SEQUENTIAL HYPOTHESIS TESTS FOR r-DEPENDENT MARGINALLY STATIONARY PROCESSES[1]

By M. Tainiter[2]

The Johns Hopkins University

1. Introduction. In [3], and more generally in [4] and [5], Robbins and Samuel respectively treated the following recurring statistical decision problem. Suppose we sequentially observe a sequence of independent random variables $\{X_n\}$: $n = 0, 1, 2, \cdots$, where the distribution of X_n depends on a random variable θ_n. The process $\{\theta_n\}$ is a sequence of independent identically distributed random variables. After each observation X_i we test a hypothesis H_0 : concerning the value of θ_i. If the common *a priori* distributions of the θ_i were known, one could find and use the "Bayes' decision function which would be optimal (as defined). However, if the *a priori*' distribution of the θ_i was unknown, Robbins has shown that under certain general conditions one could find a sequence of decision functions whose risk asymtotically approached the "Bayes" risk (i.e., the risk in using the "Bayes" decision function). Such sequences of decision functions are called *asymptotically optimal*.

In [6] the results obtained by Robbins were applied to problems of pulse detection in noisy environments. They were shown to be quite "good" (i.e., the risk in using the asymptotically optimal sequence differed slightly from the Bayes risk) under certain conditions. One of the difficulties in using Robbins' results was that the noise was assumed to be uncorrelated. Thus it is of interest (theoretically and practically) to extend certain results of Robbins and Samuel to processes which are dependent. In this paper some results are extended to r-dependent strictly stationary processes. We will consider in detail the problem of testing a completely specified simple hypothesis against a simple alternative. This will show how basically the results of Robbins can be extended to mildly dependent processes. The treatment of this problem will parallel Robbins' treatment of the same problem for independent processes (see [3]). In the remarks and extensions section we will show that other problems and ideas given by Robbins in [4] can be similarly extended. Also the processes can be extended to rth order Markov processes under suitable conditions.

2. Definitions and notation. We observe a sequence of random variables $\{Y_n\}: n = 0, 1, 2, \cdots$, sequentially. After observing $Y_0, Y_1, \cdots, Y_k(k = 0, 1, 2 \cdots)$ we are required to make a decision about the value of a random parameter $\theta_k(\theta_0, \theta_1, \cdots, \theta_k$ remain forever unknown). We suppose that the $\{\theta_n\}$ are a sequence of independent identically distributed random variables with two possible values, say "0" and "1" with probabilities,

Received 28 December 1964; revised 25 August 1965.

[1] This research was supported by the United States Air Force, Avionics Laboratory, Systems Command under Contract AF 33(657)-11029.

[2] Now at Brookhaven National Laboratories, Upton, L. I., New York.

$$p = \Pr\{\theta_n = 0\}, \qquad 1 - p = \Pr\{\theta_n = 1\}, \qquad n = 0, 1, 2, \cdots .$$

We suppose that the distribution of Y_k is P_0 if $\theta_k = 0$ and P_1 if $\theta_k = 1$. It is further required that the sequence $\{Y_n\}$ be r-dependent. More precisely, let $B_n = B(Y_n, Y_{n+1}, \cdots)$ be the σ-field generated by $(Y_n, Y_{n+1}, \cdots)$ and let $C_n = C(Y_0, \cdots, Y_n)$ be the σ-field generated by $(Y_0, \cdots, Y_n)$ for all $n \geqq 0$. Then we define,

Definition 2.1. The sequence of random variables $\{Y_n\} : n = 0, 1, 2, \cdots$ is said to be r-dependent if C_n is independent of B_{n+r+1} for all $n = 0, 1, \cdots$, e.g., Y_{r+1} is independent of Y_0, Y_{r+2} is independent of Y_1, etc.

We will further require that the sequence $\{Y_n\} : n = 0, 1, 2, \cdots$ be strictly stationary in the usual sense. However, since we are actually dealing with a sequence of two "dimensional" random variables, i.e., (Y_n, θ_n) we will define precisely the strict stationarity of the sequence $\{Y_n\}$.

