Bernoulli **6**(4), 2000, 679–697

Sampling from a stationary process and detecting a change in the mean of a stationary distribution

EITAN GREENSHTEIN[1] and YA'ACOV RITOV[2]

[1]*Faculty of Industrial Engineering and Management, Technion – Israel Institute of Technology, Haifa 32000, Israel. Email: eitang@ie.technion.ac.il*
[2]*Department of Statistics, The Hebrew University of Jerusalem, Jerusalem 91905, Israel. Email: yaacov@mscc.huji.ac.il*

Let $\{X_t\}$, $t \in \mathbb{R}$, be a stochastic process. Suppose that the process may not be continuously observed, yet an inference which is related to its probabilistic parameters, or to its sample path, is required. The main purpose of this paper is to study sampling plans. A sampling plan is a method for deciding about time instants $T_1, T_2, \ldots$ at which the process is observed. We study the effect of various sampling plans and sampling rates on the expected time to an alarm in change-point problems (of the mean). Our main effort is studying the asymptotic variance of the sum of the sampled observations until time t. This variance determines asymptotically the expected time to an alarm. As a by-product, we obtain the asymptotic variances of natural estimators for $p = \mathrm{E}(X_t)$ and for $S_t = \int_0^t X_s \, \mathrm{d}s$. Obviously, as the sampling rate is increased, a better estimation is possible. Our study enables us to decide on the 'right' sampling rate. This is analogous to the problem of deciding on the 'right' sample size in the case of independently and identically distributed observations.

Keywords: change point; sampling; sampling rate

1. Introduction

Let $\{X_t\}$, $t \in \mathbb{R}$, be a stochastic process observed only at the discrete points $T_1, T_2, \ldots$. Based on the sampled data, we want to estimate probabilistic parameters of the process such as $p = \mathrm{E}X_t$, the time ν at which there is a change in the mean, or quantities that are related to the sample path such as $S_t = \int_0^t X_s \, \mathrm{d}s$. A decision at time t may be based only on $X_{T_1}, X_{T_2}, \ldots, X_{T_k}$, $T_1 \leqslant T_2 \leqslant \cdots \leqslant T_k \leqslant t$. In this paper we discuss different sampling plans, i.e. methods of choosing the instants $T_1, T_2, \ldots$.

Our motivation comes from problems related to management of a fast network. Suppose that the network has many users, and each user sends messages at random times. One could be interested in the proportion of time accounted for by two users i and j communicating between themselves, or one could be interested in the process defined by the busy and idle periods of the network, etc. Sampling the network activity too often imposes an extra load on the network and may be undesirable.

This leads us to investigate the model where at some instants $T_1, T_2, \ldots$ we obtain

1350–7265

sampled information. The network example motivates us to put a special emphasis on 0–1 processes; for this reason we use the non-standard notation p for $\mathrm{E}X_t$.

There are other examples of processes, not necessarily 0–1, where the observations are dependent; for example, $\{X_t\}$ may be the level of air pollution at time t. When deciding on the frequency of the observation process or the sampling plan, one should obviously take into account the dependence structure among the X_t. Thus, the additional information gained from sampling every second rather than every 2 seconds is clearly negligible.

When the sampled data $X_{T_1}, X_{T_2}, \ldots$ are given, natural estimators for $p = \mathrm{E}X_t$ and for S_t are the sample average $\mathbb{P}_t$ and $\mathbb{S}_t = t\mathbb{P}_t$, respectively; see (2.1). As will be seen in Section 4, the variance of the sampled sum $\mathbb{S}_t$ affects the expected time to detect a change in the mean of the process. In order to obtain minimum variance for those quantities we should space the observations optimally. When the covariance structure of the process is completely known, an optimal spacing is determined. Cases where the covariance structure and the probabilistic mechanism of the process are known and sampling is still required, are when the unknown quantity being estimated is related to the sample path and not to the probabilistic mechanism. An example of an optimal design/spacing under such an approach in terms of minimizing $\mathrm{var}(\mathbb{S}_t)$ is given in Theorem 3.6. If the probabilistic mechanism is not completely known or (as in change-point problems) is expected to change, such an approach is usually impossible or could be highly non-robust and misleading. Moreover, this approach violates the randomization principle in experimental design, and may lead to systematic errors (see Example 2.1). The above considerations lead us to study different types of sampling design, in particular those in which the sampled instants are random. Robustness considerations lead us to search for sampling plans with minimax properties (see Theorem 3.4).

In light of the above examples, if we want to estimate S_t, the accumulated amount of air pollution in the time interval $(0, t)$, we should determine the observational instants taking into account the covariance structure, which may be reasonably assumed to be known from long experience. When we want to estimate S_t, the accumulated busy time of one among many current users of a network, it is unlikely that we will know the covariance structure related to the process that is induced by this user.

The sampling designs that we study should be understood as routine surveillance schemes. Thus the kind of asymptotics that seems relevant is as t approaches infinity.

Our main interest is in considering different sampling methods, and studying the relation between different sampling rates and the accuracy of their associated estimators. The study of the asymptotic variance of the estimators enables us to determine the 'right' sampling rate. This is analogous to the problem of deciding on the 'right' sampling size when sampling independently and identically distributed (i.i.d.) observations. Such a decision is usually based on the variance of the observations and their cost. The analogue in change-point problems is less apparent; it will be better understood and demonstrated in Section 4.

There are three natural definitions for sampling plans, which we will discuss in an increasing order of generalization.

1. *Deterministic sampling plans*. The instants $T_1, T_2, \ldots$ are fixed numbers. An important subclass is when the value of the increments $\{T_{i+1} - T_i\}$ is a constant independent of i.

2. *Random sampling plans*. In this case the increments $\{T_{i+1} - T_i\}$ are random variables independent of the process. An important subclass is renewal sampling plans, that is, where the increments are i.i.d. Another important subclass is of certain stratified sampling plans.
3. *Dynamic sampling plans*. Here the next sampling instant T_{i+1} is a random variable whose distribution is a function of $X_{T_1}, \ldots, X_{T_i}$ and $T_1, \ldots, T_i$.

In this paper we study only deterministic and random sampling plans, and postpone the study of dynamic sampling plans to a future paper. See also Remark 4.2.

Sacks and Ylvisaker (1966) considered a design problem where we observe $\{X_t = \sum_{i=1}^{k}\beta_i f_i(t) + \varepsilon_t,\ t \in (0, 1)\}$, and wish to estimate the coefficients $\beta_1, \ldots, \beta_k$; see Ylvisaker (1987) for references. Cambanis and Masry (1983) investigated a similar problem, testing $X_t = \varepsilon_t$ versus $X_t = f_1(t) + \varepsilon_t,\ t \in (0, 1)$. In both models, the autocorrelation function of the error terms and the functions $f_1, \ldots, f_k$ are considered known. The related design problem for prediction was considered by Schoenfelder and Cambanis (1982) and Ylvisaker (1987). In a typical prediction design problem, points $a < T_1 < T_2 < \cdots < T_n < b$ are selected such that $\int_a^b X_t \, \mathrm{d}t$ can be estimated using the observations $X_{T_1}, \ldots, X_{T_n}$ from the random process $\{X_t\}$. One main difference between these papers and our approach is the type of asymptotics. They considered the case where the number of design points in a given finite interval increases to infinity, whereas we study the asymptotics where the time interval approaches infinity, while the number of design points per unit interval is kept constant. A sample design similar to our stratified sampling, but for a different model and different objectives, was investigated by Roll and Yadin (1986). Their motivation, however, is relevant to our study; they considered a typical problem of industrial engineering where a single 'inspector' studies the long-run behaviour of many production inputs. Thus, instead of a continuous perfect inspection of one production input, the inspector takes repeated snapshots at random times from many production inputs.

