*The Annals of Probability*
1975, Vol. 3, No. 2, 358-364

# DIFFUSION APPROXIMATION OF NON-MARKOVIAN PROCESSES

BY M. FRANK NORMAN

*University of Pennsylvania*

General diffusion approximation theorems are established for sequences of non-Markovian processes. These theorems cover certain genetic models previously considered by Watterson. It follows that Watterson's conclusions concerning these models are correct, even though there is a gap in his proof.

**1. Introduction.** Let $I = [d_0, d_1]$ be a closed bounded interval. For each $N \geqq 1$, let $\{X_n^N, n \geqq 0\}$ be a stochastic process in $I$, adapted to an increasing sequence $\{\mathscr{F}_n^N, n \geqq 0\}$ of $\sigma$-fields. The processes need not be Markovian. The conditional moments of $\Delta X_n^N = X_{n+1}^N - X_n^N$ are supposed to satisfy conditions of the form

$$\begin{aligned} E(\Delta X_n^N \mid \mathscr{F}_n^N) &= \tau_N a(X_n^N) + e_{1,n}^N , \\ E((\Delta X_n^N)^2 \mid \mathscr{F}_n^N) &= \tau_N b(X_n^N) + e_{2,n}^N , \\ E(|\Delta X_n^N|^3 \mid \mathscr{F}_n^N) &= e_{3,n}^N , \end{aligned} \tag{1}$$

where $\tau_N > 0$ and $\tau_N \to 0$ as $N \to \infty$, and the error terms $e_{i,n}^N$ are $o(\tau_N)$ in the sense that, for any $t < \infty$,

$$\textstyle\sum_{n<[t/\tau_N]} E(|e_{i,n}^N|) \to 0 \qquad \text{as } N \to \infty . \tag{2}$$

Let $X^N(t) = X_{[t/\tau_N]}^N$, let $0 \leqq t_1 < t_2 < \cdots < t_k$, and let $\Rightarrow$ denote convergence in distribution. Our main result, Theorem 1, gives conditions on $a$ and $b$ that insure that $(X^N(t_1), \cdots, X^N(t_k)) \Rightarrow (X(t_1), \cdots, X(t_k))$ as $N \to \infty$ and $X^N(0) \Rightarrow X(0)$, where $X(t)$ is a diffusion whose transition kernel, $P(t; x, A) = P(X(t) \in A \mid X(0) = x)$, satisfies the following conditions, which are analogous to (1):

$$\begin{aligned} \textstyle\int_I (y - x)P(\tau; x, dy) &= \tau a(x) + e_1(\tau, x) , \\ \textstyle\int_I (y - x)^2 P(\tau; x, dy) &= \tau b(x) + e_2(\tau, x) , \\ \textstyle\int_I |y - x|^3 P(\tau; x, dy) &= e_3(\tau, x) . \end{aligned} \tag{3}$$

Here the error terms $e_i(\tau, x)$ are $o(\tau)$ in the uniform sense:

$$\sup_{x \in I} |e_i(\tau, x)|/\tau \to 0 \qquad \text{as } \tau \to 0 . \tag{4}$$

Let $C^j$ be the set of functions with $j$ continuous derivatives throughout $I$. At the boundaries $d_i$ these derivatives are one-sided.

THEOREM 1. *Suppose that $a \in C^3$, $a(d_0) \geqq 0$, and $a(d_1) \leqq 0$; and that $b$ admits a factorization $b(x) = \sigma_0(x)\sigma_1(x)$, where $\sigma_i$ satisfies the following conditions:* $\sigma_i \in C^3$,

Received January 23, 1974; revised June 10, 1974.

*AMS* 1970 *subject classifications.* Primary 60B10, 60J60, 60J70, 92A10.

*Key words and phrases.* Diffusion approximation, non-Markovian processes, genetic models.

