The Annals of Statistics
1996, Vol. 24, No. 5, 2199–2214

DYNAMIC SAMPLING POLICY FOR DETECTING A CHANGE IN DISTRIBUTION, WITH A PROBABILITY BOUND ON FALSE ALARM

BY BENJAMIN YAKIR

Hebrew University of Jerusalem

We show that if dynamic sampling is feasible, then there exist surveillance schemes that satisfy a probability constraint on false alarm. Procedures are suggested for detecting a change of a normal mean from 0 to a (unknown) positive value. These procedures are optimal (up to a constant term) when the post-change mean is known, and almost optimal [up to an $o(\log(1/\alpha))$ term] when the post-change mean is unknown.

1. Introduction. The subject of this article is the detection of a change in distribution, when a probability constraint on false alarm is required. To illustrate the problem, consider the following example. An agency monitors periodically the level of contamination of the soil in the neighborhood of a factory that produces some hazardous materials. The factory is considered safe if the contamination level does not exceed the safety level set up by regulations. If, however, the contamination does exceed the safety level, severe measures may be taken against the owners of the factory.

Monitoring is based on the measured levels of contamination in samples of soil; hence, it is subject to chance error. Since the declaration by the agency that the contamination level is above the safety limit may have severe consequences, the agency would like to minimize the possibility that such a declaration is made due to sampling error alone. Hence, the statistical procedure employed should keep the probability of a type I error small. (A type I error, in this example, is to declare incorrectly that the safety limits were crossed.)

The classical change point problem setup is one where independent observations $X_1, X_2, \dots, X_{\nu-1}$ are identically distributed according to some F_0, and $X_\nu, X_{\nu+1}, \dots$ are independent and identically distributed according to $F_1(\neq F_0)$. The change point ν is unknown. A detection scheme consists of a stopping time N for the process of observations $X_1, X_2, \dots$ at which one stops and declares a change to have occurred.

The speed of detection of a scheme may be measured by Lorden's functional [Lorden (1971)], namely

$$\sup_{1\le \nu<\infty} \operatorname{ess\,sup} E_\nu\left[(N-\nu+1)^+ \mid X_1,\dots,X_{\nu-1}\right],$$

Received July 1992; revised May 1996.

AMS 1991 *subject classifications*. Primary 62L10; secondary 62N10.

Key words and phrases. Quality control, control charts, change point detection, dynamic sampling.

or by the functional suggested by Pollak and Siegmund (1975)

$$\sup_{1 \le \nu < \infty} E_\nu(N - \nu | N \ge \nu).$$

The rate of false alarm is customarily controlled by allowing only stopping times that satisfy the expectation constraint $E_\infty N \ge B$ for some prespecified constant B. This constraint is adequate when the cost of false alarm and the cost of delay in detection are of the same order, or when the process (and its surveillance) is renewed after N. If this is not the case, and we want to protect ourselves against a false alarm, it is reasonable to consider only stopping times that satisfy the probability constraint

$$P_\infty(N < \infty) \le \alpha, \tag{1}$$

where α is a given constant, $0 < \alpha < 1$.

All the classical procedures for detecting a change in distribution (i.e., Shewhart, CUSUM, Shiryayev–Roberts) have $P_\infty(N < \infty) = 1$. In fact, if a stopping time N satisfies (1) with $\alpha < 1$, then the maximal expected delay (measured by both functionals) is infinite [see Pollak and Siegmund (1975) and Yakir (1995)]. Thus, using a probability bound as a constraint on the rate of false alarm in the traditional setup is problematic.

In order to overcome the problem of an infinite maximal delay when a probability constraint is required, it was suggested that procedures having a more flexible sampling structure be used. Assaf (1988) introduced dynamic sampling into the surveillance context. A dynamic sampling policy allows the statistician to determine the number of samples of soil that are sampled in each monitoring period. The statistican can decide, based on past observations, to sample a larger number of samples in a single period, as long as some budget constraint on the long-run number of samples is satisfied. Assaf (1988) and Assaf and Ritov (1989) showed that employing dynamic sampling policies can result in a dramatic reduction in the maximal expected delay (measured by the Pollak–Siegmund functional). In the non-Bayesian setting, however, their work considered only stopping times that satisfy $P_\infty(N < \infty) = 1$.

When dynamic sampling is feasible, one can allow a probability constraint on the rate of false alarm but at the same time keep the maximal expected delay bounded. This fact was shown in Assaf, Pollak, Ritov and Yakir (1993) for the special model where the observations in each period are distributed as a Brownian motion. The drift of the Brownian motion in that model is 0 before the change and is $\mu > 0$ after the change, where μ is a *known* constant. The speed of detection is measured by Pollak and Siegmund's functional. The budget constraint ensures that as long as $n \le \nu$ the average number of samples sampled up to the nth sampling period is (at most) proportional to n and that the average of the total number of samples is (at most) proportional to the average of the number of monitoring periods. Formally, if t_i is the length of time during which the statistician observes the process on the ith period, then these t_i's must satisfy

$$E_\infty \sum_{i=1}^{n} t_i \le \gamma n, \qquad n = 1, 2, \ldots,$$

$$E_\nu \sum_{i=1}^{N} t_i \le \gamma E_\nu N, \qquad \nu = 1, 2, \ldots,$$

for some prespecified constant γ. It was shown in Assaf, Pollak, Ritov and Yakir (1993) that for this model and under these constraints the probability constraint (1) can still be met, and at the same time the detection of the change point can be done efficiently.

The case expected to be encountered in practice is the case where μ is unknown. In such a case the budget constraint must be met for a wide range of possible μ's. This requirement is not satisfied by the scheme described in Assaf, Pollak, Ritov and Yakir (1993).

In this work we propose the consideration of a mathematically simpler, yet stronger, set of constraints on the sampling budget—an almost sure bound rather than an average bound. Thus, the t_i's are required to satisfy the constraint

$$\sum_{i=1}^{n} t_i \le \gamma n, \qquad n = 1, 2, \ldots . \tag{2}$$

This constraint is natural in the problem of surveillance with a probability bound on false alarm, because a probability bound is likely to be needed when the setting of the problem is a "one shot" setting (as opposed to a renewal setting). Hence, exceeding the budget cannot be compensated for in future repetitions of the monitoring process. (The probability of false alarm and the maximal average delay in this context are over all possible occurrences, rather than "in the long run.") Nevertheless, it will be shown that even though the budget constraint in (2) is more restrictive than the average budget constraint, efficient surveillance schemes are still applicable.

