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 \( \Delta X_n^\theta \), \( \Delta X_n^\theta \), and \( \Delta X_n^\theta \), 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 \( (X_n^\theta - f(n\theta))/\theta \) 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 \( \{X_n^\theta\}_{n \geq 0} \) 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:

\[
E(\Delta X_n^\theta | X_n^\theta = x) = \theta w(x) + O(\theta^2)
\]

\[
\text{Var}(\Delta X_n^\theta | X_n^\theta = x) = \theta^3 s(x) + o(\theta^3)
\]

\[
E(|\Delta X_n^\theta|^3 | X_n^\theta = x) = O(\theta^3),
\]

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

\[
\sup_{\theta \in J, x \in I_\theta} |O(\theta^3)|/\theta^2 < \infty
\]

and

\[
\sup_{x \in I_\theta} |o(\theta^3)|/\theta^2 \to 0
\]

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 \( x \in I \),

\[
\inf_{y \in I_\theta} |y - x| \to 0
\]

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

\[
g'(t) = 2w(f(t))g(t) + s(f(t))
\]
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

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

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

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

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

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

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 $L(Z_n^\theta)$ 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

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

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 $d(L_n, L) \to 0$ whenever $L_n \to L$ weakly. Then

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

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

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

as $\theta \to 0$, where

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

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
\begin{equation}
D(n_k, \theta_k, x_k) \geq c
\end{equation}
for all $k \geq 1$. We could, moreover, choose these sequences in such a way that $n_k \theta_k \to t \leq \infty$ and $x_k \to x$. It follows from (1.5) and (1.6) that $\mathcal{L}(Z_n^\theta) \to \mathcal{N}(0, g(t, x))$ 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 $\mathcal{N}(0, g(n_k \theta_k, x_k)) \to \mathcal{N}(0, g(t, x))$. 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 $\mathcal{L}(X_n^{\theta})$ converges weakly as $n \to \infty$, for every fixed $\theta$. Let $\mathcal{L}(\theta)$ be the corresponding limit of $\mathcal{L}(Z_n^\theta)$, or, equivalently, of $\mathcal{L}(z_n^{\theta})$, where
\begin{equation}
z_n^{\theta} = (X_n^{\theta} - \lambda)/\theta
\end{equation}
($f(t) \to \lambda$ as $t \to \infty$). It follows easily from (1.6) that
\begin{equation}
\mathcal{L}(\theta) \to \mathcal{N}(0, g(\infty))
\end{equation}
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 $\mathcal{L}(\theta)$ 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, $\theta = (2N)^{-i}$, 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, $E((Z_n^{\theta})^5)$ is bounded over all $\theta \in J$, $x_\theta \in I_\theta$, and $n \geq 0$.

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 $\alpha > 0$,
\begin{equation}
|f(t, x) - \lambda| \leq Ke^{-at}
\end{equation}
for all $x \in I$ and $t \geq 0$. Since $w'$ and $s$ satisfy Lipschitz conditions, we have
\begin{equation}
|w'(f(t, x)) - w'(\lambda)| \leq Ke^{-at}
\end{equation}
and
\begin{equation}
|s(f(t, x)) - s(\lambda)| \leq Ke^{-at}
\end{equation}
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.**
\begin{align}
E(\Delta Z_n \mid Z_n) &= \theta w'(\nu_n)Z_n + o(\theta) \\
E((\Delta Z_n)^2 \mid Z_n) &= \theta s(\nu_n) + o(\theta) \\
E(|\Delta Z_n|^3 \mid Z_n) &= o(\theta),
\end{align}
where the quantities \( o(\theta) \) satisfy \( E(|o(\theta)|)/\theta \to 0 \) as \( \theta \to 0 \), uniformly over \( x_\theta \in I_\theta \) and \( n \geq 0 \).

**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{equation}
E(\Delta Z_n \mid Z_n) = \theta^{-1}(E(\Delta X_n \mid X_n) - \Delta \nu_n) \\
= \theta^{-1}(w(X_n) - w(\nu_n)) + O(\theta^2).
\end{equation}
Since \( w' \) is Lipschitz, this yields
\[ E(\Delta Z_n \mid Z_n) = \theta w'(\nu_n)Z_n + \theta^3 O(|Z_n|^3) + O(\theta^4), \]
which, in view of Lemma 1, is of the form (2.3).

