Evaluation of univariate surveillance procedures for some multivariate problems

Peter Wessman

Evaluation of univariate surveillance procedures for some multivariate problems.

Peter Wessman
Department of Statistics, Göteborg University.
S - 411 80 Göteborg.

ABSTRACT

The continual surveillance to detect changes has so far received large attention in the area of industrial quality control, where the monitoring of manufacturing processes to detect decreases in quality play an important role. However, also in other areas important examples can be found such as the surveillance of intensity care patients or the monitoring of economic trends.

Often more than one measurement is made, resulting in a multivariate observation process. Many surveillance procedures for multivariate observation processes are based on summarizing statistics that reduces the multivariate process to a univariate process. This thesis studies such surveillance procedures when a change to a specific alternative is of interest. We give special attention to procedures based on likelihood ratio statistics of the observation vectors since these are known to have several optimality properties. Also, many procedures in use today can be formulated in terms of likelihood ratios.

In report **I** we consider the surveillance of a multivariate process with a common change point for all component processes. We show that the univariate reduction using the likelihood ratio statistic for the observation vector from each observation time is sufficient for detecting the change. Furthermore, the use of a likelihood ratio-based method, the LR method, for constructing surveillance procedures is suggested for multivariate surveillance situations. The LR procedure, as several other multivariate surveillance procedures, can be formulated as univariate procedures based on the univariate process of likelihood ratios. Thus, evaluating these multivariate surveillance procedures which are based on this reduction can be done by using results for univariate procedures, for example those given in report **II**. The effects of not using a sufficient univariate statistic is also

illustrated.

In the second report a simulation study of some methods based on likelihood ratios of univariate processes is made. The LR method and the Roberts procedure are compared with two methods that today are in common use, the Shewhart and the CUSUM methods. Several different measurements of performance are used, such as the probability of successful detection, the predictive value and the expected delay of an alarm. The evaluation is made for geometrically distributed change points. For this situation the LR procedure meets several optimality criteria and is therefore suitable as a benchmark. The LR procedure is shown to be robust against misspecifications of the intensities. The CUSUM method appears in the simulations to be closer to the Shewhart method than to the Roberts method in several of the properties investigated, for example the run length distribution and the predictive value. Furthermore, the Roberts procedure is shown to have properties close to the LR procedure for moderately large intensities. It has therefore near optimal properties in these cases.

This thesis consists of two parts:

I Some principles for surveillance adopted for multivariate processes with a common change point.

II Evaluations of likelihood ratio methods for surveillance.

Some principles for surveillance adopted for multivariate processes with a common change point.

Peter Wessman

Department of Statistics, University of Göteborg,

S - 411 80 Göteborg.

Abstract

The surveillance of multivariate processes has received growing attention during the last decade. Several generalizations of well-known methods such as Shewhart, CUSUM and EWMA charts have been proposed. Many of these multivariate procedures are based on a univariate summarized statistic of the multivariate observations, usually the likelihood ratio statistic. In this paper we consider the surveillance of multivariate observation processes for a shift between two fully specified alternatives. The effect of the dimension reduction using likelihood ratio statistics are discussed in the context of sufficiency properties. Also, an example of the loss of efficiency when not using the univariate sufficient statistic is given. Furthermore, a likelihood ratio method, the LR method, for constructing surveillance procedures is suggested for multivariate surveillance situations. It is shown to produce univariate surveillance procedures based on the sufficient likelihood ratios. As the LR procedure has several optimality properties in the univariate, it is also used here as a benchmark for comparisons between multivariate surveillance procedures.

Key words: Multivariate surveillance, sufficiency, Likelihood ratio, CUSUM.

1

1 Introduction

The surveillance of random processes to detect a change has received much attention during the last decades. Contributions have been made especially in industrial quality control, but also in other areas such as medicine, epidemiology or economy. The term surveillance is mainly used in the medical and epidemiological fields. Other names for this type of problem are monitoring, change point detection or statistical process control.

In surveillance the process is continually observed through a sequence of observations made continuously or at specific intervals. These observations can be either the original measurements or a function of these, for example the mean of a sample from each time point. Whenever the process is observed through more than one observation sequence we have a multivariate surveillance situation. Consider for example the monitoring of a manufacturing process to detect decreases in the product quality. In this situation several measurements are often taken simultaneously. For example, the quality of a product can be defined through several different attributes or there may be parallel lines of production monitored independently. In other cases the process is observed at different stages in the production.

When the surveillance starts the process is considered in control with observations from a known and acceptable distribution. At some unknown random time point a change occurs in the process, resulting in a change in distribution of the observations. This change usually consists of a change in the parameter vector either to a specific new point or a general shift away from the in-control parameter vector. The observation sequence is usually assumed to be independent, both before and after the change, and to have a well known distribution family such as the Gaussian family.

Of primary interest in surveillance is the random change point where a change occurs. Based on the observation process a decision must be made after each time point whether or not an event has occurred. To do this, an alarm-procedure of some sort is used. Examples of well-known procedures for univariate observation processes include the Shewhart, CUSUM and EWMA procedures, see for example Whetherill and Brown (1991).

In multivariate surveillance situations the change in the underlying process, depending on the way it is measured, can affect the observations differently. For example, if the quality in the example above depends only on the

raw materials used, then the change points in the different lines are likely to occur simultaneously in the case when parallel production lines are monitored. If instead the quality is measured at different stages of the production line then the change points for the different observation sequences differ, depending on what stage they are taken.

Multivariate surveillance has received an increased interest during the last decade with many new alarm procedures suggested. There are several situations that have earlier been treated as univariate ones but where a multivariate perspective might be useful. For examples in the post marketing surveillance of medical drugs, Sveréus (1995), where normally several different adverse events are monitored simultaneously or in the monitoring of economic trends, Frisén (1994), where several economic indexes are used. Another area where multivariate surveillance approach may be of use is in the surveillance of ecological systems, Pettersson (1996).

The alarm procedures for surveilling multivariate observation processes can be divided into those using some univariate procedure based on a summarizing statistic of some kind and others, for example those where the component processes are monitored separately. In this paper we discuss some properties, such as the sufficiency property, and compare the two types of procedures. We will also suggest a type of likelihood ratio-based surveillance procedure shown to have certain optimality properties. We restrict our attention to processes where a sudden shift occurs between two fully specified alternatives. A consequence of this limitation is that only situations where the change occurs simultaneously in all components processes are considered.

In Section 2 we will give a more formal definition of the multivariate surveillance problem with some limitations and assumptions made in this paper. We shall also mention shortly some useful measurements of performance. Section 3 contains a short review of the literature on multivariate surveillance and a description of some types of alarm procedures. In Section 4 we consider the effect of using a summarizing statistic to reduce the observation-vector to a univariate observation sequence. We discuss the likelihood ratio-based multivariate surveillance procedures in Section 5. A summary and some conclusions are finally given in Section 6.

3

2 Preliminaries

2.1 Specifications and Notation

In multivariate surveillance the process of interest is observed through a p-dimensional vector

$$\mathbf{X}(t) = \begin{pmatrix} X_1(t) \\ \vdots \\ X_p(t) \end{pmatrix}, t = 1, 2, \ldots \tag{1}$$

of observations. These observations, either the original measurement or transformations of these, are considered independent, with some distribution $F(t)$. At unknown random time points, $\tau_1, \ldots, \tau_p$, a change occurs in the different component processes, thus in multivariate surveillance the random change point is a random vector

$$\tau = \begin{pmatrix} \tau_1 \\ \vdots \\ \tau_p \end{pmatrix} \tag{2}$$

of change points. Depending on the application of interest the structure of this vector assumes different forms. The change in the underlying process can affect the component change points simultaneously, with deterministic delays for some components, or stochastically, where the change point of each component has some distribution $G_{\tau_i}, i = 1, \ldots, p$. We shall here consider only the special case with one simultaneous change point, $\tau_i = \tau, i = 1, \ldots, p$, for all component processes following some distribution G_τ. In some sections, G_τ is specified to be the geometric distribution with intensity ν. This is the most commonly used assumption in the theoretical literature and it is also a reasonable assumption in many practical applications. Some useful results for this case can be found in for example Shiryaev (1963).

In addition to a common change point for component processes we also assume that there exist two fully specified alternatives. Thus, we consider an observation process X with distribution F^0 for all $\mathbf{X}(1), \mathbf{X}(2), \ldots, \mathbf{X}(\tau - 1)$ and F^1 for $\mathbf{X}(\tau), \mathbf{X}(\tau + 1), \ldots$.

As mentioned above we are primarily interested in a sequence of critical events concerning τ. Usually we are interested in detecting a change whenever it has occurred. However, sometimes only some of the possible change points of interest at each time s are of interest, for example we might only be interested in detecting a change that has occurred in the latest d observation points. We therefore define two events in the sample space of

4

τ, $C(s) = \{\tau \in I_s \subseteq \{1, \ldots, s\}\}$, the set of time points where we, at time s, want to detect a change, and $D(s) = \{\tau > s\}$, where no change has yet occurred at time s. It should be noted that $D(s)$ is not always the complement event of $C(s)$. In fact, only if $C(s) = \{\tau \leq s\}$, that is we are interested in detecting a change occurring at any possible change point up to time s, is $D(s)$ the complement event.

A surveillance procedure can be defined through a stopping time t_A, which determines when an alarm should be given. We can define the stopping time through an alarm function $p(\cdot; C(s))$ and a critical limit $K(s)$ as

$$t_A = \min[s; p(\mathbf{x}_s; C(s)) > K(s)]. \tag{3}$$

or through the alarm set

$$A(s) = \{\mathbf{x}_s \in \Omega_{\mathbf{X}_s} | p(\mathbf{x}_s; C(s)) > K(s)\} \tag{4}$$

of outcomes leading to an alarm. The critical limit $K(s)$ regulates the probabilities of making a (false) alarm and the alarm function $p(\cdot)$ is a function of the available information at time s, $X_s = \{\mathbf{X}(1), \ldots, \mathbf{X}(s)\} \in \Omega_{X_s}$, (and possibly also of G_τ). Note the different use of the index in X_i and $X_i(t)$, the former indicating a truncated sequence and the latter the observations made at time t of component process i. A simple example of an alarm function is the alarm function for the Shewhart chart $p(\mathbf{x}_s; C(s)) = \mathbf{x}(s)$. Using only the latest observation, the Shewhart procedure is designed to detect the particular critical event $C(s) = \{\tau = s\}$.

2.2 Measurements of Performance

The performance of a procedure can be studied through the relationship between the change point τ and the stopping time t_A. To be able to construct, or choose between, surveillance procedures some measure of performance or criterion of optimality is necessary.

Two important properties for a surveillance procedure are: $i)$ that the procedure should have a high probability to detect a change and do so within reasonable time and $ii)$ the procedure should have a low probability for a false alarm. These two properties are unfortunately in conflict with each other, as for example a lower false alarm probability will usually give a lower probability of detecting a change. The desirable properties differ between applications and some sort of measurement of the performance of a procedure is therefore useful.

5

One measurement of performance dealing with the balance between the two properties i) and ii) is the Predictive Value, Frisén (1992). This is defined as

$$PV(t) = \Pr(\tau < t \,|\, t_A = t) = \frac{\Pr(t_A = t, \tau \leq t)}{\Pr(t_A = t, \tau \leq t) + \Pr(t_A = t, \tau > t)}, \quad (5)$$

$t = 1, 2, \ldots$, and measures the balance between motivated and unmotivated alarms, giving a measure of the confidence we can have in an alarm given at a certain time. Other measurements are concerned with just one of these two properties. For example, the average run length, ARL, where $ARL^0 := E\left[t_A \,|\, \tau = \infty\right]$, the ARL when no change occurs, deals with property ii) and $ARL^1 = E\left[t_A \,|\, \tau = 1\right]$ deals with property i).

In quality control the most commonly used measure of performance is the ARL where a procedure is sometimes considered optimal if it minimizes the ARL^1 for fixed ARL^0. The limitations of this optimality criterion have been discussed by Åkermo, (1995) and Frisén (1996).

Another way of defining an optimal procedure is by:

Definition 2.1 If $A \subseteq \Omega_{\mathbf{X}_s}$ is an alarm set, C a critical event concerning τ and D a proper subset of the complement of C, then A is the optimal alarm set if for all $B \subseteq \Omega_{\mathbf{X}_s}$;

$$\Pr(B \,|\, C) > \Pr(A \,|\, C) \Rightarrow \Pr(B \,|\, D) > \Pr(A \,|\, D) \quad (6)$$

Thus the procedure is deemed optimal if any other surveillance procedure with higher probability to detect a change also has a higher probability of giving a false alarm.

