*The Annals of Probability*
1974, Vol. 2, No. 6, 1065–1074

# A CENTRAL LIMIT THEOREM FOR MARKOV PROCESSES THAT MOVE BY SMALL STEPS

BY M. FRANK NORMAN

*University of Pennsylvania*

We consider a family $X_n^\theta$ of discrete-time Markov processes indexed by a positive "step-size" parameter $\theta$. The conditional expectations of $\Delta X_n^\theta$, $(\Delta X_n^\theta)^2$, and $|\Delta X_n^\theta|^3$, given $X_n^\theta$, are of the order of magnitude of $\theta$, $\theta^2$, and $\theta^3$, respectively. Previous work has shown that there are functions $f$ and $g$ such that $(X_n^\theta - f(n\theta))/\theta^{\frac{1}{2}}$ is asymptotically normally distributed, with mean 0 and variance $g(t)$, as $\theta \to 0$ and $n\theta \to t < \infty$. The present paper extends this result to $t = \infty$. The theory is illustrated by an application to the Wright–Fisher model for changes in gene frequency.

**1. Introduction and overview.** Let $J$ be a bounded set of positive numbers with infimum 0. For every $\theta \in J$, let $\{X_n^\theta\}_{n\geqq 0}$ be a Markov process with stationary transition probabilities in a Borel subset $I_\theta$ of the real line $R$. The parameter $\theta$ is an index of the magnitude of $\Delta X_n^\theta = X_{n+1}^\theta - X_n^\theta$. We will be concerned with the asymptotic behavior of the distribution of $X_n^\theta$ as $n \to \infty$ and $\theta \to 0$.

The following assumptions, or their higher dimensional analogs, are in force throughout the paper:

$$E(\Delta X_n^\theta \mid X_n^\theta = x) = \theta w(x) + O(\theta^2) \tag{1.1}$$

$$\mathrm{Var}\,(\Delta X_n^\theta \mid X_n^\theta = x) = \theta^2 s(x) + o(\theta^2) \tag{1.2}$$

$$E(|\Delta X_n^\theta|^3 \mid X_n^\theta = x) = O(\theta^3)\,, \tag{1.3}$$

uniformly over $x \in I_\theta$. Thus the error terms in (1.1) and (1.2) satisfy

$$\sup_{\theta\in J, x\in I_\theta} |O(\theta^2)|/\theta^2 < \infty$$

and

$$\sup_{x\in I_\theta} |o(\theta^2)|/\theta^2 \to 0$$

as $\theta \to 0$. Let $I$ be the closed convex hull of $\bigcup_{\theta\in J} I_\theta$. We assume that $I_\theta$ approximates $I$ as $\theta \to 0$ in the sense that, for any $x \in I$,

$$\inf_{y\in I_\theta} |y - x| \to 0 \tag{1.4}$$

as $\theta \to 0$. The functions $w$ and $s$ are defined throughout $I$, $s$ is Lipschitz, and $w$ has a bounded Lipschitz derivative.

Under these assumptions the differential equations

$$f'(t) = w(f(t))$$

and

$$g'(t) = 2w'(f(t))g(t) + s(f(t))$$

Received June 25, 1973; revised November 10, 1973.
*AMS* 1970 *subject classifications*. Primary 60F05, 60J05; Secondary 92A10.
*Key words and phrases*. Central limit theorem, Markov process, step-size parameter.



Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Probability*. www.jstor.org

have unique solutions $f(t) = f(t, x)$ and $g(t) = g(t, x)$ with $f(0) = x$ and $g(0) = 0$, where $x$ is an arbitrary point of $I$. Suppose that $x_\theta \in I_\theta$ and $X_0^\theta = x_\theta$ a.s., and let

$$Z_n^\theta = (X_n^\theta - f(n\theta, x_\theta))/\theta^{\frac{1}{2}} .$$

Let $\mathscr{L}(Z)$ be the distribution of a random variable $Z$, and let $\mathscr{N}(\mu, \sigma^2)$ be the normal distribution with mean $\mu$ and variance $\sigma^2$. It has been established previously ([6] Theorem 8.1.1) that

$$\mathscr{L}(Z_n^\theta) \to \mathscr{N}(0, g(t, x)) \tag{1.5}$$

as $\theta \to 0$, $x_\theta \to x$, and $n\theta \to t < \infty$. Moreover, it can be shown that the distribution over $C[0, T]$ of the random polygonal line $Z^\theta$ with vertices $Z^\theta(n\theta) = Z_n^\theta$ converges weakly to the distribution of the diffusion $Z$ satisfying the stochastic differential equation

$$dZ(t) = w'(f(t))Z(t)\,dt + s(f(t))^{\frac{1}{2}}\,dB(t)$$

and the initial condition $Z(0) = 0$ a.s., where $B$ is Brownian motion. Weak convergence theorems of this type have been established in similar contexts by Rosén [9] and Kurtz [4].

