ASYMPTOTIC RATE OF DISCRIMINATION FOR MARKOV PROCESSES[1]

BY LAMBERT H. KOOPMANS

Sandia Corporation

1. Summary. Simple hypotheses H_P and H_Q, specifying two distinct positive transition densities $p(x \mid y)$ and $q(x \mid y)$ and initial densities $p(x)$ and $q(x)$ with respect to a finite Lebesgue-Stieltjes measure, are assumed for a discrete time parameter Markov process. Let R_n be the likelihood ratio static based on the first $n + 1$ observations of the process, and consider the class of sequences of likelihood ratio tests $T(a, \alpha) = \{[R_n > n^{\alpha}a]: n = 0, 1, 2, \cdots\}$ generated by letting a and α vary over the real numbers. Under certain regularity assumptions on $K_t(x, y) = p^{1-t}(x \mid y)q^t(x \mid y)$ and the initial densities p and q, the subclass of consistent sequences is determined, and the limiting rates at which the error probabilities tend to zero for tests in this subclass are found.

A definition of the best asymptotic rate for distinguishing between H_P and H_Q is made for the class of consistent tests. This "asymptotic rate of discrimination" is evaluated and is shown to be attained by a certain subclass of these tests.

Some applications and extensions of the theory to infinite Lebesgue-Stieltjes measures are given.

2. Introduction. This paper is primarily a study of the asymptotic properties of the tail probabilities for the likelihood ratio statistic as applied to testing simple hypotheses for discrete time parameter Markov processes. Similar investigations have been carried out by Cramér [2], Chernoff [1], and Thomasian [6] for sums of independent random variables and from more general points of view. In the special case that the Markov process reduces to a sequence of independent, identically distributed random variables, most of the results proved herein may be obtained from these papers.

Let $\mathfrak{X}$ be the real line, $\mathfrak{B}$ the Borel sets, and μ a Lebesgue-Stieltjes measure defined on $(\mathfrak{X}, \mathfrak{B})$. Attention is restricted to the situation in which the distribution of a Markov process $\mathfrak{M} = \{X_n : n = 0, 1, 2, \cdots\}$ is determined by a transition density $p(x \mid y)$ (measurable in y for fixed x and a probability density with respect to μ in x for fixed y) and an initial density $p(x)$ with respect to μ.

Let H_P and H_Q be simple hypotheses for $\mathfrak{M}$ specifying transition densities $p(x \mid y)$ and $q(x \mid y)$ and initial densities p and q. The likelihood ratio statistic R_n based on the first $n + 1$ observations is then

$$(1.1) \qquad R_n(X_0, X_1, \cdots, X_n) = \log \frac{q(X_0)q(X_1 \mid X_0) \cdots q(X_n \mid X_{n-1})}{p(X_0)p(X_1 \mid X_0) \cdots p(X_n \mid X_{n-1})},$$

Received September 28, 1959; revised June 4, 1960.

[1] The major portion of the work on this paper was carried out while the author was a graduate student in the Department of Statistics, University of California, Berkeley. The work was supported (in part) by funds provided under contract AF41(657)-29 with the USAF School of Aviation Medicine, Randolph Field, Texas.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Mathematical Statistics.* www.jstor.org

and a (nonrandomized) likelihood ratio test is a set $[R_n > k_n]$ for some $k_n \varepsilon \mathfrak{X}$, where the square bracket indicates the set in $\mathfrak{B}^{n+1}$ for which the inequality is satisfied.

The error probabilities associated with $[R_n > k_n]$ are $P[R_n > k_n]$ and $Q[R_n \leqq k_n] = 1 - Q[R_n > k_n]$, where P and Q are the probability distributions induced on $\mathfrak{M}$ under H_P and H_Q. For simplicity it is assumed that k_n is of the form $n^\alpha a$ for real a and α. The results remain valid if k_n is taken to be $O(n^\alpha a)$. The tail probabilities to be studied are the error probabilities for the consistent members of the class of test sequences $T(a, \alpha) = \{[R_n > n^\alpha a]: n = 0, 1, 2, \cdots\}$ for a and α in $\mathfrak{X}$.

As in [1], [2], and [6], the moment-generating function, in this case the moment-generating function $M_n(t) = E_P[\exp(R_n t)]$, where E_P denotes expectation with respect to P, will play an important role in evaluating the limiting rates at which these tail probabilities tend to zero. The possibility of expressing $M_n(t)$ in terms of the nth iterate of a certain integral operator to be defined later motivates the use of operator theory in this investigation. For a complete treatment of the linear space theory pertinent to this study, the reader is referred to Zaanen [9].

The following assumptions apply throughout the paper:

A1. *Let* $K_t(x, y) = p^{1-t}(x \mid y)q^t(x \mid y)$. *There exists a set* $\Delta \varepsilon \mathfrak{B}$, *such that*

(i). $\mu(\Delta) < \infty$,

(ii). $K_t(x, y) > 0$ *on* $\Delta \times \Delta$ *and* $K_t(x, y) = 0$ *on* $(\Delta \times \Delta)^c$ *almost surely with respect to* $\mu \times \mu(\text{a.s.}(\mu \times \mu))$ *for* $t = 0$ *and* 1.

Hereafter all functions and measures will be restricted to Δ, its products, and their restricted Borel sets. For integrals over the range Δ, no domain of integration will be indicated.

Clearly, Part (ii) could equivalently be stated for the range $0 \leqq t \leqq 1$.

A2. *There exists* $\delta_1 > 0$, *such that the functions* $p^{1-t}(x)q^t(x)$ *are in* $\mathfrak{L}_2 = L_2(\Delta, \mu)$ *for* $-\delta_1 \leqq t \leqq 1 + \delta_1$.

