Stochastics, 1979, Vol. 2, pp. 139–153

Printed in Great Britain

Chernoff Bounds for Discriminating Between Two Markov Processes

C. M. NEWMAN†
Department of Mathematics, Indiana University, Bloomington, Indiana 47401

and

B. W. STUCK
Bell Laboratories, Murray Hill, New Jersey 07974

(Received March 3, 1978; in final form July 24, 1978)

We study a statistical hypothesis testing problem, where a sample function of a Markov process with one of two sets of known parameters is observed over a finite time interval. When a log likelihood ratio test is used to discriminate between the two sets of parameters, we give bounds on the probability of choosing an incorrect hypothesis, and on the total probability of error, for both discrete and continuous time parameter, and discrete and continuous state space. The asymptotic behavior of the bounds is examined as the observation interval becomes infinite.

1. INTRODUCTION

We are concerned with a classical problem in mathematical statistics, classifying a sample function of a random process, observed over a finite time interval, into one of two classes (called hypotheses). For a wide variety of cost criteria, it is well known, e.g. Grenander [7], that a minimum cost test is the so-called likelihood ratio test, where the likelihood functional is computed from the observations, and the result is compared with a threshold to choose one or the other hypothesis. Evaluating performance of the likelihood functional test involves analysis of the probability distribution of the likelihood functional, which is often the much more difficult part of the problem. Moreover, in many practical situations, it is well-known that a likelihood functional may be too

†Research supported in part by NSF Grant MPS 74–04870 A01

complicated and hence too expensive to compute, so suboptimum functionals are calculated which are much less costly to calculate; however, it is mandatory to compare the performance of the suboptimum test with that of the optimum test, and again this can lead to great analytical or computational difficulties.

We restrict attention to the special case where the observation is a Markov process whose probability distribution contains one of two sets of known parameters (this is called the sure signal in the literature, as opposed to the more general case which has attracted much interest over the last decade where the parameters themselves may be stochastic). When a likelihood functional test is used to discriminate between hypotheses, we give explicit new formulae for Chernoff-type bounds [1], [5], [8], [13] on the probability of choosing one hypothesis when the other is true, and on the total probability of error; this was previously an unsolved problem. We study the behavior of our bounds as the duration of the observation interval becomes infinite, and, under suitable conditions, show that all the probabilities of error decay exponentially, where the decay rate is governed by the largest eigenvalue of a particular operator that arises naturally from our approach. It is easy to construct examples where if one uses a suboptimum test for discriminating between the two hypotheses, the error probabilities need not decay exponentially but may decay rather much more slowly.

Some of the earlier work (see Evans [5] and the references therein) on this topic dealt mainly with Markov processes with continuous sample paths, i.e. diffusion processes. Here, we deal mainly with Markov jump processes, and treat diffusion processes separately. Skorokhod [15] has obtained sufficient conditions, equivalent to the hypotheses of Proposition 4.3 below, for the mutual nonorthogonality of the probability measures associated with two continuous time Markov processes; our results are clearly related to his, but were obtained by different methods, and appear to be much more useful in that they can be used to obtain upper and lower bounds on various error probabilities almost immediately. Finally, we are able to simplify, unify, and generalize much previous work on this topic by using a semi-group approach throughout. The extension of these results to time inhomogeneous Markov processes is straightforward, but is omitted for brevity.

2. MATHEMATICAL PRELIMINARIES AND PROBLEM STATEMENT

Let $x_1(t), x_2(t)$ be separable versions of two different stochastic processes taking values in a common state space, where t is time, $t \in E$, E a

bounded set. Time will be either discrete, $E = \{0,1,\ldots,N\}$, or continuous, $E = (0,T)$. ρ and ν denote the respective probability measures of the two processes on D, the space of right continuous paths with left hand limits everywhere defined. All probability measures are defined on the measure space (Ω, A), where Ω is the set of elementary events, and A is the σ algebra of Borel measurable subsets of Ω. The observation is denoted $r(t,\omega), t \in E, \omega \in \Omega$ where under hypothesis j, $r(t,\omega) = x_j(t,\omega)$. We use the Lebesque decomposition to write Ω as a disjoint union of measurable subsets $\{\Omega_\rho, \Omega_\nu, \Omega_{\rho\nu}\}$ such that $\rho(\Omega_\nu) = \nu(\Omega_\rho) \equiv 0$ and such that $\rho \approx \nu$ when restricted to $\Omega_{\rho\nu}$. The logarithm of the likelihood functional, $\Lambda = \Lambda[r: t \in E, \omega \in \Omega]$ may then be defined by

$$\Lambda[r:t \in E] = \begin{cases} +\infty & r \in \Omega_\rho \\ \dfrac{d\rho}{d\nu} & r \in \Omega_{\rho\nu} \\ -\infty & r \in \Omega_\nu \end{cases}$$

The test for discriminating between ρ and ν is to choose ρ if Λ is greater than or equal to a threshold L, and to choose ν otherwise.

