Proc. Natl. Acad. Sci. USA
Vol. 74, No. 11, pp. 5096–5098, November 1977
Genetics

Error estimate for the diffusion approximation of the Wright–Fisher model

(population genetics/stochastic processes/random genetic drift)

S. N. ETHIER* AND M. FRANK NORMAN†

* Department of Statistics and Probability, Michigan State University, East Lansing, Michigan 48824; and † Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania 19174

Communicated by James F. Crow, August 11, 1977

ABSTRACT The Wright–Fisher model and its approximating diffusion model are compared in terms of the expected value of a smooth but arbitrary function of nth-generation gene frequency. In the absence of selection, this expectation is shown to differ in the two models by at most a linear combination (with coefficients depending only on the derivatives of the smooth function involved) of the maximum mutation rate and the reciprocal of the population size.

Consider a single diallelic locus, with alleles A_1 and A_2, in a monoecious population of M haploid individuals (or, alternatively, N diploids, in which case it suffices to replace M by $2N$ throughout). Let u and v be the probabilities of the mutations $A_1 \rightarrow A_2$ and $A_2 \rightarrow A_1$, respectively, and suppose that there are no selective differences between genotypes. Then the Markov chain

$$\{X_n^{(M,u,v)}\}_{n=0,1,2,\ldots}$$

with state space

$$I_M = \left\{\frac{i}{M} : i = 0,1,2,\ldots,M\right\}$$

and binomial transition probabilities

$$P\left\{X_{n+1}^{(M,u,v)} = \frac{j}{M} \middle| X_n^{(M,u,v)} = \frac{i}{M}\right\} = \binom{M}{j} p_i^j(1-p_i)^{M-j}, \quad [1]$$

where

$$p_i = (1-u)\frac{i}{M} + v\left(1 - \frac{i}{M}\right),$$

represents the successive proportions of A_1 genes in generations 0, 1, 2, . . . ; it is known as the Wright–Fisher model.

A useful technique in population genetics, developed by Fisher (1, 2) and Wright (3, 4), consists of approximating discrete stochastic models by diffusion processes, the latter being more amenable to analysis; for applications of such diffusion approximations, see refs. 5–8. In particular, one usually approximates the Wright–Fisher model by the diffusion process

$$\{Y_t^{(M,u,v)}\}_{t\geq 0}$$

with state space

$$I = \{x: 0 \leq x \leq 1\}$$

and backward operator

$$L = \frac{x(1-x)}{2M}\frac{\partial^2}{\partial x^2} + [-ux + v(1-x)]\frac{\partial}{\partial x}. \quad [2]$$

Strictly speaking, L does not, in general, uniquely determine a diffusion process in I. There are, however, several ways to characterize the appropriate diffusion, perhaps the most familiar of which is through the imposition of *reflecting* boundary conditions at *regular* boundaries (9, 10). The theory of stochastic differential equations provides an alternative characterization (11). Observe that, if $\alpha,\beta \geq 0$, the diffusion process

$$\{Z_t^{(\alpha,\beta)}\}_{t\geq 0}$$

in I, defined in terms of scaled time and mutation rates by

$$Z_t^{(\alpha,\beta)} = Y_{Mt}^{(M,\alpha/M,\beta/M)}, \quad [3]$$

is independent of M; indeed, its backward operator is

$$\Gamma = \frac{1}{2}x(1-x)\frac{\partial^2}{\partial x^2} + [-\alpha x + \beta(1-x)]\frac{\partial}{\partial x}.$$

Feller (12) was the first to study the problem of providing a mathematical justification for this diffusion approximation. It was suggested by him, and later proved by Trotter (13), that, for each $\alpha,\beta \geq 0$, $K > 0$, and f continuous on I,

$$E_x[f(X_{[Mt]}^{(M,\alpha/M,\beta/M)})] - E_x[f(Z_t^{(\alpha,\beta)})] \rightarrow 0 \quad [4]$$

as $M \rightarrow \infty$, uniformly over $0 \leq t \leq K$ and x in I_M, where $[Mt]$ denotes the integral part of Mt. (Here and elsewhere, the subscript x signifies the starting point of the Markov process occurring in the expectation.) In fact, the convergence in Eq. 4 has recently been shown to be uniform over $0 \leq t < \infty$ (14). Thus, by substituting Eq. 3 into Eq. 4 and replacing t by n/M, where $n = 0, 1, 2, \ldots$, we see that, for each $\alpha,\beta \geq 0$ and f continuous on I,

$$E_x[f(X_n^{(M,\alpha/M,\beta/M)})] - E_x[f(Y_n^{(M,\alpha/M,\beta/M)})] \rightarrow 0 \quad [5]$$

as $M \rightarrow \infty$, uniformly over $0 \leq n < \infty$ and x in I_M.

