SIAM J. APPL. MATH.
Vol. 29, No. 2, September 1975

APPROXIMATION OF STOCHASTIC PROCESSES BY GAUSSIAN DIFFUSIONS, AND APPLICATIONS TO WRIGHT-FISHER GENETIC MODELS*

M. FRANK NORMAN†

Abstract. For each $N \geqq 1$, let $\{X_n^N, n \geqq 0\}$ be a discrete-time stochastic process, and let $\Delta X_n^N = X_{n+1}^N - X_n^N$. Suppose that $E(\Delta X_n^N | X_n^N) = O(\varepsilon^N)$ and $\operatorname{var}(\Delta X_n^N | X_n^N) = O(\tau^N)$, where $\varepsilon^N \to 0$ and $\tau^N/\varepsilon^N \to 0$ as $N \to \infty$. Conditions are given under which there are constants γ_n^N such that $Z_n^N = (X_n^N - \gamma_n^N)(\varepsilon^N/\tau^N)^{1/2}$ can be approximated by a Gaussian diffusion when N is large. It is shown that these conditions are satisfied by the Wright–Fisher models for fluctuations in gene frequency under the influence of mutation, selection and random drift. For these models, N is the population size and the constants γ_n^N are the gene frequencies specified by Haldane's deterministic theory of evolution.

1. Introduction. Let $\{X_n, n \geqq 0\}$ be a discrete-time stochastic process. This paper deals with an approximation to the distribution of X_n by a distribution associated with a continuous-time Gaussian process. More precisely, we have a sequence of processes, $\{X_n^N, n \geqq 0\}$, $N \geqq 1$. The dependence on N is such that $E(\Delta X_n^N | X_n^N) = O(\varepsilon^N)$ and $\operatorname{var}(\Delta X_n^N | X_n^N) = O(\tau^N)$, where $\Delta X_n^N = X_{n+1}^N - X_n^N$, and ε^N and τ^N are positive sequences that converge to 0 as $N \to \infty$. The case $\varepsilon^N = \tau^N$ has been studied rather thoroughly. This paper deals with the case $\tau^N = o(\varepsilon^N)$. We reduce the latter case to the former by centering and scaling to form $Z_n^N = (X_n^N - \gamma_n^N)(\varepsilon^N/\tau^N)^{1/2}$. Under conditions given in § 2, $E(\Delta Z_n^N | Z_n^N) = O(\varepsilon^N)$ and $E((\Delta Z_n^N)^2 | Z_n^N) = O(\varepsilon^N)$. Moreover, there is a sequence, $\{z^N(t), t \geqq 0\}$, $N \geqq 1$, of continuous-time Gaussian processes such that $\mathscr{L}(Z_n^N) \sim \mathscr{L}(z^N(n\varepsilon^N))$ as $N \to \infty$. Here $\mathscr{L}(Y)$ is the distribution of the random variable Y. The process $z^N(t)$ is a diffusion, i.e., a Markov process with continuous sample paths.

A result of this type was derived heuristically by Feller [5, § 9] in the context of the Wright–Fisher model for fluctuations in gene frequency in a finite population. We now describe this model. The population is monoecious (each individual combines both sexes) and diploid (chromosomes occur in pairs), and has non-overlapping generations. There are two alleles, A_1 and A_2, at the locus in question, hence there are three genotypes, A_1A_1, A_1A_2 and A_2A_2. The population consists of N individuals, hence $2N$ genes. Let $X_n = x$ be the relative frequency of A_1 genes among adults of generation n. Assuming random mating, the genotype probabilities at the time of conception of the next generation are x^2, $2x(1-x)$ and $(1-x)^2$ for A_1A_1, A_1A_2 and A_2A_2. If these genotypes have relative fitnesses $1 + \nu_1$, 1 and $1 - \nu_2$, the genotype frequencies after selection are proportional to $(1+\nu_1)x^2$, $2x(1-x)$ and $(1-\nu_2)(1-x)^2$, and the expected A_1 gene frequency after selection is

$$x^* = \frac{(1+\nu_1)x^2 + x(1-x)}{(1+\nu_1)x^2 + 2x(1-x) + (1-\nu_2)(1-x)^2}. \tag{1.1}$$

If an A_1 gene mutates to A_2 with probability α_1 and an A_2 mutates to A_1 with probability α_2, then the expected A_1 gene frequency in adults is

* Received by the editors April 26, 1974, and in revised form January 31, 1975.

† Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania 19174.

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

The model allows for random variation in gene frequency by assuming, in effect, that the $2N$ genes in adults of generation $n + 1$ constitute a random sample from an infinite population of genes, of which a proportion $\gamma(x)$ are of type A_1. Thus the distribution of A_1 gene frequency, X_{n+1}, in generation $n + 1$ is binomial with parameters $\gamma(x)$ and $2N$, i.e.,

$$P(X_{n+1} = j/2N | X_n = x) = \binom{2N}{j} \gamma(x)^j(1 - \gamma(x))^{2N-j}.$$

The process $X_0, X_1, \cdots$ is Markovian.

We mention that Feller [5] considered only the special case $\nu_1 = \nu_2 = 0$. Our description of the more general model follows [4, § 4.8].

It is easy to see that

$$E(\Delta X_n | X_n = x) = \gamma(x) - x = O(\varepsilon),$$

where $\varepsilon = \max(|\nu_1|, |\nu_2|, \alpha_1, \alpha_2)$. Assuming $\varepsilon > 0$, we let

(1.3) $$w(x) = (\gamma(x) - x)/\varepsilon,$$

so that

(1.4) $$E(\Delta X_n | X_n = x) = \varepsilon w(x).$$

Similarly, letting $\tau^N = 1/N$ and

(1.5) $$s(x) = \gamma(x)(1 - \gamma(x))/2,$$

we see that

(1.6) $$\operatorname{var}(\Delta X_n | X_n) = \operatorname{var}(X_{n+1} | X_n) = \tau^N s(x).$$

We are interested in the case where the population size, N, is large. Moreover, in most applications, one expects the selection differentials, ν_i, and mutation rates, α_j, to be small, so that ε is small. Thus we wish to study the asymptotic behavior of X_n when $N \to \infty$ and $\varepsilon \to 0$. To approach this problem within the framework described in the opening paragraph of the paper, we must embed X_n in a sequence of processes, X_n^N. This is accomplished by the slightly artificial device of choosing arbitrary sequences, ν_i^N and α_j^N, of parameters converging to zero, and an arbitrary sequence of distributions of X_0. The corresponding quantities X_n, ε, γ, w and s are then denoted X_n^N, ε^N, γ^N, w^N and s^N.

The standard diffusion approximation of the Wright–Fisher model is applicable when ν_i^N, α_j^N, and thus ε^N are proportional to $1/N$ (see [7] and [11, § 18.1]). In this paper, we shall be concerned with the case $\tau^N = o(\varepsilon^N)$ or $N\varepsilon^N \to \infty$. Thus the diffusion approximation developed in subsequent sections should be more appropriate than the standard one when $N\varepsilon$ is relatively large.

2. The main results. Our main theorems pertain to sequences of stochastic processes satisfying certain rather general hypotheses, not just to the Wright–Fisher model. We are now ready to present these hypotheses. They are numerous, and some appear complex, but most of them are very easy to verify in the type of application considered in this paper.

