i.d. random maps, success of the method by Propp and Wilson of "perfect sampling" from the stationary distribution of a Markov chain (understood in the sense of condition (A) below) implies the Doeblin condition (1) for some iterate $\mathbf{P}^{n_0}$ of $\mathbf{P}$. Thus the Doeblin condition is necessary and sufficient for the simulation algorithm by Propp and Wilson to be successful for Markov chains that can be generated by iteration of i.i.d. random maps. This includes Markov chains with a countable state space and more generally Markov chains with a Polish (=complete, separable, metric) state space. This result has been obtained (independently) by Foss and Tweedie (1998) (see Remark 11 below).

In the next section, we review some relevant concepts from the theory of iteration with i.i.d. random maps and prove some preliminary useful facts.

In Section 3 we apply these concepts to establish the above claims about Doeblin chains. A numerical example is presented at the end.

2. Iteration of i.i.d. Random Maps and Markov Chains

The simulation of Markov chains in discrete time is often accomplished by representing the Markov chain in the form

$$X_{n+1} = f(X_n, I_n) \tag{4}$$

where f is a function and $\{I_n\}$ is a sequence of independent and identically distributed random variables. Under mild conditions it is possible to represent a general state space Markov chain in this form. Conversely a random dynamical system of the form (4) where $\{I_n\}$ is a sequence of i.i.d. random variables generates a Markov chain under appropriate measurability conditions.

Sequences of the form (4) in the case when $\{I_n\}$ is stationary has been considered by many authors. See e.g. Brandt, Franken and Lisek (1990), Elton (1990), Arnold (1998) and Borovkov (1998) for an overview. See Silvestrov and Stenflo (1998) for the case when $\{I_n\}$ is a regenerative sequence. The particular case when $\{I_n\}$ is i.i.d. allows a richer analysis. See Kifer (1986), Stenflo (1998) and Diaconis and Freedman (1999) for surveys of this literature.

2.1. *Random dynamical systems.* Let $(X, \mathcal{B})$ and $(S, \mathcal{S})$ be two measurable spaces and $w : X \times S \to X$ be jointly measurable, i.e. for any $A \in \mathcal{B}$, $w^{-1}(A) \in \mathcal{B} \times \mathcal{S}$. Let $\{I_j\}_{j=1}^{\infty}$ be a sequence of random elements of S defined on the same probability space $(\Omega, \mathcal{F}, P)$. Consider the random dynamical system defined by

$$Z_n(x,\omega) = w(Z_{n-1}(x,\omega), I_n(\omega)),\ n \geq 1, \qquad Z_0(x,\omega) = x \tag{5}$$

If we write

$$w_s(x) = w(x,s), \tag{6}$$

then (5) can be rewritten (suppressing ω)

$$Z_n(x) := w_{I_n} \circ w_{I_{n-1}} \circ \cdots \circ w_{I_1}(x),\ n \geq 1, \qquad Z_0(x) = x. \tag{7}$$

Consider also the reversed iterates

$$\hat{Z}_n(x) := w_{I_1} \circ w_{I_2} \circ \cdots \circ w_{I_n}(x),\ n \geq 1, \qquad \hat{Z}_0(x) = x. \tag{8}$$

The assumption that $w : X \times S \to X$ is jointly measurable is crucial in rendering both $Z_n(x)$ and $\hat{Z}_n(x)$ random variables on $(\Omega, \mathcal{F}, P)$ for any fixed n and x.

2.2. *I.I.D. random maps.* Of particular importance is the special case when $\{I_j\}_{j=1}^{\infty}$ of 2.1 are independent and identically distributed with common distribution μ. It is intuitively clear from (7) that for each fixed x, $\{Z_n(x)\}_{n=0}^{\infty}$ is a Markov chain starting at x, with transition probability

$$\mathbf{P}(x,A)=\mu(s:w(x,s)\in A),\ x\in X,\ A\in\mathcal{B}. \tag{9}$$

We call the set of objects $\{(X,\mathcal{B}),(S,\mathcal{S},\mu),w(x,s)\}$ an Iterated Function System (IFS) with probabilities. (This generalizes the usual definition, see e.g. Barnsley and Demko (1985), where S typically is a finite set and the functions $w_s=w(\cdot,s):X\to X$ typically have (Lipschitz) continuity properties.)

The above suggests the question: Given a transition probability $\mathbf{P}$ on some state space $(X,\mathcal{B})$ does there exist an IFS with probabilities that generates a Markov chain with $\mathbf{P}$ as its transition probability? (We call such an IFS with probabilities an IFS representation of $\mathbf{P}$.) The answer is yes under general conditions including the case when X is a Polish space. The following proposition and its proof are essentially as in Kifer (1986), Theorem 1.1.

