The Annals of Probability
1974, Vol. 2, No. 1, 40-50

ON THE WEAK CONVERGENCE OF INTERPOLATED MARKOV CHAINS TO A DIFFUSION[1]

BY HAROLD J. KUSHNER

Brown University

Let $\{\xi_k^n, k = 0, 1, \cdots\}$ denote a R^r valued discrete parameter Markov process for each n. For each real $T > 0$, it is shown that suitable piecewise interpolations in $D^r[0, T]$ converge weakly as $n \to \infty$, to the diffusion given by

$$x(t) = x + \int_0^t f(x(s), s)\, ds + \int_0^t \sigma(x(s), s)\, dw(s)\,, \tag{*}$$

under essentially the condition that the solution to (*) is unique in the sense of multivariate distributions, $f(\cdot, \cdot)$, $\sigma(\cdot, \cdot)$ are bounded and continuous, and the scaled "infinitesimal" coefficients of the $\{\xi_k^n\}$ are close to $f(\cdot, \cdot)$ and $\sigma(\cdot, \cdot)$. It is not required that $f(\cdot, \cdot)$ or $\sigma(\cdot, \cdot)$ satisfy a uniform Lipschitz condition, nor that $\sigma(\cdot, \cdot)\sigma'(\cdot, \cdot)$ be positive definite. The result extends the result of Gikhman and Skorokhod (1969). Two examples arising in population genetics are given, where $\sigma(\cdot, \cdot)$ is not uniformly Lipschitz.

1. Introduction. For each n, let $\{\xi_k^n, k = 0, 1, \cdots\}$ denote a discrete parameter Markov process with values in R^r, Euclidean r-space, where ξ_0^n converges in distribution to a random variable ξ_0 as $n \to \infty$. Let $\mathscr{B}_k^n$ denote the minimum σ-algebra which measures ξ_i^n, $i \leqq k$, and let δt_i^n, $i = 0, 1, \cdots$, denote a sequence of positive real numbers. Define

$$t_k^n = \sum_{i=0}^{k-1} \delta t_i^n\,.$$

Let T denote a fixed positive real number, and let $f(\cdot, \cdot)$ and $f_n(\cdot, \cdot)$ denote bounded R^r valued functions on $R^r \times [0, T]$, and $\sigma_n(\cdot, \cdot)$, and $\sigma(\cdot, \cdot)$ bounded $r \times r$ matrices on $R^r \times [0, T]$. Suppose that

$$E_{\mathscr{B}_k^n}[\xi_{k+1}^n - \xi_k^n] = f_n(\xi_k^n, t_k^n)\delta t_k^n$$
$$\operatorname{Cov}_{\mathscr{B}_k^n}[\xi_{k+1}^n - \xi_k^n] = \sigma_n(\xi_k^n, t_k^n)\sigma_n'(\xi_k^n, t_k^n)\delta t_k^n\,,$$

where $'$ denotes the transpose. Let $N_n = \min\{k : t_k^n > T\}$ and suppose that (where Euclidean norms are used)

$$E \sum_{k=0}^{N_n-1} \{|f_n(\xi_k^n, t_k^n) - f(\xi_k^n, t_k^n)|^2 + |\sigma_n(\xi_k^n, t_k^n) - \sigma(\xi_k^n, t_k^n)|^2\}\delta t_k^n \to 0 \quad \text{as } n \to \infty\,. \tag{1}$$

Under some additional conditions Gikhman and Skorokhod (1969) showed, for $r = 1$, that suitable continuous time interpolations of the ξ_k^n processes con-

Received February 12, 1973; revised May 31, 1973.

[1] This research was supported in part by the National Science Foundation under grant GK 31073X and in part by the Office of Naval Research under grant NONR N00014-67-A-0191-0018.

AMS 1970 *subject classifications*. 6030, 6062, 6065, 9220.

Key words and phrases. Weak convergence of Markov chains to a diffusion, diffusions with degenerate differential generators, diffusion models in genetics.

verge weakly to a diffusion satisfying

(2) $$x(t) = x + \int_0^t f(x(s), s)\, ds + \int_0^t \sigma(x(s), s)\, dw(s)\,,$$

as $n \to \infty$, where w is a standard Wiener process. Crucial to their proof were the conditions that $f(\cdot, t)$ and $\sigma(\cdot, t)$ satisfy a Lipschitz condition uniformly in t, and that $|\sigma(\cdot, t)|$ be strictly positive. If $\sigma'(\cdot, \cdot)\sigma(\cdot, \cdot)$ is strictly positive definite and the uniform Lipschitz condition continues to hold, there is no difficult in extending their proof for general r.

There are many applications where $\sigma(\cdot, \cdot)\sigma'(\cdot, \cdot)$ is neither strictly positive definite, nor does $\sigma(\cdot, \cdot)$ satisfy a uniform Lipschitz condition. An example arising in population genetics is discussed at the end of the paper. We also note that the strict positive definiteness rarely holds in examples arising in stochastic control theory when $r > 1$. In this paper, we prove weak convergence under essentially the conditions that $f(\cdot, \cdot)$, $\sigma(\cdot, \cdot)$ are bounded and continuous, and that the solution to (2) is unique in the sense that any two non-anticipative solutions (perhaps each corresponding to a different Wiener process) have the same multivariate distributions.

2. Some preliminaries. For the most part, we follow the terminology of Billingsley (1968). Let D denote the space of real valued functions on $[0, T]$ which are continuous from the right on $[0, T)$, have left hand limits on $(0, T]$, and are continuous from the left at $t = T$. D^m denotes the m-fold product of D. There is a metric (which we use here) $d_0(x(\cdot), y(\cdot))$ (see Billingsley (1968) pages 112–116) on D^m, with respect to which D^m is a complete separable metric space, and convergence of $y_n(\cdot)$ to $y(\cdot)$ in the metric d_0 implies convergence at all points of continuity of $y(\cdot)$, and if $y(\cdot) \in D^m$ then $y(\cdot)$ has at most a countable number of discontinuities. Let $\mathscr{D}^m$ denote the Borel algebra induced on D^m by the metric d_0.

