TESTING HYPOTHESES FOR MARKOV CHAINS WHEN THE PARAMETER SPACE IS FINITE

BY KEN-ICHI YOSHIHARA

1. Summary.

In Chernoff [2], a procedure was presented for the sequential design of experiments where the problem was one of testing a hypothesis. When there were only a finite number of states of nature and a finite number of available experiments, the procedure was shown to be "asymptotically optimal" as the cost of sampling approached zero. An analogous procedure can be applied to the problem of testing a hypothesis with respect to a Markov process, and this procedure will also be shown to be "asymptotically optimal".

2. Assumptions.

Let Θ be a parameter space which consists of finite elements. We shall test the hypothesis $\theta \in H_1$ against the alternative $\theta \in H_2$ where H_1 and H_2 are two non-null disjoint subsets of Θ. In what follows, we make the following assumptions.

A1. Θ is a finite set. H_1 and H_2 are two non-null subsets of Θ and $H_1 \cup H_2 = \Theta$ and $H_1 \cap H_2 = \phi$.

A2. $\{X_k: k=0, 1, 2, \cdots\}$ is a Markov chain with state space $(\mathfrak{X}, \mathfrak{A})$ and with stationary transition measures $p_\theta(\xi, \cdot)$ $(\theta \in \Theta)$ which satisfy the following conditions:

(a) For each $\theta \in \Theta$, the transition measures $p_\theta(\xi, \cdot)$ satisfy Doeblin's condition (D) in Doob [3], and there exists only one ergodic set and the transient set is empty;

(b) For each $\theta \in \Theta$, the transition measures $p_\theta(\xi, \cdot)$ admit of a unique stationary probability measure $p_\theta(\cdot)$.

In what follows $E_\theta(\cdot)$ will denote an expected value computed under the assumption that θ is true and that $p_\theta(\cdot)$ is the initial distribution.

A3. There is a measure λ on $\mathfrak{A}$, not necessarily finite, with respect to which all the transition measures $p_\theta(\xi, \cdot)$ have densities $f(\xi, \eta: \theta)$ and the initial distribution $p_\theta(\cdot)$ has a density $f(\xi: \theta)$ for each θ. These densities satisfy the following conditions:

(a) For each $\theta, f(\xi, \eta: \theta)$ is measurable in ξ and η, and $f(\xi: \theta)$ is measurable in ξ;

(b) If θ, φ are in Θ and $\theta \neq \varphi$, then

$$P_\theta\{f(X_0, X_1: \theta) \neq f(X_0, X_1: \varphi) \mid X_0 = \xi\} > 0 \qquad \text{for almost all } \xi.$$

Received March 15, 1963.

A 4. There exists a positive number $\delta(<1)$ such that for any $t \in (0, \delta)$ and for any pair θ, φ in Θ

$$\sup_{\xi \in \mathfrak{X}} E_\theta \left\{ \exp\left(t \log \frac{f(X_{k-1}, X_k: \theta)}{f(X_{k-1}, X_k: \varphi)} \right) \middle| X_{k-1} = \xi \right\} < \infty.$$

We remark that, by A 4 (b), if θ, φ are in Θ and $\theta \neq \varphi$, then

$$(1) \qquad E_\theta \left\{ \log \frac{f(X_{k-1}, X_k: \varphi)}{f(X_{k-1}, X_k: \theta)} \right\} = -I(\theta, \varphi) < 0 \qquad k = 1, 2, \cdots)$$

where

$$(2) \qquad I(\theta, \varphi) = E_\theta \left\{ \log \frac{f(X_{k-1}, X_k: \theta)}{f(X_{k-1}, X_k: \varphi)} \right\}.$$

3. "Asymptotically optimal" procedure.

Define

$$Z_0(\theta, \varphi) = \log \frac{f(X_0: \theta)}{f(X_0: \varphi)},$$

$$S_n(\theta, \varphi) = \log \frac{f(X_{j-1}, X_j: \theta)}{f(X_{j-1}, X_j: \varphi)}, \qquad j = 1, 2, 3, \cdots,$$

$$S_n(\theta, \varphi) = \sum_{j=0}^{n} Z_j(\theta, \varphi)$$

and

$$S_n = \sum_{j=0}^{n} Z_j(\hat{\theta}_n, \tilde{\theta}_n)$$

where $\hat{\theta}_n$ is a maximum likelihood estimate of θ under Θ, based on the first $n+1$ observations and $\tilde{\theta}_n$ is a maximum likelihood estimate of θ under the hypothesis alternative to $\hat{\theta}_n$.

We define "procedure A" as follows. Stop sampling at the $n+1$-st observation and select the hypothesis $\hat{\theta}_n$ if $S_n > -\log c$, where $0 < c < 1$.

4. Bounds for $E_\theta\{e^{-tS_n(\theta, \varphi)}\}$ and $E_\theta\{e^{tS_n(\theta, \varphi)}\}$.

At first we shall prove the theorem which is needed later.

