The Annals of Statistics
1989, Vol. 17, No. 2, 793–800

DYNAMIC SAMPLING PROCEDURES FOR DETECTING A CHANGE IN THE DRIFT OF BROWNIAN MOTION: A NON-BAYESIAN MODEL[1]

BY DAVID ASSAF AND YA'ACOV RITOV

The Hebrew University of Jerusalem

We consider dynamic procedures for sampling from a process (a Brownian motion) and stopping it after a change is detected. The basic idea is to conduct a sequence of similar SPRT's, each one of them done in negligible time, while not sampling at all between them. The procedures detect the change point much faster than the standard procedures with the same sampling rate and time to false alarm, but hold the sampling rate constant.

1. Introduction. Consider a production process in which the quality of the product changes at some unknown time ν, $0 \leq \nu \leq \infty$. Assume that the quality of the output can be sampled on a continuous basis without any restriction on the instantaneous rate. There are, however, economic considerations which restrict the average sampling rate. To fix ideas, consider either a chemical reactor or the quality control of a transistor production line. We seek a statistical procedure which determines both a sampling procedure and a stopping rule to "detect" the change point ν. The performance of a procedure is evaluated on the basis of the average time until the detection of the change point and the average rate of sampling. For application of this model the reader may consult Wilson, Griffiths, Kemp, Nix and Rowlands (1979) and Pollak and Siegmund (1985).

The same model but with a constant sampling rate is well studied. Important references include Page (1954), van Dobben de Bruyn (1968), Lorden (1971) Roberts (1966), Shiryayev (1963) and Pollak and Siegmund (1985). The standard procedures are the Page (or CUSUM) procedure and the Roberts–Shiryayev procedure, which are to be described now.

Let $L(t', t)$ be the likelihood that $\nu = t'$ given the history of the process up to time t. Then the CUSUM procedure stops at

$$\inf\left\{t: \sup_{0 \leq s \leq t} L(s, t) \geq A_C\right\}$$

for some constant A_C. The Roberts–Shiryayev stopping time stops at

$$\inf\left\{t: \int_0^t L(s, t)\, ds \geq A_{RS}\right\}$$

for some constant A_{RS}.

Received July 1987; revised July 1988.

[1]Research supported by the Basic Research Foundation administered by the Israel Academy of Sciences and Humanities.

AMS 1980 *subject classifications*. 62L10, 62L20.

Key words and phrases. CUSUM procedure, Roberts–Shiryayev procedure, SPRT, Brownian motion, dynamic sampling.

Assaf (1988) and Assaf and Ritov (1988) study similar problems in a Bayesian framework with "memoryless" priors. The optimality of the "dynamic" procedure suggested in this paper stems from the fact that it is actually equivalent to some Bayes procedure with appropriate generalized prior (see Appendix). The procedure can be described as a limit of procedures in which the maximal permissible sampling rate converges to infinity. In the limit we either sample as fast as we can or we do not sample at all. As a result we conduct a sequence of SPRT's, each of them is done in "zero" time. The limiting procedure is determined when the time interval between these SPRT's shrinks to zero (keeping the average sampling rate constant). The procedures are described in Section 2 and analyzed in Section 3.

Finally, our procedures are compared to the CUSUM procedure under constant sampling rate. To make the comparison meaningful we compare procedures with the same average sampling rate before the change point ν and with the same expected time until false alarm. The comparison shows that the dynamic procedure stops the process after sampling the same amount on the average as the procedure with constant sampling rate. The superiority of the dynamic procedure is revealed when we compare the expected time from the change time to its detection. Under typical conditions the sampling rate after ν is much faster than it was on $[0, \nu)$. As a result the dynamic procedure detects the change within a much shorter time. A numerical example for the case that Pollak and Siegmund (1985) consider as typical shows that the time until detection under the dynamic scheme is approximately 20% of its value under the corresponding case with constant sampling rate.

2. Description of the procedure. We consider the limiting case when the instantaneous sampling rate can be infinite. This represents the situation in which we can sample and analyze the data as much as needed in a time which is very short relative to the time constant of the production process.

