Journal of Statistical Physics, Vol. 39, Nos. 1/2, 1985

Nonstationary Markov Chains and Convergence of the Annealing Algorithm

Basilis Gidas[1]

Received May 29, 1984; revised August 22, 1984 and October 2, 1984

We study the asymptotic behavior as time $t \to +\infty$ of certain nonstationary Markov chains, and prove the convergence of the annealing algorithm in Monte Carlo simulations. We find that in the limit $t \to +\infty$, a nonstationary Markov chain may exhibit "phase transitions." Nonstationary Markov chains in general, and the annealing algorithm in particular, lead to biased estimators for the expectation values of the process. We compute the leading terms in the bias and the variance of the sample-means estimator. We find that the annealing algorithm converges if the temperature $T(t)$ goes to zero no faster than $C/\log(t/t_0)$ as $t \to +\infty$, with a computable constant C and t_0 the initial time. The bias and the variance of the sample-means estimator in the annealing algorithm go to zero like $O(t^{-1+\varepsilon})$ for some $0 \leqslant \varepsilon < 1$, with $\varepsilon = 0$ only in very special circumstances. Our results concerning the convergence of the annealing algorithm, and the rate of convergence to zero of the bias and the variance of the sample-means estimator, provide a rigorous procedure for choosing the optimal "annealing schedule." This optimal choice reflects the competition between two physical effects: (a) The "adiabatic" effect, whereby if the temperature is lowered *too abruptly* the system may end up not in a ground state but in a nearby metastable state, and (b) the "super-cooling" effect, whereby if the temperature is lowered *too slowly* the system will indeed approach the ground state(s) but may do so extremely slowly.

KEY WORDS: Nonstationary Markov chains; annealing algorithm; annealing schedule; unbiased estimators.

1. INTRODUCTION

In this paper we study the asymptotic behavior as time $t \to +\infty$ of certain nonstationary Markov chains, and prove the convergence of the annealing

[1] Department of Mathematics, Rutgers University, New Brunswick, New Jersey 08903.

0022-4715/85/0400-0073$04.50/0

algorithm in Monte Carlo simulations. We find that in the limit $t \to +\infty$, a nonstationary Markov chain may exhibit "phase transitions," and the law of large numbers may fail, in the sense that the sample means do not form a consistent sequence of estimators for the $t \to +\infty$ stationary state. However, under appropriate conditions on the decay rate as $t \to +\infty$ of the one-step transition probabilities for nonstationary Markov chains in general, and the annealing algorithm in particular, we show that the estimators are in fact consistent, albeit biased. We compute explicitly the leading terms in the bias and the variance of such an estimator. We find that the annealing algorithm converges if the temperature $T(t)$ goes to zero as $t \to +\infty$ no faster than $C/\log(t/t_0)$. We given an (in general optimal) expression for the constant C in terms of the energies. Here t_0 is the initial time (in the rest of the paper we set $t_0 = 1$). The bias and the variance of the sample-means estimator in the annealing algorithm go to zero like $O(t^{-1+\varepsilon})$ for some $0 \leqslant \varepsilon < 1$, with $\varepsilon = 0$ only in very spcial circumstances.

The Metropolis algorithm was originally introduced[17] for studying numerically the equilibrium properties of statistical-mechanical systems at a given temperature. Simulations based on Metropolis-type Monte Carlo techniques have been used extensively in the study[18] of time evolution of spin and other lattice systems. The annealing algorithm is a modification of the Metropolis algorithm, in which the temperature is varied with time according to an "annealing schedule" $T(t)$. Simulated annealing has been important in Monte Carlo studies of "random systems" (in particular "spin glasses") in statistical mechanics,[15,20,1] and it has been used as an empirical test for a first-order phase transition in lattice gauge theories.[4,5] Recently, it has been proposed[16] for use as an optimization technique, and it has been applied successfully on a number of combinatorial optimization problems including the traveling salesman problem and certain other problems (known as NP-complete problems) arising in computer design. In Ref. 8, the annealing algorithm was introduced as a tool in computer vision, and the first rigorous result, concerning the convergence of the algorithm, was established. Our present mathematical work grew out of Metropolis-type Monte Carlo numerical experiments we are currently performing, concerning the restoration of degraded images, and edge and object detection in digital images.

A basic question in statistical-mechanical systems concerns their low-temperature behavior, which is controlled by the ground states and other states near them in energy. Experimentally, the ground state of a system can be reached by first "melting" the substance and then cooling it *slowly*, being careful to pass especially slowly through the "freezing" temperature (if any). If the temperature is lowered *too abruptly*, then the system may end up not in a ground state, but in a nearby metastable state, i.e., in a local but not global minimum of the energy (we refer to this phenomenon as the "adiabatic" effect). If, on the other hand, the temperature is lowered *too slowly*, then the system will indeed approach the ground state(s), but may do so extremely slowly (we refer to this phenomenon as the "super-cooling" effect). The optimal choice of the annealing schedule in Monte Carlo simulations is determined by the competition between these two effects. [For random systems (spin glasses), simulations are complicated further by the fact that these systems seem to have not one but many nearby almost-degenerate random ground states.]

In this paper, we treat the annealing algorithm as a special case of the theory of nonstationary Markov chains, and provide an (in general optimal) lower bound on the rate at which the temperature must be lowered in order to reach the ground state. Our results concerning the convergence of the annealing algorithm, and the rate of convergence to zero of the bias and the variance of the sample-means estimator, provide a rigorous procedure for choosing the best annealing schedule.

We now describe briefly our main results: Let $\{X^{(t)}\}$ be a discrete-time ($t = 0, 1, 2,...$), nonstationary Markov chain with *finite* state space

$$\Omega = \{s_1, s_2, ..., s_n\} \tag{1.1}$$

one-step transition probabilities

$$p_{ij}(t) \equiv p_{ij}^{(t-1,t)} = P(X^{(t)} = s_j \mid X^{(t-1)} - s_i), \qquad i, j = 1, ..., n \tag{1.2}$$

and initial probability distribution

$$\alpha_i^{(0)} = P(X^{(0)} = s_i), \qquad i = 1, ..., n \tag{1.3a}$$

$$\alpha_i^{(0)} \geqslant 0, \qquad \sum_{i=1}^{n} \alpha_i^{(0)} = 1 \tag{1.3b}$$

We assume that the limit

$$p_{ij}^{(t-1,t)} \to p_{ij} \text{ as } t \to +\infty, \qquad i, j = 1, ..., n \tag{1.4}$$

exists, and that the limiting matrix $P = P(\infty) = (p_{ij})$ has one or more (irreducible) ergodic (recurrent) sets, and perhaps some transient states. Aggregating the states properly, P takes the form

$$P = \begin{array}{c} \begin{array}{ccccc} \underbrace{r_1} & \underbrace{r_2} \cdots & & \underbrace{r_m} & \underbrace{n-\sum_{\gamma=1}^{m} t_\gamma} \end{array} \\ \left[\begin{array}{c:c:c:c:c} S^{(1)} & 0 & 0 & 0 & 0 \\ \hdashline 0 & S^{(2)} & 0 & 0 & 0 \\ \hdashline 0 & 0 & \ddots & 0 & 0 \\ \hdashline 0 & 0 & 0 & S^{(m)} & 0 \\ \hdashline L^{(1)} & L^{(2)} & & L^{(m)} & R \end{array}\right] \end{array} \begin{array}{l} \} \, r_1 \\ \} \, r_2 \\ \\ \} \, r_m \\ \} \, n-\sum_{\gamma=1}^{m} r_\gamma \end{array} \tag{1.5}$$

where $S^{(\gamma)}$, $\gamma = 1,..., m$ are the $r_\gamma \times r_\gamma$ transition matrices for the m ergodic sets, R concerns the process as long as it stays in the $n - \sum_{\gamma=1}^{m} r_\gamma$ transient states, and $L^{(\gamma)}$, $\gamma = 1,..., m$ concern transitions from the transient states into the ergodic sets $S^{(\gamma)}$, $\gamma = 1,..., m$, respectively. The regions 0 consist entirely of zeros. We will be concerned mainly with the case when the ergodic states are aperiodic. Corresponding to the form (1.5) on P, the matrix $P(t) = P^{(t-1,t)} = (p_{ij}^{(t-1,t)})$ has the form

$$P(t) = p^{(t-1,t)} =$$

$$\left[\begin{array}{c:c:c:c:c} S^{(1)} + V_{(t)}^{(1,1)} & V_{(t)}^{(1,2)} & V_{(t)}^{(1,3)} \cdots & V_{(t)}^{(1,m)} & V_{(t)}^{(1,m+1)} \\ \hdashline V_{(t)}^{(2,1)} & S^{(2)} + V_{(t)}^{(2,2)} & V_{(t)}^{(2,3)} \cdots & V_{(t)}^{(2,m)} & V_{(t)}^{(2,m+1)} \\ \hdashline & & \ddots & & \\ \hdashline V_{(t)}^{(m,1)} & V_{(t)}^{(m,2)} & V_{(t)}^{(m,3)} \cdots & S^{(m)} + V_{(t)}^{(m,m)} & V_{(t)}^{(m,m+1)} \\ \hdashline L^{(1)} + V_{(t)}^{(m+1,1)} & L^{(2)} + V_{(t)}^{(m+1,2)} & & L^{(m)} + V_{(t)}^{(m+1,m)} & R + V_{(t)}^{(m+1,m+1)} \end{array}\right] \tag{1.6}$$

where the matrices

$$V^{(k,l)}(t) \to 0 \text{ as } t \to +\infty, \qquad k, l = 1, 2,..., m$$

By the well-known Peron–Frobenius theorem, at each epoch t, the matrix $P(t)$ has an invariant ("equilibrium") probability vector $\pi(t) = (\pi_1(t),..., \pi_n(t))$, i.e., there exists a vector $\pi(t)$ such that

$$\pi_j(t) = \sum_{i=1}^{n} \pi_i(t)\, p_{ij}^{(t-1,t)} \tag{1.7a}$$

$$\sum_{i=1}^{n} \pi_i(t) = 1 \tag{1.7b}$$

$$\pi_j(t) \geq 0, \qquad j = 1, \ldots, n \tag{1.7c}$$

The higher transitions probabilities $p_{ij}^{(t_0,t)}$ are defined by

$$p_{ij}^{(t_0,t)} = P(X^{(t)} = s_j \mid X^{(t_0)} = s_i) \qquad \text{for} \quad t_0 < t$$

and satisfy the Chapman–Kolmogorov equation

$$p_{ij}^{(t_0,t)} = \sum_{l=1}^{n} p_{il}^{(t_0,t')} p_{lj}^{(t',t)}, \qquad \text{for} \quad t_0 < t' < t \tag{1.8}$$

and also

$$\sum_{j=1}^{n} p_{ij}^{(t_0,t)} = 1 \tag{1.9}$$

We will also use the absolute (or unconditional) probabilities

$$\alpha_j^{(t)} = P(X^{(t)} = s_j) = \sum_{i=1}^{n} \alpha_i^{(0)} p_{ij}^{(0,t)} \tag{1.10a}$$

which satisfy

$$\alpha_j^{(t)} = \sum_{i=1}^{n} \alpha_i^{(t_0)} p_{ij}^{(t_0,t)}, \qquad \text{for} \quad 0 \leq t_0 \leq t-1 \tag{1.10b}$$

$$\sum_{j=1}^{n} \alpha_j^{(t)} = 1 \tag{1.10c}$$

Assuming that the limit (1.4) exists and that $\pi(t)$ is unique, we are concerned mainly with four questions:

(i) Does the limit of $p_{ij}^{(0,t)}$ exist as $t \to +\infty$, and if yes, is the limit independent of the initial state i?

(ii) Does $\lim_{t \to +\infty} \pi_j(t)$ exist?

(iii) If both of the above limits exist, are they the same, i.e., does

$$\lim_{t \to +\infty} p_{ij}^{(0,t)} = \lim_{t \to +\infty} \pi_j(t) \tag{1.11}$$

hold?

(iv) Do the bias and the variance of the sample-means estimator (ergodic average)

$$\frac{1}{t}\sum_{s=0}^{t} f(X^{(s)})$$

converge to zero as $t \to +\infty$, and if yes, what is the rate of convergence? Here f is a function on the Markov chain $\{X^{(t)}\}$.

If the limiting matrix P has only one ergodic component (and possibly transient states), then $p^{(0,t)}$ and $\pi(t)$ have limits as $t \to +\infty$, and (1.11) holds always. But if P has two or more ergodic components then everything can happen. The examples of the Appendix show that any of the following possibilities may occur: (a) neither $\pi(t)$ nor $P^{(0,t)}$ has a limit; (b) $\pi(t)$ has no limit, but $P^{(0,t)}$ has a limit, and furthermore the limit of of $p_{ij}^{(0,t)}$ may or may not depend on the initial state i; (c) both $\pi(t)$ and $P^{(0,t)}$ have limits, but the limit of $p_{ij}^{(0,t)}$ depends on the initial state i, and therefore (1.11) does not hold. We do not know whether the following possibility occurs: (d) $\pi(t)$ has a limit but not $P^{(0,t)}$. We believe that case (d) does not occur. In the special case of the annealing algorithm $\pi(t)$ and $P^{(0,t)}$ always have a limit, but (1.11) may fail because the limit of $p_{ij}^{(0,t)}$ depends on i (this occurs when the temperature goes to zero sufficiently fast). All the limits above are ordinary limits, because we assume that the states are aperiodic with respect to the matrix P. Some of our results hold for periodic states provided that the limits are taken in the sense of some summability method such as Euler or Cesaro means.

If P has more than one ergodic component, then the limit of $\pi(t)$ may fail to exist no matter how fast the decay rate in (1.4) is, while if the decay rate in (1.4) is fast enough $p_{ij}^{(0,t)}$ always has a limit which may, however, depend on the initial state i. If both $\pi(t)$ and $P^{(0,t)}$ have limits, then a necessary condition for (1.11) to hold is

$$\sum_{t=1}^{+\infty} \operatorname{Tr}(I - P^{(t-1,t)}) = \sum_{t=1}^{+\infty} \sum_{j=1}^{n} (1 - p_{jj}^{(t-1,t)}) = +\infty \tag{1.12}$$

where I is the identity matrix. Sufficient conditions for (1.11) to hold are given in Theorems 1.1 and 1.2 below and in Section 2. If (1.11) holds, then the bias and the variance of the sample-means estimator go to zero as $t \to +\infty$, but they may do so slowly. Theorem 1.3 provides conditions under which the variance converges to zero like $O(t^{-1+\varepsilon})$ for some $0 \leqslant \varepsilon < 1$. These conditions yield a procedure for choosing the optimal annealing schedule for the annealing algorithm.

The intuitive reason for the nonexistence of $\lim_{t\to+\infty} \pi(t)$ (as well as of $\lim_{t\to+\infty} P^{(0,t)}$) is the occurrence of "phase transitions" when P has more than one ergotic component: Let P have the form (1.5) with $m \geqslant 2$, and let

$$\tilde{\mu}^{(\gamma)} = (\mu_1^{(\gamma)}, ..., \mu_{r_\gamma}^{(\gamma)}), \qquad \gamma = 1, 2, ..., m$$

be the unique equilibrium probability distributions (Ref. 7, p. 394) of the ergodic matrices $S^{(\gamma)}$, $\gamma = 1, ..., m$, i.e., the unique probability vectors that satisfy

$$\begin{gathered} \mu_j^{(\gamma)} = \sum_{i=1}^{r_\gamma} \mu_i^{(\gamma)} S_{ij}^{(\gamma)}, \qquad \gamma = 1, ..., m \\ \mu_j^{(\gamma)} > 0, \qquad \sum_{j=1}^{r_\gamma} \mu_j^{(\gamma)} = 1 \end{gathered} \tag{1.13a}$$

Let

$$\mu^{(\gamma)} = (\underbrace{0, ..., 0}_{r_1 + \cdots + r_{\gamma-1}}, \mu_1^{(\gamma)}, ..., \mu_{r_\gamma}^{(\gamma)}, \underbrace{0, ..., 0}_{n - \sum_{l=1}^{m} r_l}), \qquad \gamma = 1, ..., m \tag{1.13b}$$

Then any equilibrium distribution μ of P is a convex combination of $\mu^{(1)}, ..., \mu^{(m)}$, i.e.,

$$\begin{gathered} \mu = \xi^{(1)} \mu^{(1)} + \cdots + \xi^{(m)} \mu^{(m)} \\ \sum_{\gamma=1}^{m} \xi^{(\gamma)} = 1, \qquad \xi^{(\gamma)} \geqslant 0, \qquad \gamma = 1, ..., m \end{gathered} \tag{1.14}$$

What then may happen is that different subsequences of $\pi(t)$ [or $p_{i\cdot}^{(0,t)}$] may converge to different convex combinations of the $\mu^{(\gamma)}$s. (In the annealing algorithm this cannot occur for $\pi(t)$, but $p_{i\cdot}^{(0,t)}$ may converge to a convex combination that depends on i).

In order to state our first main result, we will need the probability that a transient state falls eventually into an ergodic component: Here and through this paper we shall denote the state in the γth, $\gamma = 1, ..., m$, ergodic component by $\bar{S}^{(\gamma)}$, and the transient states by $\bar{R}$. Let $\{\hat{X}^{(t)}\}$ be the stationary Markov chain associated with the limiting transition probability matrix P in (1.5). Let

$$z_{j\gamma} = P\{\hat{X}^{(t)} \in S^{(\gamma)} \text{ for some } t = 1, 2, ... \mid \hat{X}^{(0)} = s_j\},$$
$$j = r_1 + \cdots + r_m + 1, ..., n, \quad \gamma = 1, ..., m \tag{1.15}$$

Clearly

$$z_{j\gamma} \geqslant 0, \qquad \gamma = 1,..., m, j \in \bar{R} \ (j = r_1 + \cdots + r_m + 1,..., n) \qquad (1.16a)$$

$$\sum_{\gamma=1}^{m} z_{j\gamma} = 1, \qquad \text{for every} \quad j = r_1 + \cdots + r_m + 1,..., n \qquad (1.16b)$$

The probabilities $z_{j\gamma}$ can be computed explicitly by considering the $\lim_{k \to +\infty} P^k$. See formula (3.17) for the case when P has two ergodic components. There is a similar formula for the general case.

Here is our first result:

Theorem 1.1. Let $p_{ij}^{(t-1,t)}$ be the one-step transition probabilities of a discrete-time, nonstationary, finite Markov chain which converges to p_{ij} as $t \to +\infty$.