Assumptions. The random variable X_n^N takes on values in a closed, but not necessarily bounded, real interval I. It is measurable with respect to a σ-field $\mathscr{F}_n^N$, and $\mathscr{F}_{n+1}^N \supset \mathscr{F}_n^N$. We suppose that

$$E(\Delta X_n^N | \mathscr{F}_n^N) = \varepsilon^N w^N(X_n^N) + e_{1,n}^N \tag{2.1}$$

and

$$\operatorname{var}(\Delta X_n^N | \mathscr{F}_n^N) = \tau^N s^N(X_n^N) + e_{2,n}^N, \tag{2.2}$$

where $\varepsilon^N > 0$, $\tau^N > 0$, $\varepsilon^N \to 0$ and $(\tau^N/\varepsilon^N) \to 0$ as $N \to \infty$. For every $N \geqq 1$, w^N and s^N are real-valued functions defined throughout I. Let

$$|w|_\infty = \sup_{x \in I} |w(x)|$$

and

$$m(w) = \sup_{x \neq y} |w(y) - w(x)|/|y - x|.$$

The functions w^N, $Dw^N = dw^N/dx$, and s^N are uniformly bounded, i.e.,

$$W = \sup_{N \geqq 1} |w^N|_\infty < \infty, \tag{2.3}$$

$$W' = \sup_{N \geqq 1} |Dw^N|_\infty < \infty, \tag{2.4}$$

$$S = \sup_{N \geqq 1} |s^N|_\infty < \infty; \tag{2.5}$$

and Dw^N and s^N are uniformly Lipschitz, i.e.,

$$W'' = \sup_{N \geqq 1} m(Dw^N) < \infty, \tag{2.6}$$

$$S' = \sup_{N \geqq 1} m(s^N) < \infty. \tag{2.7}$$

Moreover, $s^N(x) \geqq 0$ for all $x \in I$, and

$$\gamma^N(x) = x + \varepsilon^N w^N(x) \tag{2.8}$$

maps I into I. Concerning the error terms $e_{i,n}^N$ in (2.1) and (2.2), we assume that, for some $0 < T < \infty$,

$$\sum_{n < [T/\varepsilon^N]} E(|e_{1,n}^N|^3)^{1/3} \leqq r^N(\tau^N/\varepsilon^N)^{1/2} \tag{2.9}$$

and

$$\sum_{n < [T/\varepsilon^N]} E(|e_{2,n}^N|) \leqq r^N(\tau^N/\varepsilon^N), \tag{2.10}$$

where $[y]$ is the largest integer that does not exceed y and r^N is our generic notation for a positive sequence that converges to 0 as $N \to \infty$. Finally, letting

$$e_{3,n}^N = \Delta X_n^N - E(\Delta X_n^N | \mathscr{F}_n^N) = X_{n+1}^N - E(X_{n+1}^N | \mathscr{F}_n^N), \tag{2.11}$$

we assume that

$$\sum_{n<[T/\varepsilon^N]} E(|e_{3,n}^N|^3) \leqq K(\tau^N)^{3/2}/\varepsilon^N, \tag{2.12}$$

where K is our generic notation for a quantity that does not depend on N. This is the last of our assumptions.

The σ-field $\mathscr{F}_n^N$ was introduced to permit us to deal with non-Markovian processes, like the one described in the next section. We might, for example, let $\mathscr{F}_n^N$ be the smallest σ-field with respect to which $X_0^N, \cdots, X_n^N$ are measurable, so that conditioning on $\mathscr{F}_n^N$ is equivalent to conditioning on $X_0^N, \cdots, X_n^N$. In Markovian cases, like the model considered in § 1, this is tantamount to conditioning on X_n^N.

Conditions (2.9), (2.10) and (2.12) are a bit weaker than the slightly simpler conditions

$$E(|e_{1,n}^N|^3)^{1/3} \leqq r^N(\varepsilon^N\tau^N)^{1/2}, \tag{2.9'}$$

$$E(|e_{2,n}^N|) \leqq r^N\tau^N, \tag{2.10'}$$

$$E(|e_{3,n}^N|^3)^{1/3} \leqq K(\tau^N)^{1/2} \tag{2.12'}$$

for $n < [T/\varepsilon^N]$. In the Wright–Fisher model, $e_{1,n}^N = 0$ and $e_{2,n}^N = 0$ [see (1.4) and (1.6)], so (2.9′) and (2.10′) hold, and (2.12′) follows from

$$E(|e_{3,n}^N|^3|X_n^N) \leqq KN^{-3/2},$$

which reflects the fact that, given X_n^N, $e_{3,n}^N$ has a binomial distribution, centered at its expectation.

The functions w^N and s^N in the Wright–Fisher model are defined throughout $I = [0, 1]$ by formulas (1.1), (1.2), (1.3) and (1.5). The boundedness conditions (2.3)–(2.7) are easily verified, assuming $\varepsilon^N \leqq \frac{1}{2}$, as is the fact that γ^N maps I into I. Thus the Wright–Fisher model satisfies all our assumptions.

The main consequences of our assumptions are Theorems 1 and 2 below. Let $\{\gamma_n^N, n \geqq 0\}$ be defined by the difference equation $\gamma_{n+1}^N = \gamma^N(\gamma_n^N)$ or

$$\Delta\gamma_n^N = \varepsilon^N w^N(\gamma_n^N), \tag{2.13}$$

and the initial condition $\gamma_0^N = X_0^N$. Let

$$Z_n^N = (X_n^N - \gamma_n^N)(\varepsilon^N/\tau^N)^{1/2}.$$

THEOREM 1. $E(\max_{n\leqq T/\varepsilon^N}|Z_n^N|^3) \leqq K$.

Let $\mathscr{N}(\mu, \sigma^2)$ be the normal distribution with mean μ and variance σ^2.

THEOREM 2. *For any $x \in I$, the differential equation*

$$\frac{df^N(t, x)}{dt} = w^N(f^N(t, x)) \tag{2.14}$$

has a unique solution with $f^N(0, x) = x$. Let

$$B^N(t, x) = \exp\left[\int_0^t Dw^N(f^N(\xi, x))\, d\xi\right] \tag{2.15}$$

and

$$g^N(t, x) = \int_0^t \exp\left[2\int_u^t Dw^N(f^N(\xi, x))\,d\xi\right] s^N(f^N(u, x))\,du. \tag{2.16}$$

Suppose that $\operatorname{var}(X_0^N) = 0$, *and let* $X_0^N = x_0^N$ *and* $f^N(t) = f^N(t, x_0^N)$. *There is a diffusion* $\{z^N(t), t \geqq 0\}$ *with transition probability*

$$\begin{aligned} P^N(t, z, t + \delta, \cdot) &= \mathscr{L}(z^N(t + \delta) \mid z^N(t) = z) \\ &= \mathscr{N}(B^N(\delta, f^N(t))z, g^N(\delta, f^N(t))) \end{aligned} \tag{2.17}$$

and $z^N(0) = 0$ *almost surely. For any* $J \geqq 1$, *we have*

$$\mathscr{L}(Z_{n_1}^N, \cdots, Z_{n_J}^N) \sim \mathscr{L}(z^N(n_1\varepsilon^N), \cdots, z^N(n_J\varepsilon^N)) \tag{2.18}$$

as $N \to \infty$, *uniformly over* $0 \leqq n_1 \leqq n_2 \leqq \cdots n_J \leqq T/\varepsilon^N$. *By this we mean that, for any bounded continuous function* F *of* J *real variables,*

$$\max_{n_1 \leqq n_2 \leqq \cdots \leqq T/\varepsilon^N} |E[F(Z_{n_1}^N, \cdots, Z_{n_J}^N)] - E[F(z^N(n_1\varepsilon^N), \cdots, z^N(n_J\varepsilon^N))]| \to 0 \tag{2.19}$$

as $N \to \infty$. *In particular,*

$$\mathscr{L}(Z_n^N) \sim \mathscr{N}(0, g^N(n\varepsilon^N, x_0^N)), \tag{2.20}$$

uniformly over $n \leqq T/\varepsilon^N$. *Furthermore,*

$$\max_{n \leqq T/\varepsilon^N} |E(Z_n^N)| \to 0 \tag{2.21}$$

and

$$\max_{n \leqq T/\varepsilon^N} |\operatorname{var}(Z_n^N) - g^N(n\varepsilon^N, x_0^N)| \to 0 \tag{2.22}$$

as $N \to \infty$.

