JOURNAL OF MATHEMATICAL PSYCHOLOGY 8, 1–21 (1971)

Slow Learning with Small Drift in Two-Absorbing-Barrier Models[1]

M. FRANK NORMAN

University of Pennsylvania, Philadelphia, Pennsylvania 19104

Certain two-choice learning models have the property that both the first and second moments of increments Δp_n in A_1 response probability p_n are of the order of magnitude of a parameter τ. It is shown that the distribution of p_n can be approximated by that of $X_{n\tau}$, where X_t is a Markov process that is continuous in space and time.

1. INTRODUCTION

Let p_n be a subject's probability of response A_1 (rather than A_2) on trial $n \geqslant 0$ in a two-choice learning experiment where reward follows A_i with probability π_i. In the standard linear model for experiments where non-reward is assumed to have no effect, p_n is a Markov process in the closed unit interval with transition probabilities

$$\Delta p_n = \begin{cases} \theta_1 q_n & \text{with prob. } \pi_1 p_n \\ -\theta_2 p_n & \text{with prob. } \pi_2 q_n \\ 0 & \text{with prob. } 1 - \pi_1 p_n - \pi_2 q_n, \end{cases} \tag{1.1}$$

where $q_n = 1 - p_n$ and $1 \geqslant \theta_i$, $\pi_i > 0$. The first row applies to a rewarded A_1, the second to a rewarded A_2, and the thrid to non-reward. For this model p_n converges to one or the other of the absorbing barriers 0 and 1 as $n \to \infty$. Also

$$\begin{aligned} E[\Delta p_n \mid p_n = p] &= (\pi_1\theta_1 - \pi_2\theta_2)\, pq, \\ E[(\Delta p_n)^2 \mid p_n = p] &= (\pi_1\theta_1^2 q + \pi_2\theta_2^2 p)\, pq, \end{aligned} \tag{1.2}$$

and

$$E[| \Delta p_n |^3 \mid p_n = p] = (\pi_1\theta_1^3 q^2 + \pi_2\theta_2^3 p^2)\, pq,$$

for all $0 \leqslant p \leqslant 1$.

We wish to study the dependence of p_n on π_i and θ_i when θ_i is small and the *drift* $E[\Delta p_n \mid p_n = p]$ is even smaller, $O(\theta_1\theta_2)$, although π_i is bounded away from 0. Consider for example, the important special case $\theta_1 = \theta_2 = \theta$, which arises when the same reward is used for A_1 and A_2. Fix b and $\pi_2 > 0$, and define π_1 by means of the equation

[1] This research was supported by the National Science Foundation under Grant GP-7335.

© 1971 by Academic Press, Inc.

480/8/1-1

$b = (\pi_1 - \pi_2)/\theta$. Thus b is the advantage of the reward probability of A_1 over that of A_2, relative to the learning rate parameter θ, and $E[\Delta p_n \mid p_n = p] = \theta^2 bpq$. More generally, if we let d_i be a fixed real number and c_i and e_i be fixed positive numbers such that $c_1 e_1 = c_2 e_2$, and let $\theta_i = c_i \theta$ and $\pi_i = e_i + \theta d_i$, (1.2) yields

$$\begin{aligned} E[\Delta p_n \mid p_n = p] &= \tau bpq + o(\tau), \\ E[(\Delta p_n)^2 \mid p_n = p] &= \tau(a_1 q + a_2 p)\, pq + o(\tau), \end{aligned} \tag{1.3}$$

and

$$E[| \Delta p_n |^3 \mid p_n = p] = o(\tau),$$

where $\tau = \theta^2$, $b = c_1 d_1 - c_2 d_2$, $a_i = c_i^2 e_i$, and $o(\tau)$ in the first equation is actually 0.

In Sec. 2 it is shown that, if $y_n = y_n^\tau$ is any other family of Markov processes satisfying (1.3) with $y_0 = p_0$, then $\Lambda(p_n) - \Lambda(y_n)$ converges (weakly) to 0 as $\tau \to 0$, uniformly in p_0 and $n\tau \leqslant M$, for any $M < \infty$. $\Lambda(z)$ is the distribution or law of the random variable z. In a special case, convergence is uniform over all n. This is an instance of an *invariance theorem* for asymptotic distributions (see Breiman, 1968, Sec. 8.6, for a simple example). As is shown in Sec. 3, the A_1 response probability y_n in the N element stimulus sampling model with fixed sample size $s > 1$ is such a sequence ($\tau = \epsilon^2/(1 - \epsilon) \sim \epsilon^2$, $\epsilon = 1/N$) if the parameters of the model are constrained analogously to those in the linear model above.

In Sec. 4 it is noted that there is a continuous time Markov process X_t with continuous sample paths (i.e., a *diffusion*) such that $x_n^\tau = X_{n\tau}$ satisfies (1.3). From this and the invariance theorem it follows that $\Lambda(y_n) \to \Lambda(X_t)$ as $\tau \to 0$, $n\tau \to t$, and $y_0 \to x = X_0$, for any family y_n of processes satisfying (1.3). In fact, the distribution of the entire process $\{y_n^\tau\}_{n\tau \leqslant M}$ converges to that of the process $\{X_t\}_{t \leqslant M}$ as $\tau \to 0$. Our approach yields as a dividend some new results concerning the semigroup of transition operators associated with the limiting process (Sec. 5). Extension and sharpening of these results would permit corresponding improvement in our limit theorems.

Diffusion approximations to discrete parameter Markov processes were considered systematically by Khintchine (1948), and similar approximations have established their usefulness in population genetics (Kimura, 1964; Watterson, 1962) and other areas (Iglehart, 1968). Previous results (Norman, 1968c, Theorem 3.1 and Lemma 5.3) show that, for the linear and stimulus sampling models considered in this paper, the A_1 response probability p_n is approximately normally distributed with mean p_0 and variance $r(nr)\, a(p_0)$ [see (1.4) below for $a(p)$] when $r = \theta$ or ϵ is small and $nr \leqslant M$. Since all the mass of $\Lambda(p_n)$ accumulates at 0 and 1 as $n \to \infty$, this approximation clearly becomes useless for larger values of n, and this failure was a major impetus to the present study. The approximation $\Lambda(X_{n\tau})$ to $\Lambda(p_n)$ obtained in this paper, valid at least in $nr^2 \leqslant M$, is much more satisfactory in this regard.

In Sec. 6 it is shown that

$$\gamma(p) = P(\lim_{n\to\infty} p_n = 1 \mid p_0 = p)$$

converges as $\tau \to 0$ to

$$\Gamma(p) = P(\lim_{t\to\infty} X_t = 1 \mid X_0 = p)$$

under hypotheses that include linear and stimulus sampling models, as well as families of processes with other functions

$$\begin{aligned} b(p) &= \lim_{\tau\to 0} \tau^{-1} E[\Delta p_n \mid p_n = p] \\ a(p) &= \lim_{\tau\to 0} \tau^{-1} E[(\Delta p_n)^2 \mid p_n = p] \end{aligned} \tag{1.4}$$

than those arising in (1.3). $\Gamma(p)$ is the solution of the differential equation,

$$b(p)\, \Gamma'(p) + \tfrac{1}{2} a(p)\, \Gamma''(p) = 0, \tag{1.5}$$

for which $\Gamma(0) = 0$ and $\Gamma(1) = 1$. Mosteller and Tatsuoka (1960) proposed that $\gamma(p)$ for certain linear models could be approximated by the solution Φ of

$$E[\Delta p_n \mid p_n = p]\, \Phi'(p) + \tfrac{1}{2} E[(\Delta p_n)^2 \mid p_n = p]\, \Phi''(p) = 0$$

with $\Phi(0) = 0$ and $\Phi(1) = 1$. It is easy to show that, for the families of linear models considered above, $\Phi(p) \to \Gamma(p)$ as $\theta \to 0$, hence $\Phi(p) - \gamma(p) \to 0$ as $\theta \to 0$. Thus, our results yield an analytical justification for Mosteller and Tatsuoka's approximation. When $a(p)$ and $b(p)$ are given by (1.3) and $a_1 = a_2 = a$,

$$\Gamma(p) = \frac{e^{-2bp/a} - 1}{e^{-2b/a} - 1}, \tag{1.6}$$

and $\gamma(p) \to \Gamma(p)$ as $\theta \to 0$ for the linear model with $c_1 = c_2$ follows from Norman (1968b, Theorem 4, or 1968a, Theorem 1 of Sec. III).