Proposition 1. *Suppose* $\mathbf{P}$ *is a transition probability on a metric space* (X,d) *that is a standard Borel space, i.e., Borel measurably isomorphic to a Borel subset of the real line. Then there exist a jointly measurable function* $w:X\times(0,1)\to X$ *such that*

$$\mathbf{P}(x,A)=\mu(s\in(0,1):w_s(x)\in A), \tag{10}$$

for any $x\in X$ *and Borel set* A *in* X *where* μ *is the Lebesgue measure restricted to the Borel subsets of* $(0,1)$.

If X *is (a Borel subset of)* $\mathbb{R}$, *then* $w:\mathbb{R}\times(0,1)\to\mathbb{R}$ *defined by*

$$w(x,s)=\inf\{y:\mathbf{P}(x,(-\infty,y])\geq s\},$$

is an explicit expression for a map as above.

If X *is a standard Borel space, and* $\phi:X\to\mathbb{R}$ *is a one-to-one Borel map such that* $M=\phi(X)$ *is a Borel subset of* $\mathbb{R}$ *with the property that* $\phi^{-1}:M\to X$ *is also Borel measurable then we can do the following. Define* $\psi:\mathbb{R}\to X$ *as* ϕ^{-1} *on* M *and* x_f *on* $\mathbb{R}\backslash M$ *for some point* $x_f\in X$. *For each* $x\in\mathbb{R}$ *and Borel subset* B *of* $\mathbb{R}$ *define* $\tilde{\mathbf{P}}(x,B)=\mathbf{P}(\psi(x),\phi^{-1}(B\cap M))$. *Define* $g:\mathbb{R}\times(0,1)\to\mathbb{R}$ *by* $g(x,s)=\inf\{y:\tilde{\mathbf{P}}(x,(-\infty,y])\geq s\}$, *and* *let* $w(x,s) = \psi(g(\phi(x),s))$*. Then* w *is an explicit expression for a map as above.*

REMARK 1. If (X,d) is a Polish (=complete, separable, metric) space then (X,d) is a standard Borel space.

REMARK 2. Note that an IFS representation for a transition probability is typically not unique.

Since $\{I_j\}_{j=1}^{\infty}$ is i.i.d. it follows that $Z_n(x)$ and $\hat{Z}_n(x)$ defined in (7) and (8) respectively are identically distributed random variables for each fixed n and x. Thus in order to prove distributional limit results for the Markov chain $\{Z_n(x)\}$ as n tends to infinity we may instead study the pointwise more well behaved (but non-Markovian) sequence $\{\hat{Z}_n(x)\}$.

The following proposition is part of the folklore in this subject. The proof is straightforward, and we omit it here.

PROPOSITION 2. *Let* (X,d) *be a metric space.*

(i) Suppose for some $x \in X$ *there exists a random variable* $\hat{Z}(x)$ *such that*

$$\hat{Z}_n(x) \to \hat{Z}(x), \text{ in distribution.}$$

Let π_x *denote the probability distribution of* $\hat{Z}(x)$*, i.e.* $\pi_x(\cdot) = P(\hat{Z}(x) \in \cdot)$*. Then*

$$Eh(Z_n(x)) \to Eh(\hat{Z}(x)) := \int_X h d\pi_x,$$

for any $h \in C(X)$*, the space of real-valued, bounded and continuous functions on* X*, i.e.* $Z_n(x)$ *converges in distribution to* $\hat{Z}(x)$*.*

(ii) Suppose in addition that $\mathbf{P}$ *(defined as in (9) above) has the Feller property i.e. the map* $Th(x) := \int_X h(y)\mathbf{P}(x,dy)$ *is continuous for any* $h \in C(X)$*. Then* π_x *is invariant for* $\mathbf{P}$*.*

(iii) If $\mathbf{P}$ *has the Feller property and* π_x *in (i) is independent of* $x \in X$*, then* $\pi_x = \pi$ *is the unique invariant probability measure for* $\mathbf{P}$*.*

REMARK 3. As a corollary of the above proposition we obtain that if the maps, $w_s, s \in S$, are all continuous and the limit

$$\hat{Z} := \lim_{n\to\infty} \hat{Z}_n(x) \tag{11}$$

exists and does not depend on $x \in X$ *a.s.*, then π defined by $\pi(\cdot)=P(\hat{Z}\in\cdot)$ is the unique invariant probability measure for the Markov chain with transition probability $\mathbf{P}$ defined as in (9) above. This was formulated as a principle in Letac (1986) and follows since the Markov chains obtained in this case will have the Feller property and almost sure convergence implies convergence in distribution.

REMARK 4. In the last 15 years, there has been a considerable interest for the case when S is a finite set and the maps $w_s, s \in S$, are (affine) uniform contractions. In this case the limit in (11) exists also in the deterministic sense and the compact limit point set $\hat{Z}(\Omega)$ (called the associated fractal set) typically has an intricate self-similar geometry. This set is approached by any trajectory with an exponential rate. The invariant probability measures obtained for these chains are supported on the associated fractal set. See Barnsley (1993) for more on this and an inspiring account on how to generate fractals such as flowers and landscapes as well as applications to image encoding. The Markov chains generated in this way are typically not Harris recurrent. (See e.g. Meyn and Tweedie (1993) for the definition of Harris recurrent Markov chains).