Theorems 1 and 2 are proved in §§ 4–7.

A result like (2.20) was obtained previously [11, Thm. 8.1.1] for Markov processes, under the rather restrictive condition $\tau^N = (\varepsilon^N)^2$. This equality is satisfied in a wide variety of mathematical learning models. Conditions for (2.20) to hold uniformly over $n \geqq 0$ have been given in the context of this equality [12]. These conditions can be generalized to $\tau^N = o(\varepsilon^N)$.

Let $Z^N(t)$ be the random polygonal line with vertices $Z^N(n\varepsilon^N) = Z_n^N$. It is shown in § 8 that under assumptions slightly stronger than those of Theorems 1 and 2, the distribution of the process $\{Z^N(t), t \leqq T\}$ is asymptotically equivalent to that of the process $\{z^N(t), t \leqq T\}$. Generalizations of this result and of Theorems 1 and 2 to multidimensional processes X_n^N are described in § 9. These generalizations apply to the Wright–Fisher model with multiple alleles.

Returning to the Wright–Fisher model with two alleles, we note that γ_n^N is the gene frequency specified by the classical deterministic theory of evolution developed by Haldane (see [2, Chap. 4]). Theorem 1 implies that X_n^N tends to be close to γ_n^N when ε^N is small and $N\varepsilon^N$ is large. In other words, the Wright–Fisher theory degenerates into the Haldane theory as $\varepsilon^N \to 0$ and $N\varepsilon^N \to \infty$. Theorem 2 specifies the asymptotic distribution of the deviation of X_n^N from γ_n^N.

It is worthwhile to mention certain approximations to γ_n^N that appear frequently in the genetic literature. First, comparison of the difference equation

(2.13) with the differential equation (2.14) shows that $\varphi_n^N = f^N(n\varepsilon^N, X_0^N)$ is a natural approximation to γ_n^N. Second, it follows from (1.1)–(1.3) that $|w^N - \tilde{w}^N|_\infty \leqq K\varepsilon^N$, where

$$\varepsilon^N \tilde{w}^N(x) = \alpha_2 - (\alpha_1 + \alpha_2)x + x(1 - x)(v_1 x + v_2(1 - x)).$$

This suggests consideration of the quantities $\tilde{\gamma}_n^N$ and $\tilde{\varphi}_n^N$, obtained by replacing w^N by $\tilde{w}^N$ in the definitions of γ_n^N and φ_n^N. Lemma 2 in § 5 shows that $\varphi_n^N - \gamma_n^N = O(\varepsilon^N)$, and similar arguments yield the same estimate of $\tilde{\gamma}_n^N - \gamma_n^N$ and $\tilde{\varphi}_n^N - \gamma_n^N$. It follows that the centering constant γ_n^N can be replaced by φ_n^N, $\tilde{\gamma}_n^N$ or $\tilde{\varphi}_n^N$ in the definition of Z_n^N without altering the validity of Theorems 1 and 2, provided that $(\varepsilon^N)^3 = o(\tau^N)$.

In closing this section, we note that we have treated parameter sequences α_i^N and v_j^N in the Wright–Fisher model that are restricted only by $\varepsilon^N \to 0$ and $N\varepsilon^N \to \infty$. It follows that we may dispense with the sequences v_i^N and α_j^N if we replace the limiting condition $N \to \infty$ in Theorem 2 by $\varepsilon \to 0$ and $N\varepsilon \to \infty$. Since Theorem 2 places no constraint on the sequence of initial values x_0^N, the maxima in (2.19), (2.21) and (2.22) may be taken over all possible values $j/2N$ of this parameter. Inspection of the proof of Theorem 1 shows that $E(\max_{n \leqq T/\varepsilon} |Z_n|^3)$ is bounded over all $N \geqq 1$, $\mathscr{L}(X_0)$ and $\varepsilon \leqq \frac{1}{2}$.

3. Populations with distinct sexes. Moran [9] described a Wright–Fisher type model for dioecious populations, to which the theory of the preceding section can be applied. We shall only mention the aspects of this model that are pertinent to our analysis.

The population consists of N individuals, N_1 males and N_2 females. The relative frequencies of genotypes A_1A_1, A_1A_2 and A_2A_2 among males in generation n are K_n, $1 - K_n - L_n$ and L_n; among females, the frequencies are R_n, $1 - R_n - S_n$ and S_n. We will apply Theorems 1 and 2 to the arithmetic average of the relative A_1 gene frequencies in the two sexes, which is

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

As in the monoecious Wright–Fisher model, the genotypes A_1A_1, A_1A_2 and A_2A_2 have fitnesses $1 + v_1$, 1 and $1 - v_2$, mutations from A_1 to A_2 and from A_2 to A_1 have probabilities α_1 and α_2, and $\varepsilon = \max(\alpha_1, \alpha_2, |v_1|, |v_2|)$. There is, in addition, a nonrandom-mating parameter $f \in [0, 1]$ that remains fixed as $\varepsilon \to 0$ in our limiting process. Superscript N's are omitted throughout this section.

Let $\mathscr{F}_n$ be the σ-field generated by $Y_n, Y_{n-1}, \cdots, Y_0$, where Y_n is the Markov process $Y_n = (K_n, L_n, R_n, S_n)$. Let

$$\varepsilon w(x) = \alpha_2 - (\alpha_1 + \alpha_2)x + x(1 - x)[(v_1 + v_2 f)x + (v_2 + v_1 f)(1 - x)],$$

$s(x) = 2^{-1}(1 + f)x(1 - x)$ and $\tau = 4^{-1}(N_1^{-1} + N_2^{-1})$. It can be shown that the error terms $e_{i,n}$ in (2.1), (2.2) and (2.11) admit the following estimates, valid for $\varepsilon \leqq \frac{1}{2}$:

$$E(|e_{1,n}|^3)^{1/3} \leqq \begin{cases} K\varepsilon(\varepsilon + \tau^{1/2}) & \text{if } n \geqq 2, \\ K\varepsilon & \text{if } n \geqq 0; \end{cases}$$

$$E(|e_{2,n}|) \leqq \begin{cases} K\tau(\tau + \varepsilon) & \text{if } n \geqq 1, \\ K\tau & \text{if } n \geqq 0, \end{cases}$$

and $E(|e_{3,n}|^3)^{1/3} \leqq K\tau^{1/2}$ if $n \geqq 0$. Clearly, (2.10) and (2.12) present no difficulties, and (2.9) obtains if $\varepsilon^3 = o(\tau)$. Conditions (2.3)–(2.7) are trivial, and $\gamma(x) = x + \varepsilon w(x)$ maps I into I if $\varepsilon \leqq \frac{1}{3}$.

Thus the assumptions of § 2 are all satisfied. It follows that the limits in Theorem 2 obtain, uniformly over x_0, as $\varepsilon \to 0, (\tau/\varepsilon) \to 0$, and $(\varepsilon^3/\tau) \to 0$.

4. Proof of Theorem 1. The proof of Theorem 1 is based on the following lemma.

LEMMA 1. *Let* $\{Y_n, n \geqq 0\}$ *be a martingale with* $Y_0 = 0$. *Then*

$$E(|Y_n|^3) \leqq 6n^{1/2} \sum_{j=0}^{n-1} E(|\Delta Y_j|^3)$$

for all $n \geqq 0$.

