Ann. Inst. Statist. Math.
Vol. 48, No. 2, 229–246 (1996)

EMPIRICAL BAYES DETECTION OF A CHANGE IN DISTRIBUTION

ROHANA J. KARUNAMUNI* AND SHUNPU ZHANG

Department of Mathematical Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2G1

(Received August 22, 1994; revised March 31, 1995)

Abstract. The problem of detection of a change in distribution is considered. Shiryayev (1963, *Theory Probab. Appl.*, **8**, pp. 22–46, 247–264 and 402–413; 1978, *Optimal Stopping Rules*, Springer, New York) solved the problem in a Bayesian framework assuming that the prior on the change point is Geometric (p). Shiryayev showed that the Bayes solution prescribes stopping as soon as the posterior probability of the change having occurred exceeds a fixed level. In this paper, a myopic policy is studied. An empirical Bayes stopping time is investigated for detecting a change in distribution when the prior is not completely known.

Key words and phrases: Empirical Bayes, change points, Bayes sequential rules, stopping times, statistical process control.

1. Introduction

The change-point problem has received considerable attention in recent years. The methods with which this problem and related problems have been dealt are fixed-sample and sequential methods. A review which includes both the fixed-sample and the sequential methodology, and comprehensive bibliography can be found in Zacks (1983). An excellent review, in particular, on sequential methods is given in the recent article of Zacks (1991).

Procedures of sequential detection of changes in distribution laws are of special importance for statistical process control. When a process is "in control", observations are distributed according to F_0. At an unknown point in time, the process jumps "out of control" and ensuing observations are distributed according to F_1. The aim is to raise an alarm "as soon as possible" after the process jumps "out of control". Early work in this area is due to Shewhart (1931).

The sequential methods are classified into three categories: Bayes sequential procedures, cumulative sum procedures, and tracking methods. The interest of this

* Research was supported in part by the Natural Sciences and Engineering Research Council of Canada under grant GP 7987.

paper is focused on Bayes sequential methods. Bayes stopping rules for the present problem were studied by Shiryayev (1963*a*, 1963*b*, 1978) and Bather (1963, 1967), and later developed by Zacks and Barzily (1981), Zacks (1983), Pollak (1985) and others.

Consider a sequence of independent observable random variables $X_1, X_2, \ldots, X_{m-1}, X_m, \ldots$, where, conditional on m, $X_1, \ldots, X_{m-1}$ have an identical distribution function $F_0(x)$, and (conditional on m) $X_m, X_{m+1}, \ldots$ have an identical distribution function $F_1(x)(\neq F_o(x))$. In the present paper we assume that the two distributions are known, but the index of shift (change point), m, is *unknown*. The sequence of variables is observed sequentially. Let N be a *stopping variable* associated with a detection procedure which stops soon after the shift occurs without too many "false alarms". (It is generally assumed that observations resume immediately on the same process if the alarm is false; a correct alarm terminates the process.) Let $\pi(\cdot)$ denote the prior distribution of the shift index m. Assume that the penalty for stopping before the change-point is 1 (loss unit), and the penalty for delayed stopping after the change-point is c (loss unit) per observation. The Bayesian sequential problem is then to find a stopping rule N for which the prior risk

$$R(\pi, N) = P_\pi(N < m) + cE_\pi\{\max(N - m, 0)\} \tag{1.1}$$

is minimized; see Shiryayev (1963*a*, 1963*b*). Also, see Chapter 4 of Shiryayev (1978). The quantity $P_\pi(N < m)$ is naturally interpreted as the probability of a false alarm and $E_\pi\{\max(N - m, 0)\}$ as the average delay of detecting the occurrence of disruption correctly, i.e., when $N \geq m$.

The shift index m is considered a realization of a random variable θ, having a prior distribution $\pi(\theta)$ concentrated on the set of nonnegative integers $\{0, 1, 2, \ldots\}$. Specifically, Shiryayev assumed that the prior probability function is of the form

$$\pi(\theta) = \begin{cases} \pi, & \text{if } \theta = 0 \\ (1-\pi)p(1-p)^{j-1}, & \text{if } \theta = j \geq 1, \end{cases} \tag{1.2}$$

$0 < \pi < 1$, $0 < p < 1$. The optimal stopping (detection) time for the Bayes criterion (1.1) based on the prior (1.2) is given by (Shiryayev (1963*a*, 1963*b*, 1978))

$$\tau^*_{opt} = \inf\{n; \pi_n(\boldsymbol{X}_n) \geq A^*\},$$

where $\pi_n(\boldsymbol{X}_n)$ is the posterior probability of $\{\theta \leq n\}$, given $\mathcal{F}_n = \sigma\{\pi, X_1, \ldots, X_n\}$, and A^* $(0 < A^* < 1)$ is a function of (c, π, p). An explicit expression of A^* in terms of (c, π, p) is not given by Shiryayev. Only a Wiener process approximation has been obtained. Bather (1967) proved that, for the variational problem (i.e., minimize $E_\pi\{N - m \mid N \geq m\}$ under the constraint $P_\pi(N < m) \leq \alpha$, $0 < \alpha < 1$), if the expected number of false alarms is bounded by B, then $A^* = (B+1)^{-1}$. Motivated by this fact and due to the Markovian structure of the optimal rule τ^*_{opt}, in this paper we study a stopping rule τ^* of the form

$$\tau^* = \inf\{n \geq 1; \pi_n(\boldsymbol{X}_n) \geq p/(p+c)\}.$$

Clearly, if $B = p^{-1}(c-p)$ with $c > p$, then τ^* is the optimal rule for the variational problem. The stopping rule τ^* is equivalent to the one-step look-ahead stopping rule for the present problem; see, e.g., Zacks (1991) or Zacks and Barzily (1981). Optimality of myopic rules, such as τ^*, have been examined by Chow and Robbins (1961) and Abdel-Hameed (1977), among others.

Assume now that the prior (1.2) is not completely known. Then the stopping rules τ^*_{opt} and τ^* are not available. One would still like to approximate the optimal and suboptimal procedures. But some auxiliary information is needed. Suppose that one has available a number of independent observations $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_\nu$, where $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{ik_i})$. Conditional on θ_i, the observations $Y_{i1}, \ldots, Y_{i,\theta_i-1}$ are independent with distribution F_0 and are independent of $Y_{i\theta_i}, \ldots, Y_{ik_i}$ which (conditional on θ_i) are independent with distribution F_1, $1 \le k_i < \infty$, $\nu \ge i \ge 1$. Furthermore, we assume that $(\theta_1, \boldsymbol{Y}_1), \ldots, (\theta_\nu, \boldsymbol{Y}_\nu)$ are independent of (θ, X)'s, $\nu \ge 1$. The θ_i's are unobservables, as is θ. Then it may be possible to use empirical Bayes methods to estimate unknown parameters of the prior and to construct a stopping rule for the present detection problem. The proposed empirical Bayes (EB) stopping time is given in Section 3. The performance of the EB stopping rule w.r.t. the rule τ^* is measured by comparing their respective Bayes risks. In particular, we show that the Bayes risk of the EB stopping time is asymptotically smaller than that of stopping rule τ^* as ν (-number of independent data vectors available from the past) goes to infinity. Broadly speaking, this exhibits that the EB stopping time is comparable and may be even better in performance compared with the component stopping time τ^* when ν is large. This result is given in Section 3 as well. Proofs of the main results are deferred to Section 4.

2. The Bayesian detection procedure

In this section we develop the stopping rules τ_{opt} and τ^* discussed above in detail. The exposition follows that of Zacks ((1991), §2.2).