2. Invariance Theorems

The distribution of a Markov process $\{X_n\}_{n\geq 0}$ with state space S is determined, up to specification of the distribution of X_0, by its *transition probability* or ***kernel***

$$L(x, A) = P(X_{k+1} \in A \mid X_k = x),$$

or, equivalently, by the *transition operator*

$$Vf(x) = \int_S L(x, dy) f(y) = E[f(X_{k+1}) \mid X_k = x], \tag{2.1}$$

on bounded measurable real valued functions f. Clearly V is positive ($Vf \geqslant 0$ if $f \geqslant 0$), preserves constant functions, and is a contraction ($|Vf| \leqslant |f|$ where $|g| = \sup_{x \in S} |g(x)|$). For all $n \geqslant 0$ the n step transition probabilities

$$L^{(n)}(x, A) = P(X_{k+n} \in A \mid X_k = x)$$

bear the same relation to the n th iterate V^n of V that $L = L^{(1)}$ does to $V = V^1$:

$$V^n f(x) = \int_S L^{(n)}(x, dy) f(y) = E[f(X_{k+n}) \mid X_k = x].$$

The operator V^n is our basic tool for studying the distribution $L^{(n)}(x, \cdot)$ of X_n when $X_0 = x$. If P and Q are two probability distributions on a real interval I, then $P - Q$ is regarded as close to 0 (in the weak sense) in so far as

$$\int_I f(x)(P(dx) - Q(dx)) = \int_I f(x)\, P(dx) - \int_I f(x)\, Q(dx)$$

is close to 0 for bounded continuous functions f ($f \in C(I)$). Thus, if $S \subset I$, $L^{(n)}(x, \cdot) - Q$ is close to 0 in so far as $V^n f(x) - \int_I f(x)\, Q(dx)$ is close to 0 for $f \in C(I)$.

For $x \in S$ let

$$v_1(x) = \int_S L(x, dy)(y - x) = E[\Delta X_n \mid X_n = x],$$

$$v_2(x) = \int_S L(x, dy)(y - x)^2 = E[(\Delta X_n)^2 \mid X_n = x],$$

and

$$v_3(x) = \int_S L(x, dy) \mid y - x \mid^3 = E[|\Delta X_n|^3 \mid X_n = x],$$

be the *transition moments* of L. Let $u_1(x)$, $u_2(x)$, and $u_3(x)$ be the transition moments of a kernel K with transition operator U in an interval I that includes S. Suppose that U preserves smoothness to the extent that $Uf \in C^3(I)$ whenever $f \in C^3(I)$. ($C^k(I)$ is the set of real valued functions on I with k bounded continuous derivatives; $C^0(I) = C(I)$.) Lemma 2.1 and Corollary 2.2 show that, for $f \in C^3(I)$, $U^n f$ and $V^n f$ will be approximately equal if u_i and v_i are approximately equal, $i = 1, 2$, u_3 and v_3 are small, and the derivatives $(U^k f)^{(j)}$, $j = 1, 2, 3$, $k < n$ are not large. These very general bounds will be combined with bounds on $(U^k f)^{(j)}$ for the linear model (Lemma 2.3) to yield our invariance theorems (Theorem 2.4 and Corollaries 2.5 and 2.6).

LEMMA 2.1. *For $f \in C^3(I)$ and $n \geqslant 0$*

$$| U^n f - V^n f | \leqslant \sum_{k=0}^{n-1} \sum_{j=1}^{3} h_j \, |(U^k f)^{(j)}|, \tag{2.2}$$

where

$$h_1 = | u_1 - v_1 |, \qquad h_2 = | u_2 - v_2 |/2, \qquad \textit{and} \qquad h_3 = (| u_3 | + | v_3 |)/6.$$

Throughout the paper $| g |$ denotes the supremum of $| g(x)|$ over the largest set on which g is defined, e.g., I in the case of u_3 and S in the case of $u_1 - v_1$.

Proof. For $x \in S$

$$U^n f(x) - V^n f(x) = Ug(x) - Vg(x) + V(U^{n-1}f - V^{n-1}f)(x), \tag{2.3}$$

where $g = U^{n-1}f$. Since V is a contraction,

$$| V(U^{n-1}f - V^{n-1}f)(x)| \leqslant | U^{n-1}f - V^{n-1}f |. \tag{2.4}$$

The other term on the right in (2.3) is rewritten

$$\begin{aligned} Ug(x) - Vg(x) &= [Ug(x) - g(x) - Fg(x)] \\ &\quad - [Vg(x) - g(x) - Gg(x)] + [Fg(x) - Gg(x)], \end{aligned} \tag{2.5}$$

where

$$Fg(x) = u_1(x)\, g'(x) + 2^{-1}u_2(x)\, g''(x)$$

and

$$Gg(x) = v_1(x)\, g'(x) + 2^{-1}v_2(x)\, g''(x).$$

Substituting the third order Taylor expansion

$$g(y) = g(x) + (y - x)\, g'(x) + (y - x)^2 g''(x)/2 + \omega \,| y - x |^3 \,| g^{(3)} |/6,$$

$| \omega | \leqslant 1$, of $g \in C^3(I)$ for f in (2.1) we obtain the basic inequality

$$| Vg(x) - g(x) - Gg(x)| \leqslant | v_3 | \, | g^{(3)} |/6. \tag{2.6}$$

Similarly,

$$| Ug(x) - g(x) - Fg(x)| \leqslant | u_3 | \, | g^{(3)} |/6. \tag{2.7}$$

Finally, it is clear that

$$| Fg(x) - Gg(x)| \leqslant | u_1 - v_1 | \, | g' | + | u_2 - v_2 | \, | g'' |/2.$$

Taking absolute values on both sides of (2.3) and using (2.4)–(2.7), we obtain

$$| U^n f - V^n f | \leqslant | U^{n-1} f - V^{n-1} f | + \sum_{j=1}^{3} h_j \, |(U^{n-1} f)^{(j)} |,$$

from which (2.2) follows by induction. Q.E.D.

In Corollary 2.2 we add assumptions about how much a single application of U increases $f^{(j)}$.

COROLLARY 2.2.a. *If* $\sum_{j=1}^{3} |(Uf)^{(j)} | \leqslant \zeta \sum_{j=1}^{3} | f^{(j)} |$ *for all* $f \in C^3(I)$, *then*

$$| U^n f - V^n f | \leqslant \max(h_1, h_2, h_3) \frac{\zeta^n - 1}{\zeta - 1} \sum_{j=1}^{3} | f^{(j)} |$$

for all $n \geqslant 0$ *and* $f \in C^3(I)$.

b. *If* $h_1 = 0$ *and* $\sum_{j=2}^{3} |(Uf)^{(j)} | \leqslant \zeta \sum_{j=2}^{3} | f^{(j)} |$, *then*

$$| U^n f - V^n f | \leqslant \max(h_2, h_3) \frac{\zeta^n - 1}{\zeta - 1} \sum_{j=2}^{3} | f^{(j)} |.$$

Proof.

$$| U^n f - V^n f | \leqslant \max(h_1, h_2, h_3) \sum_{k=0}^{n-1} \sum_{j=1}^{3} |(U^k f)^{(j)} |,$$

and, under a, $\sum_{j=1}^{3} |(U^k f)^{(j)} | \leqslant \zeta^k \sum_{j=1}^{3} | f^{(j)} |$. The proof of b is similar. Q.E.D.

