from Mathematical Psychology and Psychophysiology, edited by S. Grossberg

SIAM-AMS Proceedings
Volume 13
1981

A "Psychological" Proof that Certain Markov Semigroups Preserve Differentiability

M. FRANK NORMAN

1. Introduction. Unlike the preceding paper [1], the present one is directed toward a purely mathematical result. This result is linked to genetics because it can play a role in analyzing genetic models. It is linked to psychology because my proof makes essential use of a psychological model. However, the result itself is a fairly general theorem about a well-studied class of mathematical objects. The objects I have in mind are the semigroups of operators associated with certain diffusions.

For our purposes, a diffusion, $X(t)$, is a real-valued strong Markov process with continuous sample paths. (See [2, Chapter 2] for background information and terminology regarding diffusions.) We assume that the process is defined for all $t \geq 0$ and that its state space, $I = [r_0, r_1]$, is closed and bounded. A central role in the description and analysis of a finite Markov chain, X_n, is played by its transition matrix, P, with components

$$p_{xy} = P(X_{n+1} = y | X_n = x).$$

Associated with this matrix is a transformation of functions (regarded as column vectors) defined by matrix multiplication,

$$Pf(x) = \sum_y p_{xy} f(y),$$

or, equivalently, $Pf(x) = E(f(X_{n+1})|X_n = x)$. Both the transition matrix and the transition operator can be generalized to diffusions, but the transition operator, defined by

$$T_t f(x) = E(f(X_{t+s})|X_s = x),$$

is far more tractable and thus occupies a more prominent place in the theory of diffusions.

1980 *Mathematics Subject Classification*. Primary 47D07, 35K15, 35K65; Secondary 60F05, 60J60, 60J70.

These operators form a semigroup, i.e., $T_t T_u = T_{t+u}$ and T_0 is the identity operator. Moreover, T_t is linear, positive ($T_t f \geq 0$ if $f \geq 0$), conservative ($T_t 1 = 1$), and contractive ($\|T_t f\|_0 \leq \|f\|_0$, where $\|f\|_0 = \sup_{x \in I}|f(x)|$). Although T_t is well defined for bounded measurable functions, it is advantageous to restrict it to continuous functions ($f \in C^0$). For most diffusions, if $f \in C^0$ then $T_t f \in C^0$ and $T_t f$ is continuous in t with respect to the supremum norm. The latter property is called strong continuity of T_t. A strongly continuous semigroup of positive linear contractions mapping C^0 into C^0 is called a Fellerian semigroup (after W. Feller). The theory of such semigroups is presented in [3, Chapters I and II].

Diffusions are highly amenable to mathematical analysis, so there is considerable interest in using them as approximations for stochastic processes that are more recalcitrant. These processes often have discrete time scales. A prime example is the finite population version of the infinite population genetic model considered in [1]. Let $X_{K,n}$ be the proportion of A_1 genes in the nth generation for a population of size K, and suppose that selection differentials (and mutation rates, if nonzero) are of the order of magnitude of K^{-1}. Then $X_{K,n}$ can be approximated by $X(n/K)$ for a certain diffusion, $X(t)$. More precisely, if $f \in C^0$, then

$$E\big(f(X_{K,n})|X_{K,0} = x\big) - T_{n/K}f(x) \to 0$$

as $K \to \infty$, uniformly over x and $n \leq KC$, for any $C < \infty$ [4, §18.1].

General theorems of this type [5] can be proved very simply by an argument growing out of Khintchine's work [6, §1 of Chapter 3]. I will refer to this argument as Khintchine's, although modern versions are simpler than the original in certain respects. An essential ingredient of this approach is a suitable bound on x-derivatives of $T_t f$. If f is infinitely differentiable, then, typically, $T_t f$ is too, and $|T_t f|_k \leq \exp(\lambda_k t)|f|_k$ for all $k \geq 1$ and $t \geq 0$, where λ_k is a constant, $|f|_k = \sum_{j=1}^{k}\|f^{(j)}\|_0$, and $f^{(j)}$ is the jth derivative of f. To prove diffusion approximation theorems, this bound is needed only for $k = 3$. However, for a broad class of diffusions, it holds for all positive integers. Our main objective in this paper is to establish such bounds by means of an argument that uses mathematical learning models.

2. Statement of results. We shall first attempt to motivate the conditions that define the class of semigroups to which our results apply. These semigroups correspond to diffusions, $X(t)$, that satisfy equations of the form

$$E(\Delta_\tau X(t)|X(t) = x) = \tau b(x) + o(\tau), \tag{2.1}$$

$$E\big((\Delta_\tau X(t))^2|X(t) = x\big) = \tau a(x) + o(\tau), \tag{2.2}$$

$$E\big(|\Delta_\tau X(t)|^3|X(t) = x\big) = o(\tau), \tag{2.3}$$

where $\Delta_\tau X(t) = X(t+\tau) - X(t)$. The drift and diffusion functions, b and a, are continuous, and $o(\tau)$ is uniform over I, i.e.,

$$\lim_{\tau \to 0}\|\tau^{-1}o(\tau)\|_0 = 0. \tag{2.4}$$

It follows from (2.1) and (2.2) that $b(r_0) \geq 0$, $b(r_1) \leq 0$, and $a(x) \geq 0$ for all $x \in I$. We shall assume that $a(x) > 0$ in the interior of I. However, applying $E(Y^2)^2 \leq E(|Y|)E(|Y|^3)$ to $Y = \Delta_\tau X(t)$, we see that $a(r_0) = a(r_1) = 0$.

A Fellerian semigroup is characterized by its generator, Γ, defined by

$$\Gamma f = \lim_{t \to 0} t^{-1}(T_t f - f).$$

Functions for which this limit exists (with respect to $\|\cdot\|_0$) constitute the domain, $\mathfrak{D}(\Gamma)$, of Γ. Let C^k be the collection of real-valued continuous functions on I with k derivatives continuous throughout I (even at endpoints). It follows easily from (2.1)–(2.4) that

$$\mathfrak{D}(\Gamma) \supset C^2 \tag{2.5}$$

and

$$\Gamma f(x) = 2^{-1}a(x)f''(x) + b(x)f'(x) \quad \text{for } f \in C^2. \tag{2.6}$$

These conditions provide the required characterization of the semigroups to be treated.