Let $k_1, k_2, \cdots, k_s$ be any s subscripts of the $\{Y_n\}$ sequence. Let $g(\theta_{k_1}, \cdots, \theta_{k_s})$ be the joint density of the corresponding s θ's. (Note that g can be written as a product.) Let $F_{\theta_{k_1}, \cdots, \theta_{k_s}}(y_{k_1}, \cdots, y_{k_s})$ be the conditional distribution function of $(Y_{k_1}, \cdots, Y_{k_s})$ given $\theta_{k_1}, \cdots, \theta_{k_s}$. Then we define,

Definition 2.2. The stochastic process $\{Y_n\} : n = 0, 1, 2, \cdots$ is said to be marginally strictly stationary if its marginal distribution function $F(y_{k_1}, \cdots, y_{k_s})$ for any s subscripts satisfies

$$F(y_{k_1}, \cdots, y_{k_s}) = F(y_{k_1+h}, y_{k_2+h}, \cdots, y_{k_s+h}), \qquad h = 0, 1, 2, \cdots .$$

We assume that the sequence $\{Y_n\}$ is marginally strictly stationary, thus by definition it is strictly stationary in the usual sense. Furthermore, by the definitions given above we have

$$(2.1) \quad F(y_{k_1}, \cdots, y_{k_s}) = \sum_{i=1}^{s} \sum_{\theta_{k_i}=0}^{1} F_{\theta_{k_1}, \cdots, \theta_{k_s}}(y_{k_1}, \cdots, y_{k_s}) g(\theta_{k_1}, \cdots, \theta_{k_s})$$

where

$$g(\theta_{k_1}, \cdots, \theta_{k_s}) = (1 - p)^{(\sum_{j=1}^{s} \theta_{k_j})} \cdot p^{(s - \sum_{j=1}^{s} \theta_{k_j})}$$

by the definition of the $\{\theta_n\}$ process.

Thus the general problem can be stated as follows. Each time we observe a Y, say Y_n, we must decide on the value of θ_n which is unknown. Since $Y_m (m \geqq r + 1)$ depends on $(y_{m-1}, \cdots, y_{m-r})$ the decision about $\theta_m (m \geqq r + 1)$ will depend on y_m and conditionally on the r-latest past.

Hence at each observation of Y we take one of two decisions or actions, say A_0 and A_1, where A_0 is correct if $\theta = 0$ and A_1 is correct if $\theta = 1$.

We assume the following simple loss structure.

$$\begin{aligned} L_{A_i}(\theta) &= 0 && \text{if} \quad \theta = i, \\ &= a_2 && \text{if} \quad i = 0, \quad \theta = 1, \\ &= a_1 && \text{if} \quad i = 1, \quad \theta = 0, \end{aligned}$$

and suppose $L_{A_i}(\theta) \geqq 0$, where $L_{A_i}(\theta)$ is the loss incurred when taking action A_i

when $\theta_n = \theta(\theta = 0, 1)$. (We assume the loss structure is the same for each decision.)

Thus we seek a sequence of decision functions $T = \{t_k\}, k = 0, 1, 2, \cdots$, with values in $A = \{A_0, A_1\}$ (which henceforth we denote by $\{0, 1\} = A$) which is optimal. The sense of optimality will be that it minimizes the Bayes risk at each decision.

Let $R_n(p)$ be the Bayes risk at the nth decision, and although t_k depends on the last $(r + 1)$ y's we will denote it by t_k suppressing the functional dependence: it will be clear from the context what t_k depends upon.

We proceed to solve the Bayes problem with $p = \Pr\{\theta_n = 0\}$ known, and then consider the empirical Bayes problem with p unknown.