(1) If $P = (p_{ij})$ has a single ergodic aperiodic component $\bar{S}^{(1)}$ and possibly transient states $\bar{R}$, then $\pi(t)$ and $P^{(0,t)}$ have limits as $t \to +\infty$. Furthermore, if $\mu = (\mu_1, \mu_2,..., \mu_{r_1}, 0,..., 0)$ is the unique equilibrium probability distribution of P, then

$$\lim_{t \to +\infty} p_{ij}^{(0,t)} = \mu_j, \qquad j = 1,..., n, \text{ and all } i = 1,..., n \qquad (1.17a)$$

and

$$\lim_{t \to +\infty} \pi(t) = \mu \qquad (1.17b)$$

(2) Suppose that P has exactly two ergodic aperiodic components $\bar{S}^{(1)}$ and $\bar{S}^{(2)}$, and possibly transient states $\bar{R}$, i.e., P is of the form (1.5) with $m = 2$. Let $\mu^{(1)}$, $\mu^{(2)}$ be as in (1.13b), and z_{j1}, z_{j2}, $j = r_1 + r_2 + 1,..., n$ as in (1.15). Let

$$p_{ij}^{(t-1,t)} = p_{ij} + V_{ij}(t), \qquad i, j = 1,..., n \qquad (1.18)$$

and

$$\phi(t) = \sum_{i=1}^{r_2} \mu_i^{(2)}(V_{r_1+i,1}(t) + \cdots + V_{r_1+i,r_1}(t) + V_{r_1+i,r_1+r_2+1}(t) z_{r_1+r_2+1,1} + \cdots + V_{r_1+i,n}(t) z_{n,1}) \qquad (1.19a)$$

$$\psi(t) = \sum_{i=1}^{r_1} \mu_i^{(1)}(V_{i,r_1+1}(t) + \cdots + V_{i,r_1+r_2}(t) + V_{i,r_1+r_2+1}(t) z_{r_1+r_2+1,2} + \cdots + V_{i,n}(t) z_{n,2}) \qquad (1.19b)$$

Then:

(i) If

$$\sum_{t=1}^{+\infty} [\phi(t)+\psi(t)] < +\infty \tag{1.20}$$

then for each $1 \leqslant i \neq k \leqslant n$,

$$\limsup_{t\to+\infty} |p_{ij}^{(0,t)} - p_{kj}^{(0,t)}| \neq 0, \qquad j=1,..., r_1+r_2 \tag{1.21}$$

(ii) If

$$\sum_{t=1}^{+\infty} [\phi(t)+\psi(t)] = +\infty \tag{1.22}$$

then for each $1 \leqslant i,\ k \leqslant n$,

$$\lim_{t\to+\infty} \sup_{i,k} |p_{ij}^{(0,t)} - p_{kj}^{(0,t)}| = 0, \qquad j=1,..., n \tag{1.23}$$

Furthermore, if in addition to (1.22), the invariant probability vector $\pi(t)$ has a limit as $t \to +\infty$, then $\lim_{t\to+\infty} p_{ij}^{(0,t)}$ exists, and we have

$$\lim_{t\to+\infty} p_{ij}^{(0,t)} = \lim_{t\to+\infty} \pi_j(t), \qquad j=1,..., n \tag{1.24}$$

independently of the initial state i.

Note that $\phi(t)$ and $\psi(t)$ involve only the entries of the matrices $V^{(2,1)}(t)$, $V^{(2,3)}(t)$, and $V^{(1,2)}(t)$, $V^{(1,3)}(t)$, respectively [in the representation (1.6) with $m=2$]. Hence $\phi(t)$, $\psi(t)$ are strictly positive (and go to zero as $t \to +\infty$). Theorem 1.1, together with some other results, are proven in Section 3. Part (1) is intuitively obvious because of the nonoccurrence of phase transitions [i.e., because of the uniqueness of the equilibium vector μ of P (Ref. 13, Theorem 6.2.1)]. The proof of this part is simple. In contrast, the proof of part (2) is more delicate. We do not known whether such a sharp theorem holds when P has more than two ergodic aperiodic components (see related remarks in Section 3). The example of the Appendix shows that $\pi(t)$ may fail to have a limit under either condition (1.20) or condition (1.22). This example also shows that (1.22) alone does not necessarily imply the existence of $\lim_{t\to+\infty} p_{ij}^{(0,t)}$. For the annealing algorithm, condition (1.22) provides a sharp value of the constant C we mentioned in the beginning of this Introduction.

The next theorem is weaker than Theorem 1.1, but it holds even if the limit (1.4) does not exist.

Theorem 1.2. Let $p_{ij}^{(t-1,1,t)}$ be the one-step transition probabilities of a discrete-time, nonstationary, finite Markov chain with an invariant probability vector $\pi(t)$ satisfying (1.7).

(1) Suppose that there exists an integer N such that for a fixed integer $t_0 = 0, 1, 2,...,$ we have

$$\min_{i,k} \sum_{j=1}^{n} \min\{ p_{ij}^{(t_0+(v-1)N, t_0+vN)}, p_{kj}^{(t_0+(v-1)N, t_0+vN)} \} = C_v(t_0) \quad (1.25a)$$

with

$$\sum_{v=1}^{+\infty} C_v(t_0) = +\infty \quad (1.25b)$$

then

$$\lim_{t \to +\infty} \max_{i,k} | p_{ij}^{(t_0,t)} - p_{kj}^{(t_0,t)} | = 0 \quad (1.26)$$

(2) Suppose now that (1.25) holds for *every* integer $t_0 = 0, 1, 2,...$. Also assume that for some $T \geqslant 0$

$$\sum_{t=T}^{+\infty} \sum_{j} |\pi_j(t) - \pi_j(t+1)| < +\infty \quad (1.27)$$

Then $\lim_{t \to +\infty} \pi(t)$, and $\lim_{t \to +\infty} P^{(0,t)}$ exist, and if

$$\lim_{t \to +\infty} \pi_j(t) = \pi_j, \qquad j = 1,..., n \quad (1.28)$$

then

$$\lim_{t \to +\infty} p_{ij}^{(0,t)} = \pi_j, \qquad j = 1,..., n \quad (1.29)$$

independently of the initial state i.

The proofs of Theorems 1.2 and 1.2 are entirely different. Theorem 1.2 (with a minor change, see Section 2) is apparently known in the literature (Ref. 11, Theorems V.3.2 and V.4.3).[2] Accordingly, in Section 2, we only

[2] We thank one of the referees for bringing to our attention Refs. 6, 9, and 11, and further references on nonstationary Markov chains contained in Refs. 9 and 11. The quantity (1.25a) was introduced by Dobrushin[(6)] and is known as Dobrushin's ergodic coefficient. Our independent introduction was motivated by the proof of Theorem 4.1.3 of Ref. 13 for stationary Markov chains. This led us to the computation of the best constant in Lemma 2.1 which provides the basic estimate in the proof of Theorem 1.2. This lemma is equivalent to Lemma V.2.4 of Ref. 11. Reference 9 contains an alternative proof of (1.26).

Part (1) of Theorem 1.1 is also contained in Ref. 11, Theorem V.4.5, but our proof in Section 3 is new. Also, the proof of the entire Theorem 1.1 is very different from the circle of ideas involved in the proof of Theorem 1.2.

outline the proofs of Theorem 1.2 and some variants of it, and present mainly some technical estimates which are needed in the rest of the paper. Condition (1.22) is, in general, sharper than condition (1.25), and much easier to verify in practice. The two conditions are equivalent for the Example of the Appendix.

Condition (1.27) is clearly satisfied if the limit (1.28) exists and is achieved monotonically (perhaps for some js from above and for other js from below). This is the case in the annealing algorithm (see Theorem 1.4 below and Section 5). Condition (1.27) is not needed in Theorem 1.1 or the Example of the Appendix. However, we suspect that Theorem 1.2 does not hold in general without condition (1.27) or some other alternative condition. Theorem 2.3 is a variant of Theorem 1.2, where condition (1.27) is replaced by a condition concerning the divergence rate of (1.25b) [see condition (2.11)]. Condition (1.27) alone [i.e., without condition (1.25)] implies (see Proposition 2.1) that the limiting vector π_j is, in a sense, an asymptotic equilibrium vector for the nonstationary Markov chain. It would be interesting to know wether the physically reasonable result of Proposition 2.1 holds under a more natural condition than (1.27).

Our next result ammounts to an ergodic theorem for nonstationary Markov chains: Let f be a function of the Markov chain $\{X^{(t)}\}$. We set

$$Y^{(t)}=\frac{1}{t}\sum_{s=0}^{t} f(X^{(s)}) \tag{1.30}$$

and denote by $E_\sigma\{\cdot\}$ expectation values in the nonstationary Markov chain with transition probabilities (1.2) and with initial probability vector σ.

Theorem 1.3. Let $p_{ij}^{(t-1,t)}$ be the one-step transition probabilities of a discrete-time, nonstationary, finite Markov chain with a unique invariant probability vector $\pi(t)$. Assume that $p_{ij}^{(t-1,t)}$ converges to p_{ij} as $t\to+\infty$. Assume further that either (i) $P=(p_{ij})$ has a single ergodic aperiodic component and possibly transient states, or (ii) P has two ergodic components and possibly transient states, and $P^{(t-1,t)}$ satisfies (1.22) and $\lim_{t\to+\infty}\pi(t)$ exists (call it π), or (iii) P has the form (1.5) with $m\geqslant 2$, and $P^{(t-1,t)}$ satisfy (1.25), and (1.27). Then we have, for any probability vector σ:

(a)

$$\lim_{t\to+\infty} E_\sigma\{Y^{(t)}\}=\langle f\rangle \tag{1.31a}$$

where

$$\langle f\rangle=\sum_{i=1}^{n} f_i\pi_i,\qquad f_i=f(s_i) \tag{1.31b}$$

and

$$\lim_{t \to +\infty} E_\sigma \left\{ \left(Y^{(t)} - \sum_i f_i \pi_i \right)^2 \right\} = 0 \tag{1.32}$$

(b) Suppose that for some $0 \leqslant \varepsilon < 1$, the limit

$$w_{ij} = \lim_{t \to +\infty} \frac{1}{t^\varepsilon} \sum_{s=1}^{t} (p_{ij}^{(0,s)} - \pi_j) \tag{1.33}$$

exists (and is finite), then

$$\lim_{t \to +\infty} t^{1-\varepsilon} E_\sigma \left\{ Y^{(t)} - \sum_i f_i \pi_i \right\} = \sum_{i,j} \sigma_i w_{ij} f_j \tag{1.34}$$

Furthermore, if in addition to (1.33), we have

$$\limsup_{t \to +\infty} \frac{1}{t^{1+\varepsilon}} \sum_{\substack{s,\tau=1 \\ s<\tau}}^{t} |p_{ij}^{(s,\tau)} - \pi_j| < +\infty \qquad \text{for each } i, j \tag{1.35}$$

then (a) for $\varepsilon = 0$,

$$\begin{aligned} &\lim_{t \to +\infty} t E_\sigma \{ (Y^{(t)} - \langle f \rangle)^2 \} \\ &= \sum_{i,j} f_i f_j \Bigg[(\pi_i \delta_{ij} - \pi_i \pi_j) \\ &\quad + 2\pi_i \lim_{t \to +\infty} \frac{1}{t} \sum_{\substack{s,\tau=1 \\ s<\tau}}^{t} (p_{ij}^{(s,\tau)} - \pi_j) \\ &\quad + 2 \lim_{t \to +\infty} \frac{1}{t} \sum_{\substack{s,\tau=1 \\ s<\tau}}^{t} \left(\sum_k \sigma_k p_{ki}^{(0,s)} - \pi_i \right) (p_{ij}^{(s,\tau)} - \pi_j) \Bigg] \end{aligned} \tag{1.36}$$

and (b) for $0 < \varepsilon < 1$

$$\begin{aligned} &\lim_{t \to +\infty} t^{1-\varepsilon} E_\sigma \{ (Y^{(t)} - \langle f \langle)^2 \} \\ &= \sum_{i,j} f_i f_j \Bigg\{ 2\pi_i \lim_{t \to +\infty} \frac{1}{t^{1+\varepsilon}} \sum_{\substack{s,\tau=1 \\ s<\tau}}^{t} (p_{ij}^{(s,\tau)} - \pi_j) \\ &\quad + 2 \lim_{t \to +\infty} \frac{1}{1+\varepsilon} \sum_{\substack{s,\tau=1 \\ s<\tau}}^{t} \left(\sum_k \sigma_k p_{ki}^{(0,s)} - \pi_i \right) (p_{ij}^{(s,\tau)} - \pi_j) \Bigg\} \end{aligned} \tag{1.37}$$

This theorem is proven in Section 4. The limit (1.31) and (1.32) together imply that $Y^{(t)}$ is a consistent estimator of $\langle f \rangle$, i.e., $Y^{(t)}$ converges in probability to $\langle f \rangle$: for every $\delta > 0$

$$\lim_{t \to +\infty} E_\sigma\{|Y^{(t)} - \langle f \rangle| > \delta\} = 0$$

This estimator is biased, and (1.34) gives the leading term in the large-t expression of the bias. A particular feature of Theorem 1.3 is that the leading terms in both the bias (1.34) and the variance (1.37) may be of the order $t^{-1+\varepsilon}$ with $\varepsilon > 0$, and both terms depend on the intial distribution σ. For stationary Markov chains the leading term of the variance is of order t^{-1}, and is independent of σ, while the leading term in the bias is of order t^{-1} or smaller, and, in general, it depends on σ. For nonstationary Markov chains the size of ε in (1.33) [and (1.35)] depends on two different effects: (i) the rate of divergence of the series (1.22), or (1.25b), which control the rate of the limit (1.23), and (ii) the rate of convergence of $\pi(t)$ to π as $t \to +\infty$. The faster the series (1.22), (1.25b) diverge [i.e., the slower $\phi(t) + \psi(t)$, and $C_v(t_0)$ go to zero as $t \to +\infty$, $v \to +\infty$, respectively] the faster the limit (1.23) [or (1.26)] is achieved. In turn, the faster the limits (1.23) [or (1.26)] and (1.28) are attained, the faster the limit (1.29) is attained and the smaller the ε is. In the annealing algorithm the two effects are competitive: The series (1.22) [or (1.25b)] diverges fast if the temperature $T(t)$ goes to zero *slowly*. On the other hand, $\pi(t)$ converges to π fast if $T(t)$ goes to zero *slowly*. Thus the first effect requires that as $t \to +\infty$, $T(t) \geqslant C + \delta_1/\log t$ for some $\delta_1 > 0$, where C is the constant we mentioned in the beginning of this Introduction. This corresponds to the "adiabatic" effect we mentioned before. On the other hand, $\pi(t)$ converges to π like $\exp\{-[1/T(t)](U_2 - U_1)\}$ where U_1 is the energy of the ground state(s), and U_2 the nergy of the next excited state(s). Thus the second effect above requires that $T(t) \leqslant (U_2 - U_1 - \delta_2)/\log t$ as $t \to +\infty$, with some $\delta_2 > 0$. This corresponds to the "super-cooling" effect we mentioned before. However, only in very spcial circumstances do we have $C < U_2 - U_1$. Thus in general it is impossible to choose the annealing schedule $T(t)$ so that the limits (1.33) and (1.35) exist with $\varepsilon = 0$. Hence the mean-square error, i.e., $(\text{bias})^2 + \text{variance}$, can be minimized by choosing $T(t) = C/\log t$, where $C + \delta < \bar{C} < C + \delta'$ for some $0 < \delta \leqslant \delta'$, and C the optimal constant (see Theorem 1.4 below).

Dobrushin has established Ref. 6[3] a central limit theorem (CLT) for

[3] We thank again the referee who pointed out Dobrushin's work.

nonstationary Markov chains. For the annealing algorithm, his results yield that the CLT (with a normal limiting distribution) holds if

$$T(t) \geqslant \frac{3\bar{C}}{\log t}, \qquad \text{large } t \tag{1.38}$$

where $\bar{C}$ is the constant above [actually Dobrushin's results yeild a little bit better than (1.38)]. We suspect that for the annealing algorithm, the result can be improved so that (1.38) holds without the factor 3.

We end this Introduction by stating our convergence theorem for the annealing algorithm in a nonstationary version of the sampling method of Metropolis *et al.*[17] Let $Q = (q_{ij})$ be the transition matrix of an arbitrary symmetric (i.e., $q_{ij} = q_{ji}$) and irreducible Markov chain; we refer to Q as the "proposal matrix." Let

$$U_j = U(s_j), \qquad j = 1,..., n, \; U_j > 0$$

be the energies associated with the states $s_1,..., s_n$. We define a time-dependent positive probability vector on the states by

$$\pi_j(t) = \frac{e^{-\beta(t) U_j}}{\sum_{i=1}^{n} e^{-\beta(t) U_i}}, \qquad j = 1,..., n \tag{1.39}$$

where $0 \leqslant \beta(t) = 1/T(t) < +\infty$. Here $T(t)$ is the temperature of the system at time t. (We use units in which Boltzmann's constant is 1.) We order the states to that

$$U_1 \leqslant U_2 \leqslant U_2 \leqslant \cdots \leqslant U_n$$

Following Metropolis *et al.*,[17] we define the one-step transition probabilities of a nonstationary Markov chain by

$$\begin{aligned} i \neq j, \; p_{ij}^{(t-1,t)} &= q_{ij} \min\{1, e^{-\beta(t)(U_j - U_i)}\} \\ &= \begin{cases} q_{ij}, & \text{if} \quad U_j \leqslant U_i \\ q_{ij}\, e^{-\beta(t)(U_j - U_i)}, & \text{if} \quad U_j > U_i \end{cases} \end{aligned} \tag{1.40a}$$

$$\begin{aligned} p_{ii}^{(t-1,t)} &= 1 - \sum_{j \neq i} p_{ij}^{(t-1,t)} \\ &= 1 - \sum_{\substack{j \neq i \\ U_j \leqslant U_i}} q_{ij} - \sum_{\substack{j \neq i \\ U_j > U_i}} q_{ij}\, e^{-\beta(r)(U_j - U_i)} \\ &= q_{ii} + \sum_{j : U_j > U_i} q_{ij}(1 - e^{-\beta(t)(U_j - U_i)}) \end{aligned} \tag{1.40b}$$

