

# Detecting the emergence of a signal in a noisy image\*

DAVID SIEGMUND AND BENJAMIN YAKIR

---

We study sequential change-point detection when observations form a sequence of independent Gaussian random fields, and the change-point is the time at which a signal of known functional form involving a finite number of unknown parameters appears. Building on Siegmund and Yakir (2008), which identifies in a simpler problem a detection procedure of Shiriyayev-Roberts type that is asymptotically minimax up to terms that vanish as the false detection rate converges to zero, we compare easily computed approximations to the Shiriyayev-Roberts detection procedure with similar approximations to CUSUM type procedures. Although the CUSUM type procedures are suboptimal, our studies indicate that they compare favorably to the asymptotically optimal procedures.

AMS 2000 SUBJECT CLASSIFICATIONS: Primary 62L15; secondary 60G40.

KEYWORDS AND PHRASES: Change-point, Sequential detection, Image detection.

---

## 1. INTRODUCTION

Sequential change-point detection, going back to Shewhart's [6] classic paper, has played an important role in quality control of continuous manufacturing processes. More recently change-point and other related models have been developed to deal with an array of problems in a variety of applied fields: detection of regions of activation in brain scans [13], DNA/Protein sequence alignment [10], scanning statistics [9], disease surveillance statistics ([1, 4], etc.) In many of these applications one scans a noisy field of data to find a signal occupying a relatively small part of the field of observation.

In this paper we consider the problem of sequentially scanning a random field to detect the emergence of a signal, say a localized change in the incidence of a particular disease due to the accidental or intentional introduction of a toxic agent into the environment or the appearance of an object in the field of vision of a radar. The signal might take on an almost arbitrary shape and appear either suddenly or

gradually. Generally speaking, the problem is to detect the signal as soon as possible after it appears, under the constraint that (false) detection occurs very rarely if no signal appears. To illustrate the principles involved we will assume that the observations consist of a sequence of uncorrelated Gaussian random fields and that the signal is a parameterized function of a known form superimposed on the noisy observations at an unknown change-point. The discussion will be formulated in the context of image analysis. Yet the results and the principles that we introduce are meaningful in other settings as well.

Following the developments in the companion to this paper [12], we measure detection delay by the expected Kullback-Leibler information accumulated between the change-point and its detection. We begin in the next subsection with a precise description of the model. In Subsection 1.2 the criterion for asymptotic minimax optimality, which is stated and proved in the companion paper, is reformulated to fit the current context.

The asymptotic minimax policy uses a randomized form of the Shiriyayev-Roberts monitoring scheme. The alternative Cumulative Sum (CUSUM) monitoring scheme is a better known approach. In Section 2 expressions describing the asymptotic performance of optimal CUSUM and optimal Shiriyayev-Roberts rules are obtained. Suboptimal formulations of these procedures are also assessed.

It turns out that the natural candidate for optimal Shiriyayev-Roberts and CUSUM rules may require substantial computation. In Section 3 we propose alternative rules which are asymptotically equivalent but require less computational effort. A simulation study is conducted in order to investigate the finite-sample properties of these simplified rules. The paper concludes with a discussion of related open problems.

The analysis of the different detection methods draws on a substantial literature for its justification. The calculations are only sketched here and emphasize new features of the present formulation. See [12] and the references cited there, and [3].

### 1.1 The model

Denote monitoring periods by  $i$ ,  $i = 1, 2, \dots$ . An image is composed of a finite collection of pixels  $\mathcal{J}$ . At each monitoring period  $i$  and pixel location  $j$  an observation  $X_{ij}$  is

---

\*This research has been supported by the NSF and by the US-Israel Binational Science Foundation.

recorded. We will assume in this paper that all measurements are Gaussian, independent, and have a fixed, known variance  $\sigma^2$ . Initially they also have mean 0.

The presence of a signal of a given shape and intensity in an image is reflected in the mean of the pixel measurements of which the image is composed. The shape of a signal is given by a function  $g$ ,  $g : \mathcal{J} \rightarrow \mathbb{R}$  and its intensity is a positive number  $\mu$ . When a signal is present, the expectation of the measurement  $X_{ij}$  is  $\mu g(j)$ . When the signal is absent the expectation of all observations is zero. The image at period  $i$  is the random field  $X_i = \{X_{ij} : j \in \mathcal{J}\}$ . The log-likelihood ratio for testing a given image for the presence of a signal takes the form  $\{\mu \langle g, X_i \rangle - \mu^2 \langle g, g \rangle / 2\} / \sigma^2$ . Note that  $\mu$  and  $\langle g, g \rangle$  cannot be separately estimated, so we assume that  $g$  is standardized by the condition  $\langle g, g \rangle = 1$ .

We restrict the possible locations and shapes of the signals to belong to a smooth parametric family  $\{g_t : t \in T\}$ , for a compact  $T \subset \mathbb{R}^d$ . Hence a signal is parameterized by  $\theta = (\mu, t)$ . The conditional log-likelihood for the sequence of images up to time  $n$  for the case that the signal emerged at period  $k$  is

$$(1) \quad \begin{aligned} \ell_k^\theta(n) &= \sum_{i=k}^n \{\mu \langle g_t, X_i \rangle - \mu^2 / 2\} / \sigma^2 \\ &= \frac{\mu}{\sigma^2} \langle g_t, X_k(n) \rangle - (n - k + 1) \frac{\mu^2}{2\sigma^2}, \end{aligned}$$

where  $X_k(n) = \sum_{i=k}^n X_i$ .

For a simple example that we will use later for numerical illustration let  $T$  be the unit square  $[0, 1]^2$  and let  $\mathcal{J}$  be a regular grid of points. Put  $g_t(j) = C_t \exp(-K \|t - j\|^2)$ , where  $\|\cdot\|$  denotes  $d$ -dimensional Euclidean norm and  $C_t$  is the normalizing constant. Thus the signal  $g_t(\cdot)$  is a Gaussian-shaped hill centered at  $t$  with spherical contours. The normalizing constant depends on  $t$ , but it is approximately independent of  $t$  if  $K$  is sufficiently large, the distance between points of the grid sufficiently small, and the boundary of  $\mathcal{J}$  sufficiently regular that summation over the grid is well approximated by integration over the entire Euclidean space.

For illustration consider the panel of images shown in Figure 1. They are associated with the example given above. The grid has  $2^5 \times 2^5 = 1024$  pixels in the unit square,  $t = (0.3, 0.5)$ ,  $K = 50$ , and  $\mu = 2^{1/2}$ . In the upper-left im-