These results are the background for the present study. We shall consider the limit of $\mathscr{L}(Z_n^\theta)$ as $\theta \to 0$ and $n\theta \to \infty$ under certain additional assumptions. Our main project is to prove the following theorem, which was announced in ([8] Theorem 3.2(ii)).

THEOREM 1. *Suppose that $I$ is bounded, $w$ has a unique zero $\lambda$, and $w'(\lambda) < 0$. Then*

$$\mathscr{L}(Z_n^\theta) \to \mathscr{N}(0, g(\infty)) \tag{1.6}$$

*as $\theta \to 0$ and $n\theta \to \infty$, where*

$$g(\infty) = \lim_{t\to\infty} g(t, x) = s(\lambda)/2|w'(\lambda)| .$$

The limiting process in Theorem 1 places no constraint on $x_\theta$. This implies that (1.6) holds uniformly over $x_\theta$ in the following sense. Let $d$ be a metric (or pseudometric) on probability distributions over $R$, such that $d(\mathscr{L}_n, \mathscr{L}) \to 0$ whenever $\mathscr{L}_n \to \mathscr{L}$ weakly. Then

$$\sup_{x_\theta \in I_\theta} d(\mathscr{L}(Z_n^\theta), \mathscr{N}(0, g(\infty))) \to 0$$

as $\theta \to 0$ and $n\theta \to \infty$. Proceeding further in this direction, we may combine (1.5) with Theorem 1 to obtain the rather striking conclusion that

$$\sup_{n\geq 0, x_\theta \in I_\theta} D(n, \theta, x_\theta) \to 0 \tag{1.7}$$

as $\theta \to 0$, where

$$D(n, \theta, x_\theta) = d(\mathscr{L}(Z_n^\theta), \mathscr{N}(0, g(n\theta, x_\theta))) .$$

For if (1.7) were not true, there would be a $c > 0$ and sequences $\theta_k$, $n_k$, and

$x_k \in I_{\theta_k}$ such that $\theta_k \to 0$ as $k \to \infty$, but

$$D(n_k, \theta_k, x_k) \geqq c \tag{1.8}$$

for all $k \geqq 1$. We could, moreover, choose these sequences in such a way that $n_k \theta_k \to t \leqq \infty$ and $x_k \to x$. It follows from (1.5) and (1.6) that $\mathscr{L}(Z_n^\theta) \to \mathscr{N}(0, g(t, x))$ as $k \to \infty$, where $\theta = \theta_k$, $n = n_k$, and $g(\infty, x) = \lim_{t\to\infty} g(t, x)$. Furthermore, it can be shown that $g$ is continuous over $[0, \infty] \times I$, so $\mathscr{N}(0, g(n_k \theta_k, x_k)) \to \mathscr{N}(0, g(t, x))$. Therefore, by the triangle inequality, $D(n_k, \theta_k, x_k) \to 0$ as $k \to \infty$, contradicting (1.8).

Another corollary of Theorem 1 is obtained by permitting $\theta$ to approach 0 after $n \to \infty$. Suppose that $\mathscr{L}(X_n^\theta)$ converges weakly as $n \to \infty$, for every fixed $\theta$. Let $\mathscr{L}(\theta)$ be the corresponding limit of $\mathscr{L}(Z_n^\theta)$, or, equivalently, of $\mathscr{L}(z_n^\theta)$, where

$$z_n^\theta = (X_n^\theta - \lambda)/\theta^{\frac{1}{2}}$$

($f(t) \to \lambda$ as $t \to \infty$). It follows easily from (1.6) that

$$\mathscr{L}(\theta) \to \mathscr{N}(0, g(\infty)) \tag{1.9}$$

as $\theta \to 0$. Some special results of this type were established in [5]. It is also a consequence of Theorem 1 that (1.9) holds for an arbitrary family $\mathscr{L}(\theta)$ of stationary distributions of $z_n^\theta$. This implies Part B of Theorem 10.1.1 (i) of [6].