It is easily verified that $\pi(t)$ is an invariant probability vector of the matrix $P^{(t-1,t)}$ defined by (1.40). In fact $P^{(t-1,t)}$ satisfies the detailed balance (reversibility) condition

$$\pi_i(t)\, p_{ij}^{(t-1,t)} = \pi_j(t)\, p_{ji}^{(t-1,t)} \tag{1.41}$$

Clearly (1.41) implies (1.7a). We choose the proposal matrix Q so that $\pi(t)$ [defined by (1.39)] is the *only* invariant probability vector of $P^{(t-1,t)}$. Now assume that $\lim_{t\to+\infty}\beta(t)\equiv\beta_\infty$ exists. The case $\beta_\infty<+\infty$ is in most respects trivial compared with the case $\beta_\infty=+\infty$. Here we consider only the case $\beta_\infty=+\infty$. It is easily verifid that

$$\lim_{t\to+\infty} p_{ij}^{(t-1,t)} = p_{ij} \tag{1.42a}$$

where

$$i\neq j,\qquad p_{ij}=\begin{cases} q_{ij}, & \text{if } U_j\leqslant U_i\\ 0, & \text{if } U_j>U_i\end{cases} \tag{1.42b}$$

$$p_{ii}=1-\sum_{\substack{j\neq i\\ U_j\leqslant U_i}} q_{ij} \tag{1.42c}$$

Furthermore,

$$\lim_{t\to+\infty}\pi_j(t)=\pi_j\equiv\begin{cases}\dfrac{1}{|\bar{S}^{(1)}|}, & \text{if } j\in\bar{S}^{(1)}\\[2ex] 0, & \text{if } j\in\bar{S}^{(1)}\end{cases} \tag{1.43}$$

where $\bar{S}^{(1)}$ denotes the set of ground states (i.e., the states of energy U_1) and $|\bar{S}^{(1)}|$ the number of ground states. If the proposal matrix Q is such that the limiting matrix P defined by (1.42) has $m\geqslant 2$ (irreducible) ergodic components, then we reorder the states so that P takes the form (1.5). We denote by $\bar{S}^{(1)},\ldots,\bar{S}^{(m)}$ the ergodic states corresponding to the matrices $S^{(1)},\ldots,S^{(m)}$ of (1.5), and by $\bar{R}$ the transient states. Our assumptions on the matrix $P^{(t-1,t)}$ imply that starting from any state i, any other state j can be reached via a finite chain

$$A_{i\to j}\colon i=l_0\to l_1\to l_2\to\cdots\to l_{k-1}\to j=l_k \tag{1.44}$$

of allowable transitions (i.e., $q_{l_\alpha,l_{\alpha-1}}>0$). We denote by $\{A_{i\to j}\}$ all possible finite chains of the form (1.44), and define

$$E_{ij}=\min_{\{A_{i\to j}\}}\sum_{\alpha=1}^{k}\max\{0,\,U_{l_\alpha}-U_{l_{\alpha-1}}\} \tag{1.45}$$

$$\tilde{E}=\max_{i,j} E_{ij} \tag{1.46}$$

Let

$$E_\gamma \equiv E_{\bar{S}(\gamma),\bar{R}} = \min_{\substack{i \in \bar{S}^{(\gamma)} \\ j \in \bar{R}}} E_{ij}, \qquad \gamma = 1,..., m \tag{1.47a}$$

and

$$E = \min_\gamma E_\gamma \tag{1.48a}$$

Clearly $E \leqslant \tilde{E}$. We will see that the constants E and $\tilde{E}$ control the rate at which $\beta(t)$ is allowed to tend to infinity as $t \to +\infty$. In a more general sampling method which we introduce in Section 5, the proposal matrix Q is not symmetric [see (5.7) and (5.8)], and the corresponding constant E is defined as follows: Let

$$E_{\gamma\gamma'} \equiv E_{\bar{S}^{(\gamma)},\bar{S}^{(\gamma')}} = \min_{\substack{i \in \bar{S}^{(\gamma)} \\ j \in \bar{S}^{(\gamma')}}} E_{ij}, \qquad \gamma \neq \gamma' \tag{1.47b}$$

then

$$E = \min\{\min_\gamma E_\gamma, \min_{\gamma,\gamma'} E_{\gamma\gamma'}\} \tag{1.48b}$$

Here is our theorem concerning the convergence of the annealing algorithm for the nonstationary Metropolis sampling (1.40).

Theorem 1.4. Let $P^{(t-1,t)}$ be the one-step transition probabilities matrix defined by (1.40). Assume that the proposal matrix Q is chosen so that $\pi(t)$ [defined by (1.39)] is the unique invariant probability vector of $P^{(t-1,t)}$. Assuming $\beta(t) \to +\infty$ as $t \to +\infty$, we have the following: (1) If the proposal matrix Q is such that the limiting matrix P given by (1.42) has a single ergodic aperiodic component and possibly a set of transient states, then (1.29), (1.31), and (1.32) hold. Furthermore (i) if

$$\beta(t) \geqslant \frac{1+\delta}{U_2 - U_1} \log t, \qquad \text{for sufficiently large } t, \text{ some } \delta > 0 \tag{1.49}$$

then we have (1.34) with $\varepsilon = 0$, and (1.36).

(ii) If

$$\frac{1-\delta_1}{U_2 - U_1} \log t \leqslant \beta(t) \leqslant \frac{1-\delta_2}{U_2 - U_1} \log t, \qquad \text{for sufficiently large } t \tag{1.50}$$

with $0 < \delta_2 \leqslant \delta_1 < 1$, then (1.34) and (1.37) hold with some $0 < \delta_2 \leqslant \varepsilon \leqslant \delta_1 < 1$

(iii) If

$$\beta(t)=\frac{1}{U_2-U_1}\log t, \qquad \text{for sufficiently large } t \tag{1.51}$$

then we have (1.34) and (1.37) with the factor $t^{-\varepsilon}$ replaced by $(\log t)^{-1}$.

(2) If the proposal matrix Q is such that P has two or more ergodic aperiodic components and possibly a set of transient states, then there exists an (optimal) constant C_0 [see (1.56) and remarks following it] such that if

$$\beta(t)\leqslant C_0\log t, \text{ for sufficiently large } t \tag{1.52}$$

then)1.29), (1.31), and (1.32) hold, while if

$$\beta(t)\geqslant (C_0+\delta)\log t, \text{ as } t\to +\infty, \text{ for some } \delta>0 \tag{1.53}$$

then (1.29) cannot hold. Furthermore, (a) if

$$C_0\leqslant\frac{1}{U_2-U_1}$$

and

$$(C_{0-\delta})\log t\leqslant\beta(t)\leqslant C_0\log t, \qquad \text{for sufficiently large } t, \text{ some } \delta>0 \tag{1.54}$$

then we have (1.34) and (1.37) with some $0<\varepsilon<1$.

(b) If

$$C_0>\frac{1}{U_2-U_1}$$

and

$$\frac{1+\delta}{U_2-U_1}\log t\leqslant\beta(t)\leqslant(C_0-\delta)\log t, \qquad \text{sufficiently large } t \tag{1.55}$$

with

$$\delta=\left(C_0-\frac{1}{U_2-U_1}\right)\frac{U_2-U_1}{1+U_2-U_1}$$

then we have (1.34) with $\varepsilon=0$, and (1.36).

If P has exactly two ergodic components, then Theorem 1.1 yields that the optimal constant C_0 is given by

$$C_0=\frac{1}{E} \tag{1.56}$$

where E is defined by (1.48a). We conjecture that the optimal constant C_0 is given by (1.56) even in the case when P has more than two ergodic components. Part (2) of Theorem 1.4 holds if we replace C_0 by $\tilde{C}_0 = \tilde{E}^{-1}$ where $\tilde{E}$ is defined in (1.46). Of course, (b) does not occur in this case, since $\tilde{E} > U_2 - U_1$. Theorem 1.2 yields in general a worse constant than (1.56) (see Section 5). Theorem 1.4 together with similar results for a general class of sampling methods which include the Metropolis sampling method, as well as results concerning multidimensional random Markov fields are proven in Section 5.

2. DISCRETE NONSTATIONARY FINITE MARKOV CHAINS

Parts (1) and (2) of Theorem 1.2 are essentially Theorems V.3.2 and V.4.3, respectively, of Ref. 11. The only difference between Theorem 1.2 and the above theorems of Ref. 11, is that the quantity $C_v(t_0)$ in (1.25a) (i.e., Dobrushin's ergodic coefficient!) involves a fixed number N of one-step transition matrices, while the number $n_{j+1} - n_j$ of one-step transition matrices in the blocks of Theorem V.3.2 of Ref. 11 varies with j, and it may go to infinity as $j \to +\infty$. The condition of Theorem V.3.2 of Ref. 11 is necessary and sufficient for weak ergodicity, while our condition (1.25b) is only sufficient. However, condition (1.25b) is easier to verify in practice, and covers the case of the annealing algorithm.

Here, we only outline the proof of Theorem 1.2, and present mainly some technical estimates which are needed in the rest of the paper, and in particular in the determination of the number ε in Theorem 1.3. Also, because of its physical interpretation, we isolate (as we did in the first version of the paper) Proposition 2.1.

The following lemma provides the basic estimate in the proof of Theorem 1.2.

Lemma 2.1. Let $Q = (q_{ij})$ be a stochastic $n \times n$ matrix, $x = (x_1, ..., x_n)$ an n vector, and $y = xQ$. Let osc x denote the oscillation of a vector x, i.e.,

$$\operatorname{osc} x \equiv \max_{i,j} |x_i - x_j| = \max_i x_i - \min_i x_i$$

Then

$$\operatorname{osc} y \leqslant [1 - C(Q)] \operatorname{osc} x \tag{2.1}$$

where

$$C(Q) = \min_{i,k} \sum_{j=1}^{n} \min(q_{ij}, q_{kj}) \tag{2.2a}$$

$$= 1 - \frac{1}{2} \max_{i,k} \sum_{j=1}^{n} |q_{ij} - q_{kj}| \tag{2.2b}$$

Remark. This lemma was motivated by the proof of Theorem 4.1.3 of Ref. 13. The constant $1-C(Q)$ is the best constant for estimate (2.1). Lemma 2.1 is equivalent to Lemma V.2.4 of Ref. 11, and we refer to Ref. 11 for its proof.

Proof of (1.26). A straightforward repeated application of (2.1) yields

$$\max_{i,k} |p_{ij}^{(t_0,t)} - p_{kj}^{(t_0,t)}|$$

$$\leqslant \prod_{\nu=1}^{[(t-t_0)/N]} [1-C_\nu(t_0)] \max_{i,k} |p_{ij}^{(t_0+[(t-t_0)/N]N,t)} - p_{kj}^{(t_0+[(t-t_0)/N]N,t)}| \tag{2.3}$$

where $[\xi]$ denotes the greatest integer smaller or equal to ξ. If $(t-t_0)/N$ is an integer then the last factor in (2.3) does not appear, otherwise, we bound this factor by 2. Since

$$\prod_{\nu=1}^{+\infty} [1-C_\nu(t_0)] = 0 \tag{2.4}$$

if and only if (1.25b) holds, estimate (2.3) yields (1.26). ∎

If (1.27) holds, then the sequence $\{\pi_j(t)\}$, $j=1,..., n$, is a (bounded) Cauchy sequence, and therefore the limit (1.28) exists. As we mentioned in the Introduction, the following proposition says that (1.27) alone [i.e., without condition (1.25)] implies that $\pi=(\pi_1,..., \pi_n)$ is an equilbrium vector "asymptotically."

Proposition 2.1. If (1.28) holds then

$$\lim_{t_0 \leftarrow +\infty} \operatorname{sub}_{t \geqslant t_0} \sum_j \left| \sum_i p_{ij}^{(t_0,t)} \pi_i - \pi_j \right| = 0 \tag{2.5}$$

Proof. A slight variation of the procedure in the proof of Theorem V.4.3 of Ref. 11 yields

$$\sum_j \left| \sum_i p_{ij}^{(t_0,t)} \pi_i - \pi_j \right|$$

$$\leqslant \sum_{s=t_0+1}^{t-1} \sum_j |\pi_j(s) - \pi_j(s+1)| + \sum_j |\pi_j - \pi_j(t_0+1)| + \sum_j |\pi_j(t) - \pi_j| \tag{2.6}$$

This easily implies (2.5). ∎

Proof of Theorem 1.2:

$$\begin{aligned}\sum_j |p_{ij}^{(0,t)} - \pi_j| &= \sum_j \left| \sum_l (p_{il}^{(0,t_0)} - \pi_l)\, p_{lj}^{(t_0,t)} + \sum_i P_{ij}^{(t_0,t)} \pi_i - \pi_j \right| \\ &= \sum_j \left| \sum_l (p_{il}^{(0,t_0)} - \pi_l)(\max_m p_{mj}^{(t_0,t)} - p_{lj}^{(t_0,t)}) \right. \\ &\qquad \left. + \sum_i P_{ij}^{(t_0,t)} \pi_i - \pi_j \right| \\ &\leqslant \sum_j \left\{ (\max_m p_{mj}^{(t_0,t)} - \min_m p_{mj}^{(t_0,t)}) \sum_l |p_{il}^{(0,t_0)} - \pi_l| \right\} \\ &\qquad + \sum_j \left| \sum_i p_{ij}^{(t_0,t)} \pi_i - \pi_j \right|\end{aligned}$$

Using $\sum_l \pi_l = 1 = \sum_l p_{il}^{0,t_0)}$, we obtain

$$\sum_j |p_{ij}^{(0,t)} - \pi_j| \leqslant 2 \max_{i,kj} |p_{ij}^{(t_0,t)} - p_{kj}^{(t_0,t)}| + \sum_j \left| \sum_i p_{ij}^{(t_0,t)} \pi_i - \pi_j \right| \tag{2.7}$$

This together with (1.26) and (2.5), yields (1.29).

From (1.10a) we see that: if $p_{ij}^{(0,t)}$ has a limit as $t \to +\infty$, and the limit is independent of i, then $\lim_{t \to +\infty} p_{ij}^{(0,t)} = \lim_{t \to +\infty} \alpha_j^{(t)}$. We have also the following theorem:

Theorem 2.1. Suppose that condition (1.25) holds. Furthermore, suppose that the absolute probabilities $\alpha_j^{(t)}$ have a limit as $t \to +\infty$, say,

$$\lim_{t \to +\infty} \alpha_j^{(t)} = \alpha_j, \qquad j = 1, \dots, n \tag{2.8}$$

Then $p_{ij}^{(0,t)}$ has a limit as $t \to +\infty$, and

$$\lim_{t \to +\infty} p_{ij}^{(0,t)} = \alpha_j, \qquad j = 1, \dots, n \tag{2.9}$$

Remark. The interesting feature of Theorem 2.1 is that it does not require any condition analogous to (1.27). The proof of this theorem is the same as the proof of Theorem 1.2 once we establish

$$\lim_{t_0 \to +\infty} \sup_{t \geqslant t_0} \sum_j \left| \sum_i p_{ij}^{(t_0,t)} \alpha_i - \alpha_j \right| = 0 \tag{2.10}$$

This is obtained from (1.10b) in a straightforward manner.

Next we prove a variant of Theorem 1.2 by replacing condition (1.27) by condition (2.11) below.

Theorem 2.2. Suppose that $P^{(t-1,t)}$ is as in Theorem 1.2, and that (1.25) [but not (1.27)] holds. Furthermore, assume that the limit (1.28) exists, and that

$$1+\sum_{l=1}^{n}\prod_{v=l}^{n}[1-C_v(t_0)]\leqslant C<+\infty \tag{2.11}$$

is bounde as, $n\to+\infty$, uniformly in $t_0\geqslant 1$. Then (1.29) holds.

Remark. This theorem is essentially Theorem V.4.4. of Ref. 11. Condition (2.11) implies a result stronger than Proposition 2.1, i.e., it implies

$$\lim_{t\to+\infty}\sum_j\left|\sum_i P_{ij}^{(\bar t_0,t)}\pi_i-\pi_j\right|=0 \tag{2.12}$$

for every $\bar t_0\geqslant 0$. Since the proof of (2.12) is not directly transparent from the proof of Theorem V.4.4 of Ref. 11 because of the fixed number N in (1.25a), we spell out the details:

Proof of (2.12). Let $t_0\geqslant\bar t_0$, and consider

$$\begin{aligned}
&\sum_j\left|\sum_i p_{ij}^{(\bar t_0,t_0+vN)}\pi_i-\pi_j\right|\\
&\quad\leqslant\sum_j\left|\sum_l\left[\sum_i p_{il}^{(\bar t_0,t_0+(v-1)N)}\pi_i+\pi_l\right]p_{lj}^{[t_0+(v-1)N,t_0+vN]}\right|\\
&\qquad+\sum_j\left|\sum_l\{\pi_l-\pi_l[t_0+(v-1)N+1]\}\,p_{lj}^{[t_0+(v-1)N,t_0+vN]}\right|\\
&\qquad+\sum_j\left|\sum_l\pi_l[t_0+(v-1)N+1]\,p_{lj}^{[t_0+(v-1)N,t_0+vN]}-\pi_j\right|
\end{aligned}\tag{2.13}$$

We use Lemma 2.1 to bound the first two terms, and obtain

$$\begin{aligned}
&\sum_j\left|\sum_i p_{ij}^{(\bar t_0,t_0+vN)}\pi_i-\pi_j\right|\\
&\quad\leqslant[1-C_v(t_0)]\sum_j\left|\sum_i p_{ij}^{[\bar t_0,t_0+(v-1)N]}\pi_i-\pi_j\right|\\
&\qquad+[1-C_v(t_0)]\sum_j|\pi_j-\pi_j[t_0+(v-1)N+1]|\\
&\qquad+\sum_j\left|\sum_i\pi_i[t_0+(v-1)N+1]\,p_{ij}^{(t_0+(v-1)N,t_0+vN)}-\pi_j\right|
\end{aligned}\tag{2.14}$$

The last term is bounded by

$$\sum_j \left| \sum_i \pi_i[t_0+(\nu-1)N+1]\, p_{ij}^{[t_0+(\nu-1)N,t_0+\nu N]} - \pi_j \right|$$

$$\leqslant \sum_j |\pi_j[t_0+(\nu-1)N+1]-\pi_j|$$

$$+2\sum_{s=2}^{N} \sum_j |\pi_j[t_0+(\nu-1)N+2]-\pi_j| \tag{2.15}$$

This, together with (2.14), yields

$$\sum_j \left| \sum_i p_{ij}^{(\bar{t}_0,t_0+\nu N)}\pi_i - \pi_j \right| \leqslant [1-C_\nu(t_0)] \sum_j \left| \sum_i p_{ij}^{[\bar{t}_0,t_0+(\nu-1)N]}\pi_i - \pi_j \right|$$

$$+\ 2\sum_{s=1}^{N} \sum_j |\pi_j[t_0+(\nu-1)N+s]-\pi_j| \tag{2.16}$$

By (1.28), given $\varepsilon>0$ there exists $t_0(\varepsilon)>0$ such that

$$\sum_j |\pi_j(t)-\pi_j| < \varepsilon, \qquad \text{for} \quad t \geqslant t_0(\varepsilon) \tag{2.17}$$

Taking t_0 large enough, iterating the first term in the right-hand side of (2.16), and using (2.17) we obtain

$$\sum_j \left| \sum_i p_{ij}^{(\bar{t}_0,t_0+\nu N)}\pi_i - \pi_j \right|$$

$$\leqslant \sum_j \left| \sum_i p_{ij}^{(\bar{t}_0,t_0)}\pi_i - \pi_j \right| \prod_{l=1}^{\nu} [1-C_l(t_0)]$$

$$+\, 2N\varepsilon \left\{ 1 + \sum_{l=1}^{\nu} \sum_{k=l}^{\nu} [1-C_k(t_0)] \right\} \tag{2.18}$$

Using (1.25) and (2.11), we have from (2.18) for every $t_0 \geqslant 0$

$$\lim_{\nu \to +\infty} \sum_j \left| \sum_i p_{ij}^{(\bar{t}_0,t_0+\nu N)}\pi_i - \pi_j \right| = 0 \tag{2.19}$$

Applying this with t_0, t_0+1, t_0+2,..., t_0+N-1, in place of $\bar{t}_0$, we obtain (2.12). ∎

Remark. Clearly (2.11) is satisfied if $C_\nu(t_0) \geqslant C > 0$. This is the case when the limit (1.4) exists, and the matrix P has a single ergodic component with aperiodic states, and possibly some transient states. Also, since every stochastic matrix P with a single ergodic component has (Ref. 13, Theorem 6.2.1] a *unique* equilibrium distribution μ, every subsequence $\{\pi(t_n)\}$ converges, by (1.7a), as $n \to +\infty$ to μ, and therefore $\pi(t)$ converges to μ as $t \to +\infty$. These facts provide, via Theorem 2.2, a proof of part (1) of Theorem 1.1 (see also Theorem V.4.5 of Ref. 11) alternative to the one given in the next section. This result holds even if the ergodic states of P are periodic, provided that the limit in (1.17a) is taken in the sense that Cesaro means (Ref. 13, p. 101).