Shirayev (1963) defined optimality, S−optimality, using the speed of detection and the false alarm rate. He defined the optimal alarm procedure as the one with minimal expected cost, for a specific cost function of the delay of an alarm and a false alarm.

3 A Short Review of Literature and Methods

During the last decade several new procedures for surveilling multivariate observation processes have been suggested. Most of them are generalizations of

procedures developed for monitoring univariate observation processes. However, a common approach is still to monitor each component separately, usually using a union intersection type of procedure, Hochberg & Tamhane (1987), to decide on an alarm. A simple example is the union intersection Shewhart procedure, (UI Shewhart), where an alarm is made whenever one component exceeds its predefined limit.

The first one to consider the surveillance of several observation processes as a multivariate problem seems to have been Hotelling (1947). He suggested the monitoring of a shift in the mean of a multivariate Gaussian process using a Shewhart procedure based on a summarizing statistic, the so-called $T^2-statistic$:

$$T^2(t) = \left(\mathbf{x}(s) - \mu^0\right)' \Sigma^{-1} \left(\mathbf{x}(s) - \mu^0\right), t = 1, 2, \ldots \qquad (7)$$

where μ^0 is the *in-control* mean and Σ is the covariance matrix. When Σ is known the procedure is called the Shewhart χ^2-chart. Further development of procedures based on the Shewhart chart has been done, for example using principal components, Alt (1985).

Several different generalizations of the CUSUM procedure based on the T^2- statistic have also been suggested, see for example Pigniatello & Runger (1990), Alwan (1986) and Crosier (1988). These are, as the Shewhart χ^2 procedure, directionally invariant and constructed to detect shifts in the mean vector in any direction.

In recent years also generalizations of exponential weighted moving average procedures, EWMA, have been made. For example Lowry et al. (1992) proposed the use of $\mathbf{Z}(t) = R\mathbf{X}(t) + (I - R)\mathbf{Z}(t-1), t = 1, 2, \ldots$, where $\mathbf{Z}(0) = \mathbf{0}$ and $R = diag(r_1, \ldots, r_p)$ and to signal an alarm when

$$\min\left\{t; \frac{2-r}{r}\mathbf{Z}(t)' \Sigma^{-1}\mathbf{Z}(t) > K^e\right\}. \qquad (8)$$

Thus, also this procedure is based on the T^2 statistic and directional invariant for changes in the mean vector.

Woodal and Ncube (1985) used the principal components of the observations when they suggested a union intersection type of procedure based on univariate CUSUM charts, one for each independent principal component. The stopping time for the CUSUM procedure they used is defined as

$$\min\{t; \max(0, S_i(t-1) + X_i(t) - k_i) > h_i\}, i = 1, \ldots, p, \qquad (9)$$

7

with $0 \leq S_i(0) < h_i$. The reference value, $k_i \geq 0$ and the critical limit h_i are chosen for each component process so that the change can be detected quickly.

Other examples of procedures to monitor changes to a specific alternative using summarizing statistics are more rare. One example is the procedure suggested by Healy (1987) which will be discussed later in section 4. Hawkins (1991 and 1993), considering monitoring of regression adjusted variables, has extended Healy's work to the surveillance of a set of specified alternatives.

4 Summarizing Statistics in Surveillance Procedures of Multivariate Processes

4.1 Sufficiency

Many of the suggested surveillance procedures proposed so far for monitoring multivariate processes are based the univariate likelihood ratios of the observation vector from each observation time point. This raises the question of possible information loss caused by using the likelihood ratio statistic to reduce the dimension of the multivariate process. It is therefore of interest to see if the use of the likelihood ratio statistic is sufficient for detecting the change. Following Cox & Hinkley (1974) we define sufficiency as:

Definition 4.1 A statistic T is *sufficient* for a family of distributions $\mathcal{F}$ if and only if the conditional density $f_{X|T}(x\,|t)$ is the same for all $F \in \mathcal{F}$. In addition, a sufficient statistic is said to be *minimal sufficient* if no sufficient statistic of lower dimension can be found.

Since we are here interested in sequential decisions a definition of *sufficiency of a sequence of statistics* is also useful. Following Arnold (1988) we define this as:

Definition 4.2 A sequence $T_1(X_1), T_2(X_2), \ldots$ of statistics is a *sufficient sequence of statistics* for the families $\mathcal{F}_1, \mathcal{F}_2, \ldots$ of distributions if for all s, $T_s(X_s)$ is a sufficient statistic for the family $\mathcal{F}_s$.

Consider as before a surveillance situation where we monitor a process through a sequence, $X = \{X(t)\,; t = 1, 2, \ldots\}$, of independent observations for which a change in the process introduces the shift in the distribution:

$$\mathbf{X}(t) \sim F(t) = \begin{cases} F^0 & \tau > t \\ F^1 & \tau \leq t \end{cases}, \tag{10}$$

8

where F^0 and F^1 are two completely specified distributions. Notice that when considering this type of change we also assume a simultaneous change point τ for the component processes.

The available information $X_s = (\mathbf{X}(1), \ldots, \mathbf{X}(s))$ for a decision between the events $C(s) = \{\tau \leq s\}$ and $D(s) = \{\tau > s\}$ has, at time s, a distribution from the family:

$$X_s \sim \mathcal{F}_s = \left\{ \begin{array}{ll} \mathcal{F}^{D(s)} = \{F^0\}^s & \tau > s \\ \mathcal{F}^{C(s)} = \left\{ \bigcup_{\tau < s} (F^0)^{\tau-1} \times (F^1)^{\tau-s+1} \right\} & \tau \leq s \end{array} \right. \qquad (11)$$

Thus the decision whether or not a change has occurred , based on the observations sequence $\{\mathbf{X}(t) ; t = 1, 2, \ldots\}$, can be formulated as a decision between whether X_s has a distribution from the family $\mathcal{F}^{D(s)}$ or $\mathcal{F}^{C(s)}$. We therefore want to find a statistic $T_s(X_s)$ that is sufficient for the family $\left\{\mathcal{F}^{C(s)}, \mathcal{F}^{D(s)}\right\}$ for each s, thus making the sequence $\{T_s(X_s) ; s = 1, 2, \ldots\}$ a sufficient sequence of statistics for the sequence $\left\{\mathcal{F}^{C(s)}, \mathcal{F}^{D(s)}; s = 1, 2, \ldots\right\}$. We prove the following statement.

Statement 4.1 For the model above with $\tau_i = \tau$, $i = 1, \ldots, p$ the univariate process $\{lr(\mathbf{x}(t)) ; t = 1, 2, \ldots\}$ of likelihood ratios, as defined below, is a sufficient reduction of the multivariate observation process for the sequence of families $\left\{\mathcal{F}^{C(s)}, \mathcal{F}^{D(s)}; s = 1, 2, \ldots\right\}$.

Proof: Let us first define $lr(\mathbf{x}(t)) = f^1(\mathbf{x}(t)) / f^0(\mathbf{x}(t))$ and $\mathbf{lr}_t(x_s) = \{lr(\mathbf{x}(t)), \ldots, lr(\mathbf{x}(s))\}$.

(i) Let τ be fixed. Then, at time s, we can write the density of X_s as:

$$f_s(x_s \,|\, \tau = t) = \prod_{i=1}^{t-1} f^0(\mathbf{x}(i)) \prod_{i=t}^{s} f^1(\mathbf{x}(i)) = \prod_{i=1}^{s} f^0(\mathbf{x}(i)) \prod_{i=t}^{s} \frac{f^1(\mathbf{x}(i))}{f^0(\mathbf{x}(i))}$$

$$= h(x_s) \prod_{i=t}^{s} lr(\mathbf{x}(i)) = h(x_s) \, k(\mathbf{lr}_t(x_s)), \qquad (12)$$

where h and k are two real valued functions. We also use the definition $\prod_{i=s+1}^{s} = \prod_{i=1}^{0} = 1$. The factorization theorem gives that the vector $\mathbf{lr}_1(x_s)$ is sufficient for the distribution family of X_s defined by the parameter τ.

(ii) If τ is stochastic with some distribution G_τ then the density of X_s at time s can be written:

$$f_s(x_s) = \sum_{t=1}^{\infty} g_\tau(t) f_s(x_s \,|\, \tau = t)$$

9

$$= h\left(x_s\right)\left[\sum_{t=1}^{s} g_\tau\left(t\right) k\left(\mathbf{lr}_t\left(x_s\right)\right) + \left(1 - G_\tau\left(t\right)\right) k\left(\mathbf{lr}_1\left(x_s\right)\right)\right] \qquad (13)$$

and again using the factorization theorem we have that $\mathbf{lr}_1\left(x_s\right)$ is sufficient for the family $\left\{\mathcal{F}^{C(s)}, \mathcal{F}^{D(s)}\right\}$.

(iii) Finally, since $\mathbf{lr}_1\left(x_s\right)$ is sufficient for any s, we have according to definition 4.2 that the sequence $\left\{\mathbf{lr}_1\left(x_s\right); s = 1, 2, \ldots\right\}$ is a sufficient sequence for the sequence of families $\left\{\mathcal{F}^{C(s)}, \mathcal{F}^{D(s)}; s = 1, 2, \ldots\right\}$. Thus to monitor for a simultaneous, fully specified, shift in a multivariate observation process it is possible to construct a univariate surveillance procedure based on the sufficient sequence of likelihood ratios. $\square$

For the surveillance of a univariate observation process there exists in general no sufficient statistic of a lower dimension than the sample itself. Thus, a procedure based on $\left\{\mathbf{lr}_1\left(x_s\right); s = 1, 2, \ldots\right\} = \left\{lr\left(\mathbf{x}\left(t\right)\right); t = 1, 2, \ldots\right\}$ is in these cases also minimal sufficient. For example in the case with observations of a univarite Gaussian distribution there is no sufficient statistic with a lower dimension than the sample itself, see Cox and Hinkley (1974, p.30).

4.1.1 Sufficient Reduction for Observation from a $k-$Dimensional Exponential Family

Most procedures proposed so far in literature have been constructed for observations from the exponential family. We consider observations from such a $k-$dimensional exponential family where we are monitoring for a shift in the parameter vector between two (natural) parameter vectors θ^0 and θ^1:

$$\theta\left(t\right) = \begin{cases} \left(\theta_1^0, \ldots, \theta_k^0\right) & \tau \geq t \\ \left(\theta_1^1, \ldots, \theta_k^1\right) & \tau < t \end{cases}. \qquad (14)$$

According to statement 4.1 we have that it is enough to monitor the univariate process of likelihood ratio statistics for observations from the exponential family;

$$lr\left(t\right) = \sum_{j=1}^{k}\left(\theta_j^1 - \theta_j^0\right) T_j\left(\mathbf{x}\left(t\right)\right) = \sum_{j:\theta_j^1 \neq \theta_j^0}\left(\theta_j^1 - \theta_j^0\right) T_j\left(\mathbf{x}\left(t\right)\right) \qquad (15)$$

where $T_j\left(\mathbf{x}\left(t\right)\right)$ is the minimal natural sufficient statistic for θ_j.

From this follows that when surveilling a process of multivariate Gaussian distributed observations for a sudden shift in the mean vector

$$\mu\left(t\right) = \begin{pmatrix} \mu_1\left(t\right) \\ \vdots \\ \mu_p\left(t\right) \end{pmatrix} = \begin{cases} \mu^0 = \mathbf{0} & \tau > t \\ \mu^1 = \mu & \tau \leq t \end{cases} \qquad (16)$$

with a constant covariance matrix Σ, then it is sufficient to monitor the sequence of likelihood ratio statistics $T^\mu\left(\mathbf{x}\left(t\right)\right) = \left\{\mu'\Sigma^{-1}\mathbf{x}\left(t\right); t = 1, 2, \ldots\right\}$. Thus, we have that the statistic is simply a weighted sum of the observations. In the case with standardized bivariate observations with a correlation ρ where unit shifts are of interest this statistic reduces to

$$T^\mu\left(\mathbf{x}\left(t\right)\right) = \frac{X_1\left(t\right) + X_2\left(t\right)}{1 + \rho} \qquad (17)$$

If instead a shift in the covariance is monitored:

$$\mathbf{X}\left(t\right) \sim \left\{ \begin{array}{ll} N_p\left(\mu, \Sigma_0\right) & \tau > t \\ N_p\left(\mu, \Sigma_1\right) & \tau \leq t \end{array} \right. \qquad (18)$$

the minimal sufficient process to monitor would be

$$T^\Sigma\left(\mathbf{x}\left(t\right)\right) = \left\{\left(\mathbf{x}\left(t\right) - \mu\right)'\left(\Sigma_0^{-1} - \Sigma_1^{-1}\right)\left(\mathbf{x}\left(t\right) - \mu\right)\right\}_t \qquad (19)$$

if the matrix $\left(\Sigma_0^{-1} - \Sigma_1^{-1}\right)$ is not singular. Notice for example that if $\Sigma_1 = c\Sigma$, c a scalar, then the sufficient statistic to monitor is

$$T^\Sigma\left(\mathbf{x}\left(t\right)\right) = \left(\left(c - 1\right)/c\right)\left(\mathbf{x}\left(t\right) - \mu\right)'\Sigma_0^{-1}\left(\mathbf{x}\left(t\right) - \mu\right).$$

Thus, we have that procedures based on monitoring the Hotelling T^2-statistic are sufficient for surveilling a proportionate shift in the covariance matrix of the observation process.