THEOREM 1. *For any pair θ and φ in Θ and any $\varepsilon > 0$, there exist two positive numbers $A = A(\varepsilon)$ and $B = B(\varepsilon)$ such that, for all $n(\geqq 1)$ and for sufficiently small $t > 0$,*

$$(3) \qquad E_\theta\{e^{-tS_n(\theta, \varphi)}\} \leqq A e^{-nt[I(\theta, \varphi)/(1+\varepsilon) - O(t)]}$$

and

$$E_\theta\{e^{tS_n(\theta,\varphi)}\} \leqq B e^{nt[I(\theta,\varphi)/(1-\varepsilon)+O(t)]}. \tag{4}$$

Proof. At first we show (3). Let

$$R_n = -\sum_{j=1}^{n} Z_j(\theta, \varphi).$$

Since, for any t $(0 \leqq t \leqq 1)$ and for any $k(\geqq 1)$,

$$\begin{aligned} &E_\theta\{e^{-tZ_k(\theta,\varphi)} \mid X_{k-1}=\xi\} \\ &= E_\theta\left\{\left(\frac{f(X_{k-1}, X_k: \varphi)}{f(X_{k-1}, X_k: \theta)}\right)^t \,\middle|\, X_{k-1}=\xi\right\} \leqq 1 \qquad \text{for all } \xi, \end{aligned} \tag{5}$$

so, we can define

$$\phi(n, t) = \sup_{\xi \in \mathfrak{X}} E_\theta\{e^{tR_n} \mid X_0=\xi\} \tag{6}$$

If $n=j+k$, then

$$\begin{aligned} E_\theta\{e^{tR_n} \mid X_0\} &= E_\theta\left\{E_\theta\left[\exp\left(tR_j - t\sum_{i=j+1}^{n} Z_i(\theta, \varphi)\right) \middle| X_1, \cdots, X_j\right] \middle| X_0\right\} \\ &= E_\theta\left\{e^{tR_j} E_\theta\left[\exp\left(-t\sum_{i=j+1}^{n} Z_i(\theta, \varphi)\right) \middle| X_j\right] \middle| X_0\right\} \\ &\leqq \phi(j, t) \cdot \phi(k, t) \end{aligned}$$

Thus, $\phi(n, t) \leqq \phi(j, t) \cdot \phi(k, t)$ and if $n=dm+l$ $(0 \leqq l < d)$, then

$$\phi(n, t) \leqq \phi(md, t) \cdot \phi(l, t) \leqq [\phi(d, t)]^m \cdot \phi(l, t). \tag{7}$$

Since, by Doeblin's condition (D) in A 2 (a)

$$E_\theta\left\{\frac{1}{n} R_n \mid X_0=\xi\right\} \to -I(\theta, \varphi) \qquad \text{uniformly in } \xi,$$

so, for any $\varepsilon > 0$, there is an integer d_0 such that

$$E_\theta\left\{\frac{1}{d_0} R_{d_0} \mid X_0=\xi\right\} \leqq -\frac{I(\theta, \varphi)}{1+\varepsilon} \qquad \text{uniformly in } \xi,$$

Therefore, for sufficiently small $t > 0$, we have

$$\begin{aligned} &E_\theta\{e^{tR_{d_0}} \mid X_0=\xi_0\} \\ &\leqq 1 + td_0 E_\theta\left\{\frac{R_{d_0}}{d_0} \middle| X_0=\xi_0\right\} + t^2 E_\theta\left\{[R_{d_0}]^2 \exp\left(t\sum_{j=1}^{d_0} |Z_j(\theta, \varphi)|\right) \middle| X_0=\xi_0\right\} \\ &\leqq 1 - \frac{I(\theta, \varphi)}{1+\varepsilon} td_0 + t^2 E_\theta\{R_{d_0}\}^2 \exp\left(t\sum_{j=1}^{n} |Z_j(\theta, \varphi)|\right) \middle| X_0=\xi_0\Big\}. \end{aligned} \tag{8}$$

Now, we evaluate

$$E_\theta\left\{[R_{d_0}]^2 \exp\left(t\sum_{j=1}^{d_0} |Z_j(\theta, \varphi)|\right) \middle| X_0=\xi_0\right\}.$$

Let $\eta>0$ and let

$$B_n^+(x_{k-1})=\left\{x_k\colon \log\frac{f(x_{k-1},\ x_k\colon \varphi)}{f(x_{k-1},\ x_k\colon \theta)}\geqq n\eta \qquad \text{for fixed} \quad x_{k-1}\right\}$$

$$B_n^-(x_{k-1})=\left\{x_k\colon \log\frac{f(x_{k-1},\ x_k\colon \varphi)}{f(x_{k-1},\ x_k\colon \theta)}\leqq -n\eta \quad \text{for fixed} \quad x_{k-1}\right\}$$

and

$$B_n(x_{k-1})=B_n^+(x_{k-1})\cup B_n^-(x_{k-1}).$$

Then, for any $t\in(0,\ \delta)$ (δ being the one in A 4)

$$\begin{aligned}
&P_\theta\{B_n(X_{k-1})\mid X_{k-1}=\xi\}\\
&=P_\theta\{B_n^+(X_{k-1})\mid X_{k-1}=\xi\}+P_\theta\{B_n^-(X_{k-1})\mid X_{k-1}=\xi\}\\
&\leqq e^{-n\eta t}E_\theta\left\{\exp\left[t\log\frac{f(X_{k-1},\ X_k\colon \theta)}{f(X_{k-1},\ X_k\colon \varphi)}\right]\middle| X_{k-1}=\xi\right\}\\
&\quad+e^{-n\eta t}E_\theta\left\{\exp\left[t\log\frac{f(X_{k-1},\ X_k\colon \varphi)}{f(X_{k-1},\ X_k\colon \theta)}\right]\middle| X_{k-1}=\xi\right]
\end{aligned}$$

for all ξ. Thus, by (5) and A 4, there is a constant A_1, such that

$$P_\theta\{B_n(X_{k-1})\mid X_{k-1}=\xi\}\leqq A_1\, e^{-n\eta t} \qquad \text{for all } k \text{ and } n, \tag{9}$$

independent of ξ.