In the next section we study the case where μ is known. A scheme is constructed for which the expected delay is, for any given α, $0 < \alpha < 1$, within a constant of the lowest possible (α-dependent) expected delay. In Sections 3 and 4 the case where μ is unknown is considered. A scheme is constructed for which the expected delay is bounded for all $\mu > 0$, and for which the expected delay is (almost) the minimal possible [uniformly in μ over a compact interval that is bounded away from 0 and up to an $o(\log(1/\alpha))$ term as $\alpha \to 0$].

The functional we use to evaluate the performance of a surveillance scheme is the functional of Pollak and Siegmund. However, some of the results in this paper are still valid if Lorden's functional is used instead.

2. Almost optimality, when the value of the post-change drift is known. We consider the same model as described in Assaf, Pollak, Ritov and Yakir (1993), assuming that if one would take an infinite number of samples during each sampling period, then one would observe independent

Brownian motions with unit variance, having a zero drift before a change, and a positive drift thereafter. The post-change drift parameter μ is assumed to be known. Let ν denote the first sampling period for which the drift parameter is μ. Let $\{B_n(t)\colon 0 \le t < \infty\}$ be the Brownian motion on the nth sampling period and assume that the processes $B_n(t)$, $n = 1, 2, \dots,$ are independent.

If t_n is the duration of time the statistician chooses to observe the process $B_n(t)$, the resulting observed sample path at the nth sampling period is $B_n(t)$ for $0 \le t \le t_n$. Thus, unlike the classical fixed sampling rate, the dynamic sampling setup allows variation on the part of $\{B_n(t)\colon 0 \le t < \infty\}$ actually observed. Formally, t_n is a stopping time for $\{B_n(t)\colon 0 \le t < \infty\}$ such that $\{t_n > t\}$ is in the σ-field generated by $\{B_i(s)\colon 0 \le s \le t_i\}$, $i = 1, 2, \dots, n-1$, and $\{B_n(s)\colon 0 \le s \le t\}$. Denote by $\mathscr{F}_n$ the σ-field generated by $\{B_i(s)\colon 0 \le s \le t_i\}$, $i = 1, 2, \dots, n$. As a restriction on the duration of time the Brownian motions can be observed, consider the set of constraints given in (2).

A change point detection policy (in short, "a policy") is a couple $(N, \mathbf{t})$, where the sampling policy $\mathbf{t} = t_1, t_2, \dots$ is a sequence of sampling times as described above, and N is an integer-valued stopping time with respect to the observed process; that is, $\{N > n\}$ is in the σ-field $\mathscr{F}_n$. Notice that N follows a probability constraint on false alarm for a given α if (1) is true. Furthermore, we expect N to declare that a change has occurred if $\nu < \infty$. For this purpose the following constraints are placed on N:

$$P_\nu(N < \infty) = 1, \qquad 1 \le \nu < \infty. \tag{3}$$

Let $G(\alpha) = \{(N, \mathbf{t})\colon \mathbf{t}$ satisfies (1); N satisfies (2) and (3)$\}$. Then we have the following result.

THEOREM 1. *There exists a constant C (which depends on μ and γ) such that*

$$0 \le \inf_{(N,\mathbf{t}) \in G(\lambda)} \sup_{1 \le \nu < \infty} E_\nu(N - \nu \mid N \ge \nu) - \frac{2}{\gamma\mu^2}\log\frac{1}{\alpha} \le C. \tag{4}$$

PROOF. If $(N, \mathbf{t}) \in G(\alpha)$, then N can be used to define a power-one test of H_0: $\nu = \infty$ vs. H_1: $\nu = 1$, with significance level α. One rejects H_0 if and only if $N < \infty$. As a result of the optimality of the SPRT, it can be concluded that the H_1-expected sampling time until stopping satisfies

$$E_1 \sum_{i=1}^{N} t_i \ge \frac{2}{\mu^2}\log\frac{1}{\alpha}.$$

From (2) we obtain the first inequality in (4).

In order to prove the second inequality in (4), it is enough to construct a policy $(N, \mathbf{t}) \in G(\alpha)$ such that

$$\sup_{1 \le \nu < \infty} E_\nu(N - \nu \mid N \ge \nu) \le \frac{2}{\gamma\mu^2}\log\frac{1}{\alpha} + C \tag{5}$$

for some constant C. In the remainder of the proof such a policy is constructed.

The main idea of the proof is to look at the connected sampling process and to declare a change to have occurred as soon as the connected process crosses a given threshold. The sampling policy is constructed by looking again at the connected process. As long as the process is above a given line, γ units are sampled in each passing sampling period. If the process is below the line, fewer samples are taken; hence, options for future sampling periods are saved.

We begin by constructing the sampling policy $\mathbf{t}$. The t_i's are defined recursively. Let

$$t_1 = \inf\left\{t \geq 0\colon B_1(t) \leq \frac{3}{4}\mu t - \frac{\mu t}{4}\right\} \wedge \gamma.$$

Given that $B_1(t_1) = x_1$, t_2 is defined by the equation

$$t_1 + t_2 = \inf\left\{t \geq t_1\colon B_2(t - t_1) + x_1 \leq \frac{3}{4}\mu t - \frac{2\mu t}{4}\right\} \wedge 2\gamma.$$

In a similar manner, for $n \geq 2$, if $B_{n-1}(t_{n-1}) + x_{n-2} = x_{n-1}$, then t_n is required to satisfy

$$\sum_{i=1}^{n} t_i = \inf\left\{t \geq \sum_{i=1}^{n-1} t_i\colon B_n\left(t - \sum_{i=1}^{n-1} t_i\right) + x_{n-1} \leq \frac{3}{4}\mu t - \frac{n\mu t}{4}\right\} \wedge n\gamma.$$

It is clear from the definition of $\mathbf{t}$ that (2) is met.