The proof of Theorem 1 is given in Sections 2 and 3. Section 4 presents an application to the Wright–Fisher model for the evolution of gene frequency under the influence of mutation, selection, and random drift. In that context, $\theta = (2N)^{-\frac{1}{2}}$, where $N$ is the population size, and the function $f$ is the classical deterministic approximation to gene frequency for very large populations (see [2] Section 2.3). The results (1.6) and (1.7), which relate to the distribution of the error of this approximation, appear to be new even for this much studied model.

Section 5 gives a multidimensional analog of Theorem 1, and an illustrative application to a mathematical learning model.

## 2. Conditional moments of $\Delta Z_n^\theta$.

A basic component of the proof of Theorem 1 is Lemma 1.

LEMMA 1. *Under the hypotheses of Theorem* 1, $E((Z_n^\theta)^2)$ *is bounded over all* $\theta \in J$, $x_\theta \in I_\theta$, *and* $n \geqq 0$.

This result follows immediately from Theorem 3.2 (i) of [8]. The latter theorem assumes that $J$ is an interval and $I_\theta = I$ for all $\theta$, but these assumptions are not used in the proof. It emerges in the course of the proof that, for some $K < \infty$ and $\alpha > 0$,

$$|f(t, x) - \lambda| \leqq Ke^{-\alpha t}$$

for all $x \in I$ and $t \geqq 0$. Since $w'$ and $s$ satisfy Lipschitz conditions, we have

$$|w'(f(t, x)) - w'(\lambda)| \leqq Ke^{-\alpha t} \tag{2.1}$$

and

$$|s(f(t, x)) - s(\lambda)| \leqq Ke^{-\alpha t} \tag{2.2}$$

for suitable new constants $K$.

Henceforth we suppress the $\theta$ superscript on $Z_n^{\theta}$, and let $\nu_n = f(n\theta, x_\theta)$. The purpose of this section is to establish Lemma 2.

LEMMA 2.

$$E(\Delta Z_n \mid Z_n) = \theta w'(\nu_n)Z_n + o(\theta) \tag{2.3}$$

$$E((\Delta Z_n)^2 \mid Z_n) = \theta s(\nu_n) + o(\theta) \tag{2.4}$$

$$E(|\Delta Z_n|^3 \mid Z_n) = o(\theta)\,, \tag{2.5}$$

*where the quantities $o(\theta)$ satisfy $E(|o(\theta)|)/\theta \to 0$ as $\theta \to 0$, uniformly over $x_\theta \in I_\theta$ and $n \geqq 0$.*

PROOF. Since $w$ and $w'$ are bounded,

$$f''(t) = w'(f(t))w(f(t))$$

is too. Thus

$$\Delta\nu_n = \theta w(\nu_n) + O(\theta^2)$$

uniformly over $x_\theta$ and $n$. This expression and (1.1) imply that

$$\begin{aligned} E(\Delta Z_n \mid Z_n) &= \theta^{-\frac{1}{2}}(E(\Delta X_n \mid X_n) - \Delta\nu_n) \\ &= \theta^{\frac{1}{2}}(w(X_n) - w(\nu_n)) + O(\theta^{\frac{3}{2}})\,. \end{aligned} \tag{2.6}$$

Since $w'$ is Lipschitz, this yields

$$E(\Delta Z_n \mid Z_n) = \theta w'(\nu_n)Z_n + \theta^{\frac{3}{2}}O(|Z_n|^2) + O(\theta^{\frac{3}{2}})\,,$$

which, in view of Lemma 1, is of the form (2.3).