Lemma 2.3 gives suitable constants ζ for the transition operator U associated with the transition equations (1.1). As before we assume that θ_i and π_i depend on θ according to the equations

$$\theta_i = \theta c_i \quad \text{and} \quad \pi_i = e_i + \theta d_i, \tag{2.8}$$

where $c_i, e_i > 0$ and $c_1 e_1 = c_2 e_2$. Clearly $0 < \theta_i, \pi_i$ and $\theta_i \leqslant 1$ for θ sufficiently small. To achieve $\pi_i \leqslant 1$ for θ sufficiently small we assume that either $e_i < 1$, or $e_i = 1$ and $d_i \leqslant 0$. Let $\delta > 0$ be small enough that $0 < \theta_i, \pi_i \leqslant 1$ for $0 < \theta < \delta$.

LEMMA 2.3.a. *There are constants* $\gamma_k > 0$ *such that*

$$\sum_{j=1} |(Uf)^{(j)} | \leqslant (1 + \theta^2 \gamma_k) \sum_{j=1}^{k} | f^{(j)} | \tag{2.9}$$

for all $k \geqslant 1$, $f \in C^k[0, 1]$, *and* $\theta < \delta$.

b. *If* $c_1 = c_2$ *and* $d_1 = d_2$ (*so that* $\theta_1 = \theta_2$ *and* $\pi_1 = \pi_2$) *then*

$$|(Uf)^{(k)} | \leqslant B_k | f^{(k)} | \tag{2.10}$$

where

$$B_k = (1 - \pi_1) + \pi_1((1 - \theta_1)^k + k(1 - \theta_1)^{k-1}\,\theta_1).$$

It is easy to show that $B_1 = 1$, B_k is strictly decreasing, $B_k \to 0$ as $k \to \infty$, and

$$B_k = 1 - k(k-1)\,\theta^2 c_1 e_1/2 + O(\theta^3). \tag{2.11}$$

Proof. Let $\alpha_i = 1 - \theta_i$, $p_1 = \alpha_1 p + \theta_1$ and $p_2 = \alpha_2 p$. Differentiating

$$Uf(p) = f(p_1)\,\pi_1 p + f(p_2)\,\pi_2 q + f(p)(1 - \pi_1 p - \pi_2 q),$$

k times and rearranging, we obtain for $(Uf)^{(k)}(p)$ the expression

$$\begin{aligned}&\alpha_1{}^k f^{(k)}(p_1)\,\pi_1 p + \alpha_2{}^k f^{(k)}(p_2)\,\pi_2 q + f^{(k)}(p)(1 - \pi_1 p - \pi_2 q)\\ &\quad + k\pi_1\alpha_1^{k-1}[f^{(k-1)}(p_1) - f^{(k-1)}(p)] + k\pi_2\alpha_2^{k-1}[f^{(k-1)}(p) - f^{(k-1)}(p_2)]\\ &\quad + k[\pi_1(\alpha_1^{k-1} - 1) + \pi_2(1 - \alpha_2^{k-1})]\,f^{(k-1)}(p).\end{aligned}$$

Taking absolute values we obtain for $|(Uf)^{(k)}(p)|$ the estimate

$$\begin{aligned}&[\alpha_1{}^k\pi_1 p + \alpha_2{}^k\pi_2 q + (1 - \pi_1 p - \pi_2 q)]\,|f^{(k)}|\\ &\quad + k[\pi_1\alpha_1^{k-1}\theta_1 q + \pi_2\alpha_2^{k-1}\theta_2 p]\,|f^{(k)}|\\ &\quad + k\,|\pi_1(\alpha_1^{k-1} - 1) + \pi_2(1 - \alpha_2^{k-1})|\,|f^{(k-1)}|.\end{aligned}$$

b follows immediately. In any case the binomial expansions of $(1 - \theta_i)^k$ and $(1 - \theta_i)^{k-1}$ yield

$$\begin{aligned}|(Uf)^{(k)}(p)| \leqslant{}& [1 + k(\theta_1\pi_1 - \theta_2\pi_2)(q - p) + O(\theta^2)]\,|f^{(k)}|\\ &+ k\,|(k-1)(\theta_2\pi_2 - \theta_1\pi_1) + O(\theta^2)|\,|f^{(k-1)}|,\end{aligned}$$

or, since

$$\theta_1\pi_1 - \theta_2\pi_2 = \theta^2(c_1 d_1 - c_2 d_2),$$

$$|(Uf)^{(k)}| \leqslant (1 + K_k\theta^2)\,|f^{(k)}| + C_{k-1}\theta^2\,|f^{(k-1)}|,$$

for some constants K_k and C_{k-1} ($C_0 = 0$). Adding these inequalities, we obtain

$$\begin{aligned}\sum_{j=1}^{k} |(Uf)^{(j)}| \leqslant{}& (1 + K_k\theta^2)\,|f^{(k)}|\\ &+ \sum_{j=1}^{k-1} (1 + (K_j + C_j)\,\theta^2)\,|f^{(j)}|,\end{aligned}$$

which yields (2.9) with

$$\gamma_k = \max(K_1 + C_1, \ldots, K_{k-1} + C_{k-1}, K_k). \qquad \text{Q.E.D.}$$

We recall that the transition moments $u_i = u_{i,\tau}$ of $U = U_\tau$ are given by

$$\begin{aligned} u_1(p) &= \tau b p q + o(\tau), \\ u_2(p) &= \tau(a_1 q + a_2 p) p q + o(\tau), \end{aligned} \tag{2.12}$$

and

$$u_3(p) = o(\tau),$$

where

$$\tau = \theta^2, \qquad b = c_1 d_1 - c_2 d_2, \qquad \text{and} \qquad a_i = c_i^2 e_i, \tag{2.13}$$

and each $o(\tau)$ is understood to be uniform over p, e.g., $\tau^{-1} \mid u_3 \mid \to 0$ as $\tau \to 0$. Since c_i and e_i are positive, a_i is too. Furthermore, any real b and positive a_1 and a_2 can arise in (2.12). [Given a_i and b, let $c_i = \beta a_i$, $e_1 = \alpha a_2$, and $e_2 = \alpha a_1$ where $\alpha < 1/\max(a_1, a_2)$ and $\beta = 1/(\alpha a_1 a_2)^{1/2}$. Then choose d_1 and d_2 such that $c_1 d_1 - c_2 d_2 = b$. These choices satisfy (2.13) and the restrictions given just after (2.8).] The condition $b = 0$, $a_1 = a_2$ is equivalent to $c_1 = c_2$, $d_1 = d_2$, the hypothesis in Lemma 2.3.b. In this case the transition moments can be written

$$u_1(p) = 0, \qquad u_2(p) = \tau a_1 p q + o(\tau), \qquad \text{and} \qquad u_3(p) = o(\tau). \tag{2.14}$$

Let J be a set of positive real numbers with infimum 0, and, for every $\tau \in J$ let $V = V_\tau$ be a transition operator in a subset S_τ of [0, 1], with transition moments $v_i = v_{i,\tau}$.

THEOREM 2.4.a. *If v_i satisfies* (2.12) (*for $p \in S_\tau$*) *then*

$$\mid U^n f - V^n f \mid \leqslant o(1)[\exp(\gamma_3 n \tau) - 1] \sum_{j=1}^{3} \mid f^{(j)} \mid, \tag{2.15}$$

for $\tau \in J \cap (0, \delta^2)$, $n \geqslant 0$, and $f \in C^3[0, 1]$.

b. *If v_i satisfies* (2.14) *then*

$$\mid U^n f - V^n f \mid \leqslant o(1)[1 - \exp(-\pi_1 c_1^2 n \tau)] \sum_{j=2}^{3} \mid f^{(j)} \mid. \tag{2.16}$$

Here $o(1)$ depends only on τ and converges to 0 as $\tau \to 0$.

