Change-point Problems
IMS Lecture Notes – Monograph Series (Volume 23, 1994)

OPTIMAL DETECTION OF A CHANGE IN DISTRIBUTION WHEN THE OBSERVATIONS FORM A MARKOV CHAIN WITH A FINITE STATE SPACE

BY BENJAMIN YAKIR
University of Rochester

Consider a process $X_1, X_2, \ldots$. When the process is in control the sequence of observations is distributed like a stationary Markov chain with a (known) probability transition matrix A. If the process goes out of control, the probability transition matrix becomes B (which is a known matrix as well). We describe the structure of the minimax policy (in the Pollak-Siegmund sense) and of the Bayes rule for detecting the change in the distribution.

1. Introduction. Suppose one is able to observe sequentially a series of observations $X_1, X_2, \ldots$ whose distribution possibly changes at some unknown point in time. The objective is to raise an alarm as soon as possible after the change-point, subject to a restriction on the rate of false alarm.

An important motivation to such a problem is the on-line quality control of a manufacturing process. Imagine a machine that produces some product. The machine might break-down at some point in time. The propose of an on-line quality control scheme is to determine, based on the observation of the manufacturing process, whether the machine is functioning properly or not. One would like to have a scheme which detects the break-down of the machine as fast as possible but does not stop the production if the machine is working well.

In the classical setting of this problem it is assumed that the observations are independent. Their common distribution function before the change is F_0. After the change the distribution function is a different distribution function F_1.

The assumption of independence of the observations might be too restrictive in many important applications, where some dependence structure between the observations is apparent. In this article we consider the simplest

AMS 1991 Subject Classification: Primary 62L15; Secondary 62N10.

Key words and phrases: Quality control, change-point detection, optimal stopping, Bayes rule, Markov chain.

of all dependence structures : the Markov chain process with a finite space of possible states.

In the next section the theory of the Bayesian formulation is developed for the problem of detection of a change in distribution in the Markov chain case.

Section 3 deals with the structure of the optimal detection policy in a minimax sense.

2. Bayes Rules. Let us consider the Bayesian formulation of a disruption in a Markov chain process. In order to do so, we have to specify the structure of the probability space and of the loss function. Let $(\Omega, \mathbf{F})$ be a measure space. On that space we are given random variables $\nu, X_0, X_1, \dots$ and a probability measure $P^{(\pi,x)}$ such that:

$$P^{(\pi,x)}(X_0 = x) = 1,$$

$$P^{(\pi,x)}(\nu = 1) = \pi$$

and:

$$P^{(\pi,x)}(\nu = k) = (1-\pi)pq^{k-2}, \quad \text{for } \ k \geq 2;$$

where p and π and x are known constants with $0 < p \leq 1$, $0 \leq \pi \leq 1$ and x one of the possible states of the process ($x \in \mathcal{X}$). The parameter ν is the (unknown) point of change of the process.

The distribution of the observations is determined by the value of ν. Conditional on the event $\{\nu = k\}$, the distribution of the random process $X_1, X_2, \dots$ satisfies:

$$P^{(\pi,x)}(x_1, \dots, x_n | \nu = k) = \begin{cases} \prod_{i=1}^{k-1} a(x_i|x_{i-1}) \prod_{i=k}^{n} b(x_i|x_{i-1}) & \text{if } k \leq n \\ \prod_{i=1}^{n} a(x_i|x_{i-1}) & \text{if } k > n. \end{cases}$$

where $x_0 = x$.

The matrices $A = (a(i|j))$ and $B = (b(i|j))$ are two transition probability matrices. The interpretation of the above is that when the process is in control, its distribution is like that of a Markov chain process with transition matrix A. If the process goes out of control, the transition matrix changes to the matrix B.

Let N be a stopping time adapted to the system of σ-algebras $\{\mathbf{F_n}\}_{n=0}^{\infty}$, where $\mathbf{F_0} = \{\emptyset, \Omega\}$ and $\mathbf{F_n} = \sigma(\mathbf{F_0}, X_1, \dots, X_n)$. The stopping rule N can be interpreted as the detection policy, i.e. the time at which the "alarm" is sounded to signal the change in distribution. It is desirable to choose N in such a way that it is as close as possible to the time of disruption ν. Following the formulation of Shiryayev [1978] (and the modification of Yakir [1991]), the risk associated with the detection policy N is:

$$\rho(N,\nu) = P^{(\pi,x)}(N < \nu - 1) + cE^{(\pi,x)}(N - \nu + 1)^+, \tag{1}$$

where $c > 0$ is a fixed constant that represents the relative cost of taking observations from the uncontrolled process.