REMARK 5. For an overview of well known sufficient average contraction and stability conditions ensuring (11) with an (almost surely) exponential rate of convergence, or as in condition (A) below uniform in $x \in X$, see e.g. Stenflo (1998), Diaconis and Freedman (1999) and Steinsaltz (1999).

If the convergence in (11) is in the discrete metric and uniform in $x \in X$, then Propp and Wilson (1996) gave an algorithm for exact sampling from π. The following proposition may be viewed as a (slightly weaker) alternative formulation of the simulation algorithm by Propp and Wilson (1996).

The algorithm by Propp and Wilson (1996) for exact simulation: Suppose there exists a random variable $\hat{Z} : (\Omega, \mathcal{F}, P) \to (X, \mathcal{B})$ with the property that

$$(A): \qquad \sup_{x\in X} \hat{d}(\hat{Z}_n(x), \hat{Z}) \stackrel{a.s.}{\to} 0, \text{ as } n \to \infty,$$

where $\hat{d}$ denotes the discrete metric ($\hat{d}(x,y) = 1 \iff x \neq y$) then (equivalently formulated) there exists a random integer N, with $P(N < \infty) = 1$, such that $\hat{Z}_n(x) = \hat{Z}$ for all $n \geq N$ and $x \in X$ and thus $\hat{Z}_N(x)$ is a π distributed random point.

In practice, we continue to simulate i.i.d. random variables $I_1, ..., I_N$ until the first moment when the function $\hat{Z}_N(x)$ does not depend on $x \in X$. (It is clear that $\hat{Z}_n(x) = \hat{Z}$ for all $n \geq N$ since for $n > N$ we have that $\hat{Z}_n(x) := \hat{Z}_N(w_{I_{N+1}} \circ \cdots \circ w_{I_n}(x))$ and $\hat{Z}_N(y)$ does not depend on $y \in X$.)

REMARK 6. In order for this algorithm to be effective we need a good tool to determine whether $\hat{Z}_n(x)$ does depend on x or not.

In the case when X is a partially ordered set with the additional property of the existence of a largest and smallest element y_l and y_s respectively and w_s are monotone with respect to this ordering for any $s \in S$, then $\hat{Z}_n(x, \omega)$ is a map monotone in x and we only need to check whether $\hat{Z}_n(y_l) = \hat{Z}_n(y_s)$ since all other $\hat{Z}_n(x)$ will be sandwiched between these values. See e.g. Diaconis and Freedman (1999) for further details and examples.

If (11) holds and X has a smallest element, y_s, and all maps w_s are decreasing with respect to the given partial order, then we only need to check if $\hat{Z}_n(y_s) = \hat{Z}_{n+1}(y_s)$. (A "symmetric" analogue of this statement when only a largest element y_l exists and all maps are increasing can of course also be stated.) This can be regarded as an extension of the Propp and Wilson algorithm. Interesting examples and extensions of this idea for "dominating chains" can be found in e.g. Häggström and Nelander (1998), Møller (1999) and Kendall and Møller (2000).

If w_{I_n} is constant for some n then $\hat{Z}_n$ will also be constant. This simple property is an essential property we are going to use here.

REMARK 7. If the convergence in (A) is only true with the metric $\hat{d}$ replaced by the metric d, we obtain an algorithm for simulation of points from a distribution, $\hat{\pi}$, close to π in the Prokhorov metric for probability measures. The algorithm can be formulated as follows; Fix a point $x_0 \in X$ and an $\epsilon > 0$. Let $N := \min\{n : \sup_{x,y \in X} d(\hat{Z}_n(x), \hat{Z}_n(y)) < \epsilon\}$. Then $\hat{Z}_N(x_0)$ will have the desired property, with $\hat{\pi}(\cdot) := P(\hat{Z}_N(x_0) \in \cdot)$ being ϵ-close to π in the Prokhorov metric. This extension of the Propp and Wilson algorithm thus makes sense also in cases when we do not have convergence in total variation norm which e.g. is the typical case for fractal supported invariant probability measures. Note however that N need not be measurable in general. In the case when the metric space (X, d) is separable and partially ordered with the additional property of the existence of a largest and smallest element y_l and y_s respectively and w_s are monotone with respect to this ordering for any $s \in S$ then N will be measurable.

REMARK 8. Note that any map on X into itself is continuous if X is given the topology induced by the discrete metric. Thus $\pi(\cdot) = P(\hat{Z} \in \cdot)$ is invariant if condition (A) holds.

REMARK 9. Versions of the Propp and Wilson algorithm can be stated also in cases when $\{I_n\}$ is not i.i.d. but has an underlying i.i.d. structure. See

Stenflo (2001) for a version of the Propp and Wilson algorithm for Markov chains in random environments.

