Adv. Appl. Prob. **7**, 561–575 (1975)
Printed in Israel

LIMIT THEOREMS FOR STATIONARY DISTRIBUTIONS

M. FRANK NORMAN, *University of Pennsylvania*

Abstract

For every $N \geqq 1$, let $X_0^N, X_1^N, X_2^N, \cdots$ be a real-valued stationary process, and let $\Delta X_n^N = X_{n+1}^N - X_n^N$. Suppose that $E(\Delta X_n^N \mid X_n^N) = O(\varepsilon^N)$ and $\operatorname{var}(\Delta X_n^N \mid X_n^N) = O(\tau^N)$, where ε^N and τ^N are positive null sequences. Limiting distributions of X_n^N as $N \to \infty$ are obtained for the cases $\tau^N = \varepsilon^N$ and $\tau^N = o(\varepsilon^N)$. These results are established by an extension of a method due to Moran. The theory is illustrated by a variety of applications to genetic models.

STATIONARY PROCESSES; LIMIT THEOREMS; DIFFUSION APPROXIMATION; GENETIC MODELS

1. Introduction and summary

For every $N \geqq 1$, let $X_0^N, X_1^N, X_2^N, \cdots$ be a stationary process in a real interval I. The distribution of X_n^N is denoted $\mathscr{L}(X_n^N)$. Suppose that the first two conditional moments of the increment $\Delta X_n^N = X_{n+1}^N - X_n^N$ can be represented in the form

$$E(\Delta X_n^N \mid X_n^N) = \tau^N a(X_n^N) + e_{1,n}^N, \tag{1.1}$$

$$E((\Delta X_n^N)^2 \mid X_n^N) = \tau^N b(X_n^N) + e_{2,n}^N, \tag{1.2}$$

where $\tau^N > 0$, $\tau^N \to 0$ as $N \to \infty$, and $E(|e_{i,n}^N|) = o(\tau^N)$. Moreover, for every $\delta > 0$,

$$E(|\Delta X_n^N|^2 I_{n,\delta}^N) = o(\tau^N), \tag{1.3}$$

where $I_{n,\delta}^N$ is the indicator of $\{|\Delta X_n^N| > \delta\}$.

Under such conditions, we might expect that $\mathscr{L}(X_n^N) \to \mu$ as $N \to \infty$, where μ is the distribution whose density, $f = kf_0$, is given by Wright's (1938) formula

$$f_0(x) = b(x)^{-1} e^{B(x)}, \tag{1.4}$$

with

$$B'(x) = 2b(x)^{-1} a(x). \tag{1.5}$$

Received in revised form 5 March 1975.

The purpose of this paper is to spell out some simple auxiliary conditions under which this expectation is fulfilled. These conditions are sufficiently general to encompass a wide variety of applications.

Moran ((1958a, b), (1962), Chap. VII) has considered problems of this type in the context of certain genetic models. The proof of our main result, Theorem 1, is an extension of Moran's method.

Note that Eqs. (1.1)–(1.3) involve only X_n^N and X_{n+1}^N. Moreover, it turns out that stationarity is needed only in so far as it implies that $\mathcal{L}(X_n^N) = \mathcal{L}(X_{n+1}^N)$. Thus it is natural to formulate Theorem 1 as a statement about pairs (X^N, Y^N), corresponding to (X_n^N, X_{n+1}^N) in the preceding notation.

Theorem 1. *Let* $\{(X^N, Y^N),\ N \geqq 1\}$ *be a sequence of bivariate random variables, such that* $\mathcal{L}(X^N) = \mathcal{L}(Y^N)$. *Suppose that* X^N *and* Y^N *take on values in a real interval I that includes finite endpoints, if any, and that* X^N *is measurable with respect to a* σ*-field* $\mathcal{F}^N$. *Let* $\Delta^N = Y^N - X^N$, *and suppose that*

$$E(\Delta^N | \mathcal{F}^N) = \tau^N a(X^N) + e_1^N, \tag{1.6}$$

$$E((\Delta^N)^2 | \mathcal{F}^N) = \tau^N b(X^N) + e_2^N, \tag{1.7}$$

where $\tau^N > 0$, $\tau^N \to 0$ *as* $N \to \infty$, *and*

$$E(|e_i^N|) = o(\tau^N), \qquad i = 1, 2. \tag{1.8}$$

Also,

$$E((\Delta^N)^2 I_\delta^N) = o(\tau^N) \tag{1.9}$$

for every $\delta > 0$, *where* I_δ^N *is the indicator of* $\{|\Delta^N| > \delta\}$.

It is assumed that a and b are continuous throughout I, and that $b(x) > 0$ *if x is in the interior of I. Let* $c = \inf I$ *and* $d = \sup I$, *and let* $\gamma > 0$ *be less than the length of I.* (i) *If* $c > -\infty$, *suppose that* $a(c) > 0$ *and* $\int_c^{c+\gamma} b(x)^{-1} dx = \infty$. (ii) *If* $d < \infty$, *then* $a(d) < 0$ *and* $\int_{d-\gamma}^{d} b(x)^{-1} dx = \infty$. (iii) *If I is unbounded, assume, in addition, that* $\{\mathcal{L}(X^N),\ N \geqq 1\}$ *is tight,* $\{a(X^N),\ N \geqq 1\}$ *and* $\{b(X^N),\ N \geqq 1\}$ *are uniformly integrable, and* $E(|X^N|) < \infty$.

Under these conditions,

$$p = \int_c^d f_0(x)\ dx < \infty, \tag{1.10}$$

and $\mathcal{L}(X^N) \to \mu$ *as* $N \to \infty$, *where* μ *has density* $f = p^{-1} f_0$.