Let t_0 be any number in $(0,\ \delta)$ and $\rho=e^{-n\eta t_0}$. Then, for any $t\in(0,\ t_0)$, $\rho e^{t\eta}<1$ and

$$\begin{aligned}
&E_\theta\{\exp(t\mid Z_1(\theta,\ \varphi)\mid)\mid X_0=\xi_0\}\\
&=\int_{\mathfrak{X}}\exp\left(t\left|\log\frac{f(\xi_0,\ \xi_1\colon \varphi)}{f(\xi_0,\ \xi_1\colon \theta)}\right|\right)p_\theta(\xi_0,\ d\xi_1)\\
&=\sum_{n=1}^{\infty}\int_{B_{n-1}(\xi_0)-B_n(\xi_0)}\exp\left(t\left|\log\frac{f(\xi_0,\ \xi_1\colon \varphi)}{f(\xi_0,\ \xi_1\colon \theta)}\right|\right)p_\theta(\xi_0,\ d\xi_1)\\
&=\sum_{n=1}^{\infty}e^{nt\eta}P_\theta\{B_{n-1}(X_0)\mid X_0=\xi_0\}\\
&\leqq\sum_{n=1}^{\infty}A_1\rho^n e^{nt\eta}=K_1(1-\rho e^{t\eta})^{-1},
\end{aligned} \tag{10}$$

independent of ξ_0.

Similarly, for suitably chosen constants K_2 and K_3, we have

$$E_\theta\{-Z_1(\theta,\ \varphi)\exp(t\mid Z_1(\theta,\ \varphi)\mid)\mid X_0=\xi_0\}\leqq K_2(1-\rho e^{t\eta})^{-2} \tag{11}$$

and

$$E_\theta\{\mid Z_1(\theta,\ \varphi)\mid^2\exp(t\mid Z_1(\theta,\ \varphi)\mid)\mid X_0=\xi_0\}\leqq K_3(1-\rho e^{t\eta})^{-3}, \tag{12}$$

independent of ξ_0. For brevity, let

$$g_{i-1}(t)=K_i\,(1-\rho e^{t\eta})^{-i} \qquad (i=1, 2, 3).$$

Let

$$I_{d_0}(x_{d_0-1})=\int e^{t|z_{d_0-1}(\theta,\varphi)|}\cdot\left\{z_{d_0-1}^2(\theta,\varphi)+2z_{d_0-1}(\theta,\varphi)\sum_{k=1}^{d_0-1}z_k(\theta,\varphi)+\left[\sum_{k=1}^{d_0-1}z_k(\theta,\varphi)\right]^2\right\}p(x_{d_0-1},dx_{d_0})$$

and for $k=1, 2, \cdots, d_0-1$

$$I_{d_0-k}(x_{d_0-k-1})=\int e^{t|z_{d_0-k}(\theta,\varphi)|}\,I_{d_0-k+1}(x_{d_0-k})\,p(x_{d_0-k-1},dx_{d_0-k})$$

Let $y_j=-\sum_{k=1}^{j}z_k(\theta,\varphi)$. Then, we have

$$I_{d_0}(x_{d-1})\leqq g_2(t)+2g_1(t)y_{d_0-1}+g_0(t)y_{d_0-1}^2$$

and

$$\begin{aligned}I_{d_0-1}(x_{d_0-2})&\leqq g_2(t)g_0(t)+2g_1(t)\int e^{t|z_{d_0-1}(\theta,\varphi)|}[-z_{d_0-1}(\theta,\varphi)+y_{d_0-2}]\,p(x_{d_0-2},dx_{d_0-1})\\&\quad+g_0(t)\int e^{t|z_{d_0-1}(\theta,\varphi)|}[z_{d_0-1}^2(\theta,\varphi)-2y_{d_0-2}z_{d_0-1}(\theta,\varphi)+y_{d_0-2}^2]\,p(x_{d_0-2},dx_{d_0-1})\\&\leqq g_0(t)g_2(t)+2g_1^2(t)+2g_1(t)g_0(t)y_{d_0-2}+g_0(t)g_2(t)\\&\quad+2g_0(t)g_1(t)y_{d_0-2}+g_0^2(t)y_{d_0-2}^2\\&=2g_0(t)g_2(t)+2g_1^2(t)+4g_0(t)g_1(t)y_{d_0-2}+g_0^2(t)y_{d_0-2}^2.\end{aligned}$$