Turning to the proof of (2.4), we begin by writing

$$E((\Delta Z_n)^2 \mid Z_n) = \theta^{-1}\operatorname{Var}(\Delta X_n \mid X_n) + E(\Delta Z_n \mid Z_n)^2\,. \tag{2.7}$$

As a consequence of (2.6),

$$E(\Delta Z_n \mid Z_n) = \theta O(|Z_n|) + O(\theta^{\frac{3}{2}})\,, \tag{2.8}$$

so that

$$E(\Delta Z_n \mid Z_n)^2 \leqq K(\theta^2|Z_n|^2 + \theta^3)$$

and

$$E(\Delta Z_n \mid Z_n)^2 = o(\theta) \tag{2.9}$$

by Lemma 1. Next, (1.2) yields

$$\begin{aligned} \theta^{-1}\operatorname{Var}(\Delta X_n \mid X_n) &= \theta s(X_n) + o(\theta) \\ &= \theta s(\nu_n) + \theta^{\frac{3}{2}}O(|Z_n|) + o(\theta) \\ &= \theta s(\nu_n) + o(\theta) \end{aligned} \tag{2.10}$$

by Lemma 1. Substituting (2.9) and (2.10) into (2.7), we obtain (2.4).

Finally,

$$E(|\Delta Z_n|^3 \mid Z_n) \leqq 4\theta^{-\frac{3}{2}}(E(|\Delta X_n|^3 \mid X_n) + |\nu_n|^3)$$
$$\leqq K\theta^{\frac{3}{2}}$$

as a consequence of (1.3) and the boundedness of $w$. This implies (2.5).

**3. A general central limit theorem.** In view of (2.1), (2.2), Lemma 1, and Lemma 2, Theorem 1 is a corollary of Theorem 2.

THEOREM 2. *Suppose that $Z_n^\theta$, $n \geqq 0$, $\theta \in J$, is a family of real-valued stochastic processes such that*

$$E(\Delta Z_n^\theta \mid Z_n^\theta) = \theta a(n, \theta) Z_n^\theta + o(\theta) \tag{3.1}$$

$$E((\Delta Z_n^\theta)^2 \mid Z_n^\theta) = \theta b(n, \theta) + o(\theta) \tag{3.2}$$

$$E(|\Delta Z_n^\theta|^3 \mid Z_n^\theta) = o(\theta)\,, \tag{3.3}$$

*where*

$$\sup_{n \geqq 0} E(|o(\theta)|)/\theta \to 0$$

*as $\theta \to 0$,*

$$a(n, \theta) \to a \quad \text{and} \quad b(n, \theta) \to b \tag{3.4}$$

*as $\theta \to 0$ and $n\theta \to \infty$, and $a < 0$. Suppose also that*

$$\sup_{n \geqq 0, \theta \in J} E((Z_n^\theta)^2) < \infty\,. \tag{3.5}$$

*Then $\mathscr{L}(Z_n^\theta) \to \mathscr{N}(0, \sigma^2)$ as $\theta \to 0$ and $n\theta \to \infty$, where $\sigma^2 = b/2|a|$.*

PROOF. Let

$$h_n(\gamma) = h_n^\theta(\gamma) = E(\exp(i\gamma Z_n))\,.$$

Then

$$h_{n+1}(\gamma) = E(\exp(i\gamma Z_n)) E(\exp(i\gamma\, \Delta Z_n \mid Z_n))\,. \tag{3.6}$$

Expanding $\exp(i\gamma\, \Delta Z_n)$ up to terms of third order in $\gamma$ and using (3.1)—(3.3) we obtain

$$\Delta h_n(\gamma) = \theta\gamma a(n, \theta) h_n'(\gamma) - \theta 2^{-1}\gamma^2 b(n, \theta) h_n(\gamma) + d_n(\gamma)\,, \tag{3.7}$$

where

$$|d_n(\gamma)| \leqq \theta\varepsilon_\theta |\gamma| \tag{3.8}$$

and $\varepsilon_\theta$ is our generic notation for a quantity that depends only on $\theta$ and approaches 0 as $\theta$ approaches 0. This estimate is valid for all $n \geqq 0$ as long as $\gamma$ is bounded, $|\gamma| \leqq \Gamma$. From (3.7) it follows that

$$\Delta h_n(\gamma) = \theta\gamma a h_n'(\gamma) - \theta 2^{-1}\gamma^2 b h_n(\gamma) + d_n(\gamma) + e_n(\gamma)\,, \tag{3.9}$$

where, in view of (3.4) and (3.5),

$$|e_n(\gamma)| \leqq \theta c(n, \theta)|\gamma|\,, \tag{3.10}$$

and $c(n, \theta)$ is a quantity that approaches 0 as $\theta \to 0$ and $n\theta \to \infty$. The inequality (3.10) presupposes $|\gamma| \leqq \Gamma$.