A direct measure of the performance of this test is to calculate the probabilities of an error of the first and second kind, P_{12} and P_{21}, defined as

$$P_{12} = Pr[\text{choose } H_1 | H_2 \text{ true}] = \int_L^\infty d\mu[\Lambda | H_2]$$
$$P_{21} = Pr[\text{choose } H_2 | H_1 \text{ true}] = \int_{-\infty}^L d\mu[\Lambda | H_1]$$

where μ is the probability distribution of the log likelihood functional. An indirect measure of performance is the total probability of error, P_E, defined as

$$P_E = \pi_1 P_{21} + \pi_2 P_{12}$$

where π_j is the *a priori* probability that hypothesis j is true.

H. Chernoff [1] was apparently one of the first in the mathematical statistics literature to obtain the following simple upper bounds on the probabilities of an error of the first and second kind:

$$P_{21} \leqq \inf_{s<1} h_s(\rho,\nu) e^{-(s-1)L}$$
$$P_{12} \leqq \inf_{s>0} h_s(\rho,\nu) e^{-sL}$$

where $h_s(\rho,\nu)$ is a conditional moment generating function associated with the distribution of the log likelihood functional,

$$h_s(\rho,v)=E[e^{s\Lambda}|H_2]=E[e^{(s-1)\Lambda}|H_1]$$

An alternate expression for $h_s(\rho,v)$ is

$$h_s(\rho,v)=\int (d\rho/d\kappa)^s(dv/d\kappa)^{(1-s)}d\kappa \quad \rho,v\ll\kappa$$

which is independent of the choice of κ.

These ideas have been extended in several different directions. Shannon *et al.* [14] obtained lower bounds on the probabilities of an error of the first and second kind, for discrete measures ρ,v and showed their lower bounds and the Chernoff upper bounds are identical to within a factor that approaches unity as the observation becomes infinite.

It is well known, e.g. Kraft [10], Hellman and Raviv [8], that evaluating $h_{1/2}(\rho,v)$ yields simple, easy to calculate (but often crude) upper and lower bounds on the total probability of error,

$$\tfrac{1}{2}\min(\pi_1,\pi_2)[h_{1/2}(\rho,v)]^2\leq P_E\leq(\pi_1,\pi_2)^{1/2}h_{1/2}(\rho,v)$$

where the threshold L is chosen to be $\ln(\pi_2/\pi_1)$ in order to obtain the upper bound on P_E.

We emphasize that in many cases of practical interest, it is difficult or impossible to obtain closed form analytic expressions for P_{12}, P_{21}, P_E, and, furthermore, it is also quite difficult or expensive to accurately numerically approximate these quantities. The bounds just discussed are often possible to treat both analytically and numerically. Thus, we concentrate from this point only on explicitly calculating $h_s(\rho,v)$, which provide upper and lower bounds on P_E immediately, and can be further manipulated to obtain Chernoff-type bounds on P_{12}, and P_{21}. Although we are finding bounds on a likelihood ratio test, many of the same ideas developed here can be extended to bound performance of suboptimum tests, or, equivalently, to assess the performance of a suboptimum versus optimum test.

