The Annals of Mathematical Statistics
1971, Vol. 42, No. 2, 451–465

WEAK CONVERGENCE RESULTS FOR A CLASS OF MULTIVARIATE MARKOV PROCESSES

BY SIEGFRIED SCHACH[1]

The Johns Hopkins University

1. Introduction. In this paper we obtain weak convergence results for several variations and generalizations of the classical Ehrenfest process. The basic model considered here can be described as follows: N balls are distributed among K urns, with $n_k(t)$ balls being in urn k at time t. Balls move among the urns according to the following rules: The probability that a ball shifts from urn k to urn $l \neq k$ during $(t, t+\Delta t)$ is $n_k \lambda_k p_{kl} \Delta t + o(\Delta t)$, $(1 \leqq k, l \leqq K)$, and the probability of more than one transition during $(t, t+\Delta t)$ is of order $o(\Delta t)$. We are interested in the limiting behavior of suitably normalized versions of the process $(n_1(t), \cdots, n_K(t))$, as the number of balls gets large. Our main result states that, suitably normalized, these processes converge to what we call multivariate Ornstein–Uhlenbeck processes in the sense of weak convergence of probability measures [2]. In particular, we arrive at a diffusion approximation for $(n_1(t), \cdots, n_K(t))$ as N gets large.

The technique of a "random change of time" enables us to obtain weak convergence results for a large class of discrete-time multivariate Ehrenfest models also. As a special case, we obtain a new proof of Iglehardt's [5] limit theorems.

A diffusion approximation to the two-urn model has been derived by Kac [6]. Karlin and McGregor [7] analyze several multivariate extensions of those results. Iglehardt's limit theorems are closest to ours, but they require that the time-parameter be discrete and that the probabilities p_{kl} depend on l only.

Several applications of multivariate Ehrenfest urn-type models are mentioned in [5], page 875. The model has also been found useful in describing the distribution of N vehicles over the K lanes of a long stretch of a unidirectional K-lane freeway [9]. In this application variables of interest are the usage of different lanes, the return to equilibrium after a bottleneck situation, etc.

Weak convergence results can be useful in two directions: For large N the distribution of a certain functional of the process can be approximated by the distribution of the same functional of the limiting process, provided that the latter is known. On the other hand, if an approximate distribution of a functional under the limiting process is desired, it might be possible to obtain the distribution of the same functional under the Nth approximating process, where N is large. In our case this procedure is particularly useful if simulation techniques are to be employed. Since the approximating processes have sample paths which are step

Received March 4, 1970.

[1] Research sponsored by the Air Force Office of Scientific Research, Office of Aerospace Research, United States Air Force, under AFOSR Contract No. AF 49(638)-1302 and by the United States Department of Transportation, Federal Highway Administration, Bureau of Public Roads under Contract FH-11-6884.

functions, they can be represented by finitely many "coordinates", whereas the limiting processes have continuous sample paths, for which this is not possible.

2. The basic model. Let $n(t) = (n_1(t), \cdots, n_K(t))$ be a multivariate stochastic process, where the $n_k(t)$ are nonnegative integer-valued and satisfy $\sum_k n_k(t) = N$. We assume that only transitions of the form

$$(n_1, \cdots, n_K) \to (n_1, \cdots, n_{k-1}, n_k - 1, n_{k+1}, \cdots, n_{l-1}, n_l + 1, n_{l+1}, \cdots, n_K) l \neq k$$

are possible, and that the probability of such a change during $(t, t+\Delta t)$ is of the form $n_k \lambda_k p_{kl} \Delta t + o(\Delta t)$ for some constants $\lambda_k > 0$, and some irreducible transition probability matrix $P = (p_{kl})$ (with $p_{kk} = 0$ for all k). Suppose also that the probability of two or more changes during $(t, t+\Delta t)$ is of order $o(\Delta t)$. Under these assumptions the random process $(n_1(t), \cdots, n_K(t))$ is a K-variate continuous-time Markov chain.

Let $p(n, t)$ be the probability that at time t the process is in state $n = (n_1, \cdots, n_K)$, given that it started from some fixed state $(m_1, \cdots, m_K)$ at time 0. Since we are dealing with a finite chain, the Kolmogorov forward equations determine the transition probabilities uniquely. These equations are:

$$\frac{\partial p(n, t)}{\partial t} = -p(n, t) \sum_{k=1}^{K} n_k \lambda_k + \sum_{k \neq l} \sum p(n_1, \cdots, n_k + 1, \cdots, n_l - 1, \cdots, n_K, t) \lambda_k (n_k + 1) p_{kl} \tag{2.1}$$

where it is understood that $p(n, t) = 0$ whenever $n_k = -1$ or $n_k = N+1$ for some k.

We can gain more insight into the behaviour of the process by examining the generating function

$$\xi(u, t) = \sum_n p(n, t) u_1^{n_1} \cdots u_K^{n} \tag{2.2}$$

where the summation extends over all possible states $n = (n_1, \cdots, n_K)$. It is easy to verify that the system of equations (2.1) yields the partial differential equation

$$\frac{\partial \xi(u, t)}{\partial t} = -\sum_k \lambda_k u_k \frac{\partial \xi(u, t)}{\partial u_k} + \sum_k \sum_l \lambda_k p_{kl} u_l \frac{\partial \xi(u, t)}{\partial u_k}. \tag{2.3}$$

Let P be the matrix with elements p_{kl}, and let Λ be the diagonal matrix with $\Lambda_{kl} = \delta_{kl} \lambda_k$. Set

$$Q = \Lambda(P' - I). \tag{2.4}$$

The standard method (e.g., [11], page 53) of solving the differential equation (2.3) consists in first solving the associated system of ordinary linear differential equations

$$\frac{dV(t)}{dt} = -QV(t), \tag{2.5}$$

where $V(t)$ is a K-dimensional vector function. The solutions of this system are of the form

(2.6) $$V(t) = e^{-Qt}V_0.$$

The general solution to (2.3) is then given by

(2.7) $$\xi(u, t) = \chi(e^{Qt}u),$$

where $\chi(\cdot)$ is a function to be determined by initial conditions.