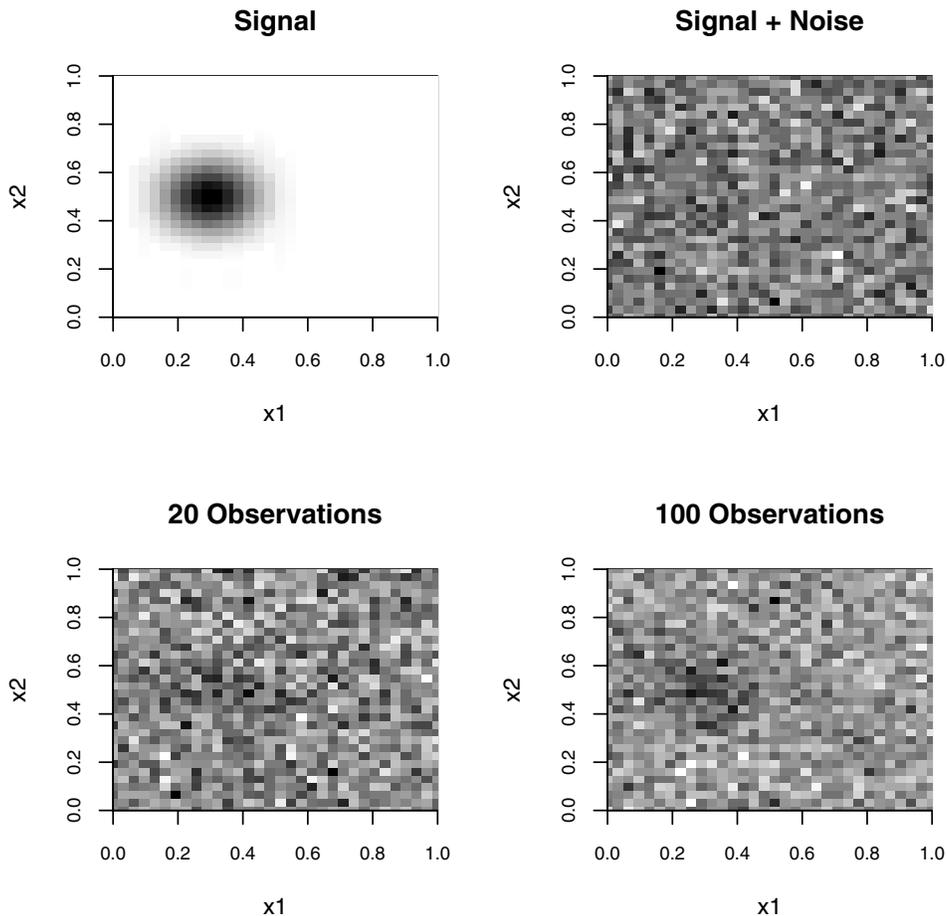


Figure 1. An Example. Two Dimensional Grid of Observations Is  $2^5 \times 2^5$ ,  $t = (0.3, 0.5)$ ,  $K = 50$ ,  $\mu = 2^{1/2}$ . The Upper Left Panel Gives Just the Signal. In the Other Panels Independent Gaussian Noise with Unit Standard Deviation Have Been Added to Each Observation.

age the signal is plotted. It should be noted that the height of the peak of the signal is approximately 0.253. A typical image is obtained by convolving the signal with Gaussian white noise of unit standard deviation. The result is shown in the upper-right panel. There is a 75% chance of obtaining observations larger than 3. Not surprisingly, the presence of a signal is difficult to detect. The lower-left plot is obtained by averaging 20 independent images that contain the signal. Knowing the presence and location of the signal, we observe an elevated region associated with the signal. The presence and the location of the signal become more apparent after the averaging of 100 independent images. This is shown in the lower-right image.

The goal of sequential change-point detection is to detect the presence of a signal as soon as possible after its emergence. No prior knowledge is assumed about the time of emergence or of the location of the signal. The procedures described below all detect the given signal with high probability well before the 20th observation. Yet some procedures may perform better than others. The investigation of the relative performance of the different schemes is the focus of this paper.

## 1.2 Optimal rates of detection

We use  $P_\infty$  to denote the probability measure when there is no change and  $P_k^\theta$  to denote probability parameterized by a change at time  $k$  to a parameter value  $\theta$ . Note that the sequence of alternative-specific log-likelihood ratios,  $\ell_k^\theta(n)$  given in (1), as a function of  $n$ , for  $n \geq k$ , is a Gaussian random walk. The expectation of its increments is the Kullback-Leibler information  $I(\theta) = I(\mu) = \mu^2/2\sigma^2$ , which is a function of the intensity alone.

For future reference we define the stopping rule of a one-sided sequential probability ratio test:  $\tau^\theta = \inf\{n : \ell_1^\theta(n) \geq a\}$ . It follows from the renewal theorem that

$$(2) \quad \begin{aligned} v(\mu) &= \lim_{a \rightarrow \infty} E_1^\theta[\ell_1^\theta(\tau^\theta) - a] \quad \text{and} \\ \lambda(\mu) &= \lim_{a \rightarrow \infty} E_1^\theta \exp\{-[\ell_1^\theta(\tau^\theta) - a]\} \end{aligned}$$

exist, can be computed numerically, and depend only on  $\mu$ . Another quantity of interest that is related to the one-sided sequential probability ratio test and depends only on  $\mu$  is defined via an infinite sum of likelihood ratios:

$$(3) \quad w(\mu) = E_1^\theta \left[ \log \left( 1 + \sum_{n=1}^{\infty} \exp\{-\ell_1^\theta(n)\} + \sum_{j=0}^{\infty} \exp\{\ell_{-j}^\theta(0)\} \right) \right].$$

Consider now  $\langle g_t, g_t \rangle$ , which has been standardized to be identically equal to one. By differentiation one gets that  $\langle \dot{g}_t, g_t \rangle$  is the zero vector and that  $-\langle \ddot{g}_t, g_t \rangle = \langle \dot{g}_t, \dot{g}_t \rangle$  is a non-negative definite  $(d \times d)$ -matrix. Note that the square root of the determinant of this matrix is the volume element for the manifold  $g_t, t \in T$ , which is a submanifold of the unit sphere in  $d$ -dimensional Euclidean space.

We consider a collection of post-change parameter values  $\Theta = [\mu_0, \mu_1] \times T$ , where  $\mu_0 > 0$ . Let  $\Theta_\varepsilon = \Theta + B(\varepsilon)$ , where  $B(\varepsilon)$  is the closed ball in  $\mathbb{R}^{d+1}$  of radius  $\varepsilon$ . Assume that  $\varepsilon > 0$  is such that  $\Theta_\varepsilon$  is compact and  $0 \notin \Theta_\varepsilon$ . We also assume that  $\langle \dot{g}_t, \dot{g}_t \rangle$  and  $\langle \dot{g}_t, \dot{g}_t \rangle^{-1}$  are uniformly Lipschitz-continuous with eigenvalues that are bounded away both from zero and from infinity. Another assumption is that the measure of the difference  $\Theta_\varepsilon \setminus \Theta$  converges to zero as  $\varepsilon \rightarrow 0$ .

Detection delay is measured by the amount of Kullback-Leibler information between change-point  $k$  and detection time  $N$ . A uniform lower bound on the expected delay is proved in [12]. After adaptation to the current situation the lower bound may be stated as:

**Theorem 1.** *Let  $N$  be a stopping rule parameterized by  $A > 1$ . Assume that*

$$(4) \quad P_\infty(N \leq l + m | N > l) \leq (m/A)[1 + o(1)]$$

for some  $l \geq 0$ , perhaps depending on  $A$ , and for some  $m = m(A)$  such that  $(\log A)^{1+\delta} \ll m \ll \exp\{(\log A)^\delta\}$ . Then, as  $A \rightarrow \infty$  and up to an  $o(1)$  term,

$$(5) \quad \begin{aligned} &\sup_{1 \leq k < \infty, \theta \in \Theta} \frac{\mu^2}{2\sigma^2} E_k^\theta [N - k + 1 | N \geq k] \\ &\geq \log A + \frac{d+1}{2} \log \log A - \frac{d+1}{2} \log(\pi\sigma^2 e) \\ &\quad + \log \left\{ \int_{\mu_1}^{\mu_2} \frac{\lambda(\mu)}{\mu} e^{v(\mu) - w(\mu)} d\mu \times \int_T \sqrt{\det(\langle \dot{g}_t, \dot{g}_t \rangle)} dt \right\}. \end{aligned}$$