Let $\{\mu_n\}$ denote a tight sequence of measures on $(D^m, \mathscr{D}^m)$, corresponding to a sequence of processes $\{x^n(\cdot)\}$ with paths in D^m w.p. 1. Then $\{\mu_n\}$ has a weakly convergent subsequence $\{\mu_{n'}\}$ (converging to a measure μ on $(D^m, \mathscr{D}^m)$) and there is a separable process $x(\cdot)$ with paths in D^m w.p. 1, corresponding to the measure μ. Such convergence will be denoted by either $\mu_{n'} \to_w \mu$ or $x^{n'}(\cdot) \to_D x(\cdot)$, and if $\{\mu_n\}$ is tight we may write $\{x^n(\cdot)\}$ is tight.

The following (slightly reworded) lemma of Skorokhod (1956), page 281 will be helpful later.

LEMMA 1. *Let $\{v_n\}$ and v denote random variables with values in a complete separable metric space X. Let $v_n \to_D v$. Then there exist random variables $\{\tilde v_n\}$, $\tilde v$ with values in X such that for any Borel set A in X,*

$$P\{v_n \in A\} = P\{\tilde v_n \in A\}\,, \qquad P\{v \in A\} = P\{\tilde v \in A\}\,.$$

The random variables $\{\tilde v_n\}$, $\tilde v$, are defined on the same probability space, where $\Omega =$

$[0, 1]$, *and the probability measure is the Lebesgue measure, and*

$$P(\tilde{v}_n \to \tilde{v}, n \to \infty) = 1 ,$$

where the convergence is in the metric on X.

(In our case $X = D^m$ for some m, and the metric will be d_0. To each random variable $x(\cdot)$ with values in D^m corresponds a process ($t \in [0, T]$), also denoted by $x(\cdot)$, with paths in D^m w.p. 1. We can (and always will) suppose that the process $x(\cdot)$ is separable.)

Criterion for tightness on D^m. Let $\{x^n(\cdot)\}$ denote a family of processes with paths in D^m w.p. 1. A sufficient condition for tightness is (see Billingsley (1968) Chapter 3, Theorem 15.3, and proof of Theorem 15.6) that (i)—(iv) hold.

There is a real K so that, for all $0 \leqq t_1 < t \leqq t_2 \leqq T$,

(i) $E|x^n(t) - x^n(t_1)|^2|x^n(t_2) - x^n(t)|^2 \leqq K|t_2 - t_1|^2$. For each $\varepsilon > 0$, $\eta > 0$, there is a $\delta \in (0, T)$ and an integer n_0 so that for $n \geqq n_0$

(ii) $P\{\sup_{0 \leqq s \leqq t \leqq \delta} |x^n(t) - x^n(s)| \geqq \varepsilon\} \leqq \eta$.

(iii) $P\{\sup_{T-\delta \leqq s \leqq t \leqq T} |x^n(t) - x^n(s)| \geqq \varepsilon\} \leqq \eta$.

(iv) $P\{\sup_{t \leqq T} |x^n(t)| > a\} \to 0$ as $a \to \infty$, uniformly in n.

3. Assumptions. We will require (1) and (A1)—(A6).

(A1) $\max_{0 \leqq k \leqq N_n - 1} |\delta t_k^n| \to 0$ as $n \to \infty$.

(A2) $f(\cdot, \cdot)$ and $\sigma(\cdot, \cdot)$ are bounded and continuous on $R^r \times [0, T]$, $f_n(\cdot, \cdot)$, $\sigma_n(\cdot, \cdot)$ are uniformly bounded.

(A3) ξ_0^n converges in distribution to a random variable ξ_0, as $n \to \infty$.

(A4) $E \sum_{k=0}^{N_n - 1} |\xi_{k+1}^n - \xi_k^n - f_n(\xi_k^n, t_k^n)\delta t_k^n|^{2+\alpha} \to 0$ as $n \to \infty$, for some $\alpha > 0$.

(A5) There is a real $K_1 > 0$, so that $K_1 < \delta t_{k+1}^n/\delta t_k^n < K_1^{-1}$, for all n, k.

(A6) Let $\xi^i(\cdot)$, $i = 1, 2$, be R^r valued processes, non-anticipative with respect to the R^r valued Wiener processes $W^i(\cdot)$, $i = 1, 2$, resp. If $(W^i(\cdot), \xi^i(\cdot))$, $i = 1, 2$, satisfy (2), and $\xi^1(0)$ and $\xi^2(0)$ have the same distribution, then the multivariate distributions of $\xi^1(\cdot)$ are the same as those of $\xi^2(\cdot)$.

4. Interpolations in D^m. We write (the equation defines δY_k^n)

$$\xi_{k+1}^n = \xi_k^n + f_n(\xi_k^n, t_k^n)\delta t_k^n + \delta Y_k^n .$$

$\{\delta Y_k^n\}$ is an orthogonal sequence for each n, and $E_{\mathscr{B}_k^n}\delta Y_k^n(\delta Y_k^n)' = \sigma_n(\xi_k^n, t_k^n)\sigma_n'(\xi_k^n, t_k^n)\delta t_k^n$. Define Y_k^n by

$$Y_k^n = \sum_{i=0}^{k-1} \delta Y_i^n .$$

For each n and $\varepsilon > 0$, let $\{\delta U_k^{n,\varepsilon}\}$ denote a sequence of independent random variables in R^r with mean zero, and which are independent of the $\{\xi_k^n\}$. Let $E(\delta U_k^{n,\varepsilon})(\delta U_k^{n,\varepsilon})' = \varepsilon^2 I \delta t_k^n$, where I is the identity matrix and $E|\delta U_k^{n,\varepsilon}|^4 \leqq K|\delta t_k^n|^2$ for some real[2] K. Define

$$U_k^{n,\varepsilon} = \sum_{i=0}^{k-1} \delta U_i^{n,\varepsilon} .$$

[2] The value of K may vary from usage to usage. It will always denote a positive real number, independent of ω, n, ε, k, etc.

Define $X_k^{n,\varepsilon} = \xi_k^n + U_k^{n,\varepsilon}$, and $\delta Q_k^{n,\varepsilon} = \delta Y_k^n + \delta U_k^{n,\varepsilon}$, and let $Q_k^{n,\varepsilon}$ denote $Y_k^n + U_k^{n,\varepsilon}$. Then

$$X_{k+1}^{n,\varepsilon} = \xi_0^n + \sum_{i=0}^k f_n(\xi_i^n, t_i^n)\delta t_i^n + Y_k^n + U_k^{n,\varepsilon} .$$

Let $\mathscr{A}_k^{n,\varepsilon}$ denote the minimal σ-algebra which measures $\{\xi_i^n, U_i^{n,\varepsilon}, i \leqq k\}$.