The following function spaces figure prominently in our results.

$$D^0 = \text{essentially bounded functions on } I,$$

and, for $k \geq 1$,

$$D^k = \{ f \in C^{k-1} : f^{(k-1)} \text{ is absolutely continuous, and } f^{(k)} \in D^0 \}.$$

Since I is compact, $C^k \subset D^k$. For $f \in D^0$, let $\|f\|_0$ = essential supremum of $|f(x)|$, and, for $f \in D^k$, $k \geq 1$, let $|f|_k = \sum_{j=1}^k \|f^{(j)}\|_0$ and $\|f\|_k = \sum_{j=0}^k \|f^{(j)}\|_0$. For all $k \geq 0$, D^k is a Banach space with respect to the norm $\|\cdot\|_k$.

THEOREM 1. *Suppose that* $a, b \in D^m$ *for some* $m \geq 2$, $a(r_0) = 0$, $a(r_1) = 0$, $a(x) > 0$ *for* $r_0 < x < r_1$, $b(r_0) \geq 0$, *and* $b(r_1) \leq 0$. *Then there is one and only one conservative Fellerian semigroup,* T_t, *whose generator,* Γ, *satisfies* (2.5) *and* (2.6). *Moreover, for any* $1 \leq k \leq m$ *and* $t \geq 0$, T_t *maps* D^k *into* D^k *and*

$$|T_t f|_k \leq \exp(\lambda_k t)|f|_k, \tag{2.7}$$

where

$$\lambda_k = \max_{1 \leq j \leq k} \sum_{i=j}^{k} \|d_{ij}\|_0 \tag{2.8}$$

and

$$d_{ij} = \binom{i}{j-1} b^{(i-j+1)} + \frac{1}{2}\binom{i}{j-2} a^{(i-j+2)} \qquad (d_{i1} = b^{(i)}). \tag{2.9}$$

Since $a(r_i) = 0$, the elliptic operator (2.6) is degenerate. A typical example of a diffusion function to which Theorem 1 applies is $a(x) = 2^{-1}x(1 - x)$. This is the diffusion function for the diffusion approximation to the genetic model mentioned in §1.

Since $\|T_t f\|_0 < \|f\|_0$, the seminorm, $|\cdot|_k$, can be replaced by the norm, $\|\cdot\|_k$, in (2.8), so that $\|T_t\|_k < \exp(\lambda_k t)$.

Theorem 1 is similar to a result of Ethier [7, Theorem 1], in which C^k takes the place of our D^k. Ethier's proof, involving a judicious combination of semigroup and martingale methods, is quite different from ours. Unlike our result, Ethier's is not restricted to finite intervals I. In order to apply our result to infinite intervals, it is necessary to make a preliminary spatial transformation of the type used in the proof of Lemma 6 in §7.

Readers with backgrounds in diffusion theory may be struck by the fact that (2.5) and (2.6) (or the equivalent conditions (2.1)–(2.4)) uniquely determine a semigroup (or diffusion) without reference to boundary conditions. This is worth considering in more detail. We begin with a review of Feller's boundary classification. This classification depends on the scale function, p, and speed function, m, defined by

$$p(x) = \int_r^x e^{-B(y)}\,dy \quad \text{and} \quad m(x) = \int_r^x \frac{2}{a(y)} e^{B(y)}\,dy,$$

where $B(x) = \int_r^x 2b(y)/a(y)\,dy$ and $r \in (r_0, r_1)$. Let

$$u(z) = \int_r^z m(x)\,dp(x) \quad \text{and} \quad v(z) = \int_r^z p(x)\,dm(x).$$

The classification of r_i is given by Table 1.

TABLE 1. Boundary classification

$u(r_i)$	$v(r_i)$	Boundary type
$< \infty$	$< \infty$	regular
$< \infty$	$= \infty$	exit
$= \infty$	$< \infty$	entrance
$= \infty$	$= \infty$	natural

Boundary conditions enter into the complete description of $\mathfrak{D}(\Gamma)$, which is as follows. Let $\mathfrak{D}$ be the subset of C^0 consisting of functions, f, with two continuous derivatives on (r_0, r_1) and for which

$$(d/dm)(d/dp)f(x) = 2^{-1}a(x)f''(x) + b(x)f'(x) \tag{2.10}$$

has finite limits at r_0 and r_1. Let

$$\mathfrak{D}_i = \begin{cases} \mathfrak{D}, & \text{if } r_i \text{ is entrance or natural,} \\ \{f \in \mathfrak{D} : (d/dp)f(r_i) = 0\}, & \text{if } r_i \text{ is regular,} \\ \{f \in \mathfrak{D} : (d/dm)(d/dp)f(r_i) = 0\}, & \text{if } r_i \text{ is exit.} \end{cases} \tag{2.11}$$

Then

$$\mathfrak{D}(\Gamma) = \mathfrak{D}_0 \cap \mathfrak{D}_1. \tag{2.12}$$

The interesting point is that, if r_i is regular, the boundary condition given above is not the only possibility for a conservative Fellerian semigroup. For example, any condition of the form

$$(-1)^{i+1}c_i(d/dp)f(r_i) + d_i(d/dm)(d/dp)f(r_i) = 0,$$

where $c_i \geqslant 0$, $d_i \geqslant 0$, and $c_i + d_i > 0$ is admissible. The diffusions obtained by boundary conditions other than those in (2.11) will satisfy (2.1)–(2.3) for each $x \in (r_0, r_1)$, but they will not satisfy the uniformity condition (2.4) or the closely related semigroup condition (2.5). Thus these conditions serve in lieu of boundary conditions in our approach. The semigroup thereby determined is subject to the bound (2.7), which, in the light of Khintchine's work, marks this semigroup as the appropriate diffusion approximation for a wide variety of processes.