In Theorem 1, $E(\Delta^N | \mathcal{F}^N)$ and $\operatorname{var}(\Delta^N | \mathcal{F}^N)$ are of the same order of magnitude. Theorem 2 deals with a situation where the mean tends to be larger than the variance. In this case, X^N converges in distribution to a constant, λ, and a linear function, $\bar{X}^N$, of X^N (see (1.16)) has a limiting normal distribution.

Theorem 2. *Suppose that X^N is $\mathcal{F}^N$|=|measurable, that $\mathcal{L}(X^N) = \mathcal{L}(Y^N)$, and that the state space, I, of X^N and Y^N is a closed bounded interval. In place of* (1.6)–(1.8), *assume that*

$$E(\Delta^N \mid \mathcal{F}^N) = \varepsilon^N a^N(X^N) + e_1^N, \tag{1.11}$$

$$\operatorname{var}(\Delta^N \mid \mathcal{F}^N) = \tau^N b^N(X^N) + e_2^N, \tag{1.12}$$

where $\varepsilon^N > 0$, $\tau^N > 0$, $\varepsilon^N \to 0$ as $N \to \infty$, $\tau^N/\varepsilon^N \to 0$ as $N \to \infty$, and

$$E((e_1^N)^2) = o(\varepsilon^N \tau^N), \tag{1.13}$$

$$E(|e_2^N|) = o(\tau^N). \tag{1.14}$$

Furthermore, letting $\delta^N = \Delta^N - E(\Delta^N \mid \mathcal{F}^N) = Y^N - E(Y^N \mid \mathcal{F}^N)$, and letting $\tilde{I}_\delta^N$ be the indicator of $\{|\delta^N| > \delta(\tau^N/\varepsilon^N)^{\frac{1}{2}}\}$,

$$E((\delta^N)^2 \tilde{I}_\delta^N) = o(\tau^N). \tag{1.15}$$

Suppose that a^N is continuously differentiable (even at the endpoints of I), and that there is a continuously differentiable function a such that $|a^N - a|_I = \sup_{x \in I} |a^N(x) - a(x)| \to 0$ and $|(a^N)' - a'|_I \to 0$ as $N \to \infty$. The function a has $a(c) > 0$ and $a(d) < 0$. Moreover it has a unique zero, λ, and $a'(\lambda) < 0$ (i.e., λ is stable). There is a continuous function b such that $|b^N - b|_I \to 0$.

Under these conditions, for N sufficiently large, a^N has a unique root, λ^N, and $\lambda^N \to \lambda$. Furthermore $b(\lambda) \geqq 0$, and

$$\bar{X}^N = (X^N - \lambda^N)\ (\varepsilon^N/\tau^N)^{\frac{1}{2}} \tag{1.16}$$

is asymptotically normally distributed, with mean 0 *and variance*

$$\sigma^2 = b(\lambda)/2|a'(\lambda)|, \tag{1.17}$$

as $N \to \infty$.

This theorem was suggested by the heuristic considerations in Sec. 9 of Feller (1951). We shall show that it is a corollary of Theorem 1. A result of this type was obtained previously [Norman ((1972), Theorem 10.1.1), (1974)] under the unnecessarily restrictive condition $(\varepsilon^N)^2 = \tau^N$.

Proofs of Theorems 1 and 2 are presented in Sections 2 and 3, respectively. Section 4 contains a number of applications to genetic models. Let α_1 and α_2 be the mutation (or migration) rates in these models, let ν_1 and ν_2 be the selection intensities and let N be the population size. Moran's papers (1958a, b) treat the case $\alpha_i = O(N^{-1})$, $\nu_i = O(N^{-1})$. Theorem 1 is applicable under these conditions, as would be expected. In addition, Theorem 2 permits us to treat the case where α_i and ν_i are of larger order of magnitude than N^{-1}. This is presumably an appropriate analytical representation of a very large population.

Theorems 1 and 2 are concerned with the stationary aspect of the general problem of diffusion approximation. For closely related recent results pertaining to diffusion approximation of pre-asymptotic behavior, see Norman (1975a, b).

2. Proof of Theorem 1

Suppose that N_j is a subsequence of the positive integers such that $\mathscr{L}(X^{N_j}) \to \nu$ as $j \to \infty$. We shall show that $\nu = \mu$. Since our assumptions ensure compactness of $\{\mathscr{L}(X^N), N \geqq 1\}$, it follows that $\mathscr{L}(X^N) \to \mu$ as $N \to \infty$.

Since the proof is lengthy, we shall break it up into a sequence of lemmas. Our first major objective is to show that $\int_I (ah' + 2^{-1}bh'')\, d\nu = 0$ for a large class of functions h (Lemma 2.2). From this it will be deduced that ν has density kf_0 in the interior, I°, of I (Lemma 2.5), and that $\nu(I - I^\circ) = 0$ (Lemma 2.6).