Let us now consider the following conditions:

(B) : There is a set A of positive probability and an n_0 such that for $\omega \in A$, $Z_{n_0}(x, \omega)$ is a constant function.

and

(C) : (The (general) Doeblin hypothesis): There exist a probability measure ν on $(X, \mathcal{B})$, and constants $0 < \alpha < 1$ and $n_0 \geq 1$ such that

$$\mathbf{P}^{n_0}(x, \cdot) \geq \alpha\nu(\cdot), \text{ for all } x \in X.$$

We call an IFS regular if the sets $\{Z_n$ is a constant function$\}$, and $\{\hat{Z}_n$ is a constant function$\}$ are measurable for each n.

Note that if X is separable and $\{q_i\}_{i=1}^{\infty}$ is a countable dense set in X we have that $\{\omega : Z_{n_0}(q_1) = Z_{n_0}(q_i), \forall i\} = \cap_n \cup_{q_i} \cap_{q_j} \{\omega : Z_{n_0}(q_j, \omega) \in \{x : d(x, q_i) < 1/n\}\}$, and $\{\omega : \hat{Z}_{n_0}(q_1) = \hat{Z}_{n_0}(q_i), \forall i\}$ are measurable and thus any IFS representation on a separable metric space with all $w_s, s \in S$ being continuous is necessarily regular.

If S is a finite or countable set and if w_s is measurable for each fixed $s \in S$, then the IFS is regular and no further topological assumptions on X is needed.

THEOREM 1. *For a regular IFS we have the following relations between our conditions:*

$$(A) \Leftrightarrow (B) \Rightarrow (C)$$

and conversely any transition probability on a standard Borel space satisfying condition (C) *with* $n_0 = 1$ *can be represented by an IFS satisfying conditions* (A) *and* (B).

REMARK 10. Many extensions of the Propp and Wilson algorithm exist in the literature. For instance it is easy to see that if there is a set K invariant for all w_s, $s \in S$ i.e. $\cup_{s \in S} w_s(K) \subseteq K$, and the Markov chain $\{Z_n(x)\}$ hits K with probability one for any starting point $x \in X$, then we can view K as the state space when applying the Propp and Wilson algorithm. It is thus also possible to make perfect sampling for Markov chains that are not necessarily uniformly ergodic.

The first nontrivial examples of perfect sampling from the invariant probability distributions for non-uniformly ergodic Markov chains was presented by Kendall (1998). (For further papers and extensions, see the papers cited in Remark 6 and their references.)

REMARK 11. If we understand the perfect sampling method of Propp and Wilson as successful if and only if condition (A) holds, then Doeblin's hypothesis also holds for some n_0. (This result has also been obtained by Foss and Tweedie (1998)). Thus we cannot perform perfect sampling from invariant distributions of general Harris chains using Propp and Wilson's method.

PROOF $(A) \Rightarrow (B)$: Let $A_n = \{\hat{Z}_n$ is a constant function $\}$. The sets A_n are measurable by assumption and increasing i.e. $A_n \subseteq A_{n+1}$, for any $n \geq 1$. Assume condition (A) holds. Condition (A) is equivalent to $P(\cup A_n) = 1$. This implies that $\lim_{n\to\infty} P(A_n) = 1$ and hence $P(A_{n_0}) > 0$ for some $n_0 \geq 1$ which is the same as (B) since for each fixed n, $P(A_n) = P(Z_n$ is a constant function).

PROOF $(B) \Rightarrow (A)$: Assume $\alpha := P(Z_{n_0}$ is a constant function$) > 0$ for some n_0. For integers $m \geq 1$, define the independent random functions $\mathbf{w_m} = w_{I_{mn_0}} \circ w_{I_{mn_0-1}} \circ \cdots \circ w_{I_{(m-1)n_0+1}}$. Thus $Z_{mn_0} = \mathbf{w_m} \circ \cdots \circ \mathbf{w_1}$, and consequently $P(Z_{mn_0}$ is not a constant function$) \leq P(\mathbf{w_i}$ is not a constant function for any $i = 1, \dots, m) = \Pi_{i=1}^{m} P(\mathbf{w_i}$ is not a constant function$) = (1-\alpha)^m$. Thus $P(\hat{Z}_{mn_0}$ is a constant function$) = P(Z_{mn_0}$ is a constant function$) \geq 1 - (1-\alpha)^m \to 1$, as $m \to \infty$, and consequently condition (A) holds. □

PROOF $(B) \Rightarrow (C)$: Assume $\alpha := P(Z_{n_0}$ is a constant function$) > 0$. Define $\nu(\cdot) := P(Z_{n_0} \in \cdot \mid Z_{n_0}$ is a constant function$)$. Then ν is a probability measure on $(X, \mathcal{B})$. It follows that

$$\begin{aligned} \mathbf{P}^{n_0}(x, \cdot) &= P(Z_{n_0}(x) \in \cdot) \\ &\geq P(Z_{n_0}(x) \in \cdot,\ Z_{n_0} \text{ is a constant function}) \\ &= \alpha\nu(\cdot), \end{aligned}$$

and thus condition (C) holds. □

The converse of Theorem 1 will be proved as a part of the proof of Theorem 2 below, stated and proved in the next section.