We turn now to the determination of the stopping time N. Let

$$n(t) = \max\left\{n \geq \colon \sum_{i=1}^{n} t_i \leq t\right\}.$$

Thus, $n(t) + 1$ is the sampling period that follows the accumulation of a total of t units of sampling time. Denote by $B(t)$ the process obtained by concatenating $\{B_n(t)\colon 0 \leq t \leq t_n\}$, $n = 1, 2, \dots$. In other words, write, for $0 \leq t < \infty$,

$$B(t) = \sum_{i=1}^{n(t)} B_i(t_i) + B_{n(t)+1}\left(t - \sum_{i=1}^{n(t)} t_i\right).$$

Let $T = \inf\{t\colon B(t) \geq (1/\mu)\log(1/\alpha) + \mu t/2\}$ ($T = \infty$ if no such t exist) and let $N = n(T) + 1$.

In order to show that N satisfies (1) and (3), notice that if $\nu = \infty$, then $B(t)$ is a standard Brownian motion. Therefore, $P_\infty(T < \infty) = \alpha$, hence (1). If $\nu < \infty$, then $B(t)$ can be presented as the concatenation of two processes. For t such that $0 \leq t \leq \sum_{i=1}^{\nu-1} t_i$, $B(t)$ is distributed like a standard Brownian motion, and for t such that $\sum_{i=1}^{\nu-1} t_i < t$, $B(t)$ has a positive drift. As a result of (2) it follows that if $\nu < \infty$ and $t \geq \gamma(\nu - 1)$, then the drift of $B(t)$ is μ. By continuity of the sample paths, it follows that $P_\nu(T < \infty) = 1$, $\nu = 1, 2, \dots$, which leads to (3).

We turn now to the investigation of the performance of $(N, \mathbf{t})$ as a surveillance scheme. The following proof shows that $(N, \mathbf{t})$ actually satisfies a stronger version of (5), namely

$$\sup_{1 \le \nu < \infty} \operatorname{ess\,sup} E_\nu\big[(N-\nu)^+ \,|\mathscr{F}_{\nu-1}\big] \le \frac{2}{\gamma\mu^2}\log\frac{1}{\alpha} + C. \tag{6}$$

In order to prove (6), notice that from the strong Markov property of Brownian motions it follows that

$$E_\nu\big[(N-\nu)^+ \,|\mathscr{F}_{\nu-1}\big] = E_\nu\left[(N-\nu)^+ \,\middle|\, B\left(\sum_{i=1}^{\nu-1} t_i\right)\right]. \tag{7}$$

Let $\{B_x(t)\colon t \ge s\}$ be the process $B(t)$ for $t \ge s$, conditioned on the event $\{\sum_{i=1}^{\nu-1} t_i = s,\ B(s) = x\}$. The origin of this process can belong to one of two sets:

$$A_1(\nu-1) = \big\{(x,s)\colon 0 \le s \le \gamma(\nu-1),\ x = \tfrac{3}{4}\mu s - (\nu-1)(\mu\gamma/4)\big\},$$
$$A_2(\nu-1) = \big\{(x,s)\colon s = \gamma(\nu-1),\ \mu\gamma(\nu-1)/2 \le x\big\}.$$

If the origin is in $A_1(\nu-1)$, then not all of the sampling budget has been used by the end of the $\nu-1$ sampling period. On the other hand, if the origin is in $A_2(\nu-1)$, then there are no saved sampling options, but the (vertical) distance from the origin to the stopping line of T is less than the distance from the point $(0,0)$ to that line.

Define an integer-valued r.v. M by the relation

$$\{M < n\} = \left\{B_x(t) > \frac{3}{4}\mu t - n\left(\frac{\mu\gamma}{4}\right),\ \forall t \text{ such that } t \ge s\right\}.$$

Here $M-(\nu-1)$ is a bound on the number of sampling periods that are going to pass until there are no savings at the end of each sampling period. If $(x,s) \in A_1(\nu-1)$, then M is a geometric random variable. If $(x,s) \in A_2(\nu-1)$, then M is bounded by such a geometric random variable.

In order to evaluate (7) we condition on M. If $M-(\nu-1) = n$, then until n sampling periods have elapsed, either N declares a change to have occurred or

$$\left(B_x\left(\sum_{i=1}^{\nu+n-1} t_i\right), \sum_{i=1}^{\nu+n-1} t_i\right) \in A_2(\nu-1).$$

Hence, the process is above the line $\frac{3}{4}\mu t - (n+\nu-1)\mu\gamma/2$ for all $t \ge \sum_{i=1}^{\nu+n-1} t_i$. If $N > \nu+n-1$, then $N-(\nu+n-1)$ is $1/\gamma$ times the real sampling time that passes until N detects the change. This sampling time is less than the (random) time it takes a Brownian motion with a drift parameter μ to reach the line $\mu t/2 + (1/\mu)\log(1/\alpha)$, where the Brownian process is conditional on the event that the process is above the line $\mu t\frac{3}{4}$. In Lemma 1, which is stated and proved next, it is shown that the expected sampling time of the conditional Brownian motion is less than $(2/\mu^2)\log(1/\alpha)$. The proof of the theorem thus follows. □

LEMMA 1. *Let $X(t)$ be a Brownian motion with a drift parameter $\theta > 0$. Consider, for $b > 0$, the random time $T = \inf\{t\colon X(t) \ge b\}$. Given a, $0 < a < \theta$, and c, $c \le 0$,*

$$E(T|X(s) > c + as,\ \forall s,\ 0 < s < \infty) \le b/\theta.$$

PROOF. Let $Y(t)$ be the process $X(t) - \theta t$, conditioned on the event $\{X(s) > c + as,\ \forall\ s,\ 0 \le s < \infty\}$. It will be shown that $Y(t)$ is an F_t-submartingale, where

$$F_t = \sigma\{X(s)\colon 0 \le s \le t\}, \qquad 0 \le t.$$

Indeed, if $X(t) = x$, then

$$\begin{aligned} &E(Y(t+u)|F_t) - Y(t) \\ &\quad = E(X(t+u) - X(t) - \theta u|X(s) > c + as, \forall s, 0 \le s < \infty) \\ &\quad =_{\text{dist.}} E(X(u) - \theta u|X(s) - \theta s > c + at - x + (a-\theta)s, \\ &\qquad\qquad \forall s, 0 \le s < \infty). \end{aligned} \tag{8}$$

Notice that $X(u) - \theta u$ is a standard Brownian motion, $a - \theta < 0$ and $c + at - x < 0$.