We end this section by establishing a theorem concerning the rate at which the limit (1.29) is attained. First, we observe that $C_v(t_0)$ [see (1.25a)], depends on $t_0 + vN$ only, i.e., $C_v(t_0) = C(t_0 + vN)$.

Theorem 2.3. Suppose that the assumptions of part (2) of Theorem 1.2 hold. Furthermore, suppose that $\pi_j(t)$ converges monotonically to π_j (for some js from above and the others from below). Suppose that for sufficiently large t, we have

$$C(t) \geqslant \frac{1-\kappa}{t}, \qquad \text{some } 0<\kappa<1 \tag{2.20}$$

$$|\pi(t)-\pi| \equiv \sum_j |\pi_j(t)-\pi_j| \leqslant \frac{a}{t^{1-\delta}}, \qquad \text{some } 0<\delta<1,\ a=\text{const} \tag{2.21}$$

Then for sufficiently large t

$$\sum_j |p_{ij}^{(0,t)}-\pi_j| \leqslant O\left(\frac{1}{t^{1-\varepsilon}}\right), \qquad \varepsilon=\min(\kappa,\delta) \tag{2.22}$$

Proof. Combining (2.7) with (2.3) and (2.6), and using the fact that $\pi_j(t)$ converges monotonically to π_j, we obtain

$$\sum_j |p_{ij}^{(0,t)}-\pi_j| \leqslant 2 \prod_{l=t_0+N}^{[t]} [1-C(l)] + \sum_i |\pi_j(t^0+1)-\pi_j| \tag{2.23}$$

The first term is estimated by using

$$\prod_{l=t_0+N}^{[t]} [1-C(l)] \leqslant \exp\left\{-\prod_{l=t_0+N}^{[t]} [1-C(l)]\right\} \tag{2.24}$$

and (2.20). The second term is estimated by (2.21). The two estimates together easily yield (2.22).

Remark. If the limiting matrix p_{ij} has a single ergodic aperiodic component, then by the remark below (2.19), the first term on the right-hand side of (2.23) converges to zero geometrically. Thus the rate of convergence of $p_{ij}^{(0,t)}$ to π_j is determined only by the rate of convergence of $\pi_j(t)$ to π_j.

3. CONTINUOUS-TIME NONSTATIONARY MARKOV CHAINS

In this section we prove Theorem 1.1 and establish a consequence of (1.12). Although in numerical simulations the time is always discrete, we introduce in this section continuous-time Markov chains, and prove Theorem 1.1 for such chains. The proof for discrete-time Markov chains is similar. We outline the modifications needed at the end of this section.

The basic problem of the theory of continuous-time Markov chains consists in finding all solutions of the Chapman–Kolmogorov identity (1.8) subject to the constraints (1.9) and $p_{ij}^{(t_0,t)} \geqslant 0$.

Given a stochastic matrix $p_{ij}(t)$, $t \geqslant 0$, $i, j = 1,..., n$, we define a solution of (1.8) and (1.9) via Kolmogorov's system of "forward" differential equations (Ref. 7, p. 472)

$$\frac{d}{dt} p_{ij}^{(t_0,t)} = -f_j(t)\, p_{ij}^{(t_0,t)} + \sum_{l=1}^{n} f_l(t)\, r_{lj}(t)\, p_{il}^{(t_0,t)}, \qquad t > t_0 \tag{3.1a}$$

$$p_{ij}^{(t_0,t_0)} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases} \tag{3.1b}$$

where

$$f_j(t) = 1 - p_{jj}(t), \qquad j = 1,..., n \tag{3.2a}$$

$$f_i(t)\, r_{ij}(t) = p_{ij}(t), \qquad \text{for } i \neq j,\ i, j = 1,..., n \tag{3.2b}$$

$$r_{jj}(t) \equiv 0 \qquad \text{for all } j = 1, 2,..., n \tag{3.2c}$$

Note that

$$0 \leqslant r_{ij}(t), \qquad \sum_j r_{ij}(t) = 1 \tag{3.3}$$

Setting

$$P^{(t_0,t)} = (p_{ij}^{(t_0,t)})$$

$$P(t) = [p_{ij}(t)]$$

equations (3.1) read

$$\frac{d}{dt} P^{(t_0,t)} = P^{(t_0,t)}[-I + P(t)] \tag{3.4a}$$

$$P^{(t_0,t_0)} = I \tag{3.4b}$$

where I is the $n \times n$ identity matrix. Also setting

$$x^{(i)}(t) = (p_{i1}^{(t_0,t)}, \dots, p_{in}^{(t_0,t)}) \tag{3.5}$$

and denoting by e_i the row vector with 1 in the ith entry and zero everywhere else, Eq. (3.1) reads

$$\frac{dx^{(i)}(t)}{dt} = x^{(i)}(t)[-I + P(t)] \tag{3.6a}$$

$$x^{(i)}(0) = e_i \tag{3.6b}$$

Taking differences, we derive an equation for $p_{ij}^{(t_0,t)} - p_{kj}^{(t_0,t)}$

$$\frac{d}{dt}[x^{(i)}(t) - x^{(k)}(t)] = [x^{(i)}(t) - x^{(k)}(t)][-I + P(t)] \tag{3.7a}$$

$$x^{(i)}(0) - x^{(k)}(0) = e_i - e_k \tag{3.7b}$$

Thus we are led to study the differential equation

$$x(t) = [x_1(t), \dots, x_n(t)]$$

$$\frac{dx(t)}{dt} = x(t)[-I + P(t)] \tag{3.8}$$

subject to the conditions

$$x_i(t) \geqslant 0, \qquad i = 1, \dots, n \tag{3.9a}$$

$$x(0) = \text{probability vector} \tag{3.9b}$$

or to some initial condition $x(0)$ which satisfies [see (3.7b)]

$$\sum_{i=1}^{n} x_i(0) = 0 \tag{3.10}$$

or to some general initial condition

$$x(0) = x_0 \tag{3.11}$$

It is an easy consequence of a classical theorem in ordinary differential equations[2] that the system (3.8)–(3.10) has a unique solution [provided that $P(t)$ has continuous coefficients] for all $t \geqslant 0$. The same is true for the systems (3.8), (3.10), and (3.8), (3.11). The problem we address here is the large-time behavior of all solutions of (3.8), especially when $x(t)$ is a probability vector [see (3.9)] or it satisfies (3.10).

Proposition 3.1. Suppose that $P(t)$ has continuous coefficients for $t \in [0, \infty]$. Suppose that

$$\lim_{t \to +\infty} \int_0^t \mathrm{Tr}(I - P(s))\, ds < +\infty \tag{3.12}$$

Then $P^{(0,t)}$ and its inverse $(P^{(0,t)})^{-1}$ are uniformly bounded as $t \to +\infty$. Furthermore, no non-identically zero solutions of (3.8) goes to zero as $t \to +\infty$. In particular,

$$\lim_{t \to +\infty} (P_{ij}^{(0,t)} - p_{kj}^{(0,t)}) \neq 0, \qquad i \neq k$$

Proof. Let

$$D(t) = \det P^{(t_0,t)}$$

Then by Theorem 7.3 of Ref. 2 we have from (3.4)

$$\frac{d}{dt} D(t) = D(t)\, \mathrm{Tr}(-I + P(t)) \tag{3.13a}$$

and therefore

$$D(t) = D(t_0) \exp\left[-\int_{t_0}^t ds\, \mathrm{Tr}(I - P(s)) \right] = \exp\left[-\int_{t_0}^t ds\, \mathrm{Tr}(I - P(s)) \right] \tag{3.13b}$$

By (3.12), $D(t)$ is nonzero as $t \to +\infty$. Therefore $P^{(0,t)}$ and $(P^{(0,t)})^{-1}$ are uniformly bounded as $t \to +\infty$. Since $P^{(0,t)}$ is a fundamental solution of (3.8), this implies that no solution of this equation goes to zero as $t \to +\infty$.

For the continuous-time analog of Theorem 1.1, we interpret the stochastic matrix $P(t) =]p_{ij}(t)]$ defined by (1.6) as the "infinitesimal matrix" of the chain specified by (3.4). Also, in (1.20) and (1.22), we replace the sums oer t with integrals, i.e., for continuous-time chains conditions (1.20) and (1.22) are replaced by

$$\int_1^{+\infty} [\phi(t) + \psi(t)]\, dt < +\infty \tag{3.14}$$

and

$$\int_1^{+\infty} [\phi(t) + \psi(t)]\, dt = +\infty \tag{3.15}$$

respectively. We write

$$P(t) = P + V(t) \tag{3.16a}$$

as in (1.18), and

$$A = -I + P \tag{3.16b}$$

If P is of the form (1.5) with $m = 1$ (i.e., it has a single ergodic aperiodic component and $n - r_1$ transient states), then $\lim_{k \to \infty} (P^k)_{ij} = \mu_j$, while if it is of the form (1.5) with $m = 2$ (i.e., it has two ergodic aperiodic component and $n - r_1 - r_2$ transient states) then

$$\lim_{k \to +\infty} P^k =$$

$$\left[\begin{array}{c:c:c}
\mu_1^{(1)} \ldots\ldots\ldots \mu_{r_1}^{(1)} & 0 \ldots\ldots\ldots 0 & 0 \ldots\ldots 0 \\
\ldots\ldots\ldots & \ldots\ldots\ldots & \ldots \\
\mu_1^{(1)} \ldots\ldots\ldots \mu_{r_1}^{(1)} & 0 \ldots\ldots\ldots 0 & 0 \ldots\ldots 0 \\
\hdashline
0 \ldots\ldots\ldots 0 & \mu_1^{(2)} \ldots\ldots\ldots \mu_{r_2}^{(2)} & 0 \ldots\ldots 0 \\
\ldots\ldots\ldots & \ldots\ldots\ldots & \ldots \\
0 \ldots\ldots\ldots 0 & \mu_1^{(2)} \ldots\ldots\ldots \mu_{r_2}^{(2)} & 0 \ldots\ldots 0 \\
\hdashline
\mu_1^{(1)} z_{r_1+e_2+1,1}, \ldots, \mu_{r_1}^{(1)} z_{r_1+r_2} + 1,1 & \mu_1^{(2)} z_{r_1+r_2+1,2}, \ldots, \mu_{r_2}^{(2)} z_{r_1+r_2+1,2} & 0 \ldots\ldots 0 \\
\ldots\ldots\ldots & \ldots\ldots\ldots & \ldots \\
\mu_1^{(1)} z_{n,1} \ldots\ldots\ldots \mu_{r_1}^{(1)} z_{n,1} & \mu_1^{(2)} z_{n,2} \ldots\ldots\ldots \mu_{r_2}^{(2)} z_{n,2} & 0 \ldots\ldots 0
\end{array}\right] \tag{3.17}$$

where $z_{j,\gamma}$, $\gamma = 1, 2$ $j = r_1 + r_2 + 1, \ldots, n$ are defined by (1.15). We will need some spectral properties of the matrix $A = -I + P$. $\lambda_1 = 0$ is an eigenvalue of A, and by Theorem 2.1 of Ref. 12 (Vol. II, p. 4), the algebraic multiplicity of $\lambda_1 = 0$ is equal to the number m (here $m = 1$ and $m = 2$) of the ergodic components of P. If the ergodic components of P are aperiodic (as we assume here), then all other eigenvalues $\lambda_{m+1}, \ldots, \lambda_n$ satisfy

$$\operatorname{Re} \lambda_i < 0, \qquad i = m + 1, \ldots, n \tag{3.18}$$

$$|\lambda_i + 1| < 1, \qquad i = m + 1, \ldots, n \tag{3.19}$$

Inequalities (3.18) and (3.19) are not true if the ergodic components of P are periodic (see Theorem 3.1 of Ref. 12, Vol. II). By Jordan's theorem there exists a nonsingular $n \times n$ real matrix Q such that

$$A = Q^{-1} J Q \tag{3.20}$$

with

$$J = \begin{pmatrix} J_0 & & & 0 \\ & J_1 & & \\ & & \ddots & \\ 0 & & & J_l \end{pmatrix} \tag{3.21}$$

where J_0 is an $m \times m$ zero matrix, J_1 is a diagonal matrix with diagonals the eigenvalues $\lambda_{m+1}, ..., \lambda_{m+m_1}$ whose algebraic and geometric multiplicities are equal, and

$$J_i = \begin{pmatrix} \lambda_{m+m_1+i} & 1 & 0 & 0 & 0 \\ 0 & \lambda_{m+m_1+i} & 1 & 0 & 0 \\ \cdots & \cdots & & \cdots & \cdots \\ 0 & \cdots & \cdots & \lambda_{m+m_1+i} & 1 \\ 0 & \cdots & \cdots & 0 & \lambda_{m+m_1+i} \end{pmatrix}, \quad i = 2, ..., l \tag{3.22}$$

are $m_i \times m_i$, $i = 2, ..., l$ matrices corresponding to eigenvalues whose geometric multiplicity is smaller than their algebraic multiplicity. Clearly $m + m_1 + m_2 + \cdots + m_l = n$. Setting

$$y(t) = x(t) Q^{-1} \tag{3.23}$$

and

$$\tilde{V}(t) = Q V(t) Q^{-1} \tag{3.24}$$

Eq. (3.8) becomes

$$\frac{dy}{dt} = y(t) J + y(t)\, \tilde{V}(t) \tag{3.25}$$

which is equivalent to

$$y(t) = y(0) e^{tJ} + \int_0^t y(s)\, \tilde{V}(s) e^{(t-s)J}\, ds \tag{3.26}$$

It follows from (3.21) that

$$e^{tJ} = \begin{pmatrix} e^{tJ_0} & & & 0 \\ & e^{tJ_1} & & \\ & & \ddots & \\ 0 & & & e^{rJ_l} \end{pmatrix} \tag{3.27}$$

where

$$e^{tJ_0} = \begin{pmatrix} 1 & & & 0 \\ & 1 & & \\ & & \ddots & \\ 0 & & & 1 \end{pmatrix} = m \times m \text{ unit matrix} \tag{3.28}$$

$$e^{tJ_1} = \begin{pmatrix} e^{t\lambda_{m+1}} & & 0 \\ & \ddots & \\ 0 & & e^{t\lambda_{m+m_1}} \end{pmatrix} \tag{3.29}$$

$$e^{tJ_i} = e^{t\lambda_{m+m_1+i}} \begin{bmatrix} 1 & 1 & \frac{t^2}{2!} & \cdots\cdots\cdots & \frac{t^{m_i-1}}{(m_i-1)!} \\ 0 & 1 & t & \cdots\cdots\cdots & \frac{t^{m_i-2}}{(m_i-2)!} \\ \cdots & \cdots & \cdots & & \\ 0 & 0 & 0 \cdots\cdots\cdots & 1 & t \\ 0 & 0 & 0 \cdots\cdots\cdots & 0 & 1 \end{bmatrix}, \qquad i = 2, \dots, l \tag{3.30}$$

Proof of Theorem 1.1 (Continuous-Time). (1) If P has a single ergodic aperiodic component ($m = 1$), then the similarity matrix Q has the form

$$Q = \begin{pmatrix} \mu^{(1)} \\ q_2 \\ \vdots \\ q_n \end{pmatrix} = \begin{pmatrix} \mu_1^{(1)}, \dots, \mu_{r_1}^{(1)} \quad 0, \dots, 0 \\ q_2, \cdots\cdots\cdots, q_{2n} \\ \cdots\cdots\cdots \\ q_{\eta_1} \cdots\cdots\cdots q_{nn} \end{pmatrix} \tag{3.31}$$

where the row vectors $q_i = (q_{i1}, \dots, q_{in})$, $i = 2, \dots, n$, satisfy

$$\begin{aligned} &q_2 A = \lambda_2 q_2 \\ &\cdots\cdots\cdots \\ &q_{1+m_1} A = \lambda_{1+m_1} q_{1+m_1} \\ &q_{m_1+2} A = \lambda_{m_1+2} q_{m_1+2} \\ &q_{m_1+3} A = q_{m_1+2} + \lambda_{m_1+2} q_{m_1+3} \\ &\cdots\cdots\cdots \\ &q_{m_1+m_2} A = q_{m_1+m_2-1} + \lambda_{m_1+2} q_{m_1+m_2} \\ &q_{m_1+m_2} A = \lambda_{m_1+3} q_{m_1+m_2+1} \\ &\cdots\cdots\cdots \\ &q_n A = q_{n-1} + \lambda_{m_1+l} q_n \end{aligned} \tag{3.32}$$

The inverse matrix Q^{-1} reads

$$Q^{-1}=(z^{(1)}, z^{(2)},..., z^{(n)})=\left[\begin{array}{cccc} 1 & 0 & 0 & \vdots\ 0 \\ \vdots & \vdots & \vdots & \vdots\ \vdots \\ 1 & 0 & 0 & \vdots\ 0 \\ & & & \vdots \\ 1 & z_{r_1+1,2} & z_{z_1+1,3} & \vdots\ z_{r_1+1,n} \\ \vdots & \vdots & \vdots & \vdots\ \vdots \\ 1 & z_{n,2} & z_{n,3} & \vdots\ z_{n,n} \end{array}\right]\begin{array}{l} \left.\begin{array}{l} \\ \\ \\ \end{array}\right\} r_1 \\ \\ \left.\begin{array}{l} \\ \\ \\ \end{array}\right\} n-r_1 \end{array} \tag{3.33}$$

This and (3.23) give

$$y_1(t)=x_1(t)+\cdots x_n(t)\equiv 1 \tag{3.34}$$

Also, from $x(t)=y(t)Q$ with Q given by (3.31), we have

$$x_i(t)=\mu_i^{(1)}y_1(t)+\sum_{j=2}^{n} y_j(t)q_{ji}, \qquad i=1,..., r_1 \tag{3.35a}$$

$$x_i(t)=\sum_{j=2}^{n} y_j(t)q_{ji}, \qquad i=r_1+1,..., n \tag{3.35b}$$

Using (3.18), (3.27), the uniform boundedness of $|y(t)|$, and the fact that $\tilde{V}(t)=QV(t)Q^{-1}\to 0$ as $t\to +\infty$, we easily deduce from (3.26) that

$$\lim_{t\to +\infty} y_i(t)=0, \qquad i=2,..., n \tag{3.36}$$

This together with (3.34) and (3.35) yields

$$\lim_{t\to +\infty} x(t)=\mu^{(1)}$$

This is equivalent [see (3.31)] to (1.17a) (continous time). The proof of (1.17b) is obtained from the compactness of $\pi(t)$ and the uniqueness of $\mu^{(1)}$ (Ref. 13, Theorem 6.2.1).

