A CENTRAL LIMIT THEOREM FOR FAMILIES OF STOCHASTIC PROCESSES INDEXED BY A SMALL AVERAGE STEP SIZE PARAMETER, AND SOME APPLICATIONS TO LEARNING MODELS

M. Frank Norman and Norma V. Graham

Reprinted from
Psychometrika
December 1968, Vol. 33, No. 4
The Psychometric Society

PSYCHOMETRIKA—VOL. 33, NO. 4
DECEMBER, 1968

A CENTRAL LIMIT THEOREM FOR FAMILIES OF STOCHASTIC PROCESSES INDEXED BY A SMALL AVERAGE STEP SIZE PARAMETER, AND SOME APPLICATIONS TO LEARNING MODELS

M. FRANK NORMAN AND NORMA V. GRAHAM

UNIVERSITY OF PENNSYLVANIA

Let $\theta > 0$ be a measure of the average step size of a stochastic process $\{p_n^{(\theta)}\}_{n=1}^{\infty}$. Conditions are given under which $p_n^{(\theta)}$ is approximately normally distributed when n is large and θ is small. This result is applied to a number of learning models where θ is a learning rate parameter and $p_n^{(\theta)}$ is the probability that the subject makes a certain response on the nth experimental trial. Both linear and stimulus sampling models are considered.

1. *Introduction*

Consider a learning experiment where response alternatives A_1 and A_2 are available to the subject on each of a sequence of trials, and, on any trial, response A_i is followed by outcome O_{ij}, presumed to be reinforcing to A_j, with probability π_{ij}. It is assumed that

$$\pi_{12} > 0 \quad \text{and} \quad \pi_{21} > 0,$$

which implies, in the learning models considered below, that a subject continues to choose both alternatives throughout the experiment. Let p_n be the probability that a subject will choose A_1 on trial n, conditional upon past events (directly observable or hypothetical) that were crucial for learning. Mathematical learning theorists are often concerned with the asymptotic distribution of p_n as $n \to \infty$. Knowledge about the predicted distribution varies from model to model. There are interesting models for which not even the asymptotic mean of p_n is known, and, even in the few cases where formulas are available for some of the asymptotic moments of p_n, these usually give little insight into the shape of the asymptotic distribution.

Learning models typically involve one or more parameters whose magnitudes are directly related to the average step size of the corresponding process $\{p_n\}$. These may be loosely described as learning rate parameters. Quite often experimental data seem to call for small values of these parameters, so approximations to the asymptotic distribution of p_n, valid when learning rates are small, are of interest. An approximation of this sort, applicable to a wide variety of models, is obtained in this paper.

Let a family of models be indexed by a single parameter θ, denote the stochastic process corresponding to θ by $\{p_n^{(\theta)}\}$, and suppose that the cumulative distribution function of $p_n^{(\theta)}$ converges as $n \to \infty$ to a distribution F_θ. The main result of this paper, the central limit theorem of Section 3 (henceforth abbreviated CLT), implies, roughly speaking, that if (a) certain means of $p_{n+1}^{(\theta)} - p_n^{(\theta)}$ are of the order of magnitude of θ, and (b) there is a point $\rho \in (0, 1)$ toward which the process $\{p_n^{(\theta)}\}$ drifts when θ is small, then the normalized distribution

$$F_\theta(\theta^{1/2}x + \rho) = \lim_{n \to \infty} P_\theta((p_n - \rho)\theta^{-1/2} \leqq x)$$

converges as $\theta \to 0$ to a normal distribution with mean 0. (Throughout the paper θ will appear as a subscript on P and E rather than as a superscript on p_n in probabilities and expectations involving the latter quantity.) A more precise statement of the theorem requires the functions