The observable random variables $X_1, X_2, \ldots$ are defined on a probability space $(\Omega, \mathcal{F})$, on which a family of probability measures $\{P_\pi, 0 < \pi < 1\}$, induce for $(X_1, \ldots, X_n)$ with $n \ge 1$, joint predictive c.d.f.

$$F_\pi(x_1, \ldots, x_n) = (\pi + (1-\pi)p) \prod_{i=1}^{n} F_1(x_i) + (1-\pi)p \sum_{j=1}^{n-1} (1-p)^j \cdot \prod_{i=1}^{j} F_0(x_i) \prod_{i=j+1}^{n} F_1(x_i) + (1-\pi)(1-p)^n \prod_{i=1}^{n} F_0(x_i). \tag{2.1}$$

Let $X_0 = 0$ and $\mathcal{F}_n = \sigma\{\pi, X_0, X_1, \ldots, X_n\}$ be the σ-field generated by the first n observations. Let $f_0(x)$ and $f_1(x)$ denote the p.d.f.'s corresponding to F_0 and F_1, respectively. Then the posterior probability function of θ on $\{n, n+1, \ldots\}$, given $x_1, \ldots, x_n$, is

$$\pi_n(\theta) = \begin{cases} \pi_n, & \text{if } \theta \le n \\ (1-\pi_n)p(1-p)^{\theta-n-1}, & \text{if } \theta \ge n+1, \end{cases} \tag{2.2}$$

where

$$\pi_n = P_\pi\{\theta \le n \mid \mathcal{F}_n\} \tag{2.3}$$
$$= \frac{(\pi + p(1-\pi)) \prod_{i=1}^{n} f_1(x_i) + (1-\pi)p \sum_{j=1}^{n-1} (1-p)^{j-1} \prod_{i=1}^{j} f_0(x_i) \prod_{i=j+1}^{n} f_1(x_i)}{D_n}$$

with

$$\begin{aligned} D_n = {} & (\pi + p(1-\pi)) \prod_{i=1}^{n} f_1(x_i) \\ & + (1-\pi)p \sum_{j=1}^{n-1} (1-p)^j \prod_{i=1}^{j} f_0(x_i) \prod_{i=j+1}^{n} f_1(x_i) \\ & + (1-\pi)(1-p)^n \prod_{i=1}^{n} f_0(x_i). \end{aligned} \tag{2.4}$$

Dividing the numerator and the denominator of π_n by $(1-\pi)\prod_{j=1}^{n} f_0(x_j)$, we get

$$\pi_n = \frac{\frac{\pi}{1-\pi} \prod_{i=1}^{n} R(x_i) + p \sum_{j=0}^{n-1} (1-p)^j \prod_{i=j+1}^{n} R(x_i)}{\frac{\pi}{1-\pi} \prod_{i=1}^{n} R(x_i) + p \sum_{j=0}^{n-1} (1-p)^j \prod_{i=j+1}^{n} R(x_i) + (1-p)^n}, \tag{2.5}$$

where $R(x_i) = f_1(x_i)/f_0(x_i)$ is the likelihood ratio. Let $q_n = 1 - \pi_n$. For $n = 0, 1, \dots$ one can write

$$q_{n+1} = \frac{(1-\pi)(1-p)^{n+1}}{R(x_{n+1})\{D_n - (1-\pi)(1-p)^n\} + B_{n+1}} \tag{2.6}$$

where

$$B_{n+1} = R(x_{n+1})(1-\pi)(1-p)^n p + (1-\pi)(1-p)^{n+1}. \tag{2.7}$$

But $(1-\pi)(1-p)^n = q_n D_n$. Hence, we obtain the recursive formulae

$$q_{n+1} = \frac{q_n(1-p)}{R(x_{n+1})(1 - q_n(1-p)) + q_n(1-p)}, \tag{2.8}$$

or, for $n = 0, 1, \dots,$

$$\pi_{n+1} = \frac{(\pi_n + (1-\pi_n)p)R(x_{n+1})}{(\pi_n + (1-\pi_n)p)R(x_{n+1}) + (1-\pi_n)(1-p)}, \tag{2.9}$$

where $\pi_0 = \pi$, $q_0 = 1 - \pi$. Then the optimal stopping (detection) time for the Bayes criterion (1.1) w.r.t. the prior (1.2) is given by (Shiryayev (1963*a*, 1963*b*))

$$\tau^*_{opt} = \inf\{n; \pi_n \ge A^*\}, \tag{2.10}$$

where $0 < A^* < 1$ is a constant depending on c, π and p only. We consider the following stopping rule

$$\begin{aligned}\tau^* &= \inf\{n \geq 1; \pi_n \geq p/(p+c)\} \\ &= \inf\{n \geq 1; c\pi_n - p(1-\pi_n) \geq 0\}.\end{aligned} \tag{2.11}$$

Finally, for later use, it is relevant that $\{\pi_n, \mathcal{F}_n^*, P_\pi\}$ is a submartingale, for $n \geq 1$,

$$E_\pi\{\pi_{n+1} \mid \mathcal{F}_n^*\} = \pi_n + p(1-\pi_n) \quad \text{a.s. } (P_\pi), \tag{2.12}$$

where $\mathcal{F}_n^* = \sigma\{\pi_0, \pi_1, \ldots, \pi_n\}$, and for any stopping time τ, the prior risk (cf. (1.1)) takes the form

$$R(\pi, \tau) = E_\pi\left\{(1-\pi_\tau) + c\sum_{k=1}^{\tau-1}\pi_k\right\} \tag{2.13}$$

with π_k given by (2.5), $k \geq 1$. Furthermore, it is easy to show that

$$P_\pi(\tau^* < \infty) = 1 \quad \text{and} \quad E_\pi\tau^* < \infty. \tag{2.14}$$

3. An empirical Bayes detection procedure

Suppose now that the prior (1.2) is not completely known. That is, suppose that π or p or both are unknown. Also, suppose that we have available a number of observations $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_\nu$ from the past, where the formulation of random vectors $\boldsymbol{Y}_i$ is as defined in the introduction $1 \leq i \leq \nu$, $\nu \geq 1$. Let us suppose that

$$\begin{cases}\hat{\pi} = \hat{\pi}(\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_\nu) \\ \hat{p} = \hat{p}(\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_\nu)\end{cases} \tag{3.1}$$

denote consistent estimates of π and p, respectively, based on the auxiliary data $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_\nu$, $\nu \geq 1$. Thus, we suppose that as $\nu \to \infty$,

$$\hat{\pi} \xrightarrow{P} \pi \quad \text{and} \quad \hat{p} \xrightarrow{P} p, \tag{3.2}$$

where $\xrightarrow{P}$ denotes convergence in probability w.r.t. random vectors $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_\nu$, $\nu \geq 1$. We first study the case where only π is unknown but p is known. Following the form of (2.11) our empirical Bayes stopping time is defined as

$$\hat{\tau}_\nu = \inf\{n \geq 1; c\hat{\pi}_n - p(1-\hat{\pi}_n) \geq 0\}, \tag{3.3}$$

where $\hat{\pi}_k$ is defined by the right side of (2.5) with π replaced by $\hat{\pi}$, i.e.,

$$\hat{\pi}_k = \text{the right side of (2.5)} \mid_{\pi=\hat{\pi}}. \tag{3.4}$$

Let $R(\pi, \hat{\tau}_\nu)$ denote the unconditional prior risk (w.r.t. the prior (1.2)) of the stopping time (3.3). Then from (2.13) we obtain

$$R(\pi, \hat{\tau}_\nu) = E\left\{(1 - \pi_{\hat{\tau}_\nu}) + c\sum_{k=1}^{\hat{\tau}_{\nu-1}} \pi_k\right\}, \tag{3.5}$$

where E denotes expectation w.r.t. all random variables involved. The next theorem compares the the prior risk $R(\pi, \tau^*)$ with the prior risk $R(\pi, \hat{\tau}_\nu)$ as $\nu \to \infty$, where τ^* is the stopping time given by (2.11).