Let $\mathscr{H}$ be the class of twice differentiable real-valued functions on $R = (-\infty, \infty)$, such that h' and h'' are bounded, and h'' is uniformly continuous. It is not assumed that h itself is bounded. Thus, for example, $h(x) \equiv x$ is in $\mathscr{H}$. Let

$$\Gamma h(x) = a(x)h'(x) + 2^{-1}b(x)h''(x).$$

Lemma 2.1. *For $h \in \mathscr{H}$, $E(\Gamma h(X^N)) \to 0$ as $N \to \infty$.*

Proof. The Taylor expansion yields

$$h(Y) = h(X) + h'(X)\Delta + 2^{-1}h''(X)\Delta^2 + D(X, Y), \tag{2.1}$$

where

$$|D(X, Y)| \leqq 2^{-1}\omega_{|\Delta|}(h'')\Delta^2 \tag{2.2}$$

and $\omega_d(g) = \sup_{|x-y| \leqq d} |g(x) - g(y)|$. (Superscript N's will be suppressed throughout most of Sections 2 and 3.) Now $|h(X)| \leqq |h(0)| + |h'|_R |X|$, and $E(|X|) < \infty$, so $E(h(X))$ and $E(h(Y))$ exist. Moreover, they are equal, since $\mathscr{L}(X) = \mathscr{L}(Y)$. Thus (2.1) yields

$$\begin{aligned} 0 &= E[h'(X)\Delta + 2^{-1}h''(X)\Delta^2 + D(X, Y)] \\ &= E[h'(X)E(\Delta \mid \mathscr{F}) + 2^{-1}h''(X)E(\Delta^2 \mid \mathscr{F}) + D(X, Y)]. \end{aligned}$$

Thus, by (1.6) and (1.7),

$$E(\Gamma h(X)) = -\tau^{-1}E[h'(X)e_1 + 2^{-1}h''(X)e_2 + D(X, Y)],$$

and

$$|E(\Gamma h(X))| \leqq \tau^{-1}[|h'|_R E(|e_1|) + 2^{-1}|h''|_R E(|e_2|) + 2^{-1}E(\omega_{|\Delta|}(h'')\Delta^2)]$$

by (2.2). In view of (1.8), it remains only to show that $E(\omega_{|\Delta|}(h'')\Delta^2) = o(\tau^N)$.

Let $\delta > 0$ be given. Then

$$E(\omega_{|\Delta|}(h'')\Delta^2) \leqq 2|h''|_R E(\Delta^2 I_\delta) + \omega_\delta(h'')E(\Delta^2). \tag{2.3}$$

Now $E(b(X^N))$ is bounded. For if I is bounded, b itself is bounded, while if I is unbounded, $b(X^N)$ is uniformly integrable. Hence (1.7) and (1.8) yield $E((\Delta^N)^2) = O(\tau^N)$. Using this and (1.9) in conjunction with (2.3), we obtain

$$\limsup_{N\to\infty} \tau^{-1}E(\omega_{|\Delta|}(h'')\Delta^2) \leqq K\omega_\delta(h'').$$

Since h'' is uniformly continuous, $\omega_\delta(h'') \to 0$ as $\delta \to 0$, and Lemma 2.1 is proved.

Lemma 2.2. *For all* $h \in \mathcal{H}$, $\int_I \Gamma h\, d\nu = 0$.

Proof. In view of Lemma 2.1, we need only show $E(\Gamma h(X^{N_j})) \to \int_I \Gamma h\, d\nu$. If I is bounded, this is trivial, since, in that case, Γh is bounded as well as continuous. If I is unbounded, $\Gamma h(X^{N_j})$ is, at least, uniformly integrable, since $a(X^N)$ and $b(X^N)$ are assumed to be uniformly integrable and h' and h'' are bounded. Uniform integrability is all that is needed to ensure that $\int_I \Gamma h\, d\nu$ exists and $E(\Gamma h(X^{N_j}))$ converges to it.

Let

$$G(x) = \int_{I_x} a\, d\nu \tag{2.4}$$

and

$$J(x) = 2^{-1}\int_{I_x} b\, d\nu,$$

where $I_x = (-\infty, x] \cap I$. Let C_c be the continuous functions on R with compact support.

Lemma 2.3. *For every* $q \in C_c$,

$$\int q(x)G(x)\, dx = \int q(x)\, dJ(x).$$

(*When limits of integration are omitted, they are understood to be* $-\infty$ *and* ∞.)

Proof. Taking $h(x) \equiv x$ in Lemma 2.2, we see that

$$\int_I a\, d\nu = 0, \tag{2.5}$$

hence

$$G(-\infty) = G(\infty) = 0. \tag{2.6}$$

In our present notation, Lemma 2.2 takes the form

$$-\int h'\,dG = \int h''\,dJ.$$

If $h'' \in C_c$, the left-hand side can be integrated by parts to obtain

$$\int h''(x)G(x)\,dx = \int h''(x)\,dJ(x). \tag{2.7}$$

(The condition that $h'' \in C_c$ is necessary at this juncture, at least when I is unbounded, since, in that case, we are not yet in a position to assert that G is integrable over R. It is clear, however, that $|G(x)| \leqq \int_I |a|\,d\nu < \infty$, so G is locally integrable. There is no similar difficulty on the right-hand side of (2.7), since the total variation of J is $\int_I b\,d\nu < \infty$.)

If $q \in C_c$, let

$$h(x) = \int_{-\infty}^{x} \left(\int_{-\infty}^{y} q(z)\,dz \right) dy.$$

Then $h'(x) = \int_{-\infty}^{x} q(z)\,dz$, so $|h'|_R < \infty$, and $h'' = q \in C_c$. Hence (2.7) applies to h and yields Lemma 2.3.

Lemma 2.4. *For all Borel subsets A of I,*

$$\int_A G(x)\,dx = 2^{-1}\int_A b\,d\nu.$$

Proof. Let α and β be real numbers with $\alpha < \beta$, and let q_n be a sequence in C_c such that $q_n(x) = 0$ for $x \leqq \alpha$ or $x \geqq \beta$, $0 \leqq q_n(x) \leqq 1$ for $\alpha < x < \beta$, and $q_n(x) \to 1$ for $\alpha < x < \beta$. Applying Lemma 2.3 to q_n, and taking the limit as $n \to \infty$, we obtain

$$\int_\alpha^\beta G(x)\,dx = dJ(\alpha, \beta) \geqq 0. \tag{2.8}$$

Dividing by $\beta - \alpha$, letting $\beta \to \alpha$, and noting that $G(\alpha) = G(\alpha+)$, we see that $G(\alpha) \geqq 0$. Letting $\alpha \to -\infty$ and $\beta \to \infty$ in (2.8), we get $\int G(x)\,dx = \int_I b\,d\nu < \infty$. Since the finite Borel measures $G(x)dx$ and dJ agree on finite open intervals, they are identical. In particular, they agree on all Borel subsets of I, as Lemma 2.4 asserts.

