Surveillance Schemes with Applications to Mass Calibration

Moshe Pollak
Carroll Croarkin
Charles Hagwood

U.S. DEPARTMENT OF COMMERCE
Technology Administration
National Institute of Standards and Technology
Computing and Applied Mathematics Laboratory
Gaithersburg, MD 20899
Surveillance Schemes with Applications to Mass Calibration

Moshe Pollak
Carroll Croarkin
Charles Hagwood

U.S. DEPARTMENT OF COMMERCE
Technology Administration
National Institute of Standards
and Technology
Computing and Applied Mathematics Laboratory
Gaithersburg, MD 20899

April 1993
Surveillance Schemes

with

Applications to Mass Calibration

Moshe Pollak*, The Hebrew University of Jerusalem
Caroll Croarkin**, National Institute of Standards and Technology
Charles Hagwood***, National Institute of Standards and Technology

Footnotes

*Work was done while visiting NIST as an ASA/NSF/NIST Fellow. Postal Address: Dept. of Statistics, The Hebrew University of Jerusalem, Mount Scopus, Jerusalem 91905

** Postal Address: NIST, Statistical Engineering Division, Div-882, Gaithersburg, MD 20899. USA.

** *Postal Address: NIST, Statistical Engineering Division, Div-882, Gaithersburg, MD 20899. USA.
## Contents

1 Abstract 1

2 Introduction 2

3 Introduction to the Shiryaev-Roberts Approach to Surveillance 5

4 Detecting a change of mean - parametric analysis of the data of Figure 2 8

5 Detecting a change of mean - nonparametric analysis of the data of Figure 2 15

6 Post-detection analysis 18

7 Surveillance of the standard deviation 29

8 Summary and Conclusion 31

9 Appendix 1. Check standard data from the mass calibration process 33

10 Appendix 2: Shewhart Chart for the data of Figure 2 38

11 Appendix 3. Rationale of the SR procedure 39

12 Appendix 4. Operating characteristics of the SR procedure: theoretical details 41

13 Appendix 5. A 2-sided SR scheme for detecting a change in a normal mean with unknown initial mean and unknown initial variance 43

14 Appendix 6. A nonparametric surveillance scheme for detecting a change 49

15 Appendix 7. Computation of \( \lim_{A \to \infty} E_{\infty} N_{A}/A \). 53
List of Figures

1. The design of the mass difference measurements. ........................................ 3
2. 217 estimates in milligrams of weight differences of two standard weights of 1 kilogram each, made at the NIST between 1975 and 1988. .............................................................. 4
3. The Shiryayev-Roberts control chart with stopping threshold $A$. 10
4. The Shiryayev-Roberts control chart with stopping threshold $A = 220$. ................................. 11
5. Parametric surveillance for a change of the mean of the mass difference data: $R_i, 1 \leq i \leq N_{220} = 23$. ................................. 12
6. Parametric surveillance for a change of the mean of the mass difference data: $R_i, 1 \leq i \leq N_{500} = 40$. ................................. 13
7. Parametric surveillance for a change of the mean of the mass difference data: $R_i, 1 \leq i \leq N_{6000} = 162$. ................................. 14
8. Parametric surveillance for a change of the mean of the mass difference data: $R_i, 1 \leq i \leq 217$. ................................. 14
9. Nonparametric surveillance for a change of the mean of the mass difference data: $R_i, 1 \leq i \leq N_{210} = 42$. ................................. 16
10. Nonparametric surveillance for a change of the mean of the mass difference data: $R_i, 1 \leq i \leq 161$. ................................. 17
11. Nonparametric surveillance for a change of the mean of the mass difference data: $R_i, 1 \leq i \leq 217$. ................................. 17
12. Post-detection parametric analysis of mass difference data: surveillance for a change of mean. ARL to false alarm=370, starting anew after each observation. (a) $R_i, 24 \leq i \leq N_{220} = 74$. (b) $R_i, 75 \leq i \leq N_{220} = 113$. (c) $R_i, 114 \leq i \leq N_{220} = 164$. (d) $R_i, 165 \leq i \leq 217$. ................................. 19
13. Post-detection nonparametric analysis of mass data: surveillance for a change of mean. ARL to false alarm=370, starting anew after each observation. (a) $R_i, 43 \leq i \leq N_{210} = 60$. (b) $R_i, 61 \leq i \leq N_{210} = 114$. (c) $R_i, 115 \leq i \leq N_{210} = 161$. (d) $R_i, 162 \leq i \leq 217$. ................................. 20
14. Surveillance segments of mass difference data, based on parametric control for change of mean. ARL to false alarm=370. Putative change= 1 standard deviation. ........................................ 21

iii
Surveillance segments of mass difference data, based on non-parametric control for change of mean. ARL to false alarm=370. Putative change= 1 standard deviation.

Post-detection parametric analysis of mass difference data also using pre-detection data for surveillance for a change in mean: $\Lambda_p^*$ as a function of $k$.

Post-detection parametric analysis of mass difference data: surveillance for a change in mean also using pre-detection data. ARL to false alarm=370.

Post-detection nonparametric analysis of mass difference data also using pre-detection data for surveillance for a change in mean: $\Lambda_p^*$ as a function of $k$.

Post-detection nonparametric analysis of mass difference data: surveillance for a change in mean also using pre-detection data. ARL to false alarm=370.

217 estimates in milligrams of the standard deviation, $\sigma_x$ of the difference between two 1 kilogram weights, made at NIST between 1975 and 1988.

Parametric analysis of mass difference data: surveillance for a change of standard deviation. ARL to false alarm=370, starting anew after each detection. (a) $R_i, 1 \leq i \leq N_{140} = 47$. (b) $R_i, 48 \leq i \leq N_{140} = 177$. (c) $R_i, 178 \leq i \leq N_{140} = 207$. (d) $R_i, 208 \leq i \leq 217$.

Parametric surveillance for a change of standard deviation, starting with observations #178, deleting observation #207; ARL to false alarm=370, $R_i, 178 \leq i \leq 216$.

A Shewhart chart for the data of Figure 2.

MATLAB-language computer program for computing $R_n$ for the parametric SR scheme of section 2.

MATLAB-language computer program for computing $R_n$ for the a (one-sided) nonparametric SR scheme for detecting an increase.

Autocorrelation plots for mass difference data: observations 166—217, with observation 207 deleted.

Detail of the mass difference data: observations 166—217, with observation 217 deleted.
29 Autocorrelation plots for residual standard deviations transformed to \( N(0,1) \) variables: (a) observations 51 – 106, (b) observations 107 – 150, (c) observations 51 – 150 (d) observations 41 – 165. .................................................. 59

30 Detail of the mass difference data: (a) observations 51 – 106 (b) observations 41 – 165. .................................................. 60

31 Autocorrelation plot for residual standard deviations transformed to \( N(0,1) \) variables: observations 166 – 217 with observation 207 deleted. .................................................. 60
Surveillance Schemes
with
Applications to Mass Calibration

Moshe Pollak, The Hebrew University of Jerusalem*
Carroll Croarkin, National Institute of Standards and Technology
Charles Hagwood, National Institute of Standards and Technology

1 Abstract

One of the activities at the NIST is calibrating mass standards. In order to ensure the quality of calibration, the NIST personnel monitor the values of check standards over time. The current standard surveillance technique is a Shewhart control chart with $3\sigma$-limits.

Here we explore the applicability of other, recently developed, control charts. While Shewhart charts are typically designed to detect large changes, the schemes regarded here are geared towards detecting medium-sized ones. Some of these procedures are parametric, others are nonparametric. They are applied here to a sequence of measurements of mass standards, made at the NIST over a period of time. Two types of surveillance problems are regarded: monitoring for a change in mean and monitoring for a change of standard deviation. The control charts considered are shown to be effective.

*ASA/NSF/NIST Fellow

1Key words and subject class: control charts, Shirvayev-Roberts, Cusum, mass standards
2 Introduction

One of the activities of the National Institute of Standards and Technology (NIST) is precision measurement of mass standards. Mass standards are calibrated at the NIST by comparison measurements which relate the mass of a client’s standard to the NIST standard kilograms, and thence to the defined unit for mass, the Paris kilogram.

The NIST has a large stake in monitoring its calibration process to ensure that the tie to the unit of mass, as quantified by the NIST statement of uncertainty, is maintained. Any significant change in this process, whether it is caused by changes in the masses of the NIST kilograms or changes in the operation of the calibration process itself, invalidates the NIST uncertainty. The check on the validity of this process is maintained by a series of check standards which are calibrated with the client’s weights.

The kilogram level is the critical level in the calibration process because weights of higher and lower denominations are calibrated relative to the NIST working kilograms through a series of intercomparison designs. Prior to 1989, the same two kilograms were used for calibration purposes, and the check standard was the measured difference between these two kilograms as estimated from a comparison design. In this report, we present an analysis of these check standard determinations made at a sequence of (nonequally spaced) time points between 1975 and 1988. The data base includes all check standard determinations which were made in the process of calibrating weight sets of 1000, 500, 300 and 200 g denominations during that period. The design pertaining to the data is illustrated in Figure 1.

The calibration design involves 6 intercomparison measurements as follows: \(y_1\) = the difference between the two the NIST 1 kg standards, \(y_2\) = the difference between one of the NIST’s 1 kg standards and the client’s, \(y_4\) = the difference between the NIST’s other 1 kg standard and the client’s, \(y_3\) = the difference between the NIST’s 1 kg standard and the sum of the client’s 500, 300, 200 g standards, \(y_5\) = the difference between the NIST’s other 1 kg standard and the sum of the client’s 500, 300, 200 g standards, and \(y_6\) = the difference between the client’s 1 kg standard and the sum of the 500, 300, 200 g standards. (See Jaeger and Davis, 1984).
Figure 1: The design of the mass difference measurements.