Proof. Apply Corollary 2.2.a. Under a above $\max(h_1, h_2, h_3) = o(\tau)$. Since $e^x \geqslant 1 + x$ we can take $\zeta = \exp(\tau \gamma_3)$, according to Lemma 2.3.a. Thus

$$U^n f - V^n f \mid \leqslant o(\tau) \frac{\exp(\gamma_3 n \tau) - 1}{\exp(\gamma_3 \tau) - 1} \sum_{j=1}^{3} \mid f^{(j)} \mid.$$

The inequality (2.15) follows on noting that $[\exp(\gamma_3\tau) - 1]/\tau$ is bounded away from 0. For (2.16), use Corollary 2.2.b with $\zeta = \exp(-\pi_1\theta_1^2)$, in accordance with Lemma 2.3.b. Q.E.D.

COROLLARY 2.5.a. *If v_i satisfies* (2.12) *then*

$$\lim_{\tau\to 0} \sup_{\substack{p\in S_\tau \\ n\tau\leqslant M}} | U^n f(p) - V^n f(p)| = 0 \tag{2.17}$$

for every $f \in C[0, 1]$ and $M < \infty$. b. *If v_i satisfies* (2.14) *then*

$$\lim_{\tau\to 0} \sup_{\substack{p\in S_\tau \\ n\geqslant 0}} | U^n f(p) - V^n f(p)| = 0. \tag{2.18}$$

Proof. For $f \in C^3[0, 1]$ these statements follow immediately from Theorem 2.4. If $f \in C[0, 1]$ and $\epsilon > 0$, let f_ϵ be a polynomial with $|f - f_\epsilon| \leqslant \epsilon$. Since U^n and V^n are contractions

$$| U^n f(p) - V^n f(p)| \leqslant 2\epsilon + | U^n f_\epsilon(p) - V^n f_\epsilon(p)|.$$

The result then follows on taking the relevant sup over p and n, then the lim sup as $\tau \to 0$, and finally letting $\epsilon \to 0$. Q.E.D.

Let K_τ and L_τ be the kernels of U_τ and V_τ, respectively. Statement (2.17) is what we mean when we say that $K_\tau^{(n)}(p, \cdot) - L_\tau^{(n)}(p, \cdot)$ converges weakly to 0 as $\tau \to 0$, uniformly in p and $n\tau \leqslant M$, while (2.18) is the definition of weak convergence of $K_\tau^{(n)}(p, \cdot) - L_\tau^{(n)}(p, \cdot)$ to 0, uniformly over p and $n \geqslant 0$.

The transition operator U_τ for the linear model plays a distinguished role in Theorem 2.4 and Corollary 2.5. There is an interesting variation of these statements that makes no reference to the linear model.

COROLLARY 2.6. *Theorem* 2.4 *and Corollary* 2.5 *remain valid if U_τ, $0 < \tau \leqslant \delta$ is any family of transition operators on* [0, 1] *that satisfies the same condition* [(2.12) *or* (2.14)] *as V_τ.*

Proof. Let $U_{0,\tau}$ be the transition operator for the linear model. Then for $p \in S_\tau$

$$| U_\tau^n f(p) - V_\tau^n f(p)| \leqslant | U_{0,\tau}^n f(p) - U_\tau^n f(p)| + | U_{0,\tau}^n f(p) - V_\tau^n f(p)|.$$

Take the appropriate sup and apply the corresponding previous result to each of the terms on the right. Q.E.D.

3. A Stimulus Sampling Model with Two Absorbing Barriers

Consider the A_1 response probability y_n for the N element stimulus sampling model with fixed sample size s (see Norman, 1968a, pp. 285–286.) A model analogous to the two-absorbing-barrier linear model is obtained by assuming that the probability $(1-\pi_1)\,c_{12}$ $((1-\pi_2)\,c_{21})$ that non-reward follows A_1 (A_2) and effectively reinforces A_2 (A_1) is zero. Let $\pi_i' = \pi_i c_{ii}$ be the probability that A_i is followed by effective reinforcement. The transition equations for the finite Markov chain y_n in $S = \{j\epsilon : 0 \leqslant j \leqslant N\}$ are

$$\Delta y_n = \begin{cases} j\epsilon & \text{with prob. } \phi_j(z_n, \epsilon)\,\dfrac{s-j}{s}\,\pi_1' \\ -k\epsilon & \text{with prob. } \phi_k(y_n, \epsilon)\,\dfrac{s-k}{s}\,\pi_2' \\ 0 & \text{with prob. } 1 - y_n\pi_1' - z_n\pi_2'. \end{cases} \tag{3.1}$$

Here $\epsilon = 1/N$, $z = 1 - y$, $0 \leqslant j,\ k \leqslant s$, and $\phi_k(y, \epsilon) = \phi_{s-k}(z, \epsilon)$ is the hypergeometric distribution with parameters y, s, and N:

$$\phi_k(y, \epsilon) = \frac{\binom{Ny}{k}\binom{Nz}{s-k}}{\binom{N}{s}}.$$

If $s = 1$ then $y_n = y_0$ for all n. Hence we assume that $s \geqslant 2$. Assuming also that $\pi_i' > 0$, y_n eventually enters one of the absorbing states 0 or 1.

We now compute the transition moments of y_n. Clearly

$$E[(\Delta y_n)^m \mid y_n = y] = \epsilon^m \frac{1}{s}\{\pi_1' M[(s-k)^m k] + \pi_2'(-1)^m M[k^m(s-k)]\}, \tag{3.2}$$

where $M[f(k)]$ is the expectation for the distribution $\phi_k(y, \epsilon)$. The only such expectation we need is

$$M[(s-k)k] = s(s-1)\,yz/1-\epsilon. \tag{3.3}$$

Equations (3.2), with $m = 1$, and (3.3) yield

$$v_1(y) = \epsilon(\pi_1' - \pi_2')(s-1)\,yz/1-\epsilon. \tag{3.4}$$

Taking $m = 2$ in (3.2), writing $\pi_i' = \alpha + \lambda_i$, and noting that

$$M[(s-k)^2 k] + M[(s-k)\,k^2] = sM[(s-k)k],$$

we obtain

$$v_2(y) = \epsilon^2 s(s-1)(\alpha + \lambda)\,yz/1-\epsilon, \tag{3.5}$$

where

$$|\lambda| \leqslant \max(|\lambda_1|, |\lambda_2|). \tag{3.6}$$

Finally, using $v_3(y) \leqslant s\epsilon v_2(y)$ in conjunction with $v_2(y) \leqslant \epsilon^2 s^{-1} M[(s-k)k]$, which comes from (3.2) on noting that $\pi_i' \leqslant 1$, we obtain

$$v_3(y) \leqslant \epsilon^3 s^2 (s-1)\, yz/1 - \epsilon. \tag{3.7}$$

These precise formulas will be needed in Sec. 6. For our present purposes we write

$$\pi_i' = \alpha + \epsilon\beta_i\,,$$

where $1 \geqslant \alpha > 0$, and $\beta_i \leqslant 0$ if $\alpha = 1$, and obtain from them

$$v_1(y) = \tau byz, \qquad v_2(y) = \tau ayz + o(\tau), \qquad v_3(y) = o(\tau), \tag{3.8}$$

where $\tau = \epsilon^2/1 - \epsilon$, $b = (s-1)(\beta_1 - \beta_2)$, and $a = s(s-1)\alpha$.