The procedures we suggest can be parameterized by three positive parameters A, C and δ. The (A, C, δ) procedure itself is a limit of procedures in which the maximal instantaneous sampling rate converges to infinity and may be described as follows. At the time instances $\delta, 2\delta, \ldots$ until the stopping time the process is being sampled. (A slight modification will be introduced later.) This can be thought of as if we temporarily stop the process and begin sampling the present output until we reach one of two possible decisions: either continue with the process until the next sampling instant or declare that the change point has been detected. During each instant the process is assumed to be a Brownian motion describing the value of the sufficient statistics after the amount τ was sampled:

$$X_i(0) = 0,$$

$$dX_i(\tau) = \begin{cases} dB_i(\tau), & \tau > 0,\ i\delta < \nu, \\ \mu\, d\tau + dB_i(\tau), & \tau > 0,\ i\delta \geq \nu, \end{cases}$$

where $B_i(\cdot)$, $i = 1, 2, \ldots$, are independent copies of a standard Brownian motion. The drift after the change point is denoted by μ. The procedure will be

constructed as if $\mu = \mu_0$ where μ_0 is a constant known to the statistician, but will be analyzed under the general assumption that the drift is some μ. For each $i \geq 1$ the process $X_i(\cdot)$ is observed until the stopping time,

$$\tau_i^* = \inf\{\tau > 0: X_i(\tau) - \tfrac{1}{2}\mu_0\tau \notin (-C\delta, A)\}.$$

The change time is declared at the stopping time,

$$T = \inf\{i\delta: X_i(\tau_i^*) = A + \tfrac{1}{2}\mu_0\tau_i^*\}.$$

We will be mainly interested in the limiting case where $\delta \to 0$ while $A, C \in (0, \infty)$ remain fixed. With some abuse of notation we denote the limiting case by $(A, C, 0)$.

When discussing the (A, C, δ) family of procedures it will be convenient to distinguish between the real time t and the "sampling time." The latter represents the amount sampled in a given period. Note that with constant rate sampling these notions differ only in scale.

We next compare the (A, C, δ) family of procedures to the CUSUM procedure. The CUSUM procedure is essentially a maximum likelihood test. It can be described as follows. We observe a diffusion process with a constant variance 1 and a drift which changes from 0 to μ at an unknown time ν. We keep at time t the minimum value of $\{X(s) - \frac{1}{2}\mu_0 s, 0 \leq s \leq t\}$ and stop at the first time the current value of $X(t) - \frac{1}{2}\mu_0 t$ exceeds the minimum record by at least A_C for some constant A_C. In other words,

$$T_C = \inf\left\{t: \sup_{0 \leq s \leq t} \{X(t) - X(s) - \tfrac{1}{2}\mu_0(t - s)\} \geq A_C\right\}.$$

Comparing this procedure to the (A, C, δ) procedure we can see that if we consider only sampling time then the difference is only that in the (A, C, δ) procedure a new minimum is recorded only if it is lower by an amount δC from the previous record. Of course, this difference is negligible as $\delta \to 0$. Hence on the sampling time scale the $(A, C, 0)$ procedures are equivalent to the CUSUM procedures with constant rate sampling.

More formally consider an (A, C, δ) procedure as above. Let

$$i(\tau) = \sup\left\{i: \sum_{j=1}^{i} \tau_j^* \leq \tau\right\}$$

and define a process $Y(\cdot)$ by $Y(0) = 0$ and

$$Y(\tau) = -\delta C i(\tau) + X_{i(\tau)+1}\left(\tau - \sum_{i=1}^{i(\tau)} \tau_i^*\right) - \tfrac{1}{2}\mu_0\left(\tau - \sum_{i=1}^{i(\tau)} \tau_i^*\right), \qquad \tau > 0.$$

The total amount sampled by the (A, C, δ) procedure until the declaration of a change is given by

$$\inf\{\tau > 0: Y(\tau) + \delta C i(\tau) = A\}. \tag{1}$$

Note that if $\nu = \infty$, then $dY(\tau) = -\frac{1}{2}\mu_0\, d\tau + dB(\tau)$ and if $\nu = 0$, then $dY(\tau) = (\mu - \frac{1}{2}\mu_0)\, d\tau + dB(\tau)$, where $B(\cdot)$ is a standard Brownian motion.

Now, a CUSUM procedure with a constant sampling rate of 1 stops at a stopping time which is distributed, both under $\nu = 0$ and $\nu = \infty$ as

$$\inf\Big\{\tau > 0: Y(\tau) - \inf_{s \le \tau} Y(s) = A_C\Big\}. \tag{2}$$

But

$$-\delta C(i(\tau) + 1) \le \inf_{s \le \tau} Y(s) \le -\delta C i(\tau). \tag{3}$$

We obtain from (1)–(3) that the total amount sampled by an (A, C, δ) procedure is stochastically smaller than the CUSUM stopping time with $A_C = A + \delta C$ and stochastically larger than the CUSUM stopping time with $A_C = A$. Letting $\delta \to 0$ we conclude:

PROPOSITION 1. *The total amount sampled by a CUSUM procedure is distributed as the total amount sampled by an $(A, C, 0)$ procedure if $\nu = 0$ or ∞ and both procedures have the same expected time until false alarm.*

The superiority of the dynamic scheme will be apparent when we compare the procedures on the real time scale. The dynamic procedure detects the change much faster on the real time scale; see the next section.