3 DISCRETE TIME, FINITE STATE MARKOV PROCESSES

Let x_1, x_2 be two time homogeneous Markov processes taking values in a finite state space $\{1,\ldots,M\}$ at discrete time instants $\{0,1,\ldots,N\}$. Let u,v be their initial distribution vectors. We note that $h_s(u,v)$ is given by

$$h_s(u,v)=\sum_{i=1}^{M} u_i^s v_i^{(1-s)}=H_s^T(u,v)e$$

where $H_s(u,v)_i=u_i^s v_i^{(1-s)} i=1,\ldots,M$ T denotes transpose and $e_i==1,\ldots,M$

Let R,Q denote the respective transition matrices of x_1,x_2

$$R_{ij}=Pr[x_1(k+1)=j|x_1(k)=i]$$

$$Q_{ij}=Pr[x_2(k+1)=j|x_2(k)=i]$$

$$\sum_{j=1}^{M}R_{ij}=\sum_{j=1}^{M}Q_{ij}=1 \quad 1\leq i,j\leq M, \quad 0\leq k\leq N$$

PROPOSITION 3.1 *Let x_1,x_2 be as just defined. Then*

a) *The log likelihood functional is given by*

$$\Lambda(r_0, \quad ,r_N)=\ln\frac{u_{j_0}}{v_{j_0}}+\sum_{k=0}^{N-1}\ln\frac{R_{j_kj_{k+1}}}{Q_{j_kj_{k+1}}}$$

where $j_k(1\leq j_k\leq M)$ is the state of $r_k, 0\leq k\leq N$.

b) *The moment generating function of the log likelihood functional is give by*

$$h_s(\rho,v)=H_s^T(u,v)[H_s(R,Q)]^N e$$

where

$$H_s(R,Q)_{ij}=[R_{ij}]^s[Q_{ij}]^{(1-s)} \quad \text{for} \quad 1\leq i,j\leq M$$

Proof (a) follows from the definition given in Section 2. To show (b), simply substitute into the definition:

$$\begin{aligned}h_s(\rho,v)&=\sum_{j_0=1}^{M}\cdots\sum_{j_N=1}^{M}(u_{j_0})^s(v_{j_0})^{(1-s)}(R_{j_0j_1})^s(Q_{j_0j_1})^{(1-s)}\\&\quad\times(R_{j_1j_2})^s(Q_{j_1j_2})^{(1-s)}\ldots(R_{j_{N-1}j_N})^s(Q_{j_{N-1}j_N})^{(1-s)}\\&=\sum_{j_0=1}^{M}\cdots\sum_{j_N=1}^{M}H_s(u,v)_{j0}H_s(R,Q)_{j_0j_1}\ldots H_s(R,Q)_{j_{N-1}j_N}\\&=H_s^T(u,v)[H_s(R,Q)]^N e \quad \text{Q.E.D.}\end{aligned}$$

We now study the role played by the eigenvalues and eigenvectors of $H_s(R,Q)$ in determining the behavior of $h_s(\rho,v)$. In particular, we find that typically $H_s(R,Q)$ has a unique largest positive real eigenvalue $r(H_s(R,Q))<1$, and that as the observation interval becomes infinite $(N\to\infty)$, $h_s(\rho,v)=O(r(H_s(R,Q))^N)$. Roughly speaking, $h_s(\rho,v)$ is a measure of the "overlap" or "similarity" of ρ,v, so that the closer $r(H_s(R,Q))$ is to unity, the greater the overlap or the greater the similarity of the two distributions, while the closer $r(H_s(R,Q))$ is to zero, the less the overlap or the less the similarity of the two distributions. Karlin [9] or Naylor and Sell [17] are suitable references for the material to follow.

The spectral radius of a matrix S with possibly complex eigenvalues $\{\lambda_i\}$ is denoted $r(S)$, where

$$r(S) = \max_{1 \leq i \leq M} \{|\lambda_i|\}$$

In the following lemma, we use the well-known fact that

$$r(S) = \limsup_{n \to \infty} \|S^n\|^{(1/n)} \leq \|S\|$$

for any suitable matrix norm $\|\cdot\|$.

LEMMA 3.2 *Given the preceding notation,*

$$\limsup_{N \to \infty} \frac{1}{N} \ln[h_s(\rho,\nu)] \leq \ln[r(H_s(R,Q))]$$

i.e.,

$$h_s(\rho,\nu) \leq O(r^N(H_s(R,Q))) \qquad N \to \infty$$

Proof Simply take norms on both sides of the expression for $h_s(\rho,\nu)$.

$$h_s(\rho,\nu)^{1/N} \leq [\|H_s(u,v)\| \, \|H_s^N(R,Q)\| \, \|e\|]^{1/N}$$

The desired result follows immediately upon taking logarithms plus the limit $N \to \infty$. Q.E.D.

