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 \( I = [d_i, d_j] \) be a closed bounded interval. For each \( N \geq 1 \), let \( \{X^N_n, n \geq 0\} \) be a stochastic process in \( I \), adapted to an increasing sequence \( \{\mathcal{F}^N_n, n \geq 0\} \) of \( \sigma \)-fields. The processes need not be Markovian. The conditional moments of \( \Delta X^N_n = X^N_n - X^N_{n-1} \) are supposed to satisfy conditions of the form

\[
E(\Delta X^N_n \mid \mathcal{F}^N_n) = \tau N a(X^N_n) + \varepsilon^N_{1,n},
\]

\[
E((\Delta X^N_n)^2 \mid \mathcal{F}^N_n) = \tau N b(X^N_n) + \varepsilon^N_{2,n},
\]

\[
E(|\Delta X^N_n|^3 \mid \mathcal{F}^N_n) = \varepsilon^N_{3,n},
\]

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

\[
\sum_{n < t/\tau_N} E(|\varepsilon^N_{i,n}|) \to 0 \quad \text{as} \quad N \to \infty.
\]

Let \( X^n(t) = X^n_{\lfloor t \tau_N \rfloor} \), let \( 0 \leq t_1 < t_2 < \cdots < t_m \), and let \( \Rightarrow \) denote convergence in distribution. Our main result, Theorem 1, gives conditions on \( a \) and \( b \) that ensure that \( (X^n(t_1), \ldots, X^n(t_m)) \Rightarrow (X(t_1), \ldots, X(t_m)) \) 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):

\[
\int_{t_i} (y - x) P(t; x, dy) = \tau a(x) + \varepsilon_i(t, x),
\]

\[
\int_{t_i} (y - x)^2 P(t; x, dy) = \tau b(x) + \varepsilon_i(t, x),
\]

\[
\int_{t_i} |y - x|^3 P(t; x, dy) = \varepsilon_i(t, x).
\]

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

\[
\sup_{x \in I} |\varepsilon_i(t, x)| / \tau \to 0 \quad \text{as} \quad \tau \to 0.
\]

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 \( a \in C^3 \), \( a(d_i) = 0 \), and \( a(d_j) \leq 0 \); and that \( b \) admits a factorization \( b(x) = a_i(x) a_i(x) \), where \( a_i \) satisfies the following conditions: \( a_i \in C^3 \).
\[ \sigma(d) = 0, \sigma(x) > 0 \text{ for } d_x < x < d_1, p(x) = \frac{\sigma(x)}{(\sigma(x) + \sigma'(x))} \text{ is non-decreasing for } d_x < x < d_1, \text{ and, letting } p(d) = \lim_{x \to d} p(x), p \in C^3. \text{ Then there is a unique transition kernel } P \text{ satisfying (3) and (4). If } X^\alpha(0) = X(0) \text{ as } N \to \infty, \text{ then } (X^\alpha(t_1), \cdots, X^\alpha(t_k)) \Rightarrow (X(t_1), \cdots, X(t_k)) \text{ as } N \to \infty. \]

Though the condition on \( b \) in Theorem 1 is undesirably complicated, it is not very restrictive in applications. The condition implies that \( b \in C^3, b(d) = 0, \text{ and } b(x) > 0 \text{ for } d_x < x < d_1. \) If, conversely, \( b \) is analytic throughout \( I \) with \( b(d) = 0 \text{ and } b(x) > 0 \text{ on the interior of } I, \) the condition of Theorem 1 is satisfied. For let \( f \), be the order of the zero at \( d \). Then \( h(x) = b(x)(x - d)^{-f} \) is analytic and positive throughout \( I, \) so \( h(x) \in C^3. \) Thus we can take \( \sigma(x) = (x - d)^{-f} h(x) \) and \( \sigma(x) = (d - x)^{-f} h(x). \) 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 \( \mathcal{D}(\Gamma) \) be the domain of \( \Gamma. \) Then \( T_t \) is the unique strongly continuous conservative semigroup on \( C \) for which \( \mathcal{D}(\Gamma) \supset C^3 \) 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 \((\Gamma f(d) = 0 \text{ for } f \in \mathcal{D}(\Gamma))\) and regular boundaries are reflecting \((d/dp)f(d) = 0 \text{ for } f \in \mathcal{D}(\Gamma), \) where \( dp(x) = e^{-ha(x)} dx \) and \( dB(x) = 2a(x)b(x) dx \) ([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 \( F_N(x, u) = P(X^u \leq x) \) and \( F(x, u) = P(X(u) \leq x). \) The transition from

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

to

\[ \lim_{N \to \infty} F_N(x, [N^s u]) = F(x, u) \]

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

\[ \int_0^\infty F_N(x, [N^s u]) du = \int_0^\infty F(x, u) du. \]
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 \( \{X^\gamma(t), t \geq 0\} \) 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 \( n \geq m \) and \( f \in C \),

\[
E[|E(f(X_n) | \mathcal{F}_m) - T_{(n-m):m} f(X_m)|] \\
\leq \sum_{j=m}^{n-1} E[|E(g_{j+1}(X_{j+1}) | \mathcal{F}_j) - T_{r} g_{j+1}(X_j)|],
\]