A3. *There exists* $\delta_2 > 0$, *such that*

$$\sum_{n=0}^{\infty} \frac{1}{n!} ||| K_{t,n} ||| \delta_2^n < \infty$$

for $t = 0, 1$, *where* $K_{t,n}(x, y) = \{\log [q(x \mid y)/p(x \mid y)]\}^n K_t(x, y)$ *and* $||| \; |||$ *is the double norm defined for (possibly complex valued)* $\mu \times \mu$ *measurable functions*

$$G(x, y) \text{ by } ||| G ||| = \left\{ \iint | G(x, y) |^2 \, d\mu(x) \, d\mu(y) \right\}^{\frac{1}{2}}.$$

This assumption is a specialization of a condition imposed by Wolf in [8] for analyticity of operators in Banach space.

LEMMA 2.1. *Assumption* A3 *implies the following conditions*:

(C1). $$||| K_t ||| < \infty \quad \textit{for} \quad 0 \leqq t \leqq 1.$$

(C2). $$\sum_{n=0}^{\infty} \frac{1}{n!} ||| K_{t,n} ||| \delta_2^n < \infty \quad \text{for} \quad 0 \leqq t \leqq 1.$$

PROOF. (C1) is an immediate consequence of (C2). (C2) follows from A3 by two applications of Hölder's inequality: if $\int |f|^r d\psi < \infty$ and $\int |g|^s d\psi < \infty$, where $r^{-1} + s^{-1} = 1$, then $\int |fg| \, d\psi \leqq [\int |f|^r d\psi]^{1/r} [\int |g|^s d\psi]^{1/s}$ with strict inequality unless $f = kg$ except on a ψ null set for some constant k. First apply the inequality to $f(x, y) = [f_n(x, y) p(x \mid y)]^{2(1-t)}$, $g(x, y) = [f_n(x, y) q(x \mid y)]^{2t}$, $r = (1 - t)^{-1}$, $s = t^{-1}$, and $\psi = \mu \times \mu$ where $f_n(x, y) = \{\log [q(x \mid y)/p(x \mid y)]\}^n$. This results in the inequality $||| k_{t,n} ||| \leqq ||| K_{0,n} |||^{1-t} ||| K_{1,n} |||^t$. Next apply Hölder's inequality to $f(n) = \{n!^{-1} ||| K_{0,n} ||| \delta_2^n\}^{1-t}$, $g(n) = \{n!^{-1} ||| K_{1,n} ||| \delta_2^n\}^t$, $r = (1 - t)^{-1}$, and $s = t^{-1}$, and to ψ, the measure which assigns mass one to the integers 0, 1, 2, $\cdots$, and is zero otherwise. Then for

$$0 \leqq t \leqq 1, \sum_{n=0}^{\infty} \frac{1}{n!} ||| K_{t,n} ||| \delta_2^n \leqq \sum_{n=0}^{\infty} \left\{ \frac{1}{n!} ||| K_{0,n} ||| \delta_2^n \right\}^{1-t} \left\{ \frac{1}{n!} ||| K_{1,n} ||| \delta_2^n \right\}^t$$

$$\leqq \left\{ \sum_{n=0}^{\infty} \frac{1}{n!} ||| K_{0,n} ||| \delta_2^n \right\}^{1-t} \left\{ \sum_{n=0}^{\infty} \frac{1}{n!} ||| K_{1,n} ||| \delta_2^n \right\}^t < \infty.$$

The following lemma, which is proved in Zaanen [9], pages 496–498, is a consequence of the above conditions.

LEMMA 2.2. *For* $0 \leqq t \leqq 1$ *the integral operator* A_t *with kernel* $K_t(x, y)$ *possesses a positive eigenvalue* $\lambda(t)$ *which is strictly larger in modulus than any other eigenvalue of* A_t. *Furthermore, every eigenfunction corresponding to* $\lambda(t)$ *is of the form* $\psi_t = k\phi_t$ *for some constant* k, *where* ϕ_t *is an* a.s. (μ) *positive function belonging to* $\mathfrak{L}_2$.

3. The Class of Consistent Likelihood Ratio Tests.

LEMMA 3.1. *There exist unique stationary densities* π_0 *and* π_1 *for* $\mathfrak{M}$ *under* H_P *and* H_Q, *respectively, which are positive* a.s. (μ).

PROOF. By Lemma 2.2, there exist numbers $\lambda(t) > 0$ and functions ϕ_t positive a.s. (μ) such that $\lambda(t)\phi_t(x) = A_t \, \phi_t(x) = \int K_t(x, y)\phi_t(y) d\mu(y)$. Taking $\mathfrak{L}_1$ ($= L_1(\Delta, \mu)$) norm on both sides of this expression, one obtains $\lambda(t) \, \| \phi_t \|_1 = \int \phi_t(y) \{ \int K_t(x, y) d\mu(x) \} d\mu(y) = \| \phi_t \|_1$ for $t = 0, 1$; hence, $\lambda(0) = \lambda(1) = 1$. The proof is concluded by taking $\pi_t = \phi_t / \| \phi_t \|_1$.

LEMMA 3.2. *There exist numbers* γ_P *and* γ_Q *independent of the initial densities* p *and* q *such that*

(i) $$\lim_{n \to \infty} R_n/n = \gamma_P \qquad \text{a.s. } (P)$$

when H_P *is true and*

$$\lim_{n \to \infty} R_n/n = \gamma_Q \qquad \text{a.s. } (Q)$$

when H_Q *is true, where* P *and* Q *are the probability distributions induced on* $\mathfrak{M}$ *by the densities specified by* H_P *and* H_Q,

(ii) $\gamma_P \leqq 0 \leqq \gamma_Q$ *with strict inequality unless* $p(x \mid y) = q(x \mid y)$ a.s. ($\mu \times \mu$).