Define the matrices $A^{n,\varepsilon}(\cdot, \cdot)$ and $A^\varepsilon(\cdot, \cdot)$ by

$$A^{n,\varepsilon}(\cdot, \cdot) = [\sigma_n(\cdot, \cdot)\sigma_n'(\cdot, \cdot) + \varepsilon^2 I]$$
$$A^\varepsilon(\cdot, \cdot) = [\sigma(\cdot, \cdot)\sigma'(\cdot, \cdot) + \varepsilon^2 I] .$$

If A is any positive definite matrix, let $A^{\frac{1}{2}}$ denote some square root of A (i.e., $A = A^{\frac{1}{2}}(A^{\frac{1}{2}})'$). Define $\sigma^\varepsilon(\cdot, \cdot) = [A^\varepsilon(\cdot, \cdot)]^{\frac{1}{2}}$, $\sigma^{n,\varepsilon}(\cdot, \cdot) = [A^{n,\varepsilon}(\cdot, \cdot)]^{\frac{1}{2}}$, and we can suppose that $\sigma^\varepsilon(\cdot, \cdot)$ and $\sigma^{n,\varepsilon}(\cdot,\cdot)$ are chosen so that $\sigma^\varepsilon(\cdot, \cdot)$ is bounded and continuous on $R^r \times [0, T]$ and that (see (1))

$$E \sum_{k=0}^{N_n-1} |\sigma^{n,\varepsilon}(\xi_k^n, t_k^n) - \sigma^\varepsilon(\xi_k^n, t_k^n)|^2 \delta t_k^n \to 0 \qquad \text{as } n \to \infty .$$

we can also suppose that $\sigma^\varepsilon(\cdot, \cdot)$ converges pointwise to $\sigma(\cdot, \cdot)$ as $\varepsilon \to 0$.

Define $b_k^{n,\varepsilon}$ by

$$b_k^{n,\varepsilon} = [\sigma^{n,\varepsilon}(\xi_k^n, t_k^n)]^{-1}[\delta Y_k^n + \delta U_k^{n,\varepsilon}] .$$

For each n, ε, $\{b_k^{n,\varepsilon}\}$ is an orthogonal sequence and $\operatorname{Cov} b_k^{n,\varepsilon} = \delta t_k^n I$. Define $B_k^{n,\varepsilon} = \sum_{i=0}^{k-1} b_i^{n,\varepsilon}$ and $Z_k^{n,\varepsilon} = \sum_{i=0}^{k-1} \sigma^\varepsilon(\xi_i^n, t_i^n) b_i^{n,\varepsilon}$. Note that

$$Q_k^{n,\varepsilon} = \sum_{i=0}^{k-1} \sigma^{n,\varepsilon}(\xi_i^n, t_i^n) b_i^{n,\varepsilon}$$

and that, using the martingale inequality of Doob (1953) VII, Theorem 3.4,

(3) $$E \max_{0 \leq k \leq N_n} |Q_k^{n,\varepsilon} - Z_k^{n,\varepsilon}|^2 \leqq 4E \sum_{k=0}^{N_n-1} |\sigma^\varepsilon(\xi_k^n, t_k^n) - \sigma^{n,\varepsilon}(\xi_k^n, t_k^n)|^2 \delta t_k^n \to 0 \qquad \text{as } n \to \infty .$$

Define the piecewise constant interpolations denoted by

(4) $$\xi^n(\cdot), Y^n(\cdot), U^{n,\varepsilon}(\cdot), Q^{n,\varepsilon}(\cdot), X^{n,\varepsilon}(\cdot), Z^{n,\varepsilon}(\cdot), B^{n,\varepsilon}(\cdot)$$

in D^r by, e.g.,

$$\xi^n(t) = \xi_k^n , \quad t_k^n \leqq t < t_{k+1}^n .$$

Let $m = 7r$ henceforth, and denote the R^m valued process whose components are listed in (4) by $\Phi^{n,\varepsilon}(\cdot)$.

LEMMA 2. *$\{\Phi^{n,\varepsilon}(\cdot)\}$ is tight in $(D^m, \mathscr{D}^m)$.*

PROOF. The proof is straightforward, and only a sketch for $Y^n(\cdot)$ will be given. Y_k^n is a martingale and the inequality

$$E \sup_{N_n \geq k \geq 0} |Y_k^n|^2 \leqq 4E|Y_{N_n}^n|^2 \leqq 4E \sum_{k=0}^{N_n-1} |\sigma_n(\xi_k^n, t_k^n)\sigma_n'(\xi_k^n, t_k^n)| \delta t_k^n \leqq KT ,$$

which holds for some real K, yields (iv). Both (ii) and (iii) follow from a very similar calculation and we omit the details.

For $s \in [0, T)$ define $m_n(s) = \max\{k : t_k^n < s\}$. Let $\mathscr{B}_s^n$ denote the minimum

σ-algebra measuring $\xi^n(t)$, $t \leqq s$. For $\tau > s$, we have

$$(5)\qquad E_{\mathscr{B}_s^n}|Y^n(\tau) - Y^n(s)|^2 = E_{\mathscr{B}^n_{m_n(s)}} |\textstyle\sum_{k=m_n(s)}^{m_n(\tau)-1} \delta Y_k^n|^2 \leqq K(t^n_{m_n(\tau)} - t^n_{m_n(s)})$$

for a real K (independent of n, τ, s, ω). Let $t_1 < s < t_2$, and let $m_n(s) = k$. If either $t_2 < t^n_{k+1}$ or $t_1 \geqq t_k^n$, then the left-hand side of (i) is zero. Thus suppose that $t_2 \geqq t^n_{k+1}$ and $t_1 < t_k^n$. By (5), the left-hand side of (i) is bounded above by

$$K^2(t^n_{m_n(t_2)} - t^n_{m_n(s)})(t^n_{m_n(s)} - t^n_{m_n(t_1)}) \leqq K^2(t_2 - t_1)(t^n_{m_n(s)} - t^n_{m_n(t_1)}),$$

and by (A5), $t^n_{m_n(s)} - t^n_{m_n(t_1)} \leqq (s - t_1) + \delta t_k^n/K_1 \leqq 2(t_2 - t_1)/K_1$, which implies (i) for some real K. □

LEMMA 3. *For each* ε, $\{B^{n,\varepsilon}(\cdot)\}$ *converges weakly to a standard r-dimensional Brownian motion* $W^\varepsilon(\cdot)$, *as* $n \to \infty$.