Let  $\rho$  denote a prior distribution over  $\Theta_\varepsilon$  and consider the (marginal) log-likelihood ratios

$$(6) \quad \ell_k(n) = \log \int_{\Theta_\varepsilon} \exp\{\ell_k^\theta(n)\} \rho(\theta) d\theta,$$

for  $k \leq n$ . The (window restricted) Shirayev-Roberts stopping rule is defined by

$$(7) \quad N_{\text{SR}}^w = \inf\{n : r^w(n) \geq \log A + \log \lambda_{\text{SR}}\},$$

where  $r^w(n) = \log \sum_{k=n-w}^n e^{\ell_k(n)}$ . The constant  $\lambda_{\text{SR}}$  is a correction factor (obtained below) to produce the asymptotic false detection rate  $E_\infty(N_{\text{SR}}^w) \sim A$  as  $A \rightarrow \infty$ . (For  $n \leq w$  one may simulate  $X_{-w+1}, X_{-w+2}, \dots, X_0$  from the null distribution and attach them to the left of the sequence of original observations. Likelihood ratios for negative  $k$  can now be formed, which enables the calculation of the scanning statistic for small, but positive, values of  $n$  and the production of a randomized stopping rule.)

An adaptation of Theorem 2 of [12] states that for a suitable  $\rho$ , the Shirayev-Roberts rule asymptotically achieves the lower bound in Theorem 1 and hence is asymptotically minimax up to terms that vanish as  $A \rightarrow \infty$ . The appropriate prior  $\rho$  makes the asymptotic expected delay constant

with regard to  $\theta$ . It follows from (17) given below that this prior is

$$(8) \quad \varrho(\theta) \propto \sqrt{\det(\langle \dot{g}_t, \dot{g}_t \rangle)} e^{v(\mu) - w(\mu)} \mu^{-1}.$$

(Note that this prior makes  $\mu$  and  $t$  independent.) The asymptotic optimality of this particular Shiriyayev-Roberts procedure is stated in:

**Theorem 2.** *Let  $\varepsilon = \varepsilon(A) = (\log A)^{-1/4}$ . Using  $\rho(\theta) = \varrho(\theta) \cdot I_{\Theta_\varepsilon} / \int_{\Theta_\varepsilon} \varrho(\vartheta) d\vartheta$  and  $\lambda_{SR} = \int_{\Theta_\varepsilon} \lambda(\theta) \rho(\theta) d\theta$ , we have  $E_\infty(N_{SR}^w) \sim A$ , the inequality (4) is satisfied for all  $l$  and for all  $m$  of the given magnitude. Moreover, the given stopping time achieves equality in (6) (up to an  $o(1)$  term) provided that  $\log A \ll w$ .*

## 2. ASYMPTOTIC APPROXIMATIONS FOR OPERATING CHARACTERISTICS

In this section we consider several detection schemes, calculate approximately their statistical properties, and compare them to the performance of the optimal procedure. The main results are the development of formulas for the asymptotic rates of false detection and the asymptotic expected delays until a true detection.

### 2.1 Likelihood ratios

All the stopping rules that will be considered are either based directly on likelihood ratios or on statistics derived from likelihood ratios. Recall that the likelihood function takes the form

$$\exp\{\ell_k^\theta(n)\} = \exp\left\{\frac{\mu}{\sigma^2} \langle g_t, X_k(n) \rangle - (n - k + 1) \frac{\mu^2}{2\sigma^2}\right\}.$$

As above, we can use a marginal likelihood, obtained by selection of a prior distribution  $\rho$  over the parameter space and the integration of the likelihood function with respect to that prior to obtain (6).

Alternatively, one may apply maximum-likelihood estimation

$$\hat{\theta} = \hat{\theta}_{k,n} = \arg \max \ell_k^\theta(n)$$

and evaluate the likelihood at that value in order to produce the generalized likelihood ratio statistic:  $\ell_k^{\hat{\theta}}(n) = \max_\theta \ell_k^\theta(n)$ . In our case one can see that, for a given  $t \in T$ , maximization with respect to  $\mu \geq 0$  produces

$$\hat{\mu}(t) = \langle g_t, X_k(n) \rangle^+ / (n - k + 1).$$

Therefore,

$$(9) \quad \ell_k^{\hat{\theta}}(n) = \frac{1}{2} (\langle g_{\hat{t}}, Z_k(n) \rangle^+)^2,$$

where  $Z_k(n) = X_k(n) / (n - k + 1)^{1/2}$  and  $\hat{t} = \arg \max_{t \in T} \langle g_t, Z_k(n) \rangle$ .

## 2.2 The CUSUM and Shiriyayev-Roberts procedures

One may apply either the marginal likelihood or the generalized-likelihood statistics given in (6) and (9) to produce detection procedures. In this paper we will consider two classes of detection procedures: The Shiriyayev-Roberts type and the Cumulative sum (CUSUM) type. Generally speaking, the former uses sums of likelihoods (or sums of exponentiated log-likelihoods) for monitoring whereas the latter uses maximization instead of summation. One may or may not restrict the summation in the one case and maximization in the other to a sliding window of potential change-point values of  $k$ . Here we will employ such a restriction and denote the width of the window by  $w$ .

For example, the window-restricted Shiriyayev-Roberts procedure when marginal likelihoods are used was given above in (7).

A CUSUM stopping rule is defined in a similar way but with maxima replacing sums. The stopping rule for the marginal likelihoods is:

$$N_{CS}^w = \inf\{n : q^w(n) \geq \log A + \lambda_{CS}\},$$

where  $q^w(n) = \max_{n-w \leq k \leq n} \ell_k(n)$  and  $\lambda_{CS}$  is the appropriate correction factor to produce the asymptotic rate of false detection of  $1/A$ .

For a CUSUM stopping rule, there is also the possibility to use the generalized log-likelihood ratio statistic, which has the marginal log-likelihood  $\ell_k(n)$  replaced by  $\ell_k^{\hat{\theta}}(n)$ , with the correction factor appropriately modified. Note that we are considering the maximum likelihood estimator restricted only by the condition that  $\mu \geq 0$ . (Some reflection will show that the analogous Shiriyayev-Roberts statistic will not prove to be useful, although a modification involving maximization only over  $\Theta$  might be.)

The generalized log likelihood ratio statistic is much simpler to compute than statistics that involve integration over a subset of the parameter space. Other possibilities for simplifying the computation of the detection statistics are discussed in Section 3.

## 2.3 The distribution when there is no change

It can be shown that for all the cases that we consider the null distribution of the stopping time is asymptotically exponential. The asymptotic mean of the exponential distribution is determined by evaluating asymptotically the probability  $P_\infty(N \leq m)$ , where  $N$  is any of the stopping rules in the list. In order to examine these probabilities in the different settings, it is useful to note that they all can be expressed in terms of tail probabilities of functionals produced by integration and maximization operations on appropriate likelihood fields.

Indeed, if we consider the  $P_\infty$  distribution of the stopping time  $N_{\text{SR}}^w$  then we may rewrite the event in question in the form:

$$\{N_{\text{SR}}^w \leq m\} = \left\{ \max_{1 \leq n \leq m} \sum_{n-w \leq k \leq n} \int_{\Theta} e^{\ell_k^\theta(n)} \rho(\theta) d\theta \geq A\lambda_{\text{SR}} \right\}.$$

Likewise, for the CUSUM procedure we will obtain

$$\{N_{\text{CS}}^w \leq m\} = \left\{ \max_{1 \leq n \leq m} \max_{n-w \leq k \leq n} \int_{\Theta} e^{\ell_k^\theta(n)} \rho(\theta) d\theta \geq A\lambda_{\text{CS}} \right\}.$$

For the profiled-based approaches the integration is replaced by maximization with respect to  $\theta$ .