If r_i is regular or exit, it is not obvious that the boundary conditions given by (2.11) are satisfied by all C^2 functions, as (2.5) requires. This fact follows easily from Theorem 2, which contains more information than is needed for our present purposes.

THEOREM 2. *Suppose that* $b \in D^1$, $a \in D^2$, $a(r_0) = 0$, $a(r_1) = 0$, $a(x) > 0$ *for* $r_0 < x < r_1$, $b(r_0) \geqslant 0$, *and* $b(r_1) \leqslant 0$. *Then the classification of* r_i *is determined by* $\alpha_i = (-1)^i a'(r_i)$ *and* $\beta_i = (-1)^i b(r_i)$ *in accordance with Table* 2.

TABLE 2. Boundary classification in terms of α_i and β_i

α_i	β_i	Boundary type
$=0$	$=0$	natural
$=0$	>0	entrance
>0	$\geqslant \alpha_i/2$	entrance
>0	$\in(0, \alpha_i/2)$	regular
>0	$=0$	exit

This result is of independent interest, since the criteria for various boundary types are much easier to check than those of Table 1. Of course the criteria of Table 1 apply to speed and scale functions other than those considered here.

3. The role of learning models. In this section we introduce relevant learning models and sketch the proof of Theorem 1. Details are given in §§4–6. The proof of Theorem 2 is in §7.

Suppose that a rat has repeated trials in which he starts at the bottom of a T-shaped alley ("T-maze") and moves to the end of the right or left arm. Food is present in both arms on every trial. Let X_n be the rat's probability of turning right on the nth trial. According to the classical linear model for such an experiment [8, p. 115],

$$\Delta X_n = \begin{cases} \theta(1 - X_n) & \text{if he turns right,} \\ -\theta X_n & \text{if he turns left,} \end{cases}$$

where θ is a "learning rate parameter" $(0 < \theta < 1)$. In other words,

$$\Delta X_n = \begin{cases} \theta(1 - X_n) & \text{with probability } X_n, \\ -\theta X_n & \text{with probability } 1 - X_n. \end{cases}$$

This equation defines a discrete parameter Markov process with continuous state space [0, 1]. The genesis of the proof of Theorem 1 was my attempt **[9]** to obtain a diffusion approximation for this learning model and others like it.

It quickly emerged (see [**4**, Chapter 9]) that the techniques of **[9]** were equally applicable to processes satisfying equations of the form

$$\Delta X_n = \begin{cases} \theta\sigma_1(X_n) + \tau b(X_n) & \text{with probability } q(X_n), \\ -\theta\sigma_0(X_n) + \tau b(X_n) & \text{with probability } 1 - q(X_n), \end{cases} \tag{3.1}$$

where $\Delta X_n = X_{n+1} - X_n$, $\tau = \theta^2$,

$$q(x) = \sigma_0(x)/\left(\sigma_0(x) + \sigma_1(x)\right) \quad \text{for } r_0 < x < r_1, \tag{3.2}$$

$$\sigma_i(r_i) = 0, \qquad \sigma_i(x) > 0 \quad \text{for } r_0 < x < r_1, \tag{3.3}$$

$$q \text{ is nondecreasing,} \tag{3.4}$$

and

$$\sigma_i, b, q \in C^\infty. \tag{3.5}$$

To identify the appropriate diffusion approximation, $X(t)$, for discrete parameter processes, one usually considers moments of ΔX_n. In the present case,

$$E(\Delta X_n | X_n = x) = \tau b(x), \tag{3.6}$$

$$E\left((\Delta X_n)^2 | X_n = x\right) = \tau\sigma_0(x)\sigma_1(x) + O(\tau^{3/2}), \tag{3.7}$$

$$E\left(|\Delta X_n|^3 | X_n = x\right) = O(\tau^{3/2}), \tag{3.8}$$

where the O's are uniform over x. If $X(n\tau)$ is to approximate X_n, then $\Delta_\tau X(t)$ in (2.1)–(2.3) is comparable to ΔX_n in (3.6)–(3.8). Comparing these two sets of equations, we see that the putative limiting process is the one for which

$$a = \sigma_0\sigma_1. \tag{3.9}$$

The proof of Theorem 1 is based on the process (3.1) in a way that I will describe momentarily. I will refer to this process as a learning model, although few special cases beyond the linear model have received much attention from psychologists and, unbeknown to psychologists, mathematicians had studied such processes earlier **[10]**.

Our sketch of the proof of Theorem 1 begins with an alteration of viewpoint. Instead of starting with a learning model and looking for a diffusion to approximate it, we now start with a diffusion and seek an approximating learning model. Suppose initially that $b \in C^\infty$ and a is a polynomial (in addition to the other assumptions of Theorem 1). Let speed and scale functions, m and p, be defined accordingly. Feller's theory [**3**, Chapter II] assures us that the operator (2.10), restricted to the space $\mathfrak{D}(\Gamma)$ defined by (2.12), generates a conservative Fellerian semigroup, T_t. In view of Theorem 2, (2.5) and (2.6) are satisfied, and (2.1) – (2.4) follow. The question now arises: Can one choose the functions σ_1 and σ_0 in (3.1) so that (3.9) holds (hence (3.7) is analogous to (2.2)), and so that (3.3), (3.4), and (3.5) also hold, where q is defined by (3.2)? If a is a polynomial, as we are presently assuming, such σ_0 and σ_1 always exist. For if the zero of a at r_i has order z_i, we can take

$$\sigma_i(x) = |x - r_i|^{z_i}(A(x))^{1/2}, \tag{3.10}$$

where $A(x)$ is the polynomial

$$A(x) = a(x)/\left[(x - r_0)^{z_0}(r_1 - x)^{z_1}\right]. \tag{3.11}$$

Since $A(x) > 0$ for all $r_0 \leqslant x \leqslant r_1$, $(A(x))^{1/2} \in C^\infty$, and (3.5) holds, as do (3.3) and (3.4).