Sampling from a stochastic process is related to Cochran's notion of sampling from a superpopulation. For surveys about sampling techniques in such a context, see Bellhouse (1988) and Murthy and Rao (1988). Finally, Brillinger (1973) studied asymptotic aspects of sampling schemes in a manner similar to ours. He studied the asymptotic variance of $\mathbb{S}_t$ for general sampling plans that may be represented as stationary point processes.

In Section 2, we will consider a few examples. We will elaborate on the case where the process $\{X_t\}$ is a homogeneous $0-1$ Markovian process. Deterministic and random sampling schemes will be examined according to the variances of their associated estimators.

In Section 3, we will study properties of sampling plans. We will investigate plans with i.i.d. increments and, in particular, exponential sampling plans, i.e., plans where $\{T_{i+1} - T_i\}$ are exponentially distributed with parameter λ, independent of each other and of the process $\{X_t\}$. The exponential distribution is an approximation to a situation in which there are many channels, and at each sampling instance one of the channels is selected at random and sampled. Another method that we will study is stratification, i.e. random sampling plans that divide the time interval into subintervals of equal length, and sample one unit from each subinterval. We also elaborate on the case where $\{X_t\}$ is a Markov process which takes $0-1$ values and the sampling plan is deterministic (see also Example 2.3).

In Section 4 we will consider the change-point problem for detecting a change in $p_t = \mathrm{E}X_t$. Our asymptotic study that is based on a Brownian motion approximation of a stationary process. Some references concerning change-point problems are Lorden (1971), Shiryayev (1983), Siegmund (1985), Pollak (1987), Basseville and Nikiforov (1993) and Brodsky and Darkhovsky (1993). The last two references are especially relevant to our work since they deal with processes with dependent observations. Chapter 7 of Basseville and Nikiforov considers change-point problems in parametric models of dependent observations (like ARMA), rather than our nonparametric approach. In Chapter 4 of Brodsky and Darkhovsky, there is a study of nonparametric cusum, which might be complementary to the brief study that we do in Section 4. Neither of the above references on change-point problems considers sampling plans; rather, they study sampling every observation (in discrete time).

2. Examples

In the following, we will consider a few examples, which will help us to illustrate the advantages and disadvantages of deterministic and random sampling plans.

Example 2.1. Consider the following process. Let $a \in [0, 1)$ be a constant, S be a Bernoulli random variable, and let τ be distributed uniformly on $[0, 1)$. Let $X_t = S$ for $t \in \cup_{n=0}^{\infty}[2n + \tau, 2n + a + \tau)$, and $1 - S$ elsewhere. Hence the process is deterministic given X_τ.

To avoid trivialities, one may define a non-deterministic version by letting the process alternate at instant i only with a very high probability. However, it will be convenient to think of the process as it was initially defined. Notice that for $a = 1$ the process is stationary, but otherwise the process is not even weakly stationary.

Example 2.2. Consider the following Markov process. The process alternates between intervals of 0 and 1, where the length of each '0 interval' is an exponential random variable with mean λ_0^{-1} and the length of each '1 interval' is an exponential random variable with mean λ_1^{-1}. Let $X_0 = 1$ with probability $\lambda_0/(\lambda_1 + \lambda_0)$.

We will now examine two deterministic sampling plans in the case of Example 2.1 (with $a = 1$). The first one samples every unit of time and the second every two units of time. The natural estimator for p, based on the sample average, performs excellently in the first case and very poorly in the second case. The weakness demonstrated in the second case is due to the violation of the randomization principle in experimental design or in sampling. The way to prevent such phenomena is to use random sampling plans.

Definition 2.1. *Let* $\mathbf{T}$ *be a sampling plan. Let* $N_t = \max\{i | T_i < t\}$ *be the number of observations until time t. Suppose* $N_t/t \xrightarrow{P} \mu^{-1}$. *We then say that* $\mathbf{T}$ *has a sampling rate* μ^{-1}.

Given a sampling plan, we will study the behaviour of

$$\mathbb{S}_t(X;\,\mathbf{T}) = X_{T_1} + \ldots + X_{T_{N_t}},$$

$$\mathbb{P}_t(X;\,\mathbf{T}) = \mathbb{S}_t(X;\,\mathbf{T})/N_t, \tag{2.1}$$

where, for the sake of definiteness, $0/0 = 0$. Those are the natural estimators for S_t and for $p = \mathrm{E}X_t$.

We will compare sampling plans with the same sampling rate by comparing the asymptotic variances of $\mathbb{P}_t$ and of $\mathbb{S}_t$.

The main purpose of this study is to determine the 'right' sampling rate. Thus, we should analyse the effect of increasing the sampling rate. Here is an example.

Example 2.3. Consider a Markovian process as described in Example 2.2. Consider the class of deterministic sampling plans, where the increments $T_{i+1} - T_i$ are of a fixed size μ. Hence the sampling rate is μ^{-1}. Let $p_{ij}^s = P(X_{t+s} = j|X_t = i)$. It may be shown (see Karlin and Taylor 1975, Exercise 7, p. 154) that

$$p_{ii}^s = \frac{\lambda_{1-i}}{\lambda_0 + \lambda_1} + \frac{\lambda_i}{\lambda_0 + \lambda_1}\exp[-(\lambda_0 + \lambda_1)s], \qquad i = 0,\,1. \tag{2.2}$$

Denote by $(p_0,\,p_1)$ the stationary distribution $p_0 = \lambda_1/(\lambda_1 + \lambda_0)$, $p_1 = 1 - p_0$. Then it may be verified (see Good 1961) for the induced two-state Markov chain with transition probabilities p_{ij}^μ that

$$\begin{aligned}\operatorname{var}(\mathbb{S}_t) &= \frac{t}{\mu}p_0 p_1 \frac{1 + p_{11}^\mu - p_{01}^\mu}{p_{10}^\mu + p_{01}^\mu} + o(t)\\ &= \frac{t}{\mu}p_0 p_1 \frac{1 + \exp(-(\lambda_0 + \lambda_1)\mu)}{1 - \exp(-(\lambda_0 + \lambda_1)\mu)} + o(t).\end{aligned} \tag{2.3}$$

(Henceforth, all $o(t)$ and $O(t)$ terms are as $t \to \infty$.) Thus we have obtained a formula relating the asymptotic variance of $\mathbb{S}_t$ to the rate of the sampling plan.

3. Properties of sampling plans

Let $\{X_t\}$ be a stochastic process. Define

$$S_t \equiv S_t(X) = \int_0^t X_s \,\mathrm{d}s.$$

In the rest of this paper we will consider processes $\{X_t\}$ such that the above integral exists with probability 1, and the variance of their associated process $\{S_t\}$ satisfies, for some constant B,

$$\operatorname{var}(S_t) = Bt + o(t). \tag{3.1}$$

If the process is weakly stationary, i.e. $R(s) \equiv \mathrm{cov}(X_t, X_{t+s})$ is independent of t (as in the case of stationary processes), (3.1) is equivalent to

$$\int_0^\infty R(s)\,\mathrm{d}s = A < \infty. \tag{3.2}$$

The variance of the process then satisfies $\mathrm{var}(S_t) = 2At + o(t)$.