Turning to the proof of (2.4), we begin by writing
\begin{equation}
E((\Delta Z_n)^2 \mid Z_n) = \theta^{-1} \text{Var}(\Delta X_n \mid X_n) + E(\Delta Z_n \mid Z_n)^2.
\end{equation}
As a consequence of (2.6),
\begin{equation}
E(\Delta Z_n \mid Z_n) = \theta O(|Z_n|) + O(\theta^3),
\end{equation}
so that
\[ E(\Delta Z_n \mid Z_n)^2 \leq K(\theta^3|Z_n|^3 + \theta^3) \]
and
\begin{equation}
E(\Delta Z_n \mid Z_n)^2 = o(\theta)
\end{equation}
by Lemma 1. Next, (1.2) yields
\[ \theta^{-1} \text{Var}(\Delta X_n \mid X_n) = \theta s(X_n) + o(\theta) \]
\begin{equation}
= \theta s(\nu_n) + \theta^3 O(|Z_n|) + o(\theta) \\
= \theta s(\nu_n) + o(\theta)
\end{equation}
by Lemma 1. Substituting (2.9) and (2.10) into (2.7), we obtain (2.4).
Finally,
\[ E(|\Delta Z_n| \mid Z_n) \leq 4n^{-1}(E(|\Delta X_n| \mid X_n) + |\nu_n|^3) \]
\[ \leq K\theta^3 \]
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 \( Z_n^\theta, n \geq 0, \theta \in J \), is a family of real-valued stochastic processes such that

\[
\begin{align*}
E(\Delta Z_n^\theta \mid Z_n^\theta) &= \theta a(n, \theta)Z_n^\theta + o(\theta) \\
E((\Delta Z_n^\theta)^2 \mid Z_n^\theta) &= \theta b(n, \theta) + o(\theta) \\
E(|\Delta Z_n^\theta|^3 \mid Z_n^\theta) &= o(\theta),
\end{align*}
\]

where
\[
\sup_{n \geq 0} E(|o(\theta)|)/\theta \to 0
\]
as \( \theta \to 0 \),

\[(3.4) \quad a(n, \theta) \to a \quad \text{and} \quad b(n, \theta) \to b\]
as \( \theta \to 0 \) and \( n\theta \to \infty \), and \( a < 0 \). Suppose also that

\[(3.5) \quad \sup_{n \geq 0, \theta \in J} E((Z_n^\theta)^2) < \infty .
\]

Then \( \mathcal{L}(Z_n^\theta) \to \mathcal{N}(0, \sigma^2) \) 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
\[(3.6) \quad h_{n+1}(\gamma) = E(\exp(i\gamma Z_n))E(\exp(i\gamma \Delta Z_n \mid Z_n)).
\]

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

\[
\Delta h_n(\gamma) = \theta \gamma a(n, \theta) h_n(\gamma) - \theta 2^{-1} \gamma^2 b(n, \theta) h_n(\gamma) + d_n(\gamma),
\]

where
\[
|d_n(\gamma)| \leq \theta \varepsilon |\gamma|
\]
and \( \varepsilon \) 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 \( n \geq 0 \) as long as \( \gamma \) is bounded, \( |\gamma| \leq \Gamma \). From (3.7) it follows that

\[
\Delta h_n(\gamma) = \theta \gamma a h_n(\gamma) - \theta 2^{-1} \gamma^2 b h_n(\gamma) + d_n(\gamma) + e_n(\gamma),
\]

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

\[
|e_n(\gamma)| \leq \theta c(n, \theta)|\gamma|,
\]
and \( c(n, \theta) \) is a quantity that approaches 0 as \( \theta \to 0 \) and \( n\theta \to \infty \). The inequality (3.10) presupposes \( |\gamma| \leq \Gamma \).
Let

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

and note that

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

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

\[ \Delta H_n(\gamma) = \theta a H_n'(\gamma) + D_n(\gamma) + E_n(\gamma). \]

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

\[ |D_n(\gamma)| \leq \theta \varepsilon_0|\gamma| \]

and

\[ |E_n(\gamma)| \leq \theta c(n, \theta)|\gamma| \]

for \( |\gamma| \leq \Gamma \).

Let

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

where \( \xi \) is fixed for the remainder of the proof. Assuming \( \theta \leq 1/|a| \),

\[ |\gamma_j| \leq e^{\theta j}|\xi|. \]

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

For any \( 0 \leq m \leq M \), define \( \mathcal{H}_m = \mathcal{H}_m(M, \theta) \) by

\[ \mathcal{H}_m = H_m(\gamma_{M-m}). \]

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

\[ h_m(\xi) \to \exp(-2^{-1}a^2), \]

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

It may assist the reader in understanding the proof that \( \mathcal{H}_M - \mathcal{H}_k \to 0 \) 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 \( \gamma \), \( (t, ge^{-at}) \) 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 \( \mathcal{H}_m = H_m(\gamma_{M-m}) \) to approximate \( H(M\theta, \gamma_M e^{-aM\theta}) \). Thus \( \mathcal{H}_m \) should vary little with \( m \).
Clearly
\[ \Delta \mathcal{X}_{m-1} = A_m - B_m \]
for \( m \geq 1 \), where
\[ A_m = H_m(\tau_{M-m}) - H_{m-1}(\tau_{M-m}) \]
and
\[ B_m = H_{m-1}(\tau_{M+1-m}) - H_{m-1}(\tau_{M-m}) \).