Recall that f_0 is defined by (1.4) and (1.5).

Lemma 2.5. *There is a constant $k \geqq 0$ such that $\nu(A) = \int_A kf_0(x)\,dx$ for all $A \subset I^\circ$.*

Proof. Since $b(x) > 0$ on I°, it follows immediately from Lemma 2.4 that, on I°, ν is absolutely continuous with respect to Lebesgue measure. Let $f^* = d\nu/dx$ be its Radon-Nikodym derivative. Lemma 2.4 implies that

$$G(x) = 2^{-1}b(x)f^*(x) \tag{2.9}$$

a.e. on I°, and modifying f^* on a set of Lebesgue measure 0, we can assume that (2.9) holds for all $x \in I^\circ$.

For $x \in I^\circ$, (2.4) and (2.9) yield

$$2^{-1}b(x)f^*(x) = G(c) + \int_c^x a(x)f^*(x)\,dx. \tag{2.10}$$

The right side is continuous, so bf^* and thus f^* must also be continuous. We can thus differentiate (2.10) to obtain

$$(bf^*)'(x) = 2a(x)f^*(x), \tag{2.11}$$

or

$$(bf^*)'(x) = B'(x)b(x)f^*(x).$$

Thus $f^* = d\nu/dx = kf_0$, for some $k \geqq 0$, as Lemma 2.5 asserts.

Lemma 2.6. $\nu(I - I^\circ) = 0$.

Proof. Clearly $I - I^\circ$ is the set of finite endpoints of I. Suppose $c = \inf I > -\infty$. We shall show that $\nu(c) = 0$. By symmetry, $\nu(d) = 0$ if $d < \infty$.

Since $a(c) > 0$ and $\int_c^{c+\gamma} b(x)^{-1}dx = \infty$, we have $\int_c^{c+\gamma} B'(x)\,dx = \infty$. Thus $b(x)f_0(x) = e^{B(x)} \to 0$ as $x \to c$. Letting $x \to c$ in (2.10), and recalling that $f^* = kf_0$, we obtain $G(c) = a(c)\nu(c) = 0$. Thus $\nu(c) = 0$, as claimed.

In view of Lemmas 2.5 and 2.6, $\int_c^d kf_0(x)\,dx = \nu(I) = 1$, hence $k = p^{-1}$. Therefore ν has density $f = p^{-1}f_0$, i.e., $\nu = \mu$. This completes the proof of Theorem 1.

3. Proof of Theorem 2

We begin by establishing the existence, uniqueness and convergence of λ^N.

Lemma 3.1. *For N sufficiently large, a^N has a unique zero λ^N, and $\lambda^N \to \lambda$ as $N \to \infty$.*

Proof. Let

$$T(x, y) = \begin{cases} (a(x) - a(y))/(x - y) & \text{if } x \neq y. \\ a'(y) & \text{if } x = y, \end{cases}$$

$$= \int_0^1 a'(y + \xi(x - y))\, d\xi,$$

and let T^N be defined similarly with a^N in place of a. Since I is compact, a' and T are uniformly continuous. Clearly $T(x, \lambda) < 0$, for all $x \in I$, hence this function is bounded away from 0. Moreover, there is a $\gamma > 0$ so small that

$$T(x, y) \leqq -t < 0 \tag{3.1}$$

for all $x \in I$ and $|y - \lambda| \leqq \gamma$. Choose γ so small that $\lambda \pm \gamma \in I^{\circ}$.

Note that

$$|T^N - T|_{I \times I} \leqq |(a^N)' - a'|_I. \tag{3.2}$$

Thus, for N_0 sufficiently large, the supremum on the left is no larger than $t/2$. Hence, in view of (3.1),

$$T^N(x, y) \leqq -t/2 \tag{3.3}$$

for $N \geqq N_0$, $x \in I$, and $|y - \lambda| \leqq \gamma$.

Now $a(\lambda - \gamma) > 0$, and $a(\lambda + \gamma) < 0$. Hence, for N sufficiently large, $a^N(\lambda - \gamma) > 0$ and $a^N(\lambda + \gamma) < 0$, so a^N has a zero, λ^N, with $|\lambda^N - \lambda| < \gamma$. Moreover, (3.3) with $y = \lambda^N$ shows that λ^N is the only zero of a^N in I. Considering $\gamma^* < \gamma$, we see that $|\lambda^N - \lambda| < \gamma^*$ for N sufficiently large, so $\lambda^N \to \lambda$, as claimed.

We shall prove Theorem 2 by showing that Theorem 1 is applicable to $\bar{X}^N$ [see (1.16)] and $\bar{Y}^N$, which is defined similarly but with Y^N in place of X^N. These variables have state space R. Lemma 3.2 implies, among other things, that $\{\mathscr{L}(\bar{X}^N), N \geqq 1\}$ is tight. In the remainder of the paper, K and r^N are our generic notations for constants and null sequences, respectively.