THEOREM 3.1. *Let τ^* and $\hat{\tau}_\nu$ be defined by* (2.11) *and* (3.3), *respectively. Let prior risks $R(\pi, \tau^*)$ and $R(\pi, \hat{\tau}_\nu)$ be defined by* (2.13) *with $\tau = \tau^*$ and* (3.5), *respectively. Suppose that only π is unknown of* (1.2). *Let $\hat{\pi}$ be defined by* (3.1) *and satisfy* (3.2). *Furthermore, suppose that* $\limsup_{\nu\to\infty} E(\hat{\tau}_\nu)^2 < \infty$. *Then,*

$$\limsup_{\nu\to\infty} R(\pi, \hat{\tau}_\nu) \leq R(\pi, \tau^*). \tag{3.6}$$

Now we examine the case where both π and p are unknown of (1.2). For this case we define

$$\hat{\hat{\tau}}_\nu = \inf\{n \geq 1; c\hat{\hat{\pi}}_n - \hat{p}(1 - \hat{\hat{\pi}}_n) \geq 0\} \tag{3.7}$$

as our empirical Bayes stopping time for the detection problem, where $\hat{\hat{\pi}}_\nu$ is defined by the right-side of (2.5) with now π and p replaced by $\hat{\pi}$ and $\hat{p}$, respectively. Let $R(\pi, \hat{\hat{\tau}}_\nu)$ denote the unconditional prior risk of $\hat{\hat{\tau}}_\nu$. Then,

$$R(\pi, \hat{\hat{\tau}}_\nu) = E\left\{(1 - \pi_{\hat{\hat{\tau}}_\nu}) + c\sum_{k=1}^{\hat{\hat{\tau}}-1})\pi_k\right\}. \tag{3.8}$$

An asymptotic result, such as (3.6), can be established again to compare $R(\pi, \hat{\hat{\tau}}_\nu)$ and $R(\pi, \tau^*)$. However, more regularity conditions are required. The next theorem is a version of such a result.

THEOREM 3.2. *Let τ^* and $\hat{\hat{\tau}}_\nu$ be defined by* (2.11) *and* (3.7), *respectively. Let $\hat{\pi}$ and $\hat{p}$ be defined by* (3.1) *and satisfy* (3.2). *Suppose that* $\limsup_{\nu\to\infty} E(\hat{\hat{\tau}}_\nu^2) < \infty$ *and that* $\limsup_{\nu\to\infty} E(1/(1-p)^{\hat{\hat{\tau}}_\nu})^2 < \infty$. *Then,* $\limsup_{\nu\to\infty} R(\pi, \hat{\hat{\tau}}_\nu) \leq R(\pi, \tau^*)$.

Remark 1. The inequality (3.6), i.e., the asymptotic comparison of risk $R(\pi, \hat{\tau}_\nu)$ to the "component risk" is termed " asymptotic superiority" of the empirical Bayes stopping time sequence $\{\hat{\tau}_\nu\}$; see Karunamuni (1988). When τ^* is the optimal rule, then (3.6) implies that $\lim_{n\to\infty} R(\pi, \hat{\tau}_\nu) = R(\pi, \tau^*)$, i.e., the "asymptotic optimality" (Robbins (1956, 1964)) of the EB stopping sequence $\{\hat{\tau}_\nu\}$.

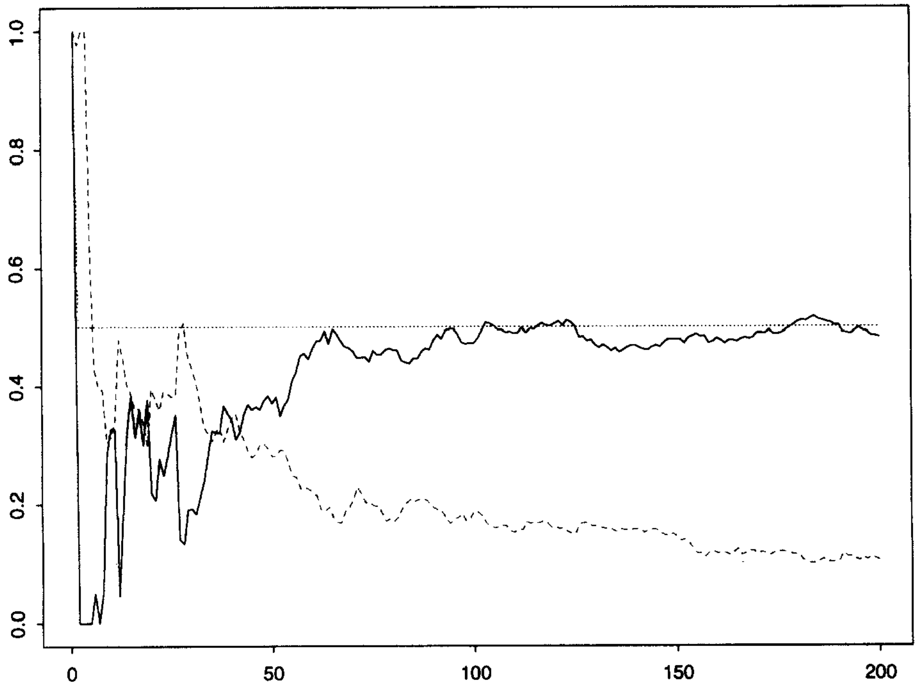

Fig. 1. Behaviors of $\hat{\pi}$ (solid curve) and $\hat{p}$ (dotted curve) w.r.t. ν, for $\pi = 0.5$, $p = 0.1$, $F_0 = N(5, 1)$, $F_1 = N(10, 1)$.

To see this, note that $R(\pi, \tau^*) \leq R(\pi, \hat{\tau})$ if τ^* is optimal. Then, $R(\pi, \tau^*) \leq \liminf_{\nu \to \infty} R(\pi, \hat{\tau})$. Hence the result.

Remark 2. Consistent estimates $\hat{\pi}$ and $\hat{p}$ satisfying (3.2) based on the previous data can be constructed in many ways. One simple set of estimates is the method of moment type: By the assumptions of auxiliary data $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_\nu$, note that the first and the second random observations of these vectors (i.e., $Y_{\cdot 1}$ and $Y_{\cdot 2}$) have the joint distribution (see (2.1))

$$F_\pi(y_1, y_2) = (\pi + (1-\pi)p)F_1(y_1)F_1(y_2) + (1-\pi)(1-p)pF_0(y_1)F_1(y_2) + (1-\pi)(1-p)^2 F_0(y_1)F_0(y_2). \tag{3.9}$$

Let $\mu_{1,j}$ denote the first moment of $F_j(y)$, $j = 0, 1$. Then, one can easily show that

$$\int y_1 dF_\pi(y_1, y_2) = (\pi + (1-\pi)p)\mu_{1,1} + (1-\pi)(1-p)\mu_{1,0} \tag{3.10}$$

and

$$\int y_2 dF_\pi(y_1, y_2) = (\pi+(1-\pi)p)\mu_{1,1} + (1-\pi)(1-p)p\mu_{1,1} + (1-\pi)(1-p)^2\mu_{1,0}. \tag{3.11}$$