Assume that at time 0 the process is in state $(m_1, m_2, \cdots, m_K)$. Hence

(2.8) $$\xi(u, 0) = \prod_{k=1}^{K} u_k^{m_k} = \chi(u).$$

Set

(2.9) $$(e^{Qt})_{kl} = p_{kl}(t).$$

Then we obtain

(2.10) $$\xi(u, t) = \prod_{k=1}^{K} (\sum_{l=1}^{K} p_{kl}(t)u_l)^{m_k}.$$

Specializing to the case $m_k = 1$, $m_l = 0$ for $l \neq k$, it follows that, for each $t > 0$

$$p_{kl}(t) \geqq 0, \sum_{l=1}^{K} p_{kl}(t) = 1 \quad \text{for each} \quad k.$$

Hence, by inspection of (2.10) we obtain

THEOREM 2.1. *Given that* $n(0) = (m_1, \cdots, m_K)$, *the distribution of* $n(t) = (n_1(t), \cdots, n_K(t))$ *is that of a sum of K independent multinomial random variables with parameters*

$$\{m_k; p_{k1}(t), \cdots, p_{kK}(t)\}, \qquad k = 1, 2, \cdots, K.$$

This result can be expressed in a somewhat different form, which will be useful later. Obviously Q is the matrix of transition intensities of a Markov process $Z'(t)$ taking values among the states $1, 2, \cdots, K$. Let $Z(t)$ be the process which is obtained from $Z'(t)$ by replacing state k by the kth unit vector $e_k = (0, 0, \cdots, 1, \cdots, 0)$, so that $Z(t)$ takes its values on the K unit vectors. Now let $\{Z_{kl}(t); k = 1, \cdots, K; l = 1, \cdots, m_k\}$ be N independent Markov processes of this type with $Z_{kl}(0) = e_k$. Set

(2.11) $$Z(t) = \sum_{k=1}^{K} \sum_{l=1}^{m_k} Z_{kl}(t).$$

Then we obtain

THEOREM 2.2. *The processes* $\{Z(t), t \geqq 0\}$ *and* $\{n(t), t \geqq 0\}$ *have the same distribution.*

PROOF. It is easy to see that $\{Z(t)\}$ is a Markov process. Since the Z_{kl} processes have transition probability matrix e^{Qt}, it follows that $P\{Z_{kl}(t) = e_l\} = p_{kl}(t)$. By independence $Z(t)$ has the same distribution as $n(t)$ for each fixed t. Now, since this is true for any initial state, $\{Z(t)\}$ and $\{n(t)\}$ have the same transition probabilities, and thus their laws coincide.

Since e^{Qt} is the transition probability matrix of a Markov process with a finite number of states, and only one communicating class, it follows that there exists a unique probability vector $p = (p_1, \cdots, p_K)'$ with

$$p' e^{Qt} = p'. \tag{2.12}$$

From Theorem 2.2 we conclude that the stationary version of our model, which we denote by $\bar{n}(t) = (\bar{n}_1(t), \cdots, \bar{n}_K(t))$, has a multinomial distribution with parameters $(N; p_1, \cdots, p_K)$. It will be useful to have formulas for the first and second moments:

$$E(n(t) \mid n(0) = m) = e^{Q't}m, \tag{2.13}$$

$$E\bar{n}(t) = Np, \tag{2.14}$$

$$\operatorname{Cov}[(n_l(s), n_{l'}(s+t)) \mid n(0) = m] = \textstyle\sum_k m_k p_{kl}(s)(p_{ll'}(t) - p_{kl'}(s+t)), \tag{2.15}$$

$$\operatorname{Cov}(\bar{n}_l(s), \bar{n}_{l'}(s+t)) = Np_l(p_{ll'}(t) - p_{l'}). \tag{2.16}$$

All these results follow directly from Theorem 2.2.

3. Multivariate Ornstein-Uhlenbeck processes. In this section we shall analyze a class of stochastic processes which includes all the limits ($N \to \infty$) of sequences of processes of the type considered in the previous section.

The stationary, univariate Ornstein–Uhlenbeck (O.U.) process can be characterized as the only stationary, continuous in probability, normal Markov process with zero mean (see e.g., [3], page 350). We will use these properties to define K-dimensional O.U. processes.

DEFINITION 3.1[2]. (a) *A K-variate stationary* O.U. *process is a K-variate Markov process* $\{Y(t)\}$ *which is stationary, Gaussian, continuous in probability, and satisfies* $EY(t) = 0$.

(b) A *K-variate* O.U. *process is a K-variate Markov process which has the same transition probabilities as a stationary* O.U. *process.*

The distribution of a Gaussian process with mean 0 is characterized uniquely by the covariance structure. In particular, K-variate stationary O.U. processes have a covariance structure which is very similar to that of the classical O.U. process:

THEOREM 3.2. *Let* $\{Y(t)\}$ *be a K-variate Gaussian process with* $EY(t) = 0$. *Then* $\{Y(t)\}$ *is a stationary* O.U. *process if and only if its covariance has the form*

$$EY(s)Y'(s+t) = Ce^{Bt}, \qquad t \geqq 0. \tag{3.1}$$

[2] This definition is more general than that given by Iglehardt [5].