PROOF. $\mathfrak{M}$, with the distribution P_t generated by $K_t(x, y)$ and π_t, is a metrically transitive stationary stochastic process for $t = 0, 1$. If E_t denotes expectation with respect to P_t, and B_t is the integral operator with kernel

$$|\log [q(x \mid y)/p(x \mid y)]| K_t(x, y),$$

then the inequality $E_t |\log [q(X \mid Y)/p(X \mid Y)]| < \infty$ for $t = 0, 1$ follows from the inequalities $(1, B_t\pi_t) \leqq \| 1 \| \| B_t\pi_t \|$ (Schwarz inequality) and

$$\| B_t\pi_t \| \leqq \| \pi_t \| \||K_{t,1}\||.$$

The last expression is finite by $A3$. The absence of a subscript on the norm symbol indicates $\mathfrak{L}_2$ norm. Under these conditions Birkhoff's ergodic theorem applies to the function $\log [q(x \mid y)/p(x \mid y)]$ yielding the result

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \log \frac{q(X_k \mid X_{k-1})}{p(X_k \mid X_{k-1})} = E_t \log \frac{q(X \mid Y)}{p(X \mid Y)} \quad \text{a.s. } (P_t) \tag{3.1}$$

when the distribution of $\mathfrak{M}$ is P_t, $t = 0, 1$.

Now

$$\frac{R_n}{n} = \frac{1}{n} \log \frac{\pi_1(X_0)}{\pi_0(X_0)} + \frac{1}{n} \sum_{k=1}^{n} \log \frac{q(X_k \mid X_{k-1})}{p(X_k \mid X_{k-1})}$$

and π_0 and π_1 are a.s. positive; hence $\lim R_n/n = E_t \log [q(X \mid Y)/p(X \mid Y)]$ a.s. (P_t) when P_t is the distribution of $\mathfrak{M}$, $t = 0, 1$. If we let $\gamma_P = E_0 \log [q(X \mid Y)/p(X \mid Y)]$ and $\gamma_Q = E_1 \log [q(X \mid Y)/p(X \mid Y)]$, Part (i) is proved for the case $P = P_0$ and $Q = P_1$.

To see that this result is valid for any pair of initial densities p and q for which $\log [q/p]$ is finite a.s. (μ), let S_t be the set in the range of the process for which (3.1) holds. Then Birkhoff's ergodic theorem implies $P_t(S_t) = 1$ for $t = 0, 1$. Write

$$P_t(S_t) = \int P_t(S_t \mid X_0 = x)\pi_t(x) \, d\mu(x),$$

where $P_t(S_t \mid X_0)$ is the conditional probability of S_t relative to the sigma field of events generated by X_0. Since $\pi_t > 0$ a.s. (μ), $P_t(S_t \mid X_0) = 1$ a.s. (μ). Thus, for arbitrary densities p and q, $P(S_0) = \int P_0(S_0 \mid X_0 = x)p(x)d\mu(x) = 1$ and $Q(S_1) = \int P_1(S_1 \mid X_0 = x)q(x)d\mu(x) = 1$. The proof of Part (i) is completed by noting that Condition A2 implies $\log [q/p]$ is finite a.s. (μ).

Because $\log x$ is concave, $E(\log X) \leqq \log E(X)$ and Part (ii) follows.

To avoid the case $\gamma_P = \gamma_Q = 0$, it is assumed that $p(x \mid y)$ and $q(x \mid y)$ differ on a set of positive $\mu \times \mu$ measure.

THEOREM 1. *Among the class of tests $T(a, \alpha)$, the consistent ones are:*

(i) $T(0, \alpha)$ *for* $\alpha > 1$,
(ii) $T(a, 1)$ *for* $\gamma_P < a < \gamma_Q$ *(and possibly* $a = \gamma_P$*)*,
(iii) $T(a, \alpha)$ *for* $\alpha < 1$ *and* $-\infty < a < \infty$.

Proof. Part (ii) is an immediate consequence of Lemma 3.2. Write

$$P[R_n > n^\alpha a] = P[R_n/n > n^{\alpha-1}a] \qquad \text{for } \alpha > 1.$$

Then for $a < 0$ and $\epsilon > 0$, $n^{\alpha-1}a < \gamma_P - \epsilon$ for n sufficiently large, which by Part (ii) implies $\lim_n P[R_n > n^\alpha a] \geqq \lim_n P[R_n/n > \gamma_P - \epsilon] = 1$. Similarly, $\lim_n Q[R_n \leqq n^\alpha a] = 1$ for $a > 0$; hence, the only consistent test when $\alpha > 1$ is $T(0, \alpha)$.

For $\alpha < 1$, write $P[R_n > n^\alpha a] = P[R_n/n > a/n^{1-\alpha}]$. Let $\epsilon > 0$ satisfy $\gamma_P + \epsilon < 0$. Then for arbitrary a, $a/n^{1-\alpha} > \gamma_P + \epsilon$ for sufficiently large n. Hence, $\lim_n P[R_n > n^\alpha a] \leqq \lim_n P[R_n/n > \gamma_P + \epsilon] = 0$. A similar argument applies to $\lim_n Q[R_n \leqq n^\alpha a]$, which completes the proof.

4. Properties of the Sequence of Moment Generating Functions. Let $M_n(t)$ be the moment generating function of R_n under the hypothesis H_P. Then $M_n(t)$ is the real restriction of the bilateral Laplace transform $M_n(z) = \int_{-\infty}^{\infty} e^{zx} dF_n(x)$, where $F_n(x) = P[R_n \leqq x]$. Hence, as is well known from the theory of the bilateral Laplace transform, $M_n(z)$ is convergent and analytic in an infinite strip $\{z = t + is : \alpha_n < t < \beta_n\}$ as is $[M_n(z)]^{1/n}$. Since $M_n(0) = M_n(1) = 1$, $\alpha_n \leqq 0$ and $\beta_n \geqq 1$. The following lemma strengthens this property.