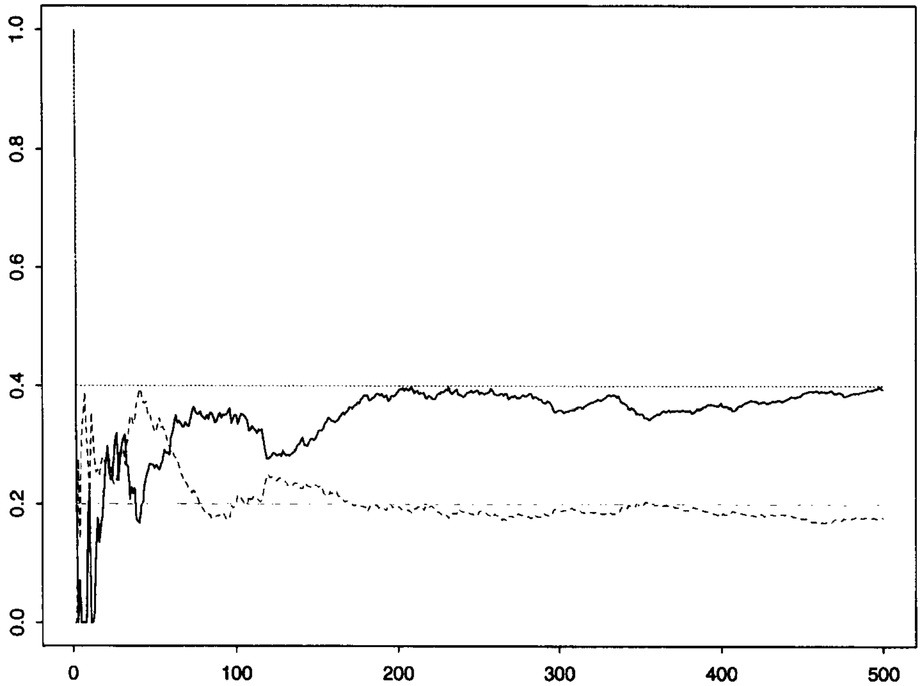

Fig. 2. Behaviors of $\hat{\pi}$ (solid curve) and $\hat{p}$ (dotted curve) w.r.t. ν, for $\pi = 0.4$, $p = 0.2$, $F_0 = N(2,1)$, $F_1 = N(10,1)$.

Let $\hat{\mu}_1 = \nu^{-1}\sum_{i=1}^{\nu} Y_{i1}$ and $\hat{\mu}_2 = \nu^{-1}\sum_{i=1}^{\nu} Y_{i2}$. Then the equalities (3.10) and (3.11) motivate the following sample moment relationships:

$$\begin{cases} \lambda\mu_{1,1} + (1-\lambda)\mu_{1,0} = \hat{\mu}_1 \\ \lambda\mu_{1,1} + (1-\lambda)p\mu_{1,1} + (1-\lambda)(1-p)\mu_{1,0} = \hat{\mu}_2, \end{cases} \tag{3.12}$$

where $\lambda = \pi + (1-\pi)p$. The solutions of equations (3.12) for π and p, restricted in the interval $(0,1)$, can be taken as a plausible set of consistent estimates for π and p. In order to study the behavior of resulting estimates (say, $\hat{\pi}$ and $\hat{p}$) with respect to ν, a Monte Carlo simulation was performed. The distributions F_0 and F_1 were chosen as normal distributions with variance 1. For several combinations of the means $\mu_{1,0}$ and $\mu_{1,1}$ of F_0 and F_1, respectively, and of prior parameters π and p, we calculated the values of $\hat{\pi}$ and $\hat{p}$ for ν ranging from 1 to 200, and sometimes from 1 to 500. The results of three separate cases, namely, (i) $\pi = 0.5$, $p = 0.1$, $\mu_{1,0} = 5$, $\mu_{1,1} = 10$; (ii) $\pi = 0.4$, $p = 0.2$, $\mu_{1,0} = 2$, $\mu_{1,1} = 10$ and (iii) $\pi = 0.5$, $p = 0.5$, $\mu_{1,0} = 3$, $\mu_{1,1} = 8$ are displayed in Figs. 1, 2 and 3, respectively. Figure 4 represents the behavior of $\hat{\pi}$ for the case (iv) $\pi = 0.5$, $p = 0.1$, $\mu_{1,0} = 3$, $\mu_{1,1} = 8$, when it is assumed that p is known and only π is unknown (i.e., the situation of Theorem 3.1 above). A number of other choices of π, p, $\mu_{1,0}$ and $\mu_{1,1}$ were also studied. Again, the behaviors of $\hat{\pi}$ and $\hat{p}$ were similar to those of Figs. 1 to 4. In each case studied, it appears that estimates $\hat{\pi}$ and $\hat{p}$ approach their respective true values as ν increases. For small values of ν (i.e., for $\nu < 25$), however, the

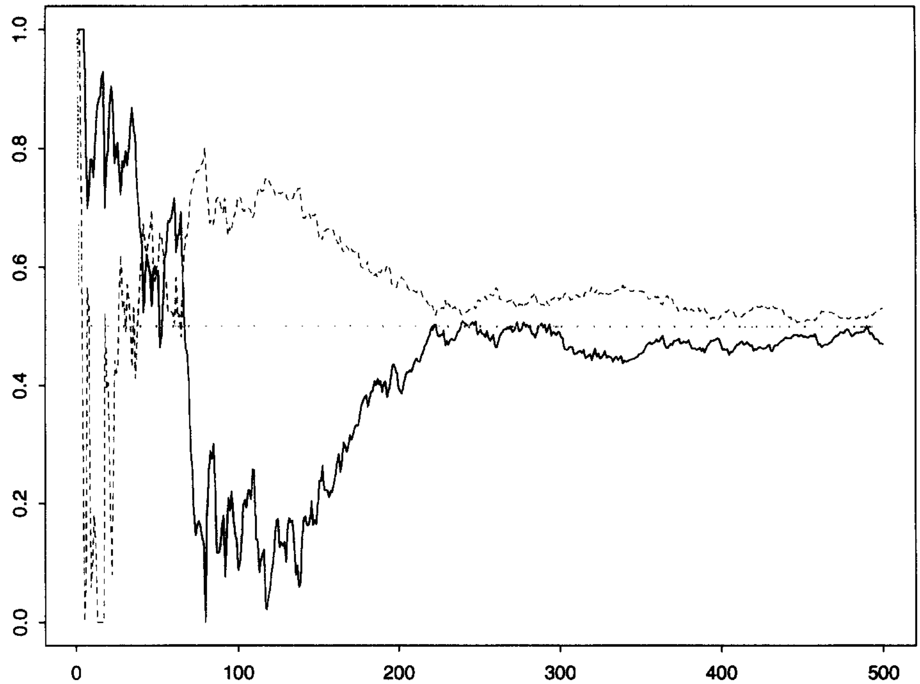

Fig. 3. Behaviors of $\hat{\pi}$ (solid curve) and $\hat{p}$ (dotted curve) w.r.t. ν, for $\pi = 0.5$, $p = 0.5$, $F_0 = N(3,1)$, $F_1 = N(8,1)$.

estimates seem to have a high fluctuation. We believe that this is partly due to the facts that the estimates were truncated into the interval $(0,1)$ in our simulation and the behavior of one estimate affects that of the other.

Remark 3. In practice, the result (3.6) means that, in the presence of a large amount of auxiliary data from the past, one can expect to do well, as good as the unknown rule τ^* and sometimes better. In most practical applications a large set of auxiliary data is available as a result of repeating the same process over and over again. For example, this is the case with many industries in which sequential methods are applied for controlling manufacturing processes and for monitoring the stationarity of a process. Of course, behavior of the proposed EB stopping time for small and moderate amount of data is of critical concern in some applications. However, the performance of an EB procedure for relatively small ν is not easy to establish except by simulation.