PROOF. *Necessity*. Let $EY(0)Y'(0) = C$. Fix an $h > 0$ and let $E[Y'(h) \mid Y(0)] = Y'(0)M(h)$ for a suitable matrix $M(h)$. This is always possible, since in a normal family the conditional expectations are linear functions of the conditioning variables. Then

$$(3.2) \qquad EY(0)Y'(h) = E(E[Y(0)Y'(h) \mid Y(0)]) = EY(0)Y'(0)M(h) = CM(h).$$

Using again the smoothing properties of conditional expectations (e.g., [3], page 74), and stationarity we obtain

$$\begin{aligned} EY(0)Y'(2h) &= E\{Y(0)E[Y'(2h) \mid Y(h)]\} + E\{Y(0)(Y'(2h) - E[Y'(2h) \mid Y(h)])\} \\ &= EY(0)Y'(h)M(h) \\ &\quad + EE[Y(0)(Y'(2h) - E[Y'(2h) \mid Y(h)]) \mid Y(0), Y(h)] \\ &= CM(h)^2 + E\{Y(0)(E[Y'(2h) \mid Y(h)] - E[Y'(2h) \mid Y(h)])\} \\ &= CM(h)^2. \end{aligned}$$

By induction we obtain

$$(3.3) \qquad EY(0)Y'(nh) = CM(h)^n.$$

We thus have a matrix function $M(\cdot)$ which satisfies

$$(3.4) \qquad M(nh) = M(h)^n \quad \text{for all integers } n, \text{ all } h > 0.$$

Since for normal families continuity in probability and continuity in quadratic mean are equivalent, $M(h)$ is continuous in h.

Now let s, t be arbitrary positive numbers, then

$$\begin{aligned} (3.5) \qquad M(s+t) &= \lim_{n\to\infty} M([sn]/n + [tn]/n) = \lim_{n\to\infty} M(1/n)^{[sn]} M(1/n)^{[tn]} \\ &= \lim_{n\to\infty} M([sn]/n) \lim_{n\to\infty} M([tn]/n) = M(s)M(t), \end{aligned}$$

where $[u]$ = integral part of u.

Hence $M(\cdot)$ is a continuous semigroup of matrices. It therefore is of the form

$$(3.6) \qquad M(t) = e^{Bt}, \qquad t \geqq 0,$$

where B is the infinitesimal generator of the semigroup (see e.g., [4], page 614). Combining (3.6) and (3.2) we obtain (3.1).

Sufficiency. Stationarity and continuity in probability are obvious. We only have to prove the Markov property.

For the moment we assume that C is nonsingular. Thus it follows from [1], page 29 that, for $t \geqq 0$,

$$(3.7) \qquad E[Y'(s+t) \mid Y(s)] = Y'(s)C^{-1}C\,e^{Bt} = Y'(s)\,e^{Bt}.$$

Now let $t_1 < t_2 < \cdots < t_n < t_{n+1}$, and set $U = Y(t_{n+1}) - E[Y(t_{n+1}) \mid Y(t_n)] = Y(t_{n+1}) - \exp(B'(t_{n+1}-t_n))Y(t_n)$. Then, for any $1 \leqq k \leqq n$, we get

$$\begin{aligned}(3.8)\qquad EY(t_k)U' &= EY(t_k)Y'(t_{n+1}) - E(Y(t_k)E[Y'(t_{n+1}) \mid Y(t_n)]) \\ &= C\exp(B(t_{n+1}-t_k)) - C\exp(B(t_n-t_k))\exp(B(t_{n+1}-t_n)) = 0,\end{aligned}$$

so that U is independent of all the $Y(t_k)$, $k = 1, \cdots, n$. Since

$$(3.9)\qquad Y(t_{n+1}) = U + \exp(B'(t_{n+1}-t_n))Y(t_n),$$

it is easy to see that the Markov property is satisfied for $\{Y(t)\}$.

We have used the nonsingularity of C only to evaluate (3.7). But even if C is singular, the left-hand term of (3.7) is equal to the right-hand term, since $Y'(s)\,e^{Bt}$ is that particular linear combination which makes $Y'(s+t) - Y'(s)\,e^{Bt}$ independent of $Y'(s)$ (equation (3.8)). This completes the proof.

THEOREM 3.3. *A K-variate Markov process $\{Y(t),\ t \geqq 0\}$ is an* O.U. *process if and only if for every s, $t \geqq 0$ the conditional distribution of $Y(s+t)$, given $Y(s)$ is $N(\mu, \Sigma)$ with moments of the form*

$$(3.10)\qquad \mu = e^{B't}Y(s),$$

$$(3.11)\qquad \Sigma = C - e^{B't}C\,e^{Bt},$$

where C is a covariance matrix.