4.2 Evaluation of Efficiency

In the previous subsection we saw that $\left\{\mathbf{lr}_1\left(x_s\right); s = 1, 2, \ldots\right\}$ is sufficient for detecting $C\left(s\right) = \left\{\tau \leq s\right\}$. In this subsection we shall give an example of the loss of efficiency that can occur with a surveillance procedure that is not based on a sufficient reduction. We will compare two procedures based on the Shewhart chart, a univariate Shewhart procedure on the likelihood ratio statistic and a UI-Shewhart procedure based on the principal components of the observations.

Several authors have suggested the use of principal components of the observations, e.g. Jackson & Bradley (1966), Woodall & Ncube (1985) and Hawkins (1991). The reason for using principal components is usually to reduce the dimension but it is also used as a mean to transform the observation vector in to a vector of independent components. We base our parallel Shewhart chart on the principal component process instead of the original data as a mean to simplify calculations.

11

As we consider Shewhart type procedures we restrict our evaluation to the use of average run length as a measure of performance. We define efficiency in terms of minimal the ARL^1 for fixed ARL^0.

Let us observe a bivariate Gaussian sequence for a sudden shift in the mean with the covariance structure remaining unchanged. We can then without loss of generality consider the standardized process

$$\mathbf{X}(t) \sim \begin{cases} N_2\left(\mathbf{0}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right) & \tau > t \\ N_2\left(\mu, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right) & \tau \leq t \end{cases}, t = 1, 2, \ldots \qquad (20)$$

with $\rho \geq 0$, for negative correlation we instead observe $(X_1(t), -X_2(t))^T$. Furthermore we here consider only equal sized shifts, $\mu = (\mu, \mu)'$. Then the principal components

$$\mathbf{Y}(t) = \frac{1}{\sqrt{2}} A \mathbf{X}(t) = \frac{1}{\sqrt{2}} \begin{pmatrix} X_1(t) + X_2(t) \\ X_1(t) - X_2(t) \end{pmatrix} \qquad (21)$$

are distributed

$$\mathbf{Y}(t) \sim \begin{cases} F^0 = N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1+\rho & 0 \\ 0 & 1-\rho \end{pmatrix}\right) & \tau > t \\ F^1 = N_2\left(\sqrt{2}\begin{pmatrix} \mu \\ 0 \end{pmatrix}, \begin{pmatrix} 1+\rho & 0 \\ 0 & 1-\rho \end{pmatrix}\right) & \tau \leq t \end{cases}. \qquad (22)$$

Notice that in this case all information concerning the shift is found in the first principal component. The normalized $lr - statistic$ is in this case

$$Z(t) = \frac{\mu'\Sigma^{-1}\mathbf{X}(t)}{\sqrt{\mu'\Sigma^{-1}\mu}}, t = 1, 2, \ldots \qquad (23)$$

and it is distributed as

$$Z(t) \sim \begin{cases} N(0,1) & \tau > t \\ N\left(\frac{\mu}{\sqrt{1+\rho}}, 1\right) & \tau \leq t \end{cases}. \qquad (24)$$

Setting $m = \mu\left(\frac{1}{1+\rho} - \sqrt{2}\right)$ and $s = \sqrt{1+\rho}$ we can see that $s(Z(t) - m) = Y_1$.

To compare the two procedures we need to set their limits so that their ARL^0 equals some value $\gamma \in Z_+$. To set the limit for the procedure based on the lr-statistic is simple as the only limit satisfying $ARL^0 = \gamma$ is $K(s) \equiv K = \Phi^{-1}\left(\frac{\gamma-1}{\gamma}\right)$. The parallel procedure is more complicated since the ARL^0 depends on two critical limits, K_1 and K_2. For a pair of limits to fix ARL^0 to

12

γ the condition $ARL^0 = 1/(1 - F^0(K_1, K_2)) = \gamma$ has to be satisfied. As we are surveilling independent Gaussian variables this condition is equivalent to

$$\Phi\left(\frac{K_1}{\sqrt{1+\rho}}\right) \Phi\left(\frac{K_2}{\sqrt{1-\rho}}\right) = \frac{\gamma - 1}{\gamma}. \qquad (25)$$

Thus we have no unique pair of critical limits to fix ARL_0. The possible choices of limits are

$$\begin{cases} K_1(p) = \sqrt{1+\rho}\Phi^{-1}\left(\left(\frac{\gamma-1}{\gamma}\right)^{1-p}\right) \\ K_2(p) = \sqrt{1-\rho}\Phi^{-1}\left(\left(\frac{\gamma-1}{\gamma}\right)^{p}\right) \end{cases}, p \in [0,1]. \qquad (26)$$

The parameter p influences the amount of interest we put in each process.

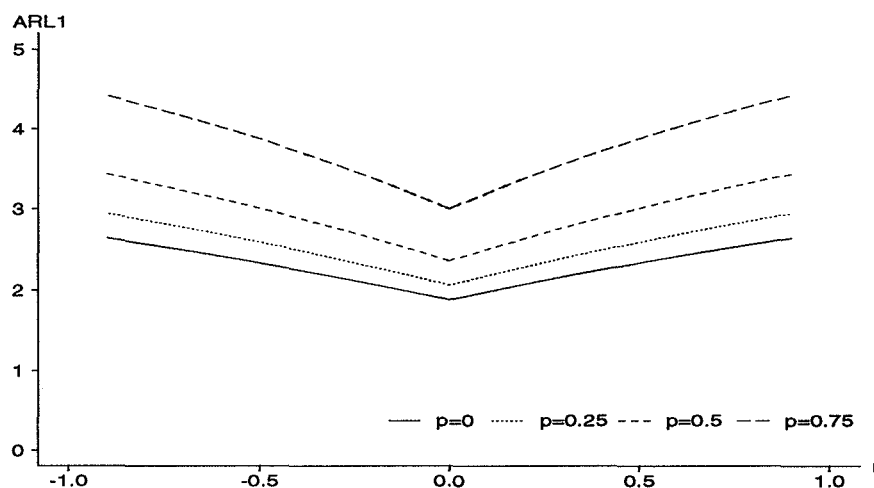

Figure 1: *The effect of the choice of critical limits, $(K_1(p), K_2(p))$, on ARL^1 for a UI Shewhart procedure. Shown for different values of the correlation, ρ.*

If we choose $p = 0$ only the first component is surveilled. Our choice of p influences ARL^1: for $p = 1$, ARL^1 equals ARL^0 and as $p \searrow 0$ decreases monotonically. In *Figure* 1 the effect of some choices of p with fixed $ARL^0 = 11$ is shown. Thus the most efficient or optimal choice of critical limits in our case is to use $(K_1(0), K_2(0))$. In practice this is equivalent to monitor only the first principal component, $Y_1(t), t = 1, 2, \ldots$, as was suggested by Jackson and Muldholkar (1979). Thus the optimal "parallel" Shewhart procedure here is the univariate Shewhart chart based on the likelihood ratio statistic $\{\mathbf{lr}_1(x_s); s = 1, 2, \ldots\}$.

5 A Likelihood Ratio Based Procedure

5.1 The General method

In this section we discuss the use of a method to construct surveillance procedures when a specific alternative is of interest,discussed Frisén and deMare (1991), in the case of multivariate processes. As before we will consider sudden simultanous shifts in all component processes between two fully specified alternatives. The method, here called the alarm method, constructs likelihood ratio based surveillance procedures designed for specific critical events and change point distributions.

The LR-procedures uses alarm functions of the form

$$p\left(x_s; C\left(s\right)\right) = \frac{d\Pr\left(x_s \left| C\left(s\right)\right.\right)}{d\Pr\left(x_s \left| D\left(s\right)\right.\right)} \tag{27}$$

and give an alarm whenever this function exceeds a critical limit $K(s)$. This critical limit usually depends both on the time of decision, s, and the distribution of the change point, G_τ. Depending on the critical event the alarm functions assume different forms. If for example we are interested in optimizing the procedures to immediate detection of a change, $C\left(s\right) = \left\{\tau = s\right\}$, the alarm function reduces to

$$p\left(x_s; \left\{\tau = s\right\}\right) = \frac{f^1\left(\mathbf{x}\left(s\right)\right)}{f^0\left(\mathbf{x}\left(s\right)\right)} \tag{28}$$

where f^0 and f^1 are the densities of $\mathbf{X}\left(t\right)$. Thus, with $K(s) \equiv K$ we have the Shewhart procedure based on the likelihood ratio statistic used in section 4.2.

If instead we are interested in optimizing for the critical event $C(s) = \left\{\tau \leq s\right\}$, the method results in an alarm function that can be written as

$$p\left(x_s; \left\{\tau \leq s\right\}\right) = \sum_{k=1}^{s} \frac{g_\tau\left(k\right)}{G_\tau\left(s\right)} \prod_{t=k}^{s} \frac{f^1\left(\mathbf{x}\left(t\right)\right)}{f^0\left(\mathbf{x}\left(t\right)\right)}. \tag{29}$$

With g_τ and G_τ being the density and distribution function of the change point τ. A choice of critical limits with good properties is here $K\left(s\right) = k \cdot \left(1 - G_\tau\left(s\right)\right)/G_\tau\left(s\right)$, where k is a constant, as the LR-procedure is then equivalent to the procedure suggested by Shiryaev (1963).

To specify the LR procedure we need here, beside the distributions of the observations, also some knowledge of the distribution of τ. The choice of distribution for τ is often the geometric distribution. It involves only one parameter, the intensity ν, and assumes equal probability of a change in each

14

point. As $\nu \to 0$ the procedure's dependence on the choice ν decreases so in many practical applications where ν is small the specific choice of ν is less important, Frisén and Wessman (1996). Also when $\nu \to 0$ the alarm function has the limit

$$\lim_{\nu \to 0} p\left(x_s; \{\tau \leq s\}\right) = \sum_{k=1}^{s} \prod_{t=k}^{s} \frac{f^1\left(\mathbf{x}\left(t\right)\right)}{f^0\left(\mathbf{x}\left(t\right)\right)} = p_r\left(x_s\right). \tag{30}$$

thus the LR-procedure has in this case the alarm function of Roberts (1966) as its limit.

Thus, we have that the LR-procedure (29), as well as the Shewhart and the Roberts (28, 30) are all based on the sequence $\{lr\left(\mathbf{x}\left(t\right)\right), t = 1, 2, \ldots\}$, which is sufficient for the family $\{C\left(s\right) = \{\tau \leq s\}, D\}$. They have therefore the same optimality properties for this situation (as the respective univariate surveillance procedure).

The LR-procedure for example is of course optimal according to definition 2.1 and satisfies several other optimality properties, see for example Frisén and Wessman (1996). These properties make the LR-procedure a good benchmark in theoretical or simulation studies for the multivariate surveillance problem discussed here. For univariate surveillance situations it has been used in this capacity by for example Frisén (1992,1996).

5.2   Surveillance of Multivariate Gaussian processes

A well-studied example in multivariate surveillance is that of a shift in the mean vector of a Gaussian process when the covariance, Σ, is known and stable throughout the surveillance. Without loss of generality we can consider shifts between $\mu^0 = (0, \ldots, 0)^T$ and $\mu^1 = (\mu_1, \ldots, \mu_p)^T$. The sufficient likelihood ratio statistic, $lr\left(t\right)$, is then, after normalizing,

$$\xi\left(\mathbf{x}\left(t\right)\right) = \frac{\mu'\Sigma^{-1}\mathbf{x}\left(t\right)}{\sqrt{\mu'\Sigma^{-1}\mu}} \tag{31}$$

and is distributed

$$\xi\left(\mathbf{x}\left(t\right)\right) \sim \begin{cases} N\left(0, 1\right) & \tau > t \\ N\left(\sqrt{\Delta}, 1\right) & \tau \leq t \end{cases} \tag{32}$$

with $\Delta = \mu'\Sigma^{-1}\mu$.