3. Analysis of (A, C, δ) procedures. When analyzing the performance of a procedure, the following quantities are of interest:

$$\begin{aligned} T_{fa} &= E(T|\nu = \infty), \\ T_d &= \sup_\nu E(T - \nu | T \ge \nu), \\ \tau_0 &= E(\tau_1^* | \nu = \infty)/\delta, \\ \tau_d &= E\Big(\sum_{\nu \le i\delta \le T} \tau_i^* | T \ge \nu\Big). \end{aligned}$$

The first quantity, T_{fa}, is the expected time until false alarm, i.e., the expected time until a change is detected while no change in the drift has actually occurred. The second, T_d, is the delay time, the expected time between the change point and its detection. The other two terms describe the expected sample size. Thus τ_0 is the average amount sampled in a unit of a real time before the change point, while τ_d is the expected amount sampled between the change point and its detection. Note that since ν is unknown and typically is much longer than T_d we measure the average rate before the change point and the total sample after it.

For mathematical simplicity we modify the sampling times slightly to be at times $U, U + \delta, U + 2\delta, \ldots$, where U is a uniform random variable on $[0, \delta]$ rather than $U \equiv \delta$ as before. This slight modification results in stationarity with respect to ν and is of course negligible as $\delta \to 0$. In terms of the delay this would result in a minimum average delay of $\frac{1}{2}\delta$ rather than δ as the worst case.

Our main result is the following.

THEOREM 1. *For the* (A, C, δ) *procedure and a drift* μ *on* $[\nu, \infty]$,

$$T_{\text{fa}} = \frac{\delta(e^{\mu_0 A} - e^{-\mu_0 \delta C})}{1 - e^{-\mu_0 \delta C}} - \frac{1}{2}\delta \to \frac{e^{\mu_0 A} - 1}{\mu_0 C} \quad \text{as } \delta \to 0,$$

$$T_{\text{d}} = \frac{\delta(e^{(2\mu-\mu_0)\delta C} - e^{-(2\mu-\mu_0)A})}{e^{(2\mu-\mu_0)\delta C} - 1} - \frac{1}{2}\delta \to \frac{1 - e^{-(2\mu-\mu_0)A}}{(2\mu - \mu_0)C} \quad \text{as } \delta \to 0,$$

$$\tau_0 = \frac{2}{\mu_0 \delta} \frac{\delta C(e^{\mu_0 A} - 1) - A(1 - e^{-\mu_0 \delta C})}{e^{\mu_0 A} - e^{-\mu_0 \delta C}}$$

$$\to \frac{2C}{\mu_0} \frac{e^{\mu_0 A} - 1 - \mu_0 A}{e^{\mu_0 A} - 1} \quad \text{as } \delta \to 0,$$

$$\tau_{\text{d}} = \frac{2}{2\mu - \mu_0} \frac{A(e^{(2\mu-\mu_0)\delta C} - 1) - \delta C(1 - e^{-(2\mu-\mu_0)A})}{e^{(2\mu-\mu_0)\delta C} - 1}$$

$$\to \frac{2}{(2\mu - \mu_0)^2}\{(2\mu - \mu_0)A - 1 + e^{-(2\mu-\mu_0)A}\} \quad \text{as } \delta \to 0.$$

PROOF. The calculations of $E(\tau_1^* | \nu)$ and $p(\nu) \equiv P_\nu(X_1(\tau_1^*) - \frac{1}{2}\mu_0 \tau_1^* = A | \nu)$ are done by standard techniques in diffusion processes [see Siegmund (1985), Theorem 3.6]. We obtain

$$p(\nu) = \frac{1 - e^{-\mu' C\delta}}{e^{\mu' A} - e^{-\mu' C\delta}}$$

and

$$E(\tau_1^* | \nu) = \frac{2}{\mu'} \frac{\delta C(e^{\mu' A} - 1) - A(1 - e^{-\mu' C\delta})}{e^{\mu' A} - e^{-\mu' C\delta}},$$

where $-\frac{1}{2}\mu'$ is the drift of the process $X_1(t) - \frac{1}{2}\mu_0 t$, i.e., $\mu' = \mu_0$ when $\nu = \infty$ and $\mu' = \mu - 2\mu_0$ when $\nu = 0$.

