Elect. Comm. in Probab. **13** (2008), 474–481

ELECTRONIC COMMUNICATIONS in PROBABILITY

PERFECT SAMPLING FROM THE LIMIT OF DETERMINISTIC PRODUCTS OF STOCHASTIC MATRICES

ÖRJAN STENFLO
Department of Mathematics, Uppsala University, 751 06 Uppsala, Sweden
email: `stenflo@math.uu.se`

Submitted July 17, 2008, *accepted in final form* September 2, 2008

AMS 2000 Subject classification: 15A51, 65C05, 60J10
Keywords: Perfect sampling, Stochastic matrices, Markov Chain Monte Carlo, Iterated Function Systems

Abstract

We illustrate how a technique from the theory of random iterations of functions can be used within the theory of products of matrices. Using this technique we give a simple proof of a basic theorem about the asymptotic behavior of (deterministic) "backwards products" of row-stochastic matrices and present an algorithm for perfect sampling from the limiting common row-vector (interpreted as a probability-distribution).

1 Preliminaries

An $N \times N$ matrix $P = (p_{ij})$ is called stochastic if $\sum_j p_{ij} = 1$, and $p_{ij} \geq 0$ for all i and j. A stochastic matrix P is called stochastic-indecomposable-aperiodic (SIA) if

$$Q = \lim_{n\to\infty} P^n$$

exists and has all rows equal. Such matrices are of interest in the theory of Markov chains since if $\{X_n\}$ is a Markov chain with a transition matrix, P, which is SIA then in the long run X_n will be distributed according to the common row vector of Q independent of the value of X_0. Basic theory of Markov chains gives soft conditions ensuring a stochastic matrix to be SIA. Essentially "periodicity" has to be ruled out, and one fixed closed irreducible set should eventually be reached by the chain from any given starting point. To have something in mind, the matrix

$$\begin{bmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0.5 & 0.5 \end{bmatrix}, \tag{1}$$

is SIA while the matrices

$$\begin{bmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0.5 & 0.5 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0.5 & 0.5 \end{bmatrix}, \tag{2}$$

are not SIA since decomposability and periodicity respectively are present in the latter two.
Wolfowitz [11] proved that if $A_1, ..., A_{n_0}$ are $N \times N$ matrices such that any finite product of them is SIA, then for any $\epsilon > 0$ there is an integer $\nu(\epsilon)$ such that $m \geq \nu(\epsilon)$ implies that every product of m A_i's has rows which differ by at most ϵ. In particular, if $\overline{\omega} = \omega_1\omega_2\ldots$, with $\omega_j \in \{1, ..., n_0\}$ for any j, then the limit

$$L^{\overline{\omega}} = \lim_{n\to\infty} A_{\omega_n}A_{\omega_{n-1}}\cdots A_{\omega_1},$$

exists with all rows of $L^{\overline{\omega}}$ being identical. Such products are known as infinite "backwards products" in the literature, see e.g. Seneta [8]. In contrast, "forwards products", defined by multiplying the matrices in reversed order will typically not converge. These notions have their roots in the theory of time-inhomogeneous Markov chains where "forwards/backwards products" corresponds to running the Markov chain forwards/backwards in "time".
It is not sufficient to require merely that $A_1, ..., A_{n_0}$ are SIA in Wolfowitz' theorem since it is e.g. easy to construct two SIA matrices A_1 and A_2 such that their product A_1A_2 is decomposable and $(A_1A_2)^n$ will therefore not have approximately identical rows for any n, see e.g. Hajnal [4], Equation 4(b). The conditions of Wolfowitz' theorem may therefore be hard to verify. Some rather weak sufficient conditions for Wolfowitz' condition however follow from Sarymsakov [7]. In addition, Thomasian [9] proved a result which implies an algorithm, requiring a bounded number of arithmetical operations, for determining whether any finite product (of whatever length) of SIA-matrices is SIA. See Paz [6] for a sharp bound.
In the present paper we will extend Wolfowitz' theorem with an algorithm for simulating from the common row vector of $L^{\overline{\omega}}$. In the particular case when $n_0 = 1$ our algorithm reduces to a "Coupling From The Past (CFTP)"-algorithm for "perfect sampling" from the stationary distribution of a homogeneous Markov chain with a SIA transition matrix.
Let us present the simple key idea in the simplest "clearest" case when $n_0 = 1$; Any probability matrix, $P = (p_{ij})$, can be represented by an Iterated Function System (IFS) with probabilities, i.e. there exists a finite family of functions $f_k : \{1, ..., N\} \to \{1, ..., N\}$, and associated probabilities p_k, $k = 1, 2, .., M$, for some constant M, ($p_k > 0$ and $\sum_{k=1}^M p_k = 1$), such that $p_{ij} = \sum_{k:f_k(i)=j} p_k$. Intuitively we may thus "run" a Markov chain with transition probability matrix P by random iterations of the functions in a representing IFS choosing functions f_k with the associated probabilities p_k. The concept of IFSs was coined by Barnsley and Demko [3] and has mainly been used on general state-spaces as a tool to construct fractals. If P has at least one column with only positive elements then we can choose the IFS with the property that (at least) one of the maps f_k, say f_{k_0} is a constant map (i.e. $f_{k_0}(x) = c$, for some constant c), see Athreya and Stenflo [2] for details in a more general setup. If $\{I_n\}_{n=1}^\infty$ is a sequence of independent random variables with $P(I_n = k) = p_k$ for any n and k then it follows that