DEFINITION: For a given pair $(\pi, x) \in [0,1] \times \mathcal{X}$ we call a stopping time N^* a (π, x)-Bayes time if

$$\rho(N^*,\nu) = \inf \rho(N,\nu),$$

where inf is taken over the class of all proper stopping times.

The following theorem characterizes the structure of the Bayes rule:

THEOREM 1. *Let $p > 0$, $c > 0$ and let:*

$$\pi_n = \pi_n^{(\pi,x)} = P^{(\pi,x)}(\nu - 1 \le n|\mathbf{F_n})$$

be the posterior probability that the next observation is governed by B. There exits a function $\delta(\cdot)$, defined on $\mathcal{X}$, such that the stopping time:

$$N^* = \inf\{n \ge 0 : \pi_n^{(\pi,x)} \ge \delta(X_n)\},$$

is the (π, x)-Bayes rule. Moreover, $\delta(\cdot)$ does not depend on π or on x.

Since the proof of the theorem is similar to Shiryayev's proof it is omited.

REMARK: The theorem remains correct when the initial pair (π, x) is random (according to a measure ϕ). Again, the threshold function does not depend on the initial state. (Notice that the stopping time does depend on the distribution of the initial state through the dependence on the initial state of the probability of a change.)

The structure of the Bayes rule plays a crucial role in the development of the optimal detection time in the non-Bayesian setting. Denote the Bayes problem described above by $B(c, p, \pi, x)$. If the pair (π, x) is random we denote the problem by $B(c, p, \phi)$.

It is convenient to reformulate the stopping time N^* in terms of a different sequence of statistics. It follows that:

$$\pi_n^{(\pi,x)} = \frac{R_{n,p}^{(r(\pi),x)}}{R_{n,p}^{(r(\pi),x)} + 1/p}, \tag{2}$$

where:

$$R_{n,p}^{(r,x)} = r\prod_{i=1}^{n}(Z_i/q) + \frac{1}{q}\sum_{k=2}^{n}\prod_{i=k}^{n}(Z_i/q) + \frac{1}{q}, \tag{3}$$

$$Z_i = Z(X_i|X_{i-1}) = b(X_i|X_{i-1})/a(X_i|X_{i-1})$$

and:

$$r(\pi) = \frac{\pi}{(1-\pi)p}. \tag{4}$$

The function $y/(y+1/p)$ is a monotone function in y, hence the Bayesian stopping time can be rewritten in terms of $R_{n,p}^{(r,x)}$:

$$\begin{aligned} N^* &= \inf\{n \geq 0 : \pi_n^{(\pi,x)} \geq \delta(X_n)\} \\ &= \inf\{n \geq 0 : R_{n,p}^{(r(\pi),x)} \geq \Delta(X_n)\}, \end{aligned}$$

for $\Delta(\cdot) = r(\delta(\cdot))$.

For a given $c > 0$, and for all $p > 0$, let $N_p = N^*$ be the the Bayes rule for the Bayes problem $B(c,p,\pi,x)$, that was discussed in the previous section. Our first aim is to find a sequence of p's, that converge to 0, for which the stopping times N_p converge to an appropriate stopping time. Furthermore, for technical reasons, we want all the stopping times in the sequence to be bounded by some stopping time with finite expectation. The main tool to achieve such goal will be the following lemma:

LEMMA. *Consider the problem $B(c,p,p,x_0)$, where x_0 is a recurrent state of the Markov chain process governed by A. There exists a constant D and some $0 < q_0 < 1$, such that for all $q_0 \leq q \leq 1$ and for all threshold functions $\Delta(\cdot)$ with the property that $\Delta(x) \geq D$, for each $x \in \mathcal{X}$, the average delay it takes the stopping time:*

$$N_{p,\Delta} = \inf\{n \geq 0 : R_{n,p}^{(r(p),x_0)} \geq \Delta(X_n)\},$$

to detect the change, satisfies:

$$E^{(p,x_0)}(N_{p,\Delta} - \nu + 1 | N_{p,\Delta} \geq \nu - 1) > 2c^{-1}. \tag{5}$$