The proof that follows is simpler than the proof of a similar inequality in [3], and yields a smaller multiplicative constant on the right.

Proof. Let $F(u) = |y + ux|^3$. By Taylor's theorem,

$$R(y, x) = |y + x|^3 - |y|^3 - 3xy|y| = \int_0^1 (1 - u)F''(u)\,du.$$

But

$$F''(u) = 6|y + ux|x^2 \leqq 6(|y| + u|x|)x^2,$$

so $|R(y, x)| \leqq 3|y|x^2 + |x|^3$.

Since Y_n is a martingale, $E(Y_j|Y_j|\Delta Y_j) = 0$. Thus, for $n \geqq 1$,

$$\begin{aligned} E(|Y_n|^3) & \\ &= \sum_{j=0}^{n-1} E(\Delta|Y_j|^3) = \sum_{j=0}^{n-1} E(R(Y_j, \Delta Y_j)) \\ &\leqq 3\sum_{j=0}^{n-1} E(|Y_j|(\Delta Y_j)^2) + \sum_{j=0}^{n-1} E(|\Delta Y_j|^3) \\ &\leqq 3\left(\sum_{j=0}^{n-1} E(|Y_j|^3)\right)^{1/3} a_n^{2/3} + a_n \end{aligned}$$

by Hölder's inequality, where $a_n = \sum_{j=0}^{n-1} E(|\Delta Y_j|^3)$. Let $b_n = E(|Y_n|^3)/a_n$ if $a_n > 0$, and $b_n = 0$ if $a_n = 0$ (e.g., $b_0 = 0$). If $a_n > 0$,

$$b_n \leqq 3\left(\sum_{j=0}^{n-1} E(|Y_j|^3)/a_n\right)^{1/3} + 1,$$

hence

$$b_n \leqq 3\left(\sum_{j=0}^{n-1} b_j\right)^{1/3} + 1. \tag{4.1}$$

If $a_n = 0$, this inequality holds trivially.

Suppose, inductively, that $n \geqq 1$ and $b_j \leqq 6j^{1/2}$ for $j \leqq n-1$. Then (4.1) implies that

$$\begin{aligned} b_n &\leqq 3 \times 6^{1/3} \times \left(\sum_{j=0}^{n-1} j^{1/2}\right)^{1/3} + 1 \\ &\leqq 3 \times 6^{1/3} \times \left(\int_0^n x^{1/2}\,dx\right)^{1/3} + 1 \\ &= 3 \times 4^{1/3} \times n^{1/2} + 1 \\ &\leqq (3 \times 4^{1/3} + 1)n^{1/2} \leqq 6n^{1/2}. \end{aligned}$$

This completes the proof of Lemma 1.

The argument that follows was suggested by Kurtz's [8] approach to related problems for continuous-time processes. Throughout the paper, superscript N's are usually suppressed in proofs.

4.1. Completion of proof of Theorem 1. Let Y_n be the martingale

$$Y_n = X_n - X_0 - \sum_{j=0}^{n-1} E(\Delta X_j | \mathscr{F}_j) = \sum_{j=0}^{n-1} e_{3,j}.$$

By Lemma 1,

$$E(|Y_n|^3) \leqq 6n^{1/2} \sum_{j=0}^{n-1} E(|e_{3,j}|^3).$$

Taking $n = k = [T/\varepsilon]$, we find that (2.12) yields $E(|Y_k|^3) \leqq K(\tau/\varepsilon)^{3/2}$. In combination with a standard martingale inequality, this implies

$$E(|Y^*|^3) \leqq K(\tau/\varepsilon)^{3/2}, \tag{4.2}$$

where $Y^* = \max_{n \leq k} |Y_n|$. Now

$$X_n = X_0 + \sum_{j=0}^{n-1} E(\Delta X_j | \mathscr{F}_j) + Y_n$$

and

$$\gamma_n = X_0 + \sum_{j=0}^{n-1} \Delta\gamma_n.$$

Subtracting, we obtain

$$\begin{aligned} X_n - \gamma_n &= \sum_{j=0}^{n-1} (E(\Delta X_j | \mathscr{F}_j) - \Delta\gamma_n) + Y_n \\ &= \varepsilon \sum_{j=0}^{n-1} (w(X_j) - w(\gamma_j)) + \sum_{j=0}^{n-1} e_{1,j} + Y_n \end{aligned}$$

by (2.1) and (2.13). In view of (2.4),

$$|X_n - \gamma_n| \leqq \varepsilon W' \sum_{j=0}^{n-1} |X_j - \gamma_j| + \sum_{j=0}^{k-1} |e_{1,j}| + Y^*$$

for $n \leqq k$. It follows that

$$|X_n - \gamma_n| \leqq e^{W'\varepsilon n}\left(Y^* + \sum_{j=0}^{k-1} |e_{1,j}|\right)$$

for $n \leqq k$; hence

$$\max_{n \leqq k} |X_n - \gamma_n| \leqq e^{W'T}\left(Y^* + \sum_{j=0}^{k-1} |e_{1,j}|\right).$$

Therefore

$$E(\max_{n \leqq k} |X_n - \gamma_n|^3)^{1/3} \leqq K\left(E(Y^{*3})^{1/3} + \sum_{j=0}^{k-1} E(|e_{1,j}|^3)^{1/3}\right) \leqq K(\tau/\varepsilon)^{1/2}$$

by (4.2) and (2.9).

5. Moments of ΔZ_n^N. We have assumed that $\gamma^N(x) = x + \varepsilon^N w^N(x) \in I$ for each $x \in I$. But I is an interval, so $x + hw^N(x) \in I$ for all $0 \leqq h \leqq \varepsilon^N$. Since $|Dw^N|_\infty < \infty$, it follows that there is one and only one function $f^N(\cdot, x)$ with values in I that satisfies (2.14) and $f^N(0, x) = x$ [13, Thm. 4]. Let $\varphi_n^N = f^N(n\varepsilon^N, X_0^N)$.

LEMMA 2. *For* $n \leqq T/\varepsilon^N$, $|\varphi_n^N - \gamma_n^N| \leqq K\varepsilon^N$.

Proof. By Taylor's theorem,

$$\varphi_n = \varphi_{n-1} + \varepsilon w(\varphi_{n-1}) + c_n,$$

where

$$|c_n| \leqq \varepsilon^2 m(w)|w|_\infty/2 \leqq WW'\varepsilon^2/2$$

by (2.3) and (2.4). Clearly,

$$\varphi_n - \gamma_n = \varphi_{n-1} - \gamma_{n-1} + \varepsilon(w(\varphi_{n-1}) - w(\gamma_{n-1})) + c_n.$$

Therefore

$$\begin{aligned}|\varphi_n - \gamma_n| &\leqq |\varphi_{n-1} - \gamma_{n-1}|(1 + W'\varepsilon) + WW'\varepsilon^2/2 \qquad \text{(by (2.4))}\\ &\leqq \frac{(1 + W'\varepsilon)^n - 1}{W'\varepsilon} WW'\varepsilon^2/2\\ &\leqq (e^{W'\varepsilon n} - 1)W\varepsilon/2 \leqq K\varepsilon\end{aligned}$$

for $n\varepsilon \leqq T$.

The next step in the proof of Theorem 2 is the derivation of suitable expressions for moments of ΔZ_n^N, conditional on $\mathscr{F}_n^N$.