$$V(p, \theta) = E_\theta[p_{n+1} - p_n \mid p_n = p]$$

and

$$M(p, \theta) = E_\theta[(p_{n+1} - p_n)^2 \mid p_n = p],$$

which are assumed to be independent of n, and

$$a_n(\theta) = E_\theta[|p_{n+1} - p_n|^3].$$

In accordance with (a), $V(p, \theta)$, $M^{1/2}(p, \theta)$, and $a_n^{1/3}(\theta)$ are assumed to be $O(\theta)$. The condition (b) means that $V(p, \theta)$ has the same sign as $(\rho - p)$ when θ is sufficiently small, or, in view of (a), that $(\partial/\partial\theta)V(p, 0)$ is positive for $p < \rho$ and negative for $p > \rho$. Thus ρ can be simply described as the unique root of $(\partial/\partial\theta)V(\rho, 0) = 0$ in (0, 1). Clearly $(\partial^2/\partial p\,\partial\theta)V(\rho, 0) \leqq 0$, and, by (a), $(\partial^2/\partial\theta^2)M(\rho, 0) \geqq 0$. The expression

$$\sigma^2 = \frac{1}{4}\,\frac{(\partial^2/\partial\theta^2)M(\rho, 0)}{-(\partial^2/\partial p\,\partial\theta)V(\rho, 0)}$$

given by CLT for the variance of the limiting normal distribution suggests that $(\partial^2/\partial p\,\partial\theta)V(\rho, 0)$ must be assumed strictly negative, and, if the asymptotic distribution is to be non-degenerate, that $(\partial^2/\partial\theta^2)M(\rho, 0)$ must be assumed strictly positive. A completely precise statement of the theorem will be given in Section 3 along with its proof. The proof is similar to the proofs of the normal convergence theorems of Norman [1966], all of which are special cases of CLT.

Some applications of the theorem to specific learning models will be given in Section 2 below. These results partially overlap those in Norman [1966].

2. *Some Corollaries of* CLT

2.1 *The Four-Operator Linear Model*

The process $\{p_n\}$ for the general four-operator or experimenter-subject controlled events linear model [Bush & Mosteller, 1955] satisfies the equation

$$p_{n+1} = (1 - \theta_{ij})p_n + \theta_{ij}\delta_{1j}$$

when A_i and O_{ij} occur on trial n. It is assumed that $1 > \theta_{ij} \geqq 0$ for all i and j, and, in addition, that $\theta_{ij} > 0$ for $i \neq j$, so that the process $\{p_n\}$ has no absorbing barriers. The distribution of p_n converges as $n \to \infty$ to a distribution that does not depend on p_1 [Norman, 1968, Theorem 3.1], but, in this generality, even the mean of the asymptotic distribution is not known.

A one parameter subfamily of this four parameter family of models is obtained by constraining the vector $\Theta = (\theta_{11}, \theta_{12}, \theta_{21}, \theta_{22})$ to a one dimensional subspace of four dimensional space. Thus for any vector $C = (c_{11}, c_{12}, c_{21}, c_{22})$ of constants with $c_{ii} \geqq 0$ and $c_{ij} > 0$ for $i \neq j$, the family of parameter-vectors Θ of the form $\Theta = \theta C$, $0 < \theta < 1/\max_{ij} c_{ij}$ is considered. The corresponding functions $V(p, \theta)$ and $M(p, \theta)$ can be written in the form

$$V(p, \theta) = \theta W(p) \quad \text{and} \quad M(p, \theta) = \theta^2 N(p), \tag{1}$$

where $W(p) = (\partial/\partial\theta)V(p, 0)$ and $N(p) = (\partial^2/\partial\theta^2)M(p, 0)/2$ are quadratic and cubic polynomials, respectively. Since $|p_{n+1}^{(\theta)} - p_n^{(\theta)}| \leqq \theta \max_{ij} c_{ij}$, $a_n(\theta) \leqq \theta^3 \max_{ij}^3 c_{ij}$. Now $W(0) = c_{21}\pi_{21} > 0$ and $W(1) = -c_{12}\pi_{12} < 0$, and, since it is quadratic, it follows that $W(p)$ has a unique root ρ in (0, 1) and $W'(\rho) < 0$. Obviously $N(\rho) > 0$. These facts and CLT have the following implication.