Remark 4. In some practical situations the distributions before and after the change are not known, i.e., $F_0(x)$ and $F_1(x)$ are unknown. Even in parametric models these distributions may depend on unknown parameters, which have to be estimated from the data during the detection processes. Bayesian solutions to such cases have been developed by Zacks and Barzily (1981) using the backward induction principle of dynamic programming. They also considered the problem of detecting a change in the success probability in a sequence of binomial trials

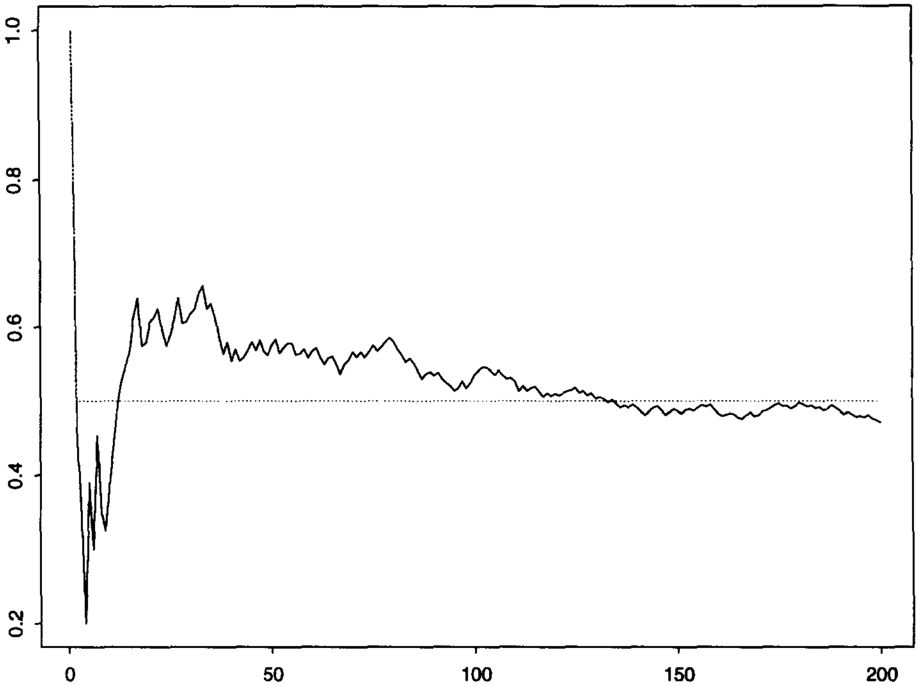

Fig. 4. Behavior of $\hat{\pi}$ (solid curve) w.r.t. ν, for $\pi = 0.5$, $p = 0.1$, $F_0 = N(3,1)$, $F_1 = N(8,1)$.

using a two-step look-ahead stopping rule. In the preceding work it is assumed that the prior probability function (1.2) for θ. It would be interesting to study the corresponding empirical Bayes problem for such cases as well. As in the present paper, the resulting procedure would then be parametric empirical Bayes. When the form of the prior is completely unknown, then a nonparametric (purely) empirical Bayes procedure should be implemented. However, the construction of such a procedure is more formidable, as for the corresponding Bayesian problem.

Remark 5. If $\hat{\tau}_\nu$ and $\hat{\hat{\tau}}_\nu$ are bounded then the regularity conditions given in Theorems 3.1 and 3.2 for $\hat{\tau}_\nu$ and $\hat{\hat{\tau}}_\nu$ are trivially satisfied. This is usually the case in practise: In most applications, it is reasonable to assume that τ^* is bounded by some known large finite number N, say. Then, instead of $\hat{\tau}_\nu$ and $\hat{\hat{\tau}}_\nu$, one can implement stopping rules (compare with (3.3) and (3.7)),

$$\tilde{\tau}_\nu = \inf\{N \geq n \geq 1; c\hat{\pi}_n - p(1 - \hat{\pi}_n) \geq 0\}$$

and

$$\tilde{\tilde{\tau}}_\nu = \inf\{N \geq n \geq 1; c\hat{\hat{\pi}}_n - \hat{p}(1 - \hat{\hat{\pi}}_n) \geq 0\}.$$

Since $\tilde{\tau}_\nu$ and $\tilde{\tilde{\tau}}_\nu$ are bounded by N, the conditions $\limsup_{\nu\to\infty} E(\tilde{\tau}_\nu^2) < \infty$, $\limsup_{\nu\to\infty} E(\tilde{\tilde{\tau}}_\nu^2) < \infty$ and $\limsup_{\nu\to\infty} E(1/(1-p)^{\tilde{\tilde{\tau}}_\nu})^2 < \infty$ are easily satisfied.

4. Proofs

First we state three preliminary lemmas. Lemma 4.1 below is known as the Datta-Singh inequality; see Lemma 4.1 of Datta (1991).

LEMMA 4.1. *For any real numbers y, z, Y, Z and L such that $z \neq 0 < L$,*

$$|z| \left\{ \left| \frac{y}{z} - \frac{Y}{Z} \right| \wedge L \right\} \leq |y - Y| + \left(\left| \frac{y}{z} \right| + L \right) |z - Z|.$$

LEMMA 4.2. *Let π_n and $\tilde{\pi}_n$ be defined by* (2.5) *and* (3.4), *respectively. Then, for $n \geq 1$,*

$$|\pi_n - \tilde{\pi}_n| \leq |\pi - \hat{\pi}| \left\{ \frac{4}{\pi(1-\pi)} + \frac{3}{(1-\pi)} \right\}.$$

PROOF. Write $\pi_n = T_n / D_n$ and $\tilde{\pi}_n = \hat{T}_n / \hat{D}_n$, where D_n is given by (2.4), T_n is equal to the numerator of (2.3), $\hat{T}_n = T_n \mid_{\pi = \hat{\pi}}$ and $\hat{D}_n = D_n \mid_{\pi = \hat{\pi}}$. Then, by an application of Lemma 4.1, we obtain

$$\begin{aligned} |\pi_n - \tilde{\pi}_n| = \{ |\pi_n - \tilde{\pi}_n| \wedge 2 \} &\leq \frac{1}{D_n} |T_n - \hat{T}_n| + \frac{1}{D_n} (\pi_n + 2) |D_n - \hat{D}_n| \\ &\leq \frac{1}{D_n} |T_n - \hat{T}_n| + \frac{3}{D_n} \left\{ |T_n - \hat{T}_n| + (1-p)^n \prod_{i=1}^{n} f_0(X_i) |\pi - \hat{\pi}| \right\}, \end{aligned}$$

since $0 < \pi_n \leq 1$ and $D_n = T_n + (1-\pi)(1-p)^n \prod_{i=1}^{n} f_0(X_i)$. Now the result follows from the following bounds:

$$\frac{1}{D_n} |T_n - \hat{T}_n| \leq \frac{(1-p)|\pi - \hat{\pi}|}{\pi + p(1-\pi)} + \frac{1}{(1-\pi)} |\pi - \hat{\pi}|$$

and

$$\frac{1}{D_n} \prod_{i=1}^{n} f_0(X_i) \leq \frac{1}{(1-\pi)(1-p)^n}. \qquad \square$$

LEMMA 4.3. *Let π_n be defined by* (2.5) *and $\hat{\tilde{\pi}}_n = \pi_n \mid_{\pi = \hat{\pi}, p = \hat{p}}$, where $\hat{p}$ and $\hat{\pi}$ are as defined in Theorem* 3.2. *Then, for $n \geq 1$,*

$$\begin{aligned} |\pi_n - \hat{\tilde{\pi}}_n| \leq |\pi - \hat{\pi}| &\left\{ \frac{8}{\pi(1-p)} + \frac{4}{p(1-\pi)} + \frac{3}{(1-\pi)} \right\} \\ &+ |\hat{p} - p| \left\{ \frac{8}{\pi(1-p)} + \frac{8}{p^2(1-\pi)} + \frac{4}{p^2(1-\pi)(1-p)^n} \right\} \\ &+ \frac{6}{(1-\pi)(1-p)^n} |(1-p)^n - (1-\hat{p})^n|. \end{aligned}$$

PROOF. Write $\pi_n = T_n/D_n$ and $\hat{\hat{\pi}}_n = \hat{\hat{T}}_n/\hat{\hat{D}}_n$, where $\hat{\hat{T}}_n = T_n \mid_{\pi=\hat{\pi},p=\hat{p}}$ and $\hat{\hat{D}}_n = D_n|_{\pi=\hat{\pi},p=\hat{p}}$. Then, again by an application of Lemma 4.1, we obtain

$$|\pi_n - \hat{\hat{\pi}}_n| \le \frac{1}{D_n}|T_n - \hat{\hat{T}}_n| + \frac{3}{D_n} \cdot \left\{|T_n - \hat{\hat{T}}_n| + \prod_{i=1}^{n} f_0(X_i)(|(1-\pi)(1-p)^n - (1-\hat{\pi})(1-\hat{p})^n|)\right\}.$$