Let L_τ be the corresponding transition probability, and let K_τ be the transition probability for any family of linear models [see (2.8)] with the same b and with $a_1 = a_2 = a$ [see (2.13)]. Since (3.8) is of the form (2.12), $L_\tau^{(n)}(p, \cdot) - K_\tau^{(n)}(p, \cdot)$ converges to 0 as $\tau \to 0$ (thus ϵ and $\theta \to 0$), uniformly in p and $n\tau \leqslant M$, by Corollary 2.5.a. If $\beta_1 = \beta_2$ so that $b = 0$, (3.8) is of the form (2.14), thus convergence is uniform over p and n by Corollary 2.5.b.

4. Diffusion Approximation

We wish to show that there is a diffusion X_t such that the transition moments of $x_n^\tau = X_{n\tau}$ satisfy (2.12). If $P(t)$ is the time t transition probability or kernel for such a process,

$$P(t, x; B) = P(X_{s+t} \in B \mid X_s = x),$$

then $P(t)$ satisfies the Chapman-Kolmogorov equation

$$P(s+t, x; B) = \int P(s, x; dy)\, P(t, y; B). \tag{4.1}$$

Also (2.12) takes the form

$$\begin{aligned} \int P(\tau, x; dy)(y-x) &= \tau b(x) + o(\tau), \\ \int P(\tau, x; dy)(y-x)^2 &= \tau a(x) + o(\tau), \\ \text{and} \quad \int P(\tau, x; dy)\, |y-x|^3 &= o(\tau), \end{aligned} \tag{4.2}$$

where

$$b(x) = bx(1-x) \qquad \text{and} \qquad a(x) = (a_1(1-x) + a_2 x)\, x(1-x). \tag{4.3}$$

As usual, it is convenient to introduce the corresponding transition operators

$$T_t f(x) = \int P(t, x; dy) f(y) = E[f(X_{t+s}) \mid X_s = x], \tag{4.4}$$

for f bounded and measurable on [0, 1]. Such operators form a *semigroup*, i.e., T_0 is the identity operator, and $T_{s+t} = T_s T_t$, the latter by virtue of the Chapman-Kolgomorov equation. From (4.2) it follows, via the third order Taylor expansion of $f \in C^3\,[0, 1]$, that

$$\frac{1}{\tau}(T_\tau f(x) - f(x)) \to \frac{1}{2} a(x) \frac{\partial^2 f}{\partial x^2}(x) + b(x) \frac{\partial f}{\partial x}(x),$$

as $\tau \to 0$, uniformly in $0 \leqslant x \leqslant 1$. The problem of showing that there is a diffusion with certain properties thus leads us to an analogous problem for semigroups. The discussion below of W. Feller's approach to this problem is based on Mandl's (1968) presentation.

A semigroup $\{T_t\}_{t \geqslant 0}$ of bounded linear operators in a Banach space B is said to be *strongly continuous* if, for every $f \in B$, $T_t f$ is continuous in t with respect to the norm of B. The *infinitesimal operator* A of such a semigroup is the linear operator defined by

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

for f in the *domain* $D(A)$ of A, the set of f for which this limit exists with respect to the norm of B. The semigroup is uniquely determined by A. If $f \in D(A)$ then $T_t f \in D(A)$ and

$$\frac{d}{dt} T_t f = A T_t f = T_t A f. \tag{4.5}$$

Let I be a closed bounded interval and $C(I)$ the Banach space of real valued continuous functions on I with the sup norm $|\cdot|$. A strongly continuous semigroup T_t on $C(I)$ is *contractive* if $|T_t f| \leqslant |f|$ for all $f \in C(I)$ and T_t maps nonnegative functions into nonnegative functions. A contractive semigroup is *conservative* if $T_t 1 = 1$.

Let $a(x)$ and $b(x)$ be as in (4.3) with $a_i > 0$. For $f \in C[0, 1]$ with two continuous (but not necessarily bounded) derivatives on (0, 1) and for $0 < x < 1$ let

$$Af(x) = \frac{1}{2} a(x) \frac{\partial^2 f}{\partial x^2}(x) + b(x) \frac{\partial f}{\partial x}(x). \tag{4.6}$$

If $Af(0^+)$ and $Af(1^-)$ exist, put $Af(0) = Af(0^+)$ and $Af(1) = Af(1^-)$.

LEMMA 4.1. *The linear transformation A on*

$$D(A) = \{f : Af(0) = Af(1) = 0\}$$

is the infinitesimal operator of a conservative semigroup T_t on $C[0, 1]$.

Note that $C^2[0, 1] \subset D(A)$, though the example $f(x) = x^{3/2}$ shows that the two sets are not equal.

Proof. The differential operator (4.6) has accessible (in fact exit) boundaries (see Mandl, 1968, 1, II). Thus, for $\lambda > 0$ and $f \in C[0, 1]$ the equation $\lambda F - AF = f$ has the unique solution

$$F = F_0 + \lambda^{-1} f(0)\, u_0 + \lambda^{-1} f(1)\, u_1$$

in $D(A)$ (Mandl, p. 34). It is easy to show that $F \geqslant 0$ if $f \geqslant 0$, $F = \lambda^{-1}$ if $f = 1$, and $D(A)$ is dense in $C[0, 1]$, so the existence of a contractive semigroup T_t with infinitesimal operator A follows from the Hille-Yosida theorem (Mandl, 1, I). The equation $F = \lambda^{-1}$ for $f = 1$ implies that T_t is conservative. Q.E.D.

Any conservative semigroup corresponds, via the first equality in (4.4), to a unique transition probability P. For any $t \geqslant 0$, $P(t, x; B)$ is a probability measure in its third argument, is measurable in its second, and satisfies (4.1).

LEMMA 4.2. *The semigroup of Lemma* 4.1 *satisfies* (4.2). *If $b = 0$ so that $b(x) \equiv 0$,*

$$\int P(\tau, x; dy)(y - x) = 0 \tag{4.7}$$

for all $\tau \geqslant 0$ and $0 \leqslant x \leqslant 1$.

Proof. For any strongly continuous semigroup on $C[0, 1]$

$$\frac{d^2}{dt^2}\, T_t f(x) = T_t A^2 f(x)$$

if $f \in D(A^2) = D(A) \cap A^{-1}(D(A))$. Hence the second order Taylor expansion of $T_t f(x)$ about $t = 0$ gives

$$| T_t f(x) - f(x) - tAf(x)| \leqslant \frac{t^2}{2} \sup_{0 < s < t} | T_s A^2 f(x)|.$$

Thus

$$| T_t f(x) - f(x) - tAf(x)| \leqslant (t^2/2)\, | A^2 f | \tag{4.8}$$

if T_t is a contractive semigroup. Applying this to $f_x(y) = (y - x)$, and noting that, in the case at hand, $Af_x(x) = b(x)$ and $\sup_{0 \leqslant x \leqslant 1} | A^2 f_x | < \infty$, we obtain the first equation in (4.2). The second is similarly obtained with $f_x(y) = (y - x)^2$. Applying (4.8) to $f_x(y) = (y - x)^4$ we get

$$T_t f_x(x) = O(t^2), \tag{4.9}$$

from which the last equation in (4.2) follows on taking the $\frac{3}{4}$ power of both sides and recalling that $E^{1/3}[|Z|^3] \leqslant E^{1/4}[Z^4]$ for any random variable Z.

Let $h(x) = x$. If $b = 0$ then $Ah = 0$, so $(d/dt)\, T_t h(x) = 0$ by (4.5). Thus $T_t h(x) = h(x)$, which is (4.7). Q.E.D.