PROOF. The proof is a straightforward vector extension of the proof of Lemma 1, page 462, Gikhman and Skorokhod (1969), and we omit the details. The cited lemma proves that the finite dimensional distributions of $\{B^{n,\varepsilon}(\cdot)\}$ converge to those of a Wiener process. That and tightness (our Lemma 2) yield Lemma 3.

5. The convergence theorem.

THEOREM 1. *Under* (A1)—(A6), *there is a Wiener process* $\tilde{W}(\cdot)$ *and a random function* $\tilde{\xi}(\cdot)$, *non-anticipative with respect to* $\tilde{W}(\cdot)$, *so that* $\xi^n(\cdot) \to_D \tilde{\xi}(\cdot)$, *and*

$$(6)\qquad \tilde{\xi}(t) = \tilde{\xi}_0 + \textstyle\int_0^t f(\tilde{\xi}(s), s)\, ds + \int_0^t \sigma(\tilde{\xi}(s), s)\, d\tilde{W}(s)$$

where $\tilde{\xi}_0$ *has the same distribution as* ξ_0.

PROOF. By Lemma 2, for each ε, there is a separable process $\Phi^\varepsilon(\cdot)$ with paths in D^m w.p. 1 and a subsequence $\{\Phi^{n',\varepsilon}(\cdot)\}$ of $\{\Phi^{n,\varepsilon}(\cdot)\}$, so that $\Phi^{n',\varepsilon}(\cdot) \to_D \Phi^\varepsilon(\cdot)$. Until further notice we will suppose that n (and not n') indexes this subsequence. *Observe that the finite dimensional distributions of* $\xi^\varepsilon(\cdot)$ *and* $Y^\varepsilon(\cdot)$ *do not depend on* ε, $(\xi^n(\cdot), Y^n(\cdot)) \to_D (\xi^\varepsilon(\cdot), Y^\varepsilon(\cdot))$.

Considered as random variables, the $\Phi^{n,\varepsilon}(\cdot)$ and $\Phi^\varepsilon(\cdot)$ take values in D^m. By Lemma 1, we can define random functions (which we can and will suppose are separable) $\tilde{\Phi}^{n,\varepsilon}(\cdot)$, $n = 1, 2, \cdots$, and $\tilde{\Phi}^\varepsilon(\cdot)$, all on the same probability space, with paths in D^m w.p. 1, and having the same multivariate[3] distributions as the $\Phi^{n,\varepsilon}(\cdot)$, $n = 1, 2, \cdots$, and $\Phi^\varepsilon(\cdot)$, resp., and for which $\tilde{\Phi}^{n,\varepsilon}(\cdot) \to \tilde{\Phi}^\varepsilon(\cdot)$ w.p. 1 in the metric d_0. In particular, this implies that $\tilde{\Phi}^{n,\varepsilon}(\cdot) \to \tilde{\Phi}^\varepsilon(\cdot)$ for almost all ω, t. Again, note that the multivariate distributions of[4] $\tilde{\xi}^\varepsilon(\cdot)$ and $\tilde{Y}^\varepsilon(\cdot)$ do not depend on ε, and that $(\xi^n(\cdot), Y^\varepsilon(\cdot)) \to_D (\tilde{\xi}^\varepsilon(\cdot), \tilde{Y}^\varepsilon(\cdot))$. Now

$$\tilde{X}^{n,\varepsilon}(t) = \xi_0^n + \textstyle\int_0^t f(\tilde{\xi}^n(s), s)\, ds + \tilde{Y}^n(t) + \tilde{U}^{n,\varepsilon}(t) + \tilde{\mathscr{E}}^{n,\varepsilon}(t),$$

---

[3] Let $x(\cdot)$ denote the generic element of D^m. For any Borel set $B \in R^m$, and $t \in [0, T]$, $\{x(\cdot): x(t) \in B\} \in \mathscr{D}^m$. Thus, e.g. $P\{\Phi^\varepsilon(t_1), \cdots, \Phi^\varepsilon(t_k) \in B_1 \times \cdots \times B_k\} = P\{\tilde{\Phi}^\varepsilon(t_1), \cdots, \tilde{\Phi}^\varepsilon(t_k) \in B_1 \times \cdots \times B_k\}$ for any k, $t_i \in [0, T]$, and Borel B_i in R^m.

[4] $\tilde{\xi}^\varepsilon(\cdot)$, $\tilde{Y}^\varepsilon(\cdot)$, etc., are the obvious components of $\tilde{\Phi}^\varepsilon(\cdot)$.

where $\mathscr{E}^{n,\varepsilon}(\cdot)$ arises due to the approximation of $\sum_{i=0}^{m_n(t)-1} f_n(\tilde{\xi}_i^n, t_i^n)\delta t_i^n$ by the integral $\int_0^t f(\tilde{\xi}^n(s), s)\, ds$. The inequality (1), the continuity of $f(\cdot, \cdot)$, and the convergence of $\tilde{\xi}^n(\cdot)$ to $\tilde{\xi}^\varepsilon(\cdot)$ w.p. 1 imply that $\mathscr{E}^{n,\varepsilon}(\cdot)$ converges to the zero function as $n \to \infty$. By Lemma 1, for almost all ω, t,

$$\tilde{X}^\varepsilon(t) = \tilde{\xi}_0^\varepsilon + \int_0^t f(\tilde{\xi}^\varepsilon(s), s)\, ds + (\tilde{Y}^\varepsilon(t) + \tilde{U}^\varepsilon(t)) . \tag{7}$$

Note that one of the components of $\tilde{\Phi}^\varepsilon(\cdot)$ is a standard Wiener process $\tilde{W}^\varepsilon(\cdot)$. It will be shown below that $\tilde{\xi}^\varepsilon(\cdot)$ is non-anticipative with respect to $\tilde{W}^\varepsilon(\cdot)$, and that

$$\tilde{Q}^\varepsilon(t) = \tilde{Y}^\varepsilon(t) + \tilde{U}^\varepsilon(t) = \int_0^t \sigma^\varepsilon(\tilde{\xi}^\varepsilon(s), s)\, d\tilde{W}^\varepsilon(s) . \tag{8}$$

Thus since $\tilde{X}^\varepsilon(\cdot)$ is separable, (7) and (8) imply that it is continuous w.p. 1, and (7) holds w.p. 1, for all $t \in [0, T]$.