As will be seen in Section 4, the asymptotic variance of $\mathbb{S}_t$ determines the asymptotic expected time to an alarm in change-point problems. Thus minimizing $\mathrm{var}(\mathbb{S}_t)$ and determining its magnitude for various sampling plans is our main goal. As a by-product, under further conditions, we also obtain the asymptotic variance of $\mathbb{P}_t$ through the relation in the following theorem. We also obtain $\mathrm{E}\,\mathrm{var}(t\mathbb{P}_t|S_t)$ asymptotically.

Here is a general result:

Theorem 3.1. *Suppose* $\mathrm{E}N_t/t \to \mu^{-1}$, $\mathrm{var}(S_t) = O(t)$ *and, for some* $\alpha \in (0, 1)$, $P(|N_t - t/\mu| > t^\alpha) = o(t^{-1})$. *Then*

$$\mathrm{var}(\mathbb{P}_t) = (\mu/t)^2(\mathrm{var}(\mathbb{S}_t) - p^2\,\mathrm{var}(N_t)) + o(t^{-1}).$$

Proof. Suppose, without loss of generality, that $\mu = 1$. Then

$$\mathrm{var}(\mathbb{P}_t) = \mathrm{E}\{(\mathbb{P}_t - p)^2 \mathbb{1}_{|N_t - t| < \xi_t}\} + \mathrm{E}\{(\mathbb{P}_t - p)^2 \mathbb{1}_{|N_t - t| > \xi_t}\},$$

where $\xi_t/t \to 0$ and $\xi_t/t^\alpha \to \infty$. The second term on the right is $o(t^{-1})$ by assumption. Thus

$$\frac{\mathrm{E}(\mathbb{S}_t - N_t p)^2}{(t - \xi_t)^2} \geqslant \mathrm{E}\{(\mathbb{P}_t - p)^2 \mathbb{1}_{|N_t - t| < \xi_t}\} \geqslant \frac{\mathrm{E}(\mathbb{S}_t - N_t p)^2}{(t + \xi_t)^2} + o(t^{-1})$$

but

$$\mathrm{var}(\mathbb{S}_t) = \mathrm{var}(\mathbb{S}_t - N_t p + N_t p) = \mathrm{var}(\mathbb{S}_t - N_t p) + p^2\,\mathrm{var}(N_t).$$

□

Remark 3.1. When estimating S_t by $t\mathbb{P}_t$, the relevant quantity to study is $\mathrm{var}(t\mathbb{P}_t|S_t)$ or, when averaging, $\mathrm{E}\,\mathrm{var}(t\mathbb{P}_t|S_t)$. Notice that

$$\mathrm{var}(t\mathbb{P}_t) = \mathrm{E}\,\mathrm{var}(t\mathbb{P}_t|S_t) + \mathrm{var}(\mathrm{E}(t\mathbb{P}_t|S_t)). \tag{3.3}$$

Write $\mathrm{E}(t\mathbb{P}_t|S_t) = S_t + R_t$. In cases where

$$\mathrm{var}(S_t + R_t) = \mathrm{var}(S_t) + o(t) = Bt + o(t), \tag{3.4}$$

we may derive the asymptotic value of $\mathrm{E}\,\mathrm{var}(t\mathbb{P}_t|S_t)$ through (3.3) and Theorem 3.1.

The results in this section are proved under two possible sets of conditions. For arbitrary valued processes $\{X_t\}$ we will assume weak stationarity. When restricting ourselves to 0–1 processes $\{X_t\}$, some of the results may be proved under weaker conditions, i.e. for a larger set of 0–1 processes denoted $\mathscr{X}_{B,p}$. This is the set of all 0–1 processes satisfying $\mathrm{E}X_t = p$ and $\mathrm{var}(S_t) = Bt + o(t)$. In Example 2.1 for $a \neq 1$ the process is not weakly stationary but it belongs to $\mathscr{X}_{B,p}$ for a proper B.

3.1. Renewal sampling plans

In the following we will study properties of renewal sampling plans, i.e. random sampling plans with i.i.d. increments. We will first give some definitions related to renewal theory. Suppose $\{T_i - T_{i-1}\}$ are distributed with distribution function F. Let $F^i(t) = P(T_i < t)$. Define the renewal function $U(t) = \sum_{i=1}^{\infty} F^i(t) = \mathrm{E}N_t$. For some background on renewal theory, see Siegmund (1985).

Theorem 3.2. *Let $\{X_t\}$ be a process such that $R(s) \equiv \mathrm{cov}(X_{t+s}, X_t)$ is independent of t. Denote $\mathrm{E}X_t = p$. Let $\mathbf{T}$ be a random sampling method with i.i.d. increments, rate μ^{-1} and renewal function U. Denote $\sigma^2 = \mathrm{var}(T_1)$. Then*

$$\mathrm{var}(\mathbb{S}_t(X;\,\mathbf{T})) = \mu^{-1} t\,\mathrm{var}(X_0) + 2\mu^{-1} t \int_0^{\infty} R(s)\,\mathrm{d}U(s) + p^2\,\mathrm{var}(N_t) + o(t)$$

$$= \mu^{-1} t\,\mathrm{var}(X_0) + 2\mu^{-1} t \int_0^{\infty} R(s)\,\mathrm{d}U(s) + p^2 \sigma^2 t \mu^{-3} + o(t).$$

Proof. The proof follows by applying Theorem 5 in Brillinger (1973). See also a direct (but similar) proof in Greenshtein and Ritov (1997). □

When the random variables $(T_{i+1} - T_i)$ have a moment generating function a large-deviation argument shows that the conditions of Theorem 3.1 are satisfied. Thus the relation $\mathrm{var}(N_t) = t\sigma^2\mu^{-3} + o(t)$ and Theorems 3.1 and 3.2 yield the asymptotic variance of $\mathbb{P}_t$.

Among renewal sampling plans, the exponential sampling plan (i.e. $(T_{i+1} - T_i)$, $i = 1, 2, \ldots,$ are independent exponential random variables) is especially appealing. In the case of exponential sampling plans, $U(t) \equiv t$, and the asymptotic variance of $\mathbb{S}_t(X;\,\mathbf{T})$ depends on the covariance structure of the process $\{X_t\}$ only through $B = \mathrm{var}(S_t) = 2\int_0^{\infty} R(s)\,\mathrm{d}s$. Exponential sampling plans may serve as an approximation to many other plans in cases where the sampling action succeeds only with small probability. An explicit expression for the asymptotic variance of $\mathbb{S}_t$ is given in the following.

Corollary 3.3. *Let $\{X_t\}$, $t \in \mathbb{R}$, be a weakly stationary process. Let $\mathbf{T}$ be an exponential sampling plan. Then, for every real number k,*

$$\mathrm{var}(\mathbb{S}_t - kN_t) = \lambda t\,\mathrm{var}(X_0) + \lambda^2 B t + (p-k)^2 \lambda t + o(t).$$

Proof. The claim follows directly by applying Theorem 3.1 to the process $\tilde{\mathbb{S}}_t = \mathbb{S}_t - kN_t$, which is induced by the process $\tilde{X}_t = (X_t - k)$, upon realizing that for an exponential sampling plan $\mathrm{d}U(s) = \lambda\,\mathrm{d}s$. □

The above result holds for a 0–1 process in $\mathscr{X}_{B,p}$, even if it is not weakly stationary; see Greenshtein and Ritov (1997).