*$\sigma_i(d_i) = 0$, $\sigma_i(x) > 0$ for $d_0 < x < d_1$, $p(x) = \sigma_0(x)/(\sigma_0(x) + \sigma_1(x))$ is non-decreasing for $d_0 < x < d_1$, and, letting $p(d_i) = \lim_{x \to d_i} p(x)$, $p \in C^3$. Then there is a unique transition kernel $P$ satisfying (3) and (4). If $X^N(0) \Rightarrow X(0)$ as $N \to \infty$, then $(X^N(t_1), \cdots, X^N(t_k)) \Rightarrow (X(t_1), \cdots, X(t_k))$ as $N \to \infty$.*

Though the condition on $b$ in Theorem 1 is undesirably complicated, it is not very restrictive in applications. The condition implies that $b \in C^3$, $b(d_i) = 0$, and $b(x) > 0$ for $d_0 < x < d_1$. If, conversely, $b$ is analytic throughout $I$ with $b(d_i) = 0$ and $b(x) > 0$ on the interior of $I$, the condition of Theorem 1 is satisfied. For let $j_i$ be the order of the zero at $d_i$. Then $h(x) = b(x)(x - d_0)^{-j_0}(d_1 - x)^{-j_1}$ is analytic and positive throughout $I$, so $h(x)^{\frac{1}{2}} \in C^3$. Thus we can take $\sigma_0(x) = (x - d_0)^{j_0} h(x)^{\frac{1}{2}}$ and $\sigma_1(x) = (d_1 - x)^{j_1} h(x)^{\frac{1}{2}}$. In the genetic applications considered in Section 4, $I = [0, 1]$, $a$ is a polynomial, and $b(x) = cx(1 - x)$, where $c > 0$.

Theorem 1 is an extension of Theorem 9.1.1 of [5] to non-Markovian processes. The latter theorem includes the existence and uniqueness of $P$. It is noteworthy that the uniformity condition (4) can be used in place of conventional boundary conditions to determine $P$ uniquely. The proof of Theorem 9.1.1 shows that the equation

$$T_t f(x) = \int_I f(y) P(t; x, dy)$$

defines a strongly continuous conservative semigroup on $C = C^0$ (supremum norm). Let $\Gamma$ be the generator of this semigroup, and let $\mathscr{D}(\Gamma)$ be the domain of $\Gamma$. Then $T_t$ is the unique strongly continuous conservative semigroup on $C$ for which $\mathscr{D}(\Gamma) \supset C^2$ and

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

for $f \in C^2$ and $x \in I$ ([5] page 150). In terms of the Feller boundary theory, exit boundaries are adhesive ($\Gamma f(d_i) = 0$ for $f \in \mathscr{D}(\Gamma)$) and regular boundaries are reflecting ($(d/dp)f(d_i) = 0$ for $f \in \mathscr{D}(\Gamma)$, where $dp(x) = e^{-B(x)}\, dx$ and $dB(x) = 2a(x) b(x)^{-1}\, dx$) ([5] page 148).

The possibility and desirability of extending Theorem 9.1.1 of [5] to non-Markovian processes was suggested by an interesting paper of Watterson [6]. Watterson's result is similar to the special case of Theorem 1 corresponding to $k = 1$, $b(x) = cx(1 - x)$ and $a(x)$ a third degree polynomial. Unfortunately, there is a large gap in Watterson's proof. Let $F_N(x, n) = P(X_n^N \leqq x)$ and $F(x, u) = P(X(u) \leqq x)$. The transition from

$$\lim_{N \to \infty} \int_0^\infty e^{-\alpha u} F_N(x, [N^m u])\, du = \int_0^\infty e^{-\alpha u} F(x, u)\, du$$

to

$$\lim_{N \to \infty} F_N(x, [N^m u]) = F(x, u) \tag{5}$$

at the bottom of page 950 of [6] cannot be justified by "the uniqueness theorem for Laplace transforms." (Watterson's $N^m$ corresponds to our $\tau_N^{-1}$.) Standard continuity theorems ([3] page 433) yield, not (5), but the integrated version

$$\lim_{N \to \infty} \int_0^t F_N(x, [N^m u])\, du = \int_0^t F(x, u)\, du .$$

Watterson's paper is oriented toward applications to two genetic models, one with overlapping generations, the other with non-overlapping generations. These applications are described in detail in a later paper [7]. In Section 4 we will show that these models fall within the scope of Theorem 1. Guess [4], using Watterson's result as a lemma, proved weak convergence of the distribution of the process $\{X^N(t), t \geqq 0\}$ for a class of models that includes the non-overlapping generation model. Theorem 2 of Section 3 is a weak convergence theorem that complements Theorem 1 and applies to both of the models considered by Watterson.

