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
Let $x(t)$ be separable versions of two different stochastic pro-
We now study the role played by the eigenvalues and eigenvectors of
\[ (\mathbf{D}(\mathbf{a})_H^* \mathbf{a})_H \]
and the distribution of \( \mathbf{D}(\mathbf{a})_H^* \mathbf{a} \) in determining the behavior of \( \mathbf{D}(\mathbf{a})_H^* \mathbf{a} \). In particular, we find that the

3 DISCRETE TIME FINITE STATE MARKOV PROCESSES

An alternative expression for \( \mathbf{H} \) given by
\[ \mathbf{H} = \sum_{n=0}^{\infty} \mathbf{a}_n \]

Let \( \mathbf{H} \) be two time homogeneous Markov processes taking values in a

CHEBYSHEF BOUNDS AND MARKOV PROCESSES

\[ \mathbf{H} = \sum_{n=0}^{\infty} \mathbf{a}_n \]

where

\[ \mathbf{H} = \sum_{n=0}^{\infty} \mathbf{a}_n \]

and that the distribution of \( \mathbf{D}(\mathbf{a})_H^* \mathbf{a} \) is a measure of the

ROBUSTLY SPEAKING, \( \mathbf{a} = (\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n) \) is a unique largest positive real eigenvalue

\[ (\mathbf{D}(\mathbf{a})_H^* \mathbf{a})_H \]

implies that \( \mathbf{D}(\mathbf{a})_H^* \mathbf{a} \) is the distribution of \( \mathbf{D}(\mathbf{a})_H^* \mathbf{a} \).
In this section, we introduce continuous-time Markov processes with a common state space $\mathcal{S}$ and with time parameter in the interval $[0,t]$. Here, $\mathcal{S}^*$ are time homogeneous Markov processes, which are a class of stochastic processes with the property that the transition probabilities between states are independent of time.

**Proposition 3.3** There exists at least one generator $\mathcal{Q}$ of $\mathcal{S}^*$ that maps nonnegative functions into nonnegative functions, and with eigenvalue $\mathcal{Q}$ of $\mathcal{S}^*$ has all the desired results for the class $\mathcal{Q}$ of functions. In particular, $\mathcal{Q}$ has all the desired results for the class $\mathcal{Q}$ of functions.

**Proof** The first statement follows from the classical Perron-Frobenius theory.

\[
\lim_{t \to \infty} \mathcal{S}^* \mathcal{Q} = \mathcal{Q}
\]

then $\mathcal{Q} = \mathcal{Q}^*$. O.F.D.

The desired results immediately follow by argument for $\mathcal{Q}^*$.

\[
\mathcal{Q}^* \mathcal{Q}^* = \mathcal{Q}^* \mathcal{Q} = \mathcal{Q}
\]

Example 3.4 Let $A$ and $B$ be two two-matrix

\[
\begin{bmatrix}
10/1 & 10/1 & \mathcal{Q}
\end{bmatrix}
\]

For any suitable matrix norm

\[
\|\mathcal{Q}\| \leq \lambda_{\text{max}}(\mathcal{Q})
\]

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

\[
\mathcal{Q}^* \mathcal{Q}^* = \mathcal{Q}^* \mathcal{Q} = \mathcal{Q}
\]

Consider the following expression:

\[
\|\mathcal{Q}\| \leq \lambda_{\text{max}}(\mathcal{Q})
\]

The spectral radius of a matrix $\mathcal{Q}$ with possibly complex eigenvalues is the largest absolute value of an eigenvalue. To see that Lemma 3.2 may be a significant improvement over the simple (often much easier to calculate) upper bound provided by the previous lemma, we use the well-known fact that

\[
\|\mathcal{Q}\| \leq \lambda_{\text{max}}(\mathcal{Q})
\]
The space process here as a special case of a simple process arising from a simple jump. We included finite state space or countable state space and a simple jump process simple function containing only simple jump functions in communication problems. In many cases, this property is true because they are fundamentally the result of a simple process, with the state transition of the system that is different from the model involved in the problem. Of course, the time process is different from the problem which is defined by the set of state transition rates from state to state.

\[ x \rightarrow (zp'x)^{a(x'x)} \]

\[ \mu(x') = (zp'x)^{b(x'x)} \]

With the measure on the map (xp'x)^{a(x'x)}

\[ (x')^b(x'x) \]

\[ \mu(x') = (zp'x)^{b(x'x)} \]

Since the measure function of the time process depends on the state transition rates from state to state, the measure function depends on the state transition rates from state to state.

The measure function depends on the state transition rates from state to state.
random amplitude occurring at random times.