(2) If P has two ergodic components ($m=2$), then

$$e^{tJ_0}=\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

The similarity matrix Q reads

$$Q = \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \\ q_3 \\ \vdots \\ q_n \end{pmatrix} = \begin{pmatrix} \mu_1^{(1)},\dots,\mu_{r_1}^{(1)} & 0\dots\dots0 & 0\dots\dots0 \\ 0\dots\dots0 & \mu_1^{2)},\dots,\mu_{r_2}^{(2)} & 0\dots\dots0 \\ q_{31}\cdots & \cdots & \cdots q_{3n} \\ \cdots & \cdots & \cdots \\ q_{n1}\cdots & \cdots & \cdots q_{nn} \end{pmatrix} \tag{3.37}$$

where the row vectors $q_i = (q_{i1},\dots, q_{1n})$, $i = 3,\dots, n$ satisfy equations similar to those in (3.32). The inverse matrix Q^{-1} reads

$$Q^{-1} = (z^{(1)},\dots, z^{(n)}) =$$

$$\left[\begin{array}{cccc} 1 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & \\ 0 & 1 & 0 & \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & 0 & 0 \\ z_{r_1+r_2+1,1} & z_{r_1+r_2+1,2} & z_{r_1+r_2+1,3} & z_{r_1+r_2+1,n} \\ \vdots & \vdots & \vdots & \vdots \\ z_{n,1} & z_{n,2} & z_{n,3} & z_{n,n} \end{array}\right] \begin{array}{l} \\ r_1 \\ \\ \\ r_2 \\ \\ \\ n-r_1-r_2 \\ \\ \end{array} \tag{3.38}$$

where $z_{j\gamma}$, $\gamma = 1, 2$, $j = r_1 + r_2 + 1,\dots, n$ are defined by (1.15) and satisfy (1.16). Using (3.38) and (3.23) we find

$$y_1(t) = x_1(t) + \cdots + x_{r_1}(t) + x_{r_1+r_2+1}(t) z_{r_1+r_2+1,1} + \cdots + x_n(t) z_{n,1} \tag{3.39a}$$

$$y_2(t) = x_{r_1+1}(t) + \cdots + x_{r_1+r_2}(t) + x_{r_1+r_2+1}(t) z_{r_1+r_2+1,2} + \cdots + x_n(t) z_{n,2} \tag{3.39b}$$

Hence

$$y_1(t) + y_2(t) = x_1(t) + \cdots + x_n(t) \tag{3.40}$$

Also from (3.37), and $x(t) = y(t)Q$, we obtain

$$x_i(t) = \mu_i^{(1)} y_1(t) + \sum_{j=3}^{n} y_j(t) q_{ji}, \qquad i = 1,\dots, r_1 \tag{3.41a}$$

$$x_{r_1+i}(t) = \mu_i^{(2)} y_2(t) + \sum_{j=3}^{n} y_j(t) q_{j,r_1+i}, \qquad i = 1,\dots, r_2 \tag{3.41b}$$

$$x_i(t) = \sum_{j=3}^{n} y_j(t) q_{ji}, \qquad i = r_1 + r_2 + 1,\dots, n \tag{3.41c}$$

Next, Eq. (3.25) in terms of components reads

$$\frac{dy_1}{dt} = y_1(t)\,\tilde{V}_{11}(t) + y_2(t)\,\tilde{V}_{21}(t) + \sum_{j=3}^{n} y_j \tilde{V}_{j1}(t) \tag{3.42}$$

$$\frac{dy_2}{dt} = y_1 \tilde{V}_{12}(t) + y_2 \tilde{V}_{22}(t) + \sum_{j=3}^{n} y_j \tilde{V}_{j2}(t) \tag{3.43}$$

$$\frac{dy_3}{dt} = \lambda_3 y_3 + \sum_{j=1}^{n} y_j(t)\,\tilde{V}_{j3}(t) \tag{3.44a}$$

$$\frac{dy_i}{dt} = \lambda_3 y_i + y_{i-1} + \sum_{j=3}^{n} y_j(t)\,\tilde{V}_{ji}(t), \qquad i = 4,..., m_1 + 2 \tag{3.44b}$$

and similar equations for the y_is, $i = m_1 + 3,..., n$. Since Re $\lambda_j < 0$, $j = 3,..., n$, and $\tilde{V}(t) = QV(t)Q^{-1} \to 0$ as $t \to +\infty$, Eqs. (3.44), and the similar equations for the y_is, $i = m_1 + 3,..., n$, yield

$$\lim_{t \to +\infty} y_i(t) = 0, \qquad i = 3,..., n \tag{3.45}$$

independently of the initial data $y(0)$.

Now using (1.6) and (3.38), a straightforward algebra gives

$$\tilde{V}_{21}(t) = -\tilde{V}_{22}(t) = \phi(t) \tag{3.46a}$$

$$\tilde{V}_{12}(t) = -\tilde{V}_{11}(t) = \psi(t) \tag{3.46b}$$

and

$$\sum_{j=3}^{n} y_j(t)\,\tilde{V}_{j1}(t) = -\sum_{j=3}^{n} y_j(t)\,\tilde{V}_{j2}(t) \equiv R(t) \tag{3.47}$$

Thus Eqs. (3.42) and (3.43) become

$$\frac{dy_1}{dt} = -\psi(t)\,y_1(t) + \phi(t)\,y_2(t) + R(t) \tag{3.48a}$$

$$\frac{dy_2}{dt} = \psi(t)\,y_1(t) - \phi(t)\,y_2(t) - R(t) \tag{3.48b}$$

First we prove (1.21) and (1.23), assuming (3.14) and (3.15), respectively. Let

$$x(t) = x^{(j)}(t) - x^{(k)}(t)$$

where $x^{(i)}(t)$, $x^{(k)}(t)$ are defined by (3.5) $(t_0 = 0)$. Then $x_1(t) + \cdots + x_n(t) = 0$, and by (3.40)

$$y_2(t) = -y_1(t) \tag{3.49}$$

Thus (3.48) are equivalent to

$$\frac{dy_1}{dt} + [\phi(t) + \psi(t)]\, y_1 = R(t) \tag{3.50}$$

$$y_2(t) = -y_1(t)$$

Hence

$$\begin{aligned} y_1(t) = {} & y_1(0) \exp\left\{-\int_0^t [\phi(s) + \psi(s)]\, ds\right\} \\ & + \int_0^t ds\, R(s) \exp\left\{-\int_s^t [\phi(\xi) + \psi(\xi)]\, d\xi\right\} \end{aligned} \tag{3.51}$$

If (3.14) holds, then (3.51) implies that $\lim_{t \to +\infty} y_1(t)$, $\lim_{t \to +\infty} y_2(t)$ exist and are different from zero. This together with (3.45), (3.41), and $x(t) = x^{(i)}(t) - x^{(k)}(t)$ implies (1.21). On the other hand, if (3.15) holds, we will show that

$$\lim_{t \to +\infty} y_1(t) = -\lim_{t \to +\infty} y_2(t) = 0 \tag{3.52}$$

Clearly the first term on the right-hand side of (3.51) goes to zero as $t \to +\infty$. We now show that the second term also goes to zero. Set $t - s = \tau$. Then

$$\begin{aligned} & \left| \int_0^t ds\, R(s) \exp\left\{-\int_s^t [\phi(\xi) + \psi(\xi)]\, d\xi\right\} \right| \\ & \leqslant \int_0^t d\tau\, |R(t-\tau)| \exp\left\{-\int_{t-\tau}^t [\phi(\xi) + \psi(\xi)]\, d\xi\right\} \\ & \leqslant \int_0^{+\infty} d\tau\, |R(t-\tau)| \exp\left\{-\int_{t-\tau}^t [\phi(\xi) + \psi(\xi)]\, d\xi\right\} \end{aligned} \tag{3.53}$$

Since $\operatorname{Re} \lambda_j < 0$, $j = 3, \ldots, n$, Eqs. (3.44) yield that $|R(t)| \to 0$ as $t \to +\infty$. Therefore we can apply the dominated convergence theorem to the right-hand side of (3.53), and deduce that it goes to zero as $t \to +\infty$. This establishes (3.53), and yields (1.23).

Next we prove (1.24) assuming (3.15) and the existence of $\lim_{t\to+\infty}\pi(t)$. We set $x(t)=(p_{i1}^{(0,t)},\ldots,p_{in}^{(0,t)})$, $i=1,\ldots,n$. Thus $x_1(t)+\cdots+x_n(t)=1$, and by (3.40)

$$y_1(t)+y_2(t)\equiv 1 \tag{3.54}$$

Thus Eqs. (3.48) become

$$\begin{gathered}\frac{dy_1}{dt}+[\phi(t)+\psi(t)]\,y_1=\phi(t)+R(t)\\ y_2(t)=1-y_1(t)\end{gathered} \tag{3.55}$$

Hence

$$\begin{aligned}y_1(t)=y_1(0)\exp\left\{-\int_0^t[\phi(s)+\psi(s)]\,ds\right\}\\ +\int_0^t ds\,\phi(s)\exp\left\{-\int_s^t[\phi(\xi)+\psi(\xi)]\,d\xi\right\}\\ +\int_0^t ds\,R(s)\exp\left\{-\int_s^t[\phi(\xi)+\psi(\xi)]\,d\xi\right\}\end{aligned} \tag{3.56}$$

The existence of $\lim_{t\to+\infty}\pi(t)$ implies the existence of an $0\leqslant\eta\leqslant 1$ such that

$$\lim_{t\to+\infty}\pi(t)=\eta\pi^{(1)}+(1-\eta)\mu^{(2)} \tag{3.57}$$

We will show below that the existence of $\lim_{t\to+\infty}\pi(t)$ implies that $\lim_{t\to+\infty}[\phi(t)/\phi(t)+\psi(t)]$ exists, and in fact that

$$\lim_{t\to+\infty}\frac{\phi(t)}{\phi(t)+\psi(t)}=\eta \tag{3.58}$$

Assuming this for a moment, we complete the proof of (1.24). We will show

$$\lim_{t\to+\infty}y_1(t)=1-\lim_{t\to+\infty}y_2(t)=\eta \tag{3.59}$$

This together with (3.45) and (3.41) quickly yield (1.24). We now prove (3.59), The first term on the right-hand side of (3.56) clearly goes to zero as $t\to+\infty$. By the argument given for (3.53), so does the third term in (3.56). We will show that

$$\lim_{t\to+\infty}\int_0^t ds\,\phi(s)\exp\left\{-\int_s^t[\phi(\xi)+\psi(\xi)]\,\xi\right\}=\eta \tag{3.60}$$

Since $\phi(t) > 0$, we may change variables by setting

$$dz = \phi(\xi)\, d\xi$$

$$T = \int_0^t \phi(s)\, ds$$

Thus

$$I(t) \equiv \int_0^t ds\, \phi(s) \exp\left\{ - \int_s^t [\phi(\xi) + \psi(\xi)]\, d\xi \right\}$$

$$= \int_0^T d\tau \exp\left[- \int_\tau^T \frac{\phi(\xi(z)) + \psi(\xi(z))}{\phi(\xi(z))}\, dz \right] \tag{3.61a}$$

By (3.15), $T \to +\infty$ as $t \to +\infty$. First, we assume $\eta > 0$, and set

$$\frac{\phi(\xi(z)) + \psi(\xi(z))}{\phi(\xi(z))} = \frac{1}{\eta} + h(z)$$

$$h(z) \to 0 \qquad \text{as} \quad z \to +\infty$$

Thus

$$I(t) \int_0^T d\tau\, e^{-(T-\tau)(1/\eta)}\, e^{-\int_\tau^T h(z)\, dz}$$

$$= \int_0^T d\tau\, e^{-\tau(1/\eta)}\, e^{-\int_{t-\tau}^T h(z)\, dz}$$

$$= \eta - \int_T^{+\infty} d\tau\, e^{-(1/\eta)\tau} + \int_0^T d\tau\, e^{-(1/\eta)\tau} [e^{-\int_{t-\tau}^T h(z)\, dz} - 1] \tag{3.61b}$$

The second term on the right-hand side of (3.61b) goes to zero as $T \to +\infty$. The third term is bounded in absolute value by

$$\int_0^{+\infty} d\tau\, |e^{-\int_{T-\tau}^T h(z)\, dz} - 1|\, e^{-(1/\eta)\tau} \tag{3.62}$$

Since $h(z) \geqslant 1 - 1/\eta$, we see that the integrand in (3.62) is bounded uniformly in T by an intergrable function. Thus the dominated convergence theorem is applicable, and it yields that (3.62) goes to zero as $T \to +\infty$. Thus, by (3.61b), $I(t) \to \eta$ as $t \to +\infty$. If $\eta = 0$, then (3.61a) easily yields $I(t) \to 0$ as $t \to +\infty$. Thus we have established (3.59). Hence the proof of Theorem 1.1 will be completed once we establish (3.58).

Let

$$\sigma(t) = \pi(t) Q^{-1} \tag{3.63}$$

Since $\pi(t)[-I + P(t)] = 0$, we have

$$\sigma(t) J + \sigma(t)\, \tilde{V}(t) = 0 \tag{3.64}$$

where $\tilde{V}(t)$ is defined by (3.24). Using the special form of (3.38), we obtain from (3.63)

$$\sigma_1(t) = \pi_1(t) + \cdots + \pi_{r_1}(t) + \pi_{r_1+r_2+1}(t) z_{r_1+r_2+1,1} + \cdots + \pi_n(t) z_{n,1}$$
$$\sigma_2(t) = \pi_{r_1+1}(t) + \cdots + \pi_{r_1+r_2}(t) + \pi_{r_1+r_2+1}(t) z_{r_1+r_2+1,2} + \cdots + \pi_n(t) z_{n,2}$$
$$\sigma_i(t) = \pi_{r_1+r_2+1}(t) z_{r_1+r_2+1,i} + \cdots + \pi_n(t) z_{n,i}, \qquad i = 3, \ldots, n$$

Thus

$$\sigma_1(t) + \sigma_2(t) = 1 \tag{3.65}$$

and by (3.57)

$$\lim_{t \to +\infty} \sigma_1(t) = 1 - \lim_{t \to +\infty} \sigma_2(t) = \eta \tag{3.66}$$

and

$$\lim_{t \to +\infty} \sigma_i(t) = 0, \qquad i = 3, \ldots, n \tag{3.67}$$

Equation (3.64) in terms of coordinates reads

$$\sigma_1 \tilde{V}_{11} + \sigma_2 \tilde{V}_{21} + \sum_{j=3}^{n} \sigma_j \tilde{V}_{j1} = 0 \tag{3.68a}$$

$$\sigma_1 \tilde{V}_{12} + \sigma_2 \tilde{V}_{22} + \sum_{j=3}^{n} \sigma_j \tilde{V}_{j2} = 0 \tag{3.68b}$$

$$\sum_{j=3}^{n} \sigma_j (J + \tilde{V})_{ji} + \sigma_1 \tilde{V}_{1i} + \sigma_2 \tilde{V}_{2i} = 0, \qquad i = 3, \ldots, n \tag{3.68c}$$

Using (3.46) and (3.65), Eqs. (3.68a, b) become

$$\sigma_1(t) = \frac{\phi(t)}{\phi(t)\,\psi(t)} + \sum_{j=3}^{n} \sigma_j \frac{\tilde{V}_{j1}}{\phi(t) + \psi(t)} \tag{3.69}$$
$$\sigma_2(t) = 1 - \sigma_1(t)$$

We note that the $(n-2)\times(n-2)$ matrix

$$B=B(t)=(J_{ji}+\tilde{V}_{ji}), \qquad i, j=3,\dots, n$$

is nonsingular for sufficiently large t. Therefore (3.68c) gives

$$(\sigma_3,\dots,\sigma_n)=[\sigma_1(\tilde{V}_{23}-\tilde{V}_{13})-\tilde{V}_{23},\dots,\sigma_1(\tilde{V}_{2n}-\tilde{V}_{1n})-\tilde{V}_{2n}]\,B^{-1}(t) \tag{3.70}$$

This and (3.69) yield

$$\sigma_1(t)=\frac{\phi(t)}{\phi(t)+\psi(t)} + \frac{1}{\phi(t)+\psi(t)}[\sigma_1(\tilde{V}_{23}-\tilde{V}_{13})-\tilde{V}_{23},\dots,\sigma_1(\tilde{V}_{2n}-\tilde{V}_{1n})-\tilde{V}_{2n}]B^{-1}\begin{pmatrix}\tilde{V}_{31}\\ \vdots\\ \tilde{V}_{n1}\end{pmatrix} \tag{3.71}$$

Since $\tilde{V}(t)\to 0$ as $t\to +\infty$, the matrices $B(t)$ and $B^{-1}(t)$ are uniformly bounded as $t\to +\infty$. Using this, the special form (3.38) of Q^{-1}, the fact that $\tilde{V}(t)\to 0$ as $t\to +\infty$, and the explicit form of $\phi(t)+\psi(t)$, it is easily seen that the second term in (3.71) goes to zero as $t\to +\infty$, if (3.15) holds. Thus the first term on the right-hand side of (3.71) has a limit and

$$\lim_{t\to+\infty}\frac{\phi(t)}{\phi(t)+\psi(t)}=\lim_{t\to+\infty}\sigma_1(t)$$

This established (3.58), and completes the proof of the theorem. ∎

The following theorem is related to the rate of convergence of (1.33) and (1.35). It should be compared with Theorem 2.3. We state the result for continuous-time Markov chains, but it holds also for discrete-time Markov chains.