Now comes a crucial point. Khintchine's diffusion approximation argument, given in detail in the next section, presupposes two rather distinct circumstances. First, it presupposes corresponding expressions for moments of ΔX_n and $\Delta_\tau X(t)$. Second, it presupposes *either* a bound like (2.7) for $k = 3$ *or* an analogous bound for the learning model. We are trying to prove (2.7), so this bound is certainly not available to us at this juncture. However, a corresponding bound for the learning model is not difficult to establish. Let U^n be the discrete parameter semigroup for the learning model, $U^n f(x) = E(f(X_n)|X_0 = x)$. Then it is obvious that U^1 and thus U^n maps C^∞ into C^∞, and we shall show that, for each $k \geqslant 1$, there is a constant, γ_k, such that $|U^n f|_k \leqslant \exp(\gamma_k n\tau)|f|_k$. Khintchine's argument then yields $\| U^n f - T_t f\|_0 \to 0$ as $\tau \to 0$ and $n\tau \to t$, for $f \in C^\infty$. It follows that, if $f \in C^\infty$, then $T_t f \in C^\infty$ and

$$|T_t f|_k \leqslant \exp(\gamma_k t)|f|_k, \tag{3.12}$$

for all $k \geqslant 1$. These are the main conclusions of §4.

The constants, γ_k, of (3.12) depend on derivatives of b and the "artificial" functions σ_i and q, rather than b and a. This deficiency is corrected in §5, where a maximum principle is applied to $(\partial/\partial x)^k T_t f$ to derive (2.7) for infinitely differentiable f and b and for polynomial a. In §6, (2.7) is extended to $a, b \in D^m$ by taking the limit along a sequence $T_{n,t}$ of semigroups defined in terms of suitable polynomial approximations, a_n and b_n, to a and b. A simpler approximation argument allows us to pass from $f \in C^\infty$ to $f \in D^k$.

Uniqueness of T_t is established by the following observations. The argument in §4 (for polynomial a and infinitely differentiable b) actually shows that any Fellerian semigroup, T_t, that satisfies (2.5) and (2.6) also satisfies $U^n f \to T_t f$ as $\tau \to 0$ and $n\tau \to t$. Thus (2.5) and (2.6) uniquely determine T_t for polynomial a and $b \in C^\infty$. Similarly, the argument in §6 (for $a, b \in D^m$) derives $T_{n,t} f \to T_t f$ from (2.5) and (2.6). Since $T_{n,t}$ is uniquely determined, T_t is too.

The argument in §4 is similar to that of [**4**, Chapter 9], while our use of the maximum principle in §5 was suggested by [**11**]. The latter paper, by the way, gives a much better bound than (2.7) for the special case of linear b and quadratic a.

Throughout the remainder of the paper we shall assume, without loss of generality, that $r_0 = 0$ and $r_1 = 1$.

4. Proof of Theorem 1. Differentiability. In this section we assume that a is a polynomial and $b \in C^\infty$, and define σ_i and q by (3.10), (3.11), and (3.2). Of course $q(i)$ is defined so that $q \in C^0$. Then (3.3), (3.4), (3.5), and (3.9) are satisfied. Let $s_i = (-1)^{i+1}$, $x_i = x + \theta s_i \sigma_i + \tau b$, $q_1 = q$, and $q_0 = 1 - q$, so that the learning model has the form

$$X_{n+1} = (X_n)_i \quad \text{with probability } q_i(X_n). \tag{4.1}$$

Since $b(0) > 0$ and $b(1) < 0$, it follows that $x_i > 0$ for $x = 0$, and $x_i < 1$ for $x = 1$. For θ sufficiently small, $dx_i/dx > 0$, hence x_i maps $I = [0, 1]$ into I and the learning model is a well-defined Markov process in I. Its transition operator is

$$Uf(x) = \sum f(x_i) q_i(x), \tag{4.2}$$

where unindexed summations in this section are over $i = 0$ and $i = 1$. If $f \in C^\infty$, then $Uf \in C^\infty$, hence $U^n f \in C^\infty$.

Suppose now that T_t is any Fellerian semigroup satisfying (2.5) and (2.6). We shall use Khintchine's argument to show that, for any $f \in C^\infty$ and $K < \infty$,

$$\max_{0 < n < K/\tau} \| T_{n\tau} f - U^n f \|_0 \to 0 \tag{4.3}$$

as $\tau \to 0$. (This, of course, ensures the uniqueness of T_t.) Let $H_j = T_{(n-j)\tau} U^j f$. Then

$$\begin{aligned} T_{n\tau} f - U^n f = H_0 - H_n &= \sum_{j=0}^{n-1} (H_j - H_{j+1}) \\ &= \sum_{j=0}^{n-1} T_{(n-j-1)\tau} (T_\tau - U) U^j f. \end{aligned}$$

Therefore

$$\| T_{n\tau} f - U^n f \|_0 < \sum_{j=0}^{n-1} \| (T_\tau - U) U^j f \|_0. \tag{4.4}$$

If $g \in C^\infty$, a third-order Taylor expansion in conjuction with (3.6)–(3.9) yields $Ug = g + \tau \Gamma g + |g|_3 o(\tau)$. Similarly, in view of (2.1) – (2.4), $T_\tau g = g + \tau \Gamma g + |g|_3 o(\tau)$, where, in both cases, $o(\tau)$ is uniform over x and g. Thus

$$\| (T_\tau - U) g \|_0 < |g|_3 \tau \varepsilon_\tau,$$

where $\varepsilon_\tau \to 0$ as $\tau \to 0$, and (4.4) implies that

$$\| T_{n\tau} f - U^n f \|_0 < \tau \varepsilon_\tau \sum_{j=0}^{n-1} | U^j f |_3. \tag{4.5}$$

Suppose now that there were a bound of the form

$$| Uf |_k < \exp(\gamma_k \tau) | f |_k \tag{4.6}$$

for $k \geqslant 1$ and $0 < \theta < \rho_k$, where $\gamma_k \geqslant 0$. Then

$$|U^j f|_k \leqslant \exp(\gamma_k j\tau)|f|_k. \tag{4.7}$$

Taking $k = 3$ and substituting into (4.5), we obtain

$$\|T_{n\tau}f - U^n f\|_0 \leqslant \varepsilon_\tau |f|_3 \int_0^{n\tau} \exp(\gamma_3 t)\, dt \tag{4.8}$$

from which (4.3) follows.