Lemma 4.1. *Let $\delta = \min(\delta_1, \delta_2)$ where δ_1 and δ_2 are specified in* A1 and A2. *Then there exists a constant $M < \infty$ such that* $|[M_n(z)]^{1/n}| < M$ for $-\delta \leqq \Re(z) \leqq 1 + \delta$, *where $\Re(z)$ denotes the real part of z.*

Proof. Let A_z be the (complex-valued) integral operator with kernel $K_z(x, y) = p^{1-z}(x \mid y)q^z(x \mid y)$. It is easily verified that $M_n(z) = (1, A_z^n k_z)$ where $k_z = p^{1-z}q^z$. Let $A_{z,n}$ be the integral operator with kernel $K_{z,n}(x, y)$ defined as in A3 with z replacing t. Then $\| A_{z,n} \| \leqq ||| K_{z,n} ||| = ||| K_{t,n} |||$ for $t = \Re(z)$ and, as a consequence of C3,

$$\sum_{n=0}^{\infty} \| A_{z_0,n} \| \, | z - z_0 |^n < \infty \quad \text{for} \quad 0 \leqq \Re(z_0) \leqq 1$$

and any z satisfying $| z - z_0 | \leqq \delta_2$. This implies (see Wolf [8]) that the definition of the operator A_z may be extended to $-\delta_2 \leqq \Re(z) \leqq 1 + \delta_2$ by the equation

$$A_z = \sum_{n=0}^{\infty} A_{z_0,n}(z - z_0)^n \quad \text{with} \quad \| A_z \| \leqq L_t = \sum_{n=0}^{\infty} ||| K_{t,n} ||| \, \delta_2^n ,$$

where $t = \Re(z_0)$. But as was shown in Lemma 2.1, $L_t \leqq L_0^{1-t}L_1^t \leqq L = \max(L_0, L_1)$; hence, for $-\delta_2 \leqq \Re(z) \leqq 1 + \delta_2$, $\| A_z \|$ is uniformly bounded by L.

Now $| M_n(z) | = | (1, A_z^n k_z) | \leqq \| 1 \| \, \| k_z \| \, \| A_z \|^n$; hence, setting

$$C = \sup\{\| k_z \| : -\delta_1 \leqq \Re(z) \leqq 1 + \delta_1\}, \qquad |[M_n(z)]^{1/n}| \leqq L(\| 1 \| \, C)^{1/n}$$

for $-\delta \leqq \Re(z) \leqq 1 + \delta$. The proof is completed by letting

$$M = L \sup_n (\| 1 \| \, C)^{1/n}.$$

COROLLARY. $\alpha_n \leqq -\delta$ *and*, $\beta_n \geqq 1 + \delta$; *hence* $[M_n(z)]^{1/n}$ *is analytic for* $-\delta < \Re(z) < 1 + \delta, n = 1, 2, \cdots$.

LEMMA 4.2. *Let* $A_t(0 \leqq t \leqq 1)$ *be the integral operator with kernel* $K_t(x, y)$ *and let* $\sigma(A_t)$ *denote the set of nonzero eigenvalues of* A_t. *Then*

(i). *there exist integral operators* T_t *and* B_t *with kernels of finite double norm such that*

$$A_t^n = \lambda^n(t)T_t + B_t^n$$

for $n = 1, 2, \cdots$, *where* $\lambda(t)$, *the largest eigenvalue of* A_t, *is positive,*

(ii). $T_t f > 0$ *for every nonnegative function* $f \varepsilon \mathfrak{L}_2$ *that is positive on a set of positive* μ *measure, and*

(iii). $$\sigma(B_t) = \{\nu: \quad \nu \neq \lambda(t), \ \nu \ \varepsilon \ \sigma(A_t)\}.$$

PROOF. The index t will be omitted in the proof of this and the next lemma.

(i) The kernel $K^*(x, y)$ of the adjoint operator A^* is $K(y, x)$; hence, both $K(x, y)$ and $K^*(x, y)$ satisfy the condition of Lemma 2.2. This implies that the largest eigenvalues of A and A^* are positive and, as is well known, are identical. The eigenfunctions ϕ and ϕ^* corresponding to this eigenvalue for A and A^* are positive and unique up to constant multiples. Since $(\phi, \phi^*) > 0$, ϕ and ϕ^* may be chosen so that $(\phi, \phi^*) = 1$. Define the operator T by $Tf = (f, \phi^*)\phi$ for $f \varepsilon \mathfrak{L}_2$ and let $B = A - \lambda T$. Then it is easily seen by induction on $B^n = (A - \lambda T)^n$ that $A^n = \lambda^n T + B^n$, provided

$$TA = AT = \lambda T \tag{4.1}$$

and

$$T^2 = T. \tag{4.2}$$

To prove (4.2), let $f \varepsilon \mathfrak{L}_2$. Then

$$T^2 f = (Tf, \phi^*)\phi = ((f, \phi^*)\phi, \phi^*)\phi = (f, \phi^*)(\phi, \phi^*)\phi = Tf.$$

A similar computation proves (4.1).

The operators T and B have kernels $K_T(x, y) = \phi^*(y)\phi(x)$ and $K_B(x, y) = K(x, y) - \lambda K_T(x, y)$. These kernels are of finite double norm since $||| K_T ||| = \| \phi^* \| \| \phi \|$ and $||| K_B ||| \leqq ||| K ||| + \lambda ||| K_T |||$.

(ii) The fact that $T_t f > 0$ is clearly a consequence of the positiveness of ϕ and ϕ^*.

(iii) It remains to be shown that $\sigma(B) = \sigma(A) \sim \{\lambda\}$ where $\sim$ denotes set theoretic difference. Let $\nu \neq 0$, $\nu \neq \lambda$ satisfy $Af = \nu f$ for some nonzero $f \varepsilon \mathfrak{L}_2$. Then $\nu Tf = \nu(f, \phi^*)\phi = (Af, \phi^*)\phi = (f, A^*\phi^*)\phi = \lambda Tf$ which implies $(\nu - \lambda)Tf = 0$ or $Tf = 0$, since $\nu \neq \lambda$. Thus $Bf = \nu f$, from which it follows that $\sigma(A) \subset \sigma(B) \sim \{\lambda\}$.

If $\nu \neq 0$ satisfies $Bf = \nu f$ for a nonzero $f \varepsilon \mathfrak{L}_2$, then $Af = \lambda Tf + \nu f$ and, applying T to both sides, $\nu Tf = \lambda Tf - \lambda Tf = 0$. Thus $Af = \nu f$ which implies $\sigma(B) \subset \sigma(A)$.