Theorem 3.1. Let $P(t)=[p_{ij}(t)]$ be the infinitesimal matrix of a continuous-time, nonstationary finite Markov chain converging to $P=(p_{ij})$ as $t\to +\infty$. Assuming that $P(t)$ has a unique invariant probability vector $\pi(t)$, then (1) if P has a single ergodic aperiodic component and a set of transient states, the rate of convergence of

$$\sum_{j=1}^{n}|p_{ij}^{(t_0,t)}-\pi_j|\to 0 \qquad \text{as} \quad t\to +\infty \tag{3.72}$$

for any fixed t_0, is the same as the rate of convergence of

$$\sum_{j=1}^{n}|\pi_j(t)-\pi_j|\to 0 \qquad \text{as} \quad t\to +\infty \tag{3.73}$$

(2) If P has the form (1.5) with $m=2$, and assume the conditions of Theorem 1.1 which ensure (1.24), then the rate of convergence of (3.72) depends on the rate of divergence of (3.15), and the rate of convergence of (3.73). In particular, if

$$\phi(t)+\psi(t)\geqslant\frac{1-\kappa}{t}, \qquad \text{for sufficiently large } t,\ 0<\kappa<1 \tag{3.74}$$

and

$$|\pi(t)-\pi|\leqslant\frac{a}{t^{1-\delta}}, \qquad \text{for sufficiently large } t,\ 0<\delta<1 \tag{3.75}$$

then

$$\sum_{j=1}^{n}|p_{ij}^{(t_0,t)}-\pi_j|\leqslant\frac{C}{t^{1-\varepsilon}}, \qquad \varepsilon=\min(\kappa,\delta),\ \text{large } t \tag{3.76}$$

Proof. (1) From (3.35), we see that the rate of convergence of $p_{ij}^{(t_0,t)}$ to π_j is the same as the rate at which $y_l(t)$, $l=2,\dots,n$ tend to zero as $t\to+\infty$ [note that $y_1(t)\equiv 1$]. Now from (3.25), $\omega(t)=[y_2(t),\dots,y_n(t)]$ satisfies

$$\frac{d\omega(t)}{dt}=\omega(t)\,B(t)+(\tilde{V}_{12},\tilde{V}_{13},\dots,\tilde{V}_{1n}) \tag{3.77}$$

where

$$[B(t)]_{ij}=J_{ij}+\tilde{V}_{ij}, \qquad i,j=2,3,\dots,n \tag{3.78}$$

and the diagonals of (J_{ij}), $i,j=2,\dots,n$ satisfy (3.18). The equation

$$\pi(t)[-I+P(t)]=0$$

in terms of $\sigma(t)=\pi(t)Q^{-1}$ reads

$$\sigma(t)J+\sigma(t)\,\tilde{V}(t)=0 \tag{3.79}$$

Using the fact that $J_0=0$, and $\sigma_1(t)\equiv 1$, Eq. (3.79) becomes

$$\rho(t)\,B(t)+(\tilde{V}_{12},\tilde{V}_{13},\dots,\tilde{V}_{1n})=0 \tag{3.80a}$$

$$\rho(t)=[\sigma_2(t),\dots,\sigma_n(t)] \tag{3.80b}$$

Comparing (3.77) and (3.80), we readily see that the rate of convergence of both (3.72) and (3.73) is the same as the rate of convergence to zero of

$$|\tilde{V}_{12}(t)|+\cdots+|\tilde{V}_{1n}(t)| \tag{3.81}$$

(2) From (3.41) we see that the rate of convergence to zero of (3.72) is the same as the rate of convergence of $y_1(t) \to \eta$, $y_2(t) \to 1-\eta$, and $y_l(t) \to 0$, $l=3,\ldots,n$. Also, from the equations below (3.64), we see that the rate of (3.73) is the same as the rate of convergence of $\sigma_1(t) \to \eta$, $\sigma_2(t) \to 1-\eta$, and $\sigma_l(t) \to 0$, $l=3,\ldots,n$. Now, Eqs. (368c) and (3.69), together with (3.75), yield

$$\frac{\phi(t)}{\phi(t)+\psi(t)} = \eta + 0\left(\frac{1}{t^{1-\delta}}\right), \qquad \text{sufficiently large } t \tag{3.82}$$

Using (3.56) [and the representation (3.61b)], a straightforward estimation implies that (3.74) and (3.82) give

$$|y_1(t) - \eta| = O\left(\frac{1}{t^{1-\varepsilon}}\right), \qquad \text{as} \quad t \to +\infty,\ \varepsilon = \min(\kappa, \delta) \tag{3.83}$$

Equations (3.44), etc., may be written in the compact form

$$\begin{aligned} &\omega(t) = [y_3(t),\ldots, y_n(t)] \\ &\frac{d\omega}{dt} = \omega(t)\, B(t) + (y_1 \tilde{V}_{13} + y_2 \tilde{V}_{23},\ldots, y_1 \tilde{V}_{1n} + y_2 \tilde{V}_{2n}) \end{aligned} \tag{3.84}$$

where $B(t)$ is defined above (3.70). From (3.84) and (3.68c) we see that the rate at which $|\omega(t)| \to 0$ is the same as the rate at which $|\sigma_3(t)| + \cdots + |\sigma_n(t)| \to 0$, i.e., the same as in (3.75). This completes the proof of the theorem. ▮

Remark. As we mentioned in the Introduction, we no not know whether there is a theorem similar to Theorem 1.1 when P has $m \geq 3$ ergodic components. In this case there are obvious generalizations of (3.37) and (3.38), which yield the following extension of (3.42) and (4.43):

$$\begin{aligned} &\frac{dy_1}{dt} = y_1 \tilde{V}_{11} + y_2 \tilde{V}_{21} + \cdots + y_m \tilde{V}_{m1} + \sum_{j=m+1}^{n} y_j \tilde{V}_{j1} \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &\frac{dy_m}{dt} = y_1 \tilde{V}_{1m} + \cdots + y_m \tilde{V}_{mm} + \sum_{j=m+1}^{n} y_j \tilde{V}_{jm} \end{aligned} \tag{3.85}$$

and equations for y_j, $j=m+1,\ldots,n$ similar to (3.44), etc. Also we have the generalization of (3.40)

$$y_1(t) + \cdots + y_m(t) = x_1(t) + \cdots + x_n(t) \tag{3.86}$$

The generalization of (3.15) here is

$$\int_1^{+\infty} \{\tilde{V}_{m1}(t)+\tilde{V}_{m2}(t)+\cdots+\tilde{V}_{m,m-1}(t)$$
$$-\tilde{V}_{11}(t)-\cdots-\tilde{V}_{m-1,m-1}(t)\}\,dt=+\infty \tag{3.87}$$

It can be shown that the integrand in (3.87) is positive. We do not know whether (3.87) suffices to control the asymptotic behavior of (3.85) and of the remaining equations for y_j, $j=m+1,..., n$.

The proof of Theorem 1.1 for discrete-time Markov chains is similar to the proof given above for continuous-time Markov chains. The basic differential equations (3.4) and (3.8) for continuous-time Markov chains are replaced by the following difference equations. Let us denote the one-step transition matrix $P^{(t-1,t)}$ by $P(t)$. Then by definition

$$P^{(t_0,t)}=P^{(t_0,t-1)}\,P(t), \qquad \text{for} \quad t_0<t$$

which implies the analog of (3.13a),

$$D(t)=D(t-1)\,d(t) \tag{3.88}$$

where

$$D(t)=\det P^{(t_0,t)}, \qquad d(t)=\det P(t)$$

Also introducing the vector $x(t)$ as in (3.5), the Chapman–Kolmogorov identity (1.8) leads to the study of the difference equation

$$x(r)=x(t-1)\,P(t) \tag{3.89}$$

subject to the initial conditions (3.9) or (3.10) or (3.11). Our study of equation (3.8) may be used to study the difference equation (3.89). The necessary changes are straightforward and we do not spell them out here. [Our analysis of the example treated in the Appendix is based on the difference equation (3.89).]

4. THE LAW OF LARGE NUMBERS AND THE VARIANCE OF A BIASED ESTIMATOR

In this section we prove Theorem 1.3.

Proof of Theorem 1.3. Let

$$\chi(X^{(t)}=s_i)=\begin{cases}1 & \text{if} \quad X^{(t)}=s_i \\ 0 & \text{if} \quad x^{(t)}\neq s_i\end{cases}$$

then

$$Y^{(t)} = \frac{1}{t} \sum_{s=1}^{n} f(X^{(s)}) = \frac{1}{t} \sum_i f_i \sum_{s=1}^{n} \chi(X^{(s)} = s_i)$$

Using this representation we compute

$$\begin{aligned} E_\sigma \left\{ Y_t - \sum_i f_i \pi_i \right\} &= \sum_i f_i \frac{1}{t} \sum_{s=1}^{t} E_\sigma \{ \chi(X^{(s)} = s_i) - \pi_i \} \\ &= \sum_i f_i \frac{1}{t} \sum_{s=1}^{t} \left(\sum_k \sigma_k p_{k_i}^{(0,s)} - \pi_i \right) \\ &= \sum_{i,k} f_i \sigma_k \frac{1}{t} \sum_{s=1}^{t} (p_{ki}^{(0,s)} - \pi_i) \end{aligned} \tag{4.1}$$

Since $p_{k_i}^{(0,s)} - \pi_i \to 0$, its Cesaro means also converges to zero, i.e.,

$$\frac{1}{t} \sum_{s=1}^{t} (p_{k_i}^{(0,s)} - \pi_i) \to 0 \qquad \text{as} \quad t \to +\infty \tag{4.2}$$

This yields (1.31). Also if the limit (1.33) exists, Eq. (4.1) easily gives (1.34).

New we prove (1.32)

$$\begin{aligned} &E_\sigma \left\{ \left(Y^{(t)} - \sum_i f_i \pi_i \right)^2 \right\} \\ &\quad = E_\sigma \left\{ \left(\sum_i f_i \frac{1}{t} \sum_{s=1}^{t} [\chi(X^{(s)=s}{}_i) - \pi_i] \right)^2 \right\} \\ &\quad = \frac{1}{t^2} \sum_{s,\tau=1}^{t} \sum_{i,j} f_i f_j E_\sigma \{ [\chi(X^{(s)} = s_i) - \pi_i][\chi(X^{(\tau)} = s_j) - \pi_j] \} \\ &\quad = \frac{1}{t^2} \sum_{i,j} f_i f_j \sum_{s,\tau=1}^{t} [E_\sigma \{ \chi(X^{(s)} = s_i)\, \chi(X^{(\tau)} = s_j) \} - \pi_j E_\sigma \{ \chi(X^{(s)} = s_i) \} \\ &\qquad - \pi_i E_\sigma \{ \chi(X^{(\tau)} = s_j) \} + \pi_i \pi_j] \end{aligned}$$

Note that

$$E_\sigma \{ \chi(X^{(s)} = s_i) \} = \sum_k \sigma_k p_{ki}^{(0,s)}$$

$$E_\sigma \{ \chi(X^{(s)} = s_i)\, \chi(X^{(\tau)} = s_j) \} = \begin{cases} \sum_k \sigma_k p_{k_i}^{(0,s)} p_{ij}^{(s,\tau)} & \text{if} \quad s < \tau \\ \sum_k \sigma_k p_{kj}^{(0,\tau)} p_{ji}^{(\tau,s)} & \text{if} \quad s > \tau \\ \sum_k \sigma_k p_{k_i}^{(0,s)} \delta_{ij} & \text{if} \quad s = \tau \end{cases}$$

Therefore

$$
\begin{aligned}
E_\sigma & \left\{\left(Y^{(t)}-\sum_i f_i \pi_i\right)^2\right\} \\
= & \frac{1}{t^2} \sum_{i,j} f_i f_j \left\{\sum_k \sigma_k \sum_{s<\tau}^{t} p_{k_i}^{(0,s)} p_{ij}^{(s,\tau)} + \sum_k \sigma_k \sum_{s>\tau} p_{k_j}^{(0,\tau)} p_{ji}^{(\tau,s)}\right. \\
& - t\pi_j \sum_k \sigma_k \sum_{s=1}^{t} p_{ki}^{(0,s)} - t\pi_i \sum_k \sigma_k \sum_{s=1}^{t} p_{ki}^{(0,s)} \\
& \left. + \sum_k \sigma_k \sum_{s=1}^{t} p_{ki}^{(0,s)} \delta_{ij} + t^2 \pi_i \pi_j \right\} \\
= & \sum_{i,j} f_i f_j \left\{ 2 \sum_k \sigma_k \frac{1}{t^2} \sum_{s<\tau}^{t} (p_{ki}^{(0,s)} - \pi_i) p_{ij}^{(s,\tau)} \right. \\
& - 2\pi_j \sum_k \sigma_k \frac{1}{t} \sum_{s=1}^{t} (p_{ki}^{(0,s)} - \pi_i) \\
& + \frac{1}{t} \sum_k \sigma_k \frac{1}{t} \sum_{s=1}^{t} (p_{ki}^{(0,s)} - \pi_i) \delta_{ij} \\
& \left. + 2 \frac{1}{t} \pi_j \frac{1}{t} \sum_{s>\tau}^{t} (p_{ij}^{(\tau,s)} - \pi_i) + \frac{1}{t} (\pi_i \delta_{ij} - \pi_i \pi_j) \right\} \\
= & \sum_{i,j} f_i f_j \left\{ 2 \sum_k \sigma_k \frac{1}{t} \sum_{s=1}^{t} \left[(p_{ki}^{(0,s)} - \pi_i) \frac{1}{t} \sum_{\tau=s+1}^{t} (p_{ij}^{(s,\tau)} - \pi_j) \right] \right. \\
& + 2\pi_j \sum_k \sigma_k \frac{1}{t} \sum_{s=1}^{t} \frac{t-s}{t} (p_{ki}^{(0,s)} - \pi_i) \\
& - 2\pi_j \sum_k \sigma_k \frac{1}{t} \sum_{s=1}^{t} (p_{ki}^{(0,s)} - \pi_i) \\
& + \sum_k \sigma_k \frac{1}{t} \left[\frac{1}{t} \sum_{s=1}^{t} (p_{ki}^{(0,s)} - \pi_i) \delta_{ij} \right] \\
& + 2\pi_i \frac{1}{t} \sum_{s=1}^{t} \left[\frac{1}{t} \sum_{\tau=s+1}^{t} (p_{ij}^{(s,\tau)} - \pi_j) \right] \\
& \left. + \frac{1}{t} (\pi_i \delta_{ij} - \pi_i \pi_j) \right\}
\end{aligned}
\tag{4.3}
$$

For fixed s, we have $p_{ij}^{(s,\tau)} - \pi_j \to 0$ as $\tau \to +\infty$. Therefore, for fixed s, we have

$$
\frac{1}{t} \sum_{\tau=s+1}^{t} (p_{ij}^{(s,\tau)} - \pi_i) \to 0 \qquad \text{as} \quad t \to +\infty \tag{4.4}
$$

This together with (4.2) imply that each term on the right-hand side of (4.3) goes to zero as $t \to +\infty$. This proves (1.32). Next, multiplying both sides of (4.3) by $t^{1-\varepsilon}$ we get

$$
\begin{aligned}
t^{1-\varepsilon}E_\sigma&\left\{\left(Y^{(t)}-\sum_i f_i\pi_i\right)^2\right\}\\
&=\sum_{i,j} f_i f_j\left\{2\sum_k \sigma_k \frac{1}{t^{1+\varepsilon}}\sum_{s=1}^{t}\sum_{\tau=s+1}^{t}(p_{ki}^{(0,s)}-\pi_i)(p_{ij}^{(s,\tau)}-\pi_j)\right.\\
&\quad+2\pi_j\sum_k \sigma_k\frac{1}{t^{\varepsilon}}\sum_{s=1}^{t}\frac{t-s}{t}(p_{ki}^{(0,s)}-\pi_i)\\
&\quad-2\pi_j\sum_k \sigma_k\frac{1}{t^{\varepsilon}}\sum_{s=1}^{t}(p_{ki}^{(0,s)}-\pi_i)\\
&\quad+\sum_k \sigma_k\frac{1}{t^{1+\varepsilon}}\sum_{s=1}^{t}(p_{ki}^{(0,s)}-\pi_i)\,\delta_{ij}\\
&\quad\left.-2\pi_i\frac{1}{t^{1+\varepsilon}}\sum_{s=1}^{t}\sum_{\tau=s+1}^{t}(p_{ij}^{(s,\tau)}-\pi_j)+\frac{1}{t^{\varepsilon}}(\pi_i\delta_{ij}-\pi_i\pi_j)\right\}
\end{aligned}
\tag{4.4}
$$

If the limit (1.33) exists, then the limit of the Cesaro means

$$\frac{1}{t^{\varepsilon}}\sum_{s=1}^{t}\frac{t-s}{t}(p_{ki}^{(0,s)}-\pi_i)$$

also exists as $t \to +\infty$, and is equal to w_{ki}. Therefore the second and third terms in the right-hand side of (4.4) cancel in the limit $t \to +\infty$. The fourth term goes to zero by (4.2). If the limit (1.35) exists, then the fifth and the first terms above also have limits as $t \to +\infty$. Thus in the limit $t \to +\infty$, we obtain (1.37) if $\varepsilon > 0$, and (1.36) if $\varepsilon = 0$.

Theorem 4.1. Let $p_{ij}^{(t-1,t)}$ be as in Theorem 1.3. Then we have the following:

(1) If P has a single ergodic component and possibly transient states, then if the series

$$\sum_{s=1}^{t}\sum_j |\pi_j(s)-\pi_j| \tag{4.5}$$

converges as $t \to +\infty$, then so does the series

$$\sum_{s=1}^{t} |p_{ij}^{(0,s)}-\pi_j| \tag{4.6}$$

Furthermore, if the series (4.5) diverges, then the series (4.6) diverges no faster than the series (4.5).

(2) Suppose that P has two or more ergodic components, and that $\pi_j(t)$ converges monotonically to π_j. Then (a) if the series

$$\sum_{s=1}^{t} \prod_{l=1}^{s} [1 - C(l)] \tag{4.7}$$

converges as $t \to +\infty$, then the series (4.6) diverges no faster than the series (4.5); (b) if

$$C(t) \geqslant \frac{1-\kappa}{t}, \qquad \text{large } t, \text{ some } 0 < \kappa < 1 \tag{4.8}$$

and

$$|\pi(t) - \pi| \leqslant \frac{a}{t^{1-\delta}}, \qquad \text{for large } t, \ 0 < \delta < 1 \tag{4.9}$$

then the series (4.6) diverges no faster than $O(t^{\varepsilon})$ with $\varepsilon = \min(\kappa, \delta)$.

Remark. If P has exactly two ergodic components, then (4.7) and (4.8) may be replaced by [compare with (1.22)]

$$\sum_{s=1}^{t} \exp\left\{-\sum_{\tau=1}^{s} [\phi(\tau) + \psi(\tau)]\right\} \tag{4.7'}$$

and

$$\phi(t) + \psi(t) \geqslant \frac{1-\kappa}{t} \tag{4.8'}$$

respectively.

Proof of Theorem 4.1. Part (1) of the theorem is a consequence of the Remark following the proof of Theorem 2.3, or part (1) of Theorem 3.1. Part (2) is a consequence of Theorem 2.3, or part (2) of Theorem 3.1.