Corollary 1. *For the four-operator linear model with* $\Theta = \theta C$ *and* $c_{ij} > 0$ *for* $i \neq j$,

$$\lim_{\theta \to 0} \lim_{n \to \infty} P_\theta((p_n - \rho)\theta^{-1/2} \leqq x) = \Phi(x/\sigma)$$

where Φ *is the normal distribution with mean* 0 *and variance* 1, ρ *is the root in* (0, 1) *of* $W(\rho) = 0$, $\sigma^2 = N(\rho)/-2W'(\rho)$, *and* N *and* W *are defined in* (1).

It can be shown that ρ is the expected operator approximation to $\lim_{n\to\infty} E[p_n]$ [Bush and Mosteller, 1955]. Thus Corollary 1 provides a justification for expected operator approximation when learning rates are small.

2.2 *Families of Models That Predict Approximate Probability Matching for Small Learning Rates*

Throughout this subsection and the next it is implicitly assumed that $O_{1j} = O_{2j} = O_j$, $j = 1, 2$, and that the outcomes O_1 and O_2 are complimentary in the sense that O_1 has the same reinforcing effect on A_1 as O_2 has on A_2.

Norman and Yellott [1966] defined abstract families of learning models indexed by a learning rate parameter ϵ by means of their axioms F1 – F4. Their paper should be consulted for statements of these axioms as well as the definitions of the italicized terms, the function w, and the number μ that appear below. If such a family predicts *approximate probability matching for small learning rates* or, equivalently [Norman and Yellott, 1966, Theorem 4] *approximate marginal constancy for small learning rates*, and if, in addition, axiom F3 is strengthened to require that the function w be thrice continuously differentiable on $[0, 1] \times [0, \mu]$, then it can be shown that the hypotheses of CLT are satisfied, and the following conclusion is obtained.

Corollary 2. *If* $0 < \pi_{12}, \pi_{21} < 1$, *then*

$$\lim_{\epsilon \to 0} \lim_{n \to \infty} P_\epsilon((p_n - \ell)\epsilon^{-1/2} \leqq x) = \Phi(x/\sigma),$$

where $\ell = \pi_{21}/(\pi_{12} + \pi_{21})$ *and*

$$\sigma^2 = \frac{1}{2} \frac{\ell(1 - \ell)}{\pi_{12} + \pi_{21}} \frac{\partial w}{\partial \epsilon} (\ell, 0).$$

2.3 *The Component Model with Fixed Sample Size*

The proportion $p_n^{(1/N)}$ of stimulus elements conditioned to A_1 on trial n in the N element component model with fixed sample size s [Estes, 1959] satisfies the stochastic difference equation

$$p_{n+1} - p_n = \begin{cases} \dfrac{j}{N} \\ 0 \\ -\dfrac{k}{N} \end{cases} \text{with prob.}$$

$$\binom{N}{s}^{-1} \begin{cases} \dbinom{(1 - p_n)N}{j}\dbinom{p_nN}{s - j}\left(\dfrac{j}{s}\pi_{21} + \dfrac{s - j}{s}\pi_{11}\right) \\ \dbinom{p_nN}{s}\pi_{11} + \dbinom{(1 - p_n)N}{s}\pi_{22} \\ \dbinom{p_nN}{k}\dbinom{(1 - p_n)N}{s - k}\left(\dfrac{k}{s}\pi_{12} + \dfrac{s - k}{s}\pi_{22}\right) \end{cases}$$

where $N \geqq s \geqq j, k \geqq 1$ and $\binom{j}{i}$ is to be interpreted as 0 if $i > j$.