$$P(f_{I_1} \circ f_{I_2} \circ \cdots \circ f_{I_n}(i) = j) = p_{ij}^{(n)},$$

where $p_{ij}^{(n)}$ denotes the element on row i and column j in the matrix P^n. If the IFS contains a constant map, then it follows that the random variable of composed maps

$$X_n(i) := f_{I_1} \circ f_{I_2} \circ \cdots \circ f_{I_n}(i),$$

will not depend on i, (i.e. will be constant) for all $n \geq T$, where $T = \inf\{j : I_j = k_0\}$ is the first time when the constant map is applied (a geometrically distributed random variable). Thus $X = f_{I_1} \circ f_{I_2} \circ \cdots \circ f_{I_T}(i)$ is a random variable having the same distribution as the common row vector of $\lim_{n\to\infty} P^n$. See Athreya and Stenflo [2] for further details of the above method and Wilson [10] for further literature related to CFTP.

2 Results

Let Ω be an arbitrary set, and let for each $\omega \in \Omega$,

$$P^{\omega} = \begin{bmatrix} p_{11}^{\omega} & \cdots & p_{1N}^{\omega} \\ \vdots & \ddots & \vdots \\ p_{N1}^{\omega} & \cdots & p_{NN}^{\omega} \end{bmatrix}, \tag{3}$$

be a row-stochastic matrix, i.e. a matrix with $\sum_j p_{ij}^{\omega} = 1$, and $p_{ij}^{\omega} \geq 0$ for all i and j.
Our basic theorem is the following:

Theorem 1. *Suppose all matrices P^{ω} satisfy a Doeblin condition uniformly in ω i.e. suppose there exists a constant $c > 0$ such that for any $\omega \in \Omega$,*

$$\sum_j c_{jmin}^{\omega} \geq c, \tag{4}$$

where $c_{jmin}^{\omega} = \min\{p_{1j}^{\omega}, p_{2j}^{\omega}, \ldots, p_{Nj}^{\omega}\}$ denotes the minimum value of the elements in the j:th column of P^{ω}.
Let $\overline{\omega} = \omega_1\omega_2\ldots$, be an arbitrary sequence of elements of Ω i.e. $\omega_j \in \Omega$, $j \geq 1$.

a) Convergence towards the limit with an exponential rate:
The limit

$$M^{\overline{\omega}} = \lim_{n\to\infty} P^{\omega_n} P^{\omega_{n-1}} \cdots P^{\omega_1} \tag{5}$$

exists, and $M^{\overline{\omega}}$ is a matrix with identical rows, i.e.

$$M^{\overline{\omega}} = \begin{bmatrix} \mu_1^{\overline{\omega}} & \cdots & \mu_N^{\overline{\omega}} \\ \vdots & \ddots & \vdots \\ \mu_1^{\overline{\omega}} & \cdots & \mu_N^{\overline{\omega}} \end{bmatrix},$$

for some probability vector $\mu^{\overline{\omega}} = (\mu_1^{\overline{\omega}}, \ldots, \mu_N^{\overline{\omega}})$.