For any $\varepsilon_0 > 0$, the set $\{\tilde{\Phi}^\varepsilon(\cdot), \varepsilon \leq \varepsilon_0\}$ is tight and

$$E \sup_{0 \leq t \leq T} |\tilde{X}^\varepsilon(t) - \tilde{\xi}^\varepsilon(t)|^2 \leq E \sup_{0 \leq t \leq T} |\tilde{U}^\varepsilon(t)|^2 \to 0 , \tag{9}$$

as $\varepsilon \to 0$. There is a process $\tilde{\Phi}(\cdot) = (\tilde{\xi}(\cdot), \cdots)$ with paths in D^m w.p. 1 so that for some sequence $\varepsilon \to 0$ (we hold the sequence fixed in the sequel) $\tilde{\Phi}^\varepsilon(\cdot) \to_D \tilde{\Phi}(\cdot)$. The processes $\tilde{\xi}(\cdot)$ and $\tilde{Y}(\cdot)$ have the same multivariate distributions as $\tilde{\xi}^\varepsilon(\cdot)$ and $\tilde{Y}^\varepsilon(\cdot)$, resp. and are the weak limits of $\xi^n(\cdot)$ and $Y^n(\cdot)$, resp. Let us suppose that (appealing to Lemma 1) all the $\tilde{\Phi}^\varepsilon(\cdot)$ and $\tilde{\Phi}(\cdot)$ are defined on the same probability space and that $\tilde{\Phi}^\varepsilon(\cdot) \to \tilde{\Phi}(\cdot)$ w.p. 1 in the metric d_0 (i.e., for almost all ω, t). Define

$$\hat{\mathscr{E}}^\varepsilon(t) = \int_0^t [\sigma^\varepsilon(\tilde{\xi}^\varepsilon(s), s) - \sigma(\tilde{\xi}^\varepsilon(s), s)]\, d\tilde{W}^\varepsilon(s)$$

and write

$$\tilde{Y}^\varepsilon(t) + \tilde{U}^\varepsilon(t) = \int_0^t \sigma^\varepsilon(\tilde{\xi}^\varepsilon(s), s)\, d\tilde{W}^\varepsilon(s) + \hat{\mathscr{E}}^\varepsilon(t) . \tag{10}$$

Then

$$\begin{aligned} E \sup_{0 \leq t \leq T} |\hat{\mathscr{E}}^\varepsilon(t)|^2 &\leq 4E \int_0^T [\sigma^\varepsilon(\tilde{\xi}^\varepsilon(s), s) - \sigma(\tilde{\xi}^\varepsilon(s), s)]^2\, ds \\ &= 4E \int_0^T [\sigma^\varepsilon(\tilde{\xi}(s), s) - \sigma(\tilde{\xi}(s), s)]^2\, ds \to 0 \end{aligned}$$

as $\varepsilon \to 0$. Observe that $\tilde{Y}(\cdot)$ is the limit (for almost all ω, t) of the sequence of random functions with values

$$\int_0^t \sigma^\varepsilon(\tilde{\xi}^\varepsilon(s), s)\, d\tilde{W}^\varepsilon(s) . \tag{11}$$

Each $\tilde{W}^\varepsilon(\cdot)$ is a standard Wiener process ($t \in [0, T]$). Since the family of $\tilde{W}^\varepsilon(\cdot)$ is tight, and they all have the same multivariate distributions, it is obvious that there is a standard Wiener process $\tilde{W}(\cdot)$ such that $\tilde{W}^\varepsilon(\cdot) \to_D \tilde{W}(\cdot)$, and $\tilde{W}^\varepsilon(\cdot) \to \tilde{W}(\cdot)$ for all $t \in [0, T]$, w.p. 1. Also, $\tilde{\xi}^\varepsilon(\cdot) \to \tilde{\xi}(\cdot)$ for almost all ω, t. Indeed, it follows from (9) that $\tilde{\xi}(\cdot)$ is continuous w.p. 1. Hence $\tilde{\xi}^\varepsilon(\cdot) \to \tilde{\xi}(\cdot)$ for all t, w.p. 1. (9) implies that $\tilde{X}(\cdot) = \tilde{\xi}(\cdot)$.

Next, we show that, w.p. 1

$$\int_0^t \sigma(\tilde{\xi}^\varepsilon(s), s)\, d\tilde{W}^\varepsilon(s) \to \int_0^t \sigma(\tilde{\xi}(s), s)\, d\tilde{W}(s) . \tag{12}$$

Fix t and let Δ denote a real number so that $t = m\Delta$ for some integer m.

Clearly

(13) $\sigma(\tilde{\xi}^\varepsilon(i\Delta), i\Delta)[\tilde{W}^\varepsilon(i\Delta + \Delta) - \tilde{W}^\varepsilon(i\Delta)] \to \sigma(\tilde{\xi}(i\Delta), i\Delta)[\tilde{W}(i\Delta + \Delta) - \tilde{W}(i\Delta)]$,

w.p. 1 as $\varepsilon \to 0$. Also

(14) $$E \sup_{i\Delta \leq t \leq i\Delta+\Delta} \left|\int_{i\Delta}^t [\sigma(\tilde{\xi}^\varepsilon(t), t) - \sigma(\tilde{\xi}^\varepsilon(i\Delta), i\Delta)]\, d\tilde{W}^\varepsilon(t)\right| \leq 4E \int_{i\Delta}^{i\Delta+\Delta} [\sigma(\tilde{\xi}^\varepsilon(t), t) - \sigma(\tilde{\xi}^\varepsilon(i\Delta), i\Delta)]^2\, ds$$

and

(15) $$\sum_{i=0}^{m-1} E \int_{i\Delta}^{i\Delta+\Delta} [\sigma(\tilde{\xi}^\varepsilon(t), t) - \sigma(\tilde{\xi}^\varepsilon(i\Delta), i\Delta)]^2\, ds \to 0$$

as $\Delta \to 0$, uniformly in ε, (since the multivariate distributions of $\tilde{\xi}^\varepsilon(\cdot)$ do not depend on ε). (13)—(15) imply (12). The limit (12), the zero limit of $\hat{\mathscr{E}}^\varepsilon(\cdot)$, and the fact that $\tilde{X}(\cdot) = \tilde{\xi}(\cdot)$, imply that (6) holds for the weak limit of the originally selected subsequence of $\{\xi^n(\cdot)\}$.