One can write $y_1 = \mu_1 + \epsilon_1, y_2 = \mu_2 + \epsilon_2, y_3 = \mu_3 + \epsilon_3, y_4 = \mu_2 - \mu_1 + \epsilon_4, y_5 =$
\[ \mu_3 - \mu_1 + \epsilon_5, y_6 = \mu_3 - \mu_2 + \epsilon_6 \] where the \( \epsilon_j \) are independent, and identically distributed and each \( \mu_i \) is a mass difference. The standard assumption is that the \( \epsilon_j \) have a \( N(0, \sigma^2) \) distribution, \( \sigma^2 \) unknown. The least squares estimates of \( \mu_1 \) (calculated separately at each point in time) are depicted in Figure 2.

![Figure 2: 217 estimates in milligrams of mass differences of two standard weights of 1 kilogram each, made at the NIST between 1975 and 1988.](image)

The mass differences are measured in milligrams. Appendix 1 contains the numerical data. (We will henceforth refer to these estimates of \( \mu_1 \) as mass difference data.)

Post facto, it seems clear that a change occurred a short time after the 150th observation with perhaps a few local fluctuations before. A second glance suggests that the change is an increase in mean, of the order of magnitude of one standard deviation. Shewhart charts (cf Shewhart, 1931, or
Alner and Keller, 1977), designed for detecting larger changes, failed to notice this increase, and it was only discovered after its occurrence by a nonroutine retrospective reappraisal. (See Appendix 2.) Our goal here is to construct monitoring schemes which would have discovered this change within a reasonably short time after its occurrence.

In this report, we describe applications of recently developed surveillance methods to these and other related data. In order to enable the reader fast access to the application, we start with a minimal technical description of the Shirayev-Roberts approach to surveillance, leaving a theoretical justification to the later sections. The paper is therefore organized in the following way. After presenting the Shirayev-Roberts approach, we apply it to the mass data portrayed in Figure 2, first parametrically and then nonparametrically. Next we investigate data related to the standard deviation of the measurements. Only after these applications do we return to more detailed explanations of the theoretical considerations involved.

3 Introduction to the Shirayev-Roberts Approach to Surveillance

The classical surveillance problem consists of being able to view sequentially a series of independent observations $X_1, X_2, X_3, \ldots$ such that $X_1, X_2, \ldots, X_{\nu-1}$ have distribution $F_0$ which changes at an unknown time $\nu$, so that $X_\nu, X_{\nu+1}, \ldots$ have distribution $F_1$. One applies a surveillance scheme which raises an alarm at time $N$, declaring that a change is in effect. Typically $N$ is a random variable; it is a stopping time, directed by the past-to-present observations when to stop and raise an alarm. A surveillance scheme is considered good if it detects a true change quickly, yet seldom raises a false alarm.

We will denote the probabilistic setup described above by $P_\nu$. Expectation will be denoted by $E_\nu$. Probability and expectation when there is no change throughout the sequence will be denoted by $P_\infty$ and $E_\infty$, respectively.

Every reasonable detection scheme may give rise to false alarms. The rate of false alarms is usually characterized by the index $E_\infty N$, the average run
length (ARL) to false alarm. The standard constraint regarding false alarms is that \( N \) satisfy
\[
E_\infty N \geq B
\]  
where \( B \) is a prespecified constant. (For example, in a problem of surveillance of a sequence of independent normally distributed observations for a change of mean, \( E_\infty N \) for the one-sided 3\( \sigma \)-limit Shewhart control chart is \( \frac{1}{1-\Phi(3)} \approx 740 \), and the two-sided chart has \( E_\infty N \approx 370 \).

The speed of detection of a surveillance scheme is typically an expression of the expected delay. A common index is
\[
\sup_{1 \leq \nu \leq \infty} E_\nu(N - \nu \mid N \geq \nu).
\]

A basic statistic when conducting a surveillance is
\[
\Lambda_k^n = \frac{f_{\nu=k}(X_1, \ldots, X_n)}{f_{\nu=\infty}(X_1, \ldots, X_n)}
\]
which is the likelihood ratio of the observations until time \( n \), for \( \nu = k \) versus \( \nu = \infty \). Cusum procedures (page, 1954; van Dobben de Bruyn, 1968) are actually maximum likelihood procedures (Lorden, 1971); a Cusum scheme can be defined as computing the sequence of statistics
\[
M_n = \max_{1 \leq k \leq n} \Lambda_k^n
\]
and raising an alarm the first time that \( M_n \) crosses a level \( A \); that is
\[
N_A = \min\{n \mid M_n \geq A\}.
\]

Given \( B \), the threshold \( A \) must be such that (1) is satisfied.

The Shiryayev-Roberts procedure (Shiryayev, 1963, and Roberts, 1966, and hence SR) is somewhat different; it requires computing the sequence of statistics
\[
R_n = \sum_{k=1}^{n} \Lambda_k^n
\]
and raising an alarm the first time that \( R_n \) exceeds a threshold \( A \); that is
\[
N = N_A = \min\{n \mid R_n \geq A\}.
\]
Again, given $B$, the threshold $A$ must be such that (1) is satisfied.

Both Cusum and SR have optimality properties in terms of speed of detection (Pollak, 1985; Moustakides, 1986; Ritov, 1990), and the differences between their performances are usually marginal (Shiryayev, 1963; Pollak and Siegmund, 1985). An advantage of SR is that it can handle dependent data much more easily than Cusum procedures. If the data are not independent, understanding the sequence of Cusum statistics becomes very complicated and standard tables become useless. For the SR technique, one can show fairly generally that

$$E_{\infty}N_A \geq A.$$  

This means that setting $A = B$ satisfies (1), with no further complications. This is true even when the observations are dependent, a case which is of wide interest, as we shall soon show. Obviously, setting $A = B$ is somewhat conservative. Often it is possible to show the existence of a constant $C$ such that

$$\lim_{n \to \infty} \frac{E_{2n}N_A}{A} = C$$

so that setting $A = \frac{B}{C}$ satisfies (1) approximately. Computation of $C$ is usually not a hard problem.

To see where the technical point concerning dependent observations makes a real difference, consider the data of Section 1. Even if we assume that the observations are normal, there is no baseline and there is no knowledge of the standard deviation. In other words, we are observing a sequence $X_1, X_2, \ldots$ where before a change $X_i \sim N(\mu_0, \sigma^2)$, and we are concerned that this may change to a $N(\mu_0 + \sigma, \sigma^2)$ or a $N(\mu_0 - \sigma, \sigma^2)$ distribution. Neither $\mu_0$ nor $\sigma$ are known. Therefore one cannot compute $\Lambda^*_n$ (since both pre- and post-change densities are unknown), and standard surveillance theory (Shewhart, Cusum, Shiryayev-Roberts) cannot be applied to the sequence $X_1, X_2, \ldots$ (Estimating the unknown parameters and applying Cusum is a tricky proposition. If there is a change, the estimates may be influenced by it. In addition, estimates have standard errors; Cusum techniques are notoriously sensitive to misspecification of parameters. See van Dobben de Bruyn, 1968, Section 2.4.
The following technique (Pollak and Siegmund, 1991) circumvents these difficulties. Let $\hat{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ and construct the sequence of standardized recursive residuals (Brown, Durbin and Evans, 1975)

$$Y_i = (X_i - \hat{X}_{i-1}) \sqrt{\frac{i-1}{i}}; \quad i = 2, 3, \ldots$$

The distribution of the sequence of $Y_i$'s is independent of $\mu_0$. Now construct the sequence

$$Z_i = \frac{Y_i}{Y_2}; \quad i = 3, 4, 5, \ldots$$

The distribution of the sequence of $Z_i$'s is independent of both $\mu_0$ and $\sigma$. If, instead of monitoring the process of $X_i$'s we monitor the sequence of $Z_i$'s, the pre-change and post-change densities of our observations (the $Z_i$'s) are completely specified. It is therefore possible to compute the likelihood ratios $\Lambda_n^k$ (for the $Z_i$ series). The technical difficulty in applying a Cusum control chart is that the $Z_i$'s are not independent. On the other hand, as mentioned above, for the SR procedure this is no obstacle.

The same technique can be applied to many other surveillance problems. Essentially, the idea is to get rid of nuisance parameters by exploiting structures of invariance inherent to the problem. Thus, one can handle surveillance for a change in a standard deviation, for a change in the parameter of an exponential distribution, as well as nonparametric problems, even when there is no in-control baseline.

In the following sections we will apply such schemes, both parametric and nonparametric, to detecting a change of mean and to detecting a change of standard deviation.

### 4 Detecting a change of mean - parametric analysis of the data of Figure 2

We continue with the notation of the previous section. We first write down the statistics $\Lambda_n^k$, the change at $\nu = k$ versus the no-change ($\nu = \infty$) likelihood ratio of the $Z_i$ values of the first $n$ observations, i.e. of $Z_3, \ldots, Z_n$. 
First note that $\Lambda_1^\nu \equiv 1$. The reason for this is that if the change is in effect at onset - that is, from the beginning all $X_i$ are distributed $N(\mu_0 + \sigma, \sigma^2)$ (or $N(\mu_0 - \sigma, \sigma^2)$) - then this will not be noticeable, as there is no baseline for comparison, and the distribution of the $Z_i$ sequence will be the same as when there is no change at all. (Another way of looking at this is that if all of the observations are $N(\mu_0 + \sigma, \sigma^2)$, then there is no change.) Hence, the likelihood when $\nu = 1$ is the same as when $\nu = \infty$, and so $\Lambda_1^\nu \equiv 1$.

Since the $Z_i$'s start with $i = 3$, $\Lambda_2^3 = 1$. In Appendix 5 it is shown that for $k \geq 3, n \geq k \geq 2$

$$\Lambda_k^n = \frac{\int_{-\infty}^{\infty} |v + a_{k,n}|^{n-2} e^{-\frac{1}{2}v^2} dv}{\int_{-\infty}^{\infty} |v|^{n-2} e^{-\frac{1}{2}v^2} dv} e^{-\frac{1}{2}(k-1)^2 \left[ \frac{1}{k+1} - \frac{1}{k+2} \right] + \frac{1}{2} a_{k,n}}$$

where $I(\cdot)$ is the indicator function of the set $(\cdot)$, $Z_2 = 1$ and

$$a_{k,n} = \frac{(k-1) \sum_{i=k}^{n} \frac{Z_i}{i(i-1)}}{\sqrt{\sum_{i=2}^{n} Z_i^2}}.$$

As detailed in Appendix 5, the ratio of the integrals in $\Lambda_k^n$ can be computed by a recursion formula, and a computer program (Figure 24) can calculate the sequence of $R_n$'s. This allows one to construct a control chart by plotting the points $(i, R_i)$ on a plane, with $i$ the values on the $x$-axis and $R_i$ on the $y$-axis (Figure 3).