3. Bayes solution: Known p. It is clear that the loss on any decision, say the νth, depends on θ_ν and y_ν and is conditionally dependent on $y_{\nu-r}, \cdots, y_{\nu-1}$. For $\nu > r$ we have immediately

$$(3.1)\quad R_\nu(p) = \int_{\Lambda_r} \sum_{\theta_\nu=0}^{1} [\int_\Lambda L_{t_\nu}(\theta_\nu)\, dF_{\theta_\nu}(y_\nu \mid y_{\nu-r}, \cdots, y_{\nu-1})] g(\theta_\nu)\, dF(y_{\nu-r}, \cdots, y_{\nu-1})$$

where $F_{\theta_\nu}(y_\nu \mid y_{\nu-r}, \cdots, y_{\nu-1})$ is the conditional distribution function of Y_ν given $Y_{\nu-1}, \cdots, Y_{\nu-r}$ and θ_ν, $g(\theta_\nu)$ is the density of θ_ν and $F(y_{\nu-r}, \cdots, y_{\nu-1})$ is the joint distribution function of $(Y_{\nu-r}, \cdots, Y_{\nu-1})$. Λ_r is the Cartesian product space of values of the y's, say Λ. To shorten notation we denote $F_{\theta_\nu}(y_\nu \mid y_{\nu-r}, \cdots, y_{\nu-1})$ by $F_{\theta_\nu}(y_\nu \mid \underline{Y}_r)$ and $F(y_{\nu-r}, \cdots, y_{\nu-1})$ by $F(Y_{\nu,r})$. Thus by the definition of $g(\cdot)$ and $L_{t_\nu}(\cdot)$ we have by elementary manipulation,

$$(3.2)\quad R_\nu(p) = (1-p)a_2 + \int_{\Lambda_r} \int_\Lambda [pa_1 t_\nu\, dF_0(y_\nu \mid \underline{Y}_r) - (1-p)a_2 t_\nu\, dF_1(y_\nu \mid \underline{Y}_r)]\, dF(Y_{\nu,r}).$$

There is no loss in generality in assuming that $F_0(y_\nu \mid \underline{Y}_r)$ and $F_1(y_\nu \mid \underline{Y}_r)$ are absolutely continuous with respect to some measure μ thus (3.2) becomes, denoting the densities by f_0, f_1 respectively,

$$(3.3)\quad R_\nu(p) = (1-p)a_2 - \int_{\Lambda_r} \int_\Lambda [(1-p)a_2 f_1(y_\nu \mid \underline{Y}_r) - pa_1 f_0(y_\nu \mid \underline{Y}_r)] t_\nu\, d\mu\, dF(Y_{\nu,r})$$

which is the function we seek to minimize, where t_ν is either zero or one. (We have assumed throughout that randomized rules were not needed to be considered.)

Let $\phi_p(y_\nu) = (1-p)a_2 f_1(y_\nu \mid \underline{Y}_r) - pa_1 f_0(y_\nu \mid \underline{Y}_r)$. Then clearly the optimal t_ν is given by

$$(3.4)\quad \begin{aligned} t_\nu(y_\nu) &= 1 \quad \text{if } \phi_p(y_\nu) \geqq 0 \\ &= 0 \quad \text{otherwise} \end{aligned}$$

and the corresponding Bayes risk is

$$(3.5)\quad R_\nu(p) = (1-p)a_2 - \int_{\Lambda_r} [\int_\Lambda [\phi_p(y_\nu)]^+\, d\mu]\, dF(Y_{\nu,r}),$$

where $[H(x)]^+ = \max[0, H(X)]$. (Note: $R_\nu(p)$ is independent of ν if $\nu > r$.)

Similarly, the obvious corresponding results hold for the first r decisions.

4. Empirical Bayes solution: Unknown p. We now ask if there exists a sequence $T^* = \{t_n^*\}$ of decision functions such that the associated risks $R_n^*(p)$ satisfies

$$\lim_{n\to\infty} R_n^*(p) = R_\nu(p) = R(p) \qquad \text{for all} \quad 0 < p < 1 \tag{4.1}$$

where p is unknown. The answer is yes and we proceed to exhibit such a sequence.

Let $p_n = p_n(y_0, \cdots, y_n)$ be a sequence of functions satisfying for every fixed $(r+1)$ tuple $y^{(1)}, y^{(2)}, \cdots, y^{(r+1)}$.

$$\Pr\{\lim_{n\to\infty} |p_n(y_0, \cdots, y_{n-r-1}, y^{(1)}, \cdots, y^{(r+1)}) - p| > \epsilon\} = 0$$

i.e., strong convergence with $0 \leqq p_n \leqq 1$.