As a consequence of (4.3) and strong continuity of T_t, $\|U^n f - T_t f\|_0 \to 0$ as $\tau \to 0$, and $n\tau \to t$. In conjunction with (4.7), this implies that $T_t f \in C^\infty$ and

$$|T_t f|_k \leqslant \exp(\gamma_k t)|f|_k \tag{4.9}$$

for all $t \geqslant 0$, $k \geqslant 1$, and $f \in C^\infty$.

It remains only to establish (4.6). This inequality follows directly from the expression

$$\begin{aligned}(d/dx)^j(f(x_i)q_i) = {} & f^{(j)}(x_i)(1 + j\theta s_i\sigma_i')q_i + \|f^{(j)}\|_0\omega_i j\theta s_i\sigma_i q_i' \\ & + \sum_{n=0}^{j-1}\binom{j}{n} f^{(n)}(x) q_i^{(j-n)} + O(|f|_j\tau) \\ & + \sum_{n=1}^{j-1} f^{(n)}(x)(\theta s_i\sigma_i q_i)^{(j-n+1)},\end{aligned} \tag{4.10}$$

where $|\omega_i| \leqslant 1$, valid for $j \geqslant 1$. The main points of the derivation of (4.10) are apparent in the case $j = 2$, to which we shall restrict our attention.

Clearly

$$(d/dx)^2(f(x_i)q_i) = \left[f''(x_i)(x_i')^2 + f'(x_i)x_i''\right]q_i + 2f'(x_i)x_i'q_i' + f(x_i)q_i''. \tag{4.11}$$

But

$$\begin{aligned}(x_i')^2 &= 1 + 2\theta s_i\sigma_i' + O(\tau),\\ f'(x_i)x_i'' &= f'(x)\theta s_i\sigma_i'' + O(|f|_2\tau),\\ f'(x_i)x_i' &= f'(x)x_i' + \|f''\|_0\omega_i(x_i - x)x_i'\\ &= f'(x)(1 + \theta s_i\sigma_i') + \|f''\|_0\omega_i\theta s_i\sigma_i + O(|f|_2\tau),\end{aligned}$$

and

$$f(x_i) = f(x) + f'(x)\theta s_i\sigma_i + O(|f|_2\tau).$$

When these expressions are substituted into (4.11), the case $j = 2$ of (4.10) is obtained.

Note that $\Sigma\, q_i = 1$ and

$$\sum s_i\sigma_i q_i = 0; \tag{4.12}$$

hence, in both cases, corresponding sums of derivatives are zero. Thus summation of (4.10) over i yields

$$\begin{aligned}(Uf)^{(j)}(x) = {} & \sum f^{(j)}(x_i)(1 + j\theta s_i\sigma_i')q_i \\ & + \sum \|f^{(j)}\|_0\omega_i j\theta s_i\sigma_i q_i' + O(|f|_j\tau).\end{aligned}$$

Now $s_i q_i' \geqslant 0$ and, for θ sufficiently small, $1 + j\theta s_i \sigma_i' \geqslant 0$; hence

$$|(Uf)^{(j)}(x)| \leqslant \|f^{(j)}\|_0 \sum (1 + j\theta s_i \sigma_i') q_i$$
$$+ \|f^{(j)}\|_0 \sum j\theta s_i \sigma_i q_i' + K_j |f|_j \tau$$
$$= \|f^{(j)}\|_0 + K_j |f|_j \tau,$$

in view of (4.12). Thus $\|(Uf)^{(j)}\|_0 \leqslant \|f^{(j)}\|_0 + K_j|f|_j\tau$. Summing over $1 \leqslant j \leqslant k$, we obtain $|Uf|_k \leqslant (1 + \gamma_k \tau)|f|_k$, from which (4.6) follows.

5. Proof of Theorem 1. Better bound for $(\partial/\partial t)^k T_t f$. Throughout this section we continue to assume that a is a polynomial and b and f are infinitely differentiable. Clearly

$$T_t f - f = \int_0^t \Gamma T_u f \, du \tag{5.1}$$

and

$$\Gamma T_u f = T_u \Gamma f. \tag{5.2}$$

Since $\Gamma f \in C^\infty$, it follows that $T_u \Gamma f \in C^\infty$ and (4.9) applies to Γf. Thus integration and $(\partial/\partial x)^i$ can be interchanged on the right in (5.1), so that

$$(\partial/\partial x)^i T_t f - f^{(i)} = \int_0^t (\partial/\partial x)^i \Gamma T_u f \, du. \tag{5.3}$$

The integrand is bounded, so $(\partial/\partial x)^i T_t f$ is continuous in t (with respect to $\|\cdot\|_0$) for $f \in C^\infty$. Applying this observation to Γf, we see that the integrand in (5.3) is continuous, from which it follows that $(\partial/\partial x)^i T_t f$ is t-differentiable, and

$$(\partial/\partial t)(\partial/\partial x)^i T_t f = (\partial/\partial x)^i \Gamma T_t f. \tag{5.4}$$

Let $w_i(t, x) = (\partial/\partial x)^i T_t f(x)$, and let

$$\Gamma_i = 2^{-1} a (\partial/\partial x)^2 + b_i (\partial/\partial x),$$

where $b_i = b + (i/2)a'$. Then (5.4) can be rewritten in the form

$$(\partial/\partial t) w_i = \Gamma_i w_i + h_i, \tag{5.5}$$

where $h_i = \sum_{j=1}^{i} d_{ij} w_j$ and d_{ij} is as in the statement of Theorem 1.