In order to specify a stopping rule, one must specify $B$ and set $A$ so that (1) is satisfied. The meaning of $B$ is the ARL to false alarm. Specifying $B$ requires consideration, as we shall see in the sequel. For the sake of a first example, suppose the alternative to a Shirayayev-Roberts procedure would be a 2-sided $3\sigma$-limit Shewhart chart. As mentioned above, the ARL to false alarm of that procedure is $\approx 370$. Hence we should specify $B = 370$ for the Shirayayev-Roberts procedure. There being an average of about 17 observations per year, $B = 370$ means that it will take an average of $370/17 \approx 22$ years to raise a false alarm.

As mentioned above, a conservative way to satisfy (1) is to set $A = B = 370$. An approximate equality in (1) may be obtained by computing the
Figure 3: The Shiryayev-Roberts control chart with stopping threshold $A$.

The limit as $A \to \infty$ of $E_{\infty}N_A/A$. By employing Theorem 1 of Gordon and Pollak (1990) and Theorem 1 of Pollak (1987), one would expect that

$$\lim_{A \to \infty} \frac{E_{\infty}N_A}{A} = 1.7$$

(see Appendix 7 in the sequel). Hence setting $A = 370/1.7 = 220$ satisfies (1) approximately. The control chart will be as in Figure 4.

One should let the process keep going as long as all the points $(i, R_i)$ are such that $R_i < A$. One should stop the process the first time $n$ that $R_n \geq A$, and declare that a change had taken place.

Applying this control scheme to the data of Figure 2 yields Figure 5. We would stop right after the 23rd observation and declare that a change is in effect. Returning to reappraise Figure 2, the decision doesn't look
Figure 4: The Shiryayev-Roberts control chart with stopping threshold $A = 220$.

What would have happened had we chosen a larger value of $B$? Suppose we were willing to risk one false alarm every 50 years, leading to $B = 850$ ($A = 500$). Figure 6 gives this picture. It would have taken another 17 observations to reach the conclusion that a change is in effect.

To what level would $B$ have had to be set in order to altogether miss calling a change in the first part of the series? Figure 7 gives a plot of $R_1$ for the first 162 observations. It seems as if $B$ would have had to be about 10000. (This way $A \approx 10000/1.7 \approx 5829 = R_{50}$.) In that case, a change would have been declared to be in effect after observation 162.

To complete the picture, the entire $R_i$ sequence is given in Figure 8.
Figure 5: Parametric surveillance for a change of the mean of the mass difference data: $R_i, 1 \leq N_{220} = 23$.

What value of $B$ should one choose? There are a number of ways of going about choosing $B$. One way is to set $B$ directly as the lowest tolerable ARL to false alarm. For instance, if Shewhart’s classical specification seems reasonable, one should set $B = 370$ or $B = 740$ for a one or two-side surveillance scheme, respectively. Sometimes the value of the lowest tolerable ARL to false alarm is nebulous. In that case, another way to set $B$ is by regarding the post-change characteristics of $N_A$. For instance, if one regards changes of the type appearing early in the Figure 2 sequence as serious, one would not set $B$ above 10000. If one were to regard such changes as mere local fluctuations which should not set off an alarm, then one would fix $B > 10000$. To fix $B$ more precisely, one can regard the expected delay $E_{\nu}(N - \nu | N \geq \nu)$. If the detection scheme is geared to detect a change of $\delta \sigma$ in the mean (i.e. the observations change from $N(\mu_0, \sigma^2)$ to $N(\mu_0 + \delta \sigma, \sigma^2)$ or $N(\mu_0 - \delta \sigma, \sigma^2)$)
Figure 6: Parametric surveillance for a change of the mean of the mass difference data: \( R_n, 1 \leq N_{500} = 40 \).

and the true change in mean is \( \mu \sigma \), then a first-order approximation to the expected delay \( E_{\nu}(N_{A} - \nu \mid N_{A} \geq \nu) \) (for \( \nu \) not close to 1 and \( \mu > \delta/2 \)) is \( (\log A) / [\delta(\mu - \delta/2)] \). In the case contemplated above, \( \delta = 1 \). If \( \mu = 1 \) (i.e. the true change equals the putative one), the expected delay is very roughly \( 2\log A \). So, if one would tolerate an expected delay of, say, no more than 10 observations, it would mean setting \( A \approx 150 \), or \( B \approx 260 \). See Pollak and Siegmund (1991) for a more precise picture.

Finally, from the discussion above there emerges a data-analytic aspect of the SR control chart not enjoyed by the other control charts. (See Kenett and Pollak, 1992.) One may regard a present value of \( R_n \) in light of the following question: what level would \( B \) have had to be in order for an alarm to be raised at time \( n \)? The answer is (approximately) \( 1.7R_n \). Thus, up to a multiplicative constant, the height of a point on a SR control chart has a simple data-analytic meaning, similar in vein to that of a \( p \)-value.
Figure 7: Parametric surveillance for a change of the mean of the mass difference data: $R_i, 1 \leq N_{6000} = 162$.

Figure 8: Parametric surveillance for a change of the mean of the mass difference data: $R_i, 1 \leq i \leq 217$. 

14
Detecting a change of mean - nonparametric analysis of the data of Figure 2

The analysis in the previous section was based on the assumption that the observations \( X_i \) are normally distributed. Not always is the distribution of the observations known. Even when it is, there may be concern that some of the observations are contaminated. In such cases, nonparametric schemes are of interest. The decision to go nonparametric need not be difficult, as the efficiency of some of the nonparametric procedures is very high.

The natural analogue of the \( Z_i \) of the previous section is the \( i \)th sequential rank \( r_i = \sum_{j=1}^{i} I(X_j \leq X_i) \), which is the rank of the \( i \)th observation among the first \( i \) observations. Surveillance will be based on the sequence \( r_1, r_2, r_3, \ldots \) instead of \( X_1, X_2, X_3, \ldots \). A nonparametric Shiryaev-Roberts (NPSR) procedure will be based on the sequence \( R_n = \sum_{k=1}^{n} \Lambda_k^\nu, n = 1, 2, \ldots \), where \( \Lambda_k^\nu \) is a \( \nu = k \) versus \( \nu = \infty \) likelihood ratio of \( r_1, \ldots, r_n \). The technical details are relegated to Appendix 6, where a recipe for choosing an appropriate nonparametric scheme is given, along with a program for computing the statistics \( R_n \).

We return to the analysis of the previous section. Suppose one is not quite sure about the distribution of the observations, but surmises a normal distribution. Then, following the recipe in Appendix 6, the parameters of the NPSR procedure should be \( p = .8413, \alpha = .53, \beta = 1.7 \).

Here \( \lim_{A \to \infty} E_\infty N_A/A = 1.89. \) (For lower values of \( A, 1.8 \) will be a better approximation than 1.89. See Gordon and Pollak, 1991.) Thus, if \( B = 370 \) (to make things comparable to the analysis of the previous section), \( A \) should be set to \( B/1.8 \approx 210 \). Figure 9 is the nonparametric analog of Figure 5.

Clearly, the NPSR procedure with \( B = 370 \) does not catch the apparent rise starting at the 15th - 20th observations. This is typical of early changes: the nonparametric technique is weaker than the parametric procedure if the change is early (within the first 30 observations). Here, it catches a change only at the 42nd observation. (It should be borne in mind that the parametric
procedure with $B = 370$ barely detects a change; the parametric $R_i$ sequence barely exceeds $A$ before the 39th observation.) Figure 10 is a nonparametric analog of Figure 7.

The configuration is similar, but the larger values of $R_i$ are usually lower for the nonparametric scheme. This is only natural: for instance, increasing the largest observation will not affect its rank, thereby not affecting the NPSR statistic, but affecting the parametric one. In other words, the NPSR scheme is less sensitive to extreme observations. (This is also the reason why the $R_i$ of the NPSR scheme have a larger "blip" at $65 \leq i \leq 95$; comparison of the ranks of $X_i$ makes the $65 \leq i \leq 95$ observations seem larger with respect to the $20 \leq i \leq 50$ observations than comparison of the actual values.) So, as usual, the nonparametric scheme is less sensitive to extreme values and is therefore robust, but it is a little slower in detecting a change (about 7% slower on the average, see Gordon and Pollak, 1991). To complete the picture, Figure 11 is a nonparametric analog of Figure 8.

![Graph](image_url)

Figure 9: Nonparametric surveillance for a change of the mean of the mass difference data: $R_i, 1 \leq i \leq N_{210} = 42$. 

16
Figure 10: Nonparametric surveillance for a change of the mean of the mass difference data: $R_i$, $1 \leq i \leq 161$.

Figure 11: Nonparametric surveillance for a change of the mean of the mass difference data: $R_i$, $1 \leq i \leq 217$. 
6 Post-detection analysis

Suppose the series $R_t$ crosses over the threshold $A$. What should one do next? In principle, one should take stock of one's new position, get one's bearings and continue. However, in practice these actions mean different things in different situations.

In some industrial contexts, crossing the threshold $A$ causes a machine to be replaced or overhauled, and surveillance will have to be completely restarted, disregarding all previous observations. Clearly, surveillance resumes under completely new circumstances. For example's sake, suppose we adopt this attitude towards the data of Figure 2.

If our scheme is the parametric setup of Section 3, then after the 23rd observation we will discard the first 23 observations, and reapply the same surveillance scheme (with $A = 220$) to the sequence starting with observation 24. As it turns out, we will stop after the 74th observation. Reapplying the scheme starting with the 75th observation we will stop after the 113th observation. The next stop will be after the 164th observation. After that, we will not stop again before the end of the data. The control charts are detailed in Figure 12. The resulting segmentation of the data is illustrated in Figure 14.