3. Perfect Sampling for Doeblin Chains

The goal of this section is to establish the claims made in Section 1 about Doeblin chains.

THEOREM 2. *Let* $\mathbf{P}$ *be a transition probability on a standard Borel space. Suppose the Doeblin hypothesis (1) holds.*

Let π *denote the unique invariant probability measure for* $\mathbf{P}$. *Then we can produce a non-trivial sample of* π*-distributed random variables of random size* M, $\{X_1, \ldots, X_M\}$, *where* $M = N!$ *and* N *is a geometric(*α*)-distributed random variable. Conditional on* $N = n$, $\{X_1, \ldots, X_{n!}\}$ *are identically distributed.*

The sample can be explicitly constructed according to the following scheme:
1. Generate a geometric(α*)-distributed random integer,* n.
2. Generate n *independent random numbers,* $i_1, \ldots, i_n$, *uniformly distributed in* $(0,1)$. *The sample can now be expressed by* $\{x_\sigma : \sigma$ *is a permutation of* $\{1, \ldots, n\}\}$, *where* $x_\sigma = f_{i_{\sigma(1)}} \circ f_{i_{\sigma(2)}} \circ \cdots \circ f_{i_{\sigma(n-1)}} \circ g_{i_{\sigma(n)}}$ *and where the functions* $f_s : X \to X$ *and* X*-valued constants* g_s, $s \in (0,1)$ *are constructed by using the algorithm described in Proposition 1 above, for* $\mathbf{Q}(x, \cdot) := (\mathbf{P}(x, \cdot) - \alpha\nu(\cdot))/(1-\alpha)$, *and* $\nu(x, \cdot) := \nu(\cdot)$ *respectively.*

REMARK 12. Murdoch and Green (1998) were the first to show how to generate a random point from the unique invariant probability measure of a Doeblin chain using the Propp and Wilson method. Their result can be considered as the special case of Theorem 2 corresponding to the point x_σ where σ is the identity permutation. The result by Murdoch and Green (1998) was expressed in the context of stochastic recursive sequences by Foss and Tweedie (1998). We believe that our IFS terminology is the most appealing terminology both with respect to intuition and also in order to be able to express results in this field in the most simple way.

REMARK 13. If a Doeblin chain is known to have some additional structure, then there are often algorithms that converges faster than the algorithm in Theorem 2. See e.g. Corcoran and Tweedie (2001). Note, however, that in compensation we here typically get huge samples from the invariant probability measure in these slowly converging cases.

PROOF OF THEOREM 2. Using the algorithm described in Proposition 1 above, let $f(\cdot, s) = f_s, s \in (0,1)$, and $g(\cdot, s) = g_s, s \in (0,1)$ together with the Lebesgue measure restricted to $(0,1)$ be IFS representations of Markov chains with transition probabilities $\mathbf{Q}$ and $\nu(\cdot)$ respectively. (We identify ν with a transition probability defined by $\nu(x,\cdot) := \nu(\cdot)$).

Let $\{I'_n\}$ be a sequence of independent random variables uniformly distributed in $(0,1)$. Let $\{I''_n\}$ be another (independent) such i.i.d. sequence.

Then $\{I_n\}$, with $I_n = (I'_n, I''_n)$ forms an independent sequence uniformly distributed in $(0,1)\times(0,1)$. If we define $w_{s,t} = f_s$ for $0 < t \le 1-\alpha$ and g_s otherwise, we obtain that

$$w_{I_n} = \chi(I''_n \le 1-\alpha) f_{I'_n} + \chi(I''_n > 1-\alpha) g_{I'_n},$$

where χ denotes the indicator function. Thus $\{(X,d), w_{s,t}, (s,t) \in (0,1)\times(0,1)\}$ together with the Lebesgue measure restricted to $(0,1)\times(0,1)$ forms an IFS representation of the transition probability $\mathbf{P}$.

Note that $g_s, s\in(0,1)$ are all constant maps chosen with positive probabi lity and thus condition (B) is fulfilled proving the converse of Theorem 1.

Let $N = \min\{n \ge 1; I''_n > 1-\alpha\}$. Then $P(N = n) = (1-\alpha)^{n-1}\alpha$ and thus $P(N > n) = (1-\alpha)^n$.

Define $Z_n(x)$ and $\hat{Z}_n(x)$ as before and note that if $N \le n$ then $\hat{Z}_n(x) = \hat{Z}_N(x)$ is a constant function. Note also that $P(N \le n) \to 1$ as $n \to \infty$. Define $\hat{Z} := \hat{Z}_N(x)$ and $\pi(\cdot) := P(\hat{Z} \in \cdot)$.