In the case of exponential sampling plans the value of R_t in (3.4) is identically zero, thus

(3.4) is trivially satisfied. Hence the asymptotic variance of $t\mathbb{P}_t$ conditional on S_t may be evaluated. We believe that (3.4) is satisfied under more general renewal sampling plans.

3.2. Stratified sampling plans

Another random sampling plan that we will study is the stratified sampling plan. Consider the real line as a union of intervals or strata of length μ. A stratified plan with rate μ^{-1} is the sampling plan that randomly samples a point from every stratum according to a uniform distribution. The stratified plan has the following appealing minimax property.

Theorem 3.4. *Let* $\mathbf{T}$ *be a stratified sampling plan with rate* μ^{-1}, *and let* $\tilde{\mathbf{T}}$ *be any random sampling plan with rate* μ^{-1}. *Then*

$$\lim_{t\to\infty} \frac{\operatorname{var}(\mathbb{S}_t(X;\, T))/t}{p(1-p)/\mu + B/\mu^2} \leqslant 1, \qquad X \in \mathscr{X}_{B,p},$$

$$\sup_{X\in\mathscr{X}_{B,p}} \limsup_{t\to\infty} \frac{\operatorname{var}(\mathbb{S}_t(X;\, \tilde{\mathbf{T}}))/t}{p(1-p)/\mu + B/\mu^2} \geqslant 1.$$

Hence, the stratified sampling plan is asymptotically minimax.

Proof. Let $A_t = \{s\colon X_s = 1,\, s < t\}$ be the random set of indices s for which $X_s = 1$. Let $L(A_t)$ be the Lebesgue measure of A_t. Then

$$\begin{aligned}
\operatorname{var}(\mathbb{S}_t(X;\, \mathbf{T})) &= \mathrm{E}(\operatorname{var}(\mathbb{S}_t|A_t)) + \operatorname{var}(\mathrm{E}(\mathbb{S}_t|A_t)) \\
&= \mathrm{E}(\operatorname{var}(\mathbb{S}_t|A_t)) + \operatorname{var}(S_t/\mu) \\
&= \mathrm{E}(\operatorname{var}(\mathbb{S}_t|A_t))I(L(A_t)/t \in (p - o(1),\, p + o(1))) + Bt/\mu^2 + o(t) \\
&\leqslant p(1-p)\frac{t}{\mu} + Bt/\mu^2 + o(t).
\end{aligned}$$

The last inequality is obtained since

$$\begin{aligned}
\frac{1}{[t/\mu]} \operatorname{var}(\mathbb{S}_t|A_t) &\leqslant \frac{1}{[t/\mu]} \sum_{i=0}^{[t/\mu]-1} \frac{L(A_{it})}{\mu}\left(1 - \frac{l(A_{it})}{\mu}\right) \\
&\leqslant \frac{L(A_t')}{\mu[t/\mu]}\left(1 - \frac{L(A_t')}{\mu[t/\mu]}\right)
\end{aligned}$$

by the Cauchy–Schwarz inequality; here $A_{it} \equiv A_t \cap [i\mu,\, (i+1)\mu)$, $A_t' = A_t \cap [0,\, \mu[t/\mu])$ and $[x]$ is the largest integer smaller than x.

We now prove the second claim by constructing a sequence of processes $\{X^n\} \in \mathscr{X}_{B,p}$ such that

$$\liminf_{n\to\infty} \limsup_{t\to\infty} \frac{\operatorname{var}(\mathbb{S}_t(X^n;\, \tilde{\mathbf{T}}))/t}{p(1-p)/\mu + B/\mu^2} \geqslant 1$$

for any sampling plan with rate μ. The idea is simple. Different sampling plans with the same sampling rate affect $\operatorname{var}(\mathbb{S}_t)$ only through the term $\mathrm{E}(\operatorname{var}(\mathbb{S}_t|A_t))$. If the process $\{X_t\}$ consists of i.i.d. Bernoulli(p) variables, then any sampling plan with rate μ^{-1} would have $\operatorname{var}(\mathbb{S}_t|A_t) = p(1-p)t/\mu + o(t)$. However, for such a process S_t is not defined, since the sample path is not measurable. We will construct a stationary process that is close to i.i.d. Bernoulli(p), but still have $\operatorname{var}(S_t) = Bt + o(t)$.

Formally, we need the following construction. Let W_1 and $W_{i+1} - W_i$, i $= 1, 2, \ldots,$ be i.i.d. exponential with mean β_n^{-1}. Let N_1, $N_{i+1} - N_i$, $i = 1, 2, \ldots,$ be i.i.d. geometric random variables. Note that $W_{N_{i+1}} - W_{N_i}$ are i.i.d. exponential, with mean α_n^{-1}. Let π_i^n, $i = \ldots, -1, 0, 1, \ldots,$ be i.i.d. random variables in $(0, 1)$, with mean p and variance σ_n^2. The process X^n is constant in the intervals $(W_i, W_{i+1}]$, its value in a given interval is independent of its value in any other interval, and is equal to 1 with probability π_j^n where $N_j \leqslant i < N_{j+1}$.

The parameters α_n, β_n and σ_n^2 are selected so that $X^n \in \mathscr{X}_{B,p}$. Note that X_0^n and X_t^n are in the same 'small interval' with probability $\mathrm{e}^{-\beta_n t}$ and not in the same 'large interval' with probability $1 - \mathrm{e}^{-\alpha_n t}$. Hence the autocorrelation is equal to

$$\begin{aligned}
R_n(t) &= \mathrm{E}(X(0)X(t)) - p^2 \\
&= \mathrm{E}(\pi_0^n \mathrm{e}^{-\beta_n t} + (\pi_0^n)^2(\mathrm{e}^{-\alpha_n t} - \mathrm{e}^{-\beta_n t}) + \pi_0^n p(1 - \mathrm{e}^{-\alpha_n t})) - p^2 \\
&= (p - p^2 - \sigma_n^2)\mathrm{e}^{-\beta_n t} + \sigma_n^2 \mathrm{e}^{-\alpha_n t}.
\end{aligned}$$

The parameters therefore satisfy

$$B/2 = \int_0^\infty R_n(t)\,\mathrm{d}t = (p - p^2 - \sigma_n^2)/\beta_n + \sigma_n^2/\alpha_n.$$

We consider the limit where $\beta_n \to \infty$, $\alpha_n \to 0$ and $\sigma_n^2/\alpha_n \to B/2$.