Similarly, if we do this for the nonparametric scheme of Section 4 with $A = 210$, the stopping times are after the 60th, 114th and 161st observations. The control charts are detailed in Figure 13. The resulting segmentation of the data is illustrated in Figure 15.

In the mass difference example discussed in the previous sections, crossing the level $A$ may bring about a recalibration of the two NIST working standards. The measuring process may continue without change only if the statement of uncertainty is expanded to account for the fact that one or both of the working standards may have changed.

Therefore, there is information in the most recent observations concerning the present level of the process; which is indicative of the level to be expected of the mass differences in future observations. Surveillance should
Figure 12: Post-detection parametric analysis of mass difference data: surveillance for a change of mean. ARL to false alarm = 370, starting anew after each observation. (a) $R_i, 24 \leq i \leq N_{220} = 74$. (b) $R_i, 75 \leq i \leq N_{220} = 113$. (c) $R_i, 114 \leq i \leq N_{220} = 164$. (d) $R_i, 165 \leq i \leq 217$. 
Figure 13: Post-detection nonparametric analysis of mass data: surveillance for a change of mean. ARL to false alarm=370, starting anew after each observation. (a) $R_4, 43 \leq i \leq N_{210} = 60$. (b) $R_4, 61 \leq i \leq N_{210} = 114$. (c) $R_4, 115 \leq i \leq N_{210} = 161$. (d) $R_4, 162 \leq i \leq 217$. 
Figure 14: Surveillance segments of mass difference data, based on parametric control for change of mean. ARL to false alarm=370. Putative change= 1 standard deviation.

Figure 15: Surveillance segments of mass difference data, based on non-parametric control for change of mean. ARL to false alarm=370. Putative change= 1 standard deviation.
now be geared towards detecting a change from the new level. Of course, one can forget the past and act as if no past exists, as in the first part of this section. This, however will make detection of an early change more difficult. It would intuitively make sense to make use of the past observations most recent to the detection time.

A strong word of caution is in order here. Though it is very tempting to use such information, there are formidable technical difficulties involved. For one thing, there's always the possibility that the detection was actually a false alarm. If this possibility cannot be ruled out, any attempt to use the information prior to detection will stand on shaky ground. Even if one is confident that a real change is in effect, there remains the question of how to make use of this information. At the time of this writing, there is no clear cut recipe of how best to estimate the point of change after a detection has been made (and how to estimate the present level), and there is virtually no discussion in the literature of how to make use of such estimates should they be available. (See Kenett and Zacks, 1992, for a Bayesian approach. See James, James and Siegmund, 1987, for estimation in the fixed sample retrospective change point problem. See Siegmund and Venkatraman, 1992, for the only paper to date dealing with estimation in the sequential case from a non-Bayesian point of view.)

Had one known the point of change, one could have regarded the post-change observations made until the time of detection as constituting a learning sample; and one could have continued with a modified Shirayev-Roberts procedure along the lines of Pollak and Siegmund (1991). Since the point of change must be estimated, at present no method is known which will produce a surveillance procedure which utilizes the pre-detection observations and honestly satisfies (1).

Nonetheless, we will now present an analysis wherein we estimate the point of change, and continue with a modified Shirayev-Roberts procedure, as suggested above. We conjecture that (1) is satisfied approximately. The reasoning: \( \Lambda_k^* \) for small \( k \) do not play a dominant role in \( R_n = \sum_k \Lambda_k^* \), when \( R_n \) is large, (see Gordon and Pollak, 1991 and Figures 16 and 18 in the sequel). It must be reemphasized that at present we have no proof of this. We hope to work on this in the future.
Our estimate $\hat{\nu}$ of the change point time $\nu$ is a maximum likelihood type estimate. Consider Figure 5 and the accompanying discussion. Our threshold is $A = 220$, and we stop after the 23rd observation. Here $\hat{\nu} = 17$. (See Figure 16(a).) Instead of starting surveillance anew from the 24th observation, we delete only the first 16 observations. Now, the "first" observation is observation 17 in the original chronology. Detection was made on the seventh observation in the new count, and the first future observation will be (new) 8. We will regard $X_1, X_2, \ldots, X_7$ (in the new count; these are the old $X_{17}, X_{18}, \ldots, X_{23}$) as a "learning sample", all of which have the same (post-first-change) distribution.

The new $Z_i$ are computed in the same way as the old ones; they will be based on $X_1, X_2, X_3, \ldots$ (of the new chronology). The likelihood ratios $\Lambda_k^n$ are calculated accordingly.

Under our assumptions, there is no (second) change prior to the 8th observation. Therefore, $\Lambda_k^n$ will not be meaningful for $k < 8$; the change point $\nu$ cannot have a value less than 8 in our present circumstances. Therefore, the statistic $R_n$ will now be

$$ R_n = \sum_{k=8}^{n} \Lambda_k^n $$

and we will stop and declare that a change is in effect at $N_{220}$, the first time that $R_n$ exceeds 220. The resulting control chart is given in Figure 17(a). (In order to facilitate reference to Figure 2 and Appendix 1, the index $i$ of the $R_i$ is translated back again to match the original serial numbers of the observations. Thus the first $R_i$ to be depicted in Figure 17(a) is not denoted as $R_8$ but as $R_{24}$. The fact that the first post-learning-sample observation is $X_8$ is coded by the input $kay = 8$. We stop after the 63rd (original count) observation.

From Figure 16(b), we obtain $\hat{\nu} = 51$. Since the next observation will be the 64th, it means that to continue we should set $kay = 14$. The resulting analysis is given in Figure 17(b) and Figure 16(b). We stop again after the 113th observation, and $\hat{\nu} = 107$. 
Therefore, for the continuation $kay = 8$. The analysis is given in Figure 17(c) and Figure 16(c). We stop again after the 164th, and $\hat{v} = 151$.

Hence, $kay = 15$ for the continuation. The analysis is given in Figure 17(d). There is no further change detected by the time of the 217th observation. (For the sake of completeness and comparison, $\Lambda_{217}^{k}$ is presented in Figure 16(e).)

The same type of analysis can be made with the nonparametric approach. The progression of analyses is given in Figures 18 and 19.
Figure 16: Post-detection parametric analysis of mass difference data also using pre-detection data for surveillance for a change in mean: $A^*_k$ as a function of $k$.

(a) $1 \leq k \leq 23; \bar{\nu} = 17; n = N_{230} = 23.$

(b) $17 \leq k \leq 63; \bar{\nu} = 51; n = N_{230} = 63.$

(c) $51 \leq k \leq 113; \bar{\nu} = 107; n = N_{230} = 113.$

(d) $107 \leq k \leq 164; \bar{\nu} = 151; n = N_{230} = 164.$

(e) $151 \leq k \leq 217; n = 217.$
Figure 17: Post-detection parametric analysis of mass difference data: surveillance for a change in mean also using pre-detection data. ARL to false alarm=370.

(a) $R_i, 1 \leq i \leq 23; \tilde{\nu} = 17$.
(b) $R_i, 24 \leq i \leq 63; kay = 8; \tilde{\nu} = 51$.
(c) $R_i, 64 \leq i \leq 113; kay = 14; \tilde{\nu} = 107$.
(d) $R_i, 114 \leq i \leq 164; kay = 8; \tilde{\nu} = 151$.
(e) $R_i, 165 \leq i \leq 217; kay = 15$. 

observation number i

observation number i

observation number i

observation number i
Figure 18: Post-detection nonparametric analysis of mass difference data also using pre-detection data for surveillance for a change in mean: $\Lambda_k^\alpha$ as a function of $k$. 

(a) $1 \leq k \leq 42; \hat{\nu} = 27; n = N_{220} = 42.$

(b) $27 \leq k \leq 62; \hat{\nu} = 51; n = N_{220} = 62.$

(c) $51 \leq k \leq 113; \hat{\nu} = 107; n = N_{220} = 113.$

(d) $107 \leq k \leq 161; \hat{\nu} = 151; n = N_{220} = 161.$

(e) $151 \leq k \leq 217; n = 217.$
(a) $R_i, 1 \leq i \leq N_{210} = 42; \hat{\nu} = 27.$  
(b) $R_i, 43 \leq i \leq N_{210} = 62; kay = 17; \hat{\nu} = 51.$  
(c) $R_i, 63 \leq i \leq N_{210} = 113; kay = 14; \hat{\nu} = 107.$  
(d) $R_i, 114 \leq i \leq N_{210} = 161; kay = 8; \hat{\nu} = 151.$  
(e) $R_i, 162 \leq i \leq 217; kay = 12.$

Figure 19: Post-detection nonparametric analysis of mass difference data: surveillance for a change in mean also using pre-detection data. ARL to false alarm $= 370.$
7 Surveillance of the standard deviation

Typically, two types of control are exercised in a quality setting: (1) the process mean is monitored via statistics computed from samples of size $m$ and (2) the precision of the process is monitored via standard deviations computed from each sample of $m$ values. In Sections 2-4, the surveillance scheme for the process mean is based on the Shiryayev-Roberts method. Invariance structures, similar to those of Sections 2-4 can be exploited to construct a surveillance scheme for the process precision, both parametrically and non-parametrically (see Gordon and Pollak, 1991, 1992).

For the mass calibration process, the statistic for monitoring the process mean is a least-squares estimate (check standard) from six difference measurements. The statistic for monitoring the process precision is the residual standard deviation of the fit to the six difference measurements. The latter characterizes the precision of the balance and any degradation or change in the balance is of special interest in this process.

Recall Figure 1 and the notation thereafter. Each set of difference measurements, $\{y_j\}_{j=1}^6$ yields an estimate $s$ of $\sigma_e$. The residual standard deviation is given by $s = \sqrt{\frac{1}{3} \sum_{j=1}^{6} (\hat{y}_j - \bar{y})^2}$, where $\hat{y}_1 = \mu_1$, $\hat{y}_2 = \mu_2$, $\hat{y}_3 = \mu_3$, $\hat{y}_4 = \mu_2 - \mu_1$, $\hat{y}_5 = \mu_3 - \mu_1$, $\hat{y}_6 = \mu_3 - \mu_2$. (Note that in the absence of components of error other than $\epsilon$, the $\sigma$ of Sections 3-6 would equal $\sigma_e/\sqrt{2}$). Estimates of $\sigma_e$ from each intercomparison design are given in the 6th column of Appendix 1 and plotted in Figure 20.