For fixed integers $n \ge 1$, and permutations σ of $\{1,\ldots,n\}$, define

$$\tilde{Z}_n^\sigma = w_{(I'_{\sigma(1)}, I''_1)} \circ \cdots \circ w_{(I'_{\sigma(n)}, I''_n)}.$$

Note that for $\sigma = id$, the identity permutation, we have that $\tilde{Z}_n^\sigma = \hat{Z}_n$. It is clear that $\tilde{Z}_n^\sigma(x)$ and $\tilde{Z}_n^{\hat{\sigma}}(x)$ are identically distributed for any pair σ and $\hat{\sigma}$ of permutations of $\{1,\ldots,n\}$ and any $x \in X$. It is also clear that conditional on the event $\{N = n\}$, $\tilde{Z}_N^\sigma$ and $\tilde{Z}_N^{\hat{\sigma}}$ have the same distribution and the value is independent of x.

Thus for any permutation σ of $\{1,\ldots,N\}$ we have that $X_\sigma := f_{I'_{\sigma(1)}} \circ f_{I'_{\sigma(2)}} \circ \cdots \circ f_{I'_{\sigma(N-1)}} \circ g_{I'_{\sigma(N)}}$ is π-distributed. From this expression we also observe that conditional on N the random variables $\{X_\sigma\}$ are identically distributed. This completes the proof of Theorem 2. □

REMARK 14. If $V(\cdot)$ is a real valued function on $(X,\mathcal{B})$ that is integrable with respect to π, then an estimate of $\lambda = \int_X V\,d\pi$ is

$$\hat{\lambda} = \frac{1}{N!}\sum_{i=1}^{N!} V(X_i),$$

where $\{X_i :\ 1 \le i \le M\}$ is as in Theorem 2.

There is an alternative formulation of Theorem 2 which can be more useful in cases when an IFS representation of $\mathbf{Q}$, defined in Theorem 2 above, is a-priori known. The following theorem states this and also gives a representation of the unique invariant probability measure.

THEOREM 3. *Let* $\mathbf{P}$ *be a transition probability on a measurable space* $(X, \mathcal{B})$ *satisfying the Doeblin condition:*

$$\mathbf{P}(x, \cdot) \geq \alpha\nu(\cdot), \text{ for all } x \in X,$$

where $0 < \alpha < 1$*, and* ν *is a probability measure on* $(X, \mathcal{B})$*. Define the transition probability,*

$$\mathbf{Q}(x, \cdot) := \frac{\mathbf{P}(x, \cdot) - \alpha\nu(\cdot)}{1 - \alpha},$$

and suppose $\{(X, \mathcal{B}), (S, \mathcal{S}, \mu), f(x, s)\}$ *is an IFS representation of* $\mathbf{Q}$*.*
Then

(a) $\mathbf{P}^n(x, \cdot) = \sum_{j=0}^{n-1}(1-\alpha)^j \alpha \int_X \mathbf{Q}^j(y, \cdot) d\nu(y) + (1-\alpha)^n \mathbf{Q}^n(x, \cdot),\ n \geq 1,$

(b) $\pi(\cdot) := \sum_{j=0}^{\infty}(1-\alpha)^j \alpha \int_X \mathbf{Q}^j(y, \cdot) d\nu(y)$ *is the unique invariant probability measure for* $\mathbf{P}$*.*

(c) Let $\{I_j\}_{j=1}^{\infty}$*,* η *and* N *be independent random variables on the same probability space* $(\Omega, \mathcal{F}, P)$ *such that for each* j*,* I_j *is an* $(S, \mathcal{S})$*-valued random variable with distribution* μ*,* η *is an* $(X, \mathcal{B})$*-valued random variable with distribution* ν*, and* N *is an integer valued random variable with geometric*(α) *distribution, i.e.,* $P(N = j) = (1-\alpha)^{j-1}\alpha$ *for* $j \geq 1$*. Let for* $n \geq 1$*,* Σ_n *be the set of all permutations of* $\{1, 2, \ldots, n\}$*. For* $n \geq 2$ *and* $\sigma \in \Sigma_{n-1}$ *let* $X_\sigma = f_{I_{\sigma(1)}} \circ f_{I_{\sigma(2)}} \circ \cdots \circ f_{I_{\sigma(n-1)}}(\eta)$*. For* $n = 1$*, set* $\Sigma_0 = \{0\}$ *and* $X_0 = \eta$*. Let for any* $n \geq 1$*,* $\{\sigma_{n,i} : i = 1, 2, \ldots, (n-1)!\}$ *be a listing of the elements of* Σ_{n-1}*. Then the collection* $\{X_{\sigma_{N,i}}\ 1 \leq i \leq (N-1)!\}$ *has the property that they are* π*-distributed. Further, conditional on* $N = n$ *and* $\eta = x$ *the collection of random variables* $\{X_{\sigma_{n,i}}\ 1 \leq i \leq (n-1)!\}$ *are also identically distributed with distribution* $\mathbf{Q}^{n-1}(x, \cdot)$*.*

REMARK 15. When $(X, \mathcal{B})$ satisfy the conditions of Theorem 2 then we can use Proposition 1 in order to find an IFS representation for $\mathbf{Q}$.