The last equation in (4.2) implies that

$$\lim_{\tau \to 0} \frac{1}{\tau} \int_{|y-x| \geqslant \delta} P(\tau, x; dy) = 0,$$

uniformly in x, for any $\delta > 0$. This, in turn, implies that there is a diffusion $\{X_t\}_{t \geqslant 0}$ with transition probability P (see Mandl, 1968, Theorem 5 of Ch. I). By Lemma 4.2 $P(\tau)$, the transition probability associated with $x_n^\tau = X_{n\tau}$, satisfies (2.12), and, if $b = 0$ and $a_1 = a_2$, (2.14). Thus we can apply Theorem 2.4 and Corollary 2.5 to $V_\tau = T_\tau$ and Corollary 2.6 to $U_\tau = T_\tau$. Note that $T_\tau^n = T_{n\tau}$.

THEOREM 4.3. *If L_τ, $\tau \in J$, is any family of transition probabilities satisfying* (2.12), *then $L_\tau^{(n)}(y, \cdot)$ converges weakly to $P(t, x; \cdot)$ as $\tau \to 0$, $n\tau \to t$, and $y \to x$.*

Proof. As we have just observed, Corollary 2.6 gives

$$\lim_{\tau \to 0} \sup_{\substack{y \in S_\tau \\ n\tau \leqslant M}} | T_{n\tau} f(y) - V_\tau^n f(y)| = 0.$$

Hence,

$$V_\tau^n f(y) - T_{n\tau} f(y) \to 0,$$

as $\tau \to 0$, $n\tau \to t$, and $y \to x$. But

$$\begin{aligned} | T_{n\tau} f(y) - T_t f(x)| &\leqslant | T_{n\tau} f(y) - T_t f(y)| + | T_t f(y) - T_t f(x)| \\ &\leqslant | T_\Delta f - f | + | T_t f(y) - T_t f(x)|, \end{aligned}$$

where $\Delta = | n\tau - t |$, and the right side converges to 0 as $n\tau \to t$ and $y \to x$. Thus

$$V_\tau^n f(y) \to T_t f(x)$$

as $\tau \to 0$, $n\tau \to t$, and $y \to x$. Q.E.D.

Let L_τ satisfy (2.12) and let y_n^τ be a corresponding family of Markov processes. The proof of Theorem 4.3 can be extended to obtain the following generalization: The joint distribution of $y_{n_1}, \ldots, y_{n_k}$, where $0 \leqslant n_1 < n_2 \ldots$, converges weakly to that of $X_{t_1}, \ldots, X_{t_k}$ as $\tau \to 0$, $n_j\tau \to t_j$, and $y_0 \to x = X_0$.

If L_τ satisfies (2.12) with all $o(\tau)$'s replaced by $O(\tau^{1+\nu})$ for some $1 \geq \nu > 0$, a more striking generalization is obtained. Reexamination of our work above shows that the stronger hypothesis, with $\nu = \frac{1}{2}$, is satisfied by the linear and stimulus sampling models, as well as by $X_{n\tau}$. Let X_t^τ, $t \leqslant M$, be the continuous time stochastic process defined by $X_{n\tau}^\tau = y_n^\tau$, and by linear interpolation for other values of t. Let $L_\tau^M(y, \cdot)$ be the distribution of the process X^τ:

$$L_\tau^M(y, B) = P(X^\tau \in B \mid y_0^\tau = y)$$

for Borel subsets B of $C[0, M]$. Let $P^M(x, \cdot)$ be the distribution of X_t, $t \leqslant M$:

$$P^M(x, B) = P(X \in B \mid X_0 = x).$$

THEOREM 4.4. *$L_\tau^M(y, \cdot)$ converges weakly to $P^M(x, \cdot)$ as $\tau \to 0$ and $y \to x$.*

Proof. Since the finite dimensional distributions of $L_\tau^M(y, \cdot)$ converge weakly to those of $P^M(x, \cdot)$ it suffices to show that there is a constant C such that

$$E[| X_{t_1}^\tau - X_{t_2}^\tau |^4 \mid X_0^\tau = y] \leqslant C \mid t_1 - t_2 \mid^{1+\nu}$$

for all $\tau \in J$, $y \in S_\tau$, and $0 \leqslant t_1, t_2 \leqslant M$ (Billingsley, 1968, Theorems 12.3 and 8.1). This follows from the special case $t_1 = k\tau$, $t_2 = (k + n)\tau$, which we now check.

When the error terms in (2.12) are $O(\tau^{1+\nu})$, the term $o(1)$ in (2.15) can be replaced by $O(\tau^\nu)$, and, as in Corollary 2.6, we can replace U in (2.15) by T_τ. Since $(e^x - 1) \leqslant xe^x$ for $x > 0$, we finally obtain

$$| T_{n\tau}f - V^n f | \leqslant K\tau^\nu(n\tau) \sum_{j=1}^{3} | f^{(j)} |,$$

for some K and all $n\tau \leqslant M$ and $f \in C^3[0, 1]$. Applying this to $f_{y'}(x) = (x - y')^4$ and noting that $\tau^\nu(n\tau) \leqslant (n\tau)^{1+\nu}$, we see that

$$| T_{n\tau}f_{y'}(y') - V^n f_{y'}(y')| \leqslant K'(n\tau)^{1+\nu},$$

for some K' and all $n\tau \leqslant M$ and $y' \in S_\tau$. It now follows from (4.9) that

$$E[(y_{k+n} - y')^4 \mid y_k = y'] = V^n f_{y'}(y') \leqslant K''(n\tau)^{1+\nu},$$

from which

$$E[(y_{k+n} - y_k)^4 \mid y_0 = y] \leqslant K''(n\tau)^{1+\nu}$$

is obtained on integrating with respect to $L^{(k)}(y, \cdot)$. Q.E.D.

5. Some Properties of T_t

In this section it will be shown that the semigroup T_t of Lemma 4.1 inherits the properties of the discrete parameter semigroup U_τ^n of the linear model that were crucial for Theorem 2.4. These properties permit an alternative approach to the results of the previous sections, and suggest some extensions.

$C^k[0, 1]$ is a Banach space with respect to the norm

$$| f |_k = \sum_{j=0}^{k} | f^{(j)} |.$$

Let U_τ be the transition operator for any family (2.8) of linear models with the same a_i and b [see (2.13)] as in the infinitesimal operator A of T_t [see (4.3) and (4.6)]. And let γ_k be as in Lemma 2.3.

Theorem 5.1. *For any $k \geqslant 0$, T_t is a strongly continuous semigroup in $C^k[0, 1]$, and U_τ^n converges strongly in $C^k[0, 1]$ to T_t as $\tau \to 0$ and $n\tau \to t$. For any $f \in C^k[0, 1]$, $k \geqslant 1$,*

$$\sum_{j=1}^{k} |(T_t f)^{(j)}| \leqslant \exp(t\gamma_k) \sum_{j=1}^{k} | f^{(j)} |, \tag{5.1}$$

and, if $b = 0$ and $a_1 = a_2$,

$$|(T_t f)^{(k)}| \leqslant \exp(-k(k-1)\, c_1 e_1 t/2)\, | f^{(k)} |. \tag{5.2}$$

Proof. Let $D = \cap_{k=0}^{\infty} C^k[0, 1]$. By Theorem 2.3, $| U^n f - T_{n\tau} f |_0 \to 0$ (as $\tau \to 0$ and $n\tau \to t$) for $f \in C^3[0, 1]$. Also $| T_{n\tau} f - T_t f |_0 \to 0$ for $f \in C[0, 1]$. Hence, certainly, $| U^n f - T_t f |_0 \to 0$ for all $f \in D$. Suppose, inductively, that $T_t f \in C^k[0, 1]$ and $| U^n f - T_t f |_k \to 0$ for some $k \geqslant 0$ and all $f \in D$. By (2.9)

$$\sum_{j=1}^{m} |(U^n f)^{(j)}| \leqslant \exp(n\tau\gamma_m) \sum_{j=1}^{m} | f^{(j)} | \tag{5.3}$$

for $m \geqslant 1$ and $f \in C^m[0, 1]$. Hence $|(U^n f)^{(k+2)}|$ is bounded as $n\tau \to t$, so that $(U^n f)^{(k+1)}$ is bounded and equicontinuous. Since $(U^n f)^{(k)}$ converges to $(T_t f)^{(k)}$, $T_t f \in C^{k+1}[0, 1]$, and any uniformly convergent subsequence of $(U^n f)^{(k+1)}$ has limit $(T_t f)^{(k+1)}$. Thus the full sequence $(U^n f)^{(k+1)}$ converges uniformly to $(T_t f)^{(k+1)}$, so that $| U^n f - T_t f |_{k+1} \to 0$. By induction T_t maps D into D and $| U^n f - T_t f |_k \to 0$ for all $k \geqslant 0$ and $f \in D$.