We shall shortly use (5.5) and a maximum principle to derive the bound

$$\|w_i\|_{0,A} \leqslant \|f^{(i)}\|_0 + A\|h_i\|_{0,A}, \tag{5.6}$$

for all $A > 0$ and $i \geqslant 1$, where $\|w\|_{0,A} = \sup_{0 < t < A, x \in I} |w(t, x)|$. Let us accept (5.6) for the moment and pursue its implications. Estimating h_i in an obvious way, summing over $1 \leqslant i \leqslant k$, and then interchanging i and j summation in the final term on the right, we find that

$$\sum_{i=1}^{k} \|w_i\|_{0,A} \leqslant |f|_k + A\lambda_k \sum_{j=1}^{k} \|w_j\|_{0,A}.$$

Thus, if $A < 1/\lambda_k$,

$$\sum_{i=1}^{k} \|(\partial/\partial x)^i T_t f\|_{0,A} \leqslant |f|_k / (1 - A\lambda_k).$$

Consequently $|T_A f|_k \leqslant |f|_k/(1 - A\lambda_k)$, and, iterating,

$$|T_{nA} f|_k \leqslant |f|_k/(1 - A\lambda_k)^n.$$

Taking $A = t/n$ and letting $n \to \infty$ we obtain (2.7).

It remains only to derive (5.6). As a consequence of (5.2), (5.4), and (4.9), both $(\partial/\partial x)w_i$ and $(\partial/\partial t)w_i$ are bounded throughout $[0, A] \times I$; hence w_i is continuous, as is h_i. Moreover $w_i(t,\cdot) \in C^2$ and b_i, like $b_0 = b$, is continuous with $b_i(0) \geqslant 0$ and $b_i(1) \leqslant 0$. Henceforth we suppress the irrelevant i subscript.

For any ζ and ξ, let $W = e^{-\zeta t}w - \xi t$. Then (5.5) implies that

$$(\partial/\partial t)W = \Gamma W - \zeta W + e^{-\zeta t}h - \xi(1 + \zeta t).$$

If $\zeta > 0$ and

$$\xi = \|h\|_{0,A}, \tag{5.7}$$

then

$$(\partial/\partial t)W \leqslant \Gamma W - \zeta W, \tag{5.8}$$

for $0 \leqslant t \leqslant A$.

Since W is continuous, it attains its maximum over $[0, A] \times I$ at some point (t_0, x_0). If $0 < x_0 < 1$, then $(\partial/\partial x)W(t_0, x_0) = 0$, so

$$\Gamma W(t_0, x_0) = 2^{-1}a(x_0)(\partial/\partial x)^2 W(t_0, x_0) \leqslant 0.$$

If $x_0 = 0$ or 1, then $\Gamma W(t_0, x_0) = b(x_0)(\partial/\partial x)W(t_0, x_0) \leqslant 0$. Thus, in any case, $\Gamma W(t_0, x_0) \leqslant 0$.

If $t_0 > 0$, then $(\partial/\partial t)W(t_0, x_0) \geqslant 0$, so (5.8) yields $W(t_0, x_0) \leqslant 0$. Moreover $W(0, x_0) \leqslant \|f^{(i)}\|_0$. Hence, regardless of the value of t_0, $W(t_0, x_0) \leqslant \|f^{(i)}\|_0$ or

$$e^{-\zeta t}w(t, x) \leqslant \|f^{(i)}\|_0 + \xi A,$$

for $t \leqslant A$. Letting $\zeta \to 0$, this yields

$$w(t, x) \leqslant \|f^{(i)}\|_0 + \xi A.$$

Applying this inequality to $-w$ in place of w and recalling (5.7), we obtain (5.6).

6. Proof of Theorem 1. From C^∞ to D^m. Suppose that $m \geqslant 2$ and that $a \in D^m$ and $b \in D^m$ satisfy the hypotheses of Theorem 1, as do polynomials a_n and b_n. Let T_t and $T_{n,t}$ be corresponding semigroups whose generators, Γ and Γ_n, satisfy (2.5) and (2.6). Since a_n and b_n satisfy the hypotheses of §§4 and 5, $T_{n,t}$ is uniquely determined and, if $f \in C^\infty$, then $T_{n,t}f \in C^\infty$ and

$$|T_{n,t}f|_k \leqslant \exp(\lambda_{n,k}t)|f|_k \tag{6.1}$$

for all $k \geqslant 1$. Of course, $\lambda_{n,k}$ is defined by (2.8) and (2.9) with a_n and b_n in place of a and b.

If $f \in C^\infty$, then $T_{n,t}f \in \mathfrak{D}(\Gamma)$, so $T_{t-u}T_{n,u}f$ can be differentiated with respect to u. Its derivative is

$$(d/du)T_{t-u}T_{n,u}f = T_{t-u}(\Gamma_n - \Gamma)T_{n,u}f;$$

hence $T_t f - T_{n,t} f = \int_0^t T_{t-u}(\Gamma - \Gamma_n) T_{n,u} f \, du$. Consequently,

$$\|T_t f - T_{n,t} f\|_0 \leq \int_0^t \|(\Gamma - \Gamma_n) T_{n,u} f\|_0 \, du$$

$$\leq \max\{2^{-1}\|a - a_n\|_0, \|b - b_n\|_0\} |f|_2 \int_0^t \exp(\lambda_{n,2} u) \, du. \quad (6.2)$$

(This inequality is analogous to (4.8).)

We shall show momentarily that the polynomials a_n and b_n can be chosen so that

$$\lim_{n\to\infty} \|a_n - a\|_0 = 0, \quad (6.3)$$

$$\lim_{n\to\infty} \|b_n - b\|_0 = 0, \quad (6.4)$$

and

$$\lim_{n\to\infty} \lambda_{n,k} = \lambda_k, \qquad 1 \leq k \leq m. \quad (6.5)$$

Since $m \geq 2$, it follows from (6.2) that

$$\lim_{n\to\infty} \|T_t f - T_{n,t} f\|_0 = 0. \quad (6.6)$$

Thus T_t is uniquely determined by (2.5) and (2.6). Furthermore, in the light of the lemma that follows, (6.1), (6.5), and (6.6) imply that, if $f \in C^\infty$, then $T_t f \in D^m$ and (2.7) holds for $1 \leq k \leq m$.