PROOF. *Necessity.* Let $\{Y(t)\}$ be an O.U. process. Then there exists a stationary O.U. process $\{\bar{Y}(t)\}$ with

$$(3.12)\qquad E\bar{Y}(t) = 0, \qquad E\bar{Y}(s)\bar{Y}'(s+t) = C\,e^{Bt}$$

which has the same transition probabilities as $\{Y(t)\}$. Since $\{\bar{Y}(t)\}$ is normal, the transition probabilities are normal. In the previous proof we have seen that (3.10) holds. Also in the previous proof set $n = 1$, $t_1 = s$, $t_2 = s+t$, then by (3.9) the conditional covariance of $\bar{Y}(s+t)$ given $\bar{Y}(s)$ is that of U, which is

$$\begin{aligned}EUU' &= E(\bar{Y}(s+t) - e^{B't}\,\bar{Y}(s))(\bar{Y}'(s+t) - \bar{Y}'(s)\,e^{Bt}) \\ &= C - e^{B't}C\,e^{Bt}.\end{aligned}$$

Sufficiency. Let $Y(0) \sim N(0, C)$. Then $Y(t)$ is also normal with mean 0. The covariance is

$$\begin{aligned}(3.13)\qquad EY(0)Y'(t) &= EE[Y(0)Y'(t) \mid Y(0)] \\ &= E(Y(0)E[Y'(t) \mid Y(0)]) = EY(0)Y'(0)\,e^{Bt} = C\,e^{Bt},\end{aligned}$$

$$\begin{aligned}(3.14)\qquad EY(t)Y'(t) &= E((Y(0) + Y(t) - Y(0))Y'(t)) \\ &= C\,e^{Bt} + E(E[(Y(t) - Y(0))Y'(t) \mid Y(0)]) \\ &= C\,e^{Bt} + E(C - e^{B't}C\,e^{Bt} + e^{B't}Y(0)Y'(0)\,e^{Bt}) - EY(0)Y'(0)\,e^{Bt} \\ &= C.\end{aligned}$$

Since the transition probabilities are stationary, we get from (3.13)

$$EY(s)Y'(s+t) = C\,e^{Bt}. \tag{3.15}$$

By Theorem 3.2 $\{Y(t)\}$ is a stationary O.U. process. $\square$

REMARKS. (a) Among the nonstationary O.U. processes only the ones with $Y(0) = \text{const.}$ are of practical interest.

(b) Standard techniques for obtaining sample path continuity (e.g., [2], Theorem 12.4) are applicable to each component of an O.U. process. Hence such processes are sample-path continuous a.s.

(c) Markov processes of the O.U. type can also be characterized in terms of their local mean $B'Y(t)$ and local variance–covariance $-(B'C+CB)$, just as in the classical univariate case. But we will not need these local properties in the sequel.

4. Weak convergence. In this section we will show that processes of the type considered in Section 2, if suitably normalized, converge to O.U. processes, as $N \to \infty$. In other words, the O.U. processes are diffusion approximations to those processes. Weak convergence of a sequence $\{X_N(t)\}$, $N = 1, 2, \cdots$ to $\{X(t)\}$ implies that $g(X_N(\cdot)) \to_{\mathscr{L}} g(X(\cdot))$ for any functional g which is continuous in a certain function space topology. We use as our basic function space the K-fold product of $D[0, 1]$ spaces with the J_1-topology of Skorokhod [10]. This topology is large enough to make functionals like $\sup_{0 \leq t \leq 1} X_{Nk}(t)$, $\int_0^1 X_N(t)\,dt$ continuous. For more details about the space $D[0, 1]$, see [2], Chapter 3. If, for any $0 < t_1 < t_2 < t_n$, we have

$$(X_N(t_1), \cdots, X_N(t_n)) \to_{\mathscr{L}} (X(t_1), \cdots, X(t_n)) \tag{4.1}$$

we say that the finite-dimensional distributions (f.d.d.) of $\{X_N(t)\}$ converge to the f.d.d. of $\{X(t)\}$. Weak convergence is stronger than convergence of the f.d.d. It requires, in addition, some compactness condition.

Let

$$X_N(t) = N^{-\frac{1}{2}}(n^{(N)}(t) - Np) \quad \text{and} \tag{4.2}$$

$$\bar{X}_N(t) = N^{-\frac{1}{2}}(\bar{n}^{(N)}(t) - Np), \tag{4.3}$$

where $\{n^{(N)}(t)\}$ is the original, nonstationary, process of Section 2, whereas $\{\bar{n}^{(N)}(t)\}$ is the corresponding stationary process. N is the total number of elements in the system. Assume that

$$n^{(N)}(0) = Np + N^{\frac{1}{2}}x_0, \tag{4.4}$$

where x_0 is an arbitrary, nonrandom vector with

$$\textstyle\sum_{k=1}^{K} x_{0k} = 0. \tag{4.5}$$

Let Π be a diagonal matrix with $\Pi_{kk} = p_k$.

Lemma 4.1. *As* $N \to \infty$ *the* f.d.d. *of* $\{X_N(t)\}$ *converge to those of a K-variate* O.U. *process* $\{X(t)\}$ *with*

$$EX(t) = e^{Q't}x_0, \qquad t \geqq 0, \tag{4.6}$$

$$EX(t)X'(t+s) - EX(t)EX'(t+s) = \Pi\, e^{Qs} - e^{Q't}\Pi\, e^{Q(s+t)} \qquad s, t \geqq 0. \tag{4.7}$$

PROOF. Using (2.12), (2.13) and (4.4) we get

$$EX_N(t) = N^{-\frac{1}{2}}(e^{Q't}(Np + N^{\frac{1}{2}}x_0) - Np) = e^{Q't}x_0.$$

Also, for the lth and l'th component of $\{X_N(t)\}$ we obtain from (2.15) and (4.4)

$$\begin{aligned} \operatorname{Cov}(X_{Nl}(t), X_{Nl'}(t+s)) &= N^{-1}\textstyle\sum_k (Np_k + N^{\frac{1}{2}}x_{0k})p_{kl}(t)(p_{ll'}(s) - p_{kl'}(t+s)) \\ &\to p_l p_{ll'}(s) - \textstyle\sum_k p_k p_{kl}(t)p_{kl'}(t+s), \qquad \text{as } N \to \infty. \end{aligned} \tag{4.8}$$