Let

$$v(\gamma) = \exp(2^{-1}\gamma^2\sigma^2) ,$$
$$H_n(\gamma) = v(\gamma)h_n(\gamma) ,$$
$$D_n(\gamma) = v(\gamma)d_n(\gamma) ,$$
$$E_n(\gamma) = v(\gamma)e_n(\gamma) ,$$

and note that

$$v(\gamma)h_n'(\gamma) = H_n'(\gamma) - \sigma^2\gamma H_n(\gamma) .$$

Thus multiplication of (3.9) by $v(\gamma)$ yields

$$\Delta H_n(\gamma) = \theta\gamma a H_n'(\gamma) + D_n(\gamma) + E_n(\gamma) . \tag{3.11}$$

As a consequence of (3.8) and (3.10),

$$|D_n(\gamma)| \leqq \theta\varepsilon_\theta|\gamma| \tag{3.12}$$

and

$$|E_n(\gamma)| \leqq \theta c(n, \theta)|\gamma| \tag{3.13}$$

for $|\gamma| \leqq \Gamma$.

Let

$$\gamma_j = (1 + \theta a)^j \xi ,$$

where $\xi$ is fixed for the remainder of the proof. Assuming $\theta \leqq 1/|a|$,

$$|\gamma_j| \leqq e^{a\theta j}|\xi| . \tag{3.14}$$

In particular, $\gamma_j$ is bounded by $|\xi| = \Gamma$ for all $j$ and $\theta$.

For any $0 \leqq m \leqq M$, define $\mathscr{H}_m = \mathscr{H}_m(M, \theta)$ by

$$\mathscr{H}_m = H_m(\gamma_{M-m}) .$$

Then $\mathscr{H}_M = H_M(\xi)$. Suppose that $\mathscr{H}_M - \mathscr{H}_k \to 0$ as $\theta \to 0$ and $k\theta \to \infty$, while $\mathscr{H}_k \to 1$ as $\theta(M - k) \to \infty$. Then choosing $k = [M/2]$ we see that $H_M(\xi) \to 1$,

$$h_M(\xi) \to \exp(-2^{-1}\xi^2\sigma^2) ,$$

and $\mathscr{L}(Z_M) \to \mathscr{N}(0, \sigma^2)$ as $\theta \to 0$ and $M\theta \to \infty$, as the theorem asserts. Thus it remains only to show that $\mathscr{H}_M - \mathscr{H}_k \to 0$ and $\mathscr{H}_k \to 1$.

It may assist the reader in understanding the proof that $\mathscr{H}_M - \mathscr{H}_k \to 0$ to regard (3.11) as an approximation to the partial differential equation

$$\frac{\partial H(t, \gamma)}{\partial t} = \gamma a \frac{\partial H(t, \gamma)}{\partial \gamma} .$$

For any constant $g$, $(t, ge^{-at})$ is a characteristic base curve of this equation ([1] page 63), so

$$\frac{d}{dt} H(t, ge^{-at}) = 0 .$$

Since $\gamma_{M-m}$ approximates $\gamma_M e^{-am\theta}$, we expect $\mathscr{H}_m = H_m(\gamma_{M-m})$ to approximate $H(m\theta, \gamma_M e^{-am\theta})$. Thus $\mathscr{H}_m$ should vary little with $m$.

Clearly

$$\Delta\mathscr{H}_{m-1} = A_m - B_m$$

for $m \geqq 1$, where

$$A_m = H_m(\gamma_{M-m}) - H_{m-1}(\gamma_{M-m})$$

and

$$B_m = H_{m-1}(\gamma_{M+1-m}) - H_{m-1}(\gamma_{M-m})\ .$$

Thus

$$|\mathscr{H}_M - \mathscr{H}_k| \leqq \textstyle\sum_{m=k+1}^{M} |A_m - B_m|\ . \tag{3.15}$$