Let

$$K = \{X(s) - \theta s > c + at - x + (a-\theta)s, \forall s, 0 \le s < \infty\}.$$

Since

$$0 = E(X(u) - \theta u) = E(X(u) - \theta u|K)P(K) + E\big(X(u) - \theta u|\overline{K}\big)P(\overline{K}),$$

it is enough, in order to prove that the last term in (8) is positive, to prove that $E(X(u) - \theta u|\overline{K}) \le 0$.

Let S be the first time the process hits the line $c + at$ (given that it does). By conditioning on S we get

$$\begin{aligned} E\big(X(u) - \theta u|\overline{K}\big) &= E[(c + at - x + (a-\theta)S)\mathbb{1}(S \le u)] \\ &\quad + \int_u^\infty E(X(u) - \theta u|S = s)\,dP(S \le s). \end{aligned}$$

The integrand is negative; hence, $Y(t)$ is indeed a submartingale.

Since the stopping time T is bounded by $(b - c)/a$, we get by the optional sampling theorem that

$$0 = Y(0) \le EY(T) = b - \theta E(T|X(s) > c + as, \forall s, 0 \le s < \infty).$$

The proof of Lemma 1 (thence the proof of Theorem 1) is concluded. □

REMARK 1. Theorem 1 is similar to Theorem 2 is Assaf, Pollak, Ritov and Yakir (1993). Notice that the current result is stronger, because (2) is a stronger set of constraints compared to the set of constraints considered in Assaf, Pollak, Ritov and Yakir (1993).

REMARK 2. From (6) it follows that Theorem 1 can be reformulated for a Lorden-type functional.

REMARK 3. If M is distributed as a geometric r.v., then the parameter of that r.v. is $p = 1 - \exp\{-\mu^2/8\}$, thus $C \le 1/p$. Notice that this bound cannot be improved by any policy $(N, \mathbf{t})$, having the same structure as the policy in the proof of Theorem 1, for which the sampling policy $\mathbf{t}$ is determined by lines of the form $b\mu t - n(1-b)\mu\tau$, $n = 1, 2, \dots,$ where $\frac{1}{2} < b < 1$ is some constant.

REMARK 4. The condition $0 < a < \theta$ in Lemma 1 is technical. Actually, a stronger version of Lemma 1 can be proved, for which the upper and lower lines are not required to be linear and the conditioned stopping time is proved to be stochastically smaller than the unconditioned stopping time [see Yakir (1990)].

3. A detection policy when the value of the post-change drift is unknown. In this section we consider the situation in which the drift parameter after the change is unknown, but positive. The model is basically the model of the previous section, but $\mu > 0$ does not have a prespecified value. We regard pairs $(N, \mathbf{t})$ as in the case for a known μ, but (4) is replaced by the more restrictive set of constraints:

$$P_{\nu,\mu}(N < \infty) = 1 \quad \text{for } 1 \le \nu < \infty,\ \mu > 0. \tag{9}$$

where $P_{\nu,\mu}$ is the distribution on the sequence of Brownian motions, for which the drift parameter changes to μ at the νth sampling period.

Let $G^*(\alpha) = \{(N, \mathbf{t})\colon \mathbf{t}$ satisfies (1), N satisfies (2) and (9)$\}$. It is not obvious that $G^*(\alpha)$ is not empty. Notice, for example, that the policy that was constructed in Theorem 1 does not belong to $G^*(\alpha)$. In the following theorem a policy that does belong to $G^*(\alpha)$ is described.

THEOREM 2. *There exists a policy* $(N, \mathbf{t}) \in G^*(\alpha)$. *Furthermore,* $(N, \mathbf{t})$ *satisfies*

$$\sup_{1 \le \nu < \infty} E_{\nu,\mu}(N - \nu \mid N \ge \nu) < \infty \quad \textit{for } \mu > 0. \tag{10}$$

PROOF. Let $\{\mu_k\}_{k=1}^{\infty}$, $\{\gamma_k\}_{k=1}^{\infty}$ and $\{\alpha_k\}_{k=1}^{\infty}$ be sequences of positive numbers, such that $\mu_k \to 0$, $\sum_{k=1}^{\infty} \gamma_k = \gamma$ and $\sum_{k=1}^{\infty} \alpha_k = \alpha$.

We construct an infinite triangular array of sampling times t_n^k, $k \le n < \infty$. For a given k, the sequence of stopping times $t_k^k, t_{k+1}^k, \dots, t_n^k, \dots$ is constructed in a similar way to the construction of $\mathbf{t}$ in the previous section, with $\mu = \mu_k$ and $\gamma = \gamma_k$. The construction is done in a recursive manner.

Consider first t_1^1, where

$$t_1^1 = \inf\left\{t \ge 0\colon B_1(t) \le \tfrac{3}{4}\mu_1 t - \tfrac{1}{4}\mu_1\gamma_1\right\} \wedge \gamma_1.$$

Given that $B_1(t_1^1) = x_1^1$, we define t_2^1 by solving the equation

$$t_1^1 + t_2^1 = \inf\left\{t \ge t_1^1\colon B_2\left(t - t_1^1\right) + x_1^1 \le \tfrac{3}{4}\mu_1 t - \tfrac{2}{4}\mu_1\gamma_1\right\} \wedge 2\gamma_1.$$

The stopping time t_2^2 is constructed by observing the process $B_2(t + t_2^1) - B_2(t_2^1)$:

$$t_2^2 = \inf\left\{t \ge 0\colon B_2\left(t + t_2^1\right) - B_2\left(t_2^1\right) \le \tfrac{3}{4}\mu_2 t - \tfrac{1}{4}\mu_2\gamma_2\right\} \wedge \gamma_2.$$