LEMMA 3.

$$E(\Delta Z_n^N | \mathscr{F}_n^N) = \varepsilon^N Dw^N(\varphi_n^N) Z_n^N + h_{1,n}^N, \tag{5.1}$$

$$E((\Delta Z_n^N)^2 | \mathscr{F}_n^N) = \varepsilon^N s^N(\varphi_n^N) + h_{2,n}^N, \tag{5.2}$$

$$E(|\Delta Z_n^N|^3 | \mathscr{F}_n^N) = h_{3,n}^N, \tag{5.3}$$

where, for $i = 1, 2, 3$,

$$\sum_{n < [T/\varepsilon^N]} E(|h_{i,n}^N|) \to 0 \quad \text{as } N \to \infty. \tag{5.4}$$

Proof. First, letting $\rho = (\varepsilon/\tau)^{1/2}$,

$$E(\Delta Z_n|\mathscr{F}_n) = \rho(E(\Delta X_n|\mathscr{F}_n) - \Delta\gamma_n) = \rho\varepsilon(w(X_n) - w(\gamma_n)) + \rho e_{1,n}. \tag{5.5}$$

Let $|X|_3 = E(|X|^3)^{1/3}$. Then

$$|E(\Delta Z_n|\mathscr{F}_n)|_3 \leqq \rho\varepsilon W'|X_n - \gamma_n|_3 + \rho|e_{1,n}|_3 \leqq K\varepsilon + \rho|e_{1,n}|_3 \tag{5.6}$$

by Theorem 1. Therefore

$$\begin{aligned} E(h_{3,n})^{1/3} = |\Delta Z_n|_3 &\leqq |E(\Delta Z_n|\mathscr{F}_n)|_3 + |\Delta Z_n - E(\Delta Z_n|\mathscr{F}_n)|_3 \\ &\leqq K\varepsilon + \rho|e_{1,n}|_3 + \rho|e_{3,n}|_3 \end{aligned}$$

by (5.6). Cubing both sides, using $(a + b + c)^3 \leqq 9(a^3 + b^3 + c^3)$ and summing over $n < k = [T/\varepsilon]$, we obtain

$$\begin{aligned} \sum_{n=0}^{k-1} E(h_{3,n}) &\leqq K\varepsilon^2 + 9 \sum_{n=0}^{k-1} \rho^3|e_{1,n}|_3^3 + K\varepsilon^{1/2} \\ &\leqq 9\left(\sum_{n=0}^{k-1} \rho|e_{1,n}|_3\right)^3 + K\varepsilon^{1/2}. \end{aligned}$$

According to (2.9), the sum on the right approaches 0 as $N \to \infty$, so (5.4) holds for $i = 3$.

Now

$$\begin{aligned} w(X_n) - w(\gamma_n) &= (X_n - \gamma_n)Dw(\gamma_n) + c_n(X_n - \gamma_n)^2 \\ &= (X_n - \gamma_n)(Dw(\varphi_n) + d_n) + c_n(X_n - \gamma_n)^2, \end{aligned}$$

where $|c_n| \leqq m(Dw)/2 \leqq K$ by (2.6), and $|d_n| \leqq m(Dw)|\gamma_n - \varphi_n| \leqq K\varepsilon$ by Lemma 2. Substituting this expression into (5.5), we see that

$$h_{1,n} = \varepsilon d_n Z_n + (\varepsilon/\rho)c_n Z_n^2 + \rho e_{1,n}.$$

Hence, by Theorem 1,

$$E(|h_{1,n}|) \leqq K\varepsilon^2 + K(\varepsilon/\rho) + \rho|e_{1,n}|_3.$$

This expression and (2.9) imply (5.4) for $i = 1$.

Finally,

$$\begin{aligned} h_{2,n} &= E((\Delta Z_n)^2|\mathscr{F}_n) - \varepsilon s(\varphi_n) \\ &= \operatorname{var}(\Delta Z_n|\mathscr{F}_n) - \varepsilon s(\varphi_n) + E(\Delta Z_n|\mathscr{F}_n)^2 \\ &= \varepsilon(s(X_n) - s(\varphi_n)) + (\varepsilon/\tau)e_{2,n} + E(\Delta Z_n|\mathscr{F}_n)^2 \end{aligned}$$

by (2.2). Hence, in view of (2.7),

$$E(|h_{2,n}|) \leqq \varepsilon K E(|X_n - \varphi_n|) + (\varepsilon/\tau)E(|e_{2,n}|) + |E(\Delta Z_n|\mathscr{F}_n)|_3^2.$$

But

$$E(|X_n - \varphi_n|) \leqq E(|X_n - \gamma_n|) + E(|\gamma_n - \varphi_n|) \leqq K(\tau/\varepsilon)^{1/2} + K\varepsilon$$

by Theorem 1 and Lemma 2. Thus, recalling (5.6),

$$E(|h_{2,n}|) \leqq K\varepsilon((\tau/\varepsilon)^{1/2} + \varepsilon) + (\varepsilon/\tau)E(|e_{2,n}|) + 2\rho^2|e_{1,n}|_3^2.$$

Summing over $n < k$ and using (2.9) and (2.10), we obtain (5.4) for $i = 2$.

6. The process $z^N(t)$. The proof of Theorem 2 depends heavily on properties of the transition probability P (see (2.17)) of the limiting process $\{z(t), t \geqq 0\}$. (As usual, we suppress N.) The proofs of these properties involve only routine analysis, so we omit many details.

Recall that $f(t, x)$ and $f(t) = f(t, x_0)$ are defined by the differential equation (2.14) and the initial condition $f(0, x) = x$. Similarly, $B(t, x)$ and $g(t, x)$ are the unique solutions of

$$\frac{dB(t, x)}{dt} = Dw(f(t, x))B(t, x), \qquad B(0, x) = 1 \tag{6.1}$$

and

$$\frac{dg(t, x)}{dt} = 2Dw(f(t, x))g(t, x) + s(f(t, x)), \qquad g(0, x) = 0. \tag{6.2}$$

The following lemma is basic to subsequent developments.

LEMMA 4. *The function P in* (2.17) *is a Markov transition probability.*

Proof. Clearly $P(t, z; u, \cdot)$ is a probability, $P(t, z; t, \cdot) = \delta_z$, and $P(t, \cdot\,; u, A)$ is measurable. Thus it remains only to verify the Chapman–Kolmogorov equality

$$\int P(t, z; u, d\zeta)P(u, \zeta; v, A) = P(t, z; v, A) \tag{6.3}$$

for $t \leqq u \leqq v$. Taking characteristic functions on both sides, we see that the distributions (2.17) satisfy (6.3) provided that

$$B(v - u, f(u))B(u - t, f(t)) = B(v - t, f(t)) \tag{6.4}$$

and

$$B(v - u, f(u))^2 g(u - t, f(t)) = g(v - t, f(t)) - g(v - u, f(u)). \tag{6.5}$$

Now $f(u - t, f(t))$ and $f(u)$ both satisfy $dh(u)/du = w(h(u))$, $u \geqq t$ and $h(t) = f(t)$. Since w is Lipschitz, we must have $f(u - t, f(t)) = f(u)$ for $u \geqq t$. Similarly, the functions on the left and right in (6.4) satisfy $dh(v)/dv = Dw(f(v))h(v)$, $v \geqq u$ and $h(u) = B(u - t, f(t))$. Since Dw is bounded, (6.4) obtains. Finally, the functions on the left and right in (6.5) satisfy $dh(v)/dv = 2Dw(f(v))h(v)$, $v \geqq u$ and $h(u) = g(u - t, f(t))$. Thus (6.5) is valid, and the proof of Lemma 4 is complete.