But λ is not an eigenvalue of B since if it were, $Bf = \lambda f$ for some nonzero $f \varepsilon \mathfrak{L}_2$. Since $B\phi = A\phi - \lambda T\phi = 0, f \neq k\phi$ for all k. Now $Af = \lambda Tf + \lambda f$ and, applying T to both sides, $Tf = 2Tf$ and $Tf = 0$. Consequently $Af = \lambda f$, but this is impossible since $f \neq k\phi$. Thus λ is not in $\sigma(B)$ and the lemma is proved.

LEMMA 4.3. $\lim_{n\to\infty} [\| B_t^n \|/(\lambda^n(t))] = 0$ *for* $0 \leqq t \leqq 1$.

PROOF. Since the eigenvalues of an integral operator with kernel of finite double norm are isolated in any region of the complex plane not containing the origin, $r_B = \sup \{| \nu | : \nu \varepsilon \sigma(B)\}$ is attained for some $\nu' \varepsilon \sigma(B)$. By Lemma 4.2, $r_B = | \nu | < \lambda$. A well-known theorem in the theory of linear operators yields $\lim_n \| B^n \|^{1/n} = r_B$. Let $\epsilon > 0$ satisfy $r_B < r_B + \epsilon < \lambda$. Then for n sufficiently large, $\| B^n \|^{1/n} \leqq r_B + \epsilon$, which implies $\lim_n \| B^n \|/\lambda^n = \lim_n[(r_B + \epsilon)/\lambda]^n = 0$.

THEOREM 2. *For every pair of initial densities p and q satisfying* A2,

(i) $$\lim_{n\to\infty} [M_n(t)]^{1/n} = \lambda(t) \quad for \quad 0 \leqq t \leqq 1,$$

(ii) $\lambda(t)$ *is convex and continuous and has a continuous first derivative for* $0 \leqq t \leqq 1$,

(iii) $$\lim_{n\to\infty} (d/dt)[M_n(t)]^{1/n} = \lambda'(t) \quad for \quad 0 \leqq t \leqq 1.$$

PROOF. Let f be a nonzero, nonnegative element of $\mathfrak{L}_2$. Then

$$(1, A_t^n f) = (1, \lambda^n(t) T_t f + B_t^n f) = \lambda^n(t)(1, T_t f + B_t^n f/\lambda^n(t)).$$

Hence, by Lemmas 4.2 and 4.3, $\lim_n (1, A_t^n f)^{1/n} = \lambda(t)$ independent of f.

If $f = \pi_t$, $(1, A_t^n f) = M_n(t)$; hence $\lim_n [M_n(t)]^{1/n} = \lambda(t)$ for $0 \leqq t \leqq 1$. As a consequence of this result, Lemma 4.1, and its corollary, Vitali's theorem (see Titchmarsh [7], page 168) implies that $[M_n(z)]^{1/n}$ tends to a limit $\lambda(z)$ uniformly in any region bounded by a contour interior to $\{z: -\delta < \mathfrak{R}(z) < 1 + \delta\}$. But then $\lambda(z)$ is analytic in this region, which implies the continuity part of (ii).

Each $[M_n(t)]^{1/n}$ is seen to be convex for $0 \leqq t \leqq 1$, since its second derivative is positive. Because convexity is preserved under passage to the limit, $\lambda(t)$ is also convex.

Part (iii) is a consequence of the analyticity and uniformity of convergence of $[M_n(z)]^{1/n}$.

COROLLARY.

$$\lim_{n\to\infty} E_P(R_n/n) = \lambda'(0) \quad and \quad \lim_{n\to\infty} E_Q(R_n/n) = \lambda'(1),$$

where E_P *and* E_Q *denote expectations with respect to the distributions defined on* $\mathfrak{M}$ *by* H_P *and* H_Q.

PROOF. Since a bilateral Laplace transform may be differentiated under the integral sign in its region of convergence, the corollary is an immediate consequence of Part (iii) of Theorem 2.

5. Limiting Rate of Convergence for the Tail Probabilities of Consistent Test Sequences.

LEMMA 5.1.

(i) $$\limsup_{n\to\infty} P[R_n > na]^{1/n} \leqq \inf_{0\leqq t\leqq 1} e^{-at}\lambda(t)$$

and

(ii) $$\limsup_{n\to\infty} Q[R_n \leqq na]^{1/n} \leqq \inf_{0\leqq t\leqq 1} e^{at}\lambda(1-t) \quad \textit{for all } a \in \mathfrak{X}.$$

PROOF. For an arbitrary random variable X with distribution P on $(\mathfrak{X}, \mathfrak{B})$, a well-known inequality (cf. Lóeve [5], page 157) is $P[X \geqq 0] \leqq Ee^{tx}$ for $t \geqq 0$. Set $X = R_n - na$ and let E_P denote expectation with respect to the distribution defined by H_P. Then $P[R_n > na] \leqq E_P \exp[(R_n - na)t] = e^{-nat}M_n(t)$ for $t \geqq 0$. Hence, by Theorem 2, $\lim\sup_n P[R_n > na]^{1/n} \leqq e^{-at} \lim_n [M_n(t)]^{1/n} = e^{-at}\lambda(t)$ for $0 \leqq t \leqq 1$ which implies Inequality (i). Part (ii) is proved in the same manner by setting $X = -(R_n - na)$ and noting that $E_Q \exp(-R_n t) = M_n(1-t)$.

The remainder of this section is devoted to showing that the limits $\lim_n P[R_n > na]^{1/n}$ and $\lim_n Q[R_n \leqq na]^{1/n}$ exist and are given by the expressions in Lemma 5.1. The method of proof will depend upon a modification of an inequality due to Thomasian [6].

To simplify the notation in the proofs of Lemmas 5.2 and 5.3, let

$$p_n = p(x_0)p(x_1 \mid x_0) \cdots p(x_n \mid x_{n-1}),$$

and

$$q_n = q(x_0)q(x_1 \mid x_0) \cdots q(x_n \mid x_{n-1}).$$

Integrals for which no measure is indicated are taken with respect to $d\mu^{n+1}$ where μ^{n+1} is the $n + 1$ dimensional product of μ.