We can note that although the lr−statistic is here constructed for a shift to a specific point it has the same properties to detect a change to all points

15

in the parameter space satisfying $\left\{\mu^* \left| \mu'\Sigma^{-1}\mu^* = \mu'\Sigma^{-1}\mu \right.\right\}$. Also, for all shifts to points $\left\{\mu^* \left| \mu'\Sigma^{-1}\mu^* > \mu'\Sigma^{-1}\mu \right.\right\}$ we have that $\xi(\mathbf{x}^*(s)) \overset{D}{=} \xi(\mathbf{x}(s))$ before a shift and, since the shift is of size $\sqrt{\mu'\Sigma^{-1}\mu^*} > \sqrt{\Delta}$, we have that $\xi(\mathbf{x}^*(s)) \overset{D}{\geq} \xi(\mathbf{x}(s))$ after a shift. Thus, for these points we have better properties for detecting the shift. A multivariate surveillance procedure based on $\{\xi(\mathbf{x}(t)) ; t = 1, 2, \ldots\}$ for detecting a specific alternative mean vector has equal or better properties for monitoring a shift to a subspace of the parameter space.

The LR alarm function derived for the critical events mentioned above becomes

$$p(x_s; \{\tau = s\}) = \xi(\mathbf{x}(t)) \tag{33}$$

and

$$p(x_s; \{\tau \leq s\}) = \sum_{k=1}^{s} \frac{g_\tau(k)}{G_\tau(s)} \prod_{t=k}^{s} e^{\sqrt{\Delta}\xi(\mathbf{x}(t)) - \frac{1}{2}\Delta} \tag{34}$$

respectively. The Roberts surveillance procedure consequently becomes

$$p_r(x_s) = \sum_{k=1}^{s} \prod_{t=k}^{s} e^{\sqrt{\Delta}\xi(\mathbf{x}(t)) - \frac{1}{2}\Delta} \tag{35}$$

Also, for the situation considered in this section Healy (1987) showed that the CUSUM-procedure

$$S(x_s) = \max\left(S(x_{s-1}) + \log\left(\frac{f^1(\mathbf{x}(t))}{f^0(\mathbf{x}(t))}\right), 0\right). \tag{36}$$

can, after normalizing, be written as

$$S(x_s) = \max\left(S(x_{s-1}) + \xi(\mathbf{x}(t)) - \frac{1}{2}\sqrt{\Delta}, 0\right). \tag{37}$$

Thus, all four multivariate surveillance procedures considered here are reduced to univariate surveillance procedures based on the univariate Gaussian process $\{\xi(\mathbf{x}(t)) ; t = 1, 2, \ldots\}$. We can therefore compare them using results for the surveillance of the mean of a univariate Gaussian process, for example a change in distribution between $N(0,1)$ and $N(1,1)$. Such a comparison can be found in Frisén and Wessman (1996).

16

6 Conclusions

In many multivariate surveillance applications detection of changes in a specific direction is of interest. It is sometimes of interest to construct a method with good properties to detect a change between two fully specified alternatives, one before and one after the change, for the case when the change points occur simultaneously. We have shown that in these cases observing the univariate likelihood ratio statistic of the observations made at each time is sufficient for detecting the critical event $C(s) = \{\tau \leq s\}$. In some cases it is also minimal sufficient. A good choice of surveillance procedure is therefore one based on this univariate process since the loss in efficiency when using a procedure not based on a sufficient reduction can be large, as was shown in section 4.2.

A multivariate surveillance procedure constructed by the LR-method is here shown to be equivalent to the surveillance procedure for the surveillance of a univariate process and to retain any optimality properties shown for univariate surveillance situations. The procedure is therefore a good candidate both as a surveillance procedure and as a benchmark in comparisons between different multivariate surveillance procedures. We also show that comparisons between the Shewhart, the CUSUM, the Roberts and the LR-procedure are especially simple when multivariate Gaussian observation processes are considered. Such results can be found in for example Frisén & Wessman (1996).

References

[1] Alt, F. B. (1985). Multivariate quality control, *Encyclopedia of Statistical Sciences*. eds. Kolz, S & Johnson N. L. New York: John Wiley & Sons.

[2] Arnold, Steven F. (1988). *Encyclopedia of Statistical Science*. eds. Kolz, S & Johnson N. L, New York: John Wiley & Sons.

[3] Cox, D. R. & Hinkley, D. V. (1974). Theoretical Statistics, Cambridge: Chapman & Hall.

[4] Crosier, R. B. (1988). Multivariate Generalizations of Cumulative Sum Quality-Control Schemes, *Technometrics*, **30**, 291-303.

[5] Frisén, M. (1992), Evaluation of methods for statistical surveillance. *Statistics in Medicine*, **11**, 1489-1502.

[6] Frisén, M. (1994). Statistical surveillance of business cycles. *Research Report*. 1994:1, Department of Statistics, University of Göteborg.

[7] Frisén, M. (1996), Characterization of Methods for Surveillance by Optimality. *Unpublished Report*.

[8] Frisén, M. & deMaré, J. (1991). Optimal Surveillance. *Biometrica*. **78**, 271-80

[9] Frisén, M. & Wessman, P. (1996). Evaluation of Likelihood Ratio Method for Surveillance. *Research Report*. 1996:3, Department of Statistics, University of Göteborg.

[10] Hawkins, D. M. (1991). Multivariate Quality Control Based on Regression-Adjusted Variables. *Technometrics*, **33**, 61-75.

[11] Hawkins, D. M. (1993). Regression Adjustment for Variables in Multivariate Quality Control. *Journal of Quality Technology*, **25**, 170-182.

[12] Healy, J. D. (1987). A Note on Multivariate CUSUM Procedures. *Technometrics*, **29**, 409-412

[13] Hotelling, H. (1947). Multivariate quality control. Illustrated by the Air Testing of Bombsights. *Techniques of Statistical Analysis*, eds. Eisenhart, C. & Hastay, M. W. New York: McGraw-Hill, 111-184.

[14] Hochberg, Y. & Tamhane, A. C. (1987). *Multiple Comparison Procedures*. New York: John Wiley & Sons.

[15] Jackson, J. E. & Bradley, R. A. (1966). Sequential Multivariate Procedures for Means with Quality Control Applications. *Multivariate Analysis, Vol 1.* eds. Krishnaian, P. R. 507-520.

[16] Jackson, J. E. & Mudholkar, G. S. (1979). Control Procedures for Residuals Associated With Principal Component Analysis. *Technometrics.* **21**, 341-349.

[17] Lowry, C. A., Woodall, W. H., Champ, C. H. & Rigdon, S. E. (1992). A Multivariate Exponentially Weighted Moving Average Control Chart. *Technometrics*, **34**, 46-53.

[18] Pettersson, M. (1996) Statistical Surveillance of the Environment. *Unpublished Report.* Department of Statistics, University of Göteborg.

[19] Pigniatello, J. J. & Runger, G. C. (1990) Comparison of Multivariate CUSUM Charts. *Journal of Quality Technology*, **22**, 173-186.

[20] Roberts, S. W. (1966). A Comparison of Some Control Chart Procedures. *Technometrics*, **8**, 411-430.

[21] Shewart W. A. (1931). Economic Control of Quality of Manufactured Product. Princeton, NJ; Van Nostrand.

[22] Shirayev, A. N. (1963). On Optimal Methods in Quickest Detection Problems. *Theory of Probability and its Applications*, **8**, 22-46.

[23] Sveréus, A. (1995), Detection of Gradual Changes. Statistical Methods in Post Marketing Surveillance. *Research Report.* 1995:2, Department of Statistics, University of Göteborg.

[24] Wetherill, G. B. & Brown, D. W. (1991). *Statistical Process Control.* London: Chapmann & Hall

[25] Woodall, W. H. & Ncube, M. M. (1985). Multivariate CUSUM Quality-Control Procedures. *Technometrics*, vol. **27**, 285-192.

[26] Åkermo, Göran (1994). On Performance of Methods for Statistical Surveillance. *Research Report.* 1994:6, Department of Statistics, University of Göteborg.

EVALUATIONS OF
LIKELIHOOD RATIO METHODS
FOR SURVEILLANCE.

By Marianne Frisén and Peter Wessman
Göteborg University

Methods based on likelihood ratios are known to have several optimality properties. When control charts are used in practice, knowledge about several characteristics of the method is important for the judgement of which action is appropriate at an alarm. The probability of a false alarm, the delay of an alarm and the predictive value of an alarm are qualities (besides the usual ARL) which are described by a simulation study for the evaluations. Since the methods also have interesting optimality properties, the results also enlighten different criteria of optimality. Evaluations are made of the "The Likelihood Ratio Method" which utilizes an assumption on the intensity and has the Shiryaev optimality. Also, the Roberts and the CUSUM method are evaluated. These two methods combine the likelihood ratios in other ways. A comparison is also made with the Shewhart method, which is a commonly used method.

Running title: Evaluations of likelihood ratio based surveillance

In many areas there is a need for continual observation of time series, with the goal of detecting an important change in the underlying process as soon as possible after it has occurred. In recent years there has been a growing number of papers in economics, medicine, environmental control and other areas dealing with the need of methods for surveillance. Examples are given in Frisén (1992) and Frisén (1994a). The timeliness of decisions is taken into account in the vast literature on quality control charts where simplicity is often a major concern. Also, the literature on stopping rules is relevant. For an overview, see the textbook by Wetherill and Brown (1990) the survey by Zacks (1983) and the bibliography by Frisén (1994 b).

Methods based on likelihood ratios are known to have several optimality properties. Evaluations are made of the LR method that is based on likelihood ratios and which in the case studied here has the Shiryaev optimality. Also, the Roberts method and the CUSUM method are evaluated. These two methods combine the likelihood ratios in other ways. A comparison is also made with the Shewhart method which is a commonly used method. When control charts are used in practice, it is necessary to know several characteristics of the method. Asymptotic properties have been studied by e.g. Srivastava and Wu (1993). Here properties for finite time of change are studied. The probability of a false alarm, the expected delay, the probability of successful detection and the predictive value are measures (besides the usual ARL) used for evaluations. Since the methods have interesting optimality properties, the results also enlighten different criteria of optimality.

In Section 1 some notations are given and there is a specification of the situation that is studied. In Section 2 the methods are described and their relations to each other and to different optimality criteria are discussed. In Section 3 comparisons based on a simulation study are reported. In Section 4 some concluding remarks are given.

1. NOTATIONS AND SPECIFICATIONS

The variable under surveillance is $X = \{X(t): t = 1,2,...\}$, where the observation at time t is $X(t)$. It may be an average or some other derived statistic. The random process which determines the state of the system is denoted $\mu = \{\mu_t: t = 1,2,...\}$.

The critical event of interest at decision time s is denoted $C(s)$. As in most literature on quality control, the case of shift in the mean of Gaussian random variables from an acceptable value μ^0 (say zero) to an unacceptable value μ^1 is considered. Only one-sided procedures are considered here. It is assumed that if a change in the process occurs, the level suddenly moves to another constant level, $\mu^1 > \mu^0$, and remains on this new level. That is $\mu_t = \mu^0$ for $t = 1,...,\tau-1$ and $\mu_t = \mu^1$ for $t = \tau, \tau+1,....$ We want to discriminate between

$$C(s) = \{\tau \leq s\} = \{\mu(s)=\mu^1\} \quad \text{and} \quad D(s) = \{\tau > s\} = \{\mu(s)=\mu^0\}.$$

Here μ^0 and μ^1 are regarded as known values and the time τ where the critical event occurs is regarded as a random variable with the density

$$\pi_t = \Pr(\tau=t)$$

and $\sum \pi_t = 1-\pi_\infty$. The intensity v_t of a change is

$$v_t = \Pr(\tau=t \mid \tau \geq t)$$

The aim is to discriminate between the states of the system at each decision time s, s=1,2,... by the observation $X_s = \{X(s): t \leq s\}$ under the assumption that $X(1) - \mu_1$, $X(2) - \mu_2$,... are independent normally distributed random variables with mean zero and with the same known standard deviation (say $\sigma=1$). In some calculations below, where no confusion is possible, μ^1 is denoted μ and $\mu^0=0$ and $\sigma=1$ for typographical clarity.