In particular, we are assuming that the corresponding operations are

\[ (x) \frac{f(x) - f(x)}{x} \| \frac{x - \bar{x}}{x} \| = -1 \]

Moreover, we make the usual \( \epsilon \)-Stochastic diffusion assumptions about \( \| \phi \| \) and \( \phi \), plus one additional

\[ \left( x, f(x) \right) \| \bar{x} \| = -1 \]

The conclusion of the previous section is explicitly that, for example, the

\[ (x) \frac{f(x) - f(x)}{x} \| \frac{x - \bar{x}}{x} \| = -1 \]

are both \( \sigma \)-measurable functions of \( x \).

In order to handle pathologic cases where \( x \) is a weak

\[ (x) \frac{f(x) - f(x)}{x} \| \frac{x - \bar{x}}{x} \| = -1 \]

In order to discuss the case of unbounded generators, we introduce a

\[ (x) \frac{f(x) - f(x)}{x} \| \frac{x - \bar{x}}{x} \| = -1 \]

which are known as the case of diffusion processes.

\[ (x) \frac{f(x) - f(x)}{x} \| \frac{x - \bar{x}}{x} \| = -1 \]

We will see below the asymptotic behavior of \( \epsilon \)-diffusion processes.

\[ (x) \frac{f(x) - f(x)}{x} \| \frac{x - \bar{x}}{x} \| = -1 \]

Then we see that

\[ (x) \frac{f(x) - f(x)}{x} \| \frac{x - \bar{x}}{x} \| = -1 \]

so that for

\[ (x) \frac{f(x) - f(x)}{x} \| \frac{x - \bar{x}}{x} \| = -1 \]

have the same space \( \epsilon \)-diffusion processes and \( \epsilon \)-Stochastic diffusion

where we deal with a class of \( \epsilon \)-diffusion processes, where \( x \) is

\[ (x) \frac{f(x) - f(x)}{x} \| \frac{x - \bar{x}}{x} \| = -1 \]

In order to prove the proposition, we use the following trick.

\[ (x) \frac{f(x) - f(x)}{x} \| \frac{x - \bar{x}}{x} \| = -1 \]

Then \( (x) \frac{f(x) - f(x)}{x} \| \frac{x - \bar{x}}{x} \| = -1 \)

The desired result follows easily from observing that

\[ (x) \frac{f(x) - f(x)}{x} \| \frac{x - \bar{x}}{x} \| = -1 \]

and \( \psi \) is the diagonal operator of multiplication by \( f(x) \frac{f(x) - f(x)}{x} \| \frac{x - \bar{x}}{x} \| = -1 \)

Theorem. -- [16] --

The conclusion of the proposition, the last equation of the proposition

\[ (x) \frac{f(x) - f(x)}{x} \| \frac{x - \bar{x}}{x} \| = -1 \]

in order to prove the proposition (the last equation of the proposition

\[ (x) \frac{f(x) - f(x)}{x} \| \frac{x - \bar{x}}{x} \| = -1 \]

is in an obvious category of

\[ (x) \frac{f(x) - f(x)}{x} \| \frac{x - \bar{x}}{x} \| = -1 \]

in a natural category of

\[ (x) \frac{f(x) - f(x)}{x} \| \frac{x - \bar{x}}{x} \| = -1 \]

Theorem. -- [16] --

The conclusion of the proposition, the last equation of the proposition

\[ (x) \frac{f(x) - f(x)}{x} \| \frac{x - \bar{x}}{x} \| = -1 \]

in order to prove the proposition (the last equation of the proposition

\[ (x) \frac{f(x) - f(x)}{x} \| \frac{x - \bar{x}}{x} \| = -1 \]

is in an obvious category of

\[ (x) \frac{f(x) - f(x)}{x} \| \frac{x - \bar{x}}{x} \| = -1 \]

in a natural category of

\[ (x) \frac{f(x) - f(x)}{x} \| \frac{x - \bar{x}}{x} \| = -1 \]
The stochastic differential equation

\[ \frac{d}{dt} \mathcal{L} \mathcal{g}(x) = \mathcal{L} \mathcal{g}(x) \]

is called a real valued stochastic process satisfying

\[ \mathcal{L} \mathcal{g}(x) \]

and are mutually orthogonal with respect to the measure \( \mathcal{L} \mathcal{g}(x) \) and \( \mathcal{L} \mathcal{g}(x) \).