Proceeding inductively we obtain

$$\begin{aligned}I_2(x_1)\leqq&(d_0-1)g_0^{d_0-2}(t)g_2^2(t)+(d_0-2)(d_0-1)g_0^{d_0-2}(t)g_1^2(t)\\&+g_0^{d_0-1}(t)y_1^2+2(d_0-1)g_0^{d_0-2}(t)g_1(t)y_1\end{aligned}$$

and finally

$$\begin{aligned}&E_\theta\left\{[R_{d_0}]^2\exp\left[t\sum_{k=1}^{d_0}\left|Z_k(\theta,\varphi)\right|\right]\middle|X_0=\xi_0\right\}\\&=\int e^{t|z_1(\theta,\varphi)|}\,p(x_0,dx_1)\int e^{t|z_2(\theta,\varphi)|}\,p(x_1,dx_2)\\&\quad\cdots\int e^{t|z_{d_0}(\theta,\varphi)|}\cdot\left\{\sum_{k=1}^{d_0}z_k(\theta,\varphi)\right\}^2 p(x_{d_0-1},dx_{d_0})\\&\leqq\int e^{t|z_1(\theta,\varphi)|}\,p(x_0,dx_1)\cdots\int e^{t|z_{d_0-2}(\theta,\varphi)|}I_{d_0}(x_{d_0-1})\,p(x_{d_0-2},dx_{d_0-1})\\&\leqq\int e^{t|z_1(\theta,\varphi)|}I_2(x_1)\,p(x_0,dx_1).\end{aligned}$$

Thus, we have

$$
\begin{aligned}
&E_\theta\left\{[R_{d_0}]^2 \exp\left(t\sum_{k=1}^{d_0} | Z_k(\theta, \varphi) |\right) | X_0=\xi_0\right\} \\
(13) \qquad &\leqq d_0\{g_0^{d_0-1}(t)g_2(t)+(d_0-1)g_0^{d_0-2}(t)g_1^2(t)\} \\
&\leqq d_0 G(t)
\end{aligned}
$$

independent of ξ_0, where $G(t)$ is defined for $(0, t_0)$ and is bounded.

By (8) and (13), we obtain

$$
\begin{aligned}
&E_\theta\{e^{tR_{d_0}} | X_0=\xi_0\} \\
&\leqq 1-\frac{I(\theta, \varphi)}{1+\varepsilon}td_0+t^2 d_0 G(t) \\
&\leqq e^{-td_0[I(\theta, \varphi)/(1+\varepsilon)-O(t)]}
\end{aligned}
$$

for sufficiently small $t\,(\in(0, t_0))$. Thus

$$
(14) \qquad \phi(d_0, t)\leqq e^{-td_0[I(\theta, \varphi)/(1+\varepsilon)-O(t)]}
$$

and, if $n=md_0+l\,(0\leqq l<d_0)$, then, by (7) and (14)

$$
\begin{aligned}
&\phi(n, t)\leqq\phi(l, t)\cdot e^{-tmd_0[I(\theta, \varphi)/(1+\varepsilon)-O(t)]} \\
(15) \qquad &\leqq A_0 e^{-nt}[I(\theta, \varphi)/(1+\varepsilon)-O(t)]
\end{aligned}
$$

where A_0 is a suitable constant. Since, for any $t\,(0\leqq t\leqq 1)$

$$
E_\theta\left\{\exp\left(t\log\frac{f(X_0\colon \varphi)}{f(X_0\colon \theta)}\right)\right\}\leqq 1
$$

so, by (15), we conclude that, for any $\varepsilon>0$ and for any sufficiently small $t\,(\in(0, t_0))$, there is a positive number A such that

$$
\begin{aligned}
&E_\theta\{e^{-tS_n(\theta, \varphi)}\} \\
(16) \qquad &\leqq E_\theta\left\{\exp\left(t\log\frac{f(X_0\colon \varphi)}{f(X_0\colon \theta)}\right)\right\}\cdot\phi(n, t) \\
&\leqq Ae^{-nt[I(\theta, \varphi)/(1+\varepsilon)-O(t)]}
\end{aligned}
$$

for all n.

In the same way, we can prove the latter half of the theorem.

5. Main results.

From now on, we represent the true state of the parameter by θ_0 which we assume to be in H_1 and all probabilities and expectations refer to θ_0 unless clearly specified otherwise.

Let $a(\theta)$ be the set alternative to θ and

$$
(17) \qquad I(\theta)=\min_{\varphi\in a(\theta)} I(\theta, \varphi).
$$

Then, by A 3 (b), $I(\theta)>0$ for all $\theta\in\Theta$.

We prove two lemmas.

LEMMA 1. *If the stopping rule is disregarded and sampling is continued, then there exist two constants K and $b>0$ such that*

$$P\{T>n\}\leqq Ke^{-bn} \tag{18}$$

where T is defined as the smallest integer such that $\hat{\theta}_n=\theta_0$ for $n\geqq T$.

Proof. Assign a priori probability $1/s$ to each value of the parameter. Then $\hat{\theta}_n$ is the value of θ which maximizes the a posteriori probability after the $n+1$-st observation. Furthermore, $\hat{\theta}_n=\theta_0$ if $S_n(\hat{\theta}_n, \theta)>0$ for all $\theta=\theta_0$. Thus, to prove Lemma 1, it suffices to show that for each $\theta\neq\theta_0$, there exist constants K and $b>0$ such that

$$P\{S_n(\theta_0, \theta)\leqq 0\}\leqq Ke^{-bn}.$$

Using an inequality in Loeve [6], p. 157, we have

$$P\{S_n(\theta_0, \theta)\leqq 0\}\leqq E\{e^{-tS_n(\theta_0, \varphi)}\}$$

for all $t>0$ and, from Theorem 1, for any $\varepsilon>0$ and for sufficiently small $t>0$, there exists $K=K(\varepsilon)$ such that

$$E\{e^{-tS_n(\theta_0, \theta)}\}\leqq Ke^{-nt[I(\theta_0, \theta)/(1+\varepsilon)-O(t)]},$$

so, for a sufficiently small $t_1>0$ such that $t_1\in(0, t_0)$ and $b=t_1\,[I(\theta_0, \theta)/(1+\varepsilon)-O(t_1)]>0$, we have

$$P\{S_n(\theta_0, \theta)\leqq 0\}\leqq Ke^{-bn}$$

and Lemma 1 is proved.