Now the result follows from the following bounds on (as in the proof of Lemma 4.2) $D_n^{-1}|T_n - \hat{\hat{T}}_n|$, $D_n^{-1}\prod_{i=1}^n f_0(X_i)$ and $|(1-\pi)(1-p)^n - (1-\hat{\pi})(1-\hat{p})^n|$:

$$\frac{1}{D_n}|T_n - \hat{\hat{T}}_n| \le |\pi - \hat{\pi}|\left\{\frac{2}{\pi(1-p)} + \frac{1}{p(1-\pi)}\right\} + |\hat{p} - p|\left\{\frac{2}{\pi(1-p)} + \frac{2}{p^2(1-\pi)} + \frac{1}{p^2(1-\pi)(1-p)^n}\right\}$$

$$|(1-\pi)(1-p)^n - (1-\hat{\pi})(1-\hat{p})^n| \le 2|(1-p)^n - (1-\hat{p})^n| + (1-p)^n|\pi - \hat{\pi}|. \qquad \square$$

PROOF OF THEOREM 3.1. Observe that the difference $R(\pi, \hat{\tau}_\nu) - R(\pi, \tau^*)$ can be written as

$$R(\pi, \hat{\tau}_\nu) - R(\pi, \tau^*) = \sum_{r=1}^{\infty}\sum_{s=1}^{\infty} E[\tau^* = s][\hat{\tau}_\nu = r] \cdot \left\{(\pi_s - \pi_r) + c\left(\sum_{i=1}^{r-1}\pi_i - \sum_{j=1}^{s-1}\pi_j\right)\right\}, \tag{4.1}$$

where $[A]$ denotes the indicator function of a set A, throughout. Also arguments of functions have been suppressed for notational convenience, and E denotes expectation w.r.t. all of the random variables involved. Write

$$R(\pi, \hat{\tau}_\nu) - R(\pi, \tau^*) = \Delta_\nu + \Delta'_\nu, \tag{4.2}$$

where

$$\Delta_\nu = \sum_{k=1}^{\infty}\sum_{\ell=k+1}^{\infty} E[\tau^* = k][\hat{\tau}_\nu = \ell]\left\{(\pi_k - \pi_\ell) + c\left(\sum_{i=1}^{\ell-1}\pi_i - \sum_{j=1}^{k-1}\pi_j\right)\right\} \tag{4.3}$$

and

$$\Delta'_\nu = \sum_{s=2}^{\infty}\sum_{r=1}^{s-1} E[\tau^* = s][\hat{\tau} = r]\left\{\pi_s - \pi_r + c\left(\sum_{i=1}^{r-1}\pi_i - \sum_{j=1}^{s-1}\pi_j\right)\right\}. \tag{4.4}$$

We shall show that $\lim_{\nu\to\infty} \Delta'_\nu = 0$ and that $\limsup_{\nu\to\infty} \Delta_\nu \le 0$. First, observe that $\sum_{r=1}^{s-1} E[\tau^* = s][\hat{\tau}_\nu = r]\{\pi_s - \pi_r - c\sum_{i=r}^{s} \pi_i\} \le E[\tau^* = s]\{2 + 2cs\}$, and

$$\sum_{s=1}^{\infty} E[\tau^* = s]\{2+2cs\} = 2P(\tau^* < \infty) + 2cE(\tau^*) < \infty, \tag{4.5}$$

where the finiteness of (4.5) follows from (2.14). This result allows us to take the limit sign (i.e. $\lim_{\nu\to\infty}$) inside the summations of (4.4). Also note that by (2.11) and (3.2), for $1 \le r < s$,

$$\begin{aligned}[t][\tau^* = s][\hat{\tau}_\nu = r] &\le [c\pi_r - p(1-\pi_r) < 0][c\hat{\pi}_r - p(1-\hat{\pi}_r) \ge 0] \\ &= [\alpha_r < 0][\hat{\alpha}_r \ge 0] \\ &\le [|\hat{\alpha}_r - \alpha_r| \ge |\alpha_r|][\alpha_r < 0],\end{aligned} \tag{4.6}$$

where

$$\hat{\alpha}_r = c\hat{\pi}_r - p(1-\hat{\pi}_r), \qquad r \ge 1, \tag{4.7}$$

and

$$\alpha_r = c\pi_r - p(1-\pi_r), \qquad r \ge 1, \tag{4.8}$$

with π_r and $\hat{\pi}_r$ are given by (2.5) and (3.4), respectively. Then it is easy to show that $\hat{\alpha}_r - \alpha_r \xrightarrow{P} 0$ under the assumption (3.2) for all $r \ge 1$. Therefore, from (4.6) we get

$$\lim_{\nu\to\infty} E[\tau^* = s][\hat{\tau}_\nu = r] = 0, \qquad \text{for} \quad 1 \le r < s < \infty. \tag{4.9}$$

It now follows from (4.4), (4.5) and (4.9) that $\lim_{\nu\to\infty} \Delta'_\nu = 0$, by an application of the dominated convergence theorem. To study the asymptotic behavior of Δ_ν, we re-arrange each term on the right-side of (4.3) and obtain

$$\sum_{k=1}^{\infty}\sum_{\ell=k+1}^{\infty} E[\tau^* = k][\hat{\tau}_\nu = \ell]\pi_k = \sum_{k=1}^{\infty} E[\tau^* = k][\hat{\tau} \ge k+1]\pi_k, \tag{4.10}$$

$$\begin{aligned}[t]&-\sum_{k=1}^{\infty}\sum_{\ell=k+1}^{\infty} E[\tau^* = k][\hat{\tau}_\nu = \ell]\pi_k \\ &= -\sum_{k=1}^{\infty} E[\tau^* = k]\left\{\sum_{\ell=k+1}^{\infty} [\hat{\tau} = \ell]\pi_\ell\right\} \\ &= -\sum_{k=1}^{\infty} E[\tau^* = k]\left\{\sum_{\ell=k+1}^{\infty} [\hat{\tau} \ge \ell]\pi_\ell - \sum_{\ell=k+1}^{\infty} [\hat{\tau} \ge \ell+1]\pi_\ell\right\}\end{aligned} \tag{4.11}$$

and

$$c\sum_{k=1}^{\infty}\sum_{\ell=k+1}^{\infty} E[\tau^*=k][\hat{\tau}_\nu=\ell]\left(\sum_{i=k}^{\ell-1}\pi_i\right) = c\sum_{k=1}^{\infty}\sum_{\ell=k+1}^{\infty} E[\tau^*=k][\hat{\tau}_\nu\geq\ell]\pi_{\ell-1}. \tag{4.12}$$

Now combining (4.10), (4.11) and (4.12) we obtain

$$\Delta_\nu=\sum_{k=1}^{\infty}\sum_{\ell=k+1}^{\infty} E[\hat{\tau}_\nu\geq\ell][\tau^*=k]\{\pi_{\ell-1}-\pi_\ell+c\pi_{\ell-1}\}. \tag{4.13}$$