5. CONVERGENCE OF THE ANNEALING ALGORITHM

In this section we prove Theorem 1.4, and establish a similar result for a class of nonstationary sampling methods which include the Metropolis method (1.40). We also consider a sampling mthod for multidimensional Random Markov Fields.

Proof of Theorem 1.4. (1) The validity of (1.29), (1.31), and (1.32) is a consequence of part (1) of Theorem 1.1. From (1.39), we see that

$$|\pi(t)-\pi| = e^{-\beta(t)(U_2-U_1)} + (\text{lower-order terms}), \qquad \text{as} \quad t \to +\infty \tag{5.1}$$

This and Theorem 4.1 imply the assertions in (i), (ii), (iii) of the theorem.

(2) If P has exactly two ergodic components, then part (2) of the present theorem is a consequence of part (2) of Theorem 1.1 and part (2) of Theorem 4.1. The constant C_0 is given by (1.56). If P has more than two ergodic components, then the validity of (1.29), (1.31), and (1.32) is a consequence of Theorem 1.2. Now we observe that $\pi_j(t)$ converges to π_j monotonically as $t \to +\infty$. Indeed, if $j \in \bar{S}^{(1)}$, then one easily sees that $\pi_j(t)$ is strictly increasing in $\beta(t)$ for all $\beta(t) > 0$, while if $j \notin \bar{S}^{(1)}$ then there exists a sufficiently large β_0 such that $\pi_j(t)$ is strictly decreasing in β for all $\beta \geqslant \beta_0$. This together with part (2) of Theorem 4.1 yields the rest of the theorem.

As we mentioned in the Introduction, Theorem 1.2 gives, in general, a worse constant than Theorem 1.1. We exhibit this is some explicit examples. First, consider the case with three states s_1, s_2, s_3 such that

$$U_1 < U_2 < U_3$$

and $q_{12} = q_{21} = 0$. The matrix $P^{(t-1,t)}$ reads

$$P^{(t-1,t)} = \begin{pmatrix} 1-q_{13}e^{-\beta(U_3-U_1)} & 0 & q_{13}e^{-\beta(U_3-U_1)} \\ 0 & 1-q_{23}e^{-\beta(U_3-U_2)} & q_{23}e^{-\beta(U_3-U_2)} \\ q_{31} & q_{32} & 1-q_{31}-q_{32} \end{pmatrix} \tag{5.2}$$

In order $\pi(t)$ to be a *unique* invariant probability vector we must have $q_{31} \neq 0$, $q_{32} \neq 0$. A straightforward computation gives

$$\tilde{E} = U_3 - U_1 \tag{5.3a}$$

$$E = U_3 - U_2 \tag{5.3b}$$

Thus $\tilde{E} > E$ and $\tilde{C}_0 < C_0$. Now we consider four states such that

$$U_1 < U_2 < U_3 < U_4 \tag{5.4a}$$

$$q_{31} = q_{32} = 0 \tag{5.4b}$$

$$q_{12} \neq 0 \tag{5.4c}$$

Since $q_{12} \neq 0$, the state s_2 [in $P(\infty)$] is transient. A straightforward computation shows that $\pi(t)$ is the only probability vector of $P^{(t-1,t)}$ if and only if $q_{34} \neq 0$. In this case [assuming (5.4)], the state s_3 is absorbing, and the state s_4 is transient. Reordering the states so that P has the form (1.5), the matrix $P^{(t-1,t)}$ reads

$$P^{(t-1,t)} = \begin{pmatrix} 1-q_{12}e^{-\beta(U_2-U_1)}-q_{24}e^{-\beta(U_4-U_1)} & 0 & q_{12}e^{-\beta(U_2-U_1)} & q_{14}e^{-\beta(U_4-U_1)} \\ 0 & 1-q_{34}e^{-\beta(U_4-U_3)} & 0 & q_{34}e^{-\beta(U_4-U_3)} \\ q_{21} & 0 & 1-q_{21}-q_{24}e^{-\beta(U_4-U_2)} & q_{24}\,e^{-\beta(U_4-U_2)} \\ q_{41} & q_{43} & q_{42} & 1-q_{41}-q_{42}-q_{43} \end{pmatrix} \tag{5.5}$$

A straightforward computation gives

$$\tilde{E} = U_4 - U_1$$

$$E = \min(U_2 - U_1, U_4 - U_3)$$

Thus we have again $\tilde{E} > E$ and $\tilde{C}_0 < C_0$.

Next we introduce a class of nonstationary sampling methods which contain the two most common methods used in statistical mechanics, i.e., the metropolis and the "heat-bath" sampling methods. The stationary versions of our sampling models appear in Ref. 10. Let $f(x)$ be a smooth function defined in the interval [0, 1] such that

$$0 \leqslant \frac{f(x)}{1+x} \leqslant 1, \qquad \text{for} \quad 0 \leqslant x \leqslant 1 \tag{5.6a}$$

$$\lim_{x \to 0+} f(x) = 1 \tag{5.6b}$$

Let $Q = (q_{ij})$ be the transition matrix of an arbitrary irreducible Markov chain. Here Q is not necessarily symmetric. We shall again refer to Q as the "proposal matrix." Let $\pi_j(t)$ be defined by (1.39). We define a generalization of (1.40) by

$$i \neq j, \; P_{ij}^{(t-1,t)} = q_{ij} \frac{f(\min\{(q_{ij}/q_{ji})[\pi_i(t)/\pi_j(t)], (q_{ji}/q_{ij})[\pi_j(t)/\pi_i(t)]\})}{1 + (q_{ij}/q_{ji})[\pi_i(t)/\pi_j(t)]} \tag{5.7a}$$

$$P_{ii}^{(t-1,t)} = 1 - \sum_{j \neq i} P_{ij}^{(t-1,t)} \tag{5.7b}$$

For $f(x) = 1 + x$, we obtain an extension of (1.40) with a nonsymmetric proposal matrix Q:

$$i \neq j, \quad P_{ij}^{(t-1,t)} = \begin{cases} q_{ij} & \text{if } \dfrac{q_{ji}}{q_{ij}} e^{-\beta(t)(U_j - U_i)} \geqslant 1 \\ q_{ji} e^{-\beta(t)(U_j - U_i)} & \text{if } \dfrac{q_{ji}}{q_{ij}} e^{-\beta(t)(U_j - U_i)} < 1 \end{cases} \tag{5.8a}$$

$$P_{ii}^{(t-1,t)} = 1 - \sum_{j \neq i} P_{ij}^{(t-1,t)} \tag{5.8b}$$

For $f(x) = 1$, we obtain

$$i \neq j, \qquad P_{ij}^{(t-1,t)} = q_{ij} \frac{q_{ji} \pi_j(t)}{q_{ij} \pi_i(t) + q_{ji} \pi_j(t)} \tag{5.9}$$

For a symmetric, $q_{ij} = q_{ji}$, proposal matrix, (5.9) is the heat-bath model

$$i \neq j, \qquad P_{ij}^{(t-1,t)} = q_{ij} \frac{\pi_j(t)}{\pi_i(t) + \pi_j(t)} \tag{5.10a}$$

$$P_{ii}^{(t-1,t)} = 1 - \sum_{j \neq i} P_{ij}^{(t-1,t)} \tag{5.10b}$$

Choosing $f(x) = 1 + 2(\frac{1}{2}x)^\gamma$, $\gamma \geqslant 1$, we obtain a sampling method which interpolates between (5.9) for $\gamma = +\infty$, and (5.8) for $\gamma = 1$.

Theorem 1.4 holds for the nonstationary Markov chain defined by (5.7). The proof of the theorem in this case is the same as before, and we will not spell out the details.

Next we consider briefly sampling methods for multidimensional Markov random fields. First we recall the definition[14] of a Markov random field (MRF) on a finite square lattice Z_m^d, with $M = m^d$ sites. A set

$$0 = \{0_a \subset Z_m^d : a \in Z_m^d\}$$

of subsets 0_a of Z_m^d is said to be a *neighborhood system* if: (a) $a \in 0_a$, (b) $a \in 0_b$ if and only if $b \in 0_a$. A subset $C \subseteq Z_m^d$ is a *clique* if every pair of distinct sites in C are neighbors. The set of all cliques will be denoted by $\mathscr{C}$. With each site $a \in Z_m^d$ we will associate a "spin" s_a with values in the spin state space

$$S = \{0, 1, ..., J-1\} \tag{5.11}$$

The set of all possible *configurations*

$$\Omega = \{s = (s_1, s_2, ..., s_M) : s_a \in S,\ a = 1, ..., M\} \tag{5.12}$$

will be referred to as the *state space*. A *potential* $V_C(s)$ associated with a clique $C \in \mathscr{C}$, is a function on the state space Ω, which depends only on those coordinates s_a of s for which $a \in C$. The *energy function* is a function on Ω defined by

$$U(s) = \sum_{C \in \mathscr{C}} V_C(s) \tag{5.13}$$

The probability measure on Ω

$$\pi^{(t)}(s) = \frac{e^{-\beta(t) U(s)}}{Z} \tag{5.14a}$$

$$Z = \sum_{s \in \Omega} e^{-\beta(t) U(s)} \tag{5.14b}$$

at temperature $T(t) = 1/\beta(t)$, is called the *Gibbs distribution* relative to the neighborhood system 0. Finally we shall use the conditional probabilities defined by

$$\pi^{(t)}(s_a \mid s_b, b \neq a) = \frac{\pi^{(t)}(s)}{\sum_{s_a \in S} \pi^{(t)}(s)}, \qquad s = (s_1, ..., s_M) \in \Omega \tag{5.15}$$

It is easily seen that the right-hand side of (5.15) depends on s_a and on s_b with $b \in 0_a$. The Gibbs distribution (5.14) relative to the neighborhood system 0 defines a Markov random field (MRF) X relative to 0, i.e., a family of random variables $X = \{X_a, a \in Z_m^d\}$, which satisfies (relative to some probability measure P on Ω)

$$P(X = s) > 0 \qquad \text{for all} \quad s \in \Omega \tag{5.16a}$$

$$P(X_a = s_a \mid X_b = s_b, b \neq a) = P(X_a = x_a \mid X_b = s_b, b \in 0_a) \tag{5.16b}$$

for every $a \in Z_m^d$, $s = (s_1, ..., s_M(\in \Omega$. Equation (5.16b) is the Markov property of the MRF. It is well known[14,8] that every MRF, X, relative to 0 [i.e., a set of random variables $X = \{X_a, a \in Z_m^d\}$ satisfying (5.16)] comes from a Gibbs distribution $\pi(s)$ relative to 0, and in fact

$$\pi(s) = P(X = s)$$

We note that the total number of states (configurations) is

$$n = |\Omega| = J^M, \qquad M = m^d \tag{5.17}$$

We will now define sampling methods via Markov chains $X^{(t)} = \{X_1^{(t)}, ..., X_M^{(t)}\}$ on the state (configuration) space Ω. Here each $X_a^{(t)}$ is a Markov chain associated with site $a \in Z_m^d$. There are at least three ways to define a Markov chain[10] associated with the MRF defined by the Gibbs distribution (5.14):

(1) $X^{(t-1)}=s^{(i)}=(s_1^{(i)},..., s_M^{(i)})$ and $X^{(t)}=s^{(j)}=(s_1^{(j)},..., s_M^{(i)})$ have all coordinates different.

(2) $X^{(t-1)}=s^{(i)}$ and $X^{(t)}=s^{(j)}$ differ only at a *single* site randomly chosen among the M sites of Z_m^d.

(3) $X^{(t-1)}=s^{(i)}$ and $X^{(t)}=s^{(j)}$ differ only at a *single* site, the site being selected from a fixed rather than random sequence.

Our basic convergence Theorem 1.4 can be adopted to handle any one of the above procedures. But we shall restrict ourselves in giving only the generalization of (1.40) and (5.10) to method (2). Let $q(a)$, $a \in Z_m^d$, be an arbitrary strictly positive probability measure defined in Z_m^d. Suppose that at time $t-1$, we have $X^{(t-1)}=s^{(i)}=(s_1^{(i)},..., s_M^{(i)})$. Choose a site $\gamma_t \in Z_m^d$ from the distribution $q(a)$, and change the spin at the site γ_t. Let $s^{(j)}$ be the new site. Then in analogy with (1.40) we define the transition probability

$$i \neq j, \quad P(X^{(t)}=s^{(j)} \mid X^{(t-1)}=s^{(i)})$$
$$=\begin{cases} q(\gamma_t) & \text{if} \quad U(s^{(j)}) \leqslant U(s^{(i)}) \\ q(\gamma_t)\exp[-\beta(t)(U(s^{(j)})-U(s^{(i)})), & \text{if} \quad U(s^{(j)}) > U(s^{(i)}) \end{cases} \tag{5.18}$$

and in analogy with (5.10)

$$P(X^{(t)}=s^{(j)} \mid X^{(t-1)}=s^{(i)})$$
$$=q(\gamma_t)\frac{\exp\{-\beta(t)[U(s^{(j)})-U(s^{(i)})]\}}{1+\exp\{-\beta(t)[U(s^{(j)})-U(s^{(i)})]\}} \tag{5.19}$$

Without spelling out the details, we note that Theorem 1.4 applies to the Markov chains defined by (5.16) and (5.17).

Next we note that if the spin state space S [see (5.12)] has only two states, i.e., $J=2$, then the right-hand side of (5.17) is equal to

$$q(\gamma_t)\,\pi^{(t)}(s_{\gamma_t}^{(j)} \mid s_a^{(j)},\ a \neq \gamma_t)$$

where the conditional probability is defined by (5.15). This leads to the following generalization of the "heat-bath" sampling method (5.17): Let $q(a)$ be as above, and suppose that $s^{(j)}=(s_1^{(j)},..., s_n^{(j)})$, $s^{(i)}=(s_1^{(i)},..., s_M^{(i)})$ differ only on one site γ_t chosen from the distribution $q(a)$. Then

$$P(X^{(t)}=s^{(j)} \mid X^{(t-1)}=s^{(i)})=q(\gamma_t)\,\pi^{(t)}(s_{\gamma_t}^{(j)} \mid s_a^{(j)}=s_s^{(i)},\ a \neq \gamma_t) \tag{5.10}$$

defines a Markov chain. With trivial modifications Theorem 1.4 applies to this Markov chain. The Markov chain (5.18) with a deterministic sequence $\{\gamma_t\}$ has been treated in the appendix of Ref. 8.

ACKNOWLEDGMENTS

This work was partially supported by NSF Grant No. MCS-8301864.

I warmly thank Alan Sokal for constant discussions concerning the problems treated in this paper. After this work was almost completed, he brought to my attention Ref. 19, which contains variations of part (1) of Theorem 1.2. I thank him for showing me this reference, as well as Ref. 3, which contains an interpretation of (1.26) in terms of the tail σ field.

APPENDIX

In this appendix we present an example with two absorbing states. The example illustrates various aspcts of nonstationary Markov chains with phase transitions. In particular, it clarifies the questions we posed in the Introduction [above (1.11)], and it illustrates the criticality of conditions (1.22) and (1.25).

We will denote the discrete time by $n = 0, 1, 2,\ldots$. The one-step transition matrix is

$$P^{(n-1,n)} = \begin{pmatrix} 1-g(n) & g(n) \\ f(n) & 1-f(n) \end{pmatrix} \tag{A1}$$

where

$$0 \leqslant f(n) < 1 \tag{A2}$$

$$0 \leqslant g(n) < 1 \tag{A3}$$

$$\lim_{n \to +\infty} f(n) = \lim_{n \to +\infty} g(n) = 0 \tag{A4}$$

The limiting matrix is

$$P = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and it defines a Markov chain with two absorbing states. Any equilibrium probability distribution of P is a convex combination of

$$\begin{aligned} \mu^{(1)} &= (1, 0) \\ \mu(2) &= (0, 1) \end{aligned} \tag{A5}$$

The unique invariant probability vector of $P^{(n-1,n)}$ is

$$\pi^{(n)} = \left[\frac{f(n)}{f(n)+g(n)}, \frac{g(n)}{f(n)+g(n)}\right] \tag{A6}$$

We readily see that no matter how fast $f(n)$ and $g(n)$ converge to zero, $\pi^{(n)}$ has no limit as $n \to +\infty$ unless $f(n)/g(n)$ [or $g(n)/f(n)$] has a limit.

Lemma A.1. For each $n_0 = 0, 1, 2, \ldots$, and $n > n_0$, we have

$$P^{(n_0,n)} = \begin{pmatrix} 1 - G(n_0, n) & G(n_0, n) \\ F(n_0, n) & 1 - F(n_0, n) \end{pmatrix} \tag{A7}$$

where $F(n_0, n) = P_{21}^{(n_0,n)}$, $G(n_0, n) = P_{12}^{(n_0,n)}$ are given by

$$F(n_0, n) = f(n) + [1 - f(n) - g(n)]\, F(n_0, n-1) \tag{A7a}$$

$$= f(n) + \sum_{k=n_0+1}^{n-1} f(k) \prod_{l=k+1}^{n} [1 - f(l) - g(l)] \tag{A7b}$$

$$G(n_0, n) = g(n) + [1 - f(n) - g(n)]\, G(n_0, n-1) \tag{A8a}$$

$$= g(n) + \sum_{k=n_0+1}^{n-1} g(k) \prod_{l=k+1}^{n} [1 - f(l) - g(l)] \tag{A8b}$$

Furthermore

$$p_{11}^{(n_0,n)} - p_{21}^{(n_0,n)} = p_{22}^{(n_0,n)} - p_{12}^{(n_0,n)} = 1 - F(n_0, n) - G(n_0, n) \tag{A9a}$$

$$1 - F(n_0, n) - G(n_0, n) = \prod_{k=n_0+1}^{n} [1 - f(k) - g(k)] \tag{A9b}$$

Proof.

$$\begin{aligned} p_{21}^{(n_0,n)} &= \sum_l p_{2l}^{(n_0,n-1)} p_{l1}^{(n-1,n)} \\ &= p_{21}^{(n_0,n-1)} p_{11}^{(n-1,n)} + p_{22}^{(n_0,n-1)} p_{21}^{(n-1,n)} \\ &= p_{21}^{(n_0,n-1)}[1 - g(n)] + (1 - p_{21}^{(n_0,n-1)})\, f(n) \end{aligned}$$

This is the same as (A7a). By iteration we obtain (A7b). The proof of (A8) is similar. Adding (A7) and (A8), we easily obtain (A9). ∎

Notation. In the remainder of this appendix we will write

$$F(n) = F(0, n) \quad \text{and} \quad G(n) = G(0, n)$$

From (A9) we get

Corollary A1. (a) If

$$\sum_{n=1}^{+\infty} [f(n)+g(n)] < +\infty \tag{A10}$$

then the following limits exists and are different from zero:

$$\lim_{n\to+\infty} (p_{11}^{(0,n)} - p_{21}^{(0,n)}) = \lim_{n\to+\infty} (p_{22}^{(0,n)} - p_{12}^{(0,n)})$$
$$= \prod_{k=1}^{+\infty} [1-f(k)-g(k)] \neq 0 \tag{A11}$$

(b) If

$$\sum_{n=1}^{+\infty} [f(n)+g(n)] = +\infty \tag{A12}$$

then the following limits exist and are zero:

$$\lim_{n\to+\infty} (p_{11}^{(0,n)} - p_{21}^{(0,n)}) = \lim_{n\to+\infty} (p_{22}^{(0,n)} - p_{12}^{(0,n)}) = 0 \tag{A13}$$

Remarks. (1) Corollary A1 has an obvious interpretation in terms of the Borel–Contelli Lemma (Ref. 7, p. 200).