We continue in this manner to define t_n^1 by conditioning on $B_{n-1}(t_{n-1}^1) + x_{n-2}^1 = x_{n-1}^1$ and solving

$$\sum_{i=1}^{n} t_i^1 = \inf\left\{t \geq \sum_{i=1}^{n-1} t_i^1 : B_n\left(t - \sum_{i=1}^{n-1} t_i^1\right) \leq \frac{3}{4}\mu_1 t - \frac{n}{4}\mu_1\gamma_1\right\} \wedge n\gamma_1.$$

In a similar fashion, for $1 < k < n$, we condition on $B_{n-1}(\sum_{j=1}^{k} t_{n-1}^j) - B_{n-1}(\sum_{j=1}^{k-1} t_{n-1}^j) + x_{n-2}^k = x_{n-1}^k$ and define t_n^k by solving

$$\sum_{i=k}^{n} t_i^k = \inf\left\{t \geq \sum_{i=k}^{n-1} t_i^k : B_n\left(t + \sum_{j=1}^{k-1} t_n^j\right) - B_n\left(\sum_{j=1}^{k-1} t_n^j\right) + x_{n-1}^k \leq \frac{3}{4}\mu_k t - \frac{n-k+1}{4}\mu_k\gamma_k\right\}$$

$$\wedge (n-k+1)\gamma_k.$$

Finally,

$$t_n^n = \inf\left\{t \geq 0 : B_n\left(t + \sum_{j=1}^{n-1} t_n^j\right) - B_n\left(\sum_{j=1}^{n-1} t_n^j\right) \leq \tfrac{3}{4}\mu_n t - \tfrac{1}{4}\mu_n\gamma_n\right\} \wedge \gamma_n.$$

The sampling policy $\mathbf{t}$ is constructed by letting

$$t_n = \sum_{j=1}^{n} t_n^j, \qquad n = 1, 2, \dots .$$

From the construction it follows that

$$\sum_{i=1}^{n} t_i = \sum_{i=1}^{n} \sum_{j=1}^{i} t_i^j = \sum_{j=1}^{n} \sum_{i=j}^{n} t_i^j \leq \sum_{j=1}^{n} (n-j+1)\gamma_j \leq n\gamma.$$

Hence, (2) is satisfied.

In order to define the stopping time N, we denote by T_k and N_k the random times constructed by observing the process $B_k(t)$. This process is formed by concatenating the sequence of processes

$$\left\{B_n\left(t + \sum_{j=1}^{k-1} t_n^j\right) - B_n\left(\sum_{j=1}^{k-1} t_n^j\right) : 0 \leq t \leq t_n^k\right\}_{n=k}^{\infty}.$$

[The construction of T_k and N_k is similar to the way T and N were constructed by observing the process $B(t)$ in the previous section.]

We return to the construction of N. If $N_1 = 1$, then $N = 1$. Otherwise, if $N_1 > 1$ but $N_1 \wedge N_2 = 2$, then $N = 2$. In general, if $\bigwedge_{i=1}^{n-1} N_i > n-1$ but $\bigwedge_{i=1}^{n} N_i = n$, then $N = n$.

The policy N is finite if and only if any of the policies N_k, $1 \leq k < \infty$, is finite. Therefore,

$$P_\infty(N < \infty) = P_\infty\left(\bigcup_{k=1}^{\infty}\{N_k < \infty\}\right) \leq \sum_{k=1}^{\infty} \alpha_k = \alpha.$$

Furthermore, since $\{N = \infty\} \subset \{N_k = \infty\}$, it follows that

$$P_{\nu,\mu}(N = \infty) \le P_{\nu,\mu}(N_k = \infty) = 0 \quad \text{for } 1 \le \nu < \infty,\ \mu \ge \mu_k.$$

Therefore, $(N, \mathbf{t}) \in G^*(\alpha)$.

In order to show (10), let $\mu > 0$ be given and let k be such that $\mu \ge \mu_k$. For $\nu \ge k$ it is true that

$$E_{\nu,\mu}(N - \nu | N \ge \nu) = \frac{E_{\nu,\mu}(N - \nu, N \ge \nu)}{P_{\nu,\mu}(N \ge \nu)} \le \frac{E_{\nu,\mu}(N - \nu, N \ge \nu)}{1 - \alpha}.$$

However, since $\{N \ge \nu\} \subset \{N_k \ge \nu\}$ and since $N \le N_k$, it follows that

$$\begin{aligned} E_{\nu,\mu}(N - \nu, N \ge \nu) &\le E_{\nu,\mu}(N_k - \nu, N_k \ge \nu) \\ &\le \sup_{1 \le \nu < \infty} E_{\nu,\mu}(N_k - \nu | N_k \ge \nu). \end{aligned}$$

Moreover, for $1 \le \nu < k$, N_k behaves as if the change has occurred in the first sampling period in which it was activated—the kth period. Therefore,

$$\sup_{1 \le \nu < \infty} E_{\nu,\mu}(N - \nu | N \ge \nu) \le \frac{k + \sup_{k \le \nu < \infty} E_{\nu,\mu}(N_k - \nu | N_k \ge \nu)}{1 - \alpha}$$

from which (10) follows. □

4. Almost optimality, when the value of the post-change drift is unknown. In this section we prove that when dynamic sampling is applicable then the average run length to detection of a change is comparable to the average run length to rejection of the null hypothesis in power-one sequential testing. Hence, one loses almost nothing by not knowing in advance the sampling period in which the drift may have changed. Formally, $(N, \mathbf{t})$ can be used as a test of H_0: $\nu = \infty$ vs. H_1: $\nu = 1$, with significance level α. It follows from Pollak (1978) that

$$\begin{aligned} &\inf_{(N,\mathbf{t}) \in G^*(\alpha)} \sup_{a \le \mu \le b} \sup_{1 \le \nu < \infty} \gamma \mu^2 E_{\nu,\mu}(N - \nu | N \ge \nu) \\ &\qquad \ge 2 \log \frac{1}{\alpha} + \log \log \frac{1}{\alpha} + O(1), \end{aligned}$$

where $0 < a < b < \infty$ and $\lim_{\alpha \to 0} |O(1)| < \infty$. We now state the main result of this section.