A sketch of a proof: Define a sequence of stopping times $T_1, T_2, \ldots$ by the recursive formula:

$$T(0) = 0,$$

$$T(j) = \inf\{n > T(j-1) : X_n = x_0\}.$$

For a given j, Let:

$$U_j = \prod_{i=T(j-1)+1}^{T(j)} (Z_i/q),$$

and:

$$V_j = \sum_{k=T(j-1)+2}^{T(j)} \prod_{i=k}^{T(j)} (Z_i/q) + \frac{1}{q}.$$

(If $T(j) = T(j-1)+1$ then $V(j) = 1/q$.)

Let $P_\infty^{x_0}$ be the distribution of the Markov chain process, with initial state x_0 and probability transition matrix A. Let $E_\infty^{x_0}(\cdot)$ be the expectation operator of that distribution. Notice that the sequence $\{U_j, V_j\}_{j=1}^{\infty}$ is a sequence of independent and identically distributed random vectors under the regime $P_\infty^{x_0}$. Furthermore, there exists a constant $0 < q_0 < 1$ such that:

$$E_\infty^{x_0} \log U_1 < \eta < 0 \tag{6}$$

and:

$$E_\infty^{x_0} V_1 < \infty \tag{7}$$

uniformly in p, for $q_0 \le 1-p \le 1$.

Consider the sequence $R(1), R(2), \ldots$, where:

$$R(j) = R_{T(j),p}^{(r(p),x_0)}.$$

It is easy to see that:

$$R(j) = V_j + U_jV_{j-1} + U_jU_{j-1}V_{j-2} + \ldots + U_jU_{j-1}\ldots U_2V_1.$$

For each j the distribution of $R(j)$ is the same as the distribution of:

$$H(j) = \sum_{k=1}^{j} U_1U_2\ldots U_{k-1}V_k.$$

Using a similar argument to the one given in Pollak [1985, Lemma 3], one can show that $H(j)$ converges to the ($P_\infty^{x_0}$-a.s. finite) random variable:

$$H = \sum_{k=1}^{\infty} U_1U_2\ldots U_{k-1}V_k.$$

The rest of the proof of the lemma follows along the lines of Lemmas 5 and 6 in Pollak [1985, pp. 211-212]. The only modification needed is to consider distributions on the space $[0,\infty) \times \mathcal{X}$, rather then the space $[0,\infty)$.

The following corollaries are easy results of Lemma 1. Notice first that the constant D does not depend on the initial state (p, x_0).

COROLLARY 1: For a given $\tilde{x} \in \mathcal{X}$ define:

$$N_{p,D,y}^{\tilde{x}} = \inf\{n : R_{n,p}^{(0,\tilde{x})} \geq D, X_n = y\},$$

and let:

$$N_{p,D,y} = \sup_{x \in \mathcal{X}} N_{p,D,y}^{x}.$$

It can be concluded that for each $0 \leq p \leq 1 - q_0$ there exists $y = y(p)$ such that

$$N_p \leq_{\text{a.s.}} N_{p,D,y} \leq_{\text{a.s.}} N_{1,D,y}, \tag{8}$$

where N_p is the Bayes rule for the problem $B(c, p, \pi, x)$.

COROLLARY 2: There exists a state y_0 and a subsequence of p's, such that (8) is true with $y(p) = y_0$, for all the p's in the subsequence.

COROLLARY 3: Let $\Delta_p(\cdot)$ be the threshold function of the problem $B(c, p, \pi, x)$, and assume that $\Delta_p(\cdot) \to_{p \to 0} \Delta_0(\cdot)$ for some function $\Delta_0(\cdot)$. Assume further that the convergence is along the subsequence of p's from the previous corollary, then:

$$\Delta_0(y_0) \leq D.$$

REMARK: Lemma 1 and the corollaries of the lemma remain true when $c = c(p)$ is allowed to vary with p, as long as $\liminf_{p \to 0} c(p) > 0$. In particular they are correct if $c(p)$ converges to some positive c.

Let $(1 - \rho(N, \nu))/p$ be the normalized risk of a stopping time N. Using the results of the last lemma we can show that for $p \to 0$, the (normalized) risk of a converging sequence of stopping times goes to a limit.