Now

$$\begin{aligned} B_m &= \Delta\gamma_{M-m} H'_{m-1}(\gamma_{M-m}) + F_{m-1} \\ &= \theta a \gamma_{M-m} H'_{m-1}(\gamma_{M-m}) + F_{m-1}\ , \end{aligned} \tag{3.16}$$

where

$$\begin{aligned} |F_{m-1}| &\leqq 2^{-1}|\Delta\gamma_{M-m}|^2 \max\nolimits_{|\gamma|\leqq|\xi|} |H''_{m-1}(\gamma)| \\ &\leqq K\theta^2 e^{2a\theta(M-m)} \end{aligned} \tag{3.17}$$

by virtue of (3.14) and (3.5). When the expression (3.16) for $B_m$ is subtracted from (3.11) for $A_m$, the leading terms cancel, so that

$$A_m - B_m = D_{m-1}(\gamma_{M-m}) + E_{m-1}(\gamma_{M-m}) - F_{m-1}\ .$$

Applying the estimates (3.12), (3.13), (3.14), and (3.17) to (3.15), we obtain

$$\begin{aligned} |\mathscr{H}_M - \mathscr{H}_k| &\leqq (\varepsilon_\theta + \sup\nolimits_{n\geqq k} c(n, \theta))\theta \textstyle\sum_{m=k+1}^{M} e^{a\theta(M-m)} \\ &\leqq (\varepsilon_\theta + \sup\nolimits_{n\geqq k} c(n, \theta))\theta/(1 - e^{a\theta})\ . \end{aligned}$$

Since $\varepsilon_\theta \to 0$ as $\theta \to 0$, and $c(n, \theta) \to 0$ as $\theta \to 0$ and $n\theta \to \infty$, it follows that $\mathscr{H}_M - \mathscr{H}_k \to 0$ as $\theta \to 0$ and $k\theta \to \infty$.

Note, finally, that

$$\begin{aligned} |h_k(\gamma_{M-k}) - 1| &\leqq |\gamma_{M-k}|E(|Z_k|) \\ &\leqq K|\gamma_{M-k}| \end{aligned}$$

by (3.5). Since $\gamma_{M-k} \to 0$ as $\theta(M-k) \to \infty$, we have $h_k(\gamma_{M-k}) \to 1$ and thus

$$\mathscr{H}_k = h_k(\gamma_{M-k})v(\gamma_{M-k}) \to 1$$

as $\theta(M-k) \to \infty$. This completes the proof.

**4. The Wright-Fisher model.** Suppose that there are two alleles, $A_1$ and $A_2$, at a certain chromosomal locus in a diploid population of $N$ individuals. Let $i$ be the number and $x = i/2N$ the proportion of $A_1$ genes in the population at any time. According to the model (see [2] Section 4.8), values $X_n$ of $x$ in successive generations form a finite Markov chain with transition probabilities

$$p_{ij} = \binom{2N}{j}\pi_i^{\,j}(1 - \pi_i)^{2N-j}\ ,$$

where

$$\pi_i = (1 - u)\pi_i^* + v(1 - \pi_i^*)$$

and

$$\pi_i^* = \frac{(1 + s_1)x^2 + (1 + s_2)x(1 - x)}{(1 + s_1)x^2 + 2(1 + s_2)x(1 - x) + (1 - x)^2}\ .$$

The constants $s_1$, $s_2$, $u$, and $v$ control selection pressure and mutation rate. The fitnesses of the genotypes $A_1A_1$ and $A_1A_2$, relative to that of $A_2A_2$, are $1 + s_1$ and $1 + s_2$, respectively. The probability that an $A_1$ gene mutates to $A_2$ is $u$, while the probability that $A_2$ mutates to $A_1$ is $v$.

To apply Theorem 1 to this model, we assume that these parameters are proportional to $\theta = (2N)^{-\frac{1}{2}}$: $s_i = \bar{s}_i\theta$, $u = \bar{u}\theta$, and $v = \bar{v}\theta$, where $\bar{u}, \bar{v} \geqq 0$. The routine verification of the assumptions in the second paragraph of Section 1 is given in ([6] Section 18.1), where it is also shown that $s(x) = x(1 - x)$ and

$$w(x) = \bar{v} - (\bar{u} + \bar{v})x + x(1 - x)(\bar{s}_2 + (\bar{s}_1 - 2\bar{s}_2)x) \tag{4.1}$$

on $I = [0, 1]$. Thus Theorem 1 applies whenever $w$ has a unique root $\lambda$ and $w'(\lambda) < 0$ (i.e., $\lambda$ is stable).