Lemma 3.2. *$E((\bar{X}^N)^2) \leqq K$. In addition, $b(\lambda) \geqq 0$, and, if $b(\lambda) = 0$, $E((\bar{X}^N)^2) \to 0$ as $N \to \infty$.*

Proof. We shall omit superscript N's whenever this will not create ambiguity. Clearly

$$E(\bar{Y}^2) = E(\bar{X}^2) + 2E(\bar{X}\bar{\Delta}) + E(\bar{\Delta}^2),$$

where $\bar{\Delta} = \bar{Y} - \bar{X} = (Y - X)\ (\varepsilon/\tau)^{\frac{1}{2}}$. Hence

$$0 = 2E(\bar{X}E(\bar{\Delta} \mid \mathcal{F})) + E(\bar{\Delta}^2). \tag{3.4}$$

We now estimate the terms on the right.

Since $a^N(\lambda^N) = 0$, (1.11) takes the form

$$E(\Delta \mid \mathcal{F}) = \varepsilon(a^N(X) - a^N(\lambda^N)) + e_1. \tag{3.5}$$

But (3.3) implies

$$(X - \lambda^N)(a^N(X) - a^N(\lambda^N)) \leqq -2^{-1}t(X - \lambda^N)^2$$

for N sufficiently large, hence

$$2\bar{X}E(\bar{\Delta} \mid \mathcal{F}) \leqq -\varepsilon t\bar{X}^2 + 2\bar{X}e_1(\varepsilon/\tau)^{\frac{1}{2}}.$$

For any $\alpha > 0$, the second term on the right does not exceed

$$\alpha\bar{X}^2 + \alpha^{-1}e_1^2\varepsilon/\tau.$$

Taking $\alpha = t\varepsilon/2$, we obtain

$$2\bar{X}E(\bar{\Delta} \mid \mathcal{F}) \leqq -2^{-1}\varepsilon t\bar{X}^2 + Ke_1^2/\tau.$$

In view of (1.13),

$$2E(\bar{X}E(\bar{\Delta} \mid \mathcal{F})) \leqq -2^{-1}\varepsilon tE(\bar{X}^2) + r\varepsilon, \tag{3.6}$$

where $r = r^N \to 0$ as $N \to \infty$.

Note that

$$|(a^N)'|_t \leqq K, \tag{3.7}$$

since $|a'|_t < \infty$ and $|(a^N)' - a'|_t \to 0$. Hence it follows from (3.5) that

$$|E(\bar{\Delta} \mid \mathcal{F})| \leqq K\varepsilon|\bar{X}| + |e_1|(\varepsilon/\tau)^{\frac{1}{2}},$$

and, by (1.13),

$$E(E(\bar{\Delta} \mid \mathcal{F})^2 \leqq r\varepsilon E(\bar{X}^2) + r\varepsilon. \tag{3.8}$$

Next,

$$\begin{aligned} E(\bar{\Delta}^2 \mid \mathcal{F}) &= \operatorname{var}(\bar{\Delta} \mid \mathcal{F} + E(\bar{\Delta} \mid \mathcal{F})^2 \\ &= \varepsilon b^N(X) + e_2\varepsilon\tau^{-1} + E(\bar{\Delta} \mid \mathcal{F})^2, \end{aligned} \tag{3.9}$$

by (1.12). Thus, by (1.14), (3.8) and $|b^N - b|_t \to 0$,

$$E(\bar{\Delta}^2) \leqq \varepsilon E(b(X)) + r\varepsilon + r\varepsilon E(\bar{X}^2). \tag{3.10}$$

It follows from (3.4), (3.6) and (3.10) that

$$\limsup_{N\to\infty} E((\bar{X}^N)^2) \leqq 2|b|_t/t,$$

hence $E((\bar{X}^N)^2)$ is bounded. Consequently $\mathscr{L}(X^N) \to \delta_\lambda$, so that $E(b(X^N)) \to b(\lambda)$. Using this in conjunction with (3.4), (3.6) and (3.10), we obtain

$$\limsup_{N\to\infty} E((\bar{X}^N)^2) \leqq 2b(\lambda)/t,$$

which implies the second statement of the lemma.

Lemma 3.3 presents the required estimates of moments of $\bar{\Delta}$.

Lemma 3.3.

$$E(\bar{\Delta}^N \mid \mathscr{F}^N) = \varepsilon^N a'(\lambda)\bar{X}^N + \bar{e}_1^N, \tag{3.11}$$

$$E((\bar{\Delta}^N)^2 \mid \mathscr{F}^N) = \varepsilon^N b(\lambda) + \bar{e}_2^N, \tag{3.12}$$

where

$$E(|\bar{e}_i^N|) = o(\varepsilon^N), \quad i = 1, 2. \tag{3.13}$$

Also, for every $\delta > 0$,

$$E((\bar{\Delta}^N)^2 \bar{I}_\delta^N) = o(\varepsilon^N), \tag{3.14}$$

where $\bar{I}_\delta^N$ *is the indicator of* $\{|\bar{\Delta}^N| > \delta\}$.

Proof. Rewriting (3.5) in the form

$$E(\Delta \mid \mathscr{F}) = \varepsilon(X - \lambda^N)T^N(X, \lambda^N) + e_1,$$

we see that (3.11) obtains with

$$\bar{e}_1 = \varepsilon \bar{X}[T^N(X, \lambda^N) - T(\lambda, \lambda)] + e_1(\varepsilon/\tau)^{\frac{1}{2}}.$$

It follows from (1.13) that the expected absolute value of the second term on the right is $o(\varepsilon)$. Moreover

$$\begin{aligned} E(|\bar{X}[T^N(X, \lambda^N) - T(\lambda, \lambda)]|) &\leqq E(\bar{X}^2)^{\frac{1}{2}} E(|T^N(X, \lambda^N) - T(\lambda, \lambda)|^2)^{\frac{1}{2}} \\ &\leqq K[|(a^N)' - a'|_I + E(|T(X, \lambda^N) - T(\lambda, \lambda)|^2)^{\frac{1}{2}}], \end{aligned}$$

by Lemma 3.2 and (3.2). The right-hand side converges to 0 as $N \to \infty$, since T is bounded and continuous, and (X^N, λ^N) converges in distribution to (λ, λ). Thus (3.13) holds for $i = 1$. The proof of (3.13) for $i = 2$ is similar but simpler, so we omit it.