But D is dense in $C^k[0, 1]$ and, for $f \in C^k[0, 1]$, $| U^n f |_k$ is bounded as $n\tau \to t$, hence by the uniform boundedness principle, T_t is a bounded linear operator on $C^k[0, 1]$ and $| U^n f - T_t f |_k \to 0$ for all $f \in C^k[0,1]$.

Inequality (5.1) follows from (5.3) on taking the limit as $n \to \infty$. Similarly, (5.2) follows from

$$|(U^n f)^{(k)}| \leqslant B_k^n |f^{(k)}|,$$

which is an immediate consequence of (2.10), and (2.11).

Inequality (5.1) implies that $| T_t f |_k \leqslant \exp(t\gamma_k) | f |_k$ for $f \in C^k[0, 1]$, so that $| T_t f |_k$ is bounded when t is. An inductive argument shows that $| T_t f - f |_k \to 0$ as $t \to 0$ for $f \in D$ and $k \geqslant 0$. Thus strong continuity of T_t on $C^k[0, 1]$ follows from the uniform boundedness principle. Q.E.D.

The properties of T_t obtained in Theorem 5.1 appear to be new. Had (5.1) and (5.2) been available at the onset, T_τ could have been used in place of U_τ in Theorem 2.4 and Corollary 2.5, thus eliminating Lemma 2.3. Were similar bounds available for other semigroups, the scope of these results could be correspondingly extended. And better bounds on $(T_t f)^{(j)}$, $j = 1, 2, 3$, would permit the conclusion

$$\sup_{n \geqslant 0} | T_{n\tau} f - V_\tau^n f | \to 0 \tag{5.4}$$

as $\tau \to 0$ for $f \in C[0, 1]$ (Corollary 2.6) to be generalized beyond the case $b(x) = 0$, $a(x) = ax(1 - x)$. Let us consider the latter possibility in more detail.

If $a(x), b(x) \in C[0, 1]$, $a(x) > 0$ on $0 < x < 1$, $a(i) = b(i) = 0$, $i = 0, 1$, and A in (4.6) has accessible boundaries, then Lemma 4.1 is valid for this operator. If $P(t, x; \{i\}) = m_i(t, x)$ for $i = 1, 2$ and

$$S_t f(x) = \int_{(0,1)} P(t, x; dy) f(y),$$

then

$$T_t f(x) = m_0(t, x) f(0) + m_1(t, x) f(1) + S_t f(x),$$

for $f \in C[0, 1]$. It can be shown that $m_1(x, t) \to \Gamma(x)$ and $m_0(x, t) \to 1 - \Gamma(x)$ as $t \to \infty$, where Γ is the solution of $A\Gamma = 0$ for which $\Gamma(0) = 0$ and $\Gamma(1) = 1$. Also $P(t, x; (0, 1)) \to 0$, so that $S_t f(x) \to 0$ as $t \to \infty$. Thus $T_t f(x) \to f(0)(1 - \Gamma(x)) + f(1)\Gamma(x)$ as $t \to \infty$. The result for $k \geqslant 2$ in (5.2) is now seen to depend on $b(x) = 0$ (hence $\Gamma(x) = x$ and $\Gamma^{(k)}(x) = 0$). However, if we consider only f in

$$C_0^k = \{f \in C^k[0, 1] : f(0) = f(1) = 0\},$$

exponential convergence of $|(T_t f)^{(k)}|$ to 0 for $k \geqslant 0$ is still a possibility. For such f, $T_t f = S_t f$. We conjecture that

$$\sum_{j=1}^{3} |(S_t f)^{(j)}| \leqslant K e^{-\mu t} \sum_{j=1}^{3} |f^{(j)}| \tag{5.5}$$

for some $\mu > 0$, $K < \infty$, and all $f \in C_0^3$ under (4.3) and, in fact, for much more general $a(x)$ and $b(x)$. Perhaps this question can be approached by way of the theory of eigenfunction expansions (McKean, 1956).

Writing $f \in C[0, 1]$ in the form

$$f = Hf + (f - Hf),$$

where

$$Hf = f(0)1 + (f(1) - f(0))\Gamma,$$

we see that (5.4) holds in general if it holds for 1, for Γ, and for functions that vanish at 0 and 1. If (5.5) is satisfied, if $a(x), b(x) \in C_0^2[0, 1]$ [this implies (4.2)], and if V_τ also satisfies (4.2), then the method of Sec. 2 yields (5.4) for functions that vanish at 0 and 1. Equation 5.4 is trivial if $f = 1$. Finally $T_t\Gamma = \Gamma$ and, analogously, $V_\tau\gamma = \gamma$, where γ is the probability that $y_n \to 1$ as $n \to \infty$. Thus,

$$| T_{n\tau}\Gamma - V_\tau^n\Gamma | \leqslant 2 | \Gamma - \gamma |,$$

so that (5.4) holds for $f = \Gamma$ if, as in the next section, $| \Gamma - \gamma | \to 0$ as $\tau \to 0$.

6. Absorption Probabilities

We have seen that the distribution of a Markov process y_n^τ in $S_\tau \subset [0, 1]$ with $v_1(y) \sim \tau b(y)$, $v_2(y) \sim \tau a(y)$, and $v_3(y) = o(\tau)$ can sometimes be approximated by the distribution of $X_{n\tau}$ for a diffusion X_t. Thus it is natural to inquire whether the probability $\gamma(y)$ of absorption at 1 starting at y can be approximated by the camparable quantity $\Gamma(y)$ for X_t. $\Gamma(y)$ is the solution of (1.5) for which $\Gamma(0) = 0$ and $\Gamma(1) = 1$. In this section we will show that such an approximation is valid under hypotheses that are more restrictive than those of Sec. 3 with respect to the mode of convergence of $\tau^{-1}v_i(y)$, but more general with respect to the form of the limit.

Our assumptions are as follows: J is a set of positive real numbers with infimum 0. For every $\tau \in J$, V_τ is a transition operator on $S_\tau \subset [0, 1]$ with transition moments v_i. Zero and one belong to S_τ and are absorbing, i.e., $L_\tau(i, \{i\}) = 1$, $i = 0, 1$, where L_τ is the kernel of V_τ. The functions $a(y)$ and $b(y)$ are in $C_0^1[0, 1]$, $a(y) > 0$ for $0 < y < 1$, and

$$r(y) = 2b(y)/a(y),$$

$0 < y < 1$, has finite limits at 0 and 1. Putting $r(0) = r(0^+)$ and $r(1) = r(1^-)$, the extended function r belongs to $C^1[0, 1]$. All of the quantities

$$\delta_1 = |(\tau^{-1}v_1 - b)/a|,$$
$$\delta_2 = |(\tau^{-1}v_2 - a)/a|,$$

and

$$\delta_3 = | \tau^{-1}v_3/a |$$

(the suprema are over $S_\tau \cap (0, 1)$) converge to 0 as $\tau \to 0$. Finally, $L_\tau^{(n)}(y, \cdot)$ converges weakly as $n \to \infty$ to a distribution concentrated at 0 and 1, i.e., there is a function $\gamma = \gamma_\tau$ on S_τ such that

$$V^n f(y) \to f(0)(1 - \gamma(y)) + f(1)\,\gamma(y)$$

as $n \to \infty$ for all $f \in C[0, 1]$. Under these conditions we have the following result.