LEMMA 2. *For procedure A, we have*

$$E(N)\leqq -(1+o(1))\frac{\log c}{I(\theta_0)} \tag{19}$$

where $N+1$ is the sample size required to reach a decision.

Proof. It is sufficient to show that for any given $\varepsilon>0$, there exists a positive number $c^*(\varepsilon)$ such that for all $c<c^*(\varepsilon)$

$$E(N)\leqq -(1+\varepsilon)\,\frac{\log c}{I(\theta_0)}\,. \tag{20}$$

For each φ, let N_φ be the smallest integer such that

$$S_n(\theta_0, \varphi)>-\log c \qquad \text{for all } n\geqq N_\varphi,$$

Since $N\leqq \max_{\varphi\in H_2}(\max(N_\varphi, T))$, so

$$P\{N>n\}\leqq P\{\max_{\varphi\in H_2}(\max(N_\varphi, T))>n\}. \tag{21}$$

To evaluate (21), we, at first, show that for any $\varphi\in H_2$ and any ε' $(0<\varepsilon'<\varepsilon)$, there exist two constants $K_1=K_1(\varepsilon, \varphi)>0$ and $b_1=b_1(\varepsilon', \varphi)>0$ such that

$$P\{N_\varphi>n\}\leqq K_1e^{-bn} \text{ for } n>n_0\geqq-(1+\varepsilon)\frac{\log c}{I(\theta_0)}, \tag{22}$$

or, for any $\varphi\in H_2$ and any $\varepsilon'>0$,

$$P\{S_n(\theta_0, \varphi)<-\log c\}\leqq K_1e^{-b_1n} \tag{23}$$

for $n>n_0$.

Let $K_1=K_1(\varepsilon'/2, \varphi)$ be the constant obtained in Theorem 1 for the above given ε' and φ. If n is so large that $n>n_0$, then, for $t\in(0, t_0)$

$$\begin{aligned}
&P\{S_n(\theta_0, \varphi)<-\log c\}\\
&\leqq P\left\{-S_n(\theta_0, \varphi)+\frac{nI(\theta_0)}{1+\varepsilon'}>0\right\}\\
&\leqq e^{nI(\theta_0)t/(1+\varepsilon')}\cdot E\{e^{-tS_n(\theta_0,\varphi)}\}\\
&\leqq e^{nI(\theta_0)t/(1+\varepsilon')}\cdot K_1e^{-nt[I(\theta_0, \varphi)/(1+\varepsilon'/2)-O(t)]}\\
&\leqq K_1e^{-nt[I(\theta_0)/2(1+\varepsilon')^2-O(t)]}.
\end{aligned}$$

If we take $t_1>0$ so small that $t_1\in(0, t_0)$ and

$$b_1=t_1\left(\frac{I(\theta_0)}{2(1+\varepsilon')^2}-O(t_1)\right)>0$$

then we obtain (23), and thus (22).

On the other hand, by Lemma 1, there are two constants K_2 and b_2 such that

$$P\{T>n\}\leqq K_2e^{-b_2n} \qquad \text{for all } n>n_0. \tag{24}$$

Thus, by (21), (22) and (24)

$$P\{N>n\}\leqq 2s\max_{\varphi\in H_2}\{\max[P\{N_\varphi>n\},\ P\{T>n\}]\}\leqq 2sKe^{-nb} \tag{25}$$

for all $n>n_0$, where s denotes the number of elements in H_2, and K and b are suitably chosen positive numbers.

Using (25), we have

$$\begin{aligned}
E(N)&=\sum_{n=0}^{\infty}P\{N>n\}\\
&=\sum_{n=0}^{n_0}P\{N>n\}+\sum_{n=n_0+1}^{\infty}P\{N>n\}\\
&\leqq n_0+\sum_{n=n_0+1}^{\infty}2sKe^{-nb}\\
&\leqq-\frac{(1+\varepsilon')\log c}{I(\theta_0)}+Mc^{b_0}
\end{aligned}$$

where M and b_0 are two positive constants. Therefore, we obtain (20), and complete the proof.

LEMMA 3. *For procedure A, the probability of error (rejecting H_1) is $\alpha=O(c)$.*

Proof. Let $A_{n\varphi}$ be the set in the sample space for which we reject the hypothesis $\theta_0 \in H_1$ at the $n+1$-st observation and for which $\hat{\theta}_n=\varphi \in H_2$. Since on the set $A_{n\varphi}$,

$$\sum_{j=0}^{n} Z_j(\varphi, \theta_0) \geqq \sum_{j=0}^{n} Z_j(\hat{\theta}_n, \tilde{\theta}_n) \geqq -\log c,$$

so we have

$$f(x_0: \theta_0) \prod_{j=1}^{n} f(x_{j-1}, x_j: \theta_0) \leqq c f(x_0: \varphi) \prod_{j=1}^{n} f(x_{j-1}, x_j: \varphi).$$