Now

$$\operatorname{var}(\mathbb{S}_t(X^n;\, \tilde{\mathbf{T}})) \geqslant \mathrm{E}\operatorname{var}(\mathbb{S}_t(X^n;\, \tilde{\mathbf{T}})|\tilde{\mathbf{T}}),$$

but

$$\operatorname{var}(\mathbb{S}_t(X^n;\, \tilde{\mathbf{T}})|\tilde{\mathbf{T}} = \mathbf{t}) = \sum_i \sum_j R_n(t_j - t_i),$$

where $\mathbf{t} = (t_1, t_2, \ldots, t_m)$. Let $\varepsilon_n \to 0$, $\varepsilon_n/\alpha_n \to \infty$ be an arbitrary sequence of constants. Then

$$\sum_{j=1}^{\infty} R_n(t_j - t_i) = R_n(0) + 2\sum_{j=i+1}^{\infty} R_n(t_j - t_i)$$

$$\geqslant p(1-p) + 2\sigma_n^2 \sum_{j=i+1}^{\infty} \mathrm{e}^{-\alpha_n(t_j - t_i)}$$

$$\geqslant p(1-p) + 2\sigma_n^2 \sum_{m=0}^{\infty} \mathrm{e}^{-m\varepsilon_n/\alpha_n} \sum_{t_j - t_i \in \left(\frac{m\varepsilon_n}{\alpha_n}, \frac{(m+1)\varepsilon_n}{\alpha_n}\right)} \mathrm{e}^{-\alpha_n(t_j - t_i) + m\varepsilon_n}$$

$$\geqslant p(1-p) + 2(1-\varepsilon_n)\frac{\sigma_n^2}{\alpha_n \mu} \sum_{m=0}^{\infty} \mathrm{e}^{-m\varepsilon_{n}\varepsilon_n}$$

$$\to p(1-p) + B/\mu,$$

for i such that t_i and $t - t_i$ are large.

The second claim follows since if $t_{n_t} \leqslant t < t_{n_t+1}$ then

$$\sum_{i=1}^{n_t}\sum_{j=1}^{n_t} R_n(t_j - t_i) = (1 + o(1))\sum_{i=1}^{n_t}\sum_{j=-\infty}^{\infty} R_n(t_j - t_i)$$

and $n_t/t \to \mu^{-1}$. □

Proposition 3.5. *Let $\{X_t\}$ be a weakly stationary process. Let* $\mathbf{T}$ *be a stratified plan with rate μ^{-1}. Then*

$$\operatorname{var}(\mathbb{S}_t(X;\, \mathbf{T})) = \frac{t}{\mu}\operatorname{var}(X_0) + B_\mu t + o(t),$$

where $B_\mu = 2\mu^{-3}\int_0^\infty (t \wedge \mu) R(t)\,\mathrm{d}t$ *and* $R(s) = \operatorname{cov}(X_t, X_{t+s})$.

Proof. The proof follows from the following calculation:

$$\sum_{j=1}^{\infty} \operatorname{cov}(X_{T_i}, X_{T_{i+j}}) = \mu^{-2}\sum_{j=1}^{\infty}\int_0^\mu \int_{j\mu}^{(j+1)\mu} R(t-s)\,\mathrm{d}t\,\mathrm{d}s$$

$$= \mu^{-2}\int_0^\mu \int_\mu^\infty R(t-s)\,\mathrm{d}t\,\mathrm{d}s.$$

$$= \mu^{-2}\int (t \wedge \mu) R(t)\,\mathrm{d}t.$$

□

Remark 3.2. It was pointed out by a referee that for the class of weakly stationary 0–1 processes with $\mathrm{E}X_t = p$ and $\operatorname{var}(X_t) = Bt + \mathrm{o}(t)$, the first part of Theorem 3.4 follows immediately from Proposition 3.5. This is since $\mathrm{var}(X_0) = p(1-p)$ and $B_\mu \leqslant (2/\mu^2)B$. Notice, however, that our proof of the first claim in Theorem 3.4 was for the wider class, $\mathscr{X}_{B,p}$.

The conditions of Theorem 3.1 are trivially satisfied for stratified sampling plans, hence Theorem 3.1 may be used to evaluate the asymptotic variance of $\mathbb{P}_t$. The value of R_t in (3.4) is identically zero, thus (3.4) is trivially satisfied and the asymptotic variance of $t\mathbb{P}_t$ conditional on S_t may be evaluated.

3.3. Optimal deterministic plans for 0–1 Markov processes

For the purpose of change-point problems, we want to design sampling plans such that $\mathrm{var}(\mathbb{S}_t(X; \mathbf{T}))$ is asymptotically minimized. When a specified covariance structure $R(s) = \mathrm{cov}(X_t, X_{t+s})$ is given, one may (try to) find a sequence of points $t_1^0, t_2^0, \ldots$ satisfying $t_i/\mu = i + o(i)$, such that $\mathrm{var}(X_{T_1} \ldots, X_{T_{N_t}})$ is asymptotically minimized among all sequences $\{t_i\}$ that satisfy $t_i/\mu = i + o(i)$. Now consider the deterministic sampling plan satisfying $T_i = t_i^0$. Such a fixed plan is good as long as the covariance structure of the process does not change much; otherwise it may be very inefficient. In any case, it has the weakness of an experimental design that violates the randomization principle. One approach could be to choose a sampling plan that is a mixture of the deterministic sampling plan described and some random sampling plan. Still there are natural cases where a certain deterministic plan is the most efficient for a class of possible processes; hence for this class the plan is robust. This subsection shows that in a situation where it is known that the process $\{X_t\}$ is a homogeneous 0–1 Markov process, the best among random sampling plans with rate μ^{-1}, in terms of $\mathrm{var}(X_{T_1} \ldots, X_{T_{N_t}})$, is the deterministic plan with equal increments. This is implied by the following Theorem 3.6. For related results see Blight (1973) and Theorem 4.1 in Bellhouse (1988) which is attributed to J. Hájek.

Theorem 3.6. *Let $\{X_t\}$ be a stationary 0–1 Markov process. Let $\mathbf{T}$ be a deterministic sampling plan with equal increments of size μ. Let $\tilde{\mathbf{T}}$ be any random sampling plan satisfying $N_t/t \xrightarrow{P} \mu^{-1}$. Then*

$$\mathrm{var}(\mathbb{S}_t(X; \mathbf{T})) \leqslant \mathrm{var}(\mathbb{S}_t(X; \tilde{\mathbf{T}})) + o(t).$$

To prove this theorem we require the following lemmas.

Lemma 3.7. *Let R be a function on the real line. Define*

$$h_n(T_1, \ldots, T_n) = \sum_{i,j} R(T_i - T_j).$$

Suppose R is symmetric around zero and convex on $[0, \infty)$. Then h_n is convex on the convex domain consisting of all the vectors $(T_1, \ldots, T_n)$ satisfying $0 \leqslant T_1 \leqslant T_2 \leqslant \cdots \leqslant T_n \leqslant \infty$.

The proof is omitted.