Active surveillance (Frisén and de Maré 1991) is assumed. That means that the surveillance is stopped at an alarm. Thus, only one alarm is possible. A study of the properties of the first alarm at passive surveillance gives the same results.

2. METHODS

2.1 The Likelihood Ratio method

The problem of finding the method which maximizes the detection probability for a fixed false alarm probability and a fixed decision time was treated by de Maré (1980) and Frisén and de Maré (1991). The likelihood ratio (LR) method, discussed below, is the solution to this criterion. Different kinds of utility functions were discussed by Frisén and de Maré (1991). An important specification of utility is that of Girshick and Rubin (1952) and Shiryaev (1963). They treat the case of constant intensity where the gain of an alarm is a linear function of the difference t_A-τ between the time of the alarm and the time of the change. The loss of a false alarm is a function of the same difference. The utility can be expressed as U= E{u(τ, t_A)}, where

$$u(\tau,t_A) \begin{cases} =l(\tau-t_A) & \text{if } \tau>t_A \\ =a_1(\tau-t_A)+a_2 & \text{else} \end{cases}$$

Their solution to the maximization of the expected utility is identical to the LR method for the situation specified in Section 1 (Frisén 1996).

The general method uses combinations of likelihood ratios. Even though methods based on likelihood ratios have been suggested earlier, for other reasons, the use in practice is (yet) rare. Here, the method is applied to the

shift case specified in Section 1. The "catastrophe" to be detected at decision time s is $C = \{ \tau \leq s\}$ and the alternative is $D = \{ \tau > s\}$.

The LR method has an alarm set consisting of those X_s for which the likelihood ratio exceeds a limit:

$$f_{Xs}(x_s \mid C) \,/f_{Xs}(x_s \mid D) = p(x_s) =$$

$$= \sum_{t=1}^{\infty} w(t)L(t) > K(s)$$

where $w(t) = \Pr(\tau=t)/\Pr(\tau \leq s)$ and $L(t)$ is the likelihood ratio when $\tau=t$.

For the case of normal distribution, $C(s)=\{\tau \leq s\}$ and $D(s)=\{\tau > s\}$ specified in Section 1 we have

$$p(x_s)=h(s)p_s(x_s)$$

where

$$h(s)=\frac{\exp(-(s+1)(\mu^1)^2/2)}{\Pr(\tau \leq s)}$$

and

$$p_s(x_s)=\sum_{k=1}^{s} \pi_k \exp\left\{\frac{1}{2}k(\mu^1)^2\right\}\exp\{\mu^1 \sum_{u=k}^{s} x(u)\}$$

which is a nonlinear function of the observations.

In order to achieve the optimal error probabilities discussed by Frisén and de Maré (1991) an alarm should be given as soon as $p(x_s) > K(s)$.

In order also to achieve maximization of the utility of Shiryaev it is required that

$$\Pr(\tau \le s | X_s = x_s) > K \quad \Rightarrow \quad p(x_s) > \frac{\Pr(\tau > s)}{\Pr(\tau \le s)} \frac{K}{1-K}$$

where K is a constant. Now we must also consider the function h(s).

2.2 Roberts method

Roberts (1966) suggested that an alarm is triggered at the first time s, for which

$$\sum_{t=1}^{s} L(t) > K$$

where K is a constant.

This is the limit of the LR method

$$\sum_{t=1}^{s} w(t)L(t) > K(s)$$

when v tends to zero since both the weights w(t) and the limit K(s) tend to constants. Thus the Roberts method approximately satisfies the optimality criterion of Shiryaev (1963) for small values of the intensity v. Roberts (1966) motivated the method by the conjecture that the intensity parameter v has very little influence on the LR method, when v is in the interval zero to 0.2 and thus the weights which depend on v can be omitted. The Roberts method is sometimes called the Shiryaev-Roberts method. Pollak (1985) demonstrated that the Roberts procedure is asymptotically (as $K \to \infty$) minimax. Mevorach and Pollak (1991) examined the expected delay (see below) by simulations in a small sample setting and concluded that the difference between this method and the CUSUM method is small.

2.3 CUSUM

The cumulative sum

$$C_t = \sum_{i=1}^{t} (X(i) - \mu^0)$$

is used in several CUSUM- variants. The most commonly advocated variant gives an alarm for the first t for which $C_t - C_{t-i} > h + ki$ (for some i=1,2,... t), where $C_0 = 0$ and h and k are chosen constants (Page 1954).

Sometimes the CUSUM test is defined in a more general way by likelihood ratios (Siegmund 1985 and Park and Kim 1990). This reduces to the method above in the case specified in Section 1.

In the simulation study below the slope $(\mu^0 + \mu^1)/2$ is used. The values $\mu^0 = 0$ and $\mu^1 = 1$ gives the slope k=1/2. The requirement of $ARL^0 = 11$ gives the parameter h=0.985. The short ARL^0 was chosen to make the computer time, necessary for the study, reasonable.

2.4 The Shewhart method

Shewhart (1931) suggested that an alarm is triggered as soon as a value which deviates too much from the target is observed. Frisén and de Maré (1991) demonstrated that this is the same as the LR method with $C = \{\tau = s\}$. That is $p(x_s) = LR(s)$. For the normal case described in Section 1 this reduces to $p(x_s) = x(s)$. The limit G for an alarm is calculated by the relation: $Pr(X(s) > G | \mu = \mu^0) = 1/ARL^0$. $ARL^0 = 11$ gives G= 1.3353.

3. RESULTS

For the Shewhart method exact calculations were made. For the other methods simulations of 10 000 000 replicates were made for each point in the diagrams.

3.1 Probability of a false alarm

In all simulations the parameters of the different methods were chosen so that ARL^0 was equal for all methods (equal to 11) to make them comparable. To fix the value of ARL^0 is not the only way to do this, but it is in accordance with most comparative studies in this area.

Equal values of ARL^0 do not imply that the run length distributions are identical when $\mu \equiv \mu^0$. The distributions can have different shapes. This is demonstrated in Figures 1 and 2. In Figure 1 the probabilities of an alarm at a specific time point, when no change has occurred, are given for some different methods. The skewness of the distributions is less pronounced for the LR method with a great intensity parameter. For the chosen parameters, the distributions for the Shewhart and the CUSUM methods are very similar except at the first point. In Figure 2, the probability of an alarm no later than at t given that no change has occurred, $\alpha_t = Pr(t_A \leq t \mid \mu = \mu^0)$, is illustrated. This is also the cumulative distribution function of the run length when the process is in control. The results of the LR method with parameters $v=0.001$ and $v=0.01$, cannot be distinguished from those of the Roberts method in the scale of the figures and are therefore not included. The results for $v=0.1$ are also very close to those of Roberts method and are included only in Figure 1 where more details can be seen.

A summarizing measure of the false alarm distribution is the probability of a false alarm, when the probability of a change has a geometric distribution with the intensity v. $Pr(t_A < \tau)$ is illustrated as a function of v in Figure 4.

$$Pr(t_A < \tau) = \sum_{t=1}^{\infty} Pr(\tau = t) Pr(t_A < t \mid \tau = t)$$

e first factor in the sum does not depend on the method but only on the true intensity v. The second factor depends only on the run length distribution when $\mu = \mu^0$. Since the values of the ARL^0 are equal for all methods only the different shapes and not their locations will influence the false alarm probability. Thus only the very modest differences as are seen in Figure 4 can be expected.

3.2 Delay of an alarm

As was seen above, the correspondence to the level of significance in an ordinary test is not a value but a distribution. For the power the correspondence is still more complicated. To describe the ability of detecting a change we need a set of run length distributions. Some kind of summarizing measure is of value.

The distribution of t_A, the time of an alarm, when the change occurred before the surveillance started ($\tau=1$), that is $\mu \equiv \mu^0$, is illustrated in Figure 3 for some methods.

The average run length under the alternative hypothesis, ARL^1, is the mean number of decisions that must be taken to detect a true level change (that occurred at the same time as the inspection started). The part of the definition in the parenthesis is seldom spelled out but seems to be generally used in the literature on quality control. The values of ARL^1 for the methods and situations examined are given in Table 1.

Since the case $\tau=1$ of is not the only case of interest the expected delay is calculated also for other values $\tau=t$.

$$E(t_A - \tau \mid t_A \geq \tau, \ \tau = t)$$

is given in Figure 5. For $\tau=1$ the values of this function equal the values of ARL^1 - 1. The differences in shapes of these curves demonstrate the need for other measures than the conventional ARL. Although the Roberts method has worse ARL^1, and thus worse delay for a change at $\tau=1$, than the Shewhart method, it is better for all other times of change. The CUSUM and the Shewhart methods are very much alike for $\tau=1$ but the CUSUM is here much better for other change points.

In some applications the loss of a delay is directly proportional to the expected value of the delay. The expected delay also with the respect to the distribution of τ is:

$$E(t_A - \tau \,|\, t_A \geq \tau)$$

It is given as a function of v in Figure 6, for the case when the distribution of τ is geometrical with the intensity v. Since the computer time necessary for reliable values for small values of v is too great, these values were not given in the figure. The shapes of the curves for small values are similar to the curve (exactly calculated) for the Shewhart method. When v tends to one, the expected delay tends to ARL^1-1.

In some applications there is a limited time available for rescuing actions. Then, the expected value of the difference τ-t_A is not of main interest. Instead of using the expected value, the probability that the difference does not exceed a fixed limit is used. The fixed limit, say d, is the time available for successful detection.

The probability of successful detection,

$$PSD(\tau,d) = Pr(t_A - \tau \leq d \,|\, t_A \geq \tau).$$

was suggested by Frisén (1992) as a measure of the performance. It is illustrated in Figure 7. The PSD is better for the CUSUM than for the Shewhart method for d=3. The shape of the curve for CUSUM with the present parameters is very similar to the constant curve for Shewhart. The Roberts and the LR methods have a worse probability of detection of a change which happens early but better for late changes.

3.3 Predicted value

The predictive value $PV(t) = Pr(\tau \leq t \,|\, t_A = t)$ has been used as a criterion of evaluation by Frisén (1992), Frisén and Åkermo (1993) and Frisén and Cassel (1994). It is illustrated in Figure 8. The price for the high probability of detection of a change in the beginning of the surveillance (as was demonstrated in Figure 7) for the CUSUM and the Shewhart method is that the early alarms are not reliable.

4. CONCLUDING REMARKS

At large, the properties differ between the LR methods on one hand and the Shewhart and the CUSUM methods on the other, in the simulations. In comparison with this, the choice of the intensity parameter of the LR method has very little influence on the performance. The results here confirm the conjecture by Roberts (1966) about the robustness of the LR method.

That the Roberts method is the limit of the LR method when the intensity v tends to zero is also seen by the simulated results. The smaller the intensity parameter, the smaller the difference between the LR method and the Roberts method. Simulations were made also for $v=0.001$ and $v=0.01$ but these results were not included in the figures since the differences to those of the Roberts method are less than the line width. The results for $v=0.1$ were included only in the first figure where more details are given. The Roberts method is a good approximation of the optimal LR method and thus approximately optimal, for small values of v.

Sometimes the LR method is considered to be a Bayesian method while the Roberts method is considered a frequentistic one. Here, however all evaluations are made in the frequentistic framework. No Bayesian assumptions are necessary for the LR method. The properties of the method will be better if the intensity parameter is not far from the actual intensity. However, since the LR method is very robust for misspecification of the value of the intensity parameter v the gain with a precise search for the best value of the intensity parameter might not be worthwhile for a specific application.

The optimal method when the intensity is great should intuitively increase the probability of early alarms. Some of the present results might seem surprising in this light and will now be discussed.

The shapes (see Figures 1-3) of the distributions of the alarm time t_A might seem surprising. The probability of a very early alarm by the LR method

with a low intensity parameter is greater than for a large value of the parameter. However, for a low intensity the probability of a late change is great and thus a thick tail of the distribution of t_A is appropriate. As the expected value ARL^0 is fixed the only possibility is a high probability in the beginning. This also causes the differences in false alarm probabilities in Figure 4.

The greater expected delay (see Figures 5 and 6) for the LR method with a great intensity parameter v might seem surprising. For a fixed false alarm loss ℓ the delay will be less for greater values of the intensity parameter v. However, now we have a fixed ARL^0, and thus a greater value of l, which implies that the delay is increased to make the locations of the distributions of t_A equal. The only difference between the distributions is the shape and this difference causes the expected delay to be greater for the greater values of the intensity parameter.

The result by Mevorach and Pollak (1991), that the expected delay is similar for the Robert and CUSUM methods, was not confirmed. They studied an artificial quasi-stationary situation where the time of change, τ, does not have any influence. For a more realistic situation and the low ARL^0 used here the CUSUM in many aspects is very similar to the Shewhart method and not to the Roberts method.