Thus

$$P\{A_{n\varphi}\} = \int_{A_{n\varphi}} f(x_0: \theta_0) \prod_{j=1}^{n} f(x_{j-1}, x_j: \theta_0) \lambda(dx_0)\lambda(dx_1)\cdots\lambda(dx_n)$$

$$\leqq c \int_{A_{n\varphi}} f(x_0: \varphi) \prod_{j=1}^{n} f(x_{j-1}, x_j: \varphi) \lambda(dx_0)\lambda(dx_1)\cdots\lambda(dx_n)$$

and the last integral is the probability of the set $A_{n\varphi}$ when φ is the true value of the parameter. Therefore,

$$\alpha(\theta_0) = \sum_{\varphi \in H_2} \sum_{n=0}^{\infty} P\{A_{n\varphi}\} \leqq \sum_{\varphi \in H_2} c < sc = O(c).$$

Thus we have the lemma.

Let the risk function $R(\theta)$ be

$$R(\theta) = r(\theta)\alpha(\theta) + cE_\theta(N) \qquad \text{for } \theta \in \Theta, \tag{26}$$

where $r(\theta)=0$ if $\theta \in H_i$ and we accept the hypothesis H_i; $r(\theta)>0$ otherwise.
Combining Lemmas 2 and 3, we have the next theorem.

THEOREM 2. *For procedure A, the risk function $R(\theta)$ satisfies*

$$R(\theta) \leqq -(1+o(1)) \frac{c \log c}{I(\theta)} \qquad \textit{for all } \theta \in \Theta. \tag{27}$$

LEMMA 4. *If $\varphi \in H_2$, $P\{accept\ H_1 \mid \theta=\varphi\} = O(-c\log c)$, $P\{reject\ H_1\} = O(-c \log c)$ and $0<\varepsilon<1$, then*

$$P\{S(\theta_0, \varphi) < -(1-\varepsilon) \log c\} = O(-c^\varepsilon \log c) \tag{28}$$

where

$$S(\theta_0, \varphi) = \sum_{j=0}^{N} Z_j(\theta_0, \varphi).$$

Proof. As

$$P\{\text{accept } H_1 \mid \theta=\varphi\}=O(-c\log c),$$

so there is a number K such that

$$-Kc\log c \geqq P\{\text{accept } H_1 \mid \theta=\varphi\}. \tag{29}$$

Let A_n be the subset of the sample space for which $S(\theta_0, \varphi)<-(1-\varepsilon)\log c$ and H_1 is accepted at the $n+1$-st step. Then

$$\begin{aligned} &P\{\text{accept } H_1 \mid \theta=\varphi\} \\ &\geqq \sum_{n=0}^{\infty}\int_{A_n} f(x_0\colon \varphi)\prod_{i=1}^{n} f(x_{i-1}, x_i\colon \varphi)\lambda(dx_0)\lambda(dx_1)\cdots\lambda(dx_n) \\ &\geqq \sum_{n=0}^{\infty}\int_{A_n} \frac{f(x_0\colon \varphi)\prod_{i=1}^{n} f(x_{i-1}\ x_i\colon \varphi)}{f(x_0\colon \theta_0)\prod_{i=1}^{n} f(x_{i-1}\ x_i\colon \theta_0)} f(x_0\colon \theta_0)\prod_{i=1}^{n} f(x_{i-1}, x_i\colon \theta_0)\lambda(dx_0)\lambda(dx_1)\cdots\lambda(dx_n) \qquad (30) \\ &= \sum_{n=0}^{\infty}\int_{A_n} e^{-S(\theta_0, \varphi)} f(x_0\colon \theta_0)\prod_{i=1}^{n} f(x_{i-1}, x_i\colon \theta_0)\lambda(dx_0)\lambda(dx_1)\cdots\lambda(dx_n) \\ &\geqq c^{1-\varepsilon}\sum_{n=0}^{\infty} P\{A_n\}. \end{aligned}$$

Combining (29) and (30), we obtain

$$\sum_{n=0}^{\infty} P\{A_n\}=O(-c^{\varepsilon}\log c) \tag{31}$$

and since, by assumption, $P\{\text{reject } H_1\}=O(-c\log c)$, so, using (31) we have

$$\begin{aligned} &P\{S(\theta_0, \varphi)<-(1-\varepsilon)\log c\} \\ &\leqq \sum_{n=0}^{\infty} P\{A_n\}+P\{\text{reject } H_1\}=O(-c^{\varepsilon}\log c). \end{aligned}$$

LEMMA 5. *If* $\varepsilon>0$, *then*

$$P\{\max_{1\leqq m\leqq n}\ \min_{\varphi\in H_2} S_m(\theta_0, \varphi)\geqq n[I(\theta_0)+\varepsilon]\}\to 0 \tag{32}$$

as $n\to\infty$.