THEOREM 3. *For dynamic sampling*

$$\inf_{(N,\mathbf{t}) \in G^*(\alpha)} \sup_{a \le \mu \le b} \sup_{1 \le \nu < \infty} \gamma \mu^2 E_{\nu,\mu}(N - \nu | N \ge \nu) = 2 \log \frac{1}{\alpha}(1 + o(1)), \tag{11}$$

where $0 < a < b < \infty$ *and* $\lim_{\alpha \to 0} o(1) = 0$.

We prepare the groundwork for the proof of Theorem 3 by stating and proving two lemmas. These lemmas deal with the H_1-expected sampling time of a mixed SPRT.

Assume that $X_i \sim N(\theta_i, 1)$, $i = 1, 2, \ldots,$ and define

$$S_{k,n} = \sum_{i=k+1}^{n+k} X_i, \qquad S_n = S_{0,n}.$$

Define, for $0 < b^* < \infty$,

$$f(x,t) = \frac{1}{b^*}\int_0^{b^*} \exp\left(\theta x - \frac{\theta^2}{2}t\right) d\theta.$$

The stopping time of a mixed SPRT (when the uniform distribution is used as the mixing distribution) is

$$T = \inf\{n\colon f(S_n, n) \geq 1/\alpha\} \qquad (T = \infty \text{ if no such } n \text{ exists}).$$

LEMMA 2. *Let π be given, $0 < a_1 \leq \pi \leq b_1 < b^*$. If $\theta_i = \pi$, for all i, then*

$$\pi^2 ET \leq 2\log\frac{1}{\alpha} + \log\log\frac{1}{\alpha} + M^*, \tag{12}$$

where M^* depends on b^* but not on α or π.

PROOF. This is merely a restatement of Lemma 3 in Pollak (1978).

LEMMA 3. *Let π be given, $0 < a_1 \leq \pi \leq b_1 < b^*$. Suppose that $\theta_i = 0$ for $1 \leq i < k$ and $\theta_i = \pi$ for $i > k$. If, for some constant M,*

$$E(S_k^- | T \geq k) \leq kM,$$

then

$$\pi^2 E(T - k | T \geq k) \leq 2\log\frac{1}{\alpha} + \log\log\frac{1}{\alpha} + kC^*, \tag{13}$$

where C^* depends on b^* only.

PROOF. Compute the left-hand side of inequality (13) by conditioning on S_k. On the event $\{T \geq k,\ S_k = y\}$ we have

$$T - k = \begin{cases} \inf\left\{n\colon f(S_{k,n} + y, n + k) \geq \dfrac{1}{\alpha}\right\}, & f(y,k) < \dfrac{1}{\alpha}, \\ 0, & f(y,k) \geq \dfrac{1}{\alpha}. \end{cases} \tag{14}$$

Now

$$\begin{aligned} f(S_{k,n} + y, n + k) &= \frac{1}{b^*}\int_0^{b^*} \exp\left\{\theta(S_{k,n} + y) - \frac{\theta^2}{2}(n + k)\right\} d\theta \\ &= \frac{1}{b^*}\int_0^{b^*} \exp\left\{\theta y - k\frac{\theta^2}{2}\right\}\exp\left\{\theta S_{k,n} - n\frac{\theta^2}{2}\right\} d\theta. \end{aligned}$$

Therefore,

$$f(S_{k,n} + y, n + k) \geq \exp\left\{b^*(y \wedge 0) - k\frac{(b^*)^2}{2}\right\} f(S_{k,n}, n).$$

Returning to (14), it can be concluded that (on the event $\{T \ge k,\ S_k = y\}$)

$$T - k \le \begin{cases} \inf\left\{n\colon f(S_{k,n}, n) \ge \dfrac{1}{\alpha}\exp\left\{-b^*(y \wedge 0) + k\dfrac{(b^*)^2}{2}\right\}\right\}, & f(y,k) < \dfrac{1}{\alpha}, \\ 0, & f(y,k) \ge \dfrac{1}{\alpha}. \end{cases}$$

Moreover, if $f(y,k) < 1/\alpha$, then $\exp\{b^*(y \wedge 0) - k(b^*)^2/2\} < 1/\alpha$. Thus, by Lemma 2 [for a significance level of $\alpha \exp\{b^*(y \wedge 0) - k(b^*)^2/2\}$],

$$\begin{aligned} \pi^2 E(T - k | T \ge k,\ S_k = y) \le{}& 2\log\Big(1/\big[\alpha\exp\{b^*(y \wedge 0) - k(b^*)^2/2\}\big]\Big) \\ &+ \log\log\Big(1/\big[\alpha\exp\{b^*(y \wedge 0) - k(b^*)^2/2\}\big]\Big) \\ &+ M^*. \end{aligned}$$

The fact that if $a > 1$ and $b > 0$ then $\log(a + b) \le b + \log a$ and some easy calculations lead to the proof of the lemma. □

PROOF OF THEOREM 3. The idea behind the proof is to notice that if the change occurred in the first period, then the stopping time of a mixed SPRT, based on sampling the process at the end of each constant time interval of length γ, satisfies (11). This statement is still correct if the change occurred after a small number of sampling periods.

Consider, for some β, $\beta \gg \alpha$, a policy $(N_1, \mathbf{t}^1)$ that satisfies (10) and $(N_1, \mathbf{t}^1) \in G^*(\beta)$. If we use this policy until it declares a change to have occurred, and then use a mixed SPRT (with significance level α) as our stopping rule, we will get that, with high probability, the number of sampling periods it takes for the detection of a change is only slightly larger than the number of sampling periods until H_0 is rejected by the SPRT.

We prove the theorem by constructing an appropriate policy $(N, \mathbf{t})$. Choose $0 < \beta < 1$. Let m be a natural number to be determined later. Let $\{(N_i, \mathbf{t}^i)\}_{i=2}^{\infty}$ be independent copies of $(N_1, \mathbf{t}^1)$.