Stated in matrix form, we have

$$\lim_{N\to\infty}(EX_N(t)X_N'(t+s) - EX_N(t)EX_N'(t+s)) = \Pi\, e^{Qs} - e^{Q't}\Pi\, e^{Q(t+s)}. \tag{4.9}$$

Using the representation of $n(t)$ given in (2.11), a straightforward application of the Cramér–Wold technique yields asymptotic normality of $\{X_N(t)\}$ with the moments given in (4.6) and (4.7). The X_N-processes have stationary transition probabilities, hence the result follows from Theorem (3.3).

For stationary processes we get

LEMMA 4.2. *As* $N \to \infty$ *the* f.d.d. *of* $\{\overline{X}_N(t)\}$ *converge to those of a K-variate stationary* O.U. *process* $\{\overline{X}(t);\, t \geqq 0\}$ *with*

$$E\overline{X}(t)\overline{X}'(t+s) = (\Pi - pp')e^{Qs}, \qquad s \geqq 0. \tag{4.10}$$

PROOF. Similar to that of Lemma 4.1.

From now on we restrict our time index to the compact interval $0 \leqq t \leqq 1$. $\{X_{Nk}(t)\}$ is the kth coordinate of the Nth process, and C is a generic constant, independent of N, X_N, t, λ. Before we prove our basic convergence result, we prove a lemma.

LEMMA 4.3. *Let* $0 \leqq t_1 \leqq t_2 \leqq 1$, $\lambda > 0$. *There exists a constant* C *such that*

$$P\{|X_N(t_2) - X_N(t_1)| \geqq \lambda \mid X_N(t_1)\} \leqq \lambda^{-2}C(t_2 - t_1)(1 + |X_N(t_1)| + |X_N(t_1)|^2). \tag{4.11}$$

PROOF. Set $\sup_{k,l}\sup_{0 \leqq t \leqq 1}(Q'e^{Q't})_{k,l} = d < \infty$. Then by the mean value theorem,

$$|(\exp(Q't_2) - \exp(Q't_1))_{k,l}| \leqq d(t_2 - t_1).$$

A straightforward application of the Chebychev inequality to the individual coordinates $X_{Nk}(t)$ in conjunction with (2.13) and (2.15) thus yields the inequality.

REMARK. Since the $\overline{X}_N$-processes have the same transition probabilities as the X_N-processes, inequality (4.11) holds for the $\overline{X}_N$-processes.

THEOREM 4.4. *As* $N \to \infty$, *the sequence* $\{X_N(t);\ 0 \leqq t \leqq 1\}$, *converges weakly to the* O.U. *process* $\{X(t);\ 0 \leqq t \leqq 1\}$ *with mean and covariance given by* (4.6) *and* (4.7), *respectively.*

PROOF. We use Theorem 15.6, page 128, Billingsley [2]. Billingsley proves the result only for univariate processes, but it is easy to check that the argument goes through in the K-variate case. In fact, a major portion of this generalization can be found in [10]. We have already obtained the convergence of the f.d.d. to the proper limit in Lemma 4.1. Obviously no t is a fixed point of discontinuity for $\{X(t),\ t \geqq 0\}$. Therefore it suffices to show that

$$P\{|X_N(t)-X_N(t_1)| \geqq \lambda,\quad |X_N(t_2)-X_N(t)| \geqq \lambda\} \leqq \frac{C}{\lambda^{10/3}}(t_2-t_1)^{4/3} \tag{4.12}$$

for $0 \leqq t_1 \leqq t \leqq t_2 \leqq 1$ and all N.

Using well-known properties of Markov processes and Lemma (4.3) we obtain

$$\begin{aligned}
&P\{|X_N(t)-X_N(t_1)| \geqq \lambda,\ |X_N(t_2)-X_N(t)| \geqq \lambda\} \\
&\quad = E[P\{|X_N(t)-X_N(t_1)| \geqq \lambda,\ |X_N(t_2)-X_N(t)| \geqq \lambda \mid X_N(t)\}] \\
&\quad = E[P\{|X_N(t)-X_N(t_1)| \geqq \lambda \mid X_N(t)\} P\{|X_N(t_2)-X_N(t)| \geqq \lambda \mid X_N(t)\}] \\
&\quad \leqq \frac{C}{\lambda^2}(t_2-t) E[P\{|X_N(t)-X_N(t_1)| \geqq \lambda \mid X_N(t)\}(1+|X_N(t)|+|X_N(t)|^2)] \\
&\quad \leqq \frac{C}{\lambda^2}(t_2-t) E^{\frac{2}{3}}[P^{\frac{3}{2}}\{|X_N(t)-X_N(t_1)| \geqq \lambda \mid X_N(t)\}] E^{\frac{1}{3}}[1+|X_N(t)|+|X_N(t)|^2]^3,
\end{aligned}$$

by the Hölder inequality.