A simple computation shows that $V(p, \psi) = \psi(\pi_{21} - (\pi_{21} + \pi_{12})p)s$, and a more complicated one yields

$$M(p, \psi) = \psi^2 s\Big\{(\pi_{21} - \pi_{11})s - 3(\pi_{21} - \pi_{11})(s - 1)p(1 - p)/(1 - \psi)$$
$$+ \pi_{11}(1 - p)\left[1 + (s - 1)\frac{(1 - p - \psi)}{1 - \psi}\right] + \pi_{22}p\left[1 + (s - 1)\frac{(p - \psi)}{1 - \psi}\right]\Big\},$$

where s is fixed and $\psi = 1/N$ is regarded as the learning rate parameter. The process $\{p_n^{(1/N)}\}$ is a finite Markov chain with only one ergodic class, and this class is aperiodic, so the distribution of $p_n^{(1/N)}$ converges as $n \to \infty$ to a distribution $F_{1/N}$ (which is independent of the distribution of $p_1^{(1/N)}$). Obviously $V(p, \psi)$, $M^{1/2}(p, \psi)$ and $a_n^{1/3}(\psi)$ are $O(\psi)$; $(\partial/\partial\psi)V(p, 0) = s(\pi_{21} - (\pi_{21} + \pi_{12})p)$ has the same sign as $\pi_{21}/(\pi_{21} + \pi_{12}) - p$, so $\rho = \pi_{21}/(\pi_{21} + \pi_{12})$; and $(\partial^2/\partial p\, \partial\psi)V(\rho, 0) = -(\pi_{21} + \pi_{12})s < 0$. Finally, it can be shown that

$$(\partial^2/\partial\psi^2)M(\rho, 0) = 2s\rho(1 - \rho)(2(\pi_{21} + \pi_{12}) + s - 1) > 0,$$

so CLT is applicable and yields this result.

Corollary 3. *If* $0 < \pi_{12}, \pi_{21} < 1$, *then*

$$\lim_{N\to\infty} \lim_{n\to\infty} P_{1/N}((p_n - \rho)N^{1/2} \leqq x) = \Phi(x/\sigma),$$

where $\rho = \pi_{21}/(\pi_{21} + \pi_{12})$ *and*

$$\sigma^2 = \rho(1 - \rho)(1 + (s - 1)/2(\pi_{21} + \pi_{12})).$$

3. *A General Central Limit Theorem*

Let S be a bounded set of positive real numbers having 0 as a limit point, $\delta = \sup S$, and $J = [0, \delta]$. For every $\theta \in S$, let $\{p_n^{(\theta)}\}_{n=1}^{\infty}$ be a stochastic process (not necessarily Markov) with state space contained in a closed bounded interval $[a, b] = I$. Suppose that there exist twice continuously differentiable real valued functions V and M on $I \times J$ such that, for every $\theta \in S$ and $n \geqq 1$, $E_\theta[p_{n+1} - p_n \mid p_n] = V(p_n, \theta)$ and $E_\theta[(p_{n+1} - p_n)^2 \mid p_n] = M(p_n, \theta)$ with probability 1. Suppose also that $(\partial^3/\partial\theta^3)M(\cdot, \cdot)$ exists and is continuous on $I \times J$, and that $(\partial^3/\partial p^2\, \partial\theta)V(\cdot, 0)$ and $(\partial^3/\partial p\, \partial\theta^2)M(\cdot, 0)$ exist and are continuous on I. Finally, assume that, for every $\theta \in S$, the distribution function of $p_n^{(\theta)}$ converges to a distribution function F_θ as $n \to \infty$.

Central Limit Theorem. If

(*i*) $V(p, \theta) = O(\theta)$ *and* $M(p, \theta) = O(\theta^2)$ *for each* $p \in I$,

(*ii*) $a_n(\theta) = E_\theta[|p_{n+1} - p_n|^3] = O(\theta^3)$ *uniformly in n, and*