Each subsequence of $\{\xi^n(\cdot)\}$ contains a further subsequence which converges weakly to a process $x(\cdot)$ satisfying (6) for some Wiener process with respect to which $x(\cdot)$ is non-anticipative. By (A3), the distribution of $x(0)$ does not depend on the subsequence and is that of ξ_0. The uniqueness condition (A6) then gives the Theorem.

We have only to show (8). Since by (3), for each ε, the weak limits of $Q^{n,\varepsilon}(\cdot)$ (resp. $\tilde{Q}^{n,\varepsilon}(\cdot)$) and $Z^{n,\varepsilon}(\cdot)$ (resp. $\tilde{Z}^{n,\varepsilon}(\cdot)$) are the same, we need only show that $\tilde{Z}^\varepsilon(\cdot)$, satisfies (8), where $\tilde{\xi}^\varepsilon(\cdot)$ is non-anticipative with respect to $\tilde{W}^\varepsilon(\cdot)$.

The proof of non-anticipativeness follows very closely the proof that $B^\varepsilon(\cdot)$ is a Wiener process in Gikhman and Skorokhod (1969) page 462 (where the scalar $B^{n,\varepsilon}(\cdot)$ process is called $w_n(\cdot)$). Let $0 \leq s_1 \leq \cdots \leq s_p \leq t_1 \leq \cdots \leq t_q \leq T$ for some integers p, q, let $\rho_1, \cdots, \rho_p, \lambda_1, \cdots, \lambda_q$ be arbitrary elements of R^r and write the characteristic function

$$\Gamma^n = E \exp \sum_{j=1}^p \rho_j'\xi^n(s_j) \exp i \sum_{j=1}^q \lambda_j'[B^{n,\varepsilon}(t_{j+1}) - B^{n,\varepsilon}(t_j)] .$$

Following the argument in Gikhman and Skorokhod, and noting that $\xi^n(t)$ is $\mathscr{B}_t^n$ measurable, we get

$$\lim_n \Gamma^n = \lim_n E \exp i \sum_{j=1}^p \rho_j'\xi^n(s_j) E_{\mathscr{B}_t^n} \exp i \sum_{j=1}^q \lambda_j'[B^{n,\varepsilon}(t_{j+1}) - B^{n,\varepsilon}(t_j)] ,$$

where the conditional expectation converges to

$$\exp - \tfrac{1}{2} \sum_{j=1}^q |\lambda_j|^2(t_{j+1} - t_j) ,$$

as $n \to \infty$. Since $\xi^n(\cdot) \to_D \tilde{\xi}^\varepsilon(\cdot)$ and $(\xi^n(\cdot), B^{n,\varepsilon}(\cdot)) \to_D (\tilde{\xi}^\varepsilon(\cdot), \tilde{W}^\varepsilon(\cdot))$,

$$\begin{aligned}\Gamma^n &\to E \exp i \sum_{j=1}^p \rho_j'\tilde{\xi}^\varepsilon(s_j) \exp - \tfrac{1}{2} \sum_{j=1}^q |\lambda_j|^2(t_{j+1} - t_j) \\ &= E \exp i \sum_{j=1}^p \rho_j'\tilde{\xi}(s_j) \exp i \sum_{j=1}^q \lambda_j'[\tilde{W}^\varepsilon(t_{j+1}) - \tilde{W}^\varepsilon(t_j)] ,\end{aligned}$$

which proves that $\tilde{\xi}^\varepsilon(\cdot)$ is non-anticipative with respect to $\tilde{W}^\varepsilon(\cdot)$.

Fix $t \in (0, T)$, and suppose that Δ is a real number such that $t = q\Delta$, for integral q. We will suppose that $\tilde{\xi}^\varepsilon(t)$ is continuous[5] w.p. 1 at $t = i\Delta$, $i = 0, \cdots, q$.

[5] Recall that the proof of the continuity of $\tilde{\xi}(\cdot)$ depended on the representation (8), which we are now proving, so we can only assume right continuity here.

This is convenient but not really necessary. Since $\tilde{\xi}^\varepsilon(\cdot)$ is continuous w.p. 1 at all t, except possibly at a countable number of them, our proof can be adjusted by dividing $[0, T]$ into slightly unequal divisions. Define $m(i, \Delta, n) = \max\{k: t_k^n \leqq t\}$. Then $\xi^n(i\Delta) = \xi^n_{m(i,\Delta,n)}$. Let $I(i, \Delta, n) = \{k: m(i, \Delta, n) \leqq k < m(i+1, \Delta, n)\}$. Then ((16) defines $C_\Delta^{n,\varepsilon}$)

$$(16) \qquad \begin{aligned} C_\Delta^{n,\varepsilon} &= \textstyle\sum_{i=0}^{q-1} \sigma^\varepsilon(\xi^n_{m(i,\Delta,n)}, t^n_{m(i,\Delta,n)}) \sum_{k\in I(i,\Delta,n)} b_k^{n,\varepsilon} \\ &= \textstyle\sum_{i=0}^{q-1} \sigma^\varepsilon(\xi^n(i\Delta), t^n_{m(i,\Delta,n)})[B^{n,\varepsilon}(i\Delta+\Delta) - B^{n,\varepsilon}(i\Delta)] . \end{aligned}$$

As $n \to \infty$, the right side of (16) tends in distribution to

$$(17) \qquad \textstyle\sum_{i=0}^{q-1} \sigma^\varepsilon(\tilde{\xi}^\varepsilon(i\Delta), i\Delta)[\tilde{W}^\varepsilon(i\Delta+\Delta) - \tilde{W}^\varepsilon(i\Delta)] ,$$

which tends to (8) in probability as $\Delta \to 0$, $q \to \infty$, with $\Delta q \equiv t$.