Next note that both under $\nu = 0$ and $\nu = \infty$, T/δ is a convolution of the $U(0, 1)$ random variable with a geometric random variable with parameter $p(\nu)$. Hence $T_{\text{fa}} = \delta/p(\infty) - \frac{1}{2}\delta$ and $T_0 = \delta/p(0) - \frac{1}{2}\delta$. The value of τ_{d} follows immediately. Finally, we obtain from Wald's lemma that $\tau_{\text{d}} = E(\tau_1^* | \nu = 0)/p(0)$. □

4. Comparison with constant rate sampling. In this section we compare the (A, C, δ) procedure to the CUSUM rule with constant rate sampling. We have already shown that on the sampling time scale there is no difference between the two procedures. Of course, while the real time scale and the sampling time scale differ only in their units when the sampling rate is constant, this is not the case with the dynamic scheme. To make the comparison meaningful we would like to compare procedures with the same T_{fa} and τ_0, i.e., procedures with the same performance under $\nu = \infty$. Let us consider the case $\delta = 0$. Proposition 1 implies as a result that both procedures have the same value of τ_{d},

and hence they differ only in their T_d value. The value of T_d(dynamic) is given in the theorem. The value of T_d (constant rate) can be found by multiplying T_d(dynamic) by the scale factor between the real time and sampling time scales. Hence,

$$T_d(\text{constant rate}) = \frac{\tau_d}{\tau_0}$$
$$= \frac{1}{\mu_0 C} \frac{\left(e^{-\mu_0 A} - 1 + \mu_0 A\right)\left(e^{\mu_0 A} - 1\right)}{e^{\mu_0 A} - 1 - \mu_0 A},$$

where τ_d and τ_0 are as given in Theorem 1 and we suppose for simplicity that $\mu = \mu_0$.

We obtain

$$\frac{T_d(\text{dynamic})}{T_d(\text{constant rate})} = \frac{\left(1 - e^{-\mu_0 A}\right)\left(e^{\mu_0 A} - 1 - \mu_0 A\right)}{\left(e^{-\mu_0 A} - 1 + \mu_0 A\right)\left(e^{\mu_0 A} - 1\right)}$$
$$= \frac{e^{-\mu_0 A}\left(e^{\mu_0 A} - 1 - \mu_0 A\right)}{e^{-\mu_0 A} - 1 + \mu_0 A}.$$

In a typical application $\mu_0 A > 1$ and the ratio is of order $(\mu_0 A - 1)^{-1}$; hence, the change point is detected much faster. For example, Pollak and Siegmund (1985) quote $T_{fa} = 793$, $\mu_0 = 1$ and $\tau_0 = 1$ as "seems appropriate for a variety of industrial inspection schemes." Solving $\mu_0 T_{fa} = 793$ we obtain $A \approx 5.977$, so that T_d(dynamic) ≈ 1.96 compared to 10 for the constant rate sampling as given in Pollak and Siegmund (1985).

In Table 1 we compare the performance of six procedures all of which have the same characteristics under $\nu = \infty$. Note that for the CUSUM procedure with

TABLE 1

Comparison of T_d and τ_d for different μ and different sampling procedures (the entries for any μ and δ are T_d over τ_d).

μ	Constant sampling rate[a]	Dynamic sampling $\delta = 0$	Dynamic sampling $\delta = 1$	Dynamic sampling $\delta = 5$	Dynamic sampling $\delta = 10$	Dynamic sampling $\delta = 20$
0.25	129.0	75.3	75.4	79.7	90.7	117.3
	129.0	129.0	129.1	137.3	158.0	207.4
0.5	36.0	11.8	11.8	12.3	13.5	17.1
	36.0	36.0	36.1	37.6	41.5	51.9
1.0	10.0	2.0	2.0	2.9	5.1	10.0
	10.0	10.0	10.0	9.5	8.7	7.4
1.5	5.5	1.0	1.1	2.5	5.0	10.0
	5.5	5.5	5.5	5.0	4.4	3.7
2.0	3.8	0.7	0.8	2.5	5.0	10.0
	3.8	3.8	3.7	3.3	2.9	2.5

[a] Results are quoted from Pollak and Siegmund (1985).

constant rate sampling $T_d = \tau_d$ and for the $(A, C, 0)$ procedure τ_d is equal to the τ_d of the CUSUM procedure. It can be seen that even with $\delta = 20$ (twice the delay of the CUSUM procedure with constant rate sampling) the delay of the dynamic procedure is equal to that of the procedure with constant rate sampling, but the amount sampled under the (A, C, δ) regime is smaller (7.4 compared to 10). With $0 \le \delta \le 20$ the dynamic procedures detect the change faster and "cheaper" under the nominal conditions. Thus, with $\delta = 10$, the (A, C, δ) procedure is superior when μ is 1 or 1.5. When $\mu = 0.25$ or 0.5, the $(A, C, 10)$ detects faster but samples somewhat more and vice versa when $\mu = 2$.

