Research Report
Department of Statistics
Göteborg University
Sweden

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

Peter Wessman

Research Report 1996:2
ISSN 0349-8034

Mailing address:	Fax	Phone	Home Page:
Department of Statistics	Nat: 031-773 12 74	Nat: 031-773 10 00	http://www.stat.gu.se
Göteborg University	Int: +46 31 773 12 74	Int: +46 31 773 10 00	
S-411 80 Göteborg			
Sweden			

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 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.

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

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

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

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.

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

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

$t = 1, 2, \ldots$, 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 $ARL^0 := E[t_A \,|\tau = \infty]$, the ARL when no change occurs, deals with property *ii)* and $ARL^1 = E[t_A \,|\tau = 1]$ 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*:

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

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

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

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

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

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

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

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

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 $C(s) = \{\tau \le s\}$ and $D(s) = \{\tau > s\}$ has, at time s, a distribution from the family:

$$X_s \sim \mathcal{F}_s = \begin{cases} \mathcal{F}^{D(s)} = \{F^0\}^s & \tau > s \\ \mathcal{F}^{C(s)} = \left\{\bigcup_{\tau<s} (F^0)^{\tau-1} \times (F^1)^{\tau-s+1}\right\} & \tau \le s \end{cases} \tag{11}$$

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:

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

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

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:

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

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

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

(iii) Finally, since $\mathbf{lr}_1(x_s)$ is sufficient for any s, we have according to definition 4.2 that the sequence $\{\mathbf{lr}_1(x_s); s = 1, 2, \ldots\}$ 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. □

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 $\{\mathbf{lr}_1(x_s); s = 1, 2, \ldots\} = \{lr(\mathbf{x}(t)); t = 1, 2, \ldots\}$ 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:

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

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;

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

where $T_j(\mathbf{x}(t))$ 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

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

with a constant covariance matrix Σ, then it is sufficient to monitor the sequence of likelihood ratio statistics $T^{\mu}(\mathbf{x}(t)) = \left\{\mu'\Sigma^{-1}\mathbf{x}(t); t = 1, 2, \ldots\right\}$. 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

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

If instead a shift in the covariance is monitored:

$$\mathbf{X}(t) \sim \begin{cases} N_p(\mu,\Sigma_0) & \tau > t \\ N_p(\mu,\Sigma_1) & \tau \leq t \end{cases} \tag{18}$$

the minimal sufficient process to monitor would be

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

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

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

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 $\{\mathbf{lr}_1(x_s); s = 1, 2, \ldots\}$ is sufficient for detecting $C(s) = \{\tau \leq s\}$. 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.

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

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

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

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

are distributed

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

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

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

and it is distributed as

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

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

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

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

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

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

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

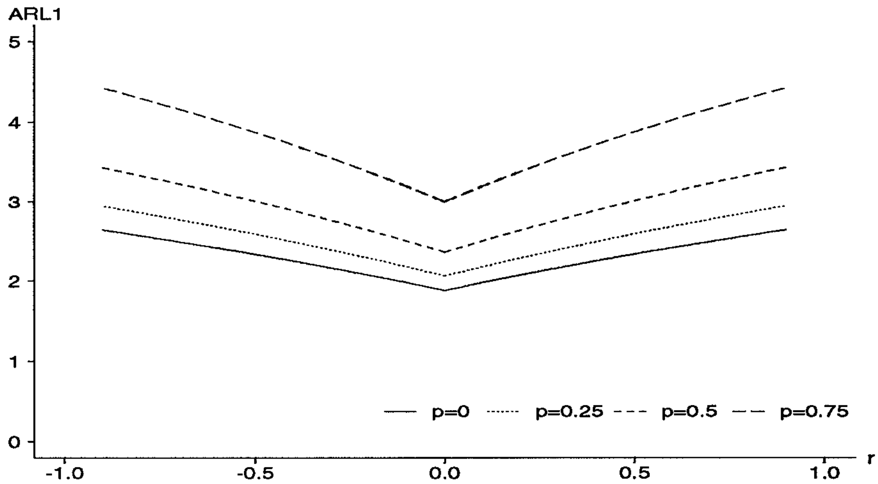

Figure 1: *The effect of the choice of critical limits, $(K_1(p), K_2(p))$, on ARL^1 for a UI Shewhart procedure. Shown for different values of the correlation, ρ.*

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, $C\left(s\right) = \{\tau = s\}$, the alarm function reduces to

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

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 $C(s) = \{\tau \leq s\}$, the method results in an alarm function that can be written as

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

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 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 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).

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.

<u>Research Report</u>

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.