PROPOSITION 3.3 *There exists at least one eigenvector w of $H_s(R,Q)$ with nonnegative entries and with eigenvalue $r(H_s(R,Q))$. If $w^T H_s(u,v) \neq 0$, then*

$$\lim_{N \to \infty} \frac{1}{N} \ln[h_s(\rho,\nu)] = \ln[r(H_s(R,Q))]$$

Proof The first statement follows from the classical Perron–Frobenius theorems for matrices with nonnegative entries. In particular, w has all nonnegative entries. This is used to write $e = cw + p$, where $c > 0$ and p has all nonnegative entries. This in turn implies that

$$\liminf_{N \to \infty} \frac{1}{N} \ln[H_s^T(u,v) H_s^N(R,Q) e]$$

$$\geq \lim_{N \to \infty} \frac{1}{N} \ln[H_s^T(u,v) c r^N(H_s(R,Q)) w] = \ln[r(H_s(R,Q))]$$

provided that $H_s^T(u,v) w \neq 0$, which together with the previous lemma completes the proof. Q.E.D.

To see that Lemma 3.2 may be a significant improvement over the simpler (often much easier to calculate) upper bound

$$h_s(\rho,\nu) \leq O(h_s(u,v) \|H_s(R,Q)\|^N$$

consider the following example:

Example 3.4 Let R and Q be two by two matrices

$$R = \begin{bmatrix} 1/2 & 1/2 \\ 9/10 & 1/10 \end{bmatrix} \quad Q = \begin{bmatrix} 1/2 & 1/2 \\ 1/10 & 9/10 \end{bmatrix}$$

where R,Q are considered to act on l_∞ for the calculation of operator norms. Then the spectral radius of $H_s(R,Q)$ is

$$r(H_s(R,Q)) = \frac{1}{4} + \frac{9^{1-s}}{20} \left[\left(\frac{1}{4} + \frac{9^{1-s}}{20} \right)^2 + \frac{9^s - 9^{1-s}}{20} \right]^{1/2}$$

e.g.

$$r(H_s(R,Q)) = 0.8 \quad \text{if } s = 1/2 \rightarrow h_s(\rho,\nu) \approx O(0.8^N \quad \text{if } s = 1/2$$

The much easier to calculate upper bound on the error probabilities is in fact useless here:

$$\|H_s(R,Q)\| = \sup_i \sum_j R_{ij}^s Q_{ij}^{1-s} =$$

i.e.

$$h_s(\rho,\nu) \leq O(1^N)$$

Variational methods can be used to numerically approximate the spectral radius of $H_s(R,Q)$.

4. CONTINUOUS TIME MARKOV PROCESSES

In this section, x_1, x_2 are time homogeneous Markov processes with common state space $\mathbf{R}^d$, and with time parameter in the interval $[0,T)$. Continuous time Markov processes have additional structure over discrete time Markov processes, because, very roughly speaking, we can write the transition probabilities of the two processes, which in discrete time were denoted R,Q, as exponentials of an operator which is called the infinitesimal generator of the transition probabilities. Our goal in this section is to take advantage of this additional structure. Suitable references are Dynkin [3,4], Feller [6].

The transition probabilities of the two processes are given by

$$Pr[x_1(s+t)\in A'|x_1(s) \qquad \int_{A'} R(t;x,dy$$
$$Pr[x_2(s+t)\in A|x_2(s) \qquad \int_A Q(t;x,dy)$$

$R(t)$ and $Q(t)$, when considered as operators, act on the space of bounded continuous functions. The infinitesimal generators associated with R and Q are denoted by A_1, A_2, where

$$A_1 f)(x \quad \lim_{t\to 0}\frac{1}{t}[R(t) \quad I]f(x$$

$$f)(x)=\lim_{t\to 0}\frac{1}{t}[Q(t \quad I]f(x)$$

where I is the identity operator. This allows us to identify the operator $R(t)$ with the operator $\exp(tA_1)$, and the operator $Q(t)$ with the operator $\exp(tA_2)$.