(2) If (A10) holds, $\pi^{(n)}$ may or may not have a limit as $n \to +\infty$, but we shall see in Theorem A1 that $P^{(0,n)}$ always has a limit.

(3) If (A12) holds, then we will prove that if $\pi^{(n)}$ has a limit, then so does $P^{(0,n)}$ [see (A33) and (A34)]. But if $\pi^{(n)}$ has no limit, $P^{(0,n)}$ may have a limit [see example (A19)], or it may not have a limit [see example (A27)].

Theorem A.1. If (A10) holds, then $p_{ij}^{(0,n)}$, $i, j = 1, 2$, have limits as $n \to +\infty$, but

$$\begin{aligned} \lim_{n\to+\infty} p_{11}^{(0,n)} &\neq \lim_{n\to+\infty} p_{21}^{(0,n)} \\ \lim_{n\to+\infty} p_{22}^{0,n)} &\neq \lim_{n\to+\infty} p_{12}^{(0,n)} \end{aligned} \tag{A14}$$

Proof. From (A7)

$$F(n) = \prod_{k=1}^{n} [1-f(k)-g(k)] \left\{ \sum_{l=1}^{n} \frac{f(l)}{\prod_{m=1}^{l} [1-f(m)-g(m)]} \right\} \tag{A15}$$

By (A10)

$$\lim_{n\to+\infty}\prod_{k=1}^{n}[1-f(k)-g(k)]=C_0, \qquad C_0<+\infty, \quad C_0\neq 0 \quad \text{(A16)}$$

Without loss of generality we may assume that $f(n)+g(n)<1$. Then

$$\sum_{l=1}^{n}\frac{f(l)}{\prod_{m=1}^{l}[1-f(m)-g(m)]}\leqslant\frac{1}{C_0}\sum_{l=1}^{n}f(l) \qquad \text{(A17)}$$

By (A10), the right-hand side of (A17) converges as $n\to+\infty$. Hence, since the left-hand side of (A17) increases with n, it has a limit as $n\to+\infty$. This together with (A16) implies that $F(n)$ has a limit as $n\to+\infty$. The same way we prove that $G(n)$ converges as $n\to+\infty$. This establishes the existence of the limits of $p_{ij}^{(0,n)}$. (A14) is now a consequence of (A11) and (A16). ∎

Here is an example where $\pi^{(n)}$ has no limit as $n\to+\infty$, but $P^{(0,n)}$ has a limit and

$$\begin{aligned}\lim_{n\to+\infty}p_{11}^{(0,n)}&=\lim_{n\to+\infty}p_{21}^{(0,n)}\\ \lim_{n\to+\infty}p_{22}^{(0,n)}&=\lim_{n\to+\infty}p_{12}^{(0,n)}\end{aligned} \qquad \text{(A18)}$$

Take

$$f(n)=\frac{2}{n}, \qquad g(n)=\frac{1+(-1)^n}{n}, \qquad \text{for} \quad n\geqslant 5 \qquad \text{(A19)}$$

Then

$$\pi^{(2n)}=(\tfrac{1}{2},\tfrac{1}{2})$$
$$\pi^{(2n+1)}=(1,0)$$

Thus $\pi^{(n)}$ has no limit as $n\to+\infty$. We will show that

$$\lim_{n\to+\infty}P^{(0,n)}=\begin{pmatrix}\frac{2}{3} & \frac{1}{3}\\ \frac{2}{3} & \frac{1}{3}\end{pmatrix} \qquad \text{(A20)}$$

Noting that

$$f(2n)=g(2n)=\frac{1}{n}$$

$$g(2n+1)=0$$

we compute

$$g(2n+1) = g(2n)[1-f(2n+1)]$$
$$+ g(2n-2)[1-f(2n+1)][1-f(2n)-g(2n)][1-f(2n+1)]$$
$$+ \cdots$$
$$+ g(2)[1-f(3)][1-f(4)-g(4)][1-f(5)]\cdots$$
$$[1-f(2n-1)][1-f(2n)-g(2n)][1-f(2n+1)]$$

A straightforward algebra gives

$$G(2n+1) = \frac{1}{(n-1)\,n(2n+1)} \{2.(2.2+1) + 3.(2.3+1) + \cdots$$
$$+ (n-2)[2(n-2)+1] + (n-1)[2.(n-1)+1]\}$$

from which we easily obtain

$$\lim_{n\to+\infty} G(2n+1) = \frac{1}{3} \tag{A21}$$

Now, from (A8a)

$$G(2n) = G(2n-1) + g(2n) - [f(2n) + g(2n)]\, G(2n-1)$$

Therefore

$$\lim_{n\to+\infty} G(2n) = \frac{1}{3}$$

This and (A21) imply that $G(n)$ has a limit and

$$\lim_{n\to+\infty} G(n) = \frac{1}{3} \tag{A22}$$

From (A9b) we have

$$\lim_{n\to+\infty} [1 - F(n) - G(n)] = 0 \tag{A23}$$

because of (A12).

Thus

$$\lim_{n\to+\infty} F(n) = \frac{2}{3}$$

This establishes (A20).

Next we give another specific example where both $\pi^{(n)}$ and $P^{(0,n)}$ have no limits as $n \to +\infty$. The choice of functions $f(n)$, $g(n)$ was motivated by the following observation: For any fixed m we have from (A7a)

$$F(n) - F(n-m) = f(n) + \sum_{k=n-m+1}^{n-1} f(k) \prod_{l=k+1}^{n} [1 - f(l) - g(l)] + \left\{ -1 + \prod_{l=n-m+1}^{n} [1 - f(l) - g(l)] \right\} F(n-m) \qquad \text{(A24)}$$

Since $F(n-m) \leqslant 1, f(n), g(n) \to 0$ as $n \to \infty$, we see from (A24) that if $F(n)$ has a limit along a subsequence $\{n_j\}$, it has the same limit along the subsequence $\{n_j - m\}$. Thus in order to construct an example where $F(n)$ has two different limits along two subsequences $\{n_j\}$ and $\{n_j\}$, the distance $n_j - \tilde{n}_j$ must go to infinity with the subsequences.

Here is our example: Let $f(n)$ be such that

$$\sum_{n=1}^{+\infty} f(n) = +\infty \qquad \text{(A25)}$$

Set

$$N = N(n) = f(1) + f(2) + \cdots + f(n) \qquad \text{(A26)}$$

and denote by $n(N)$ the inverse function. We choose $h(n)$ as follows:

$$g(n(N)) = (2 + \mu \cos N)\, f(n(N)), \qquad 0 < \mu < 1 \qquad \text{(A27)}$$

Clearly $N \to +\infty$ as $n \to +\infty$ and vice versa. We will construct two subsequence $\{N_j\}$ and $\{\tilde{N}_j\}$ which give rise to two different limits for $F(n(N))$. From (A15)

$$F(n) = \exp\left\{ \sum_{l=1}^{n} \log[1 - f(l) - g(l)] \right\} \sum_{k=1}^{n} f(k) \times \exp\left\{ - \sum_{l=1}^{k} \log[1 - f(l) - g(l)] \right\} \qquad \text{(A28)}$$

We extend a function $h(n)$ defined on the positive integers, to a function $\tilde{h}(x)$ defined on the entire half-line $x \geqslant 0$, so that

$$\tilde{h}(x) = f(k) \qquad \text{for} \quad k - 1 < x \leqslant k$$

Then

$$\sum_{k=1}^{n} h(k) = \int_1^n \tilde{h}(x)\, dx + h(1)$$

Thus

$$N = \int_1^n \bar{f}(x)\,dx + f(1)$$

and from (A28)

$$F(n) = f(1)[1 - f(2) - g(2)] \cdots [1 - f(n) - g(n)] + \int_1^n dy\, \bar{f}(y) \exp\left\{ \int_y^n dx \log[1 - \bar{f}(x) - \tilde{g}(x)] \right\}$$

Since $\bar{f}(y) > 0$, we may change variables by setting

$$d\xi = \bar{f}(x)\,dx$$

and obtain

$$\begin{aligned} F(n) &= f(1)[1 - f(2) - g(2)] \cdots [1 - f(n) - g(n)] \\ &\quad + \int_0^N dz \exp\left\{ \int_z^N d\xi \frac{1}{\bar{f}(x(\xi))} \log[1 - \bar{f}(x(\xi)) - \tilde{g}(x(\xi))] \right\} \\ &= f(1)[1 - f(2) - g(2)] \cdots [1 - f(n) - g(n)] \\ &\quad + \int_0^N dz \exp\left\{ \int_z^N d\xi \frac{1}{\bar{f}(x(\xi))} \log[1 - \bar{f}(x(\xi))(3 + \mu \cos \xi)] \right\} \end{aligned} \tag{A29}$$

The leading term as $N \to +\infty$, in the integral is

$$\begin{aligned} \int_0^N dz \exp\left[- \int_z^N d\xi (3 + \mu \cos \xi) \right] &= \int_0^N dz \exp[-3(N - z) - \mu(\sin N - \sin z)] \\ &= \int_0^N dt \exp(-3t) \exp\{-\mu[\sin N - \sin(N - t)]\} \end{aligned} \tag{A30}$$

We consider the limit of (A30) along $N = 2k\pi$ and $N = (2k+1)\pi$. For $N = 2k\pi$, we have (recall that $0 < \mu < 1$)

$$\int_0^{2\pi k} dt\, e^{-3t} e^{-\mu \sin t} \underset{k \to +\infty}{\to} \int_0^{\infty} dt\, e^{-3t} e^{-\mu \sin t} \tag{A31}$$

and for $N = (2k+1)\pi$

$$\int_0^{(2k+1)\pi} dt\, e^{-3t} e^{\mu \sin t} \underset{k \to +\infty}{\to} \int_0^{\infty} dt\, e^{-3t} e^{+\mu \sin t} \tag{A32}$$

It is easily seen that the limits (A31) and (A32) are different. The first terms on the right-hand side of (A29) goes to zero as $n \to +\infty$, and it can be verified that the error term we neglected in the integral of (A29) has no contribution as $N \to +\infty$. Thus $F(n)$ has two limits, (A31) and (A32), as $n \to +\infty$ along the subsequences $n(2\pi k)$, $n((2k+1)\pi)$, $k \to +\infty$, respectively.

By Theorem 1, if (A10) holds, then $P^{(0,n)}$ always has a limit which satisfies (A14). Therefore, in this case, if $\pi^{(n)}$ has a limit, it cannot satisfy (1.11). In contrast, we prove now, that if (A12) holds and $\pi^{(n)}$ has a limit as $n \to +\infty$, then $P^{(0,n)}$ also has a limit, and if

$$\lim_{n \to +\infty} \pi^{(n)} = \pi = (\pi_1, \pi_2) \tag{A33}$$

then

$$\lim_{n \to +\infty} P^{(0,n)} = \begin{pmatrix} \pi_1 & \pi_2 \\ \pi_1 & \pi_2 \end{pmatrix} \tag{A34}$$

i.e., (1.11) holds. To show this, we set

$$F(n) = \pi_1 + \Phi(n)$$

$$\frac{f(n) + g(n)}{f(n)} = \frac{1}{\pi_1} + \psi(n)$$

$$\psi(n) \to 0 \qquad \text{as} \quad n \to +\infty$$

From (A7a) we derive

$$\Phi(n) = -\pi_1 f(n)\,\psi(n) + \left[1 - \frac{f(n)}{\pi_1} - f(n)\,\psi(n)\right] \Phi(n-1) \tag{A35}$$

Iterating (A35), and using a representation similar to (A29), we have

$$\begin{aligned}
-\frac{1}{\pi_1}\,&\Phi(n) \\
&= \prod_{k=1}^{n} \left[1 - \frac{f(k)}{\pi_1} - f(k)\,\psi(k)\right] \left(\sum_{l=1}^{n} \frac{f(l)\,\psi(l)}{\prod_{m=1}^{l} \{1[f(m)/\pi_1] - f(m) - \psi(m)\}} \right) \\
&= f(1)\,\psi(1) \left[1 - \frac{f(2)}{\pi_1} - f(2)\,\psi(2)\right] \cdots \left[1 - \frac{f(n)}{\pi_1} - f(n)\,\psi(n)\right] \\
&\quad + \int_1^n dy\, \tilde{f}(y)\,\tilde{\psi}(y) \exp\left\{ \int_y^n dx \log\left[1 - \frac{\tilde{f}(x)}{\pi_1} - \tilde{f}(x)\,\tilde{\psi}(x)\right]\right\} \\
&= f(1)\,\psi(1) \prod_{k=2}^{n} \left[1 - \frac{f(k)}{\pi_1} - f(k)\,\psi(k)\right] + \int_0^N dz\, \tilde{\psi}(y(z)) \\
&\quad \times \exp\left(\int_z^N d\xi\, \frac{1}{\tilde{f}(x(\xi))} \log\left\{1 - \tilde{f}(x(\xi))\left[\frac{1}{\pi_1} + \tilde{\psi}(x(\xi))\right]\right\}\right)
\end{aligned} \tag{A36}$$

The first term goes to zero as $n \to +\infty$, by (A12). The integral term may be written as

$$I_N = \int_0^N dz\, \tilde{\psi}(y(z)) \exp\left[-\frac{1}{\pi_1}(N-z)\right] \tilde{\psi}(y(z)) \times \exp\left[-\int_z^N d\xi \{\tilde{\psi}(x(\xi)) + \phi(\xi)\}\right] \tag{A37}$$

where

$$\phi(\xi) = -\frac{1}{f(x(\xi))}\left(\bar{f}(x(\xi))\left[\frac{1}{\pi_1} + \tilde{\psi}(x(\xi))\right] + \log\left\{1 - \bar{f}(x(\xi))\left[\frac{1}{\pi_1} + \tilde{\psi}(x(\xi))\right]\right\}\right)$$

setting $\tilde{\psi}(y(z)) = \bar{\psi}(z)$, and $N - z = t$ we obtain

$$I_N = \int_0^N dt \exp\left(-\frac{1}{\pi_1}t\right) \bar{\psi}(N-t) \exp\left\{-\int_{N-t}^N d\xi\, [\bar{\psi}(\xi) + \phi(\xi)]\right\}$$

and

$$|I_N| \leqslant \int_0^N dt \exp\left(-\frac{1}{\pi_1}t\right) |\bar{\psi}(N-t)| \exp\left\{-\int_{N-t}^N d\xi\, [\bar{\psi}(\xi) + \phi(\xi)]\right\}$$
$$\leqslant \int_0^\infty dt \exp\left(-\frac{1}{\pi_1}t\right) |\bar{\psi}(N-t)| \exp\left\{-\int_{N-t}^N d\xi\, [\bar{\psi}(\xi) + \phi(\xi)]\right\} \tag{A38}$$

Since $\bar{\psi}(x)$, $\phi(x) \to 0$ as $x \to +\infty$, and $\bar{\psi}(x) \geqslant 1 - 1/\pi$, it is easily seen that the dominated convergence theorem is applicable in (A38). Since the integrand in (A38) goes to zero as $N \to +\infty$, we obtain

$$\lim_{N \to +\infty} I_N = 0$$

By (A36), this implies that $\Phi(n) \to 0$ as $n \to +\infty$. Hence $F(n) \to \pi_1$ as $n \to +\infty$. This together with (A9) given that $G(n) \to \pi_2$ as $n \to +\infty$. This completes the proof of (A34).

REFERENCES

1. K. Binder and D. Stauffer, Monte-Carlo studies of random systems: Spin glasses, percolation, etc., preprint.
2. E. Coddington and N. Levinson, *Theory of Ordinary Differential Equations* (McGraw-Hill, New York, 1955).

3. H. Cohn, On a paper by Doeblin on non-homogeneous Markov chains, *Adv. Appl. Prob.* **13**:388–401 (1981).
4. M. Creutz, L. Jacobs, and C. Rebbi, Experiments with a gauge-invariant Ising model, *Phys. Rev. Lett.* **42**:1390–1393 (1979).
5. M. Creutz, L. Jacobs, and C. Rebbi, Monte Carlo study of Abelian lattice gauge theories, *Phys. Rev. D* **20**:1915–1922 (1979).
6. R. L. Dobrushin, Central limit theorem for non-stationary Markov chains, I, II, *Theo. Prob. Appl.* **1**:65–80 (1956); *Theor. Prob. Appl.* **1**:329–383 (1956).
7. W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. I (John Wiley and Sons, New York, 1968).
8. S. Geman and D. Geman, Stochastic relaxation, Gibbs distributions, and Bayesian restoration of images, *IEEE transactions PAMI* (1984).
9. D. Griffeath, Uniform coupling of non-homogeneous Markov chains, *J. Appl. Prob.* **12**:753–762 (1975).
10. W. K. Hastings, Monte Carlo sampling methods using Markov chains and their applications, *Biometrika* **57**:97–109 (1970).
11. D. L. Isaacson and R. W. Madsen, *Markov Chains Theory and Applications* (John Wiley and Sons, New York, 1976).
12. S. Karlin and H. M. Taylor, *A First Course in Stochastic Processes*, Vols. I and II (Academic Press, New York, 1975).
13. J. Kemeny and J. L. Snell, *Finite Markov Chains* (Van Nostrand, New York, 1960).
14. R. Kindermann and J. L. Snell, *Markov Random Fields and Their Applications, Contemporary Mathematics*, Vol. 1 (American Mathematical Society, Providence, Rhode Island, 1980).
15. S. Kirkpatric and D. Sherrington, Infinite-ranged models of spin-glasses, *Phys. Rev. B* **17**:4384-4403 (1978).
16. S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, Optimization by simulated annealing, *Science* **220**:621–680 (13 May 1983).
17. N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, Equations of state calculations by fast computing machines, *J. Chem. Phys.* **21**:1087–1091 (1953).
18. K. Binder, ed. *Monte Carlo Methods in Statistical Physics* (Springer, New York, 1979).
19. E. Seneta, *Non-Negative Matrices and Markov Chains* (Springer, New York, 1981).
20. G. Toulouse, Frustrations et desordres: problemes nouveaux en mecanique statistique. Histoire des verres de spin, *Proceedings of the 1981 Conference of French Physical Society* (Editions de Physique, Paris, 1982).