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

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

Hence

\[
E(f(X_n) | \mathcal{F}_m) - T_{(n-m):m} f(X_m) = \sum_{j=m}^{n-1} E[|E(g_{j+1}(X_{j+1}) - g_j(X_j) | \mathcal{F}_m] \\
= \sum_{j=m}^{n-1} E[|E(g_{j+1}(X_{j+1}) | \mathcal{F}_j) - g_j(X_j) | \mathcal{F}_m].
\]

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

Let \( L \) be the subspace of \( C^4 \) 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

\[ ||g|| = ||g'|| + ||g''|| + M(g''), \]

where \( ||\cdot|| \) is the supremum norm. Any function \( g \) in \( L \) possesses a Taylor expansion

\[ g(y) = g(x) + (y-x)g'(x) + \frac{1}{2}(y-x)^2g''(x) + \lambda(y - x)^3M(g''), \]

where \( |\lambda| \leq \frac{1}{6} \). For the remainder in the first order Taylor expansion is

\[ g(y) - g(x) - \delta g'(x) = \delta^2 \int_0^1 (1 - s)g''(x + s\delta) \, ds \\
= 2^{-1} \delta^3 g''(x) + \delta^3 \int_0^1 (1 - s)g''(x + s\delta) - g''(x) \, ds, \]

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

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

Hence, in view of (1),

\[ E(g(X_{j+1}) | \mathcal{F}_j) = g(X_j) + \tau g(X_j) + \lambda ||g|| \sum_{i=1}^j |e_i|, \]
where \(|\lambda| \leq 1\). By (3), \(T, g\) has a similar expansion, so

\[ E[|E(g(X_{t+i})|, \mathscr{F}_j) - T, g(X_j)|] \leq \|g\| \sum_{i=1}^n |E(e_{i,j})| + \sup_{x \in \Omega} |e(x, x)|. \]

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.** \(T_i\) maps \(C^3\) into \(L\). Moreover, for any \(f \in C^3\) and \(K < \infty\),

\[ \sup_{t \leq K} \|T_i f\| < \infty. \]

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

\[ E[|E(f(X_{t+i})|, \mathscr{F}_m) - T_{t-m, t} f(X_{t+i})|] \]

\[ \leq K' \sum_{i=1}^n [(n - m) \sup_{x \in \Omega} |e(x, x)| + \sum_{i=1}^{n-1} E(|e_{i,j}|)] \]

for some constant \(K'\), provided that \((n - 1)r_N \leq K\).

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

\[ E[|E(f(X_{t+i})|, \mathscr{F}_N(s)) - T_{t-s, t} f(X_{t+i})|] \to 0 \]

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

\[ E(f(X_{t+i})) = E(T_{t-s} f(X_{t+i})) \]

\[ = E[f(X_{t+i})|, \mathscr{F}_N(s) - T_{t-s} f(X_{t+i})|] \to 0. \]

Taking \(s = 0\), and noting that \(T, f \in C\), so that

\[ E(T, f(X_{0+i})) = E(T, f(X_{0+i})) \]

\[ = E(f(X(t))). \]

we obtain \(E(f(X_{t+i})) = E(f(X(t)))\) as \(N \to \infty\).

Suppose, inductively, that \((X_{t_1}(t), \ldots, X_{t_k}(t)) \to (X(t_1), \ldots, X(t_k))\) for some \(k \geq 1\). If \(f_i \in C\) for \(1 \leq i \leq k + 1\), and \(0 \leq t_1 < t_2 < \ldots < t_{k+1}\), then

\[ E[\prod_{i=1}^{k+1} f_i(X_{t(i)})] = E[\prod_{i=1}^{k+1} f_i(X_{t(i)})E(f_{i+1}(X_{t(i)})|, \mathscr{F}_{t(i)} + \delta_N)] \]

\[ = E[\prod_{i=1}^{k+1} f_i(X_{t(i) + \delta_N})T_{t_{i+1} - t_i} f_{i+1}(X_{t(i)})] + \delta_N, \]

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

\[ \to E[\prod_{i=1}^{k+1} f_i(X_{t(i)})T_{t_{i+1} - t_i} f_{i+1}(X_{t(i)})], \]

by the induction hypothesis,

\[ = E[\prod_{i=1}^{k+1} f_i(X_{t(i)})]. \]

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 \( V_\tau f(x) = E(f(X_{\tau+1}^x) \mid X_\tau^x = x) \) for the discrete parameter Markov process \( X_n^x \) that moves from \( x \) to \( x + \tau a(x) + \tau b(x) \) with probability \( p(x) = \sigma(x)/(\sigma(x) + \sigma'(x)) \) and from \( x \) to \( x + \tau a(x) - \tau b(y) \) with probability \( 1 - p(x) \). It was shown by direct calculation that there is a constant \( \gamma \) such that \( \|V_\tau^n f\| \leq e^{\gamma n} \|f\| \) for \( \tau > 0 \), \( n \geq 0 \), and \( f \in C^1 \) ([5] Lemma 2.2, page 142). But \( V_\tau^n f \to T_\tau f \) uniformly as \( \tau \to 0 \) and \( \tau \to t \), and it follows that \( \|T_\tau f\| \leq e^{\gamma t} \|f\| \).