Such a sequence will be exhibited later in this section.

Let $t_n^*(y_n) = t_n^*(y_0, \cdots, y_n)$ be a sequence of decision functions defined as follows for $n > r$.

$$\begin{aligned} t_n^*(y_n) &= 1 \qquad \text{if} \quad \phi_{p_n}(y_n) \geqq 0 \\ &= 0 \qquad \text{otherwise} \end{aligned} \tag{4.2}$$

where

$$\phi_{p_n}(y_n) = (1 - p_n)a_2 f_1(y_n \mid \underline{Y}_r) - p_n \cdot a_1 f_0(y_n \mid \underline{Y}_r). \tag{4.3}$$

Thus the risk on the $(n+1)$th decision using $t_n^*(y_n)$ is clearly

$$\begin{aligned} R_n^*(p) = (1-p)a_2 - (1-p)a_2 E[t_n^*(Y_n) \mid \theta_n = 1] \\ + p a_1 E[t_n^*(Y_n) \mid \theta_n = 0] \end{aligned} \tag{4.4}$$

where expectation is with respect to the $(n+1)$ random variables $Y_0, \cdots, Y_n$. Hence by elementary computation we have using the r-dependence of $\{Y_n\}$,

$$\begin{aligned} R_n^*(p) = (1-p)a_2 - (1-p)a_2 \textstyle\int_{\Lambda(n+1)} t_n^*(y_n)\, d \\ \cdot[F(y_0, \cdots, y_{n-1}) \times F_1(y_n \mid \underline{Y}_r)] \\ + p a_1 \textstyle\int_{\Lambda(n+1)} t_n^*(y_n)\, d[F(y_0, \cdots, y_{n-1}) \times F_0(y_n \mid \underline{Y}_r)] \end{aligned} \tag{4.5}$$

where the zero and one subscripts imply that the conditional distribution functions are also conditioned on θ_n. Since all F's are by definition finite measures, the cross product measures can be broken up by the iterated integral theorem (See e.g., Loève), and we then combine the expressions and obtain

$$\begin{aligned} R_n^*(p) = (1-p)a_2 \\ - \textstyle\int_{\Lambda_n} [\int_\Lambda t_n^*(y_n)[(1-p)a_2\, dF_1(y_n \mid \underline{Y}_r) - p a_1\, dF_0(y_n \mid \underline{Y}_r)] \\ \cdot dF(Y_0, \cdots, Y_{n-1}). \end{aligned} \tag{4.6}$$

Using the absolute continuity of $F_1(y_n \mid \underline{Y}r)$, $F_0(y_n \mid \underline{Y}_r)$ with respect to μ, (4.6) becomes

$$(4.7) \quad R_n^*(p) = (1-p)a_2 - \int_{\Lambda_n} [\int_\Lambda t_n^*(y_n)\phi_p(y_n)\, d\mu]\, dF(y_0, \cdots, y_{n-1}).$$

But we now further decompose the measure $F(y_0, \cdots, y_{n-1})$ into

$$F(y_{n-r}, \cdots, y_{n-1}) \times F(y_0, \cdots, y_{n-r-1} \mid y_{n-r}, \cdots, y_{n-1}).$$

Using this decomposition and the iterated integral theorem, (4.7) becomes

$$(4.8) \quad \begin{aligned} R_n^*(p) &= (1-p)a_2 \\ &- \int_{\Lambda_r} [\int_{\Lambda(n-r)} t_n^*(y_n)\, dF(y_0, \cdots, y_{n-r-1} \mid y_{n-r}, \cdots, y_{n-1})] \\ &\cdot \int_\Lambda \phi_p(y_n)\, d\mu\, dF(Y_{n,r}). \end{aligned}$$

By strict stationarity we have for all $n > r$,

$$\begin{aligned} R_n^*(p) &= (1-p)a_2 \\ &- \int_{\Lambda_r} [\int_{\Lambda(n-r)} t_n^*(y^{(r+1)})\, dF(y_0, \cdots, y_{n-r-1} \mid y^{(1)}, \cdots, y^{(r)}) \\ &\cdot \int_\Lambda \phi_p(y^{(r+1)})\, d\mu\, dF(y^{(1)}, \cdots, y^{(r)}) \end{aligned}$$

where there is no dependence on n in the last integral, i.e.,

$$(4.10) \quad \int_{\Lambda(r)} \int_\Lambda \phi_p(y^{(r+1)})\, d\mu\, dF(y^{(1)}, \cdots, y^{(r)}).$$