To proceed further, we need certain bounds on $B(t, x)$, $g(t, x)$ and their derivatives $B'(t, x) = dB(t, x)/dt$ and $g'(t, x)$.

LEMMA 5. *The quantities $B(t, x)$, $B'(t, x)$, $g(t, x)$ and $g'(t, x)$ are bounded over $t \leqq T$, $x \in I$ and $N \geqq 1$. Moreover,*

$$M(B'(\cdot, x)) = \sup_{t \neq u \leqq T} |B'(u, x) - B'(t, x)|/|u - t| \tag{6.6}$$

and $M(g'(\cdot, x))$ are bounded over $x \in I$ and $N \geqq 1$.

Proof. It follows immediately from (2.4), (2.15) and (6.1) that $|B(t,x)| \leqq \exp(W'T)$ and $|B'(t,x)| \leqq W' \exp(W'T)$. Moreover, (6.1) yields

$$M(B'(\cdot,x)) \leqq (WW'' + W'^2)\exp(W'T),$$

via the "product rule"

$$M(ab) \leqq M(a)\sup_t |b(t)| + M(b)\sup_t |a(t)|$$

and the inequality $M(Dw(f(\cdot,x))) \leqq W''W$. Boundedness of g, g' and $M(g')$ is established by an analogous argument based on (2.16) and (6.2).

As a consequence of Lemma 4, there is Markov process, $\{z(t), t \geqq 0\}$, with transition probability P and initial state $z(0) = 0$. We can and shall assume that it is separable, and we shall see shortly that almost all of its sample paths are continuous. Theorem 2 says that $\mathscr{L}(z(n\varepsilon))$ approximates $\mathscr{L}(Z_n)$. Lemmas 3 and 6 imply that conditional moments of $z(t+\varepsilon) - z(t)$ approximate those of ΔZ_n.

LEMMA 6. *Let $h_i(\delta, t)$ be defined by*

$$E(z(t+\delta) - z(t)|z(t) = z) = \delta Dw(f(t))z + h_1(\delta,t,z), \tag{6.7}$$

$$E((z(t+\delta) - z(t))^2|z(t) = z) = \delta s(f(t)) + h_2(\delta,t,z), \tag{6.8}$$

$$E(|z(t+\delta) - z(t)|^3|z(t) = z) = h_3(\delta,t,z). \tag{6.9}$$

Then

$$|h_1(\delta,t,z)| \leqq K\delta^2|z|, \tag{6.10}$$

$$|h_2(\delta,t,z)| \leqq K\delta^2(1+z^2), \tag{6.11}$$

$$|h_3(\delta,t,z)| \leqq K\delta^{3/2}(1+|z|^3) \tag{6.12}$$

for all $N \geqq 1$, $t \geqq 0$ and $\delta \leqq T$.

Proof. Clearly,

$$h_1(\delta,t,z) = (B(\delta,f(t)) - 1 - \delta Dw(f(t)))z.$$

But $B'(0,f(t)) = Dw(f(t))$ by (6.1), and $M(B'(\cdot,f(t))) \leqq K$ by Lemma 5, so (6.10) follows.

Note next that

$$E(|z(t+\delta) - B(\delta,f(t))z(t)|^j|z(t) = z) = g(\delta,f(t))^{j/2}c_j, \tag{6.13}$$

where c_j is the jth absolute moment of the standard normal distribution. Consequently,

$$h_2(\delta,t,z) = (g(\delta,f(t)) - \delta s(f(t))) + (B(\delta,f(t)) - 1)^2 z^2.$$

But $g(0,f(t)) = 0$, $g'(0,f(t)) = s(f(t))$, and $M(g'(0,f(t))) \leqq K$, so (6.11) obtains.

Finally, since $(a+b)^3 \leqq 4a^3 + 4b^3$,

$$\begin{aligned} h_3(\delta,t,z) &\leqq 4g(\delta,f(t))^{3/2}c_3 + 4|B(\delta,f(t)) - 1|^3|z|^3 \\ &\leqq K\delta^{3/2} + K\delta^3|z|^3, \end{aligned}$$

and (6.12) is proved.

Taking $t = 0$ and $\delta = t$ in (6.13), we see that

$$E(|z(t)|^3) \leqq K \tag{6.14}$$

for $N \geqq 1$ and $t \leqq T$. In combination with (6.12), this implies that

$$E(|z(t+\delta) - z(t)|^3) \leqq K\delta^{3/2} \tag{6.15}$$

for all $t, \delta \leqq T$ and $N \geqq 1$. In particular, almost all sample paths of $z(t)$ are continuous [10, Prop. III.5.3]. Thus $z(t)$ is a diffusion.

Let $V_{t,u}$ be the transition operator corresponding to P; i.e.,

$$V_{t,u}F(z) = E(F(z(u))|z(t) = z)$$

for bounded measurable real-valued functions F. Let B^3 be the space of bounded functions with three bounded derivatives, and let

$$\|F\|_3 = \max_{1 \leqq j \leqq 3} |d^jF/dz^j|_\infty,$$

where $|G|_\infty = \sup_z |G(z)|$.

LEMMA 7. *There is a constant K such that*

$$\|V_{t,t+\delta}F\|_3 \leqq K\|F\|_3 \tag{6.16}$$

for all $t \geqq 0, \delta \leqq T, N \geqq 1$ and $F \in B^3$.

Proof. Let Q be a normally distributed random variable, with mean 0 and variance $g(\delta, f(t))$. Clearly,

$$V_{t,t+\delta}F(z) = E[F(B(\delta, f(t))z + Q)],$$

so

$$\frac{d^jV_{t,t+\delta}F(z)}{dz^j} = B(\delta, f(t))^j V_{t,t+\delta}\left(\frac{d^jF}{dz^j}\right)(z)$$

for $j \leqq 3$. According to Lemma 5, there is a constant C such that $B(\delta, f(t)) \leqq C$, hence

$$|d^jV_{t,t+\delta}F/dz^j|_\infty \leqq C^j|d^jF/dz^j|_\infty,$$

and (6.16) holds with $K = C^3$.

7. Completion of proof of Theorem 2. The remainder of the proof of Theorem 2 is a non-Markovian variant of a standard argument. See [11, Lemma 9.2.3] for a similar argument in a Markovian context. Our approach and Khintchine's [6, Chap. 3] have much in common.

For $F \in B^3$ and $0 \leqq j \leqq n \leqq T/\varepsilon$, let $G_j = V_{j\varepsilon,n\varepsilon}F$.

LEMMA 8. *If $m \leqq n$,*

$$E[|E(F(Z_n)|\mathscr{F}_m) - V_{m\varepsilon,n\varepsilon}F(Z_m)|] \leqq \sum_{j=m}^{n-1} E[|E(G_{j+1}(Z_{j+1})|\mathscr{F}_j) - G_j(Z_j)|].$$

Proof. Note first that

$$\begin{aligned} F(Z_n) - V_{m\varepsilon,n\varepsilon}F(Z_m) &= G_n(Z_n) - G_m(Z_m) \\ &= \sum_{j=m}^{n-1} (G_{j+1}(Z_{j+1}) - G_j(Z_j)). \end{aligned}$$

Hence

$$E(F(Z_n)|\mathscr{F}_m) - V_{m\varepsilon,n\varepsilon}F(Z_m) = \sum_{j=m}^{n-1} E[E(G_{j+1}(Z_{j+1})|\mathscr{F}_j) - G_j(Z_j)|\mathscr{F}_m].$$

The lemma follows on taking absolute values and expectations on both sides of this equality.