We first consider the case where the infinitesimal generators are bounded operators. This case is technically much easier to treat than the more general case of unbounded operators. Moreover, it is often reasonable to begin here in applications such as modeling signal detection in optical communication systems, or modeling detection of parameter changes in computer and data communications networks as well as in voice telephony traffic problems (see, e.g. Riordan [13]), for many such simple examples which lead to a great deal of insight into system behavior. This necessarily implies that x_1, x_2 are pure jump processes (with no drift and no diffusion component) with sample paths that are constant except for jumps occurring at random times with random amplitudes:

$$A_i f)(x) \qquad a_i(x,dz)[f(\qquad x)]$$
$$\text{with } \sup_x \int_{z\neq x} a_i(x,dz) < \infty$$

In words, the measure $a_i(x,dz)$ denotes the mean transition rate from state x to a state located in $(z,z+dz]$; e.g. for a simple Poisson process (which is often the first model investigated in optical communications or in data communications problems, in many instances) with mean rate parameter λ_i, $a_i(x,dz)=\lambda_i\delta(z-(x+1))dz$, where $\delta(.)$ is a Dirac delta function, because a simple Poisson process sample function contains only simple jump discontinuities of amplitude $+1$. We include finite state space or countable state space processes here as a special case.

In the statement and proof of the following proposition we denote by $H_s(u,v)$ for measures $u(dx)$, $v(dx)$, the measure $H_s(u,v)(dx) = (du/dw)^s(du/dw)^{1-s}w(dx)$ where $w(dx)$ is any measure such that $u,v \ll w$, and $H_s(u,v)$ is independent of the choice of w, with $h_s(u,v)=\int H_s(u,v)(dx)$. Similarly, we define the operator $H_s(R(t),Q(t))$ by

$$H_s(R(t),Q(t))f](x \quad \int_{R^n} H_s(R(\qquad Q(t;x,.))(dy)f(y$$

This n ation is consistent with that of Section 3

PROPOSITION 4.1 *Suppose A ,A_2 are bounded operators. Let $u(dx)$ be the distribution of $x_1(0)$ and (dx) be the distribution of $x_2(0)$. Define $a_3(x,dz)=H_s(a_1(x,.), a_2(x,.))(d$), and let $D_0(x)$ denote the function given b*

$$D_0(x \qquad sa_1(x,d \quad +(1 \quad)a_2(x,dz \quad a_3(x,dz)]$$

whi from Jensen's inequality is nonnegative. Then

$$h_s(\rho,v)=\int\int[H_s(u,v)(dx)\exp(T(A_3-D_0))(x,dy)$$

where

$$(A_3 f)(x \qquad {}_{\neq x} a \quad ,dz)[\qquad x)]$$

where D_0 in the last equation denotes the operator of multiplication by the function $D_0(x)$. Since A_3 can be thought of as the infinitesimal generator of a third Markov process, denoted $x_3(t)$ with initial distribution $H_s(u,v)/h_s(u,v)$, we can rewrite the moment generating function of the log likelihood functional as

$$h_s(\rho,v \quad h_s(u,v)E_{x_3}[\exp\{ \quad \int_0^T D_0(x_3)(t)dt\}$$

Proof For each $n=1,2,\ldots$, we approximate $x_i(t)$ by the discrete time process $\{x_i(kT/n): k=0,1,\ldots, n-1\}$. Letting ρ_n respectively v_n denote the probability distribution of $\{x_1(kT/n)\}$, respectively $\{x_2(kT/n)\}$, then by a calculation completely analogous to the one carried out in the proof of Proposition 3.1 above, one has

$$h_s(\rho_n v_n = \int H_s(u,v)(dx)[H_s(R(\quad /n),Q(T/n))]^n e)(x)$$

where $e(x)$ is the function identically equal to one for all x. Since

$$h_s(\rho, \qquad \lim h_s(\rho_n,v_n),$$

suffices to show that

$$\lim_{n\to\infty}[H_s(R(T/n),Q(T/n))]^n=\exp[T(A_3-D_0)]$$

in order to prove the proposition (the last equation of the proposition follows from the usual Feynman–Kac formula). To obtain the last equation, we write $A_i=A_i^0+A_i^D$, where A_i^0 is an off-diagonal coupling or interaction operator,

$$A_i^0 f(x)=\int_{z\neq x} a_i(x,dz)f(z)$$

and A_i^D is the diagonal operator of multiplication by $-\int_{z\neq x} a_i(x,dz$ Then use the fact that

$$R(t/n)=I+(t/n)A_i^0+(t/n)A_i^D+o(t/n)$$

The desired result now follows easily from observing that

$$H_s(I+\tau A_1^D+\tau A_1^0+o(\tau),I+\tau A_2^D+\tau A_2^0+o(\tau))$$
$$=I+\tau H_s(A_1^0,A_2^0+\tau(sA_1^D+(1-s)A_2^D)+o(\tau)$$
$$=I+\tau A_3-\tau D_0+o(\tau$$