To complete the proof we need a suitable bound on the difference between $C_\Delta^{n,\varepsilon}$ and $Z^{n,\varepsilon}(t)$. Thus we write

$$(18) \qquad \begin{aligned} E|C_\Delta^{n,\varepsilon} &- Z^{n,\varepsilon}(t)|^2 \\ &= E|\textstyle\sum_{i=0}^{q-1} \sum_{k\in I(i,\Delta,n)} [\sigma^\varepsilon(\xi_k^n, t_k^n) - \sigma^\varepsilon(\xi^n(i\Delta), t^n_{m(i,\Delta,n)})]^2 \delta t_k^n| \\ &\leqq E \textstyle\sum_{i=0}^{q-1} |\sum_{k\in I(i,\Delta,n)} [\sigma^\varepsilon(\xi_k^n, t_k^n) - \sigma^\varepsilon(\xi^n(i\Delta), t^n_{m(i,\Delta,n)})]^2 \delta t_k^n \\ &\quad - \textstyle\int_{i\Delta}^{i\Delta+\Delta} [\sigma^\varepsilon(\xi^n(s), s) - \sigma^\varepsilon(\xi^n(i\Delta), i\Delta)]^2 \, ds| \\ &\quad + \textstyle\sum_{i=0}^{q-1} E \int_{i\Delta}^{i\Delta+\Delta} [\sigma^\varepsilon(\xi^n(s), s) - \sigma^\varepsilon(\xi^n(i\Delta), i\Delta)]^2 \, ds . \end{aligned}$$

The argument of a typical term of the first expectation on the right-hand side of (18) has the distribution of

$$H_{n,i\Delta} = A_{n,i\Delta} - B_{n,i\Delta} ,$$

where

$$\begin{aligned} A_{n,i\Delta} &= \textstyle\sum_{k\in I(i,\Delta,n)} [\sigma^\varepsilon(\tilde{\xi}^n(t_k^n), t_k^n) - \sigma^\varepsilon(\tilde{\xi}^n(i\Delta), t^n_{m(i,\Delta,n)})]^2 \delta t_k^n \\ B_{n,i\Delta} &= \textstyle\int_{i\Delta}^{i\Delta+\Delta} [\sigma^\varepsilon(\tilde{\xi}(s), s) - \sigma^\varepsilon(\tilde{\xi}^n(i\Delta), i\Delta)]^2 \, ds . \end{aligned}$$

The sequence $\{\tilde{\xi}^n(\cdot)\}$ is tight. Thus for each $\delta > 0$, there is a compactum $A_\delta \subset D^m$ for which $P\{\tilde{\xi}^n(\cdot) \notin A_\delta\} \leqq \delta$, all n. A necessary and sufficient condition (Billingsley (1968) Theorem 14.3) for compactness of a set $A \in D^m$ is that

$$\sup_{x(\cdot)\in A} \sup_{0\leqq t\leqq T} |x(t)| < \infty$$
$$\lim_{\delta_1\to 0} \sup_{x(\cdot)\in A} \inf_{\{t_i\}} \max_i w_x[t_{i-1}, t_i) = 0$$

where, for each fixed δ_1, the inf is taken over all finite sets $\{t_i\}$ so that $0 < t_1 \leqq t_2 \cdots \leqq T$ and $t_{i+1} - t_i \geqq \delta_1$, and where we define

$$w_x[a, b) = \sup_{a\leqq s<t\leqq b} |x(s) - x(t)| .$$

These facts, together with the continuity of $\sigma^\varepsilon(\cdot, \cdot)$, and the facts that $\max_k |\delta t_k^n| \to 0$ and that $\tilde{\xi}^n(i\Delta)$ converges w.p. 1 (to $\tilde{\xi}^\varepsilon(i\Delta)$, since $\tilde{\xi}^\varepsilon(\cdot)$ is continuous w.p. 1 at $t = i\Delta$), imply that for any real $\varepsilon_1 > 0$, $\varepsilon_2 > 0$, there is an integer $n_0 < \infty$ so that $|A_{n,i\Delta} - B_{n,i\Delta}| \leqq \varepsilon_1$ with probability[6] $\geqq 1 - \varepsilon_2$ for $n \geqq n_0$.

[6] I.e., for n sufficiently large ($\geqq n_0$) the integral is uniformly (ε_1) approximated by the sum, for ω in a set whose probability is $\geqq 1 - \varepsilon_2$.

Thus $H_{n,i\Delta}$ tends to zero in probability as $n \to \infty$. Hence, by boundedness of $\sigma^\varepsilon(\cdot, \cdot)$, the expectation of $H_{n,i\Delta}$ tends to zero, as does the first term on the right side of (18), as $n \to \infty$.

Denote the last term on the right of (18) by $L_\varepsilon^{n,\Delta}$. The functional

$$\int_{i\Delta}^{i\Delta+\Delta} [\sigma^\varepsilon(X(s), s) - \sigma^\varepsilon(X(i\Delta), i\Delta)]^2\, ds$$

is bounded and continuous almost everywhere in $(D^r, \mathscr{D}^r)$ with respect to the measure of $\tilde{\xi}^\varepsilon(\cdot)$ on $(D^r, \mathscr{D}^r)$, since $\tilde{\xi}^\varepsilon(\cdot)$ is assumed continuous w.p. 1 at $t = i\Delta$. Thus, since $\xi^n(\cdot) \to_D \tilde{\xi}^\varepsilon(\cdot)$ as $n \to \infty$, $L_\varepsilon^{n,\Delta}$ converges to (as $n \to \infty$),

$$\sum_{i=0}^{q-1} E \int_{i\Delta}^{i\Delta+\Delta} [\sigma^\varepsilon(\tilde{\xi}(s), s) - \sigma^\varepsilon(\tilde{\xi}(i\Delta), i\Delta)]^2\, ds\,,$$

which in turn tends to zero as $\Delta \to 0$, since $\tilde{\xi}(\cdot)$ is right continuous. These arguments imply that as $n \to \infty$, and then $\Delta \to 0$, the left hand of (18) goes to zero. This, together with the convergence of (17) to (8) as $\Delta \to 0$, implies that $\tilde{Q}^\varepsilon(t)$ has the representation (8), as asserted. □