Suppose now that we want to compare the two extreme schemes on the basis of the average sampling rates before the change point is detected when T_{fa} and T_d are kept equal. Then we obtain for the same example ($T_{fa} = 793$ and $T_d = 10$) that τ_0(dynamic) = 0.186 compared to 1 with constant rate sampling.

It is interesting to note that as long as $\mu > \mu_0$, τ_d is less than $\{(2\mu - \mu_0)\mu_0\tau_0/2\}^{-1}$ no matter what T_{fa} is. This is in contrast to constant rate sampling where $T_d \to \infty$ as $T_{fa} \to \infty$ (with τ_0 constant).

A direct comparison of the dynamic procedure to the Roberts–Shiryayev one is more difficult. It is however well known that quantitatively there is little difference between the Roberts–Shiryayev and the CUSUM procedures [Pollak and Siegmund (1985)]. The dynamic procedures suggested here are thus highly superior to both the CUSUM and the Roberts–Shiryayev procedures with constant rate sampling.

REMARK. A direct calculation yields

$$\tfrac{1}{2}\mu_0\big(T_{fa} + \tfrac{1}{2}\delta\big)\tau_0 = C\big(T_{fa} + \tfrac{1}{2}\delta\big) - C\delta - A$$

or

$$A = \big(T_{fa} + \tfrac{1}{2}\delta\big)\big(C - \tfrac{1}{2}\mu_0\tau_0\big) - C\delta.$$

Together with the equation for T_{fa} we obtain

$$C = \tfrac{1}{2}\mu_0\tau_0 + \log\big\{1 + \big(T_{fa} + \tfrac{1}{2}\delta\big)(\exp(\mu_0\delta C) - 1)/\delta\big\}/\big(T_{fa} + \tfrac{1}{2}\delta\big)\mu_0.$$

This equation defines an iterative equation for C which converges extremely fast.

APPENDIX

Indication of optimality. Assaf (1988) investigated a similar problem except that his model incorporated an exponential distribution with parameter ϕ as the prior distribution of the change point ν. In his analysis $\mu = \mu_0$ and is known. In this paper the optimal procedure is analyzed and described. The same type of procedure would be found optimal if we consider only procedures which satisfy that the expected time to false alarm and the average sampling rate before ν are equal to given constants, and the loss function is a linear combination of the delay time T_d and τ_d. The optimal procedure in Assaf (1988) is described as the limit of a sequence of procedures as two parameters, ε and M, are converging to 0 and ∞, respectively. Now, any (A, C, δ) procedure is equal to

the limit of these procedures as $M \to \infty$ for some ϕ and ε. Considering now the limiting case of $\delta \to 0$ and $\varepsilon \to 0$ we obtain that a $(A, C, 0)$ procedure is Bayes with respect to particular prior, constraints on the set of possible procedures and loss function. Since the performance of the $(A, C, 0)$ procedure is independent of the value of ν, it follows that the dynamic procedure is admissible and is a minimax procedure as well.

REFERENCES

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

ASSAF, D. and RITOV, Y. (1988). A double sequential procedure for detecting a change in distribution. *Biometrika* **75** 715–722.

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

PAGE, E. S. (1954). Continuous inspection schemes. *Biometrika* **41** 100–115.

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

ROBERTS, S. W. (1966). A comparison of some control chart procedures. *Technometrics* **8** 411–430.

SHIRYAYEV, A. N. (1963). On optimal methods in earliest detection problems. *Theory Probab. Appl.* **8** 26–51.

SIEGMUND, D. (1985). *Sequential Analysis: Tests and Confidence Intervals.* Springer, New York.

VAN DOBBEN DE BRUYN, C. S. (1968). *Cumulative Sum Test.* Griffin, London.

WILSON, D. W., GRIFFITHS, K., KEMP, K. W., NIX, A. B. J. and ROWLANDS, R. J. (1979). Internal quality control of radioimmunoassays: Monitoring of error. *J. Endocr.* **80** 365–372.

DEPARTMENT OF STATISTICS
FACULTY OF SOCIAL SCIENCES
HEBREW UNIVERSITY
JERUSALEM 91905
ISRAEL

DEPARTMENT OF STATISTICS
THE WHARTON SCHOOL
UNIVERSITY OF PENNSYLVANIA
PHILADELPHIA, PENNSYLVANIA 19104-6302