All the probabilities may be studied with the aid of measure transformation. Specifically, if we denote a generic event by  $\{N \leq m\}$  then

$$(10) \quad \begin{aligned} & P_\infty(N \leq m) \\ &= \int_{\Theta} \sum_{n=1}^m \sum_{k=1}^n E_{k,n}^\theta \\ & \quad \left[ \frac{e^{-\ell_k^\theta(n)}}{\sum_{l=1}^m \sum_{j=1}^l \int_{\Theta} e^{\ell_j^\theta(l) - \ell_k^\theta(n)} \rho(\vartheta) d\vartheta}; N \leq m \right] d\theta, \end{aligned}$$

where  $P_{k,n}^\theta$  is the measure according to which the observations in the interval  $[k, n]$  are distributed with parameter  $\theta$  and the observations not in the interval are distributed according to the null distribution.

Analysis of (10) involves localization, where sums, integrals, and/or maxima are replaced by similar expressions involving restricted domains. Recall that  $I(\mu) = \mu^2/2\sigma^2$  denotes the Kullback-Leibler divergence measure of a single observation. For any given  $\theta$ , only expectations in (10) with indexes such that  $|n - k - [1/I(\mu)] \log A| \leq c(\log A)^{1/2}$  contribute significantly to the total sum. Additional localization is conducted for the term in the denominator of each expectation and random field associated with the definition of the event. For  $N = N_{\text{SR}}$ , we can replace the expectation appearing in (10) by

$$E_{k,n}^\theta \left[ \frac{M_{\text{SR}}}{S_{\text{SR}}} \exp\{-[\ell_k^\theta(n) \geq \log(A\lambda_{\text{SR}}) - \log M_{\text{SR}}]\}; [\dots] \geq 0 \right],$$

where

$$\begin{aligned} S_{\text{SR}} &= \sum_{|l-n| \leq c} \sum_{|j-k| \leq c} \int_V e^{\ell_j^\theta(l) - \ell_k^\theta(n)} \rho(\vartheta) d\vartheta, \\ M_{\text{SR}} &= \max_{|l-n| \leq c} \sum_{|j-k| \leq c} \int_V e^{\ell_j^\theta(l) - \ell_k^\theta(n)} \rho(\vartheta) d\vartheta, \end{aligned}$$

and the notation  $[\dots]$  denotes the quantity appearing inside square brackets in the preceding exponential.

By taking a series of differences and neglecting second order differences, we obtain the following approximation (see [12] for a partial justification):

$$(11) \quad \begin{aligned} & \int_V e^{\ell_j^\theta(l) - \ell_k^\theta(n)} \rho(\vartheta) d\vartheta \\ & \approx \int_V e^{\ell_k^\theta(n) - \ell_k^\theta(n)} \rho(\vartheta) d\vartheta \times e^{\ell_j^\theta(n) - \ell_k^\theta(n)} \times e^{\ell_k^\theta(l) - \ell_k^\theta(n)}. \end{aligned}$$

Replacing the local likelihood ratios by these approximations produces the expressions  $\hat{S}_{\text{SR}}$  and  $\hat{M}_{\text{SR}}$ , which can replace the quantities  $S_{\text{SR}}$  and  $M_{\text{SR}}$ , respectively, to produce yet other asymptotically equivalent approximations.

An application of a local limit theorem justifies the replacement of the given expectation by the quantity

$$(A\lambda_{\text{SR}})^{-1} E_{k,n}^\theta \left[ \frac{\hat{M}_{\text{SR}}}{\hat{S}_{\text{SR}}} \right] \times \frac{1}{\sigma_{k,n}^\theta} \phi \left( \frac{\log A - \mu_{k,n}^\theta}{\sigma_{k,n}^\theta} \right),$$

where  $\mu_{k,n}^\theta = E_{k,n}^\theta[\ell_k^\theta(n)]$  and  $\sigma_{k,n}^\theta = (\text{var}_{k,n}^\theta[\ell_k^\theta(n)])^{1/2}$ . Noting that the processes  $\{\ell_j^\theta(n) - \ell_k^\theta(n) : |j - k| \leq c\}$  and  $\{\ell_k^\theta(l) - \ell_k^\theta(n) : |l - n| \leq c\}$  are stochastically independent enables cancelation of the summation over  $j$  in numerator and denominator of the fraction. Hence

$$E_{k,n}^\theta \left[ \frac{\hat{M}_{\text{SR}}}{\hat{S}_{\text{SR}}} \right] = E_{k,n}^\theta \left[ \frac{\max_{|l-n| \leq c} e^{\ell_k^\theta(l) - \ell_k^\theta(n)}}{\sum_{|l-n| \leq c} e^{\ell_k^\theta(l) - \ell_k^\theta(n)}} \right] \approx E_1^\theta \left[ \frac{M_{\text{SPRT}}}{S_{\text{SPRT}}} \right],$$

which is independent of  $k$  and  $n$ . Summation with respect to  $k$  and  $n$  will thus produce the approximation

$$\begin{aligned} P_\infty(N_{\text{SR}} \leq m) &\approx \frac{m}{A\lambda_{\text{SR}}} \int_{\Theta} \frac{1}{I(\mu)} E_1^\theta \left[ \frac{M_{\text{SPRT}}}{S_{\text{SPRT}}} \right] \rho(\theta) d\theta \\ &= \frac{m}{A\lambda_{\text{SR}}} \int_{\Theta} \lambda(\mu) \rho(\theta) d\theta, \end{aligned}$$

where  $\lambda(\mu)$  was defined in (2). Consequently for the Shriyayev-Roberts rule

$$(12) \quad \lambda_{\text{SR}} = \int_{\Theta} \lambda(\mu) \rho(\theta) d\theta.$$

Now consider the CUSUM stopping rule. The same machinery can be recycled, albeit with some modifications. An important difference is that the term  $\sum_{|j-k| \leq c} \exp\{\ell_j^\theta(n) - \ell_k^\theta(n)\}$  in the numerator of the ratio that is produced after the application of the local central limit theorem is replaced by  $\max_{|j-k| \leq c} \exp\{\ell_j^\theta(n) - \ell_k^\theta(n)\}$ . Therefore, instead of cancelation, an extra factor of the form  $I(\mu)\lambda(\mu)$  is produced. As a result, one gets that

$$(13) \quad \lambda_{\text{CS}} = \int_{\Theta} I(\mu) [\lambda(\mu)]^2 \rho(\theta) d\theta.$$

For the generalized likelihood ratio statistic, from (21), we obtain

$$\exp[\ell_j^\theta(l) - \ell_k^\theta(n)] \approx e^{\ell_k(n) - \ell_k^\theta(n)} \times e^{\ell_j^\theta(n) - \ell_k^\theta(n)} \times e^{\ell_k^\theta(l) - \ell_k^\theta(n)} \\ \times \frac{(n - k + 1)^{(d+1)/2} \mu^d \sqrt{\det(\langle \dot{g}_t, \dot{g}'_t \rangle)}}{\rho(\theta) (2\pi\sigma^2)^{(d+1)/2}}.$$

We are interested in the distribution of this quantity when  $n$  is replaced by the stopping time  $N$ . A crude first order approximation shows that  $(N - k)/\log A \rightarrow 1/I(\mu)$  in  $P_k^\theta(\cdot|N \geq k)$ -probability. Substitution and calculation show that

$$\lambda_{\text{CS}}^{\hat{\theta}} = \left( \frac{\log A}{2\pi\sigma^2} \right)^{(d+1)/2} \int_0^\infty \int_T \frac{[\lambda(\mu)]^2 \mu^d \sqrt{\det(\langle \dot{g}_t, \dot{g}'_t \rangle)} d\mu dt}{[I(\mu)]^{(d-1)/2}}. \quad (14)$$

**Remarks.** Recalling that the matrix  $\langle \dot{g}_t, \dot{g}'_t \rangle$  is the metric tensor for the manifold  $g_t$ ,  $t \in T$ , so the integral with respect to  $dt$  of the square root of its determinant is the volume, say  $G_T$ , of the manifold, and that  $I(\mu) = \mu^2/2\sigma^2$ , we see that up to a numerical factor the integral in (9) simplifies to  $\int \mu \lambda^2(\mu) d\mu \times G_T$ . In the case  $d = 0$ , (14) agrees with [7].