(*iii*) $(\partial/\partial\theta)V(a, 0) > 0$, $(\partial/\partial\theta)V(b, 0) < 0$, $(\partial/\partial\theta)V(\cdot, 0)$ *has a unique root* ρ *in* (a, b), *and*

$$v = (\partial^2/\partial p\, \partial\theta)V(\rho, 0) < 0,$$

then $m = (\partial^2/\partial\theta^2)M(\rho, 0) \geqq 0$. *Denote the normalized distribution* $F_\theta(\theta^{1/2}x + \rho)$ *by* $G_\theta(x)$. *If* $m > 0$ *then* $\lim_{\theta\to 0} G_\theta(x) = \Phi(x/\sigma)$ *for all x, where* Φ *is the standard normal distribution function and* $\sigma^2 = m/-4v$. *If* $m = 0$ *then* $\lim_{\theta\to 0} G_\theta(x) = \delta_0(x)$ *for* $x \neq 0$, *where* $\delta_0(x)$ *is the distribution function with all of its mass at* 0.

Since, for every $\theta \in S$ and $n \geqq 1$, $|V(p_n^{(\theta)}, \theta)| \leqq M^{1/2}(p_n^{(\theta)}, \theta)$ with probability 1, one might hope that the first equality in (*i*) could be dispensed with.

No way of doing this has yet been found, and, at any rate, the development in Section 2 suggests that the verification of (*i*) and (*ii*) will be a triviality in most applications.

The following lemma is a basic component of the proof of CLT.

Lemma. Under hypotheses (i) *and* (iii) *of* CLT, $\int_{-\infty}^{\infty} (p - \rho)^2 \, dF_\theta(p) = O(\theta)$ *and* $m \geqq 0$. *If* $m = 0$ *then*

$$\int_{-\infty}^{\infty} (p - \rho)^2 \, dF_\theta(p) = O(\theta^{3/2}).$$

Proof of the Lemma. For $\theta \varepsilon S$,

$$\begin{aligned} E_\theta[(p_{n+1} - \rho)^2] &= E_\theta[((p_n - \rho) + (p_{n+1} - p_n))^2] \\ &= E_\theta[(p_n - \rho)^2] + 2E_\theta[(p_n - \rho)(p_{n+1} - p_n)] + E_\theta[(p_{n+1} - p_n)^2] \\ &= E_\theta[(p_n - \rho)^2] + 2E_\theta[(p_n - \rho)V(p_n , \theta)] + E_\theta[M(p_n , \theta)]. \end{aligned}$$

Letting n approach ∞ and using the Helly-Bray lemma [Loève, 1963, p. 180] we obtain

$$0 = 2 \int_a^b (p - \rho)V(p, \theta) \, dF_\theta(p) + \int_a^b M(p, \theta) \, dF_\theta(p).$$

It follows from (*i*) that $V(p, 0) = M(p, 0) = (\partial/\partial\theta)M(p, 0) = 0$ for all $p \varepsilon I$. Hence, expanding $V(p, \cdot)$ and $M(p, \cdot)$ around 0, the former up to terms of second order and the latter up to terms of third order, and then expanding $(\partial^2/\partial\theta^2)M(\cdot, 0)$ around ρ up to terms of first order we obtain

$$\begin{aligned} -2 \int_a^b (p - \rho)\theta(\partial/\partial\theta)V(p, 0) \, dF_\theta(p) \\ = 2 \int_a^b (p - \rho)(\theta^2/2)(\partial^2/\partial\theta^2)V(p, \theta^*) \, dF_\theta(p) + (\theta^2/2)m \\ + \int_a^b (\theta^2/2)(p - \rho)(\partial^3/\partial p \, \partial\theta^2)M(p^*, \theta) \, dF_\theta(p) + O(\theta^3) \end{aligned}$$

where $0 < \theta^* < \theta$ and p^* is between p and ρ. But, by (*iii*), $(\partial/\partial\theta)V(p, 0)/(p - \rho) < 0$ when $p \varepsilon I$, $p \neq \rho$. Since $(\partial^2/\partial p \, \partial\theta)V(\rho, 0) < 0$ also, it follows that there is some $\gamma > 0$ such that $(\partial/\partial\theta)V(p, 0)/(p - \rho) < -\gamma$ or $-(\partial/\partial\theta) \cdot V(p, 0)/(p - \rho) > \gamma$ for all $p \varepsilon I$, where the quotient is defined by continuity at $p = \rho$. Since