**2. Proof of Theorem 1.** We first show that, for $n \geqq m$ and $f \in C$,

$$E[|E(f(X_n) \mid \mathscr{F}_m) - T_{(n-m)\tau} f(X_m)|] \leqq \sum_{j=m}^{n-1} E[|E(g_{j+1}(X_{j+1}) \mid \mathscr{F}_j) - T_\tau g_{j+1}(X_j)|], \tag{6}$$

where $g_j = T_{(n-j)\tau} f$ and $N$'s have been suppressed. Clearly

$$f(X_n) - T_{(n-m)\tau} f(X_m) = g_n(X_n) - g_m(X_m) = \sum_{j=m}^{n-1} (g_{j+1}(X_{j+1}) - g_j(X_j)) .$$

Hence

$$\begin{aligned} E(f(X_n) \mid \mathscr{F}_m) - T_{(n-m)\tau} f(X_m) &= \sum_{j=m}^{n-1} E[g_{j+1}(X_{j+1}) - g_j(X_j) \mid \mathscr{F}_m] \\ &= \sum_{j=m}^{n-1} E[E(g_{j+1}(X_{j+1}) \mid \mathscr{F}_j) - g_j(X_j) \mid \mathscr{F}_m] . \end{aligned}$$

Taking absolute values and expectations on both sides of this equality, and noting that $g_j = T_\tau g_{j+1}$, we obtain (6).

Let $L$ be the subspace of $C^2$ consisting of those functions whose second derivatives satisfy the Lipschitz condition

$$M(g'') = \sup_{x \neq y} \frac{|g''(x) - g''(y)|}{|x - y|} < \infty .$$

For $g \in L$, let

$$||g|| = |g'|_\infty + |g''|_\infty + M(g'') ,$$

where $|\cdot|_\infty$ is the supremum norm. Any function $g$ in $L$ possesses a Taylor expansion

$$g(y) = g(x) + (y - x)g'(x) + 2^{-1}(y - x)^2 g''(x) + \lambda |y - x|^3 M(g'') ,$$

where $|\lambda| \leqq \frac{1}{6}$. For the remainder in the first order Taylor expansion is

$$\begin{aligned} g(y) - g(x) - \delta g'(x) &= \delta^2 \int_0^1 (1 - s) g''(x + s\delta)\, ds \\ &= 2^{-1}\delta^2 g''(x) + \delta^2 \int_0^1 (1 - s)(g''(x + s\delta) - g''(x))\, ds , \end{aligned}$$

where $\delta = y - x$, and the last term on the right has absolute value at most

$$|\delta|^3 M(g'') \int_0^1 (1 - s) s\, ds = 6^{-1} |\delta|^3 M(g'') .$$

Hence, in view of (1),

$$E(g(X_{j+1}) \mid \mathscr{F}_j) = g(X_j) + \tau \Gamma g(X_j) + \lambda ||g|| \sum_{i=1}^3 |e_{i,j}| ,$$

where $|\lambda| \leqq 1$. By (3), $T_\tau g$ has a similar expansion, so

$$(7) \qquad E[|E(g(X_{j+1}) \mid \mathscr{F}_j) - T_\tau g(X_j)|] \leqq ||g|| \sum_{i=1}^{3} [E(|e_{i,j}|) + \sup_{x \in I} |e_i(\tau, x)|] .$$

The following lemma is a by-product of the proof of Theorem 9.1.1 (see [5] (3.8) page 150). We shall have more to say about it at the end of the section.