## 2.4 The expected accumulated information

We measure the delay in detection of a change by the expected Kullback-Leibler information accumulated between change-point and stopping time, given that no declaration was made prior to the change point. Hence, for a given detection rule  $N$ , change-point  $k$  and post-change distribution  $\theta$  the expected rate of detection is  $E_k^\theta[\ell_k^\theta(N)|N \geq k] = I(\mu) \cdot E_k^\theta[N - k + 1|N \geq k]$ .

In every case we consider, the stopping rule is a threshold rule, with the monitoring process and the threshold level varying between cases. Given the change-point and the post-change parameter, the monitoring process is in all cases a sum of the random walk process of log-likelihoods (with respect to the change-point and the parameter) and a slowly changing process. Thus we can assess the expected Kullback-Leibler information between change-point and detection by examining the stopped process twice: in the first case considering it as a sum of the threshold and the overshoot of the process over the threshold and in the second case as a sum of a stopped random walk and a stopped slowly changing process. Non-linear renewal theory, which produces an approximation of the overshoot, and an approximation of the stopped slowly changing process can be combined in order to obtain an approximation of the expected delay for the detection rule.

Consider first the index for the Shirayev-Roberts stopping time  $N_{\text{SR}}^w$ . We assess its rate of detection by the examination of the expectation of the stopped process  $r^w(N_{\text{SR}}^w)$ . On the one hand, based on nonlinear renewal theory, one may show that

$$(15) \quad E_k^\theta[r^w(N_{\text{SR}}^w)|N \geq k] = \log A + \log \lambda_{\text{SR}} + v(\mu) + o(1),$$

where  $v(\mu)$  is the expectation of the asymptotic overshoot under  $P_k^\theta$  of the log-likelihood random walk,  $[\langle g_t, X_k(n) \rangle - \mu^2/2]/\sigma^2$ . On the other hand, the same term can be rewritten in the form  $r^w(N_{\text{SR}}^w) = w_k^{\theta,w}(N_{\text{SR}}^w) + \ell_k^\theta(N_{\text{SR}}^w)$ , where  $w_k^{\theta,w}(\cdot) = r^w(\cdot) - \ell_k^\theta(\cdot)$ . The fact that the process  $w_k^{\theta,w}(\cdot)$  is well approximated by the process  $w_k^\theta(\cdot) = r(\cdot) - \ell_k^\theta(\cdot)$  and the fact that the latter is slowly changing can be used in order to produce the approximation:

$$(16) \quad E_k^\theta[r^w(N_{\text{SR}}^w)|N \geq k] = E_k^\theta[\ell_k^\theta(N_{\text{SR}}^w)|N \geq k] \\ + E_k^\theta[w_k^\theta(\tau)|N \geq k] + o(1),$$

for

$$\tau = \tau_k^\theta = \inf\{n \geq k : \ell_k^\theta(n) \geq \log A - (\log A)^\gamma\}.$$

Using what is essentially the Laplace approximation (cf. the Appendix for some details and [12] for a more complete discussion) will produce

$$E_k^\theta[w_k^\theta(\tau)|N \geq k] \approx \\ - \frac{d+1}{2} \log \log A + \frac{d+1}{2} [\log(2\pi\sigma^2) + 1] \\ + \frac{d+1}{2} \log I(\mu) - \frac{1}{2} \log\{\det(\langle \dot{g}_t, \dot{g}'_t \rangle)\} \\ - d \log \mu + w(\mu) + \log \rho(\theta),$$

where  $w(\mu)$  is defined in (3). Combining this approximation with (15) and (16) and canceling some terms result in:

$$(17) \quad E_k^\theta[\ell_k^\theta(N_{\text{SR}}^w)|N \geq k] = \\ \log A + \frac{d+1}{2} \log \log A - \frac{d+1}{2} \log(\pi\sigma^2 e) + \log \lambda_{\text{SR}} \\ - \log \rho(\theta) + v(\mu) - w(\mu) + \frac{1}{2} \log\{\det(\langle \dot{g}_t, \dot{g}'_t \rangle)\} \\ - \log \mu + o(1).$$

The optimality claim from [12], which is quoted in Theorem 2, is a corollary of the above derivation and the fact that the Shirayev-Roberts rule is asymptotically minimax if the asymptotic expansion of the expected delay does not depend on  $\theta$ . This will be the result if one selects  $\rho = \varrho$ , for  $\varrho$  defined in (8). More generally, if we measure loss in terms of worse detection rates we conclude that the reduction in efficiency due to using  $\rho$  as a prior instead of the equalizer prior  $\varrho$  is

$$(18) \quad \log \left[ \frac{\int_{\Theta} \lambda(\eta) \rho(\vartheta) d\vartheta}{\int_{\Theta} \lambda(\eta) \varrho(\vartheta) d\vartheta} \times \max_{\theta \in \Theta} \frac{\varrho(\theta)}{\rho(\theta)} \right] \geq 0.$$

Of course, no asymptotic loss is present if one uses the equalizer prior  $\varrho$ .

Turn to the CUSUM stopping time  $N_{\text{CS}}^w$ . Here we examine  $q^w(N_{\text{CS}}^w) \approx m_k^\theta(\tau) + \ell_k^\theta(N_{\text{CS}}^w)$ , where  $m_k^\theta(\cdot) = q(\cdot) - \ell_k^\theta(\cdot)$ . The

process  $m_k^\theta(\cdot)$  is slowly changing, so the obvious analogue of (15) holds.