The derivation of (3.14) presupposes the elementary inequality

$$\begin{aligned} E(|V + W|^2 I(|V + W| > \delta)) \\ \leqq 4E(V^2 I(|V| > 2^{-1}\delta)) + 4E(W^2 I(|W| > 2^{-1}\delta)), \end{aligned}$$

valid for arbitrary random variables V and W ($I(A)$ is the indicator of A). Applying this inequality to $V = \bar{\Delta} - E(\bar{\Delta}\,|\,\mathcal{F}) = \delta^N(\varepsilon/\tau)^{\frac{1}{2}}$ and $W = E(\bar{\Delta}\,|\,\mathcal{F})$, we obtain

$$E(\bar{\Delta}^2\bar{I}_\delta) \leqq 4\varepsilon\tau^{-1}E((\delta^N)^2\tilde{I}_{2^{-1}\delta}) + 4E(E(\bar{\Delta}\,|\,\mathcal{F})^2)$$
$$\leqq r\varepsilon,$$

by (1.15), (3.8) and Lemma 3.2. This completes the proof of Lemma 3.3.

To finish the proof of Theorem 2, we must distinguish the cases $b(\lambda) = 0$ and $b(\lambda) > 0$. When $b(\lambda) = 0$, Lemma 3.2 implies that $\mathcal{L}(\bar{X}^N)$ converges to the distribution concentrated at 0. This is what the conclusion of Theorem 2 reduces to in this case. When $b(\lambda) > 0$, Theorem 1 is applicable to $(\bar{X}^N, \bar{Y}^N)$. According to Lemma 3.3, the corresponding drift and diffusion functions are $\bar{a}(\bar{x}) = a'(\lambda)\bar{x}$ and $\bar{b}(\bar{x}) \equiv b(\lambda) > 0$, and ε^N corresponds to the sequence τ^N in Theorem 1. Clearly $\{\bar{b}(\bar{X}^N), N \geqq 1\}$ is uniformly integrable, and Lemma 3.2 implies that $\{\bar{a}(\bar{X}^N)\} = \{a'(\lambda)\bar{X}^N\}$ is too. Thus all hypotheses of Theorem 1 are satisfied, and a simple computation shows that the limiting distribution is normal, with mean 0 and variance σ^2, as claimed.

4. Genetic applications

We shall first consider in detail the Wright-Fisher model for a monoecious diploid population. Then we shall comment briefly on two models due to Moran.

A. *The Wright-Fisher model*

Suppose that there are two alleles, A_1 and A_2, in a population of N diploid individuals. Mutations from A_1 to A_2 and from A_2 to A_1 have probabilities $\alpha_1 > 0$ and $\alpha_2 > 0$, respectively, and the genotypes A_1A_1, A_1A_2 and A_2A_2 have fitnesses $1 + \nu_1$, 1, and $1 - \nu_2$. In the Wright-Fisher model, the relative A_1 gene frequency X_n in the nth generation is a Markov process with stationary transition probabilities. Let

$$x^* = \frac{(1+\nu_1)x^2 + x(1-x)}{(1+\nu_1)x^2 + 2x(1-x) + (1-\nu_2)(1-x)^2}$$

and

$$p(x) = (1-\alpha_1)x^* + \alpha_2(1-x^*).$$

Given $X_n = x$, $2NX_{n+1}$ has the binomial distribution with parameters $p(x)$ and $2N$. We assume that the distribution of X_0 (hence of X_n) is the stationary distribution of the process. Theorems 1 and 2 yield limit theorems for this distribution.

To apply Theorems 1 and 2, we introduce sequences α_i^N and ν_i^N that converge to 0 at certain prescribed rates. The quantities $p(x)$ and X_n are superscripted accordingly. The following equations are the basis for our subsequent work:

$$E(\Delta X_n^N \mid X_n^N = x) = p^N(x) - x, \tag{4.1}$$

$$\operatorname{var}(\Delta X_n^N \mid X_n^N = x) = N^{-1} b^N(x), \tag{4.2}$$

where

$$b^N(x) = 2^{-1} p^N(x)(1 - p^N(x)), \tag{4.3}$$

and

$$E[(\Delta X_n^N - E(\Delta X_n^N \mid X_n^N))^4 \mid X_n^N = x] \leqq (2N)^{-2} p^N(x)(1 - p^N(x)). \tag{4.4}$$

We now consider three different cases, corresponding to different magnitudes of α_i and ν_i, *vis à vis* N^{-1}.