LEMMA 1. *$T_t$ maps $C^3$ into $L$. Moreover, for any $f \in C^3$ and $K < \infty$, $\sup_{t \leqq K} ||T_t f|| < \infty$.*

Applying (7) to $g_{j+1} = T_{(n-j-1)\tau} f$ for $f \in C^3$, using Lemma 1 to estimate $||g_{j+1}||$, and combining the result with (6), we obtain

$$(8) \qquad \begin{aligned} &E[|E(f(X_n^N) \mid \mathscr{F}_m^N) - T_{(n-m)\tau} f(X_m^N)|] \\ &\qquad \leqq K' \sum_{i=1}^{3} [(n-m) \sup_{x \in I} |e_i(\tau_N, x)| + \textstyle\sum_{j=m}^{n-1} E(|e_{i,j}^N|)] \end{aligned}$$

for some constant $K'$, provided that $(n-1)\tau_N \leqq K$.

Suppose now that $0 \leqq s \leqq t$ and let $n = [t/\tau_N]$ and $m = [s/\tau_N]$. As a consequence of (2) and (4), the quantities on the right and left in (8) approach 0 as $N \to \infty$. But $(n-m)\tau \to t - s$ as $N \to \infty$, and, as noted in Section 1, the semigroup $T_t$ on $C$ is strongly continuous with respect to the supremum norm, so $T_{(n-m)\tau} f(x) \to T_{t-s} f(x)$, uniformly over $x$, as $N \to \infty$. Thus

$$(9) \qquad E[|E(f(X^N(t)) \mid \mathscr{F}^N(s)) - T_{t-s} f(X^N(s))|] \to 0$$

as $N \to \infty$, where $\mathscr{F}^N(s) = \mathscr{F}_{[s/\tau]}^N$. Since $C^3$ is dense in $C$, (9) holds for all $f \in C$. It follows from (9) that

$$\begin{aligned} &E(f(X^N(t))) - E(T_{t-s} f(X^N(s))) \\ &\qquad = E[E(f(X^N(t)) \mid \mathscr{F}^N(s)) - T_{t-s} f(X^N(s))] \to 0 . \end{aligned}$$

Taking $s = 0$, and noting that $T_t f \in C$, so that

$$\begin{aligned} E(T_t f(X^N(0))) &\to E(T_t f(X(0))) \\ &= E(f(X(t))) , \end{aligned}$$

we obtain $E(f(X^N(t))) \to E(f(X(t)))$ as $N \to \infty$.

Suppose, inductively, that $(X^N(t_1), \cdots, X^N(t_k)) \Rightarrow (X(t_1), \cdots, X(t_k))$ for some $k \geqq 1$. If $f_i \in C$ for $1 \leqq i \leqq k+1$, and $0 \leqq t_1 < t_2 < \cdots < t_{k+1}$, then

$$\begin{aligned} E[\textstyle\prod_{i=1}^{k+1} f_i(X^N(t_i))] &= E[\textstyle\prod_{i=1}^{k} f_i(X^N(t_i)) E(f_{k+1}(X^N(t_{k+1})) \mid \mathscr{F}^N(t_k))] \\ &= E[\textstyle\prod_{i=1}^{k} f_i(X^N(t_i)) T_{t_{k+1}-t_k} f_{k+1}(X^N(t_k))] + \delta_N , \end{aligned}$$

where $\delta_N \to 0$ by (9),

$$\to E[\textstyle\prod_{i=1}^{k} f_i(X(t_i)) T_{t_{k+1}-t_k} f_{k+1}(X(t_k))] ,$$

by the induction hypothesis,

$$= E[\textstyle\prod_{i=1}^{k+1} f_i(X(t_i))] .$$

Thus $k$ in the induction hypothesis can be replaced by $k+1$, and the proof of Theorem 1 is complete.