Figure 7 demonstrates that the Roberts method has higher probability of a quick (within three time units) detection than the Shewhart method, unless the change occurs very early. For the LR(0.5) method this difference to the Shewhart method is still more pronounced.

In Figure 8 the predicted value of an alarm is given. This reflects the trust you should have in an alarm. The Roberts method has a relatively constant predicted value. This means that the same kind of action is appropriate both for early and late alarms.

REFERENCES

Frisén, M. (1992), Evaluations of methods for statistical surveillance, *Statistics in Medicine*, 11, 1489-1502.

Frisén M. (1994a) Statistical Surveillance of Business Cycles. *Research report*, 1994:1, Department of Statistics, Göteborg University.

Frisén M. (1994b) A classified bibliography on statistical surveillance. *Research report*, 1994, Department of Statistics, Göteborg University.

Frisén M. (1996) Characterization of methods for surveillance by optimality. Manuscript.

Frisén M. and Cassel C. (1994) Visual evaluations of statistical surveillance. *Research report*, 1994:3, Department of Statistics, Göteborg University.

Frisén, M. and de Maré, J. (1991), Optimal surveillance, *Biometrika*, 78, 271-280.

Frisén, M. and Åkermo, G. (1993), Comparison between two methods of surveillance: exponentially weighted moving average vs CUSUM, *Research report*, 1993:1, Department of Statistics, Göteborg University.

Girshick, M. A. and Rubin, H. (1952), A Bayes approach to a quality control model, *Ann. Math. Statist.*, 23, 114-125.

Maré, J. de (1980), Optimal prediction of catastrophes with application to Gaussian processes, *Ann. Prob.* 8, 841-850.

Mevorach, Y. and Pollak, M. (1991), A small sample size comparison of the CUSUM and Shiryaev-Roberts approaches to changepoint detection, *American J. of Mathematical and Management Sciences*, 11, 277-298.

Page, E. S. (1954), Continuous inspection schemes, *Biometrika*, 41, 100-114.

Park, C. S. and Kim, B. C. (1990) A CUSUM chart based on log probability ratio statistic, *J. Korean Statistical Society*, 19, 160-170.

Pollak, M. (1985) Optimal stopping times for detecting changes in distributions. *Ann. Statist.* 13, 206-227.

Roberts S. W. (1966) A comparison of some control chart procedures. *Technometrics*, 8, 411-30.

Shewhart, W. A. (1931) *Economic Control of Quality Control*. Reinhold Company, Princeton N.J.

Shiryaev, A. N. (1963), On optimum methods in quickest detection problems, *Theory Probab. Appl.*, 8, 22-46.

Siegmund, D. (1985), *Sequential analysis. Tests and confidence intervals*, Springer.

Srivastava, M.S. and Wu, Y (1993) Comparison of EWMA, CUSUM and Shiryaev-Roberts procedures for detecting a shift in the mean. *Ann. Statist.*, **21**, 645-670.

Wetherill, G.B. and Brown, D.W. (1990), *Statistical process control*, London: Chapman and Hall.

Wessman, P. (1996) Some principles for surveillance adopted for multivariate processes with a common change point. *Research report*, 1996:2, Department of Statistics, Göteborg University.

Zacks, S. (1983) Survey of classical and Bayesian approaches to the change-point problem: Fixed sample and sequential procedures of testing and estimation. in *Recent advances in statistics*, 245-269.

DEPARTMENT OF STATISTICS
GÖTEBORG UNIVERSITY
S-41180 GÖTEBORG
SWEDEN

Method	ARL[1]
CUSUM	2.61
Shewhart	2.71
Roberts	3.00
LR(0.001)	3.00
LR(0.01)	3.01
LR(0.1)	3.07
LR(0.5)	3.85