By definition, $[\hat{\tau}_\nu\geq\ell]=\prod_{i=1}^{\ell-1}[\hat{\alpha}_i<0]$, where $\hat{\alpha}_i$ is given by (4.7), $i\geq 1$. Hence the random function $[\hat{\tau}_\nu\geq\ell]$ is $\{\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_\nu;(\pi_0,\pi_1,\ldots,\pi_{\ell-1})\}$-measurable, and $[\tau^*=k]$ is $\{\pi_0,\pi_1,\ldots,\pi_k\}$-measurable. Also, by the assumptions, the auxiliary data $\{\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_\nu\}$ is independent of the random elements $\{\pi_0,\pi_1,\ldots,\}$. Therefore, we can rewrite (4.13) as follows:

$$\begin{aligned}\Delta_\nu&=\sum_{k=1}^{\infty}\sum_{\ell=k+1}^{\infty} E[\hat{\tau}_\nu\geq\ell][\tau^*=k]\{E(\pi_{\ell-1}-\pi_\ell+c\pi_{\ell-1}\mid\mathcal{F}^*_{\ell-1})\}\\&=\sum_{k=1}^{\infty}\sum_{\ell=k+1}^{\infty} E[\hat{\tau}_\nu\geq\ell][\tau^*=k]\{\pi_{\ell-1}+c\pi_{\ell-1}-E(\pi_\ell\mid\mathcal{F}^*_{\ell-1})\}\\&=\sum_{k=1}^{\infty}\sum_{\ell=k+1}^{\infty} E[\hat{\tau}_\nu\geq\ell][\tau^*=k]\{c\pi_{\ell-1}-p(1-\pi_{\ell-1})\},\end{aligned} \tag{4.14}$$

the last equality is obtained by $E(\pi_\ell\mid\mathcal{F}^*_{\ell-1})=\pi_{\ell-1}+p(1-\pi_{\ell-1})$; see (2.12). Now write

$$\Delta_\nu=\Delta''_\nu+\Delta'''_\nu, \tag{4.15}$$

where

$$\Delta''_\nu=\sum_{k=1}^{\infty}\sum_{\ell=k+1}^{\infty} E[\hat{\tau}_\nu\geq\ell][\tau^*=k]\{c(\pi_{\ell-1}-\hat{\pi}_{\ell-1})+p(\pi_{\ell-1}-\hat{\pi}_{\ell-1})\}\} \tag{4.16}$$

and

$$\Delta'''_\nu=\sum_{k=1}^{\infty}\sum_{\ell=k+1}^{\infty} E[\hat{\tau}_\nu\geq\ell][\tau^*=k]\{c\hat{\pi}_{\ell-1}-p(1-\hat{\pi}_{\ell-1})\}. \tag{4.17}$$

Since $\hat{\alpha}_{\ell-1}=c\hat{\pi}_{\ell-1}-p(1-\hat{\pi}_{\ell-1})<0$ on the event $\{\hat{\tau}_\nu\geq\ell\}$, we see that the term Δ'''_ν given by (4.17) is nonpositive for all ν, i.e., $\Delta'''_\nu\leq 0$ for all ν. It now remains to study the asymptotic behavior of Δ''_ν given by (4.16). Using Lemma 4.2, we can bound Δ''_ν as follows:

$$|\Delta''_\nu|\leq(c+p)\sum_{k=1}^{\infty}\sum_{\ell=k+1}^{\infty} E[\hat{\tau}_\nu\geq\ell][\tau^*=k]|\pi-\hat{\pi}|\left\{\frac{4}{\pi(1-\pi)}+\frac{3}{(1-\pi)}\right\}. \tag{4.18}$$

Now using $P(\tau^* < \infty) = 1$ followed by applying the Cauchy-Schwarz inequality,

$$(4.19)\quad \begin{aligned}\sum_{k=1}^{\infty}\sum_{\ell=k+1}^{\infty} E[\hat{\tau}_\nu \geq \ell][\tau^* = k]|\pi - \hat{\pi}| &\leq \sum_{k=1}^{\infty}\sum_{\ell=1}^{\infty} E[\hat{\tau}_\nu \geq \ell][\tau^* = k]|\pi - \hat{\pi}| \\ &= E\{\hat{\tau}_\nu|\pi - \hat{\pi}|\} \\ &\leq E^{1/2}(\hat{\tau}_\nu^2)E^{1/2}(\pi - \hat{\pi})^2.\end{aligned}$$

Thus, if $\limsup_{\nu\to\infty} E(\hat{\tau}_\nu^2) < \infty$, then from (4.18) and (4.19) one obtains $\lim_{\nu\to\infty} \Delta''_\nu = 0$, since $E(\pi - \hat{\pi})^2 \to 0$ as $\nu \to \infty$. This completes the proof of Theorem 3.1. □

PROOF OF THEOREM 3.2. The difference $R(\pi, \hat{\hat{\tau}}_\nu) - R(\pi, \tau^*)$ is equal to (cf. (4.1))

$$(4.20)\qquad R(\pi, \hat{\hat{\tau}}_\nu) - R(\pi, \tau^*) = W_\nu + W'_\nu$$

where

$$(4.21)\quad W_\nu = \sum_{k=1}^{\infty}\sum_{\ell=k+1}^{\infty} E[\tau^* = k][\hat{\hat{\tau}}_\nu = \ell]\left\{(\pi_k - \pi_\ell) + c\left(\sum_{i=1}^{\ell-1}\pi_i - \sum_{j=1}^{k-1}\pi_j\right)\right\}$$

and

$$(4.22)\quad W'_\nu = \sum_{s=2}^{\infty}\sum_{r=1}^{s-1} E[\tau^* = s][\hat{\hat{\tau}}_\nu = r]\left\{(\pi_s - \pi_r) + c\left(\sum_{i=1}^{r-1}\pi_i - \sum_{j=1}^{s-1}\pi_j\right)\right\}.$$

Following the same lines of the proof $\lim_{\nu\to\infty} \Delta'_\nu = 0$, it can be shown that $\lim_{\nu\to\infty} W'_\nu = 0$. Further, using similar steps that we used to obtain (4.14), we can show that $W_\nu = W''_\nu + W'''_\nu$, where

$$(4.23)\qquad W''_\nu = \sum_{k=1}^{\infty}\sum_{\ell=k+1}^{\infty} E[\hat{\hat{\tau}}_\nu \geq \ell][\tau^* = k]\{(c+p)(\pi_{\ell-1} - \hat{\hat{\pi}}_{\ell-1})\}$$

and

$$(4.24)\qquad W'''_\nu = \sum_{k=1}^{\infty}\sum_{\ell=k+1}^{\infty} E[\hat{\hat{\tau}}_\nu \geq \ell][\tau^* = k]\{c\hat{\hat{\pi}}_{\ell-1} - p(1 - \hat{\hat{\pi}}_{\ell-1})\}.$$

Again, it follows that $W'''_\nu \leq 0$ for all ν. So, it remains to study the asymptotic behavior of W''_ν. We shall show that $\lim_{\nu\to\infty} W''_\nu = 0$. Using Lemma 4.3, we can bound W''_ν as follows:

$$(4.25)\quad \begin{aligned}|W''_\nu| \leq (c+p)&\left\{\frac{8}{\pi(1-\pi)} + \frac{4}{p(1-\pi)} + \frac{3}{(1-\pi)}\right\} \\ &\cdot\sum_{k=1}^{\infty}\sum_{\ell=k+1}^{\infty} E[\hat{\hat{\tau}}_\nu \geq \ell][\tau^* = k]|\pi - \hat{\pi}|\end{aligned}$$

$$+ (c+p)\left(\frac{8}{\pi(1-p)} + \frac{8}{p^2(1-\pi)}\right)$$
$$\cdot \sum_{k=1}^{\infty}\sum_{\ell=k+1}^{\infty} E[\hat{\hat{\tau}}_\nu \geq \ell][\tau^* = k]|p-\hat{p}|$$
$$+ \frac{4(c+p)}{p^2(1-\pi)}\sum_{k=1}^{\infty}\sum_{\ell=k+1}^{\infty} E[\hat{\hat{\tau}}_\nu \geq \ell][\tau^* = k]|p-\hat{p}|\frac{1}{(1-p)^{\ell-1}}$$
$$+ \frac{6(c+p)}{(1-\pi)}\sum_{k=1}^{\infty}\sum_{\ell=k+1}^{\infty} E[\hat{\hat{\tau}}_\nu \geq \ell][\tau^* = k]$$
$$\cdot |(1-p)^{\ell-1} - (1-\hat{p})^{\ell-1}|\frac{1}{(1-p)^{\ell-1}}.$$