Start with the definition of $\mathbf{t}$. Let $\mathbf{t} = \mathbf{t}^1$ during the first N_1 sampling periods. After these sampling periods, sample the process for m periods at a constant rate. Therefore, $t_i = \gamma$ for $N_1 < i \le N_1 + m$ and the first cycle is complete. Set $\mathbf{t} = \mathbf{t}^2$ for the next N_2 sampling periods, after which sample at a constant rate and so on and so forth.

Here N is defined by the process over the constant rate periods. More precisely, it is defined by the sequence

$$X_k = \frac{1}{\sqrt{m\gamma}}\left[B\big(t_{\bar{\tau}(k)}\big) - B\big(t_{\underline{\tau}(k)}\big)\right], \qquad k = 1, 2, \dots,$$

where $\bar{\tau}(k) = \sum_{i=1}^{k}(N_i + m)$ and $\underline{\tau}(k) = \sum_{i=1}^{k-1}(N_i + m) + N_k$.

Let S_n and $f(x,t)$ be defined as before, and let $T = \inf\{n: f(S_n, n) \ge 1/\alpha\}$. Notice that N declares a change to have occurred at the end of the first cycle at which T rejects H_0. Hence, $N = \sum_{i=1}^{T}(N_i + m)$.

Given the sequence $N_0 = 0, N_1, N_2, \dots,$ notice that the X_k's are independent and normally distributed with variance 1 and mean parameter θ_k. If $\nu = \infty$, then $\theta = 0$ for all k. Otherwise,

$$\theta_k = \begin{cases} \mu(m\gamma)^{1/2}, & \nu \le \sum_{i=1}^{k-1}(N_i + m) + N_k, \\ \mu\left(\left(\sum_{i=1}^{k}(N_i + m) - \nu\right)\dfrac{\gamma}{m}\right)^{1/2}, & 0 < \nu - \sum_{i=1}^{k-1}(N_i + m) + N_k \le m, \\ 0, & \sum_{i=1}^{k}(N_i + m) < \nu. \end{cases}$$

Therefore, $P_\infty(T < \infty) \le \alpha$ and $P_{\nu,\mu}(T < \infty) = 1$ for $1 \le \nu < \infty$ and $\mu > 0$.

Since $\{N < \infty\} \subset \{T < \infty\}$ it follows that $P_\infty(N < \infty) < \alpha$. Furthermore, due to the fact that

$$\bigcap_{i=1}^{\infty}\{N_i < \infty\} \cap \{T < \infty\} \subset \{N < \infty\}$$

it can be shown that $P_{\nu,\mu}(N < \infty) = 1$ for $1 \le \nu < \infty$ and $\mu > 0$. Hence, indeed, $(N, \mathbf{t}) \in G^*(\alpha)$.

In order to evaluate $E_{\nu,\mu}(N - \nu | N \ge \nu)$ let

$$J = \sup\left\{k \ge 0: \sum_{i=1}^{k-1}(N_i + m) + N_k < \nu\right\}$$

be the number of times the sequence $N_0, N_1, \dots$ declares a change to have occurred before the real change point ν. For any given ν, J is a bounded r.v. Hence,

$$\begin{aligned} &E_{\nu,\mu}(N - \nu | N \ge \nu) = \sum_{j=0}^{\infty} E_{\nu,\mu}(N - \nu | N \ge \nu,\ J = j) \\ (15)\qquad &P_{\nu,\mu}(J = j | N \ge \nu) \le \frac{P_{\nu,\mu}(J = j)}{P_{\nu,\mu}(N \ge \nu)} \le \frac{\beta^j}{1 - \alpha}. \end{aligned}$$

The next step is to bound $E_{\nu,\mu}(N - \nu | N \ge \nu,\ J = j)$. In order to do that, notice that if $J = 0$, then $N_1 \ge \nu$ and $\theta_k = \mu(m\gamma)^{1/2}$ for all k. Using Lemma 2, one gets (since $\{J = 0\} \subset \{N \ge \nu\}$)

$$m\gamma\mu^2 E(T | N \ge \nu,\ J = 0) \le 2\log\frac{1}{\alpha} + \log\log\frac{1}{\alpha} + M^*$$

for all μ, $a \le \mu \le b$. Now, from the definition of N it follows that

$$\begin{aligned}
&E_{\nu,\mu}(N - \nu | N \ge \nu,\, J = 0)\\
&\qquad = E_{\nu,\mu}\left(\sum_{i=1}^{T} (N_i + M) - \nu | N_1 \ge \nu \right)\\
&\qquad \le E_{\nu,\mu}(N_1 - \nu | N_1 \ge \nu) + m + E_{1,\mu}\left(\sum_{i=2}^{T} (N_i + m) \right).
\end{aligned}$$

For $i \ge 2$ the N_i's are independent of T. Therefore, for $a \le \mu \le b$,

$$\gamma \mu^2 E_{\nu,\mu}(N - \nu | N \ge \nu,\, J = 0) \le \frac{m + L}{m}\left(2 \log \frac{1}{\alpha} + \log\log \frac{1}{\alpha} + M^* \right), \tag{16}$$

where

$$L = \sup_{a \le \mu \le b}\ \sup_{1 \le \nu < \infty} E_{\nu,\mu}(N_1 - \nu | N_1 \ge \nu).$$

In a similar fashion, for $j > 0$,

$$\theta_k = \begin{cases}
\mu(m\gamma)^{1/2}, & j < k,\\
\mu\left(\left(\sum_{i=1}^{j} (N_i + m) - \nu\right)\dfrac{\gamma}{m}\right)^{1/2}, & 0 \le \nu - \sum_{i=1}^{j-1} (N_i + m) + N_j \le m,\\
0, & 0 < \nu - \sum_{i=1}^{j} (N_i + m) \le N_{j+1},\\
0, & j > k.
\end{cases}$$

If $J = j$ and $N \ge \nu$, then either

$$\sum_{i=1}^{j-1} (N_i + m) + N_j < \nu \quad \text{and} \quad \sum_{i=1}^{j} (N_i + m) \ge \nu$$

or

$$\sum_{i=1}^{j-1} (N_i + m) < \nu \quad \text{and} \quad \sum_{i=1}^{j} (N_i + m) + N_{j+1} > \nu.$$

Nonetheless, in both cases, $\{N \ge \nu,\ J = j\} \subset \{T \ge j,\ J = j\}$.