Table 1. The expected value of the alarm time t_A, when the change occurs immediately ($\tau=1$), and the parameters are set to make the expected value, when no change occurs, the same for all methods ($ARL^0=11$).
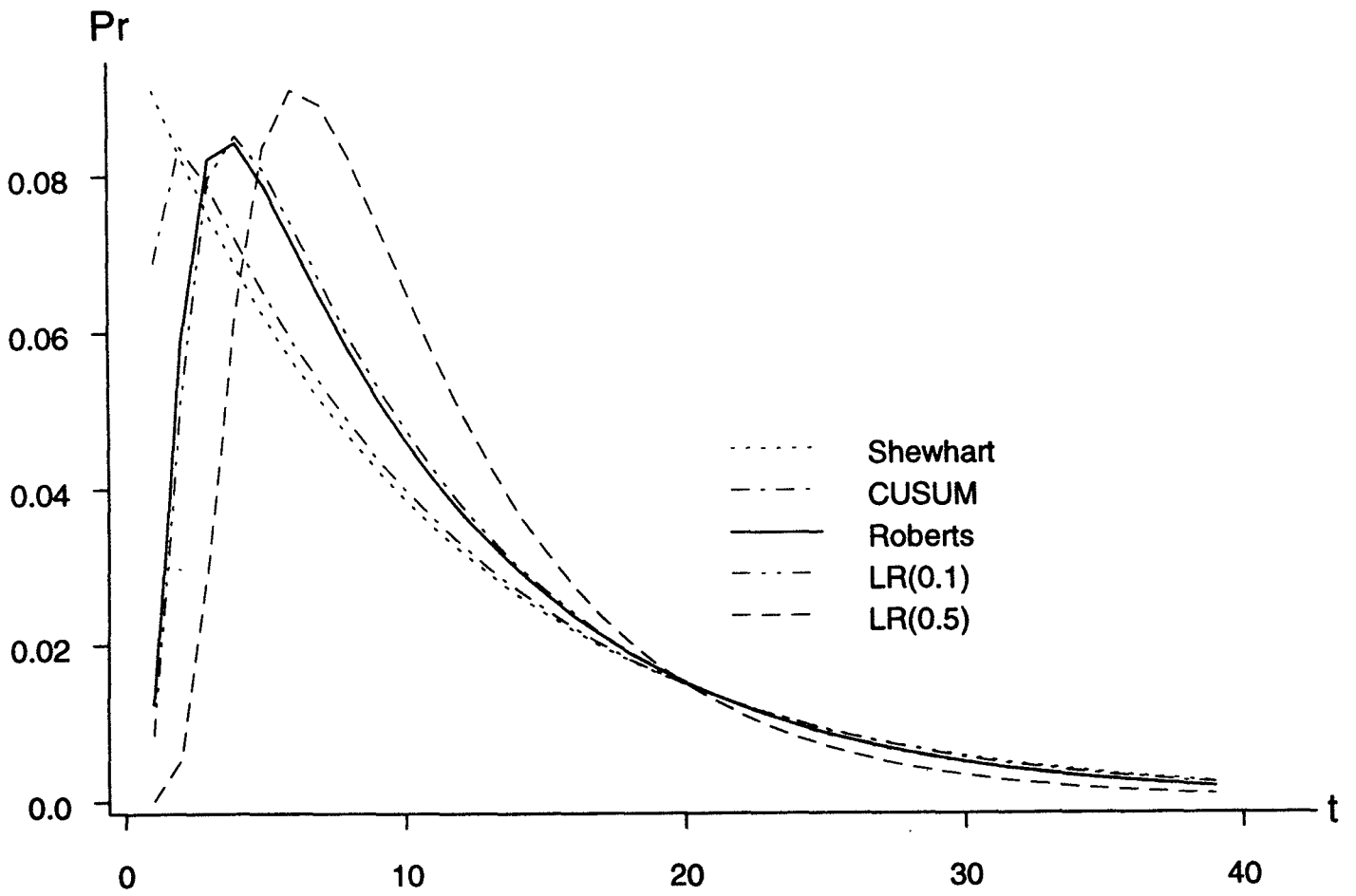
Figure 1. The density of the time of alarm, $\Pr(t_A=t|\mu_i\equiv\mu^0)$. This is the probability of an alarm at time t, when no change has occurred. The expectations, ARL^0, of the distributions are all equal to 11.
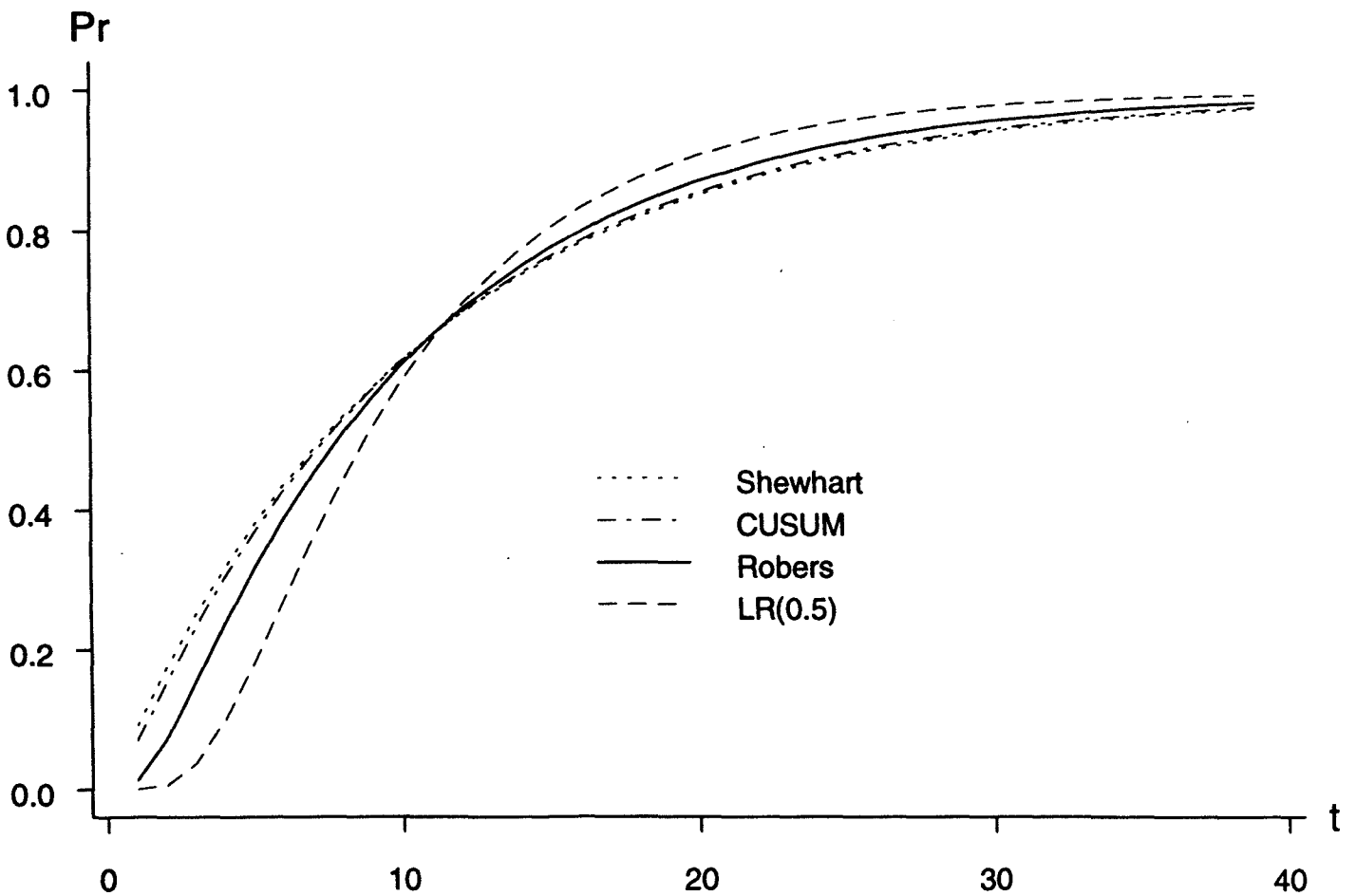
Figure 2. The distribution of the time of an alarm, when no change has occurred, $\alpha_t=\Pr(t_A\leq t|\mu_i\equiv\mu^0)$. The expectations, ARL^0, of the distributions are all equal to 11.
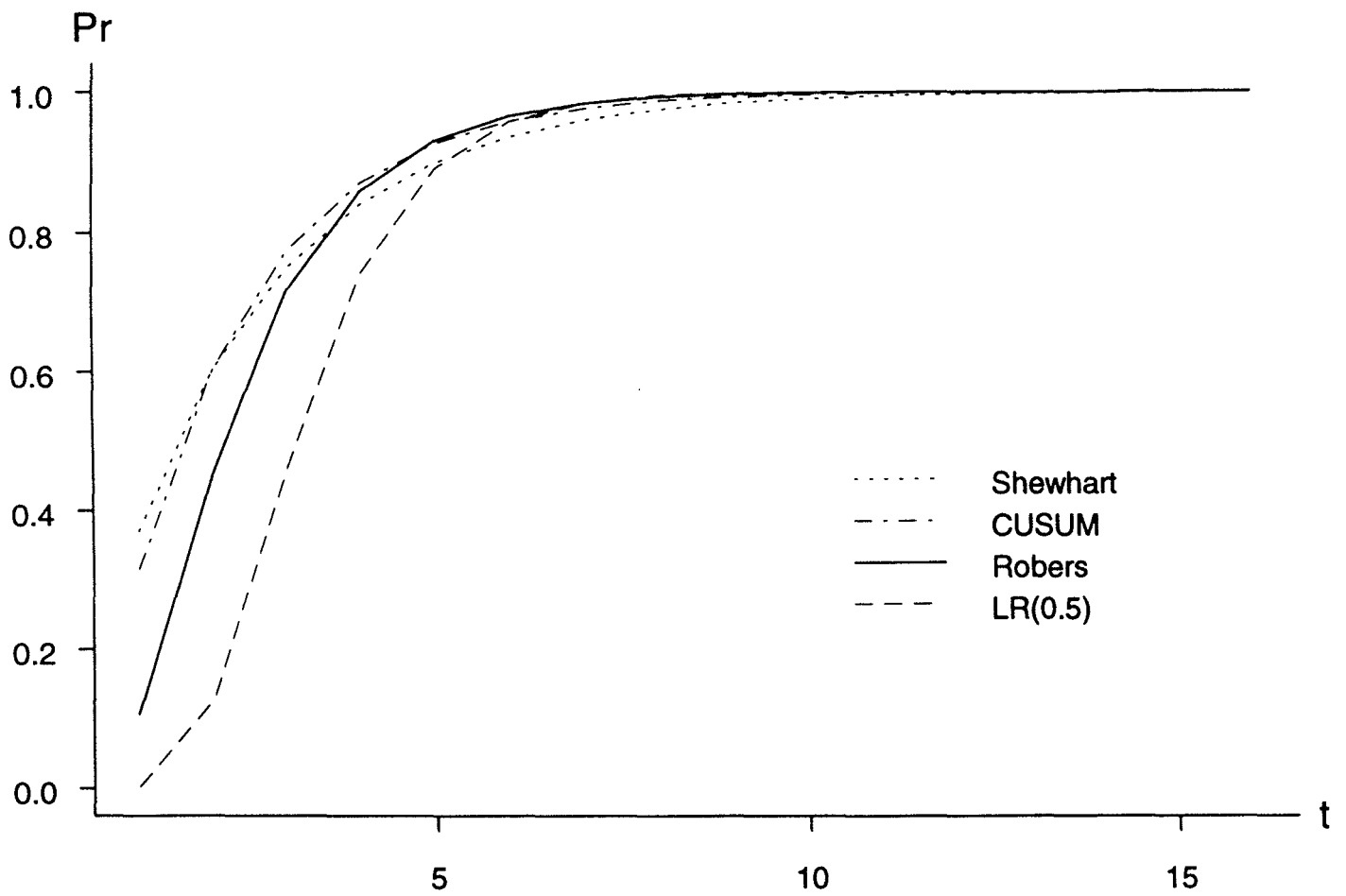
Figure 3. The distribution of the time of an alarm, when the change occurred before the surveillance started, $\Pr(t_A \leq t | \mu_i = \mu^1)$. The expectations, ARL^0, of the distributions are all equal to 11.
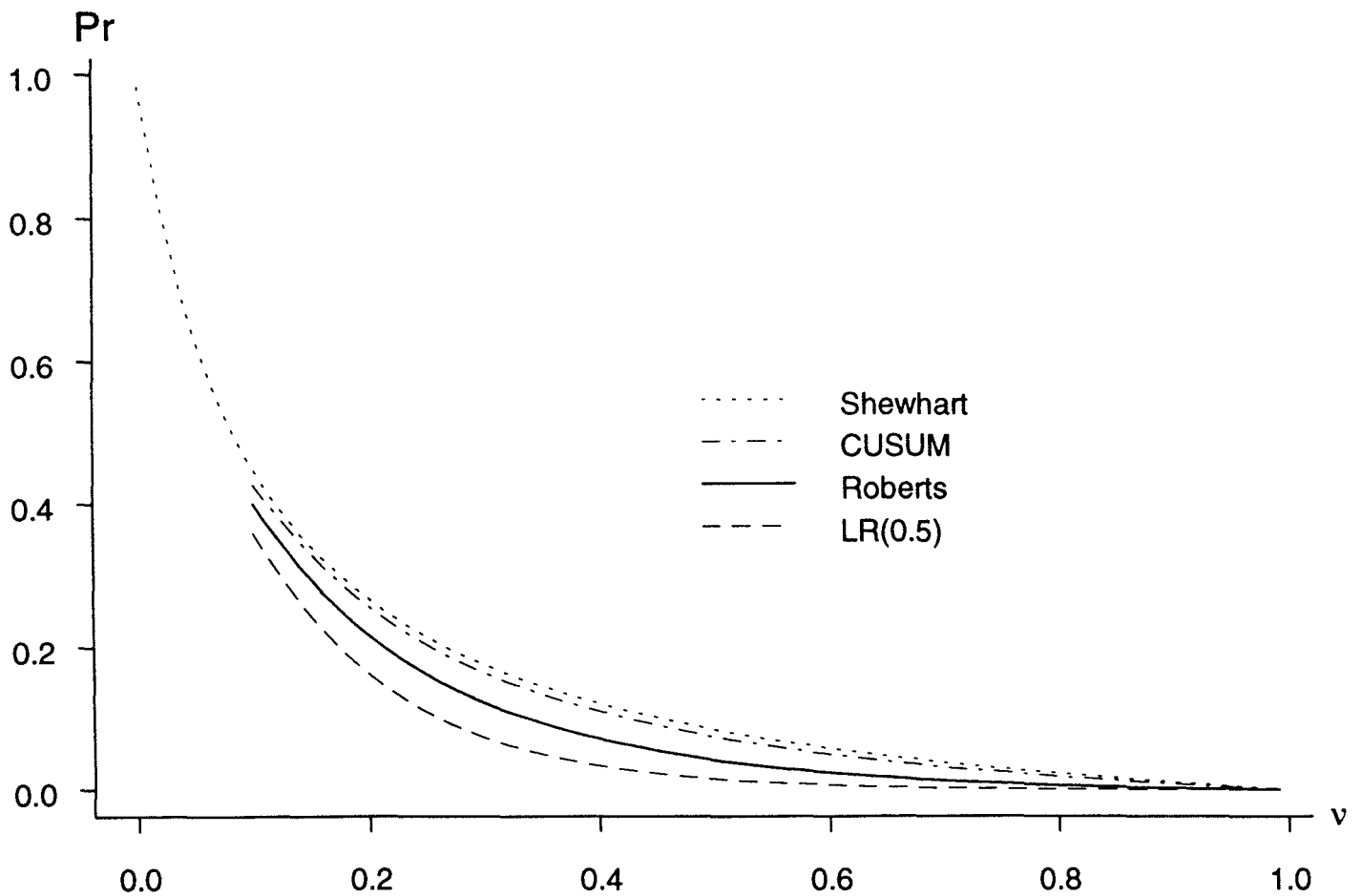
Figure 4. The expected probability of a false alarm $\Pr(\tau > t_A)$ when the distribution of τ is geometrical with intensity v.
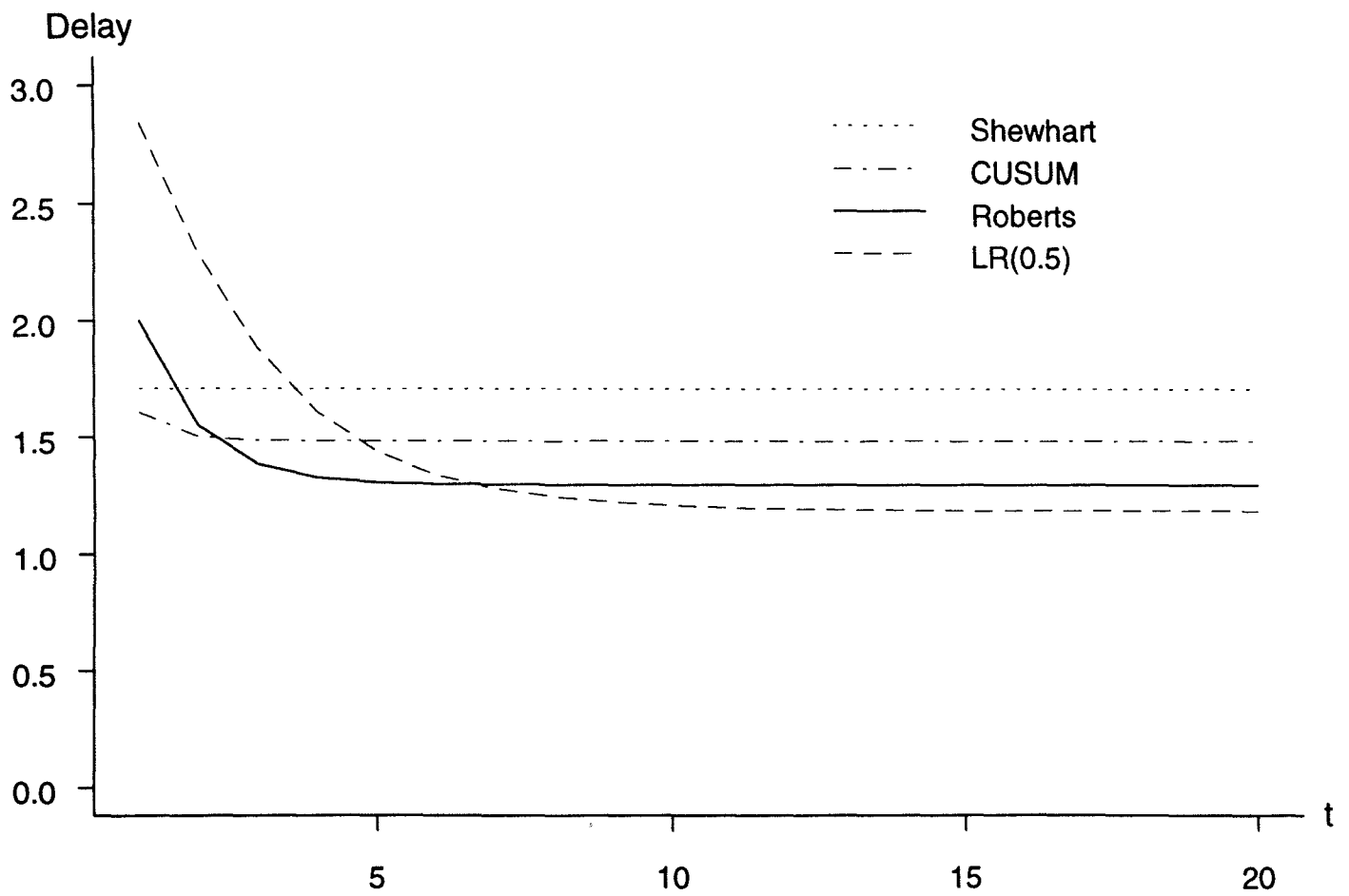
Figure 5. The expected delay after a change at time $\tau=t$. That is:

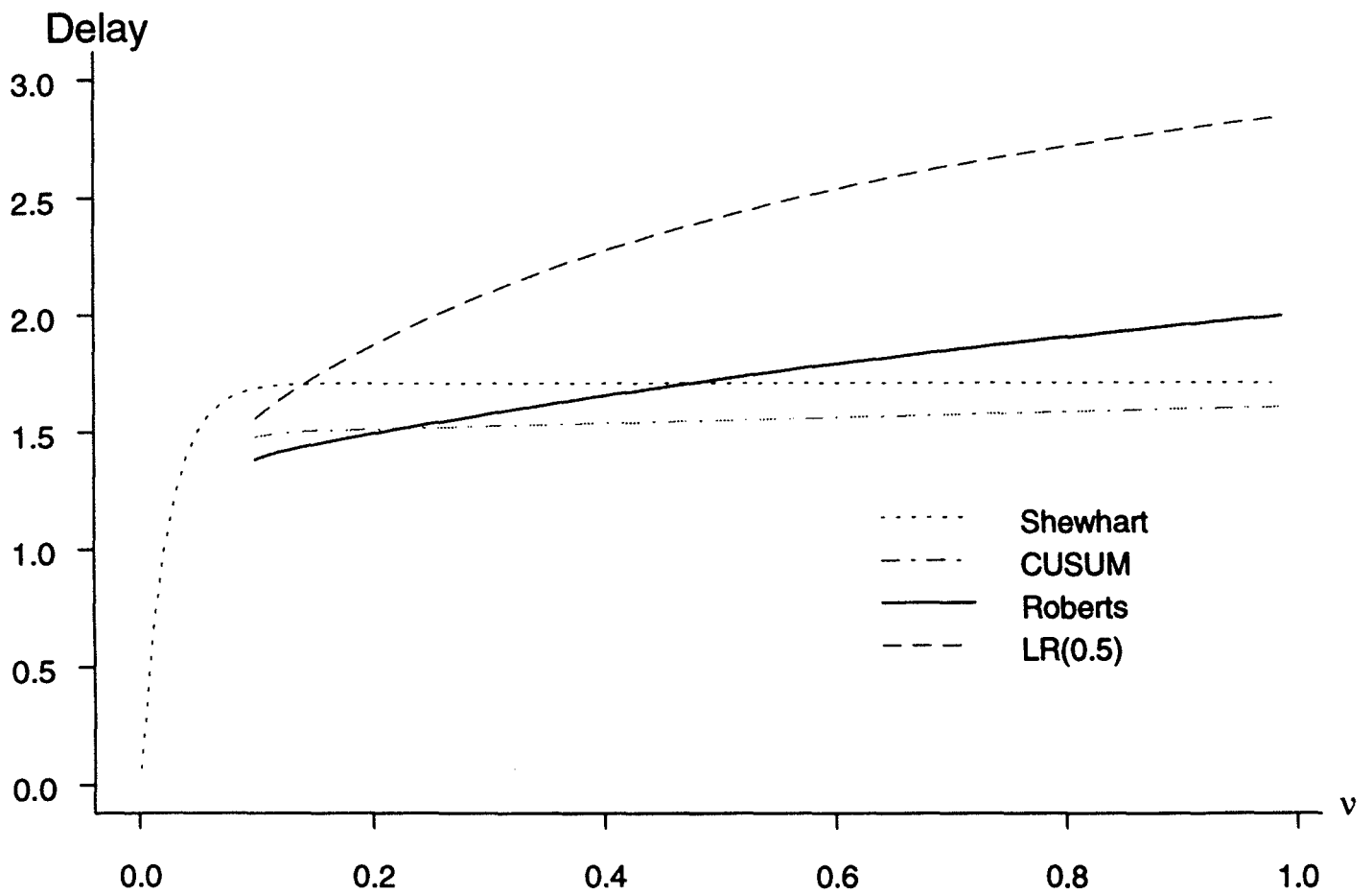$$E(t_A - \tau \mid t_A \geq \tau, \ \tau = t).$$

Figure 6. The expected delay when τ has a geometric distribution with intensity ν. That is:
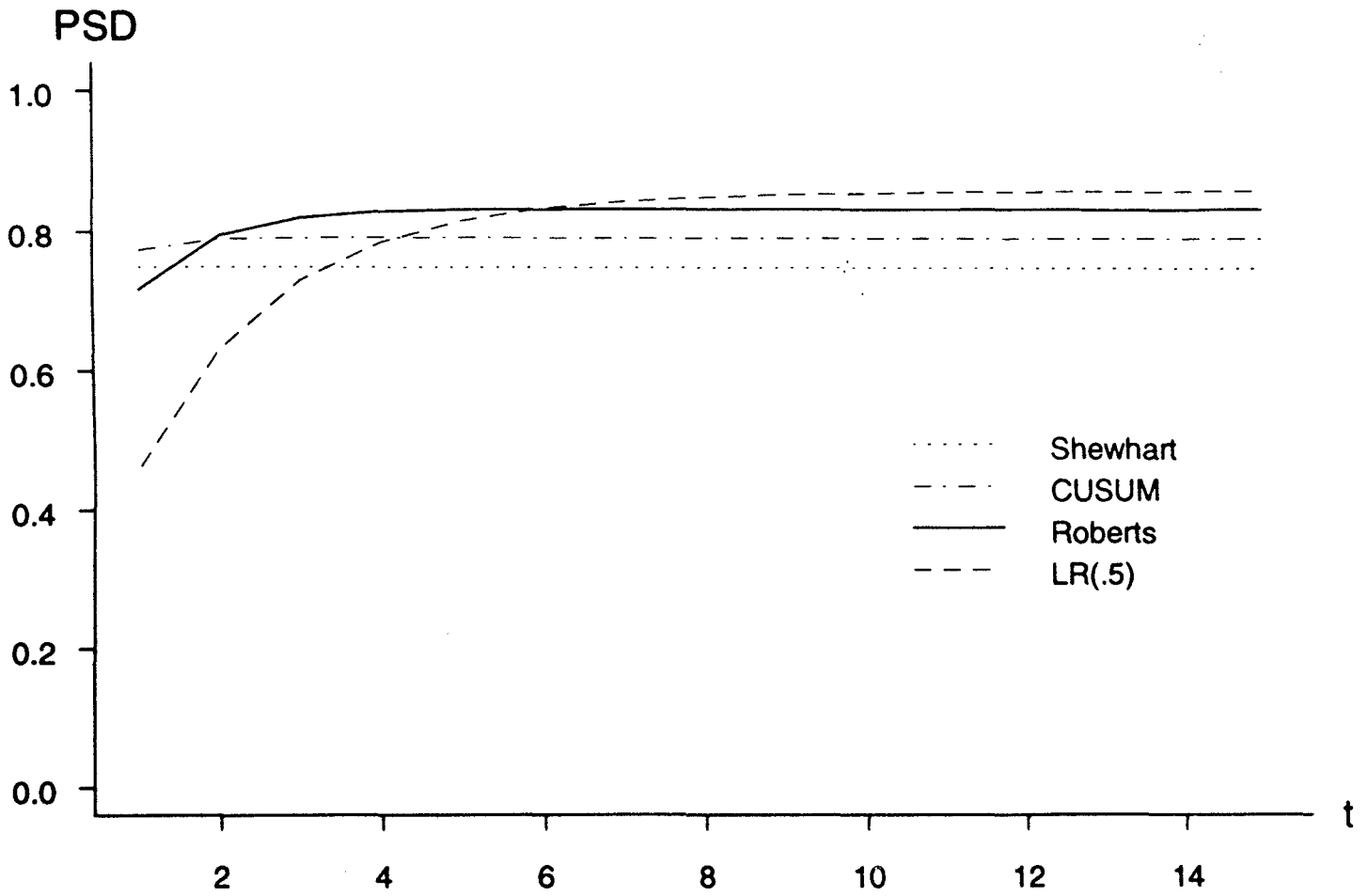
$$E(t_A - \tau \mid \tau \geq t_A)$$

Figure 7. The probability of successful detection (for d=3)

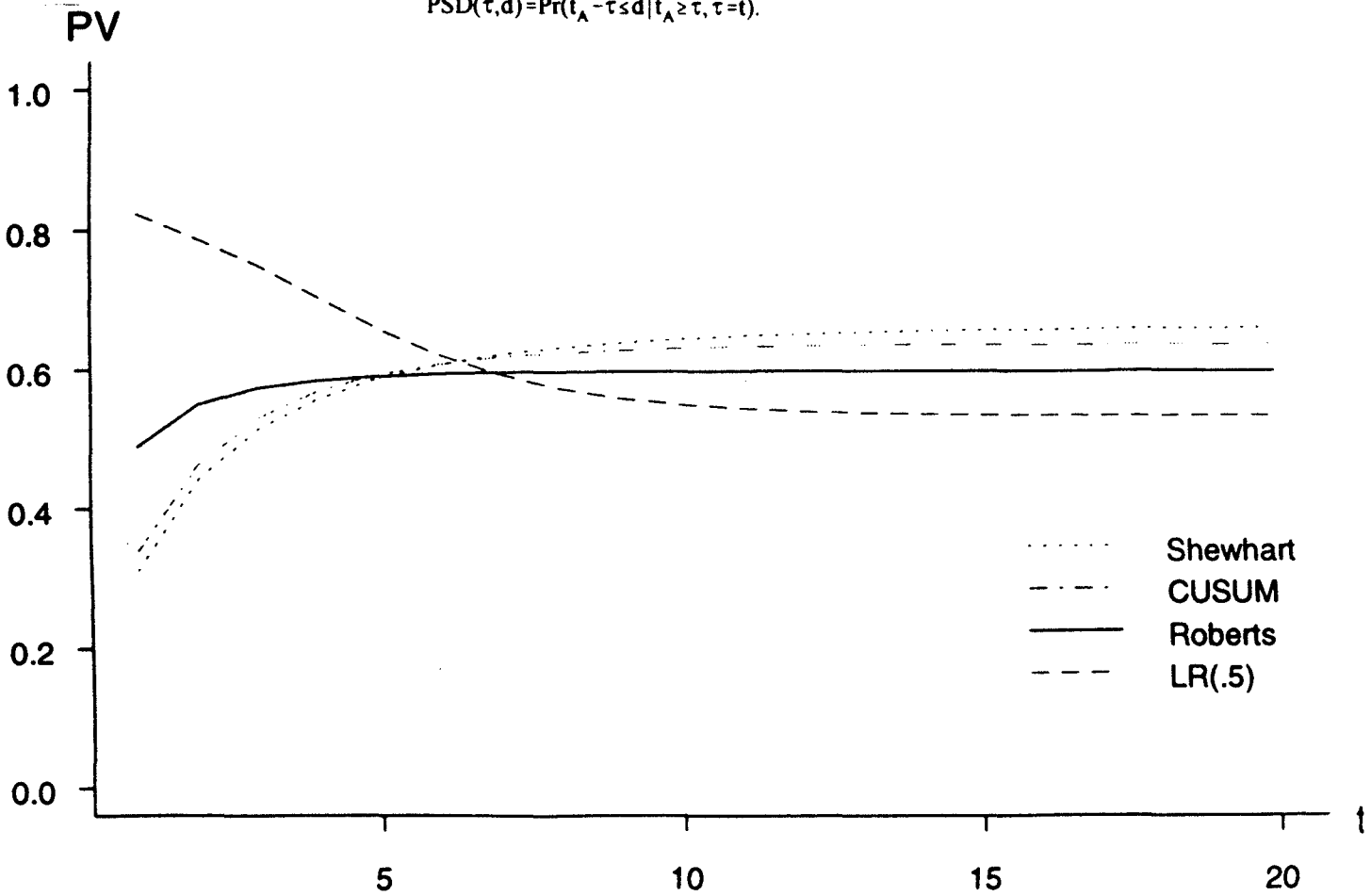$$PSD(\tau,d)=Pr(t_A-\tau \leq d | t_A \geq \tau, \tau=t).$$

Figure 8. The predictive value $PV(t) = Pr(\tau \leq t | t_A=t)$ is given for the case when τ has a geometric distribution with intensity 0.1.

Research Report

1994:1	Frisén, M	Statistical surveillance of business cycles.
1994:2	Frisén, M	Characterization of methods for surveillance by optimality.
1994:3	Frisén, M & Cassel, C	Visual evaluations of statistical surveillance.
1994:4	Ekman, C	A comparison of two designs for estimating a second order surface with a known maximum.
1994:5	Palaszewski, B	Comparing power and multiple significance level for step up and step down multiple test procedures for correlated estimates.
1994:6	Åkermo, G	Constant predictive value of an alarm.
1994:7	Åkermo, G	On performance of methods for statistical surveillance.
1994:8	Palaszewksi, B	An abstract bootstrap based step down test procedure for testing correlated coefficients in linear models.
1994:9	Ekman, C	Saturated designs for second order models.
1994:10	Ekman, C	A note on rotatability.
1995:1	Arnkelsdottir, H	Surveillance of rare events. On evaluations of the sets method.
1995:2	Sveréus, A	Detection of gradual changes. Statistical methods in post marketing surveillance.
1995:3	Ekman, C	On second order surfaces estimation and rotatability.
1996:1	Ekman, A	Sequential analysis of simple hypotheses when using play-the-winner allocation.
1996:2	Wessman, P	Some principles for surveillance adopted for multivariate processes with a common change point.
1996:3	Frisén, M. & Wessman, P	Evaluations of likelihood ratio methods for surveillance.