The first two terms on the right side of (4.25) are similar to the right-side of (4.18). So, it is easy to show that

$$\begin{aligned}&\sum_{k=1}^{\infty}\sum_{\ell=k+1}^{\infty} E[\hat{\hat{\tau}}_\nu \geq \ell][\tau^* = k]|\pi-\hat{\pi}| \leq E^{1/2}(\hat{\hat{\tau}}_\nu^2)E^{1/2}(\pi-\hat{\pi})^2\\ &\sum_{k=1}^{\infty}\sum_{\ell=k+1}^{\infty} E[\hat{\hat{\tau}}_\nu \geq \ell][\tau^* = k]|p-\hat{p}| \leq E^{1/2}(\hat{\hat{\tau}}_\nu^2)E^{1/2}(p-\hat{p})^2.\end{aligned} \tag{4.26}$$

The third term on the right-side of (4.25) can be bounded by

$$J\sum_{\ell=1}^{\infty} E[\hat{\hat{\tau}}_\nu \geq \ell]|p-\hat{p}|\frac{1}{(1-p)^{\ell-1}}, \tag{4.27}$$

since $P(\tau^* < \infty) = 1$, where $J = 4(c+p)/p^2(1-\pi)$. We can rewrite and bound (4.27) as follows:

$$\begin{aligned}&J\sum_{\ell=1}^{\infty} E[\hat{\hat{\tau}}_\nu \geq \ell]|p-\hat{p}|\frac{1}{(1-p)^{\ell-1}}\\ &\quad = J\sum_{j=1}^{\infty}\sum_{\ell=1}^{j} E[\hat{\hat{\tau}}_\nu = j]|p-\hat{p}|\frac{1}{(1-p)^{\ell-1}}\\ &\quad \leq Jp(1-p)^{-1}\sum_{j=1}^{\infty} E[\hat{\hat{\tau}}_\nu = j]|p-\hat{p}|\frac{1}{(1-p)^{\ell-1}}\\ &\quad \leq Jp(1-p)^{-1}E^{1/2}(p-\hat{p})^2E^{1/2}\left(\frac{1}{(1-p)^{\hat{\hat{\tau}}_\nu}}\right)^2.\end{aligned} \tag{4.28}$$

Using the identity $a^n - b^n = (a-b)\sum_{i=1}^{n} a^{i-1}b^{n-i}$, the fourth term on the right side of (4.25) can be bounded by

$$K\sum_{k=1}^{\infty}\sum_{\ell=k+1}^{\infty} E[\hat{\hat{\tau}}_\nu \geq \ell][\tau^* = k]|p-\hat{p}|\sum_{i=1}^{\ell-1}(1-p)^{i-1}(1-\hat{p})^{\ell-1-i}\frac{1}{(1-p)^{\ell-1}}, \tag{4.29}$$

where K is a positive constant. By re-arranging terms (with the help of Fubini's theorem), (4.29) can be again bounded by

$$
\begin{aligned}
(4.30)\quad & KE\left\{|p-\hat{p}|\sum_{k=1}^{\infty}[\tau^*=k]\right.\\
&\left.\cdot\sum_{\ell=2}^{\infty}\sum_{j=\ell}^{\infty}\sum_{i=1}^{\ell-1}[\hat{\hat{\tau}}_\nu=j](1-p)^{i-1}(1-\hat{p})^{\ell-1-i}\frac{1}{(1-p)^{\ell-1}}\right\}\\
&\le KE\left\{|p-\hat{p}|\sum_{j=2}^{\infty}\sum_{\ell=2}^{j}E[\hat{\hat{\tau}}_\nu=j]\sum_{i=1}^{\ell-1}(1-p)^{i-1}\frac{1}{(1-p)^{\ell-1}}\right\}\\
&=KE\left\{|p-\hat{p}|\sum_{j=2}^{\infty}[\hat{\hat{\tau}}_\nu=j]\sum_{\ell=2}^{j}\sum_{i=1}^{\ell-1}(1-p)^{i-1}\frac{1}{(1-p)^{\ell-1}}\right\}\\
&\le Kp^{-1}E\left\{|p-\hat{p}|\sum_{j=2}^{\infty}[\hat{\hat{\tau}}_\nu=j]\sum_{\ell=2}^{j}\frac{1}{(1-p)^{\ell-1}}\right\}\\
&\le Kp^{-2}E\left\{|p-\hat{p}|\sum_{j=2}^{\infty}[\hat{\hat{\tau}}_\nu=j]\frac{1}{(1-p)^{j-1}}\right\}\\
&\le Kp^{-2}E\left\{|p-\hat{p}|\left(1+\frac{1}{(1-p)^{\hat{\hat{\tau}}_\nu-1}}\right)\right\}\\
&\le kp^{-2}\left\{E|p-\hat{p}|+E^{1/2}(p-\hat{p})^2E\left(\frac{1}{(1-p)^{\hat{\hat{\tau}}_\nu}}\right)^2\right\}.
\end{aligned}
$$

Thus, if $\limsup_{\nu\to\infty}E(\hat{\hat{\tau}}_\nu^2)<\infty$ and $\limsup_{\nu\to\infty}E(1/(1-p)^{\hat{\hat{\tau}}_\nu})^2<\infty$, then from (4.25) to (4.30) it follows that $\lim_{\nu\to\infty}W_\nu''=0$, since $E(\pi-\hat{\pi})^2\to 0$ and $E(p-\hat{p})^2\to 0$ as $\nu\to\infty$ by assumptions. This completes the proof. □

Acknowledgements

The authors would like to thank the referees for their helpful and constructive comments, which significantly improved the presentation of the material.

References

Abdel-Hameed, M. (1977). Optimality of the one-step look-ahead stopping times, *J. Appl. Probab.*, **14**, 162–169.

Bather, J. A. (1963). Control charts and minimization of costs, *J. Roy. Statist. Soc. Ser. B*, **25**, 49–80.

Bather, J. A. (1967). On a quickest detection problem, *Ann. Math. Statist.*, **38**, 711–724.

Chow, Y. S. and Robbins, H. (1961). A martingale system theory and applications, *Proc. Fourth Berkeley Symp. on Math. Statist. Prob.*, Vol. 1, 93–104, University of California Press, Berkeley.

Datta, G. S. (1991). Asymptotic optimality of Bayes compound estimators in compact exponential families, *Ann. Statist.*, **19**, 354–365.

Karunamuni, R. J. (1988). On empirical Bayes testing with sequential components, *Ann. Statist.*, **16**, 1270–1282.

Pollak, M. (1985). Optimal detection of a change in distribution, *Ann. Statist.*, **13**, 206–227.

Robbins, H. (1956). An empirical Bayes approach to statistics, *Proc. Third Berkeley Symp. on Math. Statist. Prob.*, Vol. 1, 157–163, University of California Press, Berkeley.

Robbins, H. (1964). The empirical Bayes approach to statistical problems, *Ann. Math. Statist.*, **35**, 1–20.

Shewhart, W. A. (1931). *Economic Control of Quality of Manufactured Product*, Van Nostrand, Princeton, New Jersey.

Shiryayev, A. N. (1963*a*). On optimum methods in quickest detection problem, *Theory Probab. Appl.*, **8**, 22–46.

Shiryayev, A. N. (1963*b*). On the detection of disorder in a manufacturing process I, II, *Theory Probab. Appl.*, **8**, 247–264, 402–413.

Shiryayev, A. N. (1978). *Optimal Stopping Rules*, Springer, New York.

Zacks, S. (1983). A survey of classical and Bayesian approaches to the change-point problem: fixed sample and sequential procedures of testing and estimation, *Recent Advances in Statistics*, 245–269, Academic Press, New York.

Zacks, S. (1991). Detection and change-point problems, *Handbook of Sequential Analysis* (eds. B. K. Ghosh and P. K. Sen), 531–562.

Zacks, S. and Barzily, Z. (1981). Bayes procedures for detecting a shift in the probability of success in a series of Bernoulli trials, *J. Statist. Plann. Inference*, **5**, 107–119.