On the event $\{N \ge \nu,\ J = j\}$ it is true that

$$\begin{aligned}
N - \nu &= \sum_{i=j+1}^{T} (N_i + m) - \left(\nu - \sum_{i=1}^{j} (N_i + m) \right)\\
&\le N_{j+1} - \left(\nu - \sum_{i=1}^{j} (N_i + m) \right) + m + \sum_{i=j+2}^{T} (N_i + m).
\end{aligned}$$

Conditional on the event $\{N \ge \nu,\ J = j\}$, the sequence $N_{j+2}, N_{j+3}, \dots$ is independent of T. Therefore, for all $a \le \mu \le b$,

$$E_{\nu,\mu}\left(\sum_{i=j+2}^{T}(N_i+m)|N \ge \nu,\, J=j\right)$$
$$\le (L+m)E_{\nu,\mu}(T-j-1|N\ge \nu,\, J=j).$$

Moreover, by conditioning on $\nu - \sum_{i=1}^{j}(N_i+m)$, it can be shown that

$$E_{\nu,\mu}\left[N_{j+1} - \left(\nu - \sum_{i=1}^{j}(N_i+m)\right)\middle| N\ge \nu,\, J=j\right] \le L+m.$$

Furthermore, if $T=j$, then $N-\nu \le m$. The above results can be summarized in the inequality:

$$E_{\nu,\mu}(N-\nu|N\ge\nu,\, J=j) \le m + (m+L)E_{\nu,\mu}(T-j|N\ge\nu,\, J=j). \tag{17}$$

Given the event $\{J=j\}$, the probability of the event $\{N<\nu\}$ is less than α. Thus,

$$E_{\nu,\mu}(T-j|N\ge\nu,\, J=j) \le \frac{E_{\nu,\mu}(T-j|T\ge j,\, J=j)}{1-\alpha}.$$

Lemma 3 is applicable since

$$E_{\nu,\mu}\left(S_j^-|T\ge j,\, J=j\right) \le \frac{E_{\nu,\mu}\left(S_j^-|J=j\right)}{1-\alpha} \le \frac{jEZ^+}{1-\alpha},$$

where Z is a standard normal. From (17) it follows that

$$\begin{aligned} &\gamma\mu^2 E_{\nu,\mu}(N-\nu|N\ge\nu,\, J=j)\\ &\quad\le \gamma\mu^2 m + \frac{1}{1-\alpha}\,\frac{m+L}{m}\left(2\log\frac{1}{\alpha} + \log\log\frac{1}{\alpha} + C^*j\right). \end{aligned} \tag{18}$$

All that is needed, in order to complete the proof of Theorem 3, is to tie up the loose ends. Combining (15), (16) and (18), one gets, for all $a\le\mu\le b$,

$$\gamma\mu^2 E_{\nu,\mu}(N-\nu|N\ge\nu) \le B_1\left(2\log\frac{1}{\alpha} + \log\log\frac{1}{\alpha}\right) + B_2,$$

where

$$\begin{aligned} B_1 &= \left(1+\frac{L}{m}\right)/(1-\alpha)\\ &\ge \left(1+\frac{L}{m}\right)\frac{1-\alpha P_{\nu,\mu}(J=0|N\ge\nu)}{1-\alpha}\\ &= \left(1+\frac{L}{m}\right)\left[\sum_{j=0}^{\infty}\frac{P_{\nu,\mu}(J=j|N\ge\nu)}{1-\alpha} + P_{\nu,\mu}(J=0|N\ge\nu)\right] \end{aligned}$$

and

$$B_2 = \sup_{\alpha \le \mu \le b} \sup_{0 < \alpha < \alpha_0} \sup_{1 \le \nu < \infty} \left(1 + \frac{L}{m}\right) \left[\sum_{j=1}^{\infty} \frac{jC^* + m\gamma b^2(1-\alpha)}{(1-\alpha)^2} \beta^j + M^* P_{\nu,\mu}(J = 0 | N \ge \nu) \right].$$

Notice that B_1 and B_2 do not depend on ν or μ. Moreover, B_2 is bounded and B_1 satisfies $\lim_{\alpha \to 0} B_1 = 1 + L/m$. The limit, $1 + L/m$, can be made as close to 1 as one wishes by choosing m large. This completes the proof of the theorem. □

Acknowledgment. This work is based on a Ph.D. dissertation, written under the supervision of Professor Moshe Pollak and submitted to the Hebrew University of Jerusalem.

REFERENCES

ASSAF, D. (1988). A dynamic sampling approach for detecting a change in distribution. *Ann. Statist.* **16** 236–253.

ASSAF, D., POLLAK, M., RITOV, Y. and YAKIR, B. (1993). Detecting a change of a normal mean by dynamic sampling with a probability bound on a false alarm. *Ann. Statist.* **21** 1155–1165.

ASSAF, D. and RITOV, Y. (1989). Dynamic sampling procedures for detecting a change in the drift of Brownian motion: a non-Bayesian model. *Ann. Statist.* **17** 793–800.

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

POLLAK, M. (1978). Optimality and almost optimality of mixture stopping rules. *Ann. Statist.* **6** 910–916.

POLLAK, M. and SIEGMUND, D. (1975). Approximations to the expected sample size of certain sequential tests. *Ann. Statist.* **3** 1267–1282.

YAKIR, B. (1990). Optimal dynamic sampling policy of detecting a change in distribution, with a probability bound on false alarm. Ph.D. dissertation, Dept. Statistics, Hebrew Univ. Jerusalem.

YAKIR, B. (1995). A lower bound on the ARL to detection of a change with a probability constraint on false alarm. Unpublished manuscript.

DEPARTMENT OF STATISTICS
HEBREW UNIVERSITY OF JERUSALEM
JERUSALEM 91905
ISRAEL
E-MAIL: msby@olive.mscc.huji.ac.il