LEMMA 5.2. *For every* $a \in \mathfrak{X}$, t $(0 < t < 1)$, *and* $n (n = 1, 2, \cdots)$,

(i) $$P[R_n > na] \geqq e^{-nbt} M_n(t)P_{t,n}[a < (R_n/n) \leqq b] \quad \textit{for every } b > a,$$

and

(ii) $$Q[R_n \leqq na] \geqq e^{nbt}M_n(1-t)P_{1-t,n}[b < (R_n/n) \leqq a] \quad \textit{for every } b < a,$$

where

$$P_{t,n}(A) = \frac{1}{M_n(t)} \int_A p_n^{1-t}q_n^t \qquad \textit{for } A \in \mathfrak{B}^{n+1}.$$

PROOF. To prove Part (i), let

$$A_n = [R_n > na] \quad \text{and} \quad B_n = [a < (R_n/n) \leqq b].$$

Then $A_n \supset B_n$ and

$$P[R_n > na] = \int_{A_n} p_n = M_n(t) \int_{A_n} \left(\frac{q_n}{p_n}\right)^{-t} dP_{t,n}$$
$$\geqq M_n(t) \int_{B_n} \left(\frac{q_n}{p_n}\right)^{-t} dP_{t,n} \geqq e^{-nbt} M_n(t) P_{t,n}(B_n).$$

Part (ii) is obtained by altering this proof slightly.

LEMMA 5.3. *For every* $a \in \mathfrak{X}$, $t(0 < t < 1)$, $s(0 < s < \min(t, 1-t))$, *and* $n(n = 1, 2, \cdots)$,

$$\text{(i)} \qquad P[R_n > na] \geqq e^{-nbt} M_n(t) \left[1 - e^{nas} \frac{M_n(t-s)}{M_n(t)} - e^{-nbs} \frac{M_n(t+s)}{M_n(t)}\right]$$

for every $b > a$, *and*

$$\text{(ii)} \quad Q[R_n \leqq na] \geqq e^{nbt} M_n(1-t) \left[1 - e^{nbs} \frac{M_n(1-t+s)}{M_n(1-t)} - e^{-nas} \frac{M_n(1-t-s)}{M_n(1-t)}\right]$$

for every $b < a$.

PROOF. Let $C_n = [R_n \leqq na]$ and $D_n = [R_n > nb]$. From Lemma 5.2 (i),

$$P[R_n > na] \geqq e^{-nbt} M_n(t)[1 - P_{t,n}(C_n) - P_{t,n}(D_n)]. \tag{5.1}$$

But

$$P_{t,n}(C_n) = e^{nas}\left\{e^{-nas} \int_{C_n} \frac{p_n^{1-t} q_n^{t}}{M_n(t)}\right\} \leqq e^{nas} \int_{C_n} \frac{p_n^{1-t+s} q_n^{t-s}}{M_n(t)} \leqq e^{nas} \frac{M_n(t-s)}{M_n(t)}$$

for $0 < s < \min(t, 1-t)$. Similarly, $P_{t,n}(D_n) \leqq e^{-nbs}[M_n(t+s)/M_n(t)]$. The proof of Part (i) is completed by substituting these inequalities into Expression (5.1).

Again, the proof of Part (ii) follows from the proof of Part (i) with only minor changes.

THEOREM 3. *Let* $m_P(a, \alpha) = \lim_n P[R_n > n^\alpha a]^{1/n}$ and

$$m_Q(a, \alpha) = \lim_n Q[R_n \leqq n^\alpha a]^{1/n}.$$

Then

(i) $m_P(a, 1) = \inf_{0 \leqq t \leqq 1} e^{-at}\lambda(t)$ *and* $m_Q(a, 1) = \inf_{0 \leqq t \leqq 1} e^{at}\lambda(1-t)$ *for* $\lambda'(0) < a < \lambda'(1)$,

(ii) $m_P(a, \alpha) = m_Q(a, \alpha) = \inf_{0 \leqq t \leqq 1} \lambda(t)$ *for* $\alpha < 1$ *and all* $a \in \mathfrak{X}$.

PROOF. To prove the first part of (i), it suffices to show that for every a and b for which $\lambda'(0) \leqq a \leqq b \leqq \lambda'(1)$, there exists $t^*(0 < t^* < 1)$ such that for all sufficiently small $s > 0$,

$$\lim_{n \to \infty} e^{nas}[M_n(t^* - s)/(M_n(t^*))] = \lim_{n \to \infty} e^{-nbs}[M_n(t^* + s)/(M_n(t^*))] = 0. \tag{5.2}$$

For then by Lemma 5.3, $\liminf_n P[R_n > na]^{1/n} \geqq e^{-bt^*}\lambda(t^*) \geqq \inf_{0 \leqq t \leqq 1} e^{-bt}\lambda(t)$ for every $b > a$, which implies $\liminf_n P[R_n > na]^{1/n} \geqq \inf_{0 \leqq t \leqq 1} e^{-at}\lambda(t)$.

If $c_n(t) = (1, A_t^n \pi_t / \lambda^n(t))$, then $M_n(t) = c_n(t)\lambda^n(t)$ and, by Lemmas 4.2 and 4.3, $\lim_n c_n(t) > 0$ for $0 \leqq t \leqq 1$. Thus, if we write

$$e^{nas}[M_n(t-s)/(M_n(t))] = [c_n(t-s)/(c_n(t))]\{[\lambda(t-s)/(\lambda(t))]e^{as}\}^n$$

and

$$e^{-nbs}[M_n(t+s)/(M_n)t))] = [c_n(t+s)/(c_n(t))]\{[\lambda(t+s)/(\lambda(t))]e^{-bs}\}^n,$$

a sufficient condition for Equation (5.2) to hold is

$$[\log \lambda(t^*) - \log \lambda(t^* - s)]/s > a \quad \text{and} \quad [\log \lambda(t^* + s) - \log \lambda(t^*)]/s < b$$

for some $t^*(0 < t^* < 1)$ and sufficiently small $s > 0$. But by Theorem 2(ii), $d \log \lambda(t)/dt = \lambda'(t)/\lambda(t)$ is continuous for $0 \leqq t \leqq 1$; hence, this condition can always be met for a and b in the specified range. The second part of (i) follows by a similar argument.