We recognize immediately the relationship between (4.9) and (3.5), thus to prove $R_n^*(p)$ goes to $R(p)$ it is sufficient to prove that

$$(4.11) \quad \begin{aligned} &\int_{\Lambda(n-r)} t_n^*(y^{(r+1)})\, dF(y_0, \cdots, y_{n-r-1} \mid y^{(1)}, \cdots, y^{(r)}) \\ &\qquad \to 1 \quad \text{if} \quad \phi_p(y^{(r+1)}) > 0 \\ &\qquad \to 0 \quad \text{if} \quad \phi_p(y^{(r+1)}) < 0 \end{aligned}$$

and that the remaining integral converges for $\nu > r$ to

$$(4.12) \quad \begin{aligned} &\int_{\Lambda(r)} [\int_\Lambda [\phi_p(y_\nu)]^+\, d\mu]\, dF(Y_{\nu,r}) \\ &\qquad = \int_{\Lambda_r} [\int_\Lambda [\phi_p(y^{(r+1)})]^+\, d\mu\, dF(y^{(1)}, \cdots, y^{(r)}). \end{aligned}$$

First we have from (4.11) that

$$(4.13) \quad \begin{aligned} &\int_{\Lambda(n-r)} t_n^*(y^{(r+1)})\, dF(y_0, \cdots, y_{n-r-1} \mid y^{(1)}, \cdots, y^{(r)}) \\ &\qquad = \Pr\{\phi_{p_n}(y_0, \cdots, y_{n-r-1}; y^{(1)}, \cdots, y^{(r+1)})(y^{(r+1)}) > 0 \mid y^{(1)}, \cdots, y^{(r)}\}. \end{aligned}$$

Now since $\phi_p(y)$ is a continuous function of p and since we have assumed that

$$(4.14) \quad \Pr\{\lim_{n\to\infty} |p_n(y_0, \cdots, y_{n-r-1}, y^{(1)}, \cdots, y^{(r+1)}) - p| > \epsilon\} = 0$$

for every fixed $y^{(1)}, \cdots, y^{(r+1)}$, it follows that (4.11) is true: thus we have

$$(4.15) \quad [\int_{\Lambda(n-r)} t_n^*\, dF(y_0, \cdots, y_{n-r-1} \mid y^{(1)}, \cdots, y^{(r)})]\phi_p(y^{(r+1)}) \to [\phi_p(y^{r+1)})]^+.$$

It remains now only to prove that the entire integral in (4.9) converges to (4.12).

This will follow by the Lebesgue dominated convergence theorem once we show that the integrand is absolutely bounded by an integrable function. This is easy since the left hand side is bounded by $|\phi_p(y^{(r+1)})|$ and $|\phi_p(y^{(r+1)})|$ is integrable as follows

$$(4.16)\quad \int_\Lambda |(1-p)a_2 f_1(y^{(r+1)}|\, y^{(1)}, \cdots, y^{(r)}) - pa_1 f_0(y^{(r+1)}\,|y^{(1)}, \cdots, y^{(r)})|\, d\mu \leqq (1-p)a_2 + a_1 .$$

Thus the sequence $T^* = \{t_n^*\}$ given by (4.2) is *asymptotically* optimal if we can exhibit a sequence $\{p_n\}$ satisfying the required properties.

To this end let $F_\theta(y)$ be the conditional distribution of Y_n given $\theta_n = \theta$. (Note: the $\{Y_n\}$ all have the same marginal distribution.)