LEMMA 1. *If* $k \geq 1$, $g_n \in D^k$, $\lim_{n\to\infty} \|g_n - g\|_0 = 0$ *and* $\liminf_{n\to\infty} |g_n|_k < \infty$, *then* $g \in D^k$ *and*

$$|g|_k \leq \liminf_{n\to\infty} |g_n|_k.$$

PROOF. We may assume, without loss of generality, that $|g_n|_k \leq K < \infty$. Then the sequences $\{g_n^{(j)}\}$ are bounded and equicontinuous for $1 \leq j < k$, so a straightforward compactness argument yields $g \in C^{k-1}$ and $\|g_n - g\|_{k-1} \to 0$. Let

$$L(h) = \sup_{x \neq y} |h(x) - h(y)|/|x - y|.$$

Then $L(h) < \infty$ if and only if h is absolutely continuous and $\|h'\|_0 < \infty$. Moreover, in this case, $L(h) = \|h'\|_0$. Since $g_n^{(k-1)}(x) \to g^{(k-1)}(x)$ for all x, $L(g^{(k-1)}) \leq \liminf_{n\to\infty} L(g_n^{(k-1)})$ and the conclusions of the lemma follow immediately.

To extend (2.7) from $f \in C^\infty$ to $f \in D^k$, $1 \leq k \leq m$, assume that $f \in D^k$, and consider the Taylor expansion

$$f(x) = \sum_{j=0}^{k-1} f^{(j)}(0) \frac{x^j}{j!} + \frac{1}{(k-1)!} \int_0^x (x - y)^{k-1} f^{(k)}(y) \, dy. \quad (6.7)$$

Since $f^{(k)} \in D^0$, there is a sequence, p_n, of polynomials such that $p_n \to f^{(k)}$ almost everywhere and $\|p_n\|_0 \to \|f^{(k)}\|_0$, from which it follows that

$$\lim_{n\to\infty} \int_0^1 |p_n(x) - f^{(k)}(x)| \, dx = 0.$$

Replacing $f^{(k)}$ in (6.7) by p_n, we obtain a sequence, $\bar{f}_n$, of polynomials such that $\|\bar{f}_n - f\|_{k-1} \to 0$ and $\bar{f}_n^{(k)} = p_n$, so $\|\bar{f}_n^{(k)}\|_0 \to \|f^{(k)}\|_0$. Consequently $|\bar{f}_n|_k \to |f|_k$. Now apply (2.7) to $\bar{f}_n$. Since $\lim_{n\to\infty}\|T_t\bar{f}_n - T_tf\|_0 = 0$, Lemma 1 implies that $T_tf \in D^k$ and (2.7) holds if $f \in D^k$.

To complete the proof of Theorem 1, it remains only to show that there are polynomials a_n and b_n satisfying (6.3)–(6.5) in addition to the hypotheses of Theorem 1. Let $\bar{b}_n$ be a sequence of polynomials constructed as in the last paragraph for $k = m$, and let $b_n(x) = \bar{b}_n(x) + (b(1) - \bar{b}_n(1))x$. Then

$$\lim_{n\to\infty} \|b_n - b\|_{m-1} = 0, \tag{6.8}$$

$$\lim_{n\to\infty} \|b_n^{(m)}\|_0 = \|b^{(m)}\|_0, \tag{6.9}$$

and $\bar{b}_n(i) = b(i)$, so $\bar{b}_n(0) \geqslant 0$ and $\bar{b}_n(1) \leqslant 0$.

The construction of a_n is similar, but certain refinements are required in order to satisfy (6.5) for $k = m$ and $a_n(x) > 0$ for $0 < x < 1$. The only coefficient d_{ij}, $1 \leqslant j \leqslant i \leqslant m$, that involves $a^{(m)}$ is

$$d_{m2} = mb^{(m-1)} + 2^{-1}a^{(m)}.$$

Let $d^*_{n,m2}$ be a sequence of polynomials such that $d^*_{n,m2} \to d_{m2}$ almost everywhere and $\|d^*_{n,m2}\|_0 \to \|d_{m2}\|_0$. Let p_n be the polynomial defined by $d^*_{n,m2} = mb_n^{(m-1)} + 2^{-1}p_n$. Then $p_n \to a^{(m)}$ almost everywhere and

$$\int_0^1 |p_n(x) - a^{(m)}(x)|\, dx \to 0.$$

Let $\bar{a}_n$ be polynomial approximations to a constructed as in the penultimate paragraph, let $\tilde{a}_n(x) = \bar{a}_n(x) - \bar{a}_n(1)x$, and, finally, let

$$a_n(x) = \tilde{a}_n(x) + (16/3)\|a - \tilde{a}_n\|_1 x(1 - x).$$

Then $a_n(0) = a_n(1) = 0$, $a_n(x) \geqslant a(x)$ so that $a_n(x) > 0$ for all $0 < x < 1$, $\lim_{n\to\infty}\|a_n - a\|_{m-1} = 0$, and $\lim_{n\to\infty}\|d_{n,m2}\|_0 = \|d_{m2}\|_0$. The latter equations, together with (6.8) and (6.9), readily yield

$$\lim_{n\to\infty} \|d_{n,ij}\|_0 = \|d_{ij}\|_0$$

for all $1 \leqslant j \leqslant i \leqslant m$, from which (6.5) follows easily.

This completes the proof of Theorem 1.

7. Proof of Theorem 2. We shall restrict our attention to the upper boundary, $r_1 = 1$. Since $a \in D^2$ and $b \in D^1$, we have at our disposal the Taylor expansions $b(x) = -\beta + O(1 - x)$ and

$$a(x) = \alpha(1 - x) + O\big((1 - x)^2\big), \tag{7.1}$$

where $\beta = \beta_1$ and $\alpha = \alpha_1$ are nonnegative.