A straightforward calculation shows that $E[1+|X_N(t)|+|X_N(t)|^2]^3$ is bounded uniformly in $0 \leqq t \leqq 1$ and N. Hence, for a suitable constant C we get for the l.h.s. of (4.12)

$$\begin{aligned}
&\leqq \frac{C}{\lambda^2}(t_2-t) E^{\frac{2}{3}}[P\{|X_N(t)-X_N(t_1)| \geqq \lambda \mid X_N(t)\}] \\
&= \frac{C}{\lambda^2}(t_2-t) P^{\frac{2}{3}}\{|X_N(t)-X_N(t_1)| \geqq \lambda\} \\
&= \frac{C}{\lambda^2}(t_2-t) E^{\frac{2}{3}}[P\{|X_N(t)-X_N(t_1)| \geqq \lambda \mid X_N(t_1)\}] \\
&\leqq \frac{C}{\lambda^2 \lambda^{4/3}}(t_2-t)(t-t_1)^{\frac{2}{3}} E^{\frac{2}{3}}[1+|X_N(t_1)|+|X_N(t_1)|^2] \quad \text{(by Lemma 4.3)} \\
&\leqq \frac{C}{\lambda^{10/3}}(t_2-t)^{\frac{2}{3}}(t-t_1)^{\frac{2}{3}} \quad \text{(by above argument)} \\
&\leqq \frac{C}{\lambda^{10/3}}(t_2-t_1)^{4/3},
\end{aligned}$$

where, as always, C is a generic constant. This completes our proof.

We also obtain the corresponding result for the sequence of stationary processes $\bar{X}_N(t)$.

THEOREM 4.5. *As* $N \to \infty$, *the sequence* $\{\bar{X}_N(t);\ 0 \leqq t \leqq 1\}$ *converges weakly to the* O.U. *process* $\{\bar{X}(t); 0 \leqq t \leqq 1\}$ *with mean* 0 and *covariance given by* (4.10).

PROOF. Same as that of Theorem 4.4, except that Lemma 4.1 is replaced by Lemma 4.2.

5. Extension of weak convergence results to a class of discrete-time processes. In this section we use a random change of time argument to obtain weak convergence results for a class of urn models with a discrete time-parameter. They can be described as follows: N balls are distributed among K urns. If at time i the distribution of the N balls over the K urns is given by the vector $(n_1, \cdots, n_K)$, then the distribution at time $i+1$ is obtained by removing a single ball from the kth urn with probability n_k/N and placing it into the lth urn with probability $\bar{p}_{kl}$, where $\bar{P} = (\bar{p}_{kl})$ is a probability matrix. We assume that $\bar{p}_{kk} < 1$ for each k, and that $\bar{P}$ is irreducible. Take an arbitrary initial distribution and let $M_N(i)$ be the state of the chain at time i. Let $\{W(t), t \geqq 0\}$ be an independent Poisson process with intensity 1.
Set

$$n^{(N)}(t) = M_N(W(Nt)), \tag{5.1}$$

then it is easy to see that $n^{(N)}(t)$ is a continuous-time Markov process with total intensity out of state $(n_1, \cdots, n_K)$ given by

$$N\left(1-\frac{n_1}{N}\bar{p}_{11}-\cdots-\frac{n_K}{N}\bar{p}_{KK}\right) = \sum_k n_k(1-\bar{p}_{kk}),$$

and transition probabilities for a ball to move from the kth to the lth urn given by

$$\begin{aligned} p_{kl} &= \frac{\bar{p}_{kl}}{1-\bar{p}_{kk}} \qquad \text{if} \quad k \neq l, \\ &= 0 \qquad\qquad \text{if} \quad k = l. \end{aligned} \tag{5.2}$$

Hence, $n^{(N)}(t)$ is a process of the type considered in Section 2 with p_{kl} given by (5.2) and

$$\lambda_k = 1 - \bar{p}_{kk}. \tag{5.3}$$

Before we exploit the relationship (5.1) we prove a useful lemma.

LEMMA 5.1. *The stationary distribution of* $\{M_N(i);\ i = 0, 1, \cdots\}$ *is the same as that of* $\{n^{(N)}(t);\ t \geqq 0\}$.

PROOF. In Section 2 we obtained the result that the stationary distribution of $n^{(N)}(t)$ is multinomial with parameters N, p, where p is the left eigenvector of e^Q, corresponding to the eigenvalue 1, i.e.,

$$p' e^Q = p'. \tag{5.4}$$

By (2.4) $Q = \Lambda(P-I)$. Using (5.2) and (5.3) we obtain

$$Q = \bar{P} - I. \tag{5.5}$$

A vector p satisfies (5.4) if and only if

$$p'Q = 0, \tag{5.6}$$

by the spectral mapping theorem. But this is then equivalent to

$$p'\bar{P} = p'. \tag{5.7}$$

Now let

$$\xi_N(u, m) = u_1^{m_1} \cdots u_k^{m_k}. \tag{5.8}$$

Then the generating function $\psi_N(u, i)$ is given by

$$\psi_N(u, i) = E\xi_N(u, M_N(i)). \tag{5.9}$$

The assumptions on the probability structure of $M_N(i)$ imply that for $u_k \neq 0$, all k,

$$\begin{aligned} &E[\xi_N(u, M_N(i+1)) \mid M_N(i) = m] \\ &= \sum_{k,l} \xi_N(u, m) \frac{m_k}{N} \bar{p}_{kl} \frac{u_l}{u_k} \\ &= \frac{1}{N} \sum_{k,l} \frac{\partial \xi_N(u, m)}{\partial u_k} \bar{p}_{kl} u_l. \end{aligned} \tag{5.10}$$

If $M_N(i)$ has a multinomial distribution with parameters N, p, then we obtain

$$\begin{aligned} \psi_N(u, i+1) &= E[E(\xi_N(u, M_N(i+1)) \mid M_N(i) = m)] \\ &= \frac{1}{N} \sum_{k,l} \frac{\partial \psi_N(u, i)}{\partial u_k} \bar{p}_{kl} u_l, \end{aligned} \tag{5.11}$$

and also

$$\frac{\partial \psi_N(u, i)}{\partial u_k} = N(\textstyle\sum_l p_l u_l)^{N-1} p_k. \tag{5.12}$$

Combining (5.11), (5.12) and (5.7) yields

$$\psi_N(u, i+1) = (\textstyle\sum_l p_l u_l)^{N-1} (\sum_l \sum_k p_k \bar{p}_{kl} u_l) = \psi_N(u, i),$$

which ends our proof.