From the inequality quoted in the proof of Lemma 5.1, $P[R_n > n^\alpha a] \leqq \exp(-n^\alpha at)\, M_n(t)$ for $t \geqq 0$. Hence for $\alpha < 1$, $\limsup_n P[R_n > n^\alpha a]^{1/n} \leqq \lim_n \exp(-n^{\alpha-1}at) \lim_n [M_n(t)]^{1/n} = \lambda(t)$ for $0 \leqq t \leqq 1$ and all $a \varepsilon \mathfrak{X}$. This implies $\limsup_n P[R_n > n^\alpha a]^{1/n} \leqq \inf_{0 \leqq t \leqq 1} \lambda(t)$.

Now for every $\epsilon > 0$, $n^{\alpha-1}a < \epsilon$ for sufficiently large n whatever be a. Hence $\liminf_n P[R_n > n^\alpha a]^{1/n} \geqq \lim_n P[R_n > n\epsilon]^{1/n}$, and the first part of (ii) follows from (i). Again, a similar proof holds for the second part of (ii).

It is not yet clear that the exponential rate of convergence embraces all consistent tests. It is easy to show that $\gamma_P \leqq \lambda'(0)$ and $\lambda'(1) \leqq \gamma_Q$, but the inequalities may be strict. The following theorem resolves this point.

THEOREM 4. $\quad \gamma_P = \lambda'(0)$, *and* $\gamma_Q = \lambda'(1)$.

PROOF. The proof will be carried out only for the first equality since the proof of the second equality is quite similar.

By means of the inequality $P[X > 0] \leqq Ee^{tx}$, $t \geqq 0$, it is seen by the method of proof of Lemma 5.1 that for all $k \geqq 0$, $P[R_n/n > k] \leqq e^{-nk}$. The same inequality yields $P[R_n/n < a] \leqq e^{nta}M_n(-t)$ for $t \geqq 0$ where it will be assumed $a < \gamma_P \leqq 0$. From the proof of Theorem 2(i), $\lim_n[M_n(-\delta/2)]^{1/n} = \lambda(-\delta/2)$. Hence for $\epsilon > 0$ and n sufficiently large,

$$P[R_n/n < a] \leqq e^{n\delta a/2}[\lambda(-\delta/2) + \epsilon]^n.$$

Select a so small that $e^{\delta a/2}[\lambda(-\delta/2) + \epsilon] < 1$ and let b be any positive integer. Define the sequence of bounded random variables Z_n by $Z_n = R_n/n$ for $a \leqq R_n/n \leqq b$ and $Z_n = 0$ otherwise. Then, since $\lim_n R_n/n = \gamma_P$ a.s. (P) and $a < \gamma_P < b$, $\lim_n Z_n = \gamma_P$ a.s. (P). Hence by the bounded convergence theorem,

$\lim_n E_P Z_n = \gamma_P$. Now

$$\begin{aligned} | E_P(R_n/n) - \gamma_P | &\leqq | E_P Z_n - \gamma_P | + | E_P (R_n/n) - E_P Z_n | \\ &\leqq | E_P Z_n - \gamma_P | + \sum_{k=0}^{\infty} (|a| + k + 1) \\ &\qquad \cdot P\left[a - k - 1 \leqq \frac{R_n}{n} < a - k\right] \\ &\qquad + \sum_{k=b}^{\infty} (k+1) P\left[\frac{R_n}{n} > k\right] \\ &\leqq | E_P Z_n - \gamma_P | + e^{\frac{n\delta a}{2}} \left[\lambda\left(-\frac{\delta}{2}\right) + \epsilon\right]^n \\ &\qquad \cdot \left[| a | + \left(1 - e^{-\frac{n\delta}{2}}\right)^{-2}\right] + \frac{e^{-nb}[b(1 - e^{-n}) + 1]}{(1 - e^{-n})^2}, \end{aligned}$$

and this upper bound tends to zero with n. Hence $\lim_n E_P R_n/n = \gamma_P$ as was to be shown.

6. Definition and Evaluation of the Asymptotic Rate of Discrimination.

A possible criterion for selecting a "best" test from the class of consistent tests is the minimax principle based on a loss function involving the asymptotic values $m_P(a, \alpha)$ and $m_Q(a, \alpha)$.[2] Since the asymptotic behavior of the sequences $T(a, \alpha)$ for $\alpha < 1$ is equivalent to that of the sequence $T(0, 1)$, it suffices to restrict attention to $T(a, 1)$ for $\gamma_P < a < \gamma_Q$. The asymptotic rate of discrimination of the class of consistent likelihood ratio tests is defined to be the minimax rate

$$\rho(P, Q) = \inf_{\gamma_P < a < \gamma_Q} \max \{m_P(a, 1), m_Q(a, 1)\}.$$

THEOREM 5. $$\rho(P, Q) = \inf_{0 \leqq t \leqq 1} \lambda(t).$$

PROOF. $m_Q(a, 1) = \inf_{0 \leqq t \leqq 1} e^{at}\lambda(1 - t) = e^a \inf_{0 \leqq t \leqq 1} e^{-at}\lambda(t) = e^a m_P(a, 1)$.

Thus $m_Q(0, 1) = m_P(0, 1) = \inf_{0 \leqq t \leqq 1} \lambda(t)$. The theorem now follows from the fact that $m_Q(a, 1)$ is nondecreasing and $m_P(a, 1)$ is nonincreasing in a.

The asymptotic rate $\rho(P, Q)$ is achieved by the sequences $T(a, \alpha)$ for $\alpha < 1$ and $-\infty < a < \infty$, so the "best" test is actually an equivalence class of tests which are indistinguishable on the basis of the asymptotic rates at which their error probabilities tend to zero.