The approximation of the expectation of the stopped slowly-changing process is given by

$$\begin{aligned} E_k^\theta[m_k^\theta(\tau)|N \geq k] &\approx \\ &- \frac{d+1}{2} \log \log A + \frac{d+1}{2} [\log(2\pi\sigma^2) + 1] \\ &+ \frac{d+1}{2} \log I(\mu) - \frac{1}{2} \log\{\det(\langle \dot{g}_t, \dot{g}'_t \rangle)\} \\ &- d \log \mu + m(\mu) + \log \rho(\theta), \end{aligned}$$

where

$$(19) \quad m(\mu) = E_1^\theta \left[ 0 \vee \max_{n \geq 1} [-\ell_1^\theta(n)] \vee \max_{k \geq 0} \ell_{-k}^\theta(0) \right].$$

Using this approximation, after some cancelation we get

$$\begin{aligned} E_k^\theta[\ell_k^\theta(N_{CS}^w)|N \geq k] &= \\ \log A + \frac{d+1}{2} \log \log A - \frac{d+1}{2} \log(\pi\sigma^2 e) + \log \lambda_{CS} \\ &- \log \rho(\theta) + v(\mu) - m(\mu) + \frac{1}{2} \log\{\det(\langle \dot{g}_t, \dot{g}'_t \rangle)\} \\ &- \log \mu + o(1). \end{aligned}$$

In order to obtain an equalizer rule we need to modify the prior distribution to

$$\rho(\theta) \propto \varrho(\theta) \exp\{w(\mu) - m(\mu)\},$$

which produces the expected delay:

$$\begin{aligned} (20) \quad E_k^\theta[\ell_k^\theta(N_{CS}^w)|N \geq k] &= \\ \log A + \frac{d+1}{2} \log \log A - \frac{d+1}{2} \log(\pi\sigma^2 e) \\ &+ \log \int_{\Theta} \left[ \frac{I(\eta)\lambda(\eta)}{e^{-w(\eta)+m(\eta)}} \right] \\ &\times \lambda(\eta)\eta^{-1} \sqrt{\det(\langle \dot{g}_s, \dot{g}'_s \rangle)} e^{v(\vartheta)-w(\eta)} d\vartheta + o(1). \end{aligned}$$

Note that the expected Kullback-Leibler index for the optimal Shirayev-Roberts rule is obtained by the replacement of the term in the square brackets within the integral by one. From the representation of  $I(\eta)\lambda(\eta)$  as the expectation of the ratio between the maximum of likelihood ratios and their sum, from the definition of  $w(\eta)$  and  $m(\eta)$  and from Jensen's inequality one can conclude that the term in the brackets is strictly larger than one. Consequently, the Cusum procedure is sub-optimal.

### 3. SIMPLIFIED OPTIMAL POLICIES

Detection procedures that rely on marginal likelihoods face computational difficulties. The evaluation of such likelihoods requires integration over the parameter space  $\Theta$ , which is not a trivial task unless  $d+1$ , the dimension of the

space, is small. As a remedy one may replace the integrals by approximations. One possibility is to use the Laplace approximation of the integral. In this approximation the likelihood is expanded (with respect to  $\theta$ ) about its maximal value to produce a Gaussian kernel. A calculation given heuristically in the Appendix (with details in [3]) shows that for data such that  $\ell_k^\theta(n)$  is maximized at an interior point of  $\Theta$ , the marginal likelihood,  $\ell_k(n)$ , is given approximately by

$$(21) \quad \frac{1}{2} (\langle g_t, Z_k(n) \rangle^+)^2 + \log \left\{ \frac{\rho(\hat{\theta})(2\pi\sigma^2)^{(d+1)/2}}{(n-k+1)^{(d+1)/2} (\hat{\mu}(t))^d \sqrt{\det(\langle \dot{g}_t, \dot{g}'_t \rangle)}} \right\}.$$

The virtue of (21) is that, given the maximum likelihood estimator for  $\theta = (\mu, t)$ , the approximated log likelihood is obtained without integration. Although the approximation cannot be expected to be uniformly accurate, it should be accurate for data that put the statistic near to a (high) stopping boundary, and this is sufficient to make the approximation useful. A disadvantage of (21) is that it ignores the boundary constraints imposed by the compactness of  $\Theta_\varepsilon$ . This need not be a severe issue with respect to the location parameters since the optimal prior,  $\varrho$ , is relatively flat in  $t$ . However, it is a concern with respect to the intensity parameter  $\mu$ .

An alternative is to apply the Laplace approximation for the integration with respect to  $t$  and carry out a one-dimensional numerical integration with respect to  $\mu$ . The resulting semi-Laplace approximation is

$$(22) \quad \frac{\rho_2(\hat{t})(2\pi\sigma^2)^{d/2}}{(n-k+1)^{d/2} \sqrt{\det(\langle \dot{g}_t, \dot{g}'_t \rangle)}} \times \int_{\mu_0}^{\mu_1} e^{\mu(X_k(n), g_t) - (n-k+1)\mu^2/2} \rho(\mu|\hat{t}) \mu^{-d} d\mu,$$

where  $\rho_2(t)$  is the marginal prior density of the location parameter  $t$  and  $\rho(\mu|t)$  is the conditional density of  $\mu$ , given the location. (See Appendix A.)

If we denote either (21) or the log of (22) by  $\hat{\ell}_k(n)$  and use them as substitutes for the log-likelihoods in the formation of monitoring sequences we produce the stopping time  $N_{SR}^{\hat{\ell}_k, w}$  for Shirayev-Roberts-like procedures and  $N_{CS}^{\hat{\ell}_k, w}$  for CUSUM-like procedures. Examination of the derivations given in the previous section reveals that the two stopping times are asymptotically equivalent, in terms of their null distribution and expected delay in detection, to their counterparts that use the actual likelihoods. In particular, the use of equalizer priors will produce rules that are asymptotically minimax.

Although the rules that use approximations are asymptotically equivalent, their finite-sample properties may differ from what the theory suggests. In order to examine this issue

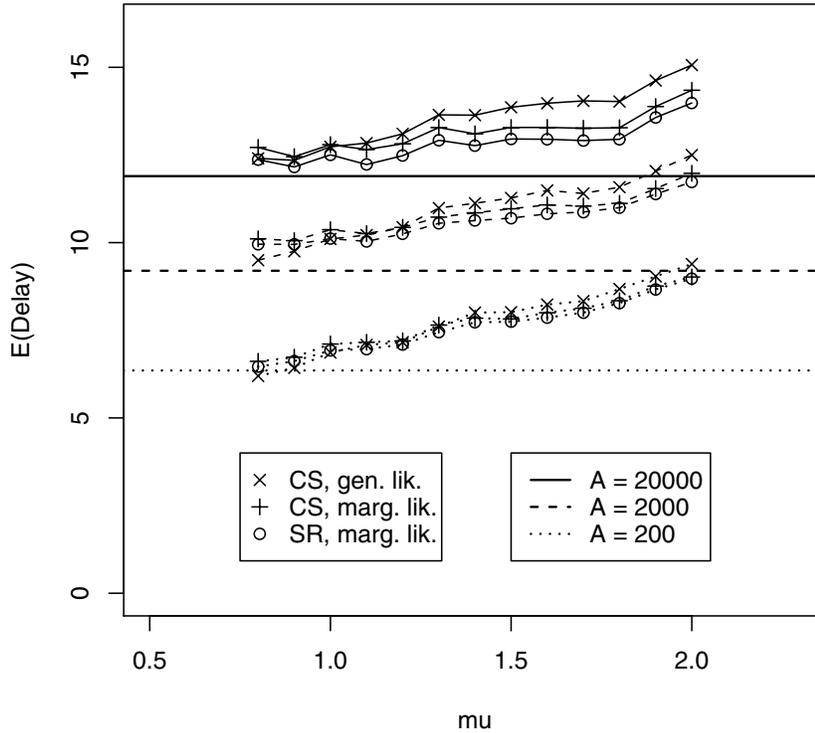


Figure 2. Expected Delays (Kullback-Leibler Information Accumulated Between Change-Point and Detection) Are Plotted as a Function of  $\mu$ . Values of  $A$  Are the Expected Number of Observations Before a False Detection (i.e., No Signal Present). Horizontal Lines Are Asymptotically Minimax Expected Delays.

we have conducted a simulation experiment. In the simulation we considered the example described in Subsection 1.1. Recall that signals there were characterized by a Gaussian kernel:  $g_t(j) \propto \exp\{-50\|j - t\|^2\}$ , for  $j = (j_1, j_2) \in \mathcal{J}$  and location parameter  $t = (t_1, t_2) \in [0, 1]^2$ .