REMARK 16. As a consequence of the representations in Theorem 3 we see that for any $x \in X$ and $n \geq 0$,

$$\begin{aligned} \|\mathbf{P}^n(x,\cdot)-\pi(\cdot)\| &= \|\alpha\sum_{k=0}^{\infty}(1-\alpha)^k\int_X \mathbf{Q}^{n+k}(y,\cdot)d\nu(y) - \mathbf{Q}^n(x,\cdot)\|(1-\alpha)^n \\ &\leq (1-\alpha)^n. \end{aligned}$$

We have thus in particular proved (3).

REMARK 17. If for any $x \in X$, $\mathbf{Q}(x,\cdot)$ is absolutely continuous with respect to λ for some measure λ then π is also absolutely continuous with respect to λ and

$$\Big(\frac{d\pi}{d\lambda}\Big)(y) = \sum_{j=0}^{\infty}(1-\alpha)^j\alpha\int_X q^{(j)}(x,y)d\nu(x),$$

where $q^{(j)}(x,\cdot)$ is the density of $\mathbf{Q}^j(x,\cdot)$ with respect to λ. This can be seen by using the representation in Theorem 3 (b).

REMARK 18. Versions of Theorems 2 and 3 can also be given under the generalized Doeblin hypothesis (C). If we consider subsequences $\{Z_{nn_0}\}_{n=0}^{\infty}$ and note that they have the same invariant probability measure as the full sequence, we see that the methods for the case $n_0 = 1$ can be used.

4. Example

We illustrate our sampling algorithm with a simple example.

EXAMPLE 1. Let

$$\mathbf{P} = \begin{pmatrix} 0.7 & 0.2 & 0.1 \\ 0.4 & 0.2 & 0.4 \\ 0.1 & 0.6 & 0.3 \end{pmatrix},$$

be a Markov transition matrix for a Markov chain on the three points state space $\{0, 1, 2\}$. This Matrix can be written as

$$\begin{pmatrix} 0.7 & 0.2 & 0.1 \\ 0.4 & 0.2 & 0.4 \\ 0.1 & 0.6 & 0.3 \end{pmatrix} = 0.4\begin{pmatrix} 1/4 & 2/4 & 1/4 \\ 1/4 & 2/4 & 1/4 \\ 1/4 & 2/4 & 1/4 \end{pmatrix} + 0.6\begin{pmatrix} 1 & 0 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 2/3 & 1/3 \end{pmatrix}. \qquad (12)$$

Using the algorithm described in Proposition 1 above for generating an IFS representation for the first matrix on the right side of (12), we obtain $g_s = 0$, if $0 < s \leq 1/4$, $g_s = 1$, if $1/4 < s \leq 3/4$, and $g_s = 2$, if $3/4 < s < 1$, and for the second matrix on the right side of (12), we obtain $f_s = h_1$, if $0 < s \leq 1/2$, $f_s = h_2$, if $1/2 < s \leq 2/3$, and $f_s = h_3$, if $2/3 < s < 1$, where the functions $h_i : \{0,1,2\} \to \{0,1,2\}$, $i = 1,2,3$, can be expressed by

$$h_1 = \begin{pmatrix} 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix},\ h_2 = \begin{pmatrix} 0 & 1 & 2 \\ 0 & 2 & 1 \end{pmatrix},\ h_3 = \begin{pmatrix} 0 & 1 & 2 \\ 0 & 2 & 2 \end{pmatrix}.$$

In order to use the algorithm described in Theorem 2, we toss a "skew coin" with probability 0.4 to obtain "head" until the first time, N, when "head" occurs.

Suppose that the value we obtain by this experiment is equal to 3. We now generate 3 random points uniformly distributed in $(0,1)$. Suppose that $0.367, 0.252$, and 0.839 are the results we obtain from this experiment. We note that $g_{0.367} = g_{0.252} = 1$, and $g_{0.839} = 2$ and $f_{0.367} = f_{0.252} = h_1$, and $f_{0.839} = h_3$. Let π denote the unique invariant probability measure for $\mathbf{P}$. We obtain the following sample of π-distributed points; $x_1 := f_{0.367} \circ f_{0.252} \circ g_{0.839} = 0, x_2 := f_{0.367} \circ f_{0.839} \circ g_{0.252} = 1, x_3 := f_{0.252} \circ f_{0.839} \circ g_{0.367} = 1, x_4 := f_{0.252} \circ f_{0.367} \circ g_{0.839} = 0, x_5 := f_{0.839} \circ f_{0.252} \circ g_{0.367} = 0, x_6 := f_{0.839} \circ f_{0.367} \circ g_{0.252} = 0$.