$$\begin{aligned} 2 \int_a^b (p - \rho)^2[-(\partial/\partial\theta)V(p, 0)/(p - \rho)] \, dF_\theta(p) \\ = O\left(\theta \int_a^b |p - \rho| \, dF_\theta(p)\right) + (\theta/2)m + O(\theta^2), \end{aligned}$$

and since $\int_a^b |p - \rho| \, dF_\theta(p) = O(1)$, we obtain $\int_a^b (p - \rho)^2 \, dF_\theta(p) = O(\theta)$. Therefore $2\gamma \int_a^b (p - \rho)^2 \, dF_\theta(p) = O(\theta^{3/2}) + (\theta/2)m$. From this it follows that $m \geqq 0$ and that, if $m = 0$, then $\int_a^b (p - \rho)^2 \, dF_\theta(p) = O(\theta^{3/2})$. Q.E.D.

Proof of CLT. If $m = 0$ the lemma implies $\int_{-\infty}^{\infty} x^2 \, dG_\theta(x) = O(\theta^{1/2})$. From this it follows that G_θ converges to δ_0, as claimed. We assume henceforth that $m > 0$ so that the quantity σ^2 defined in the statement of the theorem is positive.

For $\theta \in S$,

$$E_\theta\{\exp [it(p_{n+1} - \rho)\theta^{-1/2}]\}$$
$$= E_\theta\{\exp [it(p_n - \rho)\theta^{-1/2}] \exp [it(p_{n+1} - p_n)\theta^{-1/2}]\}$$
$$= E_\theta\{\exp [it(p_n - \rho)\theta^{-1/2}] E_\theta\{\exp [it(p_{n+1} - p_n)\theta^{-1/2}] \mid p_n\}\}.$$

Now

$$E_\theta\{\exp [it(p_{n+1} - p_n)\theta^{-1/2}] \mid p_n\} = 1 + it\theta^{-1/2}V(p_n, \theta) - (t^2/2)\theta^{-1}M(p_n, \theta) + \omega(|t|^3/3!)\theta^{-3/2}E_\theta[|p_{n+1} - p_n|^3 \mid p_n]$$

where $|\omega| \leqq 1$ [Loève, 1963, p. 199]. Therefore substituting this into the previous equation, using (*ii*), letting $n \to \infty$, using the Helly-Bray lemma, and cancelling the term on the left of the resulting equation and the first term on the right we obtain

$$0 = \int_a^b \exp [it(p - \rho)\theta^{-1/2}] it\theta^{-1/2}V(p, \theta) \, dF_\theta(p) - \int_a^b \exp [it(p - \rho)\theta^{-1/2}](t^2/2)\theta^{-1}M(p, \theta) \, dF_\theta(p) + O(|t|^3 \, \theta^{3/2}).$$

The change of variables $x = (p - \rho)\theta^{-1/2}$ yields

$$0 = \int_{-\infty}^{\infty} \exp (itx) it\theta^{-1/2}V(\theta^{1/2}x + \rho, \theta) \, dG_\theta(x) - \int_{-\infty}^{\infty} \exp (itx)(t^2/2)\theta^{-1}M(\theta^{1/2}x + \rho, \theta) \, dG_\theta(x) + O(|t|^3 \, \theta^{3/2}).$$