If $M_n^{\overline{\omega}} = P^{\omega_n} P^{\omega_{n-1}} \cdots P^{\omega_1}$, then for any i,

$$\frac{1}{2}\sum_{j=1}^{N} |M_n^{\overline{\omega}}(i,j) - \mu_j^{\overline{\omega}}| \leq (1-c)^n, \; n \geq 0. \tag{6}$$

Thus the convergence rates of the row distributions towards the limit distribution $\mu^{\overline{\omega}} = (\mu_1^{\overline{\omega}}, \ldots, \mu_N^{\overline{\omega}})$ in (5) is exponential (in total-variation norm) uniformly in i.

b) Perfect sampling:
Let U_1, U_2, U_3,... be independent random variables uniformly distributed on the unit interval, and let T be a geometrically distributed random variable with $P(T=k)=c(1-c)^{k-1}$, $k\geq 1$, independent of the U_i:s, $i\geq 1$.

Then $X^{\overline{\omega}}=g_{U_1}^{\omega_1}\circ g_{U_2}^{\omega_2}\circ\cdots\circ g_{U_{T-1}}^{\omega_{T-1}}(X_{U_T})$, where

$$X_{U_T}=\min(k:\frac{\sum_{j=1}^{k}c_{jmin}^{\omega_T}}{\sum_{j=1}^{N}c_{jmin}^{\omega_T}}\geq U_T),$$

and $g_{U_m}^{\omega_m}(i)=\min(k:\sum_{j=1}^{k}(p_{ij}^{\omega_m}-c\frac{c_{jmin}^{\omega_m}}{\sum_{j=1}^{N}c_{jmin}^{\omega_m}})\geq U_m(1-c))$, $m\geq 1$, is a random variable with distribution $\mu^{\overline{\omega}}=(\mu_1^{\overline{\omega}},\ldots,\mu_N^{\overline{\omega}})$.

Remark 1. *The result in* **b)** *suggest an obvious algorithm for generating perfect samples from the common row-distribution of $M^{\overline{\omega}}$ by simulating sets of a geometrically distributed size of uniformly distributed random numbers. See Athreya and Stenflo [2] for generalizations of the above theorem in the case when Ω contains one point.*

Proof The matrices P^{ω} may be expressed as

$$P^{\omega}=cP_{\star}^{\omega}+(1-c)P_{\diamond}^{\omega},$$

where $P_{\star}^{\omega}=(p_{ij}^{\star\omega})$, with $p_{ij}^{\star\omega}=c_{jmin}^{\omega}/(\sum_{j=1}^{N}c_{jmin}^{\omega})$ and $P_{\diamond}^{\omega}=(p_{ij}^{\diamond\omega})$, with

$$p_{ij}^{\diamond\omega}=\frac{(p_{ij}^{\omega}-c\frac{c_{jmin}^{\omega}}{\sum_{j=1}^{N}c_{jmin}^{\omega}})}{1-c}.$$

Define

$$g_s^{\diamond\omega}(i)=\min(k:\sum_{j=1}^{k}p_{ij}^{\diamond\omega}\geq s),$$

for each given $\omega\in\Omega$, and $s\in(0,1)$. The function $g^{\diamond\omega}$ is an IFS representation of $P_{\diamond}^{\omega}$ in the sense that if U is a uniformly distributed random variable on the unit interval, then $P(g_U^{\diamond\omega}(i)=j)=p_{ij}^{\diamond\omega}$. Similarly the matrix $P_{\star}^{\omega}$ can be represented by an IFS containing only the constant maps

$$g_s^{\star\omega}=g_s^{\star\omega}(i)=\min(k:\sum_{j=1}^{k}p_{ij}^{\star\omega}\geq s).$$

If for $\bar{s}=(s_1,s_2)\in(0,1)^2$,

$$g_{\bar{s}}^{\omega}=\chi(s_2\leq 1-c)g_{s_1}^{\diamond\omega}+\chi(s_2>1-c)g_{s_1}^{\star\omega},$$

where χ denotes the indicator function, then if $\mathbf{U}=(U,\hat{U})$ is a uniformly distributed random variable on $(0,1)^2$, then

$$P(g_{\mathbf{U}}^{\omega}(i)=j)=c\cdot P(g_U^{\star\omega}(i)=j)+(1-c)\cdot P(g_U^{\diamond\omega}(i)=j)=p_{ij}^{\omega}, \tag{7}$$

and we have thus created an explicit IFS representation of P^{ω}.