Proof. For any $t<0$

$$\begin{aligned} &P\{\max_{1\leqq m\leqq n}\ \min_{\varphi\in H_2} S_m(\theta_0, \varphi)\geqq n[I(\theta_0)+\varepsilon]\} \\ &\leqq n \max_{1\leqq m\leqq n} P\{\min_{\varphi\in H_2} S_m(\theta_0, \varphi)\geqq n[I(\theta_0)+\varepsilon]\} \\ &\leqq n \max_{1\leqq m\leqq n} e^{-nt[I(\theta_0)+\varepsilon]}\cdot E\{e^{t\min_{\varphi\in H_2} S_m(\theta_0,\varphi)}\}. \end{aligned}$$

Using Theorem 1, for sufficiently small $t>0$ ($t\in(0, \delta)$),

$$E\{\exp[t\min_{\varphi\in H_2} S_m(\theta_0, \varphi)]\}$$

$$\leqq \min_{\varphi \in H_2} E\{e^{tS_m(\theta_0, \varphi)}\}$$

$$\leqq \min_{\varphi \in H_2} B \cdot e^{mt[I(\theta_0, \varphi)/(1-\varepsilon')+O(t)]}$$

$$= Be^{mt[I(\theta_0)/(1-\varepsilon')+O(t)]},$$

where $0<\varepsilon'<\varepsilon/(I(\theta_0)+\varepsilon)$ and $B=B(\varepsilon')$ is the one in Theorem 1. So we have

$$P\{\max_{1\leqq m\leqq n} \min_{\varphi \in H_2} S_m(\theta_0, \varphi) \geqq n[I(\theta_0)+\varepsilon]\}$$

$$\leqq nBe^{-nt[I(\theta_0)+\varepsilon]} \cdot \max_{1\leqq m\leqq n} e^{mt[I(\theta_0)/(1-\varepsilon')+O(t)]}$$

$$= nBe^{-nt[-\varepsilon' I(\theta_0)/(1-\varepsilon')+O(t)]},$$

and if we choose a sufficiently small number $t_2>0$ such that $t_2 \in (0, t_0')$ (t_0' being a number in $(0, \delta)$) and

$$b=t_2\left(-\frac{\varepsilon' I(\theta_0)}{1-\varepsilon'}+\varepsilon-O(t_2)\right)>0$$

then

$$P\{\max_{1\leqq m\leqq n} \min_{\varphi \in H_2} S_m(\theta_0, \varphi) \geqq n[I(\theta_0)+\varepsilon]\} \leqq nBe^{-nb}.$$

Thus, we complete the proof.

Combining Lemmas 4 and 5, we have the next theorem.

THEOREM 3. *Let* $\{X_k: k=0, 1, 2, \cdots\}$ *be any Markov chain which satisfies* A 2–A 4. *Then any procedure, for which*

(33) $$R(\theta)=O(-c \log c) \qquad \textit{for all } \theta \in \Theta$$

satisfies

(34) $$R(\theta) \geqq -(1+o(1)) \frac{c \log c}{I(\theta)} \qquad \textit{for all } \theta \in \Theta.$$

Proof. Let

$$n_c^* = -\frac{(1-\varepsilon) \log c}{I(\theta)+\varepsilon} \qquad (0<\varepsilon<1).$$

Since, if $N \leqq n_c^*$ and $\min_{\varphi \in H_2} S_N(\theta, \varphi) \geqq -(1-\varepsilon) \log c$, then

$$\max_{1\leqq m\leqq n_c^*} \min_{\varphi \in H_2} S_m(\theta, \varphi) \geqq n_c^* (I(\theta)+\varepsilon),$$

so

$$P_\theta\{N \leqq n_c^*\}$$

$$\leqq P_\theta\{\mathrm{mav}_{1\leqq m\leqq n_c^*} \min_{\varphi \in H_2} S_m(\theta, \varphi) \geqq n_c^*[I(\theta)+\varepsilon]\}$$

$$+P_\theta\{\min_{\varphi\in H_2} S_N(\theta, \varphi)\leqq -(1-\varepsilon)\log c\}.$$

As, by assumption, $R(\theta)=O(-c\log c)$ for each $\theta\in\Theta$, so lemma 4 can be applied to each θ and each $\varphi\in a(\theta)$. Thus

$$P_\theta\{\min_{\varphi\in H_2} S_N(\theta, \varphi)\leqq -(1-\varepsilon)\log c\}$$
$$\leqq \sum_{\varphi\in H_2} P_\theta\{S_N(\theta, \varphi)\leqq -(1-\varepsilon)\log c\}=O(-c^\varepsilon\log c),$$

on the other hand, by lemma 5

$$P_\theta\{\max_{1\leqq m\leqq n}\ \min_{\varphi\in H_2} S_m(\theta, \varphi)\geqq n_c^*[I(\theta)+\varepsilon]\}\to 0 \qquad (n_c^*\to\infty).$$

Therefore,

$$E_\theta(N)\geqq n_c^* P_\theta\{N>n_c^*\}=-\frac{(1-\varepsilon)(1+o(1))\log c}{I(\theta)+\varepsilon}$$

and, for all $\varepsilon\ (0<\varepsilon<1)$,

$$R(\theta)\geqq cE_\theta(N)\geqq -\frac{(1+o(1))(1-\varepsilon)\ c\log c}{I(\theta)+\varepsilon}$$

and thus

$$R(\theta)\geqq -\frac{(1+o(1))\ c\ \log cc}{I(\theta)},$$

which was to be proved.

By Theorems 1 and 2, we see that procedure A is "asymptotically optimal" in the sense that for any procedure to do essentially better than procedure A for any θ' implies that its risk will be of a greater order of magnitude for some θ''.

6. Examples.

Now we consider two examples.