We now use the relationship (5.1) to obtain weak convergence results for the M_N-processes. If we include all the ω's, where ω is an element of the basic probability space, (5.1) has to be written as

$$n^{(N)}(t, \omega) = M_N(W(Nt, \omega), \omega). \tag{5.13}$$

The sample functions of the W-process are monotone increasing step functions, and therefore there exists an inverse process $\{V(t), t \geqq 0\}$ defined by

$$V(t, \omega) = \inf\{s : W(s, \omega) \geqq t\}. \tag{5.14}$$

Using this process we may write (5.13) as

$$n^{(N)}\left(\frac{1}{N} V(t, \omega), \omega\right) = M_N(t, \omega), \tag{5.15}$$

where now the sample paths of the process $\{M_N(t), t \geqq 0\}$ are step functions which are constant over intervals $(i-1, i]$, and the distribution of $\{M_N(t), t = 1, 2, 3, \cdots\}$ is the same as that of the discrete-time chain $\{M_N(i), i = 1, 2, \cdots\}$. Obviously the proper standardization of the $M_N(t)$-process is

$$X_N'(t) = N^{-\frac{1}{2}}(M_N(Nt) - Np) = N^{-\frac{1}{2}}\left(n^{(N)}\left(\frac{1}{N} V(Nt)\right) - Np\right), \quad 0 \leqq t \leqq 1, \tag{5.16}$$

where p is given by (5.6) or, equivalently, by (5.7). In the previous section (Theorem 4.4) it was shown that $X_N(t)$, given by (4.2), converges weakly to an O.U. process. Now $X_N'(t)$ differs from $X_N(t)$ only by the random change of time

$$\Phi_N : t \to \frac{1}{N} V(Nt). \tag{5.17}$$

By the definition of $V(t)$, $V(Nt) = S_{[Nt]+1}$, where $S_n = \sum_{i=1}^{n} U_i$ and the U_i are i.i.d. exponential with mean 1. Hence, by the strong law of large numbers the random perturbation caused by Φ_N should be negligible as $N \to \infty$. This argument is used in the proof of the following theorem. Assume that

$$M_N(0) = Np + N^{\frac{1}{2}} x_0. \tag{5.18}$$

THEOREM 5.2. *As* $N \to \infty$, *the sequence* $\{X_N'(t); 0 \leqq t \leqq 1\}$, *converges weakly to the* O.U. *process* $\{X(t); 0 \leqq t \leqq 1\}$ *with mean and covariance given by* (4.6) *and* (4.7), *respectively.*

PROOF. We follow the arguments given by Billingsby [2], pages 144–150. In Section 3 we obtained the result that the process $\{X(t)\}$ has continuous sample paths. The well-known invariance principle ([2], page 137) implies that

$$T_N = N^{\frac{1}{2}} \sup_{0 \leqq t \leqq 1} |\Phi_N(t) - t| \tag{5.19}$$

has a limiting distribution. Therefore

$$\sup_{0 \leqq t \leqq 1} |\Phi_N(t) - t| \to_P 0. \tag{5.20}$$

The function Φ: $t \to t$ lies in $C[0, 1]$. Set $X_N(t) = N^{-\frac{1}{2}}(n^{(N)}(t) - Np)$. Then by Theorem 4.4 above and by [2], page 27,

$$(X_N, \Phi_N) \to_D (X, \Phi), \tag{5.21}$$

where $\to_D$ stands for weak convergence.
This implies

$$X_N(\Phi_N(t)) = X_N'(t) \to_D X(t). \quad \square \tag{5.22}$$

If $\{\bar{M}_N(i)\}$ is stationary, then set

$$\bar{X}_N'(t) = N^{-\frac{1}{2}}(M_N[Nt] - Np).$$

Carrying out the same steps as in the proof of Theorem 5.2 one obtains

THEOREM 5.3. *As* $N \to \infty$, $\{\bar{X}_N'(t)\}$ *converges weakly to* $\{\bar{X}(t)\}$, *where* $\{\bar{X}(t)\}$ *is the* O.U. *process with mean* 0 *and covariance given by* (4.10).