The assumption in Eq. 4 that the mutation rates be $O(1/M)$ as $M \rightarrow \infty$ is necessary in order to obtain an approximating diffusion model of the type being considered *which is independent of M*. In Eq. 5, however, such an assumption is unnecessary because one can show that, for each f continuous on I,

$$E_x[f(X_n^{(M,u,v)})] - E_x[f(Y_n^{(M,u,v)})] \rightarrow 0 \quad [6]$$

as $M \rightarrow \infty$, $u \rightarrow 0+$, and $v \rightarrow 0+$, uniformly over $0 \leq n < \infty$ and x in I_M; we temporarily defer the proof. Nevertheless, even Eq. 6 gives no information concerning the error in the approximation for specified values of the parameters. The purpose of the following result is to provide such information.

THEOREM. *Assume that* $M \geq 1$, $0 \leq u \leq 1$, $0 \leq v \leq 1$, $n \geq 0$, x *belongs to* I_M, *and* f *is a continuous function on* I *with*

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

six continuous derivatives. Then

$$|E_x[f(X_n^{(M,u,v)})] - E_x[f(Y_n^{(M,u,v)})]|$$
$$\le \max(u,v)\left\{\frac{1}{2}\|f^{(1)}\|\right\} + \frac{1}{M}\left\{\frac{\theta}{16}\|f^{(2)}\| + \frac{1}{216\sqrt{3}}\|f^{(3)}\|\right\}$$
$$+ \max(u^2,v^2)\left\{\frac{9}{2}\sum_{j=1}^{6}\|f^{(j)}\|\right\} + \frac{1}{M^2}\left\{\frac{7}{16}\sum_{j=2}^{6}\|f^{(j)}\|\right\}, \quad [7]$$

where

$$\theta = \frac{1 + 4M\max(u,v)}{1 + 2M(u+v)}$$

and $\|f^{(j)}\|$ *denotes the maximum absolute value of the* j*th derivative of* f. *(Note that* $1 \le \theta < 2$ *and that* $\theta = 1$ *if* u = v.)

Before turning to the proof, let us consider some applications and corollaries.

By applying the theorem with $f(x) = 2x(1-x)$ and $M = 2N$, we find that the expected nth-generation heterozygosity in the Wright–Fisher model (for a diploid population of size N with mutation rates u and v) differs from that in its approximating diffusion by at most $\max(u,v) + 1/(4N)$ plus terms of higher order. One can also compare kth moments of nth-generation A_1-gene frequency in the two models by taking $f(x) = x^k$. When $k = 1$ [i.e., $f(x) = x$] and $u = v = 0$, the right side (and therefore the left side) of Eq. **7** vanishes, confirming a rather obvious result; however, as k increases, so does our bound.

Of course, many genetically interesting quantities associated with the Wright–Fisher model cannot be expressed as the expected value of a smooth function of nth-generation A_1-gene frequency. For example, when $u = 0$, the probability of fixation of the allele A_1 by generation n is the expected value of the discontinuous function $f(x) = 0$, $0 \le x < 1$, $f(1) = 1$, of nth-generation A_1-gene frequency. Other examples (e.g., expected total heterozygosity, expected fixation time) depend on gene frequencies in more than one generation; for a discussion of such problems in this context, see ref. 15.

Because the right side of Eq. **7** is independent of n, we can let $n \to \infty$ on the left without affecting the inequality. When both u and v are positive, this yields a bound on the difference between the expected values of a smooth but arbitrary function with respect to the stationary distributions in the Wright–Fisher model and its approximating diffusion. When $u = v = 0$, we find [with $f(x) = x$] that the probabilities of eventual fixation of the allele A_1 are identical in the two models, a well-known and easily derived result. In the former case, Ewens (16) has compared the stationary distribution in the Wright–Fisher model with that in its approximating diffusion in terms of their densities.