7. Application of the Theory and Extensions to Infinite Measures.

A. A class of processes for which the theory is immediately applicable is the one whose members possess densities bounded from above and away from zero

[2] The referee points out for those adverse to using the minimax principle that the general use of ρ may be justified by its relevance for any Bayes strategy corresponding to fixed nonzero *a priori* probabilities and fixed nonzero costs of error.

on a set of finite measure. More precisely, if there exists a constant $C_1 > 1$ such that for $t = 0, 1, 1/C_1 \leqq K_t(x, y) \leqq C_1$ on $\Delta \times \Delta$ and $K_t(x, y) = 0$ on $(\Delta \times \Delta)^c$ where $\mu(\Delta) < \infty$, then for p and q satisfying $1/C_2 \leqq p \leqq C_2$ and $1/C_3 \leqq q \leqq C_3$ on Δ for constants $C_2 > 1$ and $C_3 > 1$, Conditions A1, A2, A3 are satisfied. Condition A3 follows from the inequality

$$\sum_{n=0}^{\infty} ||| K_{t,n} ||| \delta^n \leqq \mu(\Delta)\, C_1 \sum_{n=0}^{\infty} \frac{(\delta \log C_1^2)^n}{n!}.$$

Important members of this class are finite Markov chains for which H_P and H_Q specify transition matrices composed of positive elements. In this case, Δ consists of a finite set of numbers, and μ is the measure assigning mass one to the elements of Δ and zero otherwise.

B. Many processes for which $\mu(\Delta) = \infty$ can be brought under the restrictions of Assumptions A1 through A3 by truncation. If there exists a subset Δ' of Δ and a constant $C > 1$ such that $1/C \leqq K_t(x, y) \leqq C$ on $\Delta' \times \Delta'$ and $0 < \mu(\Delta') < \infty$, then by defining $K_t'(x, y) = K_t(x, y)/\int_{\Delta'} K_t(x, y) d\mu(x)$ on $\Delta' \times \Delta'$ and zero on $(\Delta' \times \Delta')^c$ for $t = 0$ and 1, the process restricted to Δ' with transition densities $K_t'(x, y)$ and initial densities p' and q' constructed by truncating the original initial densities satisfies all three conditions.

The truncation of a process is reasonable in many instances. For example, if a Markov process with densities defined on a set of infinite measure is used to approximate a real system for which observations outside a bounded interval are physically impossible, the restriction of the process to this interval by truncation may yield a more appropriate model.

C. Let μ be Lebesgue measure, Δ any interval, and Δ' a subinterval of Δ for which $\mu(\Delta') < \infty$. Let ψ be a one-to-one mapping of Δ onto Δ' such that $\phi = \psi^{-1}$ has a derivative bounded away from zero on Δ a.s. (μ) by some positive constant. Then if the Markov process for which hypotheses H_P and H_Q specify $K_t(x, y)$ on $\Delta \times \Delta$ is replaced by a process for which new hypotheses $\tilde{H}_p$ and $\tilde{H}_q$ define $\tilde{K}_t(u, v) = K_t(\phi(u), \phi(v))\phi'(u)$ on $\Delta' \times \Delta'$ for $t = 0, 1$, and if Assumptions A1, A2, and A3 are satisfied by $\tilde{K}_t(u, v)$, $\tilde{p}(u) = p(\phi(u))\phi'(u)$ and $\tilde{q}(u) = q(\phi(u))\phi'(u)$, then all the results of Theorems 1 through 5 apply to the original process as well as its replacement. This follows from the fact that the probabilities $\tilde{P}[\tilde{R}_n > n^\alpha a]$ and $\tilde{Q}[\tilde{R}_n \leqq n^\alpha a]$ generated under $\tilde{H}_P$ and $\tilde{H}_Q$ are equal, respectively, to $P[R_n > n^\alpha a]$ and $Q[R_n \leqq n^\alpha a]$, defined in terms of the original process.

D. Another way of handling the case $\mu(\Delta) = \infty$ is to impose additional restrictions on the transition densities to compensate for the fact that $L_1(\Delta, \mu)$ no longer properly contains $L_2(\Delta, \mu)$ and that the function which is identically 1 is no longer integrable. The details of the theory have been carried out by Koopmans [4] for equal initial densities using slightly different methods, and the results coincide with those for $\mu(\Delta) < \infty$ except that it is not known whether the inequalities $\gamma_P \leqq \lambda'(0)$ and $\lambda'(1) \leqq \gamma_Q$ can always be strengthened to equalities as was done in Theorem 4.

8. Acknowledgments. The author is indebted to Professor David Blackwell for suggesting the problem and for his guidance during its solution and to Dr. Leo Breiman for many helpful suggestions and stimulating discussions. Thanks are also due to the referee for a number of clarifying remarks and for suggesting the simplified proof of Theorem 4.

REFERENCES

1. CHERNOFF, HERMAN, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 493–507.
2. CRAMÉR, H., "Sur un nouveau théorème-limite de la théorie des probabilités," *Actualités Scientifiqués et Industrielles*, No. 736, Paris, 1938.
3. HARDY, G. H., LITTLEWOOD, J. E., AND PÓLYA, G., *Inequalities*, Cambridge University Press, 1934.
4. KOOPMANS, L. H., "Asymptotic rate of discrimination for Markov processes," unpublished doctoral dissertation, University of California, Berkeley, 1958.
5. LOÉVE, MICHEL, *Probability Theory*, Van Nostrand, New York, 1955.
6. THOMASIAN, ARAM JOHN, "On the magnitude of the sum of error probabilities," unpublished doctoral dissertation, University of California, Berkeley, 1956.
7. TITCHMARSH, E. C., *The Theory of Functions*, 2nd ed., Oxford University Press, London, 1959.
8. WOLF, FRANTIŠEK, "Analytic perturbation of operators in Banach spaces," *Mathematical Annalen*, Vol. 124 (1952), pp. 317–333.
9. ZAANEN, ADRIAAN CORNELIS, *Linear Analysis*, North Holland Pub. Co., Amsterdam, 1953.