Case 1. Suppose that $\alpha_i^N = N^{-1}\beta_i$, where $\beta_i > 0$, and $\nu_i^N = N^{-1}\mu_i$. This is the standard assumption in diffusion approximation of genetic models. In this case, (4.1)–(4.4) yield

$$E(\Delta X_n^N \mid X_n^N = x) = N^{-1} a(x) + O(N^{-2}), \tag{4.5}$$

where

$$a(x) = \beta_2 - (\beta_1 + \beta_2)x + x(1-x)(\mu_1 x + \mu_2(1-x)); \tag{4.6}$$

$$E((\Delta X_n^N)^2 \mid X_n^N = x) = N^{-1} b(x) + O(N^{-2}), \tag{4.7}$$

where

$$b(x) = 2^{-1} x(1-x); \tag{4.8}$$

and

$$E((\Delta X_n^N)^4 \mid X_n^N = x) = O(N^{-2}).$$

The quantities $O(N^{-2})$ in (4.5), (4.7) and (4.9) are of the order of magnitude of N^{-2} uniformly over x. It follows easily that Theorem 1 applies with $I = [0, 1]$ and $\tau^N = N^{-1}$. We thus obtain the well-known result that the asymptotic distribution of X_n^N, as $N \to \infty$, has density

$$f(x) = k e^{2\mu_1 x^2 - 2\mu_2(1-x)^2} x^{4\beta_2 - 1}(1-x)^{4\beta_1 - 1}.$$

Case 2. Suppose that $\alpha_i^N = \varepsilon^N \beta_i > 0$ and $\nu_i^N = \varepsilon^N \mu_i$, where $\varepsilon^N > 0$, $\varepsilon^N \to 0$, and $N\varepsilon^N \to \infty$ as $N \to \infty$. Let $a^N(x) = (p^N(x) - x)/\varepsilon^N$, so that

$$E(\Delta X_n^N \mid X_n^N = x) = \varepsilon^N a^N(x).$$

This equality and (4.2) are exemplars of (1.11) and (1.12), with $\tau^N = N^{-1}$ and $e_i^N = 0$. Moreover (4.4) implies (1.15), and it is easily shown that $a^N \to a$, $(a^N)' \to a'$, and $b^N \to b$, uniformly, where a and b are given by (4.6) and (4.8).

It is not true, in general, that a has a unique zero, λ, in I and that $a'(\lambda) < 0$, but this condition is satisfied if, for example, the fitness of A_1A_2 is at least as great as the average of the fitnesses of A_1A_1 and A_2A_2 (see Norman (1974), Sec. 4). In our present notation, this condition takes the form $\mu_1 \leqq \mu_2$. At any rate, whenever λ is unique and stable Theorem 2 applies, and we conclude that

$$\bar{X}_n^N = (X_n^N - \lambda^N)(N\varepsilon^N)^{\frac{1}{2}} \tag{4.9}$$

is asymptotically normal, with mean 0 and variance $b(\lambda)/2|a'(\lambda)|$. In particular, the distribution of X_n^N concentrates at λ as $N \to \infty$.

If $\nu_i^N = 0$, then $a^N = a$ and $\lambda^N = \lambda$. In general, $\lambda^N - \lambda = O(\varepsilon^N)$, so we can replace λ^N by λ in (4.9) without affecting the limiting distribution, provided that $N(\varepsilon^N)^3 \to 0$. It is precisely to avoid restrictions of this type that the sequence a^N was introduced into Theorem 2. Unfortunately, this device is not helpful in the case of Moran's model for dioecious populations (see (C) below), and we have found it necessary to require $N(\varepsilon^N)^3 \to 0$ in our treatment of Case 2 in that model.

Case 3. Suppose that $\nu_1^N = \nu_2^N = 0$, that $\alpha_2^N = N^{-1}\beta$ where $\beta > 0$, and that $\alpha_1^N \to 0$ sufficiently slowly that $N\alpha_1^N \to \infty$. Thus α_2^N is as in Case 1, and α_1^N is as in Case 2. Since $\alpha_2^N = o(\alpha_1^N)$, we expect X_n^N to concentrate near 0. In fact, $\mathscr{L}(X_n^N(N\alpha_1^N))$ converges, as $N \to \infty$, to the gamma distribution with density

$$f(u) = ku^{4\beta-1}e^{-4u},$$

where

$$k = 4^{4\beta}/\Gamma(4\beta).$$

This can be proved by applying Theorem 1 to $\bar{X}^N = X_n^N/\lambda^N$ and $\bar{Y}^N = X_{n+1}^N/\lambda^N$, where $\lambda^N = \alpha_2^N(\alpha_1^N + \alpha_2^N)^{-1}$. The corresponding drift and diffusion functions are $\bar{a}(\bar{x}) = 1 - \bar{x}$ and $\bar{b}(\bar{x}) = \bar{x}/(2\beta)$ on $\bar{I} = [0, \infty)$. The role of τ^N is played by $\alpha_1^N + \alpha_2^N$. The proof is too similar to that of Theorem 2 to justify presentation here.

B. *A model with overlapping generations*

We now turn our attention to other genetic models. First, we consider Moran's ((1962), p. 133) model for haploid organisms with overlapping generations. Suppose that the population size is $2N$, that mutations from A_1 to A_2 and *vice versa* have probabilities α_1^N and α_2^N, and that the death rates ρ_1^N and ρ_2^N of A_1 and A_2 have ratio $\rho_2^N/\rho_1^N = 1 + \nu^N$. Applying Theorems 1 and 2 to the embedded chain X_n^N, we obtain results similar to those described above for the Wright-Fisher model. Assuming, as we shall, that $\nu^N \to 0$ as $N \to \infty$, these results apply without change to the stationary distribution of the continuous time process postulated by the model. Karlin and McGregor (1964) described diffusion approximations to the transient behavior of this model when there is no selection ($\rho_1 = \rho_2$). Our Cases 1, 2 and 3 correspond to their cases (ii), (i) and (iv), respectively.