Observe that Eq. **6** holds for every function f to which the theorem applies. Because continuous functions on I can be uniformly approximated arbitrarily closely by such functions (in fact, by polynomials), it follows that Eq. **6** holds for every continuous function f on I, as claimed.

Finally, we note that, of Eqs. **4–7**, only Eqs. **4** and **5** have been extended to models incorporating selection (14). However, one can estimate the rate of convergence in limit theorems such as Eq. **4** quite generally (cf. ref. 17).

Proof: The proof of the theorem is not difficult, so, aside from some reasonably straightforward but tedious calculations, we present it in detail. Let M, u, and v be fixed throughout, as in the statement of the theorem.

For $n = 0,1,2,\ldots$, define the operator S_n on $B(I_M)$, the space of (bounded) functions on I_M, by

$$(S_nf)(x) = E_x[f(X_n^{(M,u,v)})] \quad [8]$$

and note that

$$S_mS_n = S_{m+n} \quad [9]$$

for $m = 0,1,2,\ldots$; this follows immediately from the fact that the k-step transition matrix of a time-homogeneous Markov chain is the kth power of the one-step transition matrix. For each $t \ge 0$, define the operator T_t on $C(I)$, the space of continuous functions on I, by

$$(T_tf)(x) = E_x[f(Y_t^{(M,u,v)})] \quad [10]$$

and note that

$$T_sT_t = T_{s+t} \quad [11]$$

for all $s \ge 0$; this is a consequence of the Chapman–Kolmogorov identity for the transition function of a time-homogeneous Markov process (10). [The fact that each T_t is indeed a bona fide operator on $C(I)$—i.e., that it maps $C(I)$ into $C(I)$—can be deduced from ref. 10.] By Eqs. **9** and **11**,

$$S_nT_0f - S_0T_nf$$
$$= \sum_{k=0}^{n-1}(S_{n-k}T_kf - S_{n-k-1}T_{k+1}f)$$
$$= \sum_{k=0}^{n-1}(S_{n-k-1}S_1T_kf - S_{n-k-1}T_1T_kf)$$
$$= \sum_{k=0}^{n-1}S_{n-k-1}(S_1 - T_1)T_kf \quad [12]$$

for $n = 1,2,3,\ldots$ and f in $C(I)$. [We do not distinguish between functions in $C(I)$ and their restrictions to I_M.]

Denote the maximum absolute values of functions f in $B(I_M)$ and g in $C(I)$ by $\|f\|_M$ and $\|g\|$. Applying the triangle inequality to Eq. **12** and noting that both S_0 and T_0 are identity operators and each S_n is a contraction operator [i.e., $\|S_nf\|_M \le \|f\|_M$ for all f in $B(I_M)$], we obtain

$$\|S_nf - T_nf\|_M \le \sum_{k=0}^{n-1}\|(S_1 - T_1)T_kf\|_M \quad [13]$$

for $n = 1,2,3,\ldots$ and f in $C(I)$.

Let $C^m(I)$ be the subspace of $C(I)$ consisting of those functions with m continuous derivatives on I. We will need the fact that

$$\lim_{t\to 0+}\|t^{-1}\{T_tf - f\} - Lf\| = 0 \quad [14]$$

for every f in $C^2(I)$, where L is the backward operator defined by Eq. **2**. The proof of this result depends on the characterization of the diffusion process in terms of which $\{T_t\}_{t\ge 0}$ was defined; consult Lemma 9.3.1 of ref. 18 for details. It will later be seen that there is a close connection between Eq. **14** and the Kolmogorov backward equation; for now, it suffices to note that, by Eqs. **11** and **14**,

$$\frac{\partial^j}{\partial t^j}(T_tf)(x) = (T_tL^jf)(x) \quad [15]$$

for $j = 1,2,3,\ldots$, f in $C^{2j}(I)$, x in I, and $t \ge 0$.