Since three parameters or mean $(\mu_1, \mu_2, \mu_3)$ are estimated and there are six observations, the distribution of each $3\sigma_e^2/\sigma_e^2$ is $\chi^2_{(3)} = \text{Gamma}(1.5, 0.5)$. So letting $Y_i$ denote the $i$th estimate of $\sigma_e$, the sequence of observations $Y_i$ with unknown baseline $\sigma_e$ is being monitored for a change. In order to get rid of the nuisance parameter $\sigma_e$, denote $T_1 = 0$ and

$$T_n = \sum_{j=1}^{n-1} Y_j^2 / \sum_{j=1}^{n} Y_j^2$$

for $n > 1$. The distribution of the sequence $\{T_n\}$ does not depend on $\sigma_e$. (Since our procedure is based on likelihood ratios, any equivalent invariant set
Figure 20: 217 estimates in milligrams of the standard deviation, $\sigma_e$ of the difference between two 1 kilogram weights, made at NIST between 1975 and 1988.

of statistics - such as $Z_n = Y_n / Y_1$ - will yield the same $\Lambda_n^\nu$'s, hence the same control scheme. Our choice is $T_n$ in order to use the formulae of Gordon and Pollak, 1990.)

Consider first the one-sided detection problem (that is, the change can only be an increase of the standard deviation; or, alternately, the change can only be a decrease). Suppose it is of importance to detect a change of a magnitude $g \sigma_e$ (or more extreme); that is, if after change the value of the standard deviation becomes $g \sigma_e$ (or more extreme), it would be of interest to raise an alarm. Then the $\nu = k$ versus $\nu = \infty$ likelihood $\Lambda_n^\nu$ of $T_2, \ldots, T_n$ for $n \geq k \geq 1$ (Gordon and Pollak, 1990, Theorem 2) is

$$\Lambda_n^\nu = g^{3(n-k+1)} [g^2 + (1 - g^2) \prod_{j=k}^n T_j]^{-1.5n}.$$ 

To emphasize the dependence on $g$, write $R_n^g$ instead of $R_n$. The foregoing analysis uses $g = 2$ and $g = 1/2$. To put the two together - i.e. the putative change is to double the original standard deviation or to half of it - we redefine $R_n = (R_n^{g=2} + R_n^{g=1/2})/2$, and stop at $N_A = \min\{n \mid R_n \geq A\}$. In the limit (as $A \to \infty$), we expect $E_A N_A / A \approx 2.6$ (see Appendix 7). Following the previous sections, if we set $B = 370$, then $A = B/2.6 \approx 140$.

Analysis analogous to Figures 5 and 12 is portrayed in Figure 21.

30
Figure 21: Parametric analysis of mass difference data: surveillance for a change of standard deviation. ARL to false alarm=370, starting anew after each detection. (a) $R_i, 1 \leq i \leq N_{140} = 47$. (b) $R_i, 48 \leq i \leq N_{140} = 177$. (c) $R_i, 178 \leq i \leq N_{140} = 207$. (d) $R_i, 208 \leq i \leq 217$. 

31
That is, if after each detection we start anew, the stopping times are immediately after the 47th, 177th and 207th observations. (An analysis analogous to Figures 16-19 is also possible).

It is clear from an inspection of the data that observation 207 is an outlier. If this observation is deleted, one obtains Figure 22. In other words, in reality there seems not to have been a change, and the alarm raised after the 207th observation is a false alarm.

Figure 22: Parametric surveillance for a change of standard deviation, starting with observations #178, deleting observation #207; ARL to false alarm=370, $R_i$, $178 < i < 216$.

The details of a nonparametric analysis will be given elsewhere. We do remark, however, that an analogous nonparametric analysis would not have stopped after the 207th observation, even without deleting it, implicitly rec-
ognizing it for what it is (an outlier).

As a final comment, we remark that the changes seem to be somewhat smaller than the putative ones. Had we taken, for instance, \( g = \sqrt{2} \) and \( g = 1/\sqrt{2} \), (doubling or halving the variance), the stopping times for an analysis analogous to Figure 21 are 47, 166 and 207 (with no stopping at 207 if observation 207 is deleted). In other words, a better guess of the post-change value will result in (somewhat) earlier detection.

8 Summary and Conclusion

We have presented a number of surveillance schemes, both parametric and nonparametric, for detecting a change in mean and for detecting a change in standard deviation, where no baseline is known. We probed the meaning and ramifications of various surveillance schemes and choices of parameters.

We find that the surveillance schemes presented are powerful, and should be considered for use instead of a Shewhart scheme: the change in mean following the 150th observation would have been discovered within the 15 subsequent observations, by all of the schemes we presented. (The Shewhart control chart leaves one unaware of the change almost to the end of the sequence. See Appendix 2.)
9 Appendix 1. Check standard data from the mass calibration process

<table>
<thead>
<tr>
<th></th>
<th>75.884</th>
<th>-19.51836</th>
<th>12.</th>
<th>.0217</th>
<th>41.</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>75.922</td>
<td>-19.49785</td>
<td>12.</td>
<td>.0118</td>
<td>41.</td>
</tr>
<tr>
<td>4</td>
<td>76.129</td>
<td>-19.49223</td>
<td>12.</td>
<td>.0210</td>
<td>41.</td>
</tr>
<tr>
<td>5</td>
<td>76.642</td>
<td>-19.52728</td>
<td>12.</td>
<td>.0265</td>
<td>41.</td>
</tr>
<tr>
<td>6</td>
<td>76.723</td>
<td>-19.50172</td>
<td>12.</td>
<td>.0317</td>
<td>41.</td>
</tr>
<tr>
<td>7</td>
<td>77.220</td>
<td>-19.49191</td>
<td>12.</td>
<td>.0194</td>
<td>41.</td>
</tr>
<tr>
<td>9</td>
<td>77.551</td>
<td>-19.49717</td>
<td>12.</td>
<td>.0274</td>
<td>41.</td>
</tr>
<tr>
<td>10</td>
<td>77.742</td>
<td>-19.46766</td>
<td>12.</td>
<td>.0361</td>
<td>41.</td>
</tr>
<tr>
<td>11</td>
<td>77.742</td>
<td>-19.52115</td>
<td>12.</td>
<td>.0362</td>
<td>41.</td>
</tr>
<tr>
<td>12</td>
<td>77.839</td>
<td>-19.52576</td>
<td>12.</td>
<td>.0320</td>
<td>41.</td>
</tr>
<tr>
<td>13</td>
<td>77.954</td>
<td>-19.43922</td>
<td>12.</td>
<td>.0096</td>
<td>41.</td>
</tr>
<tr>
<td>14</td>
<td>77.954</td>
<td>-19.49686</td>
<td>12.</td>
<td>.0238</td>
<td>41.</td>
</tr>
<tr>
<td>15</td>
<td>78.097</td>
<td>-19.50119</td>
<td>12.</td>
<td>.0224</td>
<td>41.</td>
</tr>
<tr>
<td>16</td>
<td>78.218</td>
<td>-19.50662</td>
<td>12.</td>
<td>.0117</td>
<td>41.</td>
</tr>
<tr>
<td>20</td>
<td>79.113</td>
<td>-19.43884</td>
<td>12.</td>
<td>.0122</td>
<td>41.</td>
</tr>
<tr>
<td>26</td>
<td>79.188</td>
<td>-19.50532</td>
<td>12.</td>
<td>.0339</td>
<td>41.</td>
</tr>
<tr>
<td>31</td>
<td>79.312</td>
<td>-19.50940</td>
<td>12.</td>
<td>.0203</td>
<td>41.</td>
</tr>
<tr>
<td>32</td>
<td>79.312</td>
<td>-19.42423</td>
<td>12.</td>
<td>.0274</td>
<td>41.</td>
</tr>
<tr>
<td>38</td>
<td>79.661</td>
<td>-19.45838</td>
<td>12.</td>
<td>.0362</td>
<td>41.</td>
</tr>
<tr>
<td>46</td>
<td>80.430</td>
<td>-19.44348</td>
<td>12.</td>
<td>.0275</td>
<td>41.</td>
</tr>
<tr>
<td>49</td>
<td>81.653</td>
<td>-19.44764</td>
<td>12.</td>
<td>.0054</td>
<td>41.</td>
</tr>
<tr>
<td>50</td>
<td>82.030</td>
<td>-19.45455</td>
<td>12.</td>
<td>.0340</td>
<td>41.</td>
</tr>
<tr>
<td>52</td>
<td>82.304</td>
<td>-19.47514</td>
<td>12.</td>
<td>.0376</td>
<td>41.</td>
</tr>
<tr>
<td>53</td>
<td>82.384</td>
<td>-19.49357</td>
<td>12.</td>
<td>.0482</td>
<td>41.</td>
</tr>
<tr>
<td>54</td>
<td>82.444</td>
<td>-19.50361</td>
<td>12.</td>
<td>.0290</td>
<td>41.</td>
</tr>
<tr>
<td>55</td>
<td>82.481</td>
<td>-19.48108</td>
<td>12.</td>
<td>.0326</td>
<td>41.</td>
</tr>
<tr>
<td>56</td>
<td>82.489</td>
<td>-19.50317</td>
<td>12.</td>
<td>.0338</td>
<td>41.</td>
</tr>
<tr>
<td>57</td>
<td>82.489</td>
<td>-19.45445</td>
<td>12.</td>
<td>.0381</td>
<td>41.</td>
</tr>
<tr>
<td>58</td>
<td>82.737</td>
<td>-19.48382</td>
<td>12.</td>
<td>.0373</td>
<td>41.</td>
</tr>
<tr>
<td>59</td>
<td>82.804</td>
<td>-19.50324</td>
<td>12.</td>
<td>.0245</td>
<td>41.</td>
</tr>
<tr>
<td>60</td>
<td>82.836</td>
<td>-19.50624</td>
<td>12.</td>
<td>.0212</td>
<td>41.</td>
</tr>
<tr>
<td>61</td>
<td>82.839</td>
<td>-19.45943</td>
<td>12.</td>
<td>.0194</td>
<td>41.</td>
</tr>
<tr>
<td>63</td>
<td>83.051</td>
<td>-19.47940</td>
<td>12.</td>
<td>.0322</td>
<td>41.</td>
</tr>
<tr>
<td>64</td>
<td>83.124</td>
<td>-19.52737</td>
<td>12.</td>
<td>.0482</td>
<td>41.</td>
</tr>
<tr>
<td>Serial number</td>
<td>Date by year</td>
<td>Check standard ID</td>
<td>Check standard value</td>
<td>Balance ID</td>
<td>Residual standard deviation</td>
</tr>
<tr>
<td>---------------</td>
<td>--------------</td>
<td>-------------------</td>
<td>----------------------</td>
<td>------------</td>
<td>----------------------------</td>
</tr>
<tr>
<td>198</td>
<td>87.788 41</td>
<td>-19.41108 12</td>
<td>.0377 41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>199</td>
<td>87.798 41</td>
<td>-19.41254 12</td>
<td>.0471 41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>87.801 41</td>
<td>-19.43138 12</td>
<td>.0592 41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>201</td>
<td>87.820 41</td>
<td>-19.45783 12</td>
<td>.0102 41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>202</td>
<td>87.847 41</td>
<td>-19.47197 12</td>
<td>.0445 41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>203</td>
<td>87.866 41</td>
<td>-19.46213 12</td>
<td>.0454 41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>204</td>
<td>87.922 41</td>
<td>-19.45791 12</td>
<td>.0240 41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>205</td>
<td>87.925 41</td>
<td>-19.45214 12</td>
<td>.0409 41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>206</td>
<td>87.957 41</td>
<td>-19.46248 12</td>
<td>.0570 41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>207</td>
<td>88.027 41</td>
<td>-19.50108 12</td>
<td>.1492 41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>208</td>
<td>88.027 41</td>
<td>-19.43155 12</td>
<td>.0469 41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>209</td>
<td>88.094 41</td>
<td>-19.44742 12</td>
<td>.0275 41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>210</td>
<td>88.151 41</td>
<td>-19.43895 12</td>
<td>.0180 41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>211</td>
<td>88.204 41</td>
<td>-19.41520 12</td>
<td>.0105 41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>212</td>
<td>88.207 41</td>
<td>-19.39467 12</td>
<td>.0252 41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>213</td>
<td>88.277 41</td>
<td>-19.43458 12</td>
<td>.0523 41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>214</td>
<td>88.315 41</td>
<td>-19.41401 12</td>
<td>.0275 41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>215</td>
<td>88.339 41</td>
<td>-19.44789 12</td>
<td>.0376 41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>216</td>
<td>88.398 41</td>
<td>-19.43033 12</td>
<td>.0215 41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>217</td>
<td>88.433 41</td>
<td>-19.43883 12</td>
<td>.0403 41</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Column 1: Serial number of observation.
Column 2: Date by year.
Column 3: Check standard ID.
Column 4: Check standard value.
Column 5: Balance ID.
Column 6: Residual standard deviation.
Column 7: Design ID.
10 Appendix 2: Shewhart Chart for the data of Figure 2