Lemma 3.8. *Let h_n and R be functions as defined in Lemma 3.7, where R is the convex function $R(s) = a\exp(-b|s|)$. Suppose the minimum of h_n, over $0 \leqslant T_1 \leqslant \cdots \leqslant T_n \leqslant n-1$, is attained at the point $(T_1^0, \ldots, T_n^0)$. Then*

$$h_n(T_1^0, \ldots, T_n^0) = h(0, 1, 2, \ldots, n-1) + o(n).$$

Proof. Without loss of generality, let $a = b = 1$. We first calculate the partial derivative of h_n at a coordinate i at the point $(T_1 = 0, \ldots, T_n = n-1)$:

$$\frac{\partial h_n}{\partial T_i} = -\sum_{j<i} \exp(-i+j) + \sum_{j>i} \exp(-j+i).$$

By symmetry, terms are cancelled and we obtain, for $k\log(n) < i < n - k\log(n)$, that $|\partial h_n/\partial T_i| = O(\exp - [k\log(n)])$. Otherwise the partial derivative is of order $O(1)$. Our next step is to show that, for $i < k\log(n)$, $|T_i^0 - (i-1)| < \sqrt{n}$ (say). This follows since otherwise we could find $j, j+1 < k\log(n)$, such that $|T_j^0 - T_{j+1}^0| > \sqrt{n}/k\log(n)$. This leads to a contradiction of the fact that $(T_1^0, \ldots, T_n^0)$ is the point where h_n attains its minimum, and may be seen as follows. There are indices l, $l+1$ such that $|T_l^0 - T_{l+1}^0| < 1$. Thus the vector $(T_1^1 \ldots, T_n^1)$ satisfying

$$T_1^1 = T_1^0, \ldots, T_j^1 = T_j^0, T_{j+1}^1 = (T_j^0 + T_{j+1}^0)/2, T_{j+2}^1 = T_{j+1}^0, \ldots, T_{l+1}^1 = T_l^0,$$

$$T_{l+2}^1 = T_{l+2}^0, \ldots, T_n^1 = T_n^0$$

has a smaller value of h_n. Similarly, we show for $i > n - k\log(n)$ that $|T_i{}^0 - (i-1)| < \sqrt{n}$. Obviously $|T_i{}^0 - (i-1)| < n$ for every i. Now, by the Lagrange theorem,

$$h_n(T_1^0, \ldots, T_n^0) - h(0, \ldots, n-1) = \sum_{i=1}^{n} \frac{\partial h_n}{\partial T_i}(\theta_i)(T_i^0 - (i-1))$$

for points $\theta_i \in (T_i^0, i-1)$. By the convexity of h_n and by the above discussion, the last quantity is of an order of magnitude $n^2 \exp - (k\log(n)) + 2k\log(n)\sqrt{n} = o(n)$ for $k > 2$. $\square$

Proof of Theorem 3.6. Without loss of generality, let $\mu = 1$. Now

$$\mathrm{var}(\mathbb{S}_t(X; \tilde{\mathbf{T}})) = \mathrm{E} \sum_{\tilde{T}_i, \tilde{T}_j < t} R(\tilde{T}_i - \tilde{T}_j).$$

Here $R(\tilde{T}_i, \tilde{T}_j) = \mathrm{cov}(X_{\tilde{T}_i}, X_{\tilde{T}_j})$. It may be deduced from (2.2) in Example 2.3 that $R(s) = a\exp(-bs)$, where $a = (\lambda_0/(\lambda_0 + \lambda_1))^2$ and $b = \lambda_0 + \lambda_1$. Now

$$\mathrm{E}\sum_{\tilde{T}_i,\tilde{T}_j<t} R(\tilde{T}_i-\tilde{T}_j) = \mathrm{EE}\left(\sum_{\tilde{T}_i,\tilde{T}_j<t} R(\tilde{T}_i-\tilde{T}_j)|N_t\right)$$

$$\geqslant \mathrm{E}(N_t p_0 p_1)\left(\frac{1+p_{11}^{t/N_t}-p_{01}^{t/N_t}}{p_{10}^{t/N_t}+p_{01}^{t/N_t}}\right)+o(t) \tag{3.5}$$

$$= t(p_0 p_1)\left(\frac{1+p_{11}^{1}-p_{01}^{1}}{p_{01}^{1}+p_{01}^{1}}\right)+o(t) \tag{3.6}$$

$$= \mathrm{var}(\mathbb{S}_t(X;\mathbf{T}))+o(t). \tag{3.7}$$

Inequality (3.5) follows from (2.3) in Example 2.3 and from Lemma 3.8. Equality (3.7) follows since $N_t/t \xrightarrow{p} 1$. Equality (3.7) follows, again, from (2.3). □

4. Sampling rate and expected time to alarm

In this section we will formulate and study a change-point problem related to our processes $\{X_t\}$. Suppose that at an unknown time ν, there is a change in the stochastic mechanism of the process, resulting in a change in the mean of its stationary distribution; in the case of a 0–1 process, this means that $\lim_{t\to\infty} P(X_t = 1) \neq p$. In such a case we want to declare an alarm. Let $\mathbf{T}$ be a sampling plan. An alarm is a stopping rule τ, which at time t is a function of $X_{T_1}, \ldots, X_{T_{N_t}}$ and $T_1, \ldots, T_{N_t}$. We are interested in the expectation of $(\tau-\nu)^+$ under a change and the expectation of τ when there is no change.

When doing routine sampling (or surveillance) in order to detect a change in the stationary mean, obviously as the sampling rate becomes higher, more efficient stopping rules may be designed. Here, more efficient rules should be understood as rules with smaller expected time to an alarm and larger expected time to a false alarm. Still, as discussed in the Introduction, high rates of sampling might be undesirable. Thus, a desired rate is a compromise. In order to determine it we should study the relation between the sampling rate and the expected time to alarm or false alarm.

Another relevant issue is the type of sampling plan to be chosen – stratified, exponential, deterministic, etc. Different considerations might apply as discussed in the Introduction. When the covariance structure of the controlled process is known and no 'dramatic' change in it is expected after the change time ν, we might design a deterministic sampling plan. However, we might want to use types of randomized sampling plans for robustness reasons and to avoid systematic errors. The issue of the type of sampling plan will not be discussed further.

The main purpose of this section is to develop tools to evaluate and to decide on the 'right' sampling rate for a given type of plan.

We will study the asymptotic behaviour of sampling plans and its relation to their sampling rate. In Section 4.1, the development is for general random sampling plans. In

Sections 4.2 and 4.3, we elaborate on two cases: where **T** is an exponential sampling plan; and where **T** is a stratified sampling plan.

4.1. General random sampling plans

We will study the cusum type of stopping rule, denoted $\tau = \tau(k, b)$, and defined as follows. Let $\tilde{\tau}_0 = 0$ and define, for $i \geqslant 1$, the stopping rules $\tau_i = \tau_i(k, b)$ by

$$\tilde{\tau}_i = \inf\{t\colon t > \tilde{\tau}_{i-1}, \exists s \in (\tilde{\tau}_{i-1}, t), \mathbb{S}_t - \mathbb{S}_s - k(N_t - N_s) \notin (0, b)\}.$$

Now define the stopping rule $\tau(k, b)$ to be

$$\tau = \sum_{i=1}^{N} \tilde{\tau}_i. \tag{4.1}$$

Here N is the first index such that the stopping time of $\tilde{\tau}_N$ occurs when $\max_{t_{N-1} \leqslant s \leqslant t} \mathbb{S}_t - \mathbb{S}_s - k(N_t - N_s) \geqslant b$.

The choice of k and b in the above depends on the particular problem and constraints; see the numerical examples below.

In the following, we give a formula relating the sampling rate and the expected time to an alarm. We will approximate the process $\mathbb{S}_t - kN_t$, defined by a stationary process $\{X_t\}$ with a stationary mean p and a given sampling plan with rate μ^{-1}, by a Brownian motion with drift coefficient $\theta = (p-k)\mu^{-1}$ and variance coefficient σ_k^2. To be more precise, we assume that the processes $\mathbb{S}_{sT} - kN_{sT} - (p-k)sT/(\mu\sigma_k\sqrt{t})$, $s \in (0, a_T)$, converge weakly to a Brownian process, where X_t is weakly stationary and $a_T \to \infty$. General conditions under which such a functional central limit theorem is valid may be found in Ethier and Kurtz (1986, Theorem 3.1, p. 351). A rigorous (yet straightforward) way of incorporating these conditions into our problem may be found in Greenshtein and Ritov (1997); see also related results in Chapter 4 of Brodsky and Darkhovsky (1993).