Let $h(x)$ be any function satisfying

$$(4.17)\qquad \int_\Lambda h(x)\, dF_\theta(x) = \theta, \qquad (\theta = 0, 1),$$

for example (following Robbins ([3], p. 198),

$$(4.18)\qquad \begin{aligned} h(x) &= [1 - P_0(B)]/[P_1(B) - P_0(B)] \qquad \text{if } x \in B \\ &= [-P_0(B)]/[P_1(B) - P_0(B)] \qquad \text{if } x \notin B \end{aligned}$$

where B is any event for which $P_0(B) \neq P_1(B)$.

Now define,

$$\tilde{p}_n(y_0, \cdots, y_n) = (n+1)^{-1} \textstyle\sum_{i=0}^n h(y_i).$$

Thus since the y's are marginally identically distributed with,

$$E[h(y_i)] = p.$$

Hence we must show that $\Pr\{\lim_{n\to\infty} \tilde{p}_n = p\} = 1$, i.e., (strong convergence). For if that is true then for any fixed $y^{(1)}, \cdots, y^{(r+1)}$ we have $n^{-1} \sum_{i=1}^{r+1} h(y^{(i)}) \to 0$ as $n \to \infty$, so that convergence still holds for the sequence,

$$\tilde{p}_n(y_0, \cdots, y_{n-r-1}, y^{(1)}, \cdots, y^{(r+1)}),$$

and furthermore the convergence will hold if we restrict p_n to satisfy

$$\begin{aligned} p_n &= 0 \qquad \text{if } \tilde{p}_n < 0 \\ &= \tilde{p}_n \qquad \text{if } 0 \leqq \tilde{p}_n \leqq 1 \\ &= 1 \qquad \text{if } \tilde{p}_n > 1, \end{aligned}$$

e.g., if $F_\theta(y)$ is Gaussian with mean θ. Then we can set $h(y) = y$, i.e.,

$$\tilde{p}_n = (n+1)^{-1} \textstyle\sum_{i=0}^n y_i$$

will satisfy the requirements after we prove the strong convergence of $\tilde{p}_n$.

In order to prove the strong convergence we need the following lemma (which must be well known) which we prove since we have no reference for the lemma.

LEMMA 4.1. *Let $\{Y_n\}$ be an r-dependent strictly stationary process, and let $h(Y)$ be any function with $E[|h(Y)|] < \infty$. Then*

$$[h(Y_0) + h(Y_1) + \cdots + h(Y_n)]/(n + 1) \to_{\text{a.s.}} E[h(Y)].$$

PROOF. From the ergodic theorem we know that

$$[h(Y_0) + h(Y_1) + \cdots + h(Y_n)]/(n + 1) \to_{\text{a.s.}} E^A[h(Y)]$$

where A is the invariant σ-field and expectation is conditional with respect to A.

Now $A \subset C$ where $C = \bigcap_{n=0}^{\infty} B_n$ with B_n as in Definition 2.1, is the tail σ-field of the sequence $\{Y_n\}$.

Thus to prove the lemma we must show that, a.s., $C = (\Lambda, \phi)$ i.e., the whole space and the null set. This follows from the fact that C is independent of itself, i.e., if D is any event $P(D \cdot D) = P(D) \cdot P(D)$ hence D is Λ or ϕ. First we have that C is contained in the tail σ-field of the sequence $\{Y_n\}$. But C_n is independent of B_{n+r+1} for all $n = 0, 1, 2, \cdots$. By the known result that limits of independent σ-fields are independent it follows that C is independent of the σ-field generated by $(Y_0, Y_1 \cdots)$ but as above it is also contained in that σ-field, hence the lemma follows.||

COROLLARY. *$\hat{p}_n$ as defined above converges* a.s. *to p.*

This completes the proof of the asymptotic optimality of the sequence $T^* = \{t_n^*\}$ as defined above.

In the remaining sections we will generalize most of these results and discuss applications to signal detection in correlated noise.