LEMMA 2. *If $\beta > 0$, then $v(1) < \infty$.*

PROOF. Choose $c \in (r, 1)$ sufficiently large that $b(x) < -\beta/2$ for $x > c$. Then for $1 > z > c$,

$$(\beta/2)\int_c^z p(x)\,dm(x) \leqslant -\int_c^z p(x)b(x)\,dm(x) = -\int_c^z p(x)\,de^{B(x)}$$
$$= p(c)e^{B(c)} - p(z)e^{B(z)} + \int_c^z e^{B(x)}\,dp(x)$$
$$\leqslant p(c)e^{B(c)} + (1 - c).$$

Thus $v(1) = v(1^-) < \infty$, as claimed.

LEMMA 3. *If $\beta > 0$ and $\alpha = 0$, then 1 is entrance.*

PROOF. Since $2/a(x) = p'(x)m'(x)$, the Schwarz inequality yields

$$\int_r^1 (2/a(x))^{1/2}\,dx \leqslant \left(\int_r^1 p'(x)\,dx \int_r^1 m'(x)\,dx\right)^{1/2} = (p(1)m(1))^{1/2}.$$

If $\alpha = 0$, it follows from (7.1) that $p(1) = \infty$ or $m(1) = \infty$, so $u(1) = \infty$ or $v(1) = \infty$. In view of Lemma 2, $v(1) < \infty$, so $u(1) = \infty$ and 1 is entrance.

LEMMA 4. *If $\beta > 0$, then 1 is entrance if $0 < \alpha \leqslant 2\beta$ and regular if $\alpha > 2\beta$.*

PROOF. If $\alpha > 0$, then

$$\frac{b(x)}{a(x)} = \frac{-\beta}{\alpha(1 - x)} + O(1)$$

for $r \leqslant x < 1$. Hence $B(x) = 2(\beta/\alpha)\ln(1 - x) + O(1)$ and $p'(x) = (1 - x)^{-2(\beta/\alpha)}e^{O(1)}$. If $2\beta \geqslant \alpha$ then $p(1) = \infty$, $u(1) = \infty$, and 1 is entrance by Lemma 2. If $2\beta < \alpha$, then $p(1) < \infty$. But $m(1) < \infty$ by Lemma 2, so 1 is regular.

LEMMA 5. *If $\beta = 0$ and $\alpha > 0$, then 1 is exit.*

PROOF. Under these conditions $B(x) = O(1)$; hence $m(x) = O(-\ln(1 - x))$ for $r \leqslant x < 1$, and $u(1) < \infty$. However, for any $c \in (r, 1)$, $v(1) \geqslant K\int_c^1 (a(x))^{-1}\,dx$ for some $K > 0$, so $v(1) = \infty$.

LEMMA 6. *If $\beta = 0$ and $\alpha = 0$, then 1 is natural.*

PROOF. We shall transform $(0, 1)$ onto $(0, \infty)$ in order to apply a result of Ethier [7, Lemma 3]. Let $h(x) = x/(1 - x)$, $\tilde{p}(y) = p(h^{-1}(y))$, and $\tilde{m}(y) = m(h^{-1}(y))$. Then

$$\int_r^1 m\,dp = \int_{h(r)}^\infty \tilde{m}\,d\tilde{p},$$

and similarly for $\int_r^1 p\,dm$, so the boundary classification of ∞ for $\tilde{p}$ and $\tilde{m}$ is the same as the classification of 1 for p and m. It is easy to check that the coefficients $\tilde{a}$ and $\tilde{b}$ corresponding to $\tilde{p}$ and $\tilde{m}$ are

$$\tilde{a}(y) = a(x)(1 + y)^4$$

and

$$\tilde{b}(y) = a(x)(1 + y)^3 + b(x)(1 + y)^2,$$

where $x = h^{-1}(y) = y/(1 + y)$. If $\alpha = 0$ and $\beta = 0$, then $\tilde{a}(y) = O(y^2)$ and $\tilde{b}(y) = O(y)$ for $y \geq 1$. Hence Ethier's lemma implies that ∞ is natural for $\tilde{p}$ and $\tilde{m}$, so that 1 is natural for p and m.

This completes the proof of Theorem 2.

References

1. M. F. Norman, *Sociobiological variations on a Mendelian theme*, these PROCEEDINGS.

2. D. Freedman, *Brownian motion and diffusion*, Holden-Day, San Francisco, Calif., 1971.

3. P. Mandl, *Analytical treatment of one-dimensional Markov processes*, Springer-Verlag, New York, 1968.

4. M. F. Norman, *Markov processes and learning models*, Academic Press, New York, 1972.

5. ______, *Diffusion approximation of non-Markovian processes*, Ann. Probab. **3** (1975), 358–364.

6. A. Khintchine, *Asymptotische Gesetze der Wahrscheinlichkeitsrechnung*, Chelsea, New York, 1948.

7. S. N. Ethier, *Differentiability preserving properties of Markov semigroups associated with one-dimensional diffusions*, Z. Wahrsch. Verw. Gebiete **45** (1978), 225–238.

8. R. R. Bush and F. Mosteller, *Stochastic models for learning*, Wiley, New York, 1955.

9. M. F. Norman, *Slow learning with small drift in two-absorbing-barrier models*, J. Math. Psych. **8** (1971), 1–21.

10. O. Onicescu and G. Mihoc, *Sur les chaînes de variables statistiques*, Bull. Soc. Math. France **59** (1935), 174–192.

11. S. N. Ethier and M. F. Norman, *Error estimate for the diffusion approximation of the Wright-Fisher model*, Proc. Nat. Acad. Sci. U.S.A. **74** (1977), 5096–5098.

DEPARTMENT OF PSYCHOLOGY, UNIVERSITY OF PENNSYLVANIA, PHILADELPHIA, PENNSYLVANIA 19104