3. Minimax Detection Policies. After understanding the structure of the Bayes rules for detecting a change in the Markov chain process and the characteristics of the limits of such rules, we can turn our attention to the problem of detecting a change in a non-Bayesian setting. Let us start by describing the model we have in mind.

Let $(\Omega, \mathbf{F})$ be a measure space. On that space we are given a sequence of random variables $X_0, X_1, X_2, \ldots$, each one of them has its values in the finite space of states $\mathcal{X}$. Let A and B be two transition probability matrices. Consider the sequence of probability measures, denoted by $\tilde{P}_k(\cdot)$, for $k = \infty, 1, 2, \ldots$. Under the regime $\tilde{P}_\infty(\cdot)$ the sequence of observations form a stationary Markov chain process with the transition probability matrix A. (That is to say that the distribution Λ of X_0 is stationary under the transformation A.) For $1 \leq k < \infty$, the distribution $\tilde{P}_k(\cdot)$ is such that the distribution of the observations $X_0, \ldots, X_{k-1}$ is $\tilde{P}_\infty(\cdot)$, and the observations $X_k, X_{k+1}, \ldots$ form a Markov chain process with transition matrix B and initial state X_{k-1}.

Let $\tilde{E}_k(\cdot)$ be the expectation operator of the distribution $\tilde{P}_k(\cdot)$. It is assumed that the statistician knows what the starting value of the process (X_0) is.

Consider controlling the process of the manufacturing of some product. A situation where the above formulation is reasonable is, for example, when at the beginning of a production cycle, the repaired machine is operated for a while under close inspection. If it performs well at the end of that inspection period, regular production is continued, and the process control policy we have in mind goes into effect. Since the machine was operating for a while, the distribution of the state of production, just before the regular production starts, is approximately the stationary distribution of the controlled process.

A change-point detection policy is a stopping time, adapted to the sequence of observations $X_0, X_1, \ldots$. An optimal detection policy, or a minimax detection policy, is a policy that minimizes the maximal average delay in detection:

$$\sup_{1 \leq k < \infty} \tilde{E}_k(N - k + 1 | N \geq k - 1), \tag{9}$$

among all policies that satisfy a constraint on the rate of false alarm:

$$\tilde{E}_\infty N \geq \beta, \tag{10}$$

for a given constant β. In this section we will characterize the structure of the optimal policy.

For a given set of non-negative boundary points $\Delta = \{\Delta(x)\}_{x \in \mathcal{X}}$ (infinity is not excluded), consider the set:

$$S_\Delta = \{(r, x) : x \in \mathcal{X},\ \ 0 \leq r \leq \Delta(x)\}.$$

Let $\mathcal{F}_\Delta$ be the set of distribution functions, with support in S_Δ. Let T_Δ be the transformation defined by:

$$T_\Delta F(r, x) = \frac{1}{Q(F)} \sum_{y \in \mathcal{X}} \int_0^{\Delta(y)} I(s(\frac{b(x|y)}{a(x|y)} + 1) \leq r) a(x|y) dF(s, y),$$

where:

$$Q(F) = \sum_{x,y \in \mathcal{X}} \int_0^{\Delta(y)} I(s(\frac{b(x|y)}{a(x|y)} + 1) \leq \Delta(y)) a(x|y) dF(s, y).$$

It can be shown that with each Δ there associates a set of invariant measures Φ_Δ, i.e. $T_\Delta \phi = \phi$ for all $\phi \in \Phi_\Delta$. For each such ϕ a detection policy can be defined in the following way: Let $\tilde{\phi}$ be the transformed distribution, defined on $R^+ \times \mathcal{X}$ by the relation:

$$\tilde{\phi}(r,x) = \sum_{y \in \mathcal{X}} \int_0^{\Delta(y)} I(s(\frac{b(x|y)}{a(x|y)} + 1) \leq r) a(x|y) d\phi(s,y).$$

Given the value of the initial state $X_0 = x_0$, simulate a random vector (R_0, x_0) from the distribution $\tilde{\phi}$, conditioned on the event $\{X_0 = x_0\}$. Define sequentially:

$$R_n = R_{n-1} b(X_n|X_{n-1})/a(X_n|X_{n-1}) + 1.$$

The detection policy is:

$$N^{\phi} = \inf\{n \geq 0 : R_n \geq \Delta(X_n)\}.$$

If the $\tilde{\phi}$-distribution has atoms on the boundary Δ, we allow randomized stopping times in the sense that if the statistic falls exactly on Δ then the decision onto whether to continue sampling or not is done by some random law. This law can depend on the boundary point, but not on time. In this case, thus, we have a family of policies, associated with the invariant measure ϕ.