We examined three procedures: (i) The CUSUM based on generalized likelihoods ( $N_{CS}^{\hat{\theta}, w}$ ), (ii) The asymptotic equalizer CUSUM based on approximation (22) of the likelihood ( $N_{CS}^{\hat{\ell}, w}$ ), and (iii) The asymptotic equalizer Shirayev-Roberts with the same semi-Laplace approximation of the likelihood ( $N_{SR}^{\hat{\ell}, w}$ ). Simulations were carried out first under the null distribution  $\mu = 0$  in order to set thresholds that produce the pre-required rate of false alarm. Using these thresholds more simulations were conducted over a collection of intensity values in order to determine the expected delay in detection.

The results of the simulations are presented in Figure 2. Three thresholds were considered:  $E_\infty(N) = 200$  to represent a fairly high rate of false detection,  $E_\infty(N) = 2,000$  to represent medium rate and  $E_\infty(N) = 20,000$  to represent low rate. Horizontal lines are marked at the levels of the theoretical lower bound (6). A dotted line, a broken line and a full line are used for each of the rates of false detection, respectively. The expected total Kullback-Leibler index was computed for  $t = (0.3, 0.5)$ , for  $\mu = 0.8, 0.9, \dots, 1.9, 2.0$  and

for each of the three procedures. The results are plotted as points connected by lines. Each procedure is marked with a different character and the results associated with a given false detection rate fall just above the line that is associated with the lower bound for that rate. The procedures all used a window of 100 observations. The statistics were initiated with observations that were simulated under the null distribution (as described before Theorem 2) and the change took place at  $k = 1$ . The two asymptotic equalizer rules were geared towards values of  $\mu$  in the range  $[0.5, 2.3]$ . In order to save computing effort in determination of the thresholds, we used the fact that the  $P_\infty$  distribution of  $N$  is asymptotically exponential. Instead of simulating the mean of that distribution directly, we obtained an estimate of the mean from an estimate of the cumulative distribution function based on 2,500 iterations. The expected delays at each of the values of  $\mu$  were based on 1,000 iterations. More computational details regarding the simulations are given below.

Looking at the plots one may see that for a high rate of false detection the three procedures perform more or less the same, and the detection delay is not constant in  $\mu$ . As the rate decreases (and  $A$  increases) the curves associated with the two equalizer rules becomes flatter and approach the lower bound. Yet, even for the highest threshold there is still some discrepancy between the worst detection rate and the associated lower bound. Overall, the performance of the

Shiryayev-Roberts procedure is better, but the difference between it and the equalizer CUSUM procedure is negligible. Indeed, under the conditions of the example we get that the asymptotic difference between the two procedures is about 0.302.

The expected detection delay is bounded below by  $\mu^2/2\sigma^2$ , since even the observation at the change-point is counted in our definition. This accounts for part of the observed increase in the empirical results compared to the theoretically flat asymptotic minimax result. It would also be reasonable not to penalize the observation at the change-point, which would flatten the empirical values of the expected detection delay.

All three procedures require the estimation of the location parameter  $t$ . Let  $\bar{X}$  correspond to an image averaged over the interval  $[k, n]$ . Candidate maximizers can be found by solving the normal equations  $\langle \bar{X}, \dot{g}_t \rangle = 0$ . Notice that

$$\dot{g}_t(j) = -2Kg_t(j)[(j-t) - \langle g_t, (\mathcal{J}-t)g_t \rangle], \quad \text{for } j \in \mathcal{J}.$$

In our simulations we used the approximation  $\dot{g}_t(j) \approx -2Kg_t(j)(j-t)$ , which produce the iteration step  $t_{n+1} = \langle \bar{X}, \mathcal{J}g_{t_n} \rangle / \langle \bar{X}, g_{t_n} \rangle$ , where  $\langle \bar{X}, \mathcal{J}g_t \rangle = \sum_{j \in \mathcal{J}} j \bar{X}_j g_t(j)$ , as an approximation of the steepest-descent algorithm. The equation was iterated three times, starting with a initial uniform  $g$ . The outcome may fall outside of the range, especially under the null distribution. If so, we used the maximizer of  $\bar{X}$  as our estimate.

As an alternative to our approach one could use the full-blown steepest descent algorithm, with iterations of the form

$$t_{n+1} = t_n(1 - \langle \bar{X}, g_{t_n} \rangle) + \langle \bar{X}, \mathcal{J}g_{t_n} \rangle,$$

or the Newton-Raphson algorithm

$$t_{n+1} = t_n + \langle \bar{X}, [I + 2K(\mathcal{J} - t_n)(\mathcal{J} - t_n)']g_{t_n} \rangle^{-1} \times \langle \bar{X}, (\mathcal{J} - t_n)g_{t_n} \rangle;$$

and one could propose more sophisticated methods for producing a starting value or for stopping the iterations.

From the definition of the gradient we can compute the metric tensor of the manifold at location  $t$ , which is used in the evaluation of the asymptotic lower bound:

$$\langle \dot{g}_t, \dot{g}_t \rangle = (2K)^2 \left\{ \langle (\mathcal{J}-t)(\mathcal{J}-t)', g_t^2 \rangle - \langle \mathcal{J}-t, g_t^2 \rangle \langle \mathcal{J}-t, g_t^2 \rangle' \right\}.$$

The likelihood substitutes that were used in the formation of the statistics were computed for all combinations of  $k \leq n$ , such that  $n-k \leq w$ . Given an estimate  $\hat{t}$  of the location, for each averaged image, the generalized likelihood ratios for the CUSUM procedure (i) were computed directly from (9). For the equalizer rules (ii) and (iii), numerical integration with respect to  $\mu$  over the range  $[0.5, 2.3]$  was carried out. Observe that the volume element of the manifold cancels out in the approximation for the expected detection delay when the approximation (22) with an equalizer prior

(cf. (8)) is used. The function  $\lambda(\mu)$  was approximated in the integration using a formula that appears in [11]. The quantities  $v(\mu)$ ,  $w(\mu)$  and  $m(\mu)$  were computed by simulations at selected points and interpolated to the entire interval by fitting splines. (Analytic approximations are also possible for these quantities. They are quite accurate for small values of  $\mu$ , but deteriorate for larger values. By way of contrast, simulations are computationally more time consuming for small values, less so for large values.)

## 4. DISCUSSION

The model in this paper behaves like a random walk in the time variable and a smooth Gaussian process in the spatial variable. Thus when a change occurs, information accumulates linearly until the change is detected. In some applications it may be preferable to assume that the parameter  $\mu$ , indicating the size of the change, itself increases, say linearly, and then the information would accumulate quadratically. Another possibility is that the spatial process has rough sample paths, while the information accumulates linearly or quadratically. In principle, these variations can be treated by the methods of this paper and of its companion [12], but the details appear to be substantially more complicated.

In practice the shape of the signal can be quite different from the spherical shape of known size we have discussed. Essentially any shape described by a finite number of parameters, e.g., a spherical shape of unknown size [8] or an elliptical shape [5], is easily accommodated. It would be an interesting problem to develop a flexible set of model shapes that have sufficiently few parameters to be tractable but achieve reasonable power for a general set of actual shapes.