If $\{\mathbf{U}_n = (U_n, \hat{U}_n)\}$ is a sequence of independent uniformly distributed random variables on $(0,1)^2$, and we define

$$X_n^{\overline{\omega}}(i) = g_{\mathbf{U}_1}^{\omega_1} \circ \cdots g_{\mathbf{U}_n}^{\omega_n}(i),$$

then if $T = \inf\{n : \hat{U}_n > 1-c\}$, then it follows that T is a geometrically distributed random variable with parameter c, independent of the random variables U_n, and

$$X_n^{\overline{\omega}}(i) = g_{U_1}^{\diamond\omega_1} \circ g_{U_2}^{\diamond\omega_2} \circ g_{U_n}^{\diamond\omega_n}(i), \text{ for } n < T,$$

and

$$X_n^{\overline{\omega}}(i) = X^{\overline{\omega}} := g_{U_1}^{\diamond\omega_1} \circ g_{U_2}^{\diamond\omega_2} \circ g_{U_{T-1}}^{\diamond\omega_{T-1}}(g_{U_T}^{\star\omega_T}), \text{ for } n \geq T.$$

Let

$$M_n^{\overline{\omega}} = P^{\omega_n} P^{\omega_{n-1}} \cdots P^{\omega_1}.$$

We claim that

$$P(X_n^{\overline{\omega}}(i) = j) = M_n^{\overline{\omega}}(i,j) \tag{8}$$

for all n, $\overline{\omega}$, i and j. In order to see this we use induction; First note by construction that $P(X_1^{\overline{\omega}}(i) = j) = p_{ij}^{\omega_1} = M_1^{\overline{\omega}}(i,j)$, see (7). If $P(X_n^{\overline{\omega}}(i) = j) = M_n^{\overline{\omega}}(i,j)$ for some fixed n, then

$$\begin{aligned}
P(X_{n+1}^{\overline{\omega}}(i) = j) &= P(X_n^{\overline{\omega}}(g_{\mathbf{U}_{n+1}}^{\omega_{n+1}}(i)) = j) \\
&= c \cdot P(X_n^{\overline{\omega}}(g_{U_{n+1}}^{\star\omega_{n+1}}(i)) = j) + (1-c) \cdot P(X_n^{\overline{\omega}}(g_{U_{n+1}}^{\diamond\omega_{n+1}}(i)) = j) \\
&= c \cdot \sum_{k=1}^{N} p_{ik}^{\star\omega_{n+1}} M_n^{\overline{\omega}}(k,j) + (1-c) \sum_{k=1}^{N} p_{ik}^{\diamond\omega_{n+1}} M_n^{\overline{\omega}}(k,j) \\
&= (P^{\omega_{n+1}} M_n^{\overline{\omega}})(i,j) = M_{n+1}^{\overline{\omega}}(i,j).
\end{aligned}$$

Therefore (8) follows by induction.
Define $\mu_j^{\overline{\omega}} = P(X^{\overline{\omega}} = j)$. By construction

$$\begin{aligned}
\frac{1}{2} \sum_{j=1}^{N} |M_n^{\overline{\omega}}(i,j) - \mu_j^{\overline{\omega}}| &= \frac{1}{2} \sum_{j=1}^{N} |P(X_n^{\overline{\omega}}(i) = j) - P(X^{\overline{\omega}} = j)| \\
&\leq P(X_n^{\overline{\omega}}(i) \neq X^{\overline{\omega}}) \leq P(T > n) = (1-c)^n.
\end{aligned}$$

This completes the proof of the theorem.

Remark 2. *Let d denote the Dobrushin ergodic coefficient defined for a stochastic matrix A by*

$$d(A) = \frac{1}{2} \max_{i_1,i_2} \sum_{j=1}^{N} |A(i_1,j) - A(i_2,j)| = 1 - \min_{i_1,i_2} \sum_{j=1}^{N} \min(A(i_1,j), A(i_2,j)). \tag{9}$$

It is straightforward to prove that

$$d(A \cdot B) \leq d(A) d(B),$$

for any two stochastic matrices A and B. By using the right hand side of (9) we see that assumption (4) implies that $d(P^{\omega}) \leq 1-c$, for any ω and thus

$$d(M_n^{\overline{\omega}}) \leq (1-c)^n. \tag{10}$$

Now fix j and $\overline{\omega}$ and note that $\min_i M_n^{\overline{\omega}}(i,j)$ *is increasing in n. Also note, similarly, that* $\max_i M_n^{\overline{\omega}}(i,j)$ *is decreasing in n. By (10),*

$$(\max_i M_n^{\overline{\omega}}(i,j) - \min_i M_n^{\overline{\omega}}(i,j)) \sim O((1-c)^n) \tag{11}$$

and therefore there exists a limit $\mu_j^{\overline{\omega}}$ such that

$$\sum_{j=1}^{N} |M_n^{\overline{\omega}}(i,j) - \mu_j^{\overline{\omega}}| \sim O((1-c)^n).$$

We can thus prove that the exponential convergence rate of Theorem 1**a)** *holds without involving IFSs. A good reason to involve IFSs in the proof above is however that it simultaneously allows us to prove* **b)**.