Given g in $C^6(I)$ and x in I_M, an application of Taylor's theorem yields

$$(S_1g)(x) = g(x) + \sum_{j=1}^{5}E_x[(X_1^{(M,u,v)} - x)^j]g^{(j)}(x)/j!$$
$$+ \omega_1 E_x[(X_1^{(M,u,v)} - x)^6]\|g^{(6)}\|/6!, \quad [16]$$

where $|\omega_1| \le 1$ and $g^{(j)}$ denotes the jth derivative of g; in view of Eq. **15**, a second application of Taylor's theorem results in

$$(T_1g)(x) = g(x) + (Lg)(x) + (L^2g)(x)/2 + \omega_2\|L^3g\|/6, \quad [17]$$

where $|\omega_2| \leq 1$; here we are using the facts that T_0 is the identity operator and each T_t is a contraction operator. By Eq. **1**, we can evaluate the moments in Eq. **16** in terms of the first six moments of the binomial distribution (the central moments of which are actually easier to compute); we can also calculate the coefficients determining the fourth- and sixth-order differential operators L^2 and L^3. Upon subtracting Eq. **17** from Eq. **16** and noting the resulting cancellations, we obtain

$$\|S_1 g - T_1 g\|_M \leq \sum_{j=1}^{6} \gamma_j \|g^{(j)}\| \qquad [18]$$

for every g in $C^6(I)$, where $\gamma_1, \gamma_2, \ldots, \gamma_6$ are certain polynomials in the three variables $\max(u,v)$, $u + v$, and $1/M$.

Suppose for the moment that, for $j = 1,2,3, \ldots$ and $t \geq 0$,

$$T_t \text{ maps } C^j(I) \text{ into } C^j(I) \qquad [19]$$

and

$$\|(T_t f)^{(j)}\| \leq \exp(-\lambda_j t)\|f^{(j)}\| \qquad [20]$$

for every f in $C^j(I)$, where

$$\lambda_j = \frac{j(j-1)}{2M} + j(u+v).$$

Then, from the estimates **13**, **18**, and **20**, together with the inequality

$$\sum_{k=0}^{n-1} \exp(-\lambda_j k) < \sum_{k=0}^{\infty} \exp(-\lambda_j k) = [1 - \exp(-\lambda_j)]^{-1} < 1 + \lambda_j^{-1},$$

we conclude that

$$\|S_n f - T_n f\|_M \leq \sum_{k=0}^{n-1} \sum_{j=1}^{6} \gamma_j \|(T_k f)^{(j)}\| \leq \sum_{j=1}^{6} \gamma_j \sum_{k=0}^{n-1} \exp(-\lambda_j k)\|f^{(j)}\| \leq \sum_{j=1}^{6} \gamma_j (1 + \lambda_j^{-1})\|f^{(j)}\| \qquad [21]$$

for $n = 1,2,3, \ldots$ and f in $C^6(I)$. (If $u = v = 0$, then $\lambda_1 = 0$; but in this case $\gamma_1 = 0$, so the first term of the last sum in Eq. **21** is absent.) Using the explicit expressions for γ_j (not given) and λ_j, $j = 1,2\ldots,6$, one can check that the right side of Eq. **21** cannot exceed the right side of Eq. **7**. By substituting Eqs. **8** and **10** into the left side of Eq. **21**, we therefore arrive at the conclusion of the theorem.

Thus, to complete the proof, it suffices to verify Eqs. **19** and **20**. Because Eq. **14** holds for every f in $C^2(I)$, it follows from Theorem 1 of ref. 19 (or from the argument of Lemma 2 there) that Eq. **19** holds for $j = 1,2,3, \ldots$ and $t \geq 0$. As for Eq. **20**, we can argue as follows. Fix a positive integer j. Given f in $C^{j+2}(I)$, let $\phi(t,x) = (T_t f)(x)$ for $t \geq 0$ and x in I. By Eq. **19**, $\phi(t,\cdot)$ belongs to $C^{j+2}(I)$ for each $t \geq 0$, so, by Eqs. **11** and **14**, ϕ satisfies the Kolmogorov backward equation

$$\frac{\partial}{\partial t}\phi = L\phi, \qquad \phi(0,\cdot) = f.$$

Differentiating j times with respect to x and denoting $\partial^j\phi/\partial x^j$ by $\phi^{(j)}$, we obtain

$$\frac{\partial}{\partial t}\phi^{(j)} = L_j\phi^{(j)} - \lambda_j\phi^{(j)}, \qquad \phi^{(j)}(0,\cdot) = f^{(j)}, \qquad [22]$$