A Shewhart chart for the data of Figure 2 would typically be constructed in the following way: stop when $|\frac{X_n - \bar{X}_{n'}}{s_{n'}}|$ exceeds 3. Here $\bar{X}_{n'}$ is the mean of the first $n' = 114$ mass differences in column 6 of Appendix 1, and $s_{n'}$ is the standard deviation computed from the $n' = 114$ differences. The resulting chart is given in Figure 23.

![Shewhart Chart](image)

Figure 23: A Shewhart chart for the data of Figure 2.

The Shewhart chart raised an alarm at observation #154. This observation was initially regarded as an outlier, and since no other observations came in which crossed the $3\sigma$ limit until a half year later when #179 crossed the $3\sigma$ limit, a change in mean level was not detected until a deeper retrospective analysis was made.
The reason for Shewhart's ineffectiveness is clear, post-facto: the average of the first 114 observations is $-19.4771$ mg; the average of the last 103 is $-19.4506$ mg. The standard deviation of the first 114 observations is 0.030 mg. Hence the change of 0.026 mg was an increase of less than 1 standard deviation. Shewhart charts are known to be ineffective in detecting such changes.

11 Appendix 3. Rationale of the SR procedure

We continue with the notation of Section 3. Assume first that the observations are independent.

The idea behind the SR procedure regards the problem in a Bayesian context. Consider the following structure: one stands to lose one unit for raising an alarm and $c$ units ($c < 1$) for each observation taken after change until detection. Suppose $\nu$ has a Geometric $(p)$ prior distribution. Heuristically, because of the memoryless property of the geometric distribution, one would expect that at any point in time, the only relevant information is the posterior distribution that a change is in effect, $P(\nu \leq n | X_1, \ldots, X_n)$. Given that there was no change, the future as seen in two different points in time is stochastically the same, due to the memoryless of the geometric prior; therefore, if the posterior probability of a change being in effect is the same for two different points, one's actions should be the same. In other words, one would expect to raise an alarm whenever the posterior probability of a change in effect exceeds a certain threshold. (For a formal proof, see Shiryaev, 1963 or 1978.)

Using Bayes' theorem, letting $q = 1 - p$ and noting that $f_{\nu=k}(X_1, \ldots, X_n) = f_{\nu=\infty}(X_1, \ldots, X_n)$ for $k > n$, we obtain

$$P(\nu = k | X_1, \ldots, X_n) = \frac{f_{\nu=k}(X_1, \ldots, X_n)q^{k-1}}{\sum_{j=1}^{\infty} f_{\nu=j}(X_1, \ldots, X_n)q^{j-1}}$$

$$= \frac{f_{\nu=k}(X_1, \ldots, X_n)q^{k-1}}{\sum_{j=1}^{\infty} f_{\nu=j}(X_1, \ldots, X_n)q^{j-1} + f_{\nu=\infty}(X_1, \ldots, X_n)q^{n}}.$$
Therefore the posterior probability that a change is in effect is

\[
P(\nu \leq n \mid X_1, \ldots, X_n) = \frac{\sum_{k=1}^{n} f_{\nu=k}(X_1, \ldots, X_n)p^q^{k-1}}{\sum_{j=1}^{n} f_{\nu=j}(X_1, \ldots, X_n)p^q^{j-1} + f_{\nu=\infty}(X_1, \ldots, X_n)q^n}
\]

\[
= \frac{\sum_{k=1}^{n} \frac{f_{\nu=k}(X_1, \ldots, X_n)}{f_{\nu=\infty}(X_1, \ldots, X_n)} \left( \frac{1}{q} \right)^{n-k+1}}{\sum_{k=1}^{n} \frac{f_{\nu=k}(X_1, \ldots, X_n)}{f_{\nu=\infty}(X_1, \ldots, X_n)} \left( \frac{1}{q} \right)^{n-k+1} + \frac{1}{p}}.
\]

Since this expression is an increasing function of its numerator, the stopping rule has the form: stop the first time that

\[
\sum_{k=1}^{n} \frac{f_{\nu=k}(X_1, \ldots, X_n)}{f_{\nu=\infty}(X_1, \ldots, X_n)} \left( \frac{1}{q} \right)^{n-k+1}
\]

exceeds a prespecified threshold.

Now consider the case \( p \approx 0 \). This is approximately a noninformative prior. But \( p \approx 0 \) implies \( q \approx 1 \), so (2) \( \approx R_n \), and the Bayes rule is approximately \( N_A \) (for an appropriate \( A \)).

For a rigorous treatment in the case of independent observations, see Poljak (1985).

When the observations are not independent, nothing changes in the derivation of \( R_n \approx 2 \). What does change is the heuristics; it is not true any more that everything depends only on the posterior probability that a change is in effect. Nonetheless, one can still proceed with a SR procedure. Although it won't be optimal any more, in many cases it is almost optimal. (Cf. Poljak and Siegmund, 1991; Gordon and Pollak, 1991.)
12 Appendix 4. Operating characteristics of the SR procedure: theoretical details

We continue with the notation of Section 3. Note that under \( P_\infty \) when \( k \) is fixed, the sequence \( \Lambda^n_k, n \geq 1 \) is a martingale with unit expectation:

\[
E_\infty(\Lambda^{n+1}_k \mid X_1, \ldots, X_n) = E_\infty\left(\frac{f_{\nu=k}(X_1, \ldots, X_{n+1})}{f_{\nu=\infty}(X_1, \ldots, X_n)} \mid X_1, \ldots, X_{n+1}\right)
\]

\[
= E_\infty\left(\frac{f_{\nu=k}(X_{n+1} \mid X_1, \ldots, X_n) f_{\nu=k}(X_1, \ldots, X_n)}{f_{\nu=\infty}(X_1, \ldots, X_n) f_{\nu=\infty}(X_1, \ldots, X_n)} \mid X_1, \ldots, X_n\right)
\]

\[
= \frac{f_{\nu=k}(X_1, \ldots, X_n)}{f_{\nu=\infty}(X_1, \ldots, X_n)} E_\infty\left(\frac{f_{\nu=k}(X_{n+1} \mid X_1, \ldots, X_n)}{f_{\nu=\infty}(X_{n+1} \mid X_1, \ldots, X_n)} \mid X_1, \ldots, X_n\right)
\]

\[
= \Lambda^n_k \int \frac{f_{X_{n+1} \mid X_1, \ldots, X_n, \nu=k}(x)}{f_{X_{n+1} \mid X_1, \ldots, X_n, \nu=\infty}(x)} f_{X_{n+1} \mid X_1, \ldots, X_n, \nu=\infty}(x) dx
\]

\[
= \Lambda^n_k \int f_{X_{n+1} \mid X_1, \ldots, X_n, \nu=k}(x) dx
\]

and

\[
E_\infty \Lambda^n_k = \int \cdots \int \frac{f_{\nu=k}(x_1, \ldots, x_n)}{f_{\nu=\infty}(x_1, \ldots, x_n)} f_{\nu=\infty}(x_1, \ldots, x_n) dx_1 \cdots dx_n
\]

\[
= \int \cdots \int f_{\nu=\infty}(x_1, \ldots, x_n) dx_1 \cdots dx_n
\]

\[
= 1.
\]