since

$$(I+(T/n)(A_3-D_0)+o(T/n))^n\to\exp[T(A_3-D_0)].\quad \text{Q.E.D.}$$

Example 4.2 In many optical communications, data communications, or telephone traffic problems, one of the simplest class of models that is often first investigated is the class of birth-death processes because it is analytically tractable and because it suggests behavior of more complicated models (e.g. the simple Poisson process is a pure birth process). Here we deal with a class of generalized birth-death models, where x_1,x_2 have finite state space $\{1,\ldots,M\}$ and infinitesimal generators

$$(A_i)_{jk}=(a^i)_{jk}\quad i=1,2\qquad \leq j,k\leq M$$

so that for $j\neq k$

$$Pr[x_i(t+\Delta)=k\,|\,x_i(t)=j]=a_{jk}^i\Delta+o(\Delta).$$

Then we see that

$$(A_3-D_0)_{jk}=\begin{cases}(a_{jk}^1)^s(a_{jk}^2)^{1-s} & j\neq k\\ -\sum_{l\neq j} sa_{jl}^1+(1-s)a_{jl}^2 & j=k\end{cases}$$

As we will see below, the asymptotic behavior of $h_s(\rho,v)$ for large time is obtained directly from an analysis of the spectrum of the matrix A_3-D_0 with $h_s\approx 0(\exp(-\kappa T))$ for an appropriate κ determined from A_3-D_0.

The calculation for the log likelihood moment generating functional provides a new set of explicit examples of the Trotter product formula [16], and is a special case of more general limit theorems of P. Chernoff [2] which are now needed here for the case of unbounded operators (which are encountered in pure drift or diffusion processes).

In order to discuss the case of unbounded generators, we introduce a series of technical hypotheses. In a wide variety of practical applications it will be straightforward to check that these conditions hold (see, e.g. the examples 4.5, 4.6, 4.7).

In order to handle pathologies that arise when $x_k(t)$ has an infinite number of jumps in any finite time interval (with "most" of the jumps of vanishingly small amplitude), we assume $x_k(t)$ can be obtained as a weak limit of processes $x_k^\varepsilon(t)$, $\varepsilon\to 0$, with no jumps of magnitude less than ε. The infinitesimal generators are

[illegible]

$$+\int_{\|z-x\|\geq\varepsilon} a_k(x,dz)[f(z)-f(x)\quad \frac{(z-x)^T}{+\|z-x\|^2}\nabla f(x)]$$

Moreover, we make the usual, e.g. Skorokhod [15], Dynkin [3], [4], continuity assumptions about δ_k, σ_{lm}^k, and a_k, plus one additional assumption:

$$\sup_x\int_{\|z-x\|\geq\varepsilon} a_k(x,dz)<\infty\quad \text{for each } \varepsilon>0$$

In particular, we are assuming that the corresponding operators $R(\varepsilon,t),Q(\varepsilon,t)$ converge weakly to $R(t),Q(t)$ respectively as $\varepsilon\to 0$.

The structure of the class of processes we consider is quite rich, in that it includes all independent increment processes (both the Wiener process and the superposition of independent Poisson processes with different rates and jump amplitudes), as well as processes that are causal functionals of independent increment processes. Recall that (i) δ_k controls the drift or centering of the process, (ii) σ^k regulates the diffusion process infinitesimal variance, i.e. that component of the process with almost surely continuous sample paths, and (iii) a_k governs the jump process infinitesimal transition rate, i.e. that component of the process with almost surely discontinuous sample paths, with simple jump discontinuities of random amplitudes occurring at random times.