Verification of the hypotheses of Theorems 1 and 2 is similar to the analogous problem for the Wright-Fisher model, though somewhat simpler, so we limit ourselves to describing the dependence of parameters on N in each case. In Case 1, we take $\alpha_i = N^{-1}\beta_i > 0$ and $\nu^N = N^{-1}\mu$ just as before, but now $\tau^N = N^{-2}$. The same choice of τ^N is appropriate for Case 2. In that case, we may take $\alpha_i^N = \theta^N\beta_i > 0$, $\nu^N = \theta^N\mu$, and $\varepsilon^N = N^{-1}\theta^N$, where $\theta^N \to 0$ and $N\theta^N \to \infty$. Alternatively, we may take $\alpha_i^N = \alpha_i > 0$, $\nu^N = 0$, and $\varepsilon^N = N^{-1}$. Since there is no dominance, the function a has a unique root, λ, in I, and $a'(\lambda) < 0$. In Case 3, take $\alpha_2^N = N^{-1}\beta$, $\alpha_1^N \to 0$, $N\alpha_1^N \to \infty$, and $\nu^N = 0$ as before, but take $\tau^N = N^{-1}(\alpha_1^N + \alpha_2^N)$ in applying Theorem 1.

C. *A model for dioecious populations*

Finally, we consider Moran's (1958b) model for a dioecious population with non-overlapping generations. (Theorems 1 and 2 are equally applicable to Moran's (1958a) dioecious overlapping generation model.) There are N individuals, of whom N_1 are male and N_2 are female. The parameters α_i and ν_i are as in the Wright-Fisher model. There is, in addition, a non-random-mating parameter f that does not vary as $N \to \infty$. The relative frequencies of A_1A_1 and A_2A_2 among males in the nth generation are K_n and L_n; the comparable relative frequencies for females are R_n and S_n. The process $Y_n = (K_n, L_n, R_n, S_n)$ is Markovian, but the average

$$X_n = 2^{-1} + 4^{-1}(K_n - L_n + R_n - S_n)$$

of the A_1 gene frequencies in the two sexes is non-Markovian. Let Y_0 have its stationary distribution, and let $\mathcal{F}_n$ be the σ-field generated by Y_n.

Assume that $\alpha_i^N = \varepsilon^N\beta_i > 0$ and $\nu_i^N = \varepsilon^N\mu_i$, where $\varepsilon^N > 0$ and $\varepsilon^N \to 0$ as $N \to \infty$. Let $\tau^N = 4^{-1}(N_1^{-1} + N_2^{-1})$. It can be shown that

$$E(\Delta X_n^N \mid \mathcal{F}_n^N) = \varepsilon^N a(X_n^N) + e_{1,n}^N$$

and

$$\operatorname{var}(\Delta X_n^N \mid \mathcal{F}_n^N) = \tau^N b(X_n^N) + e_{2,n}^N,$$

where

$$a(x) = \beta_2 - (\beta_1 + \beta_2)x + x(1-x)[(\mu_1 + \mu_2 f)x + (\mu_2 + \mu_1 f)(1-x)],$$

$$b(x) = 2^{-1}(1+f)x(1-x),$$

(4.10)
$$E((e_{1,n}^N)^2)^{\frac{1}{2}} \leqq K\varepsilon^N(\varepsilon^N + (\tau^N)^{\frac{1}{2}}),$$

and

$$E(|e_{2,n}^N|) \leqq K\tau^N(\tau^N + \varepsilon^N).$$

Also,

$$E(|\Delta X_n^N - E(\Delta X_n^N \mid \mathcal{F}_n^N)|^3) \leqq K(\tau^N)^{\frac{3}{2}}.$$

It follows easily from these estimates that Theorem 1 applies to X_n^N if $\varepsilon^N = \tau^N$ (Case 1). This is the case considered by Moran (1958b). Limits are taken as $N_1 \to \infty$ and $N_2 \to \infty$.

Suppose now that $\tau^N = o(\varepsilon^N)$, that a has a unique zero, λ, and that λ is stable (e.g., $\mu_1 \leqq \mu_2$). Then all hypotheses of Theorem 2 are satisfied, with $a^N = a$ (hence $\lambda^N = \lambda$) and $b^N = b$, except perhaps (1.13). To obtain (1.13) from (4.10), it is necessary to impose the additional restriction $(\varepsilon^N)^3 = o(\tau^N)$.

In Case 3 of our treatment of the Wright-Fisher model, it was assumed that $\alpha_2^N = N^{-1}\beta$, $\alpha_1^N \to 0$, $N\alpha_1^N \to \infty$, and $\nu_i^N = 0$. We have not yet considered an analogous case for Moran's model.

References

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

KARLIN, S., AND MCGREGOR, J. (1964) On some stochastic models in genetics. *Stochastic Models in Medicine and Biology*, ed. J. Gurland, pp. 245–271. University of Wisconsin Press, Madison.

MORAN, P. A. P. (1958a) A general theory of the distribution of gene frequencies I. Overlapping generations. *Proc. Roy. Soc. London* B **149**, 102–112.

MORAN, P. A. P. (1958b) A general theory of the distribution of gene frequencies II. Non-overlapping generations. *Proc. Roy. Soc. London* B **149**, 113–116.

MORAN, P. A. P. (1962) *The Statistical Processes of Evolutionary Theory.* Clarendon Press, Oxford.

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

NORMAN, M. F. (1974) A central limit theorem for Markov processes that move by small steps. *Ann. Probability* **2**, 1065–1074.

NORMAN, M. F. (1975a) Diffusion approximation of non-Markovian processes. *Ann. Probability* **3**, 358–364.

NORMAN, M. F. (1975b) Approximation of stochastic processes by Gaussian diffusions, and applications to Wright-Fisher genetic models. *SIAM J. Appl. Math.* **29**, 225–242.

WRIGHT, S. (1938) The distribution of gene frequencies in populations of polyploids. *Proc. Nat. Acad. Sci., Wash.* **24**, 372–377.