Hence, \( R_n - n \) is a \( P_\infty \)-martingale with zero expectation; for

\[
E_\infty(R_{n+1} - (n + 1) \mid X_1, \ldots, X_n) = E_\infty(\sum_{k=1}^{n+1} \Lambda^{n+1}_k \mid X_1, \ldots, X_n) - (n + 1)
\]

\[
= \sum_{k=1}^{n} E_\infty(\Lambda^{n+1}_k \mid X_1, \ldots, X_n)
\]

\[
+ E_\infty(\Lambda^{n+1}_{n+1} \mid X_1, \ldots, X_n) - (n + 1)
\]

\[
= \sum_{k=1}^{n} \Lambda^n_k + \Lambda^n_{n+1} - (n + 1)
\]

\[
= R_n + 1 - (n + 1)
\]

\[
= R_n - n
\]
and

\[ E_\infty(R_n - n) = E_\infty \sum_{k=1}^{n} \Lambda_k^n - n = \sum_{k=1}^{n} E_\infty \Lambda_k^n - n = n - n = 0. \]

Now apply the optional sampling theorem to \( R_n - n \) with the stopping time \( N_A \) (cf. Chow, Robbins and Siegmund, 1967) to obtain

\[ E_\infty(R_{N_A} - N_A) = 0 \]

or

\[ E_\infty N_A = E_\infty R_{N_A}. \]  \hspace{1cm} (3)

Since by definition \( R_{N_A} \geq A \), this implies

\[ E_\infty N_A \geq A. \]

In case the observations are independent, a renewal theoretic analysis of the overshoot \( R_{N_A}/A \) promises the existence of the limit \( E_\infty R_{N_A}/A \), which by (3) equals \( E_\infty N_A/A \). The argument also provides a means of calculating the limit. See Pollak (1987) for details.

A somewhat more involved analysis of the overshoot does the same for many problems involving dependent observations. See Gordon and Pollak (1990) for details.

Note that the 2-sided schemes considered in previous sections are covered by the argument delineated above. For example, consider the two-sided scheme for a change in standard deviation. The one-sided scheme had \( R_n^{g=2} \) and \( R_n^{g=1/2} \) as their statistics; for the two-sided case \( R_n = (R_n^{g=2} + R_n^{g=1/2})/2 \). This is obtained by postulating a 50%–50% prior on \( (g = 2) - (g = 1/2) \) after the change; an easy calculation shows that \( \Lambda_k^n = 0.5\Lambda_k^n(g = 2) + 0.5\Lambda_k^n(g = 1/2) \), so that \( R_n = (R_n^{g=2} + R_n^{g=1/2})/2 \).
13 Appendix 5. A 2-sided SR scheme for detecting a change in a normal mean with unknown initial mean and unknown initial variance

We use the notation of Section 2; i.e.

\[ X_1, X_2, \ldots, X_{\nu-1} \sim N(\mu_0, \sigma^2) \]
\[ X_\nu, \ldots \sim N(\mu_0 + \delta \sigma, \sigma^2) \]

are independent, \( \nu, \mu_0, \sigma^2 \) are unknown, \( \delta \) is known (viewed as a representative of the post-change parameter); \( \bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i, Z_1 = 0, \)

\[ Y_i = (X_i - \bar{X}_{i-1}) \sqrt{\frac{i - 1}{i}}; \quad i = 2, 3, \ldots \]
\[ Z_i = \frac{Y_i}{Y_2}; \quad i = 2, 3, \ldots \]

Note that \( C_{\nu=k}(Y_i, Y_j) = 0 \) so that \( \{Y_i\} \) is a sequence of independent normally distributed random variables (under any of the probabilities \( P_{\nu=k}, 1 \leq k \leq \infty \)). Calculate

\[ E_{\nu=k}Y_i = \left[ \mu_0 + \delta \sigma - \frac{(i - k)(\mu_0 + \delta \sigma) + (k - 1)\mu_0}{i - 1} \right] \sqrt{\frac{i - 1}{i}} 1(i \geq k) \]
\[ = \frac{(k - 1)\delta \sigma}{i - 1} \sqrt{\frac{i - 1}{i}} 1(i \geq k) \]
\[ = \delta \sigma \frac{k - 1}{\sqrt{(i - 1)i}} 1(i \geq k). \]

For \( n \geq k > 2 \), obtain (by first conditioning on \( Y_2 \) and then integrating)

\[ f_{\nu=k; z_3, \ldots, z_n}(z_3, \ldots, z_n) \]
\[ = \frac{\partial^{n-2}}{\prod_{i=3}^{n} \partial z_i} E_{\nu=k} [\prod_{i=3}^{n} \Phi(z_i Y_2 - \frac{\delta(k - 1)}{\sqrt{(i - 1)i}} 1(i \geq k)) 1(Y_2 > 0)] \]

43
\[
+ \prod_{i=3}^{n}(1 - \Phi(z_i Y_2 - \frac{\delta(k-1)}{\sqrt{i-1}})^2 1(i \geq k)] 1(Y_2 < 0)
\]
\[
E_{\nu=k} | Y_2 | n^{-2} \left( \frac{1}{\sqrt{2\pi}} \right)^n e^{-\frac{1}{2} [Y_2^2 \sum_{i=3}^{\infty} z_i^2 - 2Y_2 \delta(k-1) \sum_{i=3}^{\infty} \frac{z_i}{\sqrt{i-1}} + \delta^2 (k-1)^2 \sum_{i=3}^{\infty} \frac{z_i^2}{i(i-1)}]}
\]
\[
= \int_{-\infty}^{\infty} \left( \frac{1}{\sqrt{2\pi}} \right)^{n-1} | y | n^{-2} e^{-\frac{1}{2} [y^2 \sum_{i=3}^{n} z_i^2 - 2y \delta(k-1) \sum_{i=3}^{n} \frac{z_i}{\sqrt{i-1}} + \delta^2 (k-1)^2 (\frac{1}{i(i-1)})]}
\]
\[
= (\frac{1}{\sqrt{2\pi}})^{n-1} e^{-\frac{\delta^2 (k-1)^2}{2} (\frac{1}{i(i-1)})} \frac{1}{\sum_{i=3}^{\infty} z_i^2} \sum_{i=3}^{n} \frac{z_i}{\sqrt{i-1}} \left( \frac{1}{\sum_{i=2}^{n} z_i^2} \right)^{n-1}
\times \int_{-\infty}^{\infty} | v - \frac{\delta(k-1) \sum_{i=2}^{n} \frac{z_i}{\sqrt{i-1}}}{\sqrt{\sum_{i=2}^{n} z_i^2}} | n^{-2} e^{-\frac{1}{2} v^2} dv
\]

where \( z_2 = 1 \), and similarly
\[
f_{\nu=\infty; z_3, \ldots, \infty} (z_3, \ldots, z_n) = \left( \frac{1}{\sqrt{2\pi}} \right)^{n-1} \left( \frac{1}{\sum_{i=2}^{n} z_i^2} \right)^{n-1} \int_{-\infty}^{\infty} | v | n^{-2} e^{-\frac{1}{2} v^2} dv.
\]

Letting
\[
a_{k,n} = -\frac{\delta(k-1) \sum_{i=3}^{n} \frac{z_i}{\sqrt{i(i-1)}}}{\sqrt{\sum_{i=2}^{n} z_i^2}},
\]

obtain for \( n > k > 2 \) that
\[
\Lambda_n^k = \frac{\int_{-\infty}^{\infty} | v - a_{k,n} | n^{-2} e^{-\frac{1}{2} v^2} dv}{\int_{-\infty}^{\infty} | v | n^{-2} e^{-\frac{1}{2} v^2} dv}
\]
\[
\text{Note that } \Lambda_n^k \text{ doesn't change if we insert } -\delta \text{ instead of } \delta; \text{ i.e. } +a_{k,n} \text{ and } -a_{k,n} \text{ will give the same value of } \Lambda_n^k.
\]

The analogous calculation when \( n > k = 2 \) yields
\[
\Lambda_n^k = \frac{\int_{-\infty}^{\infty} | v - a_{2,n} | n^{-2} e^{-\frac{1}{2} v^2} dv}{\int_{-\infty}^{\infty} | v | n^{-2} e^{-\frac{1}{2} v^2} dv}
\]
\[
\text{Clearly, } \Lambda_1^1 \equiv 1, \text{ and } \Lambda_2^2 \equiv 1.
\]
It remains to calculate the integrals on the right side of (4) and (5). Each can be computed by a recursion formula. However, for purposes of programming, computing each integral separately will cause problems, as the integrals become very large as $n$ progresses. It is better to write out a recursion-as it turns out, it's a double recursion - for the ratio of the integrals. Denote

\[
\begin{align*}
g_m(a) &= \int_a^\infty (v - a)^m e^{-\frac{1}{2}v^2} \, dv \\
f_m(a) &= \int_0^a (a - v)^m e^{-\frac{1}{2}v^2} \, dv \\
h_m &= \int_{-\infty}^\infty \left| v \right|^m e^{-\frac{1}{2}v^2} \, dv \\
u_m(a) &= \frac{g_m(a)}{h_m} \\
v_m(a) &= \frac{f_m(a)}{h_m}.
\end{align*}
\]

Thus

\[\Lambda_k^n = e^{-\frac{1}{2}k^2} \left[ \frac{1}{3} + \frac{1}{3} + \frac{1}{3} \right] + \frac{1}{2} a_{k,n} \left[ u_{n-2}(a_{k,n}) + v_{n-2}(a_{k,n}) \right].\] (6)