We shall now obtain expressions for $E(G_{j+1}(Z_{j+1})|\mathscr{F}_j)$ and $G_j(Z_j)$ that permit us to estimate their difference. To simplify notation, we write G instead of G_{j+1}. A third order Taylor expansion of G about Z_j yields

$$E(G(Z_{j+1})|\mathscr{F}_j) = G(Z_j) + G'(Z_j)E(\Delta Z_j|\mathscr{F}_j) + 2^{-1}G''(Z_j)E((\Delta Z_j)^2|\mathscr{F}_j) + \omega 6^{-1}|G'''|_\infty E(|\Delta Z_j|^3|\mathscr{F}_j),$$

where $|\omega| \leqq 1$. Using (5.1)–(5.3), we obtain

$$E(G(Z_{j+1})|\mathscr{F}_j) = G(Z_j) + \varepsilon Dw(\varphi_j)Z_jG'(Z_j) + 2^{-1}\varepsilon s(\varphi_j)G''(Z_j) + \omega \sum_{i=1}^{3} |G^{(i)}|_\infty |h_{i,j}|/i!.$$

Since $G_j(Z_j) = V_{j\varepsilon,(j+1)\varepsilon}G(Z_j)$, (6.7)–(6.9) give

$$G_j(Z_j) = G(Z_j) + \varepsilon Dw(\varphi_j)Z_jG'(Z_j) + 2^{-1}\varepsilon s(\varphi_j)G''(Z_j) + \omega \sum_{i=1}^{3} |G^{(i)}|_\infty |h_i(\varepsilon, j\varepsilon, Z_j)|/i!.$$

Subtracting these two expressions, we obtain

$$\begin{aligned} E[|E(G_{j+1}(Z_{j+1})|\mathscr{F}_j) - G_j(Z_j)|] &\leqq \sum_{i=1}^{3} (i!)^{-1}|G_{j+1}^{(i)}|_\infty [E(|h_{i,j}|) + E(|h_i(\varepsilon, j\varepsilon, Z_j)|)] \\ &\leqq K\|F\|_3 \sum_{i=1}^{3} (i!)^{-1}[E(|h_{i,j}|) + K\varepsilon^{3/2}]. \end{aligned}$$

The latter expression results from Lemma 7, applied to $G_{j+1} = V_{(j+1)\varepsilon,n\varepsilon}F$, and Lemma 6 and Theorem 1, applied to $E(|h_i|)$.

When this inequality is combined with Lemma 8 and note is taken of (5.4), we conclude that there is a sequence $r_1 = r_1^N$, independent of F, m and n, such that $r_1 \to 0$ as $N \to \infty$ and

$$E[|E(F(Z_n)|\mathscr{F}_m) - V_{m\varepsilon,n\varepsilon}F(Z_m)|] \leqq r_1\|F\|_3 \tag{7.1}$$

for $m \leqq n \leqq T/\varepsilon$.

Let $\|F\|$ be the maximum of $\|F\|_3$ and $|F|_\infty$. It can be shown by induction that, for each $J \geqq 1$, there is a sequence $r_J = r_J^N$ such that

$$\left| E\left(\prod_{j=1}^{J} F_j(Z_{n_j})\right) - E\left(\prod_{j=1}^{J} F_j(z(n_j\varepsilon))\right)\right| \leqq r_J \prod_{j=1}^{J} \|F_j\| \tag{7.2}$$

for $F_j \in B^3$ and $n_1 \leqq n_2 \leqq \cdots \leqq T/\varepsilon$. The case $J = 1$ follows from (7.1) on taking $m = 0$. The essential idea of the inductive step is contained in the case $J = 2$ treated next.

Clearly,

$$\begin{aligned}\Delta &= E(F(Z_m)G(Z_n)) - E(F(z(m\varepsilon))G(z(n\varepsilon)))\\ &= E(F(Z_m)[E(G(Z_n)|\mathscr{F}_m) - V_{m\varepsilon,n\varepsilon}G(Z_m)])\\ &\quad + E(H(Z_m)) - E(H(z(m\varepsilon))),\end{aligned}$$

where $H(z) = F(z)V_{m\varepsilon,n\varepsilon}G(z)$. Hence

$$|\Delta| \leqq |F|_\infty r_1 \|G\|_3 + r_1\|H\|$$

by (7.1). But

$$\|H\| \leqq 8\|F\|\,\|V_{m\varepsilon,n\varepsilon}G\| \leqq K\|F\|\,\|G\|$$

by Lemma 7. Therefore $|\Delta| \leqq Kr_1\|F\|\,\|G\|$, as required.

The transition from (7.2) to (2.19), which involves arbitrary bounded continuous functions $F(z_1, \cdots, z_J)$, is straightforward. We sketch this transition for the case $J = 1$. Let F be given and let L be a number greater than 1. There is a polynomial P such that $|F(z) - P(z)| \leqq L^{-1}$ for $|z| \leqq L + 1$. Let S be an infinitely differentiable function with $S(z) = 1$ for $|z| \leqq L$, $0 \leqq S(z) \leqq 1$ for $L \leqq |z| \leqq L + 1$, and $S(z) = 0$ for $|z| \geqq L + 1$. Then $F_L(z) = P(z)S(z)$ has the following properties: $|F(z) - F_L(z)| \leqq L^{-1}$ for $|z| \leqq L$, $|F_L|_\infty \leqq |F|_\infty + 1$, and F_L is infinitely differentiable. Clearly,

$$|E(F(Z_n)) - E(F_L(Z_n))| \leqq (2|F|_\infty + 1)P(|Z_n| \geqq L) + L^{-1} \leqq KL^{-1}$$

by Theorem 1, and, in view of (6.14), a similar bound applies if Z_n is replaced by $z(n\varepsilon)$. Thus the maximum in (2.19) is bounded by KL^{-1} plus the comparable maximum for F_L. By (7.2), the latter maximum converges to 0 as $N \to \infty$. Since L is arbitrary, (2.19) follows.

Analogous arguments lead from (2.19) to (2.21) and (2.22). In the first case, for example, we approximate $F(z) = z$ by the bounded continuous function F_L that agrees with F for $|z| \leqq L$ and is constant for $z \leqq -L$ and $z \geqq L$.

This completes the proof of Theorem 2.

8. Convergence of processes. Let

$$Z^N(t) = Z_n^N + (\Delta Z_n^N)(t - n\varepsilon^N)/\varepsilon^N$$

for $n\varepsilon^N \leqq t \leqq (n+1)\varepsilon^N$, and let Z^N denote the process $\{Z^N(t),\ 0 \leqq t \leqq T\}$. Similarly, let z^N denote the process $\{z^N(t), 0 \leqq t \leqq T\}$. Let $C[0, T]$ be the space of continuous functions on $[0, T]$.

THEOREM 3. *Suppose that, in place of* (2.9) *and* (2.12), *we require that the stronger inequalities* (2.9′) *and* (2.12′) *hold for* $n \leqq T/\varepsilon$. *Then the distributions of* Z^N *and* z^N *over* $C[0, T]$ *are asymptotically equivalent. That is, for any bounded continuous real-valued function* H *on* $C[0, T]$, $E(H(Z^N)) - E(H(z^N)) \to 0$ *as* $N \to \infty$.

Theorem 3 applies to the monoecious Wright–Fisher model discussed in §§ 1 and 2. However, the dioecious model of § 3 does not appear to satisfy hypothesis (2.9′) for $n = 0$ and 1, even if $\varepsilon^3 = o(\tau)$.

The proof of Theorem 3 is based on Theorem 2 and the following lemma.