6. Examples from population genetics. We first consider the scalar problem dealt with by Feller (1950). There is a population of $2n$ genes of two types. Suppose that in the current generation there are j of type **a**, and $2n - j$ of type A. The new generation, also of size $2n$ is selected by $2n$ trials with replacement, each trial yielding an **a** with probability $p_j = j/2n$ and an A with probability $q_j = 1 - j/2n$. The sequence of numbers of type **a** in successive generations form a Markov chain with transition probabilities $p_{jl} = \binom{2n}{l} p_j^l q_j^{2n-l}$. Define ξ_k^n to be the number of **a**'s in the kth generation divided by $2n$; i.e., the probability that each selection (trial) taken to form the next $(k+1)$st generation yields an **a** with probability ξ_k^n.

We have

$$E_{\mathscr{B}_k^n}[\xi_{k+1}^n - \xi_k^n] = 0$$
$$E_{\mathscr{B}_k^n}|\xi_{k+1}^n - \xi_k^n|^2 = \xi_k^n(1 - \xi_k^n)/2n$$
$$E_{\mathscr{B}_k^n}|\xi_{k+1}^n - \xi_k^n|^4 \leqq K(1/2n)^2$$

for a real K. Interpolate by defining $t_k^n = k/2n$, $\delta t_k^n = 1/2n$, $\xi^n(\cdot)$ by $\xi^n(t) = \xi_k^n$ for $t \in [k/2n, (k+1)/2n)$. Feller (1950) proves that the characteristic function of $\xi^n(t)$ converges (as the population size $2n$ increases to ∞) to that of the solution to the stochastic differential equation

$$x(t) = x + \int_0^t [x(s)(1 - x(s))]^{\frac{1}{2}}\, dw(s)\,, \tag{19}$$

which (as seen below) has a unique solution, with absorption at $\{1, 0\}$. $\sigma(x)$ does not satisfy a uniform Lipschitz condition. Our Theorem 1 is directly applicable and yields weak convergence of the processes $\xi^n(\cdot)$ to $x(\cdot)$ on any finite time interval, a strong result. In addition, our method is applicable to many other problems arising in genetics, where there is a drift due to selective advantages, where there is mutation, where the selection rule depends on time, etc.

To prove the uniqueness assertion, suppose that the pair $(\tilde{x}(t), \tilde{w}(t))$ satisfy (19) also, where $\tilde{w}(t)$ is a standard Wiener process, and $\tilde{x}(0) = x(0) = x$. Define τ_ε (resp. $\tilde{\tau}_\varepsilon$) as the first time that $x(t)$ (resp. $\tilde{x}(t)$) is either $\leqq \varepsilon$ or $\geqq 1 - \varepsilon$, $\varepsilon > 0$. Then the two processes defined by $x^\varepsilon(t) \equiv x(t \cap \tau_\varepsilon)$, $\tilde{x}^\varepsilon(t) \equiv \tilde{x}(t \cap \tilde{\tau}_\varepsilon)$ are Markov processes, and they have the same multivariate distributions. Uniqueness of the multivariate distributions follows, since $x^\varepsilon(t) \to x(t)$, $\tilde{x}^\varepsilon(t) \to \tilde{x}(t)$ for all t, w.p. 1.

The next example, following Crow and Kimura ((1970) Section 5.6, Chapter 8), and involving selective advantages due to fertility difference between the a and A, was suggested to the author by Wendell Fleming. Let h and θ denote positive real numbers, fix the (gamete) population size at $2n$, suppose that in the current generation there are j gametes of type **a** ($2n - j$ of type A), and define $p_j = j/2n$. Let the p_{ij} be defined by

$$p_{jl} = \binom{2n}{l}\tilde{p}_j{}^l\tilde{q}_j{}^l, \qquad \tilde{q}_j \equiv 1 - \tilde{p}_j,$$

where we define

$$\tilde{p}_j = p_j + f_n(p_j)/2n$$

and

$$f_n(p) = \frac{\theta pq(q + h(p - q))}{1 - (2n)^{-1}\theta q(2kp + q)}, \qquad q = 1 - p.$$

The parameters h and θ are used to represent differences in fertility among the genotypes. We have ($\xi_k{}^n$ again denotes the number of gametes of type **a** in the kth generation)

$$E_{\mathscr{B}_k{}^n}[\xi^n_{k+1} - \xi_k{}^n] = f_n(\xi_k{}^n)/2n$$
$$\text{Cov}_{\mathscr{B}_k{}^n}[\xi^n_{k+1} - \xi_k{}^n] = \sigma_n{}^2(\xi_k{}^n)/2n$$

where

$$\sigma_n{}^2(p) = \left(p + \frac{f_n(p)}{2n}\right)\left(q - \frac{f_n(p)}{2n}\right),$$

and, for a real K,

$$E\left|\xi^n_{k+1} - \xi_k{}^n - \frac{f_n(\xi_K{}^n)}{2n}\right|^4 \leqq K(1/2n)^2$$

$f_n(p)$ and $\sigma_n{}^2(p)$ converge uniformly (for $p \in [0, 1]$) to $\theta pq(q + h(p - q))$ and pq, resp., and Theorem 1 is applicable.

REFERENCES

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

[2] CROW, J. F. and KIMURA, M. (1970). *An Introduction to Population Genetics Theory.* Harper and Row, New York.

[3] DOOB, J. L. (1953). *Stochastic Processes.* Wiley, New York.

[4] FELLER, W. (1950). Diffusion processes in genetics. *Proc. Second Berkeley Symp. Math. Statist Prob.* Univ. of California Press, 227–246.

[5] GIKHMAN, I. I. and SKOROKHOD, A. V. (1969). *Introduction to the Theory of Stochastic Processes.* Saunders, Philadelphia (in trans.).

[6] Skorokhod, A. V. (1956). Limit theorems for stochastic processes. *Theor. Prob. Appl.* **1** 261–290 (trans.).

Division of Applied Mathematics
Center for Dynamical Systems
Brown University
Providence, Rhode Island 02912