The following conditions are sufficient but by no means necessary for this: $\bar{u} > 0, \bar{v} > 0$, and $\bar{s}_1 \leqq 2\bar{s}_2$. (Proof. Since $w(0) = \bar{v} > 0$ and $w(1) = -\bar{u} < 0$, $w$ has at least one zero in $(0, 1)$. If $\bar{s}_1 = 2\bar{s}_2$, $w$ is quadratic or linear, and uniqueness and stability certainly obtain. If $\bar{s}_1 < 2\bar{s}_2$, the coefficient of $x^3$ is positive, so $w$ has a root above 1 and a root below 0. Thus $w$ has only one root $\lambda$ in $(0, 1)$ and it must satisfy $w'(\lambda) < 0$.) The inequality $\bar{s}_1 \leqq 2\bar{s}_2$ admits a number of genetically significant special cases:

(i) no dominance, $\bar{s}_1 = 2\bar{s}_2$;
(ii) favored gene completely dominant, $\bar{s}_1 = \bar{s}_2 > 0$ or $\bar{s}_1 < \bar{s}_2 = 0$; and
(iii) heterozygote advantage, $\bar{s}_1 < \bar{s}_2 > 0$.

Writing $X_n^N$ and $x^N$ for $X_n^\theta$ and $x_\theta$, the conclusion of Theorem 1 can be expressed as follows:

$$(2N)^{\frac{1}{4}}[X_n^N - f(n/(2N)^{\frac{1}{2}}, x^N)] \sim \mathscr{N}(0, g(\infty))$$

as $N \to \infty$ and $n/N^{\frac{1}{2}} \to \infty$. The occurrence of the fourth root on the left is noteworthy. (We observe that the related results in lines 13 and 22 on page 259 of [6] should have fourth roots instead of square roots.)

To see the relation of Theorem 1 to other diffusion approximations of the Wright–Fisher model, suppose that the mutation and selection parameters are proportional to a parameter $\varepsilon > 0$, i.e., $s_i = \bar{s}_i\varepsilon$, $u = \bar{u}\varepsilon$, and $v = \bar{v}\varepsilon$. Theorem 1 pertains directly to $\varepsilon = (2N)^{-\frac{1}{2}}$, but it turns out that this result is typical of those obtained when $\varepsilon \to 0$ sufficiently slowly that $N\varepsilon \to \infty$. If the function $w$ given in (4.1) satisfies the hypotheses of Theorem 1, then $(\varepsilon N)^{\frac{1}{2}}X_n$, suitably centered, is asymptotically normally distributed as $\varepsilon \to 0$, $N\varepsilon \to \infty$, and $n\varepsilon \to \infty$. This generalization of Theorem 1 will be proved in a subsequent paper. For a clear heuristic analysis of the asymptotic behavior of $X_n$ when $\varepsilon \to 0$ and $N\varepsilon \to \infty$, see Section 9 of [3].

Suppose now that $\varepsilon = (2N)^{-1}$. In this case, $\mathscr{L}(X_n^N) \to \mathscr{P}(t, x)$ as $x^N \to x$, $N \to \infty$, and $n\varepsilon \to t < \infty$, where $\mathscr{P}(t, x)$ is a nondegenerate distribution associated with a diffusion on $I$ ([6] page 260). The standard diffusion approximations of population genetics are of this type ([2] Section 5.1). This result, like the analogous result (1.5), is valid whether or not the function $w$ in (4.1) satisfies the hypotheses of Theorem 1. One would like to know what auxiliary conditions, if any, must be imposed to insure that $\mathscr{L}(X_n{}^N)$ converges to $\lim_{t\to\infty} \mathscr{P}(t, x)$ as $x^N \to x$, $N \to \infty$, and $n\varepsilon \to \infty$.

**5. Multidimensional case.** Suppose that the assumptions of the first two paragraphs of Section 1 are in force, except that $X_n{}^\theta$ is $k$ dimensional, and the conditional variance in (1.2) is replaced by the conditional covariance matrix. Then (1.5) is valid, where the asymptotic covariance matrix $g(t) = g(t, x)$ satisfies

$$g'(t) = w'(f(t))g(t) + g(t)w'(f(t))^* + s(f(t)),$$

and * indicates transposition ([6] Theorem 8.1.1). Theorem 3 is the multidimensional analog of Theorem 1.