Remark 3. *Note that we have no restrictions on the number of matrices in our theorem above. The theorem corresponds to backwards limits for Markov chains in random environments or time-inhomogeneous Markov chains with state space $S = \{1,...,N\}$. It is straightforward to extend the theorem to the case when S is an arbitrary complete separable metric space. Transition matrices $P^{\omega} = (p_{ij}^{\omega})$ are then replaced by transition kernels $P^{\omega}(x,B)$ (interpreted as the transition probability of moving from state x into the set B) satisfying the Doeblin condition $P^{\omega}(x,B) \geq c\nu^{\omega}(B)$, for all sets B for some probability measure ν^{ω} depending on $\omega \in \Omega$. In the proof we then use the fact that any Markov chain on a complete separable metric space can be represented by an IFS, see Kifer [5] or Athreya and Stenflo [2].*

As a consequence of Theorem 1 we can now, as our main result, extend the theorem by Wolfowitz [11] with a perfect sampling algorithm;
Suppose $A_1,...,A_{n_0}$ are $N \times N$ matrices such that any finite product of them is SIA. If $\overline{\omega} = \omega_1\omega_2\ldots$, with $\omega_j \in \{1,...,n_0\}$ for any j, then it follows from [11] that for any $c \in (0,1)$ there exists an integer n_c such that the matrices

$$P^{(1)} = A_{\omega_{n_c}} A_{\omega_{n_c-1}} \cdots A_{\omega_1}, \quad P^{(2)} = A_{\omega_{2n_c}} A_{\omega_{n_c-1}} \cdots A_{\omega_{n_c+1}}, \quad P^{(3)} = \ldots, \ \ldots$$

satisfy the conditions of Theorem 1.
That is

$$\sum_j c_{jmin}^{(k)} \geq c,$$

where $c_{jmin}^{(k)} = \min\{p_{1j}^{(k)}, p_{2j}^{(k)}, \ldots, p_{Nj}^{(k)}\}$ denotes the minimum value of the elements in the j:th column of $P^{(k)}$, $k \geq 1$.
By applying Theorem 1 we obtain;

Corollary 1. *Suppose $A_1,...,A_{n_0}$ are $N \times N$ matrices such that any finite product of them is SIA. Then for any $\overline{\omega} = \omega_1\omega_2\ldots$, with $\omega_j \in \{1,...,n_0\}$ for any j, the limit*

$$L^{\overline{\omega}} = \lim_{n\to\infty} A_{\omega_n} A_{\omega_{n-1}} \cdots A_{\omega_1}, \tag{12}$$

exists with all rows of $L^{\overline{\omega}}$ being identical and the convergence rates of the row distributions towards the limit is exponential (in total-variation norm) uniformly in the row-index and $\overline{\omega}$.

If U_1, U_2, U_3,... is a sequence of independent random variables uniformly distributed on the unit interval, and T is a geometrically distributed random variable with $P(T=k)=c(1-c)^{k-1}$, $k \geq 1$, independent of the U_i:s, $i \geq 1$, then $X^{\overline{\omega}} = \hat{g}^{(1)}_{U_1} \circ \hat{g}^{(2)}_{U_2} \circ \cdots \circ \hat{g}^{(T-1)}_{U_{T-1}}(\hat{X}_{U_T})$, where

$$\hat{X}_{U_T} = \min(k : \frac{\sum_{j=1}^{k} c^{(T)}_{jmin}}{\sum_{j=1}^{N} c^{(T)}_{jmin}} \geq U_T),$$

and $\hat{g}^{(m)}_{U_m}(i) = \min(k : \sum_{j=1}^{k}(p^{(m)}_{ij} - c\frac{c^{(m)}_{jmin}}{\sum_{j=1}^{N} c^{(m)}_{jmin}}) \geq U_m(1-c))$, $m \geq 1$, is a random variable with distribution $l^{\overline{\omega}} = (l^{\overline{\omega}}_1, \dots, l^{\overline{\omega}}_N)$ being the common row vector of $L^{\overline{\omega}}$.