THEOREM 6.1. *As* $\tau \to 0$, $|\gamma_\tau - \Gamma| \to 0$, *where* Γ *is the solution of* (1.5) *with* $\Gamma(0) = 0$ *and* $\Gamma(1) = 1$.

For the linear and stimulus sampling models $a(y)$ and $b(y)$ are of the form (4.3), and, going back to (1.2) and (3.4)–(3.7), it is easy to show that the δ_i converge to 0. For these models one has almost sure convergence to 0 or 1, and this implies the weak convergence assumed above. Thus Theorem 6.1 is applicable to both models.

Rewriting (1.5)

$$\Gamma''(y) + r(y)\,\Gamma'(y) = 0,$$

we see that $\Gamma'(y) = K \exp(-Z(y))$ where $Z(y) = \int_0^y r(z)\,dz$ and $K \neq 0$. Hence

$$\Gamma(y) = \frac{\int_0^y \exp(-Z(x))\,dx}{\int_0^1 \exp(-Z(x))\,dx}.$$

In the special case (4.3) where $a_1 \neq a_2$ [if $a_1 = a_2$ see (1.6)]

$$\Gamma(y) = \frac{h(y) - h(0)}{h(1) - h(0)},$$

where

$$h(y) = \begin{cases} (a_1(1-y) + a_2 y)^\rho & \rho \neq 0 \\ \ln(a_1(1-y) + a_2 y) & \rho = 0, \end{cases}$$

and

$$\rho = 1 + 2b/(a_1 - a_2).$$

Proof of Theorem 6.1. Let $f = f_\epsilon$ be the solution of

$$f''(y) + r(y) f'(y) = \epsilon \tag{6.1}$$

such that $f'(0) = 1$ and $f(0) = 0$, i.e.,

$$f(y) = \int_0^y f'(x)\,dx,$$

$$f'(y) = e^{-Z(y)}[1 + \epsilon \int_0^y e^{Z(x)}\,dx].$$

Suppose that $|\epsilon|$ is sufficiently small that $f'(y) > 0$ for all $0 \leqslant y \leqslant 1$. Clearly $|f'| < \infty$. Thus $|f''| < \infty$ by (6.1) and $|f^{(3)}| < \infty$ by the derivative of (6.1).

From (2.7) we get

$$\begin{aligned} Vf(y) - f(y) = \tau\{[b(y)f'(y) + a(y)f''(y)/2] + [(\tau^{-1}v_1(y) - b(y))f'(y) \\ + (\tau^{-1}v_2(y) - a(y))f''(y)/2 + \lambda\tau^{-1}v_3(y)\,|f^{(3)}|/6]\}, \end{aligned}$$

where $|\lambda| \leqslant 1$. For $y \in S_\tau \cap (0, 1)$ this yields, on dividing by $\tau a(y)/2$ and using (6.1),

$$2(Vf(y) - f(y))/\tau a(y) = \epsilon + \omega(2\delta_1\,|f'| + \delta_2\,|f''| + \delta_3\,|f^{(3)}|/3),$$

where $|\omega| \leqslant 1$. If τ is sufficiently small that the quantity in parentheses on the right is less than $|\epsilon|$ (we write $\tau \leqslant \delta$) then $Vf(y) - f(y)$ has the same sign as ϵ. Since $Vf(i) = f(i)$, $i = 0, 1$, we conclude that

$$Vf(y) \geqslant f(y) \quad \text{if } \epsilon > 0 \qquad \text{and} \qquad Vf(y) \leqslant f(y) \quad \text{if } \epsilon < 0$$

for all $y \in S_\tau$ when $\tau \leqslant \delta$. Since V is linear

$$Vg(y) \geqslant g(y) \quad \text{if } \epsilon > 0 \qquad \text{and} \qquad Vg(y) \leqslant g(y) \quad \text{if } \epsilon < 0 \tag{6.2}$$

if $\tau \leqslant \delta$, where

$$g(y) = g_\epsilon(y) = f(y)/f(1).$$

Applying V^n to both sides of (6.2) we see that $V^n g(y)$ is a monotonic sequence, hence

$$V^n g(y) \geqslant g(y) \quad \text{if } \epsilon > 0 \qquad \text{and} \qquad V^n g(y) \leqslant g(y) \quad \text{if } \epsilon < 0.$$

Letting $n \to \infty$ and nothing that $g \in C[0, 1]$ with $g(1) = 1$ and $g(0) = 0$ we obtain

$$g_\epsilon(y) \leqslant \gamma_\tau(y) \leqslant g_{-\epsilon}(y) \tag{6.3}$$

for all $y \in S_\tau$, if $\epsilon > 0$ and $\tau \leqslant \delta$.

It is easily shown that $g(y)$ is noncreasing, hence

$$g_\epsilon(y) \leqslant g_0(y) = \Gamma(y) \leqslant g_{-\epsilon}(y). \tag{6.4}$$

Combining (6.3) and (6.4) we obtain

$$|\gamma_\tau - \Gamma| \leqslant |g_{-\epsilon} - g_\epsilon|$$

for τ sufficiently small. Since $\epsilon > 0$ is arbitrary and $|g_{-\epsilon} - g_\epsilon| \to 0$ as $\epsilon \to 0$, $|\gamma_\tau - \Gamma| \to 0$ as $\tau \to 0$. Q.E.D.

REFERENCES

BILLINGSLEY, P. *Convergence of probability measures.* New York: Wiley, 1968.

BREIMAN, L. *Probability.* Reading, Mass.: Addison-Wesley, 1968.

IGLEHART, D. L. Diffusion approximations in applied probability. In G. B. Dantzig and A. F. Veinott, Jr. (Eds), *Mathematics of the decision sciences, part 2.* Providence, R. I.: Amer. Math. Soc., 1968. Pp. 235–254.

KHINTCHINE, A. *Asymptotische Gesetze der Wahrscheinlichkeitsrechnung.* New York: Chelsea, 1948. (A translation by L. B. Kapp, supported by my NSF grant, is available from the National Translations Center, The John Crerar Library, 35 West 33rd Street, Chicago, Illinois 60616.)

KIMURA, M. *Diffusion models in population genetics.* London: Methuen, 1964.

MANDL, P. *Analytical treatment of one-dimensional Markov processes.* New York/Berlin: Springer-Verlag, 1968.

MCKEAN, H. P., JR. Elementary solutions of certain parabolic partial differential equations. *Transactions of the American Mathematical Society*, 1956, **82**, 519–548.

MOSTELLER, F., AND TATSUOKA, M. Ultimate choice between two attractive goals: predictions from a model. *Psychometrika*, 1960, **25**, 1–18.

NORMAN, M. F. Mathematical learning theory. In G. B. Dantzig and A. F. Veinott, Jr. (Eds.), *Mathematics of the decision sciences, part 2.* Providence, R. I.: Amer. Math. Soc., 1968. Pp. 283–313. (a)

NORMAN, M. F. On the linear model with two absorbing barriers. *Journal of Mathematical Psychology*, 1968, **5**, 225–241. (b)

NORMAN, M. F. Slow learning. *British Journal of Mathematical and Statistical Psychology*, 1968, **21**, 141–159. (c)

WATTERSON, G. A. Some theoretical aspects of diffusion theory in population genetics. *Annals of Mathematical Statistics*, 1962, **33**, 939–957.

RECEIVED: July 22, 1969