EXAMPLE 1. Let $\mathfrak{X}$ be the real line and let $\mathfrak{A}$ consist of the linear Borel sets. Let X_0 be arbitrarily distributed and let $\{X_1, X_2, \cdots\}$ be defined recursively by

(35) $$X_n=\alpha X_{n-1}+\beta+\gamma^{1/2}Y_n$$

where $(X_0, Y_1, Y_2, \cdots)$ is an independent sequence of random variables, the Y_n each being $\mathfrak{N}(0, 1)$ and where $|\alpha|<1$. Only the Y_n are observed. In this case, the unique stationary distribution is

$$\mathfrak{N}\left(\frac{\beta}{1-\alpha}, \frac{\gamma}{1-\alpha^2}\right).$$

Let $\Theta=\{(\beta_i, \gamma_j): \gamma_j>0, i=1, \cdots, m, j=1, \cdot\cdot, n\}$. If λ is taken to be Lebesgue measure, then the transition densities are given by

(36) $$f(\xi_{k-1}, \xi_k: \theta)=\frac{1}{\sqrt{2\pi\gamma}}e^{-(\xi_k-\alpha\xi_{k-1}-\beta)^2/2\gamma}$$

and stationary densities are given by

(37) $$f(\xi_0: \theta)=\frac{\sqrt{1-\alpha^2}}{\sqrt{2\pi\gamma}}e^{-(1-\alpha^2)(\xi_0-\beta/(1-\alpha))^2/2\gamma}.$$

The densities $f(\xi_{k-1}, \xi_k: \theta)$ and $f(\xi_0: \theta)$ obviously satisfy A 3 and A 4. Thus, we we can apply the above testing procedure A to this case. For the densities $f(\xi_{k-1}, \xi_k: \theta)$,

(38) $$\begin{aligned} I(\theta, \varphi) &= E_\theta\left\{-\frac{(X_k-\alpha X_{k-1}-\beta_1)^2}{2\gamma_1}+\frac{(X_k-\alpha X_{k-1}-\beta_2)^2}{2\gamma_2}-\frac{1}{2}\log\frac{\gamma_1}{\gamma_2}\right\} \\ &= \frac{1}{2}\left\{\log\frac{\gamma_2}{\gamma_1}-1+\frac{\gamma_1}{\gamma_2}+\frac{1}{\gamma_2}(\beta_1-\beta_2)^2\right\} \end{aligned}$$

if $\theta=(\beta_1, \gamma_1)$ and $\varphi=(\beta_2, \gamma_2)$.

Example 2. Let the state space consists of the non-negative integers and let the transition probabilities be

(39) $$p_{i, i+1}(\theta)=\theta, \qquad p_{i, i-1}(\theta)=1-\theta$$

for $i=1, 2, \cdots$, while $p_{01}(\theta)=1$. If $\lambda(i)=1$ for each $i\in\mathfrak{X}$, then the densities are just the $p_{ij}(\theta)$. If we assume that $0<\theta<1/2$, then the unique stationary distribution is given by

(40) $$p_0(\theta)=\frac{1-2\theta}{2-2\theta}, \qquad p_i(\theta)=p_0(\theta)\cdot\frac{1}{\theta}\left(\frac{\theta}{1-\theta}\right)^i.$$

Let $\Theta=\{\theta_i: 0<\theta_i<1/2, i=1, \cdots, l\}$. Then, for $\theta\in\Theta$, $p_{i, i+1}(\theta)$, $p_{i, i-1}(\theta)$ and $p_i(\theta)$ satisfy A 2–A 4, and thus we can deduce the same conclusion. In this case

(41) $$\begin{aligned} I(\theta, \varphi) &= \sum_{i=0}^{\infty}\left\{p_{i, i+1}(\theta)\log\frac{p_{i, i+1}(\theta)}{p_{i, i+1}(\varphi)}+p_{i, i-1}(\theta)\log\frac{p_{i, i-1}(\theta)}{p_{i, i-1}(\varphi)}\right\}p_i(\theta) \\ &= \left\{\theta\log\frac{\theta}{\varphi}+(1-\theta)\log\frac{1-\theta}{1-\varphi}\right\}\sum_{i=1}^{\infty}p_i(\theta) \\ &= \frac{1}{2(1-\theta)}\left\{\theta\log\frac{\theta}{\varphi}+(1-\theta)\log\frac{1-\theta}{1-\varphi}\right\}. \end{aligned}$$

Acknowledgement. The author wishes to thank Dr. M. Udagawa, the Tokyo University of Education, for his helpful suggestions and guidance.

References

[1] Billingsley, P., Statistical inference for Markov processes. Univ. of Chicago Press. (1961).

[2] Baum, L. E., M. Katz, and R. R. Read, Exponential convergence rates for the

law of large numbers. Trans. Amer. Math. Soc. **102** (1962), 187–199.

[3] CHERNOFF, H., Sequential design of experiments. Ann. Math. Stat. **30** (1959), 755–770.

[4] DOOB, J. L., Stochastic processes. John Wiley and Sons. (1953).

[5] KOOPMAN, L. H., Asymptotic rate of discrimination for Markov processes. Ann. Math. Stat. **31** (1960), 982–994.

[6] LOEVE, M., Probability theory. D. Van Nostrand Co., Inc. (1955).

[7] KULLBACK, S., Information theory and statistics. John Wiley and Sons. (1959).

FACULTY OF ART EDUCATION,
YOKOHAMA NATIONAL UNIVERSITY.