By (*i*) and the boundedness of the relevant third partial derivatives we have the Taylor expansions

$$V(p, \theta) = \theta(p - \rho)(\partial^2/\partial p \, \partial\theta)V(\rho, 0) + O(\theta(p - \rho)^2) + O(\theta^2)$$

and

$$M(p, \theta) = (\theta^2/2)(\partial^2/\partial\theta^2)M(\rho, 0) + O(\theta^2 \, |p - \rho|) + O(\theta^3)$$

where $O(\theta^2)$ and $O(\theta^3)$ are uniform in p. Therefore

$$V(\theta^{1/2}x + \rho, \theta) = \theta^{3/2}xv + O(\theta^2x^2) + O(\theta^2)$$

and

$$M(\theta^{1/2}x + \rho, \theta) = (\theta^2/2)m + O(\theta^{5/2}|x|) + O(\theta^3).$$

But the lemma implies that $\int_{-\infty}^{\infty} x^2 \, dG_\theta(x) = O(1)$ and $\int_{-\infty}^{\infty} |x| \, dG_\theta(x) = O(1)$. Thus we can write

$$0 = \int_{-\infty}^{\infty} \exp(itx) itxv \, dG_\theta(x) - \int_{-\infty}^{\infty} \exp(itx)(t^2/4)m \, dG_\theta(x) + O((|t| + |t|^3)\theta^{1/2}).$$

Therefore

$$(d/dt)g_\theta(t) + \sigma^2 t g_\theta(t) = O((1 + t^2)\theta^{1/2})$$

where g_θ is the characteristic function of G_θ. The solution of this differential equation for which $g_\theta(0) = 1$ is

$$g_\theta(t) = \exp(-\sigma^2 t^2/2)\left(1 + \int_0^t \exp(\sigma^2 s^2/2) O((1 + s^2)\theta^{1/2}) \, ds\right).$$

Thus $g_\theta(t) \to \exp(-\sigma^2 t^2/2)$ as $\theta \to 0$. It follows that $G_\theta(x) \to \Phi(x/\sigma)$ for all x as $\theta \to 0$. Q.E.D.

It is easy to verify that all of the normal convergence theorems discussed in Norman [1966] follow from CLT. From the present vantage point, however, it can be seen that the hypotheses of some of these theorems are unnecessarily strong, while others are completely superfluous. In the latter category are assumptions (45) and (49) in Theorem 4. In the former class are the important assumptions (10) of Theorem 1, (56) and (57) of Theorem 4, and (70) of Theorem 5, which imply that

$$(\partial^2/\partial p \, \partial\theta) V(p, 0) < 0 \quad \text{for all} \quad p \in I. \tag{2}$$

This condition, in conjunction with $(\partial/\partial\theta)V(a, 0) > 0$ and $(\partial/\partial\theta)V(b, 0) < 0$, implies hypothesis (*iii*) of CLT above. The inequality (2) need not be satisfied under the hypotheses of Corollaries 1 and 2. Consider, for instance, the case treated by Corollary 1 under the additional assumption that $c_{1j} = c_{2j} = c_j$ (i.e. $\theta_{1j} = \theta_{2j} = \theta_j$), $j = 1, 2$. The condition (2) can be shown to hold for all $\pi_{12} > 0$ and $\pi_{21} > 0$ if $c_1 = c_2$, but to fail if $c_1 \neq c_2$ and π_{12} and π_{21} are sufficiently small.

REFERENCES

Bush, R. R., & Mosteller, F. *Stochastic models for learning.* New York: Wiley, 1955.

Estes, W. K. Component and pattern models with Markovian interpretations. In R. R. Bush and W. K. Estes (Eds.), *Studies in mathematical learning theory.* Stanford, Calif.: Stanford Univer. Press, 1959, 9–52.

Loève, M. *Probability theory.* (3rd ed.) Princeton, N. J.: Van Nostrand, 1963.

Norman, M. F. Limiting distributions for some random walks arising in learning models. *Annals of Mathematical Statistics*, 1966, 37, 393–405.

Norman, M. F. Some convergence theorems for stochastic learning models with distance diminishing operators. *Journal of Mathematical Psychology*, 1968, 5, 61–101.

Norman, M. F., & Yellott, J. I., Jr. Probability matching. *Psychometrika*, 1966, 31, 43–60.

Manuscript received 12/19/66
Revised manuscript received 8/30/67