5. Extensions and remarks.

A. rth *order, discrete parameter Markov processes with general state space.* The first extension of the above results will be to rth order, discrete parameter Markov processes, with general state space. There is clearly no difficulty in carrying through the above results to this Markov case if one assumes that the initial distribution is the stationary absolute probability distribution. To sketch this proof we observe that the distribution of Y_n given the past depends only on the r previous observations, thus (4.5) holds. The remainder of the results are by formal manipulation, hence all results hold up to Lemma 4.1. But this lemma can be replaced by Theorem 6.1, (See Doob, [1], p. 219) if one assumes in addition that

(a) There is only one ergodic set

(b) *Doeblin's hypothesis:* (Hypothesis D, Doob)

Let $\mathfrak{F}_\Lambda$ be the Borel field of Λ sets, and let $p(\zeta, A)$, $\zeta \in \Lambda$, $A \in \mathfrak{F}_\Lambda$ be a probability measure for fixed $\zeta \in \Lambda$, and $p^{(\nu)}(\zeta, A)$ its νth iterate. Then there is a (finite-valued) measure ϕ of sets $A \in \mathfrak{F}_\Lambda$ with $Y(\Lambda) > 0$, an integer $\nu \geqq 1$, and an $\epsilon > 0$, such that

$$p^{\nu}(\zeta, A) \leqq 1 - \epsilon \quad \text{if} \quad \phi(A) \leqq \epsilon. \tag{5.1}$$

B. *Let θ_n assume a finite number of values.* The extension to this case follows immediately by direct analogue to Robbins' (see [4], p. 5) Corollary 1. This result also follows for arbitrary loss functions satisfying

$$\textstyle\sum_{i=1}^{n} L_{a_i}(\theta) p_i < \infty$$

where p_i is the *a priori* probability of θ assuming the ith value; we suppose there are n-possible values. The only difficulty that occurs in this case is the sequential estimation of the p_i, $(i = 1, \cdots, n)$. However, it develops that the construction scheme given by Robbins for this case (see [4], pp. 17–18) also carries through for the r-dependent stationary case, if we use the finite number of conditional densities of Y given θ generate the linear manifold and replace the strong law of large numbers by Lemma 4.1, (or in the Markov case by the ergodic theorem for Markov processes) to prove convergence of the estimates.

6. Applications to pulse detection in correlated noise. We will detail applications only for the simplest case discussed in Sections 3 and 4. Let $\{Y_n\}$ be a sequence of r-dependent stationary random variables and suppose

$$\begin{aligned} Y_i &= 1 + n_i \quad &&\text{if pulse say of unit amplitude is present} \\ &= n_i &&\text{if noise alone is present.} \end{aligned}$$

It is supposed that the $\{n_i\}$ sequence is r-dependent and strictly stationary (i.e., noise samples more than r units apart are uncorrelated). Thus the $\{Y_i\}$ $i = 0, 1, 2, \cdots$ is r-dependent and marginally strictly stationary, where the pulse train is random and pulses occur, or do not occur, independently from observation to observation. If the *a priori* probability of no pulse is p, and of pulse is $(1 - p)$ at each observation and p is unknown, we can use the procedure of Section 4, in particular the sequence $T^* = \{t_n^*\}$ as given by (4.2) and (4.3) which we proved was asymptotically optimal. For the estimating function we can use $h(y) = y$. If the noise is Gaussian the conditional densities have a convenient form as can be obtained from modifying expression (7.4.25) in Wilks [7] for example.

Furthermore, all remarks in Section 5 also carry over to these applications.

7. Acknowledgment. I would like to thank the referee for his patience and care in correcting the manuscript. I am indeed grateful.

REFERENCES

[1] Doob, J. L. (1953). *Stochastic Processes*. Wiley, New York.
[2] Loève, M. (1960). *Probability Theory* (2nd ed.). Van Nostrand, Princeton, New Jersey.
[3] Robbins, H. (1963). An empirical Bayes approach to testing statistical hypotheses. *Rev. Inst. Internat. Statist.* **41** 195–208.
[4] Robbins, H. (1964). The empirical Bayes approach to statistical decision problems. *Ann. Math. Statist.* **35** 1–20.
[5] Samuel, E. (1963). An empirical Bayes approach to the testing of certain parametric hypotheses. *Ann. Math. Statist.* **34** 1370–1385.
[6] Tainiter, M. (1964). Adaptive procedures for pulse detection and extraction. Johns Hopkins University Internal Note, C. Barton Lab., IMA-11.
[7] Wilks, S. S. (1962). *Mathematical Statistics*. Wiley, New York.