6. Examples.

Traffic model. Suppose that N vehicles are travelling on a K-lane uni-directional freeway with n_k vehicles on lane k at time t, and that each vehicle in lane k has probability $\lambda_k \Delta t + o(\Delta t)$ of shifting lanes during $(t, t+\Delta t)$. Assume furthermore that a shifting vehicle from lane k moves into lane $k-1$ or $k+1$ with probability q_k, $1-q_k$ respectively, that the probability of more than one change by a vehicle during $(t, t+\Delta t)$ is of order $o(\Delta t)$, and that the vehicles act independently. Then $n(t) = (n_1(t), \cdots, n_K(t))$, the numbers of vehicles on the K lanes, is a stochastic process of the type considered in Section 2. The matrix Q has elements

$$\begin{aligned} q_{kl} &= \lambda_k q_k && \text{if } l = k-1, \\ &= -\lambda_k && \text{if } l = k, \\ &= \lambda_k(1-q_k) && \text{if } l = k+1, \\ &= 0 && \text{otherwise;} \end{aligned} \tag{6.1}$$

where $0 < q_k < 1$ for $1 < k < K$, $q_1 = 0$, $q_K = 1$. Such a Q has K different real characteristic roots μ_k, the largest one being 0 ([8], page 333). Let p be the left eigenvector of Q corresponding to $\mu_1 = 0$ and such that $\sum_k p_k = 1$, and Π be the diagonal matrix with elements $\Pi_{kk} = p_k$. Then there exists a matrix M with the properties

$$M^{-1}QM = D, \qquad M'\Pi = M^{-1}, \tag{6.2}$$

where D is diagonal (see [8], page 326). Let $n^{(N)}(0) = Np + N^{\frac{1}{2}}x_0$. Thus by

Theorem 4.4 the sequence of stochastic processes $X_N(t) = N^{-\frac{1}{2}}(n^{(N)}(t)-Np)$ converges weakly to an O.U. process $X(t)$ with

$$EX(t) = e^{Q't} \tag{6.3}$$

$$\begin{aligned} EX(t)X'(t) &= \Pi - e^{Q't}\Pi e^{Qt} = \Pi-(M^{-1})' e^{Dt} M' \Pi M e^{Dt} M^{-1} \\ &= \Pi - \Pi M e^{2Dt} M'\Pi. \end{aligned} \tag{6.4}$$

It follows from these formulas, that, after a disturbance of the system, $\exp(\mu_2 t)$ is the rate of return to equilibrium, where μ_2 is the largest characteristic root < 0.

Iglehardt's urn model. Iglehardt [5] obtained a weak convergence theorem for the particular discrete time urn model, in which each ball has the same probability of being drawn, and it is then placed into the kth urn with probability p_k, irrespective of what urn it came from. This model is then a special case of our discrete-time chain $\{M_N(i)\}$ with

$$\bar{p}_{kl} = p_l. \tag{6.5}$$

The vector p_0 of equilibrium probabilities is the solution of

$$p_0' \bar{P} = p_0', \tag{6.6}$$

and its solution is $p_0' = p'$, where $p' = (p_1, \cdots, p_K)$. Let $M_N(0) = Np + N^{\frac{1}{2}}x_0$, $\sum_k x_{0k} = 0$; thus, by Theorem 5.2 the sequence $N^{-\frac{1}{2}}(M_N([Nt])-Np)$, $0 \leqq t \leqq 1$, converges weakly to the O.U. process $\{X(t); 0 \leqq t \leqq 1\}$ with mean and covariance given by (4.6) and (4.7). Hence

$$EX(t) = \exp(Q't)x_0 = \exp((\bar{P}-I)t)x_0 = \exp(-t)x_0, \tag{6.7}$$

since $\bar{P}^k = \bar{P}$ for all positive integers k,

$$\exp(\bar{P}t) = (I-\bar{P}) + \exp(t)\bar{P}, \qquad \text{and} \qquad \bar{P}'x_0 = 0.$$

Also

$$\operatorname{Cov} X(t) = \Pi - \exp((\bar{P}'-I)t)\Pi \exp((\bar{P}-I)t) = (1-\exp(-2t))\Pi(I-\bar{P}), \tag{6.8}$$

which follows easily from the fact that $\bar{P}'\Pi\bar{P} = \bar{P}'pp' = pp' = \Pi\bar{P}$. ($\Pi$ is the diagonal matrix with elements $\Pi_{kk} = p_k$.) The moments (6.7) and (6.8) are those obtained by Iglehardt.

Theorem 5.3 gives us a similar result for the corresponding sequence of stationary chains. The limit process has moments

$$E\bar{X}(t) = 0, \qquad E\bar{X}(t)\bar{X}(t+s) = \exp(-s)(\Pi - pp'). \tag{6.9}$$

REFERENCES

[1] ANDERSON, T. W. (1962). *An Introduction to Multivariate Statistical Analysis*. Wiley, New York.

[2] BILLINGSLEY P. (1968). *Convergence of Probability Measures*. Wiley, New York.

[3] BREIMAN, L. (1968). *Probability*. Addison-Wesley, Reading, Mass.

[4] DUNFORD, N. and SCHWARTZ, J. T. (1964). *Linear Operators*, Part 1. Interscience, New York.

[5] IGLEHARDT, D. L. (1968). Limit theorems for the multi-urn Ehrenfest model. *Ann. Math. Statist.* **39** 864–876.
[6] KAC, M. (1947). Random walk and the theory of Brownian motion. *Amer. Math. Monthly* **54** 369–391.
[7] KARLIN, S. and MCGREGOR, J. L. (1965). Ehrenfest urn models. *J. Appl. Prob.* **2** 352–376.
[8] LEDERMANN, W. and REUTHER, G. E. H. (1953–54). Spectral theory for the differential equations of simple birth and death processes. *Philos. Trans. Royal Soc. London, Ser. A.* **246** 321–369.
[9] SCHACH, S. (1969). Markov models for multi-lane freeway traffic. Technical Report No. 126, Department of Statistics, The Johns Hopkins Univ.
[10] SKOROKHOD, A. V. (1956). Limit theorems for stochastic processes. *Theor. Probability Appl.* **1** 261–290.
[11] SNEDDON, I. N. (1957). *Elements of Partial Differential Equations.* McGraw-Hill, New York.