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

\[
|E(\Delta X_n^x \mid \mathcal{F}_n^x)| \leq G_1 \tau_n
\]

and

\[
\text{Var}(\Delta X_n^x \mid \mathcal{F}_n^x) \leq G_2 \tau_n
\]

a.s., for \( N \geq 1 \) and \( n \geq 0 \). Then, for any \( K > 0 \), \( \{X^N(t), t \leq K\} \Rightarrow \{X(t), t \leq K\} \) 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

\[
E[(X^N(t) - X^N(t_1))^3(X^N(t_2) - X^N(t_3))^3] \leq H(t_3 - t_1)^3
\]

for all \( 0 \leq t_1 \leq t \leq t_3 \leq K \). For this it is sufficient that

\[
E[(X^N_n - X^N_m)^3 \mid \mathcal{F}_n^x] \leq H'(n - m) \tau
\]

for \( 0 \leq m \leq n \leq K/\tau_N \).

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

\[
X_n^x - X_m^x = \sum_{j=m}^{n-1} V_j + \sum_{j=m}^{n-1} W_j,
\]

where

\[
V_j = E(\Delta X_j \mid \mathcal{F}_j)
\]

and

\[
W_j = \Delta X_j - E(\Delta X_j \mid \mathcal{F}_j).
\]

By (10),

\[
(\sum_{j=m}^{n-1} V_j)^3 \leq G_1(n - m)^2 \tau^2 \leq G_1^2(K(n - m) \tau
\]

\[\leq G_1^2(K(n - m) \tau.\]
By (11)
\[ E(W_j^2 \mid \mathcal{F}_n) = E(\text{Var}(\Delta X_j \mid \mathcal{F}_j) \mid \mathcal{F}_n) \leq G_j \tau, \]
hence
\[ E(\sum_{j=1}^{n-1} W_j^2 \mid \mathcal{F}_n) = \sum_{j=1}^{n-1} E(W_j^2 \mid \mathcal{F}_n) \leq G_1(n - m) \tau. \]
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, \( N_i \) are males and \( N_s \) are females. The three genotypes, aa, aA, and AA have frequencies \( k, N_i - k - l \), and \( l \) among males, and \( r, N_s - r - s \), 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
\[ X = 2^{-1} + 4^{-1}N_i^{-1}(k - l) + 4^{-1}N_s^{-1}(r - s) \]
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, \( \nu_i \) and \( \nu_s \), and a nonrandom-mating parameter \( f \). The latter is fixed as \( N \to \infty \), but the former are assumed to be inversely proportional to \( N \); \( N\alpha_1 = \hat{\alpha}_1 \geq 0 \) and \( N\nu_i = \hat{\nu}_i \) are constant. Moreover, \( N_i/N = r_i > 0 \) is fixed. Let \( X_n^\alpha \) be the value of \( X \) after \( n \) steps, and let \( \mathcal{F}_n^\alpha \) be the \( \sigma \)-field generated by the values of \( k, l, r, \) and \( s \) after \( j \) steps, \( j \leq n \).

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, \( \tau_\alpha = N^{-1} \) and \( \tau_N = N^{-1} \).

In both models, the function \( a \) of Theorem 1 is
\[ a(x) = \hat{\alpha}_1(1 - x) - \alpha_1 x \]
\[ - x(1 - x)[\hat{\nu}_i(1 - f)x + f] + \hat{\nu}_s(1 - f)(1 - x) + f]. \]
For the overlapping generation model,
\[ b(x) = 4^{-1}(r_i^{-1} + r_s^{-1})(1 + f)x(1 - x), \]
while \( b(x) \) is half the quantity on the right for the other model. Lemmas 2 and 3 give estimates of the error terms \( e_{n,\alpha}^\alpha \) in (1).

**Lemma 2.** For the overlapping generation model,
\[ E(|e_{n,\alpha}^\alpha|) \leq c\tau_N^{\frac{1}{4}} + ce^{-n/2N}\tau_N \]
for \( i = 1 \) and \( 2 \), \( n \geq 0 \), and all \( N \). Also

\[
E(|e_{i,n}^n|) \leq e^{n^2/4}.
\]

**Lemma 3.** For the non-overlapping generation model, \( |e_{i,n}^n| \leq G \tau_N \) a.s. for \( i = 1 \) and \( 2 \) and \( n \geq 0 \), \( E(|e_{i,n}^n|) \leq e^{n^2/4} \) for \( n \geq 2 \), \( E(|e_{2,n}^n|) \leq e^{n^2/4} \) for \( n \geq 1 \), and \( E(|e_{1,n}^n|) \leq e^{n^2/4} \) for \( n \geq 0 \). 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**