Notice that $\mathbb{S}_t - kN_t = \tilde{\mathbb{S}}_t$, where $\tilde{\mathbb{S}}_t$ is induced by the stationary process $\tilde{X}_t = (X_t - k)$ and the particular sampling plan. Thus the theory developed in the previous section may be applied to evaluate the coefficient $\sigma_k^2 = \lim_{t\to\infty} \operatorname{var}(\tilde{\mathbb{S}}_t)/t$.

A straightforward adaptation of Theorem 3.6 and expression (2.57) in Siegmund (1985) (see Greenshtein and Ritov 1997, for details) implies that the expected time to a false alarm is

$$\mathrm{E}(\tau) = \begin{cases} 1/2(\theta/\sigma_k)^{-2}\left(\exp\left[-\dfrac{2\theta b}{\sigma_k^2}\right] + \dfrac{2\theta b}{\sigma_k^2} - 1\right), & \theta \neq 0, \\ \dfrac{b^2}{\sigma_k^2}, & \theta = 0. \end{cases} \tag{4.2}$$

This is the expected time until a false alarm from time 0, or the time to alarm from the change point in the worst case (see Lorden 1971). It may be seen, from the last equation, that the dependence structure of the process $\{X_t\}$ affects the expected time to an alarm only through the asymptotic variance coefficient σ_k^2.

4.2. Exponential sampling plans

In the following we will apply the above results to the case where **T** is an exponential sampling plan. In order to apply formula (4.2), we need the expression for σ_k^2. The expression is given in Corollary 3.3. In the development for exponential sampling plans we denote $\mu^{-1} = \lambda$.

In the following we illustrate the results that were obtained by a numerical study. For any procedure that detects a change in a process, if the expected time to detect a change is finite, then the expected time until false alarm is finite as well. The cusum procedure is defined by two parameters, b and k. Usually they are found so that the time to a (true) alarm under a given value p_1 of p is minimized, while the time to a (false) alarm when $p = p_0$ is constrained to a given value. This value represents the tolerable time to false alarm under normal conditions. In Figure 1 we plotted equation (4.2) when the expected time to false alarm was restricted to 10 000 under $p = 0.1$, and the parameters of the procedure were chosen to minimize the detection time under $p = 0.12$. That is, we essentially maximized the slope of the time to detection as a function of p. The equation is plotted for two different values of B, 0.02 and 0.2, and four different values of λ. It can be seen that the performance of the procedure hardly changes when the sampling rate is increased from 1 to 10. It may also be seen that the advantage in increasing the sampling rate is more significant under $B = 0.02$ than under $B = 0.2$. The explanation for this is that a large value of B means a strong positive dependence (see (3.1) and (3.2)) among the variables X_t. Thus, the extra information in more frequent sampling is small.

Also plotted are Monte Carlo estimated times to false alarm of a 0–1 Markov process as in Example 2.2, with the appropriate parameters. For example, $p = 0.1$ and $B = 0.02$ correspond to $\lambda_0 = 0.9$ and $\lambda_1 = 8.1$, and an average cycle time of 1.235. The value of λ, the sampling intensity, was taken to be equal to 1. The estimates are based on 100 simulations. Since the distribution of the detection times is approximately exponential, the standard error is approximately 10% of the mean time. Note that the estimated values for the different values of p are highly correlated. It can be seen that the asymptotic approximations are pretty good.

The Monte Carlo estimated times to detection were 9270, 216, 108, 71 and 52, for $p = 0.1, 0.2, \ldots, 0.5$, respectively. That is, the time to detection is approximately 180 times longer under $p = 0.1$ than under $p = 0.5$.

4.3. Stratified sampling plans

In the following we will examine our general development in the case where **T** is a stratified sampling plan. In this case σ_k^2 is independent of k, since N_t is deterministic, and its value is given in Proposition 3.5. In practice, when determining the sampling rate μ^{-1}, we should assume that under a change in the probabilistic mechanism of the process, which results in a small change in the mean of its stationary distribution, the change in σ_k^2 is negligible.

In the following, we take $\{X_t\}$ as a 0–1 Markovian process. Initially the process has the parameters $\lambda_0 = 1$ and $\lambda_1 = 1$ (see Example 2.1). We study the performance of stratified

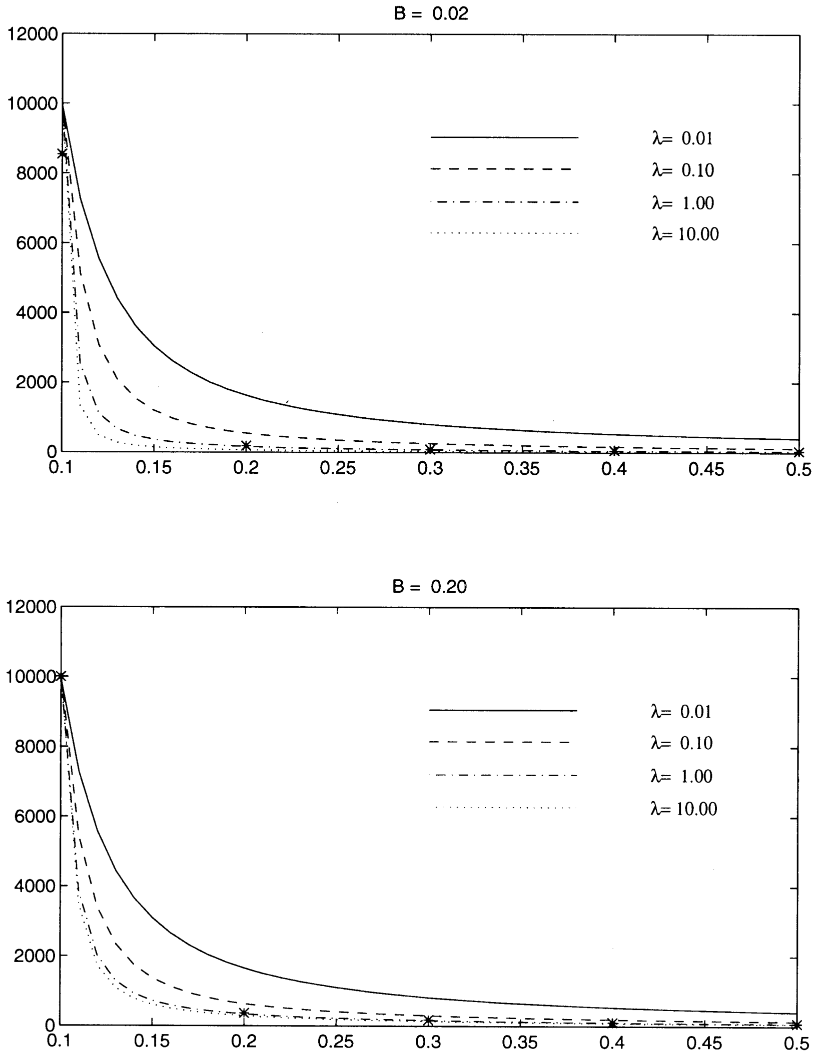

Figure 1. Time to alarm as a function of p, for two different values of the variance coefficient B and four sampling rates λ. The lines are plots of equation (4.2), while the stars denote the average from 100 Monte Carlo experiments of a 0–1 Markov process with sampling intensity 1.

sampling plans with four values of sampling rates under a change in λ_1. As in the numerical study of the exponential sampling plans, the time to false alarm was restricted to 10 000, and the parameters b and k were chosen to minimize the expected time to alarm at the point $\lambda_1 = 0.98$ under the restriction. The results may be seen in Figure 2.