As above, we define a new time homogeneous Markov process x_3 with state space $\mathbf{R}^d$, $t\in[0,T)$, initial distribution $H_s(u,v)/h_s(u,v)$ and infinitesimal generator A_3 defined when conditions (a), (b), (c) in Proposition 4.3 are met via

$$a_3(x,d \quad H_s(a_1,a_2)(x,dz \quad \sigma_{ij}^2(z \quad \sigma_{ij}^1(z \quad \sigma_{ij}^2(z$$

$$)\delta_2(z$$

$$(z,du)+ \quad s)a_2(z,du \quad a_3(z,du)] \quad \frac{z-u}{+\|z \quad u\|^2}$$

PROPOSITION 4.3 *Suppose the following conditions are satisfied*

a) *$D_0(x)$ is finite or all x, i.e*

$$D \quad x)\equiv\lim \quad sa_1(x,dy)+(1 \quad)a_2(x,dy \quad a_3(x,dy)]$$

b) *The infinitesimal variance of the diffusion process compone are identical*

$$\sigma_{ij}^1(x)= \quad \sigma_{ii}(x \quad M \quad \text{for all } x \quad R^d$$

c) *The infinitesimal drifts differ by δ which lies in the range of σ, for all ∈ R^d where*

$$\delta(x \quad \delta \quad \delta_2(x \quad \lim_{\varepsilon\to 0}\int \quad + \quad [a_1(x,dz \quad a_2(x,dz)]$$

The og likelihood moment generating functional is then given by

$$h_s(\rho \quad =h_s(u,v)E_{x_3}[\exp(-\int_0^T D(x_3(t))dt)]$$
$$=\int\int H_s(u,v)(dx)\exp[T(A_3-D)](x,dy$$

where

$$D(x \quad D_0(x \quad s(1 \quad \delta^T(x)\sigma \quad x)\delta(x)$$

denote the matrix inverse of σ(x

Proof The proof of this proposition is based upon a direct combination of Proposition 4.1 together with the techniques and results o Newman [11], [12]. We do not include the details for the sake of brevity.

The following result is a consequence of the definition of the spectra radius of the operator $\exp(A_3-D)$, and the proof of Proposition 3.3:

PROPOSITION

$$\limsup_{T\to\infty}\frac{}{T} \quad h_s(\rho,v)] \quad \ln[\lim\|\exp(\quad D))\|$$

If $_3-D$ has a nonnegati eigenfunction w(x)∈ eigenvalue equa to κ, and if

$$w(x)H_s(u,v)(dx)\neq 0$$

then

$$\limsup_{T\to\infty}\frac{}{T}\ln[h_s(\rho,v)] \quad \sup_{w(x)}\mathrm{Re}\int w(x)[(A \quad D)w](x)dx$$

where the supremum is aken over all w in the domain of D with L_2 norm unity.

Example 4.5 In the case where x_1,x_2 are independent increment processes, the infinitesimal drift, variance, and jump measure are all independent of the present state x, so that $D(x)$ is also a constant D independent of x, and

$$h_s(\rho,v)=h_s(u,v)\exp(\quad \kappa T$$

$$\delta^T\sigma \quad ^1\delta+\int_{y\neq 0}[sa_1(dy)+(1 \quad s)a_2(dy)+H_s(a_1,a_2)(dy)]$$

We note that in this case h_s has an exact exponential dependence on T moreover, ρ and v are mutually orthogonal unless $\delta\in\mathrm{range}(\sigma)$ and

$$\int d[sa \quad (1 \quad)a_2 \quad H_s(a_1,a_2)]<$$

(see 12])

Example 4.6 Suppose $x_k(t)$ is a real valued stochastic process satisfying the stochastic differential equation

$$dx_k(t)=db(t \quad \delta_k(x)dt \quad x_k(0)=0 \quad \text{a.s}$$

with $\delta_1=-\delta_2=\delta(x)$ for some given function $\delta(x)$, where $b(t)$ denote standard Brownian motion. The infinitesimal generator is given by

$$A_k \quad \frac{1}{2}\frac{d^2}{dx} \quad \delta_k(x)\frac{a}{d}$$

and the problem is to discriminate between the two possible drift directions of the state dependent diffusion processes. We set $s=\frac{1}{2}$ and note that

$$A_3 = \text{[illegible]}$$

so that

$$x_3(t)=b(t) \quad h_s(\rho,v)=E[\exp(-\tfrac{1}{2}\textstyle\int_0^T \delta^2(b(t))dt)]$$

For simplicity, we assume that $\delta(x)\to\infty$ as $|x|\to\infty$ so that A_3-D has a discrete spectrum, $h_s(\rho,v)\approx O(\exp(-\kappa T))$, and the value of κ must be found numerically in general. In particular, the assumption $\delta(x)\to\infty$ as $|x|\to\infty$ must be checked because the model is probably not valid over the range $x\in(-\infty,+\infty)$.