For simplicity we have assumed that observations occur at the points of a rectangular grid. In some applications, e.g., disease surveillance, monitoring locations are likely to be quite irregularly placed. The methods of this paper are easily generalized, although operating characteristics will in general depend on the monitoring locations and their relation to the location of a signal.

If the volume of the  $(d-1)$ -dimensional boundary of  $T$  is large relative to the volume of the interior, it may be desirable to consider the asymptotically second order effects that arise when  $\hat{t}$  lies near this boundary. Since the second order term in an asymptotic expansion for tail probabilities of maxima of smooth Gaussian fields is typically of order  $[\log(A)]^{-1/2}$  times the first order term, whereas the rate of decrease is  $[\log(A)]^{-1}$  for second order effects on the time scale, it is plausible that a second order approximation can be obtained reasonably easily for the models considered in this paper. A similar remark applies to a skewness correction for non-normality (cf. [3]).

## APPENDIX A. HEURISTIC LAPLACE APPROXIMATION TO OBTAIN (21) AND (22)

Let  $\ell_k^\theta(n) = (n - k + 1)[\mu \langle \bar{X}_k(n), g_t \rangle - \mu^2/2]/\sigma^2$  be the log-likelihood ratio for  $n - k + 1$  observations, which we consider as a function of  $\theta = (\mu, t)$ . Independently of  $\mu$ , the MLE with respect to  $t$  is found by solving the equation  $\langle \bar{X}_k(n), \dot{g}_t \rangle = 0$  and  $\hat{\mu} = \langle \bar{X}_k(n), g_t \rangle$ . The gradient of  $\ell^\theta(n)$  vanishes at  $\hat{\theta} = (\hat{\mu}, \hat{t})$  and the Hessian satisfies

$$\begin{aligned} \ddot{\ell}_k^\theta(n) &= (n - k + 1) \begin{pmatrix} -1 & 0 \\ 0 & \hat{\mu} \langle \bar{X}_k(n), \ddot{g}_t \rangle \end{pmatrix} \\ &\approx -(n - k + 1) \begin{pmatrix} 1 & 0 \\ 0 & \hat{\mu}^2 \langle \dot{g}_t, \dot{g}_t' \rangle \end{pmatrix}. \end{aligned}$$

A Taylor expansion of the log-likelihood about its maximizer produces

$$\ell^\theta(n) \approx \ell_k^{\hat{\theta}}(n) - \frac{1}{2}(\theta - \hat{\theta})' [\ddot{\ell}_k^{\hat{\theta}}(n)] (\theta - \hat{\theta}).$$

Substituting the approximation for the actual log likelihood and substituting for the function  $\rho(\theta)$  the constant  $\rho(\hat{\theta})$  lead to the Laplace approximation we have in (21), since

$$\begin{aligned} \det(\ddot{\ell}_k^{\hat{\theta}}(n)) &= (n - k + 1)^{d+1} \det(\langle \bar{X}_k(n), -\ddot{g}_t \rangle) \\ &\approx (n - k + 1)^{d+1} \hat{\mu}^{2d} \det(\langle \dot{g}_t, \dot{g}_t' \rangle). \end{aligned}$$

This approximation should produce satisfactory results whenever  $(n - k + 1)$  is large,  $\hat{\theta}$  belongs to the interior of  $\Theta$ , and the variability of  $\rho$  can be ignored.

Because of the restrictions placed on  $\mu$  in the definition of  $\Theta$ , the assumptions underlying the approximation (21) may not be satisfied. This problem can be mitigated by using (22), which involves a Laplace approximation only for integration over the location parameters. To that end we represent the log-likelihood in the form

$$\begin{aligned} \ell_k^\theta(n) &= (n - k + 1) \mu \langle \bar{X}_k(n), g_t - g_{\hat{t}} \rangle \\ &\quad + (n - k + 1) [\mu \langle \bar{X}_k(n), g_{\hat{t}} \rangle - \mu^2/2] \\ &\approx \frac{(n - k + 1) \mu}{2} (t - \hat{t}) \langle \bar{X}_k(n), \ddot{g}_{\hat{t}} \rangle (t - \hat{t}) + \ell_k^{(\mu, \hat{t})}(n) \\ &\approx -\frac{(n - k + 1) \mu^2}{2} (t - \hat{t}) \langle \dot{g}_{\hat{t}}, \dot{g}_{\hat{t}}' \rangle (t - \hat{t}) + \ell_k^{(\mu, \hat{t})}(n). \end{aligned}$$

Received 9 January 2008

## REFERENCES

- [1] KULLDORFF, M. (1997). A spatial span statistic. *Communications in Statistics: Theory and Methods* **26** 1481–1496. [MR1456844](#)
- [2] LAI, T. L. AND SIEGMUND, D. (1977). A non-linear renewal theory with applications to sequential analysis I. *Ann. Statist.* **5** 946–954. [MR0445599](#)
- [3] NARDI, Y., SIEGMUND, D. AND YAKIR, B. (2008). The distribution of maxima of approximately Gaussian random fields. *Ann. Statist.* **36**, in press.
- [4] RABINOWITZ, D. (1994). Detecting clusters in disease incidence. *Change-point Problems*, E. Carlstein, H.-G. Müller, D. Siegmund, eds., IMS, Hayward, CA, 255–275. [MR1477929](#)
- [5] SHAFIE, K., SIGAL, B., SIEGMUND, D. AND WORSLEY, K. (2003). Rotation space random fields with an application to fMRI data. *Ann. Statist.* **31** 1732–1771. [MR2036389](#)
- [6] SHEWHART, W. A. (1931). *The Economic Control of the Quality of Manufactured Product*. Macmillan, New York.
- [7] SIEGMUND, D. AND VENKATRAMAN, E. S. (1995). Using the generalized likelihood ratio statistic for sequential detection of a change-point. *Ann. Statist.* **23** 255–271. [MR1331667](#)
- [8] SIEGMUND, D. AND WORSLEY, K. (1995). Testing for a signal with unknown location and scale in a stationary Gaussian random field. *Ann. Statist.* **23** 608–639. [MR1332585](#)
- [9] SIEGMUND, D. AND YAKIR, B. (2000). Tail probabilities for the null distribution of scanning statistics. *Bernoulli* **6** 191–213. [MR1748719](#)
- [10] SIEGMUND, D. AND YAKIR, B. (2001). Approximate p-values for local sequence alignments. *Ann. Statist.* **28** 657–680. [MR1792782](#)
- [11] SIEGMUND, D. AND YAKIR, B. (2007). *The Statistics of Gene Mapping*. Springer, New York. [MR2301277](#)
- [12] SIEGMUND, D. AND YAKIR, B. (2008). Minimax optimality of the Shiryayev-Roberts change-point detection rule. *J. Statist. Planning and Inference*, to appear.
- [13] WORSLEY, K., EVANS, A., MARRETT, S. AND NEELIN, P. (1992). A three dimensional statistical analysis for cbf activation studies in human brain. *Journal of Cerebral Blood Flow and Metabolism* **12** 900–918.
- [14] YAKIR, B. AND POLLAK, M. (1998). A new representation for a renewal-theoretic constant appearing in asymptotic approximations of large deviations. *Ann. Appl. Probab.* **8** 749–774. [MR1627779](#)

David Siegmund

Department of Statistics

Stanford University

Stanford, CA 94305, USA

E-mail address: [dos@stat.stanford.edu](mailto:dos@stat.stanford.edu)

Benjamin Yakir

Department of Statistics

The Hebrew University

Jerusalem, 91905, Israel

E-mail address: [msby@mssc.huji.ac.il](mailto:msby@mssc.huji.ac.il)