THEOREM 3. *Suppose that the following additional conditions obtain: $I$ is bounded, there is a point $\lambda$ such that $w(\lambda) = 0$, and there is an inner product $[x, y]$ on $R^k$ such that*

$$[x - \lambda, w(x)] < 0 \tag{5.1}$$

*for all $x \in I$, $x \neq \lambda$, and*

$$[z, w'(\lambda)z] < 0$$

*for all $z \in R^k$, $z \neq 0$. Then*

$$\mathscr{L}(Z_n{}^\theta) \to \mathscr{N}(0, g(\infty))$$

*as $\theta \to 0$ and $n\theta \to \infty$, where $g(\infty)$ is the unique solution of the system*

$$w'(\lambda)g(\infty) + g(\infty)w'(\lambda)^* + s(\lambda) = 0$$

*of linear equations.*

Obviously (5.1) implies that $\lambda$ is the only zero of $w$. The most general inner product on $R^k$ is $[x, y] = (x, Py)$, where $(x, y)$ is the Euclidean inner product and $P$ is a positive definite matrix.

Theorem 3 can be established by a straightforward generalization of the proof of Theorem 1. This involves establishing the multidimensional generalizations of Lemmas 1 and 2 and Theorem 2. We omit details.

Theorem 3 is applicable to the Zeaman–House–Lovejoy learning model [7], which describes how a human or lower animal might learn to attend to a certain "relevant" dimension of a multidimensional stimulus. In this rather complex model, $X_n$ is two dimensional and $I$ is the closed unit square. There are six learning rate parameters, $\varphi_1$, $\varphi_2$, $\varphi_3$, $\varphi_4$, $\theta_1$, $\theta_2$, and two payoff probability parameters, $\pi_B$ and $\pi_W$. To apply Theorem 3, we assume that the learning rate parameters are all proportional to a single parameter $\theta$, i.e., $\varphi_i = \theta\bar{\varphi}_i$ and $\theta_j = \theta\bar{\theta}_j$, where $\bar{\varphi}_i$ and $\bar{\theta}_j$ are positive constants. It can be shown that the hypotheses of Theorem 3 are satisfied if and only if one of the following conditions holds: (i) $\pi_B < 1$ and $\pi_W < 1$, or (ii) $\max(\pi_B, \pi_W) = 1$, $\min(\pi_B, \pi_W) < 1$, and

$\bar{\varphi}_1 > \bar{\varphi}_3(\pi_B + \pi_W)/2$. In either case we can take $[x, x'] = x_1 x_1' + c x_2 x_2'$ for $c$ sufficiently large. Under condition (i), $\lambda$ is in the interior of $I$, while under condition (ii), it is one of the corners.

## REFERENCES


[1] COURANT, R. and HILBERT, D. (1962). *Methods of Mathematical Physics* **2**. Interscience, New York.

[2] EWENS, W. J. (1969). *Population Genetics*. Methuen, London.

[3] FELLER, W. (1951). Diffusion processes in genetics. *Proc. Second Berkeley Symp. Math. Statist. Prob.* Univ. of California Press, 227–246.

[4] KURTZ, T. G. (1971). Limit theorems for sequences of jump Markov processes approximating ordinary differential processes. *J. Appl. Probability* **8** 344–356.

[5] NORMAN, M. F. (1966). Limiting distributions for some random walks arising in learning models. *Ann. Math. Statist.* **37** 393–405.

[6] NORMAN, M. F. (1972). *Markov Processes and Learning Models*. Academic Press, New York.

[7] NORMAN, M. F. (1974a). Effects of overtraining, problem shifts, and probabilistic reinforcement in discrimination learning: predictions of an attentional model, in *Contemporary Developments in Mathematical Psychology* **1**, ed. D. H. Krantz, R. C. Atkinson, R. D. Luce, and P. Suppes. Freeman, San Francisco, 185–208.

[8] NORMAN, M. F. (1974b). Markovian learning processes. *SIAM Rev.* **16** 143–162.

[9] ROSÉN, B. (1967). On the central limit theorem for sums of dependent random variables. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete.* **7** 48–82.



DEPARTMENT OF PSYCHOLOGY
UNIVERSITY OF PENNSYLVANIA
3815 WALNUT STREET
PHILADELPHIA, PA. 19174