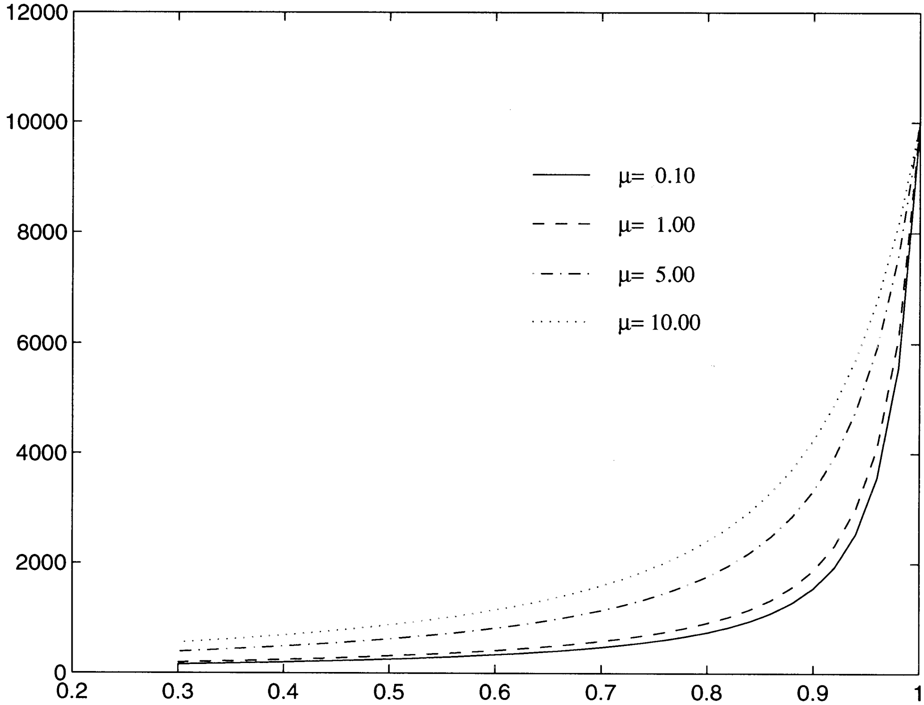

Figure 2. Time to alarm as function of λ_1 for four sampling rates μ^{-1}. The lines are plots of equation (4.2) for 0–1 Markov processes with parameters λ_i, $i = 1, 2$.

We conclude with two remarks on generalizations and extensions of the results in this section.

Remark 4.1. One generalization is to the case where there are errors in measurements – i.e. suppose that the experimenter observes the process $\tilde{X}_{T_i}$, where $\tilde{X}_{T_i} = X_{T_i} + \varepsilon_{T_i}$, and $\{\varepsilon_{T_i}\}$ are i.i.d. random errors with mean 0.

A procedure that suggests itself is first to estimate X_{T_i} by $\hat{X}_t$, where the estimator is based on the previous observations $\tilde{X}_{T_1}, \ldots, \tilde{X}_{T_{N_t}}$ and $T_1, \ldots, T_{N_t}$. Then use the statistic $\tilde{\mathbb{S}}_t - kN_t$, where $\hat{W}_t = \hat{X}_{T_1}, \ldots, \hat{X}T_{N_t}$, along the lines of Section 4. A naive estimator that suggests itself is $\hat{X}_{T_i} = \tilde{X}_{T_i}$.

Remark 4.2. Another direction of generalization is to the case of dynamic sampling. It seems plausible that a significant improvement will result if we permit sampling plans with dynamic sampling rate. It also seems plausible that the intensity of a 'good' sampling plan will be positively correlated with the 'amount of evidence' in favour of the event that a change has occurred. Such an approach will lead us to investigate more general diffusion approximations than used in this section. An approach somewhat similar in spirit may be found in Assaf and Ritov (1989).

Acknowledgement

The research of the first author was supported the fund for the promotion of research at Technion. We are grateful to Michael Schurman for suggesting and describing the problem of sampling from a network. The contribution of the referees and an Associate Editor to a critical improvement of the presentation is acknowledged.

References

Assaf, D. and Ritov, Y. (1989) A dynamic sampling approach for detecting a change in the drift of Brownian motion. A non-Bayesian model. *Ann. Statist.*, **17**, 793–800.

Basseville, M. and Nikiforov, I.V. (1993) *Detection of Abrupt Changes: Theory and Application.* Englewood Cliffs, NJ: Prentice Hall.

Bellhouse, D.R. (1988) Systematic sampling. In P.R. Krishnaiah and C.R. Rao (eds), *Handbook of Statistics 6: Sampling*, pp. 125–145. Amsterdam: North-Holland.

Blight, B.J.N. (1973) Sampling from an autocorrelated finite population. *Biometrika*, **60**, 375–385.

Brillinger, R.D. (1973) Estimation of the mean of a stationary time series by sampling. *J. Appl. Probab.*, **10**, 419–431.

Brodsky, B.E. and Darkhovsky, B.S. (1993) *Nonparametric Methods in Change-Point Problems.* Dordrecht: Kluwer Academic Publishers.

Cambanis, S. and Masry, E. (1983) Sampling designs for the detection of signals in noise. *IEEE Trans. Inform. Theory*, **IT-29**, 83–104.

Ethier, N.S. and Kurtz, T.G. (1986) *Markov Processes. Characterization and Convergence*. New York: Wiley.

Good, J.I. (1961) The frequency count of a Markov chain and the transition to continuous time. *Ann. Math. Statist.*, **32**, 41–48.

Greenshtein, E. and Ritov, Y. (1997) Sampling designs for detecting a change in the mean of a stationary distribution. Technical Report, Dept. of Industrial Engineering, Technion.

Karlin, S. and Taylor, H.M. (1975) *A First Course in Stochastic Processes*. New York: Academic Press.

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

Murthy, M.N. and Rao, T.J. (1988) Systematic sampling with illustrative examples. In P.R. Krishnaiah and C.R. Rao (eds), *Handbook of Statistics 6: Sampling*, pp. 147–185. Amsterdam: North-Holland.

Pollak, M. (1987) Average runlength of an optimal method of detecting a change in distribution. *Ann. Statist.*, **15**, 749–779.

Roll, Y. and Yadin, M. (1986) Activity sampling in a stochastic environment. *IIE Trans.*, **18**, 343–349.

Sacks, J. and Ylvisaker, D. (1966) Designs for regression problems with correlated errors. *Ann. Math. Statist.*, **37**, 66–89.

Schoenfelder, C. and Cambanis, S. (1982) Random designs for estimating integrals of stochastic processes. *Ann. Statist.*, **10**, 526–538.

Siegmund, D.O. (1985) *Sequential Analysis*. New York: Springer-Verlag.

Shiryayev, A.N. (1983) On optimal methods in earliest detection problems. *Theory Probab. Appl.*, **8**, 26–51.

Ylvisaker, D. (1987) Prediction and design. *Ann. Statist.*, **15**, 1–19.

Received October 1997 and revised December 1999