Notice that each one of these detecting policies is an "equalizer rule" in the sense that:

$$\tilde{E}_k(N^{\phi} - k + 1 | N^{\phi} \geq k - 1) = \tilde{E}_1 N^{\phi},$$

for all $k \geq 1$. The same is true for the case where $\tilde{\phi}$ has atoms on the boundary, since the randomization law is time independent.

For a given β, let $\mathcal{N}_\beta$ be the set of all detecting policies N^{ϕ}, of the above form, for which $\tilde{E}_\infty N^{\phi} = \beta$. In the next lemma it is shown that $\mathcal{N}_\beta$ is not empty. Furthermore, this set contains a stopping rule that is a limit of Bayes stopping rules:

LEMMA 2. *There exists a sequence of p's that converge to* 0, *a sequence of randomized Bayes problems* $B(c(p), p, \tilde{\phi}_p)$ *with the appropriate Bayes rules* $\tilde{N}(c(p), p, \tilde{\phi}_p)$, *a detection policy* N^{ϕ} *of the above form and a constant* $0 < c < \infty$ *such that:*

(i) $c(p) \longrightarrow_{p \to 0} c$.

(ii) $\tilde{E}_\infty N^{\phi} = \beta$.

(iii) $\varrho^{\tilde{\phi}_p}(\tilde{N}(c(p), p, \tilde{\phi}_p)) \longrightarrow_{p \to 0} (\mu(\phi) + \beta)(1 - c\tilde{E}_1 N^{\phi})$,

where $\mu(\phi) = \sum_{x \in \mathcal{X}} \int_0^\infty r d\tilde{\phi}(r,x)$, *and* $\varrho^{\tilde{\phi}_p}(\tilde{N}(c(p), p, \tilde{\phi}_p))$ *is the normelized Bayes risk.*

PROOF: For each p, it can be shown that one can choose a Bayes problem $B(c(p), p, \tilde{\phi}_p)$, such that the appropriate Bayes rule $\tilde{N}(c(p), p, \tilde{\phi}_p)$ satisfies:

$$\tilde{E}_\infty \tilde{N}(c(p), p, \tilde{\phi}_p) = \beta.$$

Furthermore, for each p, the measure $\tilde{\phi}_p$ can be picked in such a way that:

$$E^{\tilde{\phi}}(\tilde{N} - \nu + 1)^+ = \tilde{E}_1 \tilde{N} \cdot P^{\tilde{\phi}}(\tilde{N} \geq \nu - 1),$$

where $\tilde{E}_1(\cdot) = E^{\tilde{\phi}}(\cdot|\nu = 1)$. (This corresponds to the Bayesian version of the "equalizer rule".)

The first task is to show that the sequence $c(p)$ is bounded away from zero and from infinity. Consider the Bayes problems $B(c, p, 0, x)$, for $x \in \mathcal{X}$. Let $c_x(p)$ be the cost for which the P^x_∞-expectation of the Bayes rule is β. It is easy to see that the sequence $c_x(p)$ is bounded away from infinity for each x. Since $c(p) \leq \sup_{x \in \mathcal{X}} c_x(p)$ we get that the sequence $c(p)$ is bounded away from infinity as well.

In order to prove that the sequence of costs is bounded away from zero, notice that all the invariant measures ϕ, associated with the couple (c, p), are uniformly tight for all $p \leq 1 - q_0$ and all c. Hence, the function $Q(\phi)$ converges to 1 as $c \to 0$, uniformly in p, for $p \leq 1 - q_0$. But it is assumed that $\tilde{E}_\infty \tilde{N}(c(p), p, \tilde{\phi}_p) = \beta$, and hence it is bounded away from infinity for all p's, therefore c must be bounded away from zero.