where

$$L_j = L + \frac{j}{2M}(1-2x)\frac{\partial}{\partial x}.$$

To justify Eq. **22** as well as what follows, one should check that the first t-derivative and the first two x-derivatives of $\phi^{(j)}$ are jointly continuous (cf. ref. 19). Now let $\psi^{(j)}(t,x) = \phi^{(j)}(t,x)\exp(\lambda_j t)$ for $t \geq 0$ and x in I, and note that Eq. **22** becomes

$$\frac{\partial}{\partial t}\psi^{(j)} = L_j\psi^{(j)}, \qquad \psi^{(j)}(0,\cdot) = f^{(j)}.$$

Therefore,

$$\|\psi^{(j)}(t,\cdot)\| \leq \|f^{(j)}\| \qquad [23]$$

for each $t \geq 0$ by the weak maximum principle [cf. Theorem 6.3.1 in ref. 20, a slight modification of which is sufficient here when applied to $\psi^{(j)}$ and $-\psi^{(j)}$; the point is that if h in $C^2(I)$ has its maximum at x in I, then $(L_j h)(x) \leq 0$, *even if the maximum occurs at* 0 *or* 1]. We conclude from Eq. **23** that, for each $t \geq 0$, Eq. **20** holds for every f in $C^{j+2}(I)$ and thus (by a simple argument) for every f in $C^j(I)$.

We thank Prof. T. G. Kurtz for suggesting the problem and Prof. T. Nagylaki for helpful communication.

1. Fisher, R. A. (1922) "On the dominance ratio," *Proc. Roy. Soc. Edinburgh* **42**, 321–341.
2. Fisher, R. A. (1930) *The Genetical Theory of Natural Selection* (Clarendon Press, Oxford).
3. Wright, S. (1931) "Evolution in Mendelian populations," *Genetics* **16**, 97–159.
4. Wright, S. (1945) "The differential equation of the distribution of gene frequencies," *Proc. Natl. Acad. Sci. USA* **31**, 382–389.
5. Kimura, M. (1964) "Diffusion models in population genetics," *J. Appl. Probability* **1**, 177–232.
6. Ewens, W. J. (1969) *Population Genetics* (Methuen, London).
7. Crow, J. F. & Kimura, M. (1970) *An Introduction to Population Genetics Theory* (Harper & Row, New York).
8. Kimura, M. & Ohta, T. (1971) *Theoretical Aspects of Population Genetics* (Princeton Univ. Press, Princeton, NJ).
9. Feller, W. (1952) "The parabolic differential equations and the associated semi-groups of transformations," *Ann. Math.* **55**, 468–519.
10. Feller, W. (1954) "Diffusion processes in one dimension," *Trans. Am. Math. Soc.* **77**, 1–31.
11. Yamada, T. & Watanabe, S. (1971) "On the uniqueness of solutions of stochastic differential equations," *J. Math. Kyoto Univ.* **11**, 155–167.
12. Feller, W. (1951) "Diffusion processes in genetics," in *Proc. Second Berkeley Symposium Math. Statist. and Probability*, ed. Neyman, J. (Univ. of California Press, Berkeley, CA), pp. 227–246.
13. Trotter, H. F. (1958) "Approximation of semi-groups of operators," *Pac. J. Math.* **8**, 887–919.
14. Norman, M. F. (1977) "Ergodicity of diffusion and temporal uniformity of diffusion approximation," *J. Appl. Probability* **14**, 399–404.
15. Guess, H. A. (1973) "On the weak convergence of Wright-Fisher models," *Stochastic Processes Appl.* **1**, 287–306.
16. Ewens, W. J. (1965) "The adequacy of the diffusion approximation to certain distributions in genetics," *Biometrics* **21**, 386–394.
17. Ethier, S. N. (1975) Dissertation, University of Wisconsin, Madison, WI.
18. Norman, M. F. (1972) *Markov Processes and Learning Models* (Academic Press, New York).
19. Ethier, S. N. (1976) "A class of degenerate diffusion processes occurring in population genetics," *Comm. Pure Appl. Math.* **29**, 483–493.
20. Friedman, A. (1975) *Stochastic Differential Equations and Applications* (Academic Press, New York), Vol. 1.