Since the boundedness assertion of Lemma 1 is the key to the proof, it is worthwhile to review briefly how this boundedness was established in [5]. This review also gives some insight concerning the origin of the assumption concerning $b$ in Theorem 1. Let $V_\tau f(x) = E(f(X_{n+1}^\tau) \mid X_n^\tau = x)$ for the discrete parameter Markov process $X_n^\tau$ that moves from $x$ to $x + \tau a(x) + \tau^{\frac{1}{2}}\sigma_1(x)$ with probability $p(x) = \sigma_0(x)/(\sigma_0(x) + \sigma_1(x))$ and from $x$ to $x + \tau a(x) - \tau^{\frac{1}{2}}\sigma_0(x)$ with probability $1 - p(x)$. It was shown by direct calculation that there is a constant $\gamma$ such that $\|V_\tau^n f\| \leqq e^{\gamma n\tau}\|f\|$ for $\tau > 0$, $n \geqq 0$, and $f \in C^3$ ([5] Lemma 2.2, page 142). But $V_\tau^n f \to T_t f$ uniformly as $\tau \to 0$ and $n\tau \to t$, and it follows that $\|T_t f\| \leqq e^{\gamma t}\|f\|$. We remark that it would be desirable to have an alternative proof of Lemma 1 based on semigroup and differential equation theory. For an illustration of such methods in a similar context, see [2].

**3. Weak convergence of the distribution of** $X^N(\cdot)$. For any $K > 0$, let $D_K$ be the space of real-valued functions on $[0, K]$ that are right-continuous and have left-hand limits. Let $D_K$ be equipped with the Skorohod $J_1$ topology ([1] Section 14), and let $\Rightarrow$ denote convergence in distribution for random elements of $D_K$.

THEOREM 2. *Suppose that the hypotheses of Theorem* 1 *hold, and that, in addition, there are constants* $G_i$ *such that*

$$|E(\Delta X_n^N \mid \mathscr{F}_n^N)| \leqq G_1 \tau_N \tag{10}$$

*and*

$$\mathrm{Var}\,(\Delta X_n^N \mid \mathscr{F}_n^N) \leqq G_2 \tau_N \tag{11}$$

a.s., *for* $N \geqq 1$ *and* $n \geqq 0$. *Then, for any* $K > 0$, $\{X^N(t), t \leqq K\} \Rightarrow \{X(t), t \leqq K\}$ *as* $N \to \infty$.

PROOF. According to Theorem 15.6 of [1], it suffices to show that there is a constant $H = H_K$ such that

$$E[(X^N(t) - X^N(t_1))^2(X^N(t_2) - X^N(t))^2] \leqq H(t_2 - t_1)^2$$

for all $0 \leqq t_1 \leqq t \leqq t_2 \leqq K$. For this it is sufficient that

$$E[(X_n^N - X_m^N)^2 \mid \mathscr{F}_m^N] \leqq H'(n - m)\tau \tag{12}$$

for $0 \leqq m \leqq n \leqq K/\tau_N$.

Following Guess ([4] page 294), we write

$$X_n - X_m = \textstyle\sum_{j=m}^{n-1} V_j + \sum_{j=m}^{n-1} W_j\,, \tag{13}$$

where

$$V_j = E(\Delta X_j \mid \mathscr{F}_j)$$

and

$$W_j = \Delta X_j - E(\Delta X_j \mid \mathscr{F}_j)\,.$$

By (10),

$$\begin{aligned}(\textstyle\sum_{j=m}^{n-1} V_j)^2 &\leqq G_1^2(n - m)^2\tau^2 \\ &\leqq G_1^2 K(n - m)\tau\,.\end{aligned} \tag{14}$$

By (11)

$$E(W_j^2 \mid \mathscr{F}_m) = E(\mathrm{Var}\,(\Delta X_j \mid \mathscr{F}_j) \mid \mathscr{F}_m) \leqq G_2 \tau ,$$

hence

$$E((\textstyle\sum_{j=m}^{n-1} W_j)^2 \mid \mathscr{F}_m) = \sum_{j=m}^{n-1} E(W_j^2 \mid \mathscr{F}_m) \leqq G_2(n-m)\tau . \tag{15}$$

Combining (13), (14), and (15) we obtain (12).

**4. Two genetic models.** The models considered by Watterson [7] are relevant to a population of $N$ diploid individuals. Of these, $N_1$ are males and $N_2$ are females. The three genotypes, aa, aA, and AA have frequencies $k$, $N_1 - k - l$, and $l$ among males, and $r$, $N_2 - r - s$, and $s$ among females. The successive values of the vector $(k, l, r, s)$ form a Markov process in both models. However, interest centers on the average

$$X = 2^{-1} + 4^{-1}N_1^{-1}(k-l) + 4^{-1}N_2^{-1}(r-s)$$

of the relative frequencies of the a gene in the two sexes, and the trajectory of this variate is non-Markovian.