Let c be an accumulation point of the sequence $c(p)$. Assume that the threshold functions $\Delta_{c(p)}(\cdot)$ converge to a function $\Delta(\cdot)$. (If the assumption does not hold for the given sequence – choose a subsequence for which it does.) Furthermore, without loss of generality it can be assumed that:

$$\sup_{0 \leq p \leq 1 - q_0} \Delta_{c(p)}(y) < \infty, \tag{11}$$

for some $y \in \mathcal{X}$. This follows from Lemma 1 and its corollaries.

Consider next the sequence of invariant measures ϕ_p. Since this sequence is uniformly tight, it converges to a measure ϕ. It can be shown that this measure is invariant under the transformation T_Δ. By the Representation Theorem (see Pollard [1984, sec. iv.3.] it can be assumed that the random variables, that carry the distributions ϕ_p, converge almost surely to a random variable that carry the distribution ϕ. Hence, the stopping times $\tilde{N}(c(p), p, \tilde{\phi}_p)$ converge almost surely to a stopping time N^ϕ. By corollary 1 of Lemma 1 it follows that all these stopping times are uniformly bounded by an appropriate stopping time. Claims (ii) and (iii) follow from Lesbegue's Dominated Convergence theorem.

After showing that the set $\mathcal{N}_\beta$ contains a limit of Bayes rules, it is an easy task to show that this limit is the minimax detection policy. This result is stated in the next theorem:

THEOREM 2. *Let N^ϕ be a stopping time from the set $\mathcal{N}_\beta$, that minimizes $\tilde{E}_1 N$ among all stopping times N from that set. The change-point detection policy N^ϕ is a minimax policy in the sense of equations (9) and (10).*

The proof of the theorem is almost identical to the proof of Theorem 3 in Yakir [1991], and is thus omited. Notice that a limit of Bayes rules minimizes $\tilde{E}_1 N$ among all stopping times in the set $\mathcal{N}_\beta$, hence the claim of the above theorem is not empty.

4. Concluding Remarks And Directions For Further Research. In the last theorem it was shown that a minimax detection policy can be constructed by solving a minimization problem, involving the $\tilde{P}_1$-expectation of a stopping time of a given form, restricted by a constraint on the value of the $\tilde{P}_\infty$-expectation of the stopping time. Actually, by using the same methods as in Theorem 4 in Yakir [1991], the problem can be translated to a problem concerning invariant measures. It can be shown that an optimal policy can be found by solving the problem of minimizing the functional $\mu(\phi)$ among all the invariant measures of the transformation T_Δ and all boundary points Δ, for which $Q(\phi) = \beta/(\beta+1)$.

The solution of such a problem is feasible with the aid of Markov chain simulation techniques. We hope to demonstrate this approach in the near future.

Another problem is the comparison between the optimal detection policy and the policy suggested by Pollak & Yahav [1991]. The two policies are distinguished by two aspects: the structure of the boundary function and the distribution of the starting values of the statistics (R, X). Pollak & Yahav's boundary function is constant and the starting values is a fixed point $(0, x)$, for some $x \in \mathcal{X}$.

Consider the asymptotics (as $\beta \to \infty$) of the difference between the average delay in detection of their policy and of the optimal policy. Our conjecture is that the fact that one starts from a fixed starting value, rather than from the appropriate invariant distribution will affect the difference by a positive constant value. This phenomenon was observed in the independent random variables case (See Yakir [1992]). Furthermore, we conjecture that the effect of choosing a constant threshold, rather than a threshold that depends on the state of the process, is a $o(1)$ term when $\beta \to \infty$. Thus far, we were not able to prove either of these conjectures.

Another interesting area for further research is the problem of finding an optimal detection policy when the transition matrix after the change is unknown. A simple model, where the set of possible transition matrices B is finite, was considered by Pollak & Yahav. The solution they suggested, of assigning a prior distribution to the different matrices, seems to be promising.

5. An Example. We give an example of a change-point problem in a Markov chain process in which the conjectures of the previous section are true. Consider the following problem: Let $X_1, X_2, \ldots$ be a process with values in $\mathcal{X} = \{0, 1\}$. The probability transition matrix before the change is: $a(i|j) = 1/2$ for $i, j \in \mathcal{X}$. After the change it is: $b(1|1) = b(0|1) = 1/2$, $b(0|0) = 1$ and $b(1|0) = 0$. The state 0 is an absorbing state. The Pollak-Yahav policy is to observe the process until the Shiryayev-Roberts statistic crosses a constant boundary.