Example 4.7 We close by illustrating the power of an optimum log likelihood ratio test over a plausible suboptimum test. Suppose the observations are independent identically distributed random variables. Suppose each observation $r_k, 0\leq k\leq N$ is drawn from a distribution with one of two sets of scaling parameters, γ_i, $i=1,2$, where the characteristic function of x_i, $i=1,2$ is

$$E[\exp(i\omega x_j)]=\exp(-\gamma_j|\omega|^\alpha)$$

For $\alpha=2$ the likelihood ratio test involves computing a χ^2 statistic and comparing the result with a threshold to discriminate between distributions. What if the same test and threshold is used for $0<\alpha<2$? Since for $0<\alpha<2$ the distribution of x_j can be shown to behave asymptotically as $O(|X|^{-\alpha})$, elementary arguments (see Feller (1970, p. 272)) allow us to show that the probability density function of the χ^2 statistic behaves asymptotically like $O(|X|^{-(1+\alpha/2)})$, and hence P_{12}, P_{21}, and P_E decay as $O(N^{(1-\alpha/2)})$ as $N\to\infty$, while the arguments given previously show that the likelihood ratio test error probabilities decay exponentially with $N(O(r^N(H_s)), 0<r<1)$. The physical reason for this is clear: for $0<\alpha<2$ the χ^2 statistic is dominated by one or a few large excursions, while the likelihood ratio test essentially weights large excursions quite lightly. For γ_1 and γ_2 differing by a factor of ten or more, machine calculations show the asymptotic results $(N\to\infty)$ hold for $N\geq 5$.

References

[1] H. Chernoff, A measure of asymptotic efficiency for tests of a hypothesis based on a sum of observations, *Ann. Math. Stat.* **23**, 493–507 (1952).

[2] P. R. Chernoff, Note on product formulas for operator semigroups, *J. Functional Anal.* **2**, 238–242 (1968).

[3] E. B. Dynkin, Markov processes and semigroups of operators, *Thy. Prob. Appl.* **1**, 22–33 (1956a).

[4] E. B. Dynkin, Infinitesimal operators of Markov processes. *Thy. Prob. Appl.* **1**, 34–54 (1956b).

[5] J. E. Evans, Chernoff bounds on the error probability for the detection of non-Gaussian signals, *IEEE Trans. Info. Thy.* **20**, 569–577 (1974).

[6] W. Feller, *An Introduction to Probability Theory and Its Applications*, Volume II, 275–284, 574–581, Second Edition, Wiley, N.Y., 1971.

[7] U. Grenander, Stochastic processes and statistical inference, *Arkiv. För Mathematik* **1** (17), 195–277 (1950).

[8] M. E. Hellman and J. Raviv, Probability of error, equivocation, and the Chernoff bound, *IEEE Trans. Info. Thy.* **16**, 368–372 (1970).

[9] S. Karlin, Positive operators, *J. Math. Mech.* **8**, 907–938 (1959).

[10] C. H. Kraft, Some conditions for consistency and uniform consistency of statistical procedures, *Univ. California at Berkeley Pub. Statistics* **2**, (6), 125–142 (1955).

[11] C. M. Newman, The inner product of path space measures corresponding to random processes with independent increments, *Bull. Am. Math. Soc.* **78**, 268–272 (1972).

[12] C. M. Newman, The orthogonality of independent increment processes, in *Topics in Probability Theory*, D. W. Stroock, S. R. S. Varadhan (eds.), pp. 93–111, Courant Institute of Mathematical Sciences, N.Y.U., 1973.

[13] J. Riordan, *Stochastic Service Systems*, Wiley, N.Y., 1962.

[14] C. E. Shannon, R. G. Gallager and E. R. Berlekamp, Lower bounds to error probability for coding on discrete memoryless channels, I, *Info. Control* **10**, 65–103 (1967).

[15] A. V. Skorokhod, On the differentiability of measures which correspond to stochastic processes, II: Markov processes, *Thy. Prob. Appl.* **5**, 40–49 (1960).

[16] H. F. Trotter, On the product of semigroups of operators, *Proc. Am. Math. Soc.* **10**, 545–551 (1959).

[17] A. W. Naylor and G. R. Sell, *Linear Operator Theory in Engineering and Science*, Chapter 6, Holt, Rinehart, Winston, N.Y., 1971.