Variations in $X$ are controlled by two mutation parameters, $\alpha_1$ and $\alpha_2$, two selection parameters, $\nu_1$ and $\nu_2$, and a nonrandom-mating parameter $f$. The latter is fixed as $N \to \infty$, but the former are assumed to be inversely proportional to $N$; $N\alpha_i = \bar{\alpha}_i \geqq 0$ and $N\nu_i = \bar{\nu}_i$ are constant. Moreover, $N_i/N = r_i > 0$ is fixed. Let $X_n^N$ be the value of $X$ after $n$ steps, and let $\mathscr{F}_n^N$ be the $\sigma$-field generated by the values of $k$, $l$, $r$, and $s$ after $j$ steps, $j \leqq n$.

The significance of a single step is different in the two models. One is of the Moran type, with overlapping generations. Each step of the process corresponds to the death of a single individual and the birth of another. The other model is of the Wright–Fisher type. Generations are non-overlapping and the entire population is replaced at each step of the process. The step-size parameters for the overlapping and non-overlapping generation models are, respectively, $\tau_N = N^{-2}$ and $\tau_N = N^{-1}$.

In both models, the function $a$ of Theorem 1 is

$$a(x) = \bar{\alpha}_2(1-x) - \bar{\alpha}_1 x - x(1-x)\{\bar{\nu}_1[(1-f)x + f] + \bar{\nu}_2[(1-f)(1-x) + f]\} .$$

For the overlapping generation model,

$$b(x) = 4^{-1}(r_1^{-1} + r_2^{-1})(1+f)x(1-x) ,$$

while $b(x)$ is half the quantity on the right for the other model. Lemmas 2 and 3 give estimates of the error terms $e_{i,n}^N$ in (1).

LEMMA 2. *For the overlapping generation model,*

$$E(|e_{i,n}^N|) \leqq c\tau_N^{\frac{3}{2}} + c' e^{-n/2N}\tau_N$$

*for* $i = 1$ *and* 2, $n \geqq 0$, *and all* $N$. *Also*

$$E(|e^N_{3,n}|) \leqq c\tau_N^{\frac{3}{2}} .$$

LEMMA 3. *For the non-overlapping generation model,* $|e^N_{i,n}| \leqq G\tau_N$ a.s. *for* $i = 1$ *and* 2 *and* $n \geqq 0$, $E(|e^N_{1,n}|) \leqq c\tau_N^{\frac{3}{2}}$ *for* $n \geqq 2$, $E(|e^N_{2,n}|) \leqq c\tau_N^{2}$ *for* $n \geqq 1$, *and* $E(|e^N_{3,n}|) \leqq c\tau_N^{\frac{3}{2}}$ *for* $n \geqq 0$. *In all cases these estimates hold uniformly over* $N$.

These estimates can be obtained by lengthy but, for the most part, straightforward calculations, using the suggestions in [7]. It follows immediately from these estimates that (2) is satisfied. Moreover it is very easy to show that (10) and (11) hold. Thus Theorems 1 and 2 apply to both models. We conclude that the gap in Watterson's proof does not alter the correctness of his conclusion that diffusion approximation is applicable to these models.

## REFERENCES


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

[2] BREZIS, H., ROSENKRANTZ, W., and SINGER, B. (1971). On a degenerate elliptic-parabolic equation occurring in the theory of probability. *Comm. Pure Appl. Math.* **14** 395–416.

[3] FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications* **2** 2nd ed. Wiley, New York.

[4] GUESS, H. A. (1973). On the weak convergence of Wright-Fisher models. *Stochastic Proc. Appl.* **1** 287–306.

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

[6] WATTERSON, G. A. (1962). Some theoretical aspects of diffusion theory in population genetics. *Ann. Math. Statist.* **33** 939–957.

[7] WATTERSON, G. A. (1964). The application of diffusion theory to two population genetic models of Moran. *J. Appl. Probability.* **1** 233–246.



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