Given any pair of boundary points $(\Delta(0), \Delta(1))$, it follows that the invariant distribution is unique (up to randomization on the boundary) and its support is in the set $\{(n, m) : n, m \in \mathbf{N}, 2 \leq n < \Delta(0), 1 \leq m < \Delta(1)\}$. Consider the policy N_Δ based on the boundary points (Δ, ∞), where Δ is uniquely defined by β. The conditional distribution of the invariant measure, given the event $X_0 = 1$, is stochastically bounded by a geometric random variable, hence it follows that the policy N_Δ is almost optimal (up to a $o(1)$ term, where $o(1) \to 0$ as $\beta \to \infty$).

Let N_A be the policy that stops the first time the statistic R_n is larger than A and $X_n = 0$. ($R_0 = 1$.) It is just as easy to show that the difference between the behavior of the Pollak-Yahav policy, and the policy N_A is negligible provided that β is big enough. (Again the threshold A is determined by β.) We will proceed by examining the difference between N_Δ and N_A.

For the policy N_A, since the event $\{R_n = 1, X_n = 1\}$ is positive-recurrent, it follows that:

$$E(N_A | X_0 = 1) = \frac{E(M_A(1) | X_0 = 1)}{P(R_{M_A(1)} \geq A | X_0 = 1)}, \tag{12}$$

where:

$$M_A(1) = \inf\{n \geq 1 : (R_n = 1, X_n = 1) \text{ } or \text{ } (R_n \geq A, X_n = 0)\}.$$

Similar arguments and some algebra lead to the conclusion:

$$EN_A = 2A(1 + o(1)) \text{ and } EN_\Delta = 2\Delta(1 + o(1)).$$

Choosing $A = \Delta = \beta/2$ it can be shown that:

$$E_1(N_A - N_\Delta) \longrightarrow_{\beta\to\infty}$$

$$E(\log_2(T + R_0 + 1) - \log_2(T + 1)|X_0 = 1) + E(\log_2(R_0 + 1)|X_0 = 1), \quad (13)$$

where the distribution of R_0 is the limit of the distributions ϕ_Δ as $\Delta \to \infty$ and T is a *Geometric*$(1/2)$ random variable, independent of R_0.

Acknowledgment. I would like to thank M. Pollak for his suggestions and for some helpful discussions.

REFERENCES

DUNFORD, N. and SCHWARTZ, J. T. [1958]. *Linear Operator, Part 1.* Interscience, New-York.

KESTEN, H. [1973]. Random difference equations ane renewal theory for products of random matrices. *Acta Math.* **131**, 207–248.

LORDEN, G. [1971]. Procedures for reacting to a change in distribution. *Ann. Math. Statist.* **42**. 1897–1908.

MOUSTAKIDES, G. V. [1986]. Optimal stopping times for detecting a change in distribution. *Ann. Statist.* **14**, 1379–1387.

POLLARD, D. [1984]. *Convergence of Stochastic Processes.* Springer-Verlag, New York.

POLLAK, M. [1985]. Optimal detection of a change in distribution. *Ann. Statist.* **13**, 206–227.

POLLAK, M. [1987]. Average run length of an optimal method of detecting a change in distribution. *Ann. Statist.* **15**, 749–779.

POLLAK, M. and SIEGMUND, D. [1985]. A diffusion process and its application to detecting a change in the drift of a Brownian motion process. *Biometrika* **72**, 267–280.

POLLAK, M. and YAHAV, J. A. [1992]. Surveillance of Markov chain with a finite state space. The Hebrew University of Jerusalem Technical report.

RITOV, R. [1990]. Decision theoretic optimality of the CUSUM procedure. *Ann. Statist.* **18**, 1464–1469.

SHIRYAYEV, A. N. [1978]. *Optimal Stopping Rules.* Springer-Verlag, New-York.

YAKIR, B. [1991] Optimal detection of a change in distribution, when the observations are independent. The Hebrew University of Jerusalem Technical report.

YAKIR, B. [1992] An explicit description of the Bayes rule for detecting a change in distribution. Mathematical Sciences Research Institute at Berkeley Technical report.

DEPARTMENT OF BIOSTATISTICS
UNIVERSITY OF ROCHESTER
MEDICAL CENTER, PO BOX 630
ROCHESTER, NY 14642