Thus
(3.15)
\[ |\mathcal{X}_M - \mathcal{X}_k| \leq \sum_{n=k+1}^M |A_m - B_m| . \]

Now
(3.16)
\[ B_m = \Delta \gamma_{M-m} H'_{m-1}(\gamma_{M-m}) + F_{m-1} \]
\[ = \theta a_{M-m} H'_{m-1}(\gamma_{M-m}) + F_{m-1} , \]
where
(3.17)
\[ |F_{m-1}| \leq 2^{-1} |\Delta \gamma_{M-m}|^3 \max_{|\gamma| \leq |\gamma|} |H''_{m-1}(\gamma)| \]
\[ \leq K \theta^2 e^{2\theta(M-m)} \]

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
\[ |\mathcal{X}_M - \mathcal{X}_k| \leq (\varepsilon_{\theta} + \sup_{|n| \leq k} c(n, \theta)) \theta \sum_{n=k+1}^M e^{\theta(M-m)} \]
\[ \leq (\varepsilon_{\theta} + \sup_{|n| \leq k} c(n, \theta)) \theta/(1 - e^{\theta}) . \]

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 \( \mathcal{X}_M - \mathcal{X}_k \to 0 \) as \( \theta \to 0 \) and \( k \theta \to \infty \).

Note, finally, that
\[ |h_k(\gamma_{M-k}) - 1| \leq |\gamma_{M-k}| E(|Z_k|) \]
\[ \leq K |\gamma_{M-k}| \]
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
\[ \mathcal{X}_k = h_k(\gamma_{M-k}) \phi(\gamma_{M-k}) \to 1 \]
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
\[ p_{ij} = \binom{N}{j} \pi_i^j (1 - \pi_i)^{2N-j} , \]
where
\[ \pi_i = (1 - u) \pi_i^* + \nu(1 - \pi_i^*) \]
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 \( A_i A_i \) and \( A_i A_s \), relative to that of \( A_s A_s \), are \( 1 + s_i \) and \( 1 + s_s \), respectively. The probability that an \( A_i \) gene mutates to \( A_s \) is \( u \), while the probability that \( A_s \) mutates to \( A_i \) is \( v \).

To apply Theorem 1 to this model, we assume that these parameters are proportional to \( \theta = (2N)^{-1} \): \( s_i = \bar{s_i} \theta, u = \bar{u} \theta, \) and \( v = \bar{v} \theta \), where \( \bar{u}, \bar{v} \geq 0 \). 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

\[
(4.1) \quad w(x) = \bar{\theta} - (\bar{u} + \bar{v})x + x(1 - x)(\bar{s}_i + (\bar{s}_i - 2\bar{s}_s)x)
\]
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 \( \bar{s}_i \leq 2\bar{s}_s \). (Proof. Since \( w(0) = \bar{\theta} > 0 \) and \( w(1) = -\bar{u} < 0 \), \( w \) has at least one zero in \((0, 1)\). If \( \bar{s}_i = 2\bar{s}_s \), \( w \) is quadratic or linear, and uniqueness and stability certainly obtain. If \( \bar{s}_i < 2\bar{s}_s \), the coefficient of \( x^i \) 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 \( \bar{s}_i \leq 2\bar{s}_s \) admits a number of genetically significant special cases:

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

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

\[
(2N)^i[X_s^N - f(n/(2N)^i, x^N)] \sim \mathcal{N}(0, g(\infty))
\]
as \( N \to \infty \) and \( n/N^i \to \infty \). 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., \( s_i = s_i \varepsilon, u = u\varepsilon, \) and \( v = v\varepsilon \). Theorem 1 pertains directly to \( \varepsilon = (2N)^{-1} \), 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 \( (\varepsilon N)^iX_s \), 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 \( X_s \) when \( \varepsilon \to 0 \) and \( N\varepsilon \to \infty \), see Section 9 of [3].

Suppose now that \( \varepsilon = (2N)^{-1} \). In this case, \( \mathcal{L}(X_s^N) \to \mathcal{P}(t, x) \) as \( x^N \to x, N \to \infty, \) and \( n\varepsilon \to t < \infty \), where \( \mathcal{P}(t, x) \) is a nondegenerate distribution associated with a diffusion on \( I ([6] \text{ page 260}) \). 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 $\mathcal{L}(X^{n})$ converges to $\lim_{t \to \infty} \mathcal{P}(t, x)$ as
$x^{n} \to x$, $N \to \infty$, and $n \to \infty$.

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 multi-
dimensional 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

$$[x - \lambda, w(x)] < 0$$

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

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

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

$$\mathcal{L}(Z_{n}^{\theta}) \to \mathcal{N}(0, g(\infty))$$

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, $\varphi_{1}$, $\varphi_{2}$, $\varphi_{3}$, $\theta_{1}$, $\theta_{2}$, and two payoff probability
parameters, $\pi_{n}$ 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) $\pi_{n} < 1$ and $\pi_{w} < 1$, or (ii) $\max(\pi_{n}, \pi_{w}) = 1$, $\min(\pi_{n}, \pi_{w}) < 1$, and
\( \varphi_1 > \varphi_2 (\pi_B + \pi_W)/2. \) 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


Department of Psychology
University of Pennsylvania
3815 Walnut Street
Philadelphia, Pa. 19174