LEMMA 9. *The sequences* $\{Z^N, N \geqq 1\}$ *and* $\{z^N, N \geqq 1\}$ *are tight.*

Proof. Recall that $z^N(0) = 0$. Since the constant K in (6.15) can be chosen independently of N, tightness of $\{z^N, N \geqq 1\}$ follows from [1, Thm. 12.3]. To complete the proof, it suffices to show that Z^N satisfies an inequality like (6.15), and this, in turn, follows from the special case

$$E(|Z_n^N - Z_m^N|^3) \leqq K(n\varepsilon - m\varepsilon)^{3/2}, \tag{8.1}$$

$m \leqq n \leqq (T/\varepsilon) + 1$, which will now be proved.

Clearly,

$$\begin{aligned} Z_n - Z_m &= \sum_{j=m}^{n-1} E(\Delta Z_j|\mathscr{F}_j) + \sum_{j=m}^{n-1} [\Delta Z_j - E(\Delta Z_j|\mathscr{F}_j)] \\ &= A_{m,n} + B_{m,n}. \end{aligned}$$

Hence, letting $|Y|_3 = E(|Y|^3)^{1/3}$,

$$|Z_n - Z_m|_3 \leqq |A_{m,n}|_3 + |B_{m,n}|_3. \tag{8.2}$$

But

$$|A_{m,n}|_3 \leqq \sum_{j=m}^{n-1} |E(\Delta Z_j|\mathscr{F}_j)|_3 \leqq K(n - m)\varepsilon \tag{8.3}$$

by (5.6) and (2.9′), and

$$|B_{m,n}|_3^3 \leqq 6(n - m)^{3/2}\rho^3 K\tau^{3/2} = K[(n - m)\varepsilon]^{3/2} \tag{8.4}$$

by Lemma 1 and (2.12′). Combining (8.2)–(8.4), we obtain the required inequality (8.1).

8.1. Completion of proof of Theorem 3. Suppose that the distributions of Z^N and z^N are not asymptotically equivalent. Then there is a bounded continuous H and a $\delta > 0$ such that

$$|E(H(Z^N)) - E(H(z^N))| \geqq \delta \tag{8.5}$$

for a sequence $N = N_k$ of values of N. In view of Lemma 9, we may suppose, without loss of generality, that the distributions of Z^N and z^N converge to probabilities μ and ν, respectively, along this sequence. As a consequence of (8.5), $\int H\, d\mu \neq \int H\, d\nu$, so $\mu \neq \nu$.

Let $t_1 < t_2 < \cdots < t_J \leqq T$ be given and let $t_j^N = \varepsilon^N[t_j/\varepsilon^N]$. It follows from Theorem 2 that if F is bounded and continuous,

$$E[F(Z^N(t_1^N), \cdots, Z^N(t_J^N))] - E[F(z^N(t_1^N), \cdots, z^N(t_J^N))] \to 0$$

as $N \to \infty$. Moreover, if F is uniformly continuous, (6.15) and the analogue for Z^N imply that

$$E[F(z^N(t_1^N), \cdots, z^N(t_J^N))] - E[F(z^N(t_1), \cdots, z^N(t_J))] \to 0$$

and that the same limit obtains if z^N is replaced by Z^N. It follows that

$$\int F(z(t_1), \cdots, z(t_J))\mu(dz) = \int F(z(t_1), \cdots, z(t_J))\nu(dz),$$

hence $\mu = \nu$. This contradicts our earlier conclusion that $\mu \neq \nu$. Thus the supposition that Theorem 3 is false is untenable.

9. Multidimensional processes. Theorems 1, 2, and 3 and their proofs need only trivial modifications to accommodate multidimensional processes X_n. The required changes in the theorems will now be described. Changes in the proofs are left to the reader.

The assumptions stated in the second paragraph of § 2 require the following amendments. The random vector X_n^N takes on values in a closed convex subset I of M-dimensional Euclidean space R^M. In (2.2), we interpret the left side as the conditional covariance matrix of ΔX_n^N. For $y \in R^M$, $|y|$ is the Euclidean norm of y. If A is an $M \times M$ matrix (e.g., $A = Dw^N(x)$ or $A = s^N(x)$), we can take $|A|^2 = \sum_{i,j} (A_{ij})^2$. Conditions (2.3)–(2.7), (2.9), (2.10) and (2.12) then require no modification. The condition $s^N(x) \geqq 0$ is naturally taken to mean that $s^N(x)$ is nonnegative definite.

In the statement of Theorem 2, we can define $B^N(\cdot, x)$ and $g^N(\cdot, x)$ as the unique solutions of (6.1) and

$$\frac{dg(t, x)}{dt} = Dw(f(t, x))g(t, x) + g(t, x)Dw(f(t, x))' + s(f(t, x)) \tag{6.2'}$$

satisfying $B(0, x) =$ identity and $g(0, x) = 0$. Here prime indicates transposition and N is omitted. Theorem 3 applies without change to multidimensional processes.

The multiallelic generalization of the Wright–Fisher model described in § 1 provides an example of a multidimensional process satisfying our assumptions. Suppose that $M \geqq 1$ and that there are $M + 1$ alleles, $A_1, \cdots, A_{M+1}$. Let the genotype A_iA_j have fitness $1 + v_{ij}$, and let A_i mutate into A_j, $j \neq i$, with probability α_{ij}. Let X_n be the vector of relative frequencies of $A_1, \cdots, A_M$ in generation n, and suppose that $X_n = x$. Define x_{M+1} so that $\sum_{i=1}^{M+1} x_i = 1$. After selection and mutation, gene probabilities are given by

$$x_i^* = \sum_{j=1}^{M+1} x_i x_j(1 + v_{ij}) \Big/ \sum_{i,j=1}^{M+1} x_i x_j (1 + v_{ij}),$$

and

$$\gamma_i(x) = x_i^*\left(1 - \sum_{j=1}^{M+1} \alpha_{ij}\right) + \sum_{j=1}^{M+1} x_j^* \alpha_{ji},$$

where the sums in the second equation omit $j = i$. The distribution of X_{n+1} given $X_n = x$, is multinomial, with mean vector $\gamma(x)$ and sample size $2N$.

Let $\tau = 1/N$, and let $\varepsilon = \max(|v_{ij}|, \alpha_{ij})$. Then application of Theorems 1, 2, and 3 to this model proceeds just as in the case of two alleles treated previously.

REFERENCES

[1] P. Billingsley, *Convergence of Probability Measures*, John Wiley, New York, 1968.

[2] L. L. Cavalli-Sforza and W. F. Bodmer, *The Genetics of Human Populations*, W. H. Freeman, San Francisco, 1971.

[3] S. W. DHARMADHIKARI, V. FABIAN AND K. JOGDEO, *Bounds on the moments of martingales*, Ann. Math. Statist., 39 (1968), pp. 1719–1723.

[4] W. J. EWENS, *Population Genetics*, Methuen, London, 1969.

[5] W. FELLER, *Diffusion processes in genetics*, Proc. 2nd Berkeley Sympos. Math. Statist. and Probability, J. Neyman, ed., University of California Press, Berkeley, 1951, pp. 227–246.

[6] A. KHINTCHINE, *Asymptotische Gesetze der Wahrscheinlichkeitsrechnung*, Chelsea, New York, 1948.

[7] M. KIMURA, *Diffusion models in population genetics*, J. Appl. Probability, 1 (1964), pp. 177–232.

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

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

[10] J. NEVEU, *Mathematical Foundations of the Calculus of Probability*, Holden-Day, San Francisco, 1965.

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

[12] ———, *A central limit theorem for Markov processes that move by small steps*, Ann. Probability, 2 (1974), pp. 1065–1074.

[13] R. M. REDHEFFER, *The theorems of Bony and Brezis on flow-invariant sets*, Amer. Math. Monthly, 79 (1972), pp. 740–747.