References

Arnold, L. (1998). *Random Dynamical Systems*, Springer-Verlag, Berlin.

Athreya, K.B. and Ney, P. (1978). A new approach to the limit theory of recurrent Markov chains. *Trans. Amer. Math. Soc.* **245**, 493-501.

Barnsley, M.F. (1993). *Fractals Everywhere*, second edition, Academic Press, Boston.

Barnsley, M.F. and Demko, S. (1985). Iterated function systems and the global construction of fractals, *Proc. Roy. Soc. London Ser. A* **399**, 243-275.

Brandt, A., Franken, P. and Lisek, B. (1990). *Stationary Stochastic Models*, Wiley, Chichester.

Borovkov, A.A. (1998). *Ergodicity and Stability of Stochastic Processes*, Wiley, Chichester.

Corcoran, J.N. and Tweedie, R.L. (2001). Perfect sampling of ergodic Harris chains, *Ann. Appl. Probab.* **11**, 438-451.

Diaconis, P. and Freedman, D. (1999). Iterated random functions, *SIAM Rev.* **41**, 45-76.

Elton, J.H. (1990). A multiplicative ergodic theorem for Lipschitz maps, *Stochastic Process. Appl.* **34**, 39-47.

Foss, S.G. and Tweedie, R.L. (1998). Perfect simulation and backward coupling, *Comm. Statist. Stochastic Models* **14**, 187-203.

HÄGGSTRÖM, O. and NELANDER, K. (1998). Exact sampling from anti-monotone systems, *Statist. Neerlandica* **52**, 360-380.

KENDALL, W.S. (1998). Perfect simulation for the area-interaction point process. In *Probability Towards 2000 (New York, 1995)*, Lecture Notes in Statistics **128**, L. Accardi and C.C. Heyde, eds., Springer-Verlag, New York, 218-234.

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.

KIFER, Y. (1986). *Ergodic Theory of Random Transformations*, Birkhäuser, Boston.

LETAC, G. (1986). A contraction principle for certain Markov chains and its applications, *Contemp. Math.* **50**, 263-273.

MEYN, S.P. and TWEEDIE, R.L. (1993). *Markov Chains and Stochastic Stability*, Springer-Verlag, London.

MØLLER, J. (1999). Perfect simulation of conditionally specified models, *J. R. Stat. Soc. Ser. B* **61**, 251-264.

MURDOCH, D.J. and GREEN, P.J. (1998). Exact sampling from a continuous state space, *Scand. J. Statist.* **25**, 483-502.

NEVEU, J. (1965). *Mathematical Foundations of the Calculus of Probability*, Holden-Day, San Francisco.

NUMMELIN, E. (1978). A splitting technique for Harris recurrent Markov chains, *Z. Wahrsch. Verw. Gebiete* **43**, 309-318.

OREY, S. (1971). *Lecture Notes on Limit Theorems for Markov Chain Transition Probabilities*, Van Nostrand Reinhold Mathematical Studies **34**, Van Nostrand, Londan.

PROPP, J.G. and WILSON, D.B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics, *Random Structures and Algorithms* **9**, 223-252.

SILVESTROV, D.S. and STENFLO, Ö. (1998). Ergodic theorems for iterated function systems controlled by regenerative sequences, *J. Theoret. Probab.* **11**, 589-608.

STEINSALTZ, D. (1999). Locally contractive iterated function systems, *Ann. Probab.* **27**, 1952-1979.

STENFLO, Ö. (1998). *Ergodic Theorems for Iterated Function Systems Controlled by Stochastic Sequences*, *Doctoral thesis No 14*, Dept. of Mathematics, Umeå University. (Available at http://www.math.umu.se/~stenflo/avhandling.html)

STENFLO, Ö. (2001). Markov chains in random environments and random iterated function systems, *Trans. Amer. Math. Soc.* **353**, 3547-3562.

WILSON, D.B. (2002). Web site for perfectly random sampling with Markov chains, (http://www.dbwilson.com/exact/)

KRISHNA B. ATHREYA
SCHOOL OF OPERATIONS RESEARCH
AND INDUSTRIAL ENGINEERING
CORNELL UNIVERSITY, ITHACA
NY 14853-3801, USA
AND
DEPARTMENT OF MATHEMATICS AND STATISTICS
IOWA STATE UNIVERSITY
AMES, IA 50011, USA
E-mail: athreya@orie.cornell.edu
kba@iastate.edu

ÖRJAN STENFLO
DEPARTMENT OF MATHEMATICS
STOCKHOLM UNIVERSITY
SE-10691 STOCKHOLM
SWEDEN
E-mail: stenflo@math.su.se