Proof From Theorem 1 it follows that the limit

$$L^{\overline{\omega}} = \lim_{n\to\infty} P^{(n)}P^{(n-1)}\cdots P^{(1)} \tag{13}$$

exists, and $L^{\overline{\omega}}$ is a matrix with identical rows, i.e.

$$L^{\overline{\omega}} = \begin{bmatrix} l^{\overline{\omega}}_1 & \dots & l^{\overline{\omega}}_N \\ \vdots & \ddots & \vdots \\ l^{\overline{\omega}}_1 & \dots & l^{\overline{\omega}}_N \end{bmatrix},$$

for some probability vector $l^{\overline{\omega}} = (l^{\overline{\omega}}_1, \dots, l^{\overline{\omega}}_N)$.

Moreover, if $L^{\overline{\omega}}_n = A_{\omega_n}A_{\omega_{n-1}}\cdots A_{\omega_1}$, then for any i,

$$\frac{1}{2}\sum_{j=1}^{N}|L^{\overline{\omega}}_{nn_c}(i,j) - l^{\overline{\omega}}_j| \leq (1-c)^n,\ n \geq 0, \tag{14}$$

and thus, (c.f. (11)),

$$\max_j(\max_i L^{\overline{\omega}}_{nn_c}(i,j) - \min_i L^{\overline{\omega}}_{nn_c}(i,j)) \sim O((1-c)^n) \tag{15}$$

This establishes the exponential convergence rate, since if $0 \leq k < n_c$, then

$$\begin{aligned} |L^{\overline{\omega}}_{nn_c+k}(i,j) - l^{\overline{\omega}}_j| &\leq \max_i L^{\overline{\omega}}_{nn_c+k}(i,j) - \min_i L^{\overline{\omega}}_{nn_c+k}(i,j) \\ &\leq \max_i L^{\overline{\omega}}_{nn_c}(i,j) - \min_i L^{\overline{\omega}}_{nn_c}(i,j) \end{aligned}$$

and thus by (15)

$$\frac{1}{2}\sum_{j=1}^{N}|L^{\overline{\omega}}_{nn_c+k}(i,j) - l^{\overline{\omega}}_j| \sim O((1-c)^n).$$

Therefore

$$\frac{1}{2}\sum_{j=1}^{N}|L^{\overline{\omega}}_{n}(i,j) - l^{\overline{\omega}}_j| \sim O((1-c)^{[n/n_c]}),\ n \geq 0,$$

for any i, where $[x]$ denotes the largest integer less than or equal to x.

Remark 4. *The exponential convergence rate in Corollary 1 is a known consequence of the theorem in [11], see e.g. Anthonisse and Tijms [1]. See also Seneta [8].*
An advantage with our method based on IFSs is that it gives the perfect sampling algorithm "for free".

3 Acknowledgment

I am grateful to Michael Barnsley for encouraging me to write this paper.

References

[1] J. M. Anthonisse and H. Tijms, Exponential convergence of products of stochastic matrices, *J. Math. Anal. Appl.* **59** (1977), no. 2, 360–364. MR0443092

[2] K. B. Athreya and Ö. Stenflo, Perfect sampling for Doeblin chains, *Sankhyā* **65** (2003), no. 4, 763–777. MR2060404

[3] M. F. Barnsley and S. Demko, Iterated function systems and the global construction of fractals, *Proc. Roy. Soc. London Ser. A* **399** (1985), no. 1817, 243–275. MR0799111

[4] J. Hajnal, Weak ergodicity in non-homogeneous Markov chains, *Proc. Cambridge Philos. Soc.* **54** (1958), 233–246. MR0096306

[5] Y. Kifer, *Ergodic theory of random transformations*, Birkhäuser Boston, Boston, MA, 1986. MR0884892

[6] A. Paz, Definite and quasidefinite sets of stochastic matrices, *Proc. Amer. Math. Soc.* **16** (1965), 634–641. MR0177996

[7] T. A. Sarymsakov, Inhomogeneous Markov chains, *Teor. Verojatnost. i Primenen* **6** (1961), 194–201. MR0203813

[8] E. Seneta, *Nonnegative matrices and Markov chains*, Second edition, Springer, New York, 1981. MR0719544

[9] A. J. Thomasian, A finite criterion for indecomposable channels, *Ann. Math. Statist.* **34** (1963), 337–338. MR0144797

[10] D. B. Wilson, Web site for perfectly random sampling with Markov chains, (http://www.dbwilson.com/exact/)

[11] J. Wolfowitz, Products of indecomposable, aperiodic, stochastic matrices, *Proc. Amer. Math. Soc.* **14** (1963), 733–737. MR0154756