Moreover, if we define the generator \( \mathcal{L} \mathcal{g}(x) \) in the following way,

\[ \mathcal{L} \mathcal{g}(x) \]

and \( \mathcal{L} \mathcal{g}(x) \), then the generator \( \mathcal{L} \mathcal{g}(x) \) is defined as the differential operator

\[ \mathcal{L} \mathcal{g}(x) = \frac{d}{dt} \mathcal{L} \mathcal{g}(x) \]

and the generator \( \mathcal{L} \mathcal{g}(x) \) is given by

\[ \mathcal{L} \mathcal{g}(x) = \frac{d}{dt} \mathcal{L} \mathcal{g}(x) \]

with \( \mathcal{L} \mathcal{g}(x) = \frac{d}{dt} \mathcal{L} \mathcal{g}(x) \).

\[ \frac{d}{dt} \mathcal{L} \mathcal{g}(x) = \mathcal{L} \mathcal{g}(x) \]

for some given function \( \mathcal{L} \mathcal{g}(x) \).

\[ \mathcal{L} \mathcal{g}(x) = \frac{d}{dt} \mathcal{L} \mathcal{g}(x) \]

for some given function \( \mathcal{L} \mathcal{g}(x) \).

\[ \mathcal{L} \mathcal{g}(x) = \frac{d}{dt} \mathcal{L} \mathcal{g}(x) \]

for some given function \( \mathcal{L} \mathcal{g}(x) \).

\[ \mathcal{L} \mathcal{g}(x) = \frac{d}{dt} \mathcal{L} \mathcal{g}(x) \]

for some given function \( \mathcal{L} \mathcal{g}(x) \).

\[ \mathcal{L} \mathcal{g}(x) = \frac{d}{dt} \mathcal{L} \mathcal{g}(x) \]

for some given function \( \mathcal{L} \mathcal{g}(x) \).

\[ \mathcal{L} \mathcal{g}(x) = \frac{d}{dt} \mathcal{L} \mathcal{g}(x) \]

for some given function \( \mathcal{L} \mathcal{g}(x) \).

\[ \mathcal{L} \mathcal{g}(x) = \frac{d}{dt} \mathcal{L} \mathcal{g}(x) \]

for some given function \( \mathcal{L} \mathcal{g}(x) \).

\[ \mathcal{L} \mathcal{g}(x) = \frac{d}{dt} \mathcal{L} \mathcal{g}(x) \]

for some given function \( \mathcal{L} \mathcal{g}(x) \).

\[ \mathcal{L} \mathcal{g}(x) = \frac{d}{dt} \mathcal{L} \mathcal{g}(x) \]

for some given function \( \mathcal{L} \mathcal{g}(x) \).

\[ \mathcal{L} \mathcal{g}(x) = \frac{d}{dt} \mathcal{L} \mathcal{g}(x) \]

for some given function \( \mathcal{L} \mathcal{g}(x) \).
The asymptotic results \((N_f) \rightarrow \infty\) hold for \(\frac{z}{\sqrt{\lambda}} + \frac{s}{\sqrt{\lambda}} \), and \(z\) diverges by a factor of \(\sqrt{\lambda}\). The machine calculations show that the \(z\) statistic is dominated by one or few large observations while the true \(z\) statistic converges to the theoretical limit. For \(N_f \gg 1\), the physical reason for this is clear: for \(0 < \lambda < 1\), the limit distribution \(x^0\) and \(x^1\) decay as \(\lambda \rightarrow 0\), and hence the limit distribution \(x^\lambda\) is obtained by setting \(\lambda = 0\). Since \(\lambda \leq \frac{1}{\sqrt{N_f}}\), the distribution of \(x^\lambda\) can be shown to converge asymptotically to the limit distribution.

For \(N_f = 2\), the likelihood ratio tests involve computing \(x^\lambda\) statistic and

\[
E_{(x^\lambda)} = \int x^\lambda dx = e^{-\lambda} \cdot \Gamma(\lambda + 1)
\]

where \(\Gamma(\lambda + 1)\) is the Gamma function of \(\lambda + 1\). The distribution of \(x^\lambda\) is obtained as a mixture of two distributions.

Example 4.7: We close by mentioning two useful expansions in optimisation.

\[
x(\epsilon) = \lim_{\epsilon \to 0} x(\epsilon) \quad \text{where the derivatives}
\]

\[
\frac{d}{d\epsilon} x(\epsilon) \bigg|_{\epsilon = 0} = \frac{d}{d\epsilon} \lim_{\epsilon \to 0} x(\epsilon) = \lim_{\epsilon \to 0} \frac{d}{d\epsilon} x(\epsilon)
\]

must be calculated because the model is properly not valid over the range

\[
0 \to x \to \infty
\]

and the value of \(x\) must be found numerically in general. In particular, the assumption \(\frac{d}{d\epsilon} x(\epsilon) \bigg|_{\epsilon = 0} \neq 0\), and the value of \(x\) must be found numerically.

For simplicity we assume that \(\alpha = 0\) as \(\lim_{\epsilon \to 0} \frac{d}{d\epsilon} x(\epsilon) \bigg|_{\epsilon = 0} = 0\)

so that

\[
\frac{\partial}{\partial \alpha} x(\alpha) = (\alpha + 1) x(\alpha) = \frac{\partial}{\partial \alpha} \left(\frac{\partial}{\partial \alpha} x(\alpha) \right) = 0
\]

that is, the solution of the state dependent diffusion processes. We set \(\alpha = 1\) and once