Now for $m \geq 2$

\[
\begin{align*}
h_m &= 2 \int_0^\infty v^{m-1} e^{-\frac{1}{2}v^2} \, dv \\
&= 2 \left[ v^{m-1} (-e^{-\frac{1}{2}v^2}) \right]_0^\infty + (m - 1) \int_0^\infty v^{m-2} e^{-\frac{1}{2}v^2} \, dv \\
&= (m - 1) h_{m-2}
\end{align*}
\]

where

\[
\begin{align*}
h_0 &= \sqrt{2\pi} \\
h_1 &= 2 \int_0^\infty v e^{-\frac{1}{2}v^2} \, dv = 2.
\end{align*}
\]

Also, for $m \geq 2$

\[
g_m(a) = \int_a^\infty (v - a)^{m-1} (v - a) e^{-\frac{1}{2}v^2} \, dv
\]
\[
\begin{align*}
&= \int_a^\infty (v - a)^{m-1}e^{-\frac{1}{2}v^2}dv - ag_{m-1}(a) \\
&= (v - a)^{m-1}(-e^{-\frac{1}{2}x^2}) \bigg|_a^\infty + (m - 1) \int_a^\infty (v - a)^{m-2}e^{-\frac{1}{2}v^2}dv - ag_{m-1}(a) \\
&= (m - 1)g_{m-2}(a) - ag_{m-1}(a)
\end{align*}
\]

where
\[
\begin{align*}
g_0(a) &= \sqrt{2\pi}(1 - \Phi(a)) \\
g_1(a) &= e^{-\frac{1}{2}a^2} - a\sqrt{2\pi}(1 - \Phi(a)).
\end{align*}
\]

Similarly, for \( m \geq 2 \)
\[
f_m(a) = (m - 1)f_{m-2}(a) + af_{m-1}(a)
\]

where
\[
\begin{align*}
f_0(a) &= \sqrt{2\pi}\Phi(a) \\
f_1(a) &= e^{-\frac{1}{2}a^2} + a\sqrt{2\pi}\Phi(a).
\end{align*}
\]

Now
\[
\begin{align*}
u_0(a) &= \frac{g_0(a)}{h_0} = 1 - \Phi(a) \\
u_1(a) &= \frac{g_1(a)}{h_1} = \frac{1}{2}e^{-\frac{1}{2}a^2} - a\sqrt{\pi}/2(1 - \Phi(a)) \\
v_0(a) &= \Phi(a) \\
v_1(a) &= \frac{1}{2}e^{-\frac{1}{2}a^2} + a\sqrt{\pi}/2\Phi(a).
\end{align*}
\]

Denote
\[
\begin{align*}
w_m(a) &= \frac{g_m(a)}{h_{m-1}} \\
y_m(a) &= \frac{f_m(a)}{h_{m-1}}.
\end{align*}
\]

For \( m \geq 3 \)
\[
w_m(a) = \frac{(m - 1)g_{m-2}(a) - ag_{m-1}(a)}{(m - 2)h_{m-3}} = \frac{m - 1}{m - 2}w_{m-2}(a) - au_{m-1}(a) \quad (7)
\]
and

\[ y_m(a) = \frac{(m-1)f_{m-2}(a) + af_{m-1}(a)}{(m-2)h_{m-3}} = \frac{m-1}{m-2}y_{m-2}(a) + av_{m-1}(a) \]  

(8)

where

\[ w_1(a) = \frac{g_1(a)}{h_0} = \frac{e^{-\frac{1}{2}a^2} - a\sqrt{2\pi}(1 - \Phi(a))}{\sqrt{2\pi}} \]

\[ w_2(a) = \frac{g_2(a)}{h_1} = \frac{g_0(a) - ag_1(a)}{h_1} = \frac{\sqrt{2\pi}(1 - \Phi(a)) - a[e^{-\frac{1}{2}a^2} - a\sqrt{2\pi}(1 - \Phi(a))]}{2} \]

\[ y_1(a) = \frac{f_1(a)}{h_0} = \frac{e^{-\frac{1}{2}a^2} + a\sqrt{2\pi}\Phi(a)}{\sqrt{2\pi}} \]

\[ y_2(a) = \frac{f_2(a)}{h_1} = \frac{f_0(a) + af_1(a)}{h_1} = \frac{\sqrt{2\pi}\Phi(a) + a[e^{-\frac{1}{2}a^2} + a\sqrt{2\pi}\Phi(a)]}{2} \]

Finally, note that for \( m \geq 2 \)

\[ u_m(a) = \frac{g_m(a)}{h_m} = \frac{(m-1)g_{m-2} - ag_{m-1}}{(m-1)h_{m-2}} = u_{m-2}(a) - \frac{a}{m-1}w_{m-1} \]  

(9)

\[ v_m(a) = \frac{f_m(a)}{h_m} = \frac{(m-1)f_{m-2} + af_{m-1}}{(m-1)h_{m-2}} = v_{m-2}(a) + \frac{a}{m-1}y_{m-1} \]  

(10)

Formulae (7)-(10) give values of \( u_m, v_m, w_m, y_m \) in terms of previous \( u_i, v_i, w_i, y_i \). So, starting by calculating \( u_i(a), v_i(a), w_i(a), y_i(a) \) (in this order) for \( i = 0, 1, 2, \ldots, m \) yields a recursive way of computing \( u_{n-2}, v_{n-2} \) needed to calculate \( \Lambda_k^n \) (as in (6)).

All of this is brought together in the computer program given in Figure 24, which calculates the series \( R_n \). The program is written in the MATLAB language. (See The Math Works, 1989, for a description.) The program uses the homogeneity of \( \Lambda_k^n \) in \( \sum_{i=2}^n z_i^2 \) and replaces \( z_i \) by \( Y_i \). Pressing \( R \) after the program has run will display the \( R_i, 1 \leq i \leq en \). Pressing \( t \) displays the values \( \Lambda_k^n, 1 \leq k \leq en \) (when \( en \geq 3 \)).
This is a program designed for detecting an abrupt change in the mean of independent Normal observations, the variance of which is unknown. The change may be to either side of the initial unknown mean. The detection scheme is a Shiryaev-Roberts parametric control chart based on the likelihood ratios of ratios of recursive residuals.

Input Parameters: data (row vector of size en)  
\( \delta = \text{delta, the representation of the change in the mean} \)

Output: \( R \) (row vector of size en, giving the values of the Shiryaev-Roberts statistic for \( n=1:en \))

```matlab
en=length(data);  
X=cumsum(data);  
R=zeros(1:en);  
Y=R;  
W=R;  
R(1)=1;  
R(2)=2;  
for n=3:en  
    t=zeros(1:n);  
    t(1)=1;  
    for i=2:n  
        Y(i)=(data(i)-(X(i-1)/(i-1)))*sqrt((i-1)/i);  
        W(i)=Y(i)/sqrt(i*(i-1));  
    end  
    s=sqrt(Y*Y');  
    W=W(:,n:-1:1);  
    W=cumsum(W);  
    x=ones(1:n);  
    x=cumsum(x)-1;  
    a=(d/s)*x.*W;  
    u=zeros(1:n);  
    v=u;  
    w*u  
    y=n;  
    for k=2:n  
        p=(exp((-a(k)^2)/2))/2;  
        q=erf(a(k)/sqrt(2))/2+0.5;  
        u(1)=p*(1-q)*sqrt(pi/2);  
        v(1)=p*a(k)*sqrt(pi/2);  
        w(1)=u(1)*sqrt(2/pi);  
        y(1)=(1-q)^2+sqrt(2/2);  
        u(2)=1-q-a(k)*u(1);  
        v(2)=q+a(k)*y(1);  
        w(2)=(1-q)*sqrt(2/pi)-a(k)*z(1);  
        y(2)=q*sqrt(pi/2)-z(1);  
        for j=3:n  
            u(j)=u(j-2)-a(k)*w(j-1)/(j-1);  
            w(j)=(j-1)/(j-2) + w(j-2)-a(k)*u(j-1);  
            v(j)=v(j-2)+a(k)*y(j-1)/(j-1);  
            y(j)=(j-1)/(j-2) + y(j-2)+a(k)*v(j-1);  
        end  
        t(k)=u(n-2)+v(n-2)*exp(-0.5*((d*(k-1))/2)*((1/(1-k))-...  
                              (1/n)-(W(k)/s)^2));  
        t(2)=t(2)*exp(-0.25*(d^2));  
    end  
end  
R(n)=sum(t);  
```

Figure 24: MATLAB-language computer program for computing \( R_n \) for the parametric SR scheme of section 2.
Appendix 6. A nonparametric surveillance scheme for detecting a change

We follow the notation of Section 5. The foregoing is an attempt to explain the ideas behind the nonparametric scheme. For exact details and rigorous proof see Gordon and Pollak (1992).

Suppose first that the post-change observations are stochastically larger than pre-change. Clearly the problem has an invariance structure: applying any increasing function to the observations will not change the problem, and will not change the rank of the observations.

The main difficulty in constructing a SR procedure is to calculate a likelihood ratio for the ranks $r_1, r_2, \ldots, r_n$. The denominator is obvious: if $\nu = \infty$, all observations are interchangeable and every configuration $r_1, r_2, \ldots, r_n$ has the same probability (namely, $1/n!$). It is the numerator which causes problems.

The primary idea is to find any two densities, $f_0$ and $f_1$, which allow a (tractable) calculation of the numerator. If such be found, they will yield likelihood ratios $\Lambda^n_F$ of $r_1, r_2, \ldots, r_n$. While these likelihood ratios are calculated under the assumption that $f_0$ is the true pre-change density, this assumption makes no difference when regarding the $\Lambda^n_F$'s (hence the $R_n$'s) behavior under $P_\infty$: after all, if the true pre-change distribution is, say, $G_0$, transforming the observations by $F_0^{-1}(G_0(\cdot))$ (where $F_0$ is the cdf whose density is $f_0$ and $F_0^{-1}$ is the inverse transform of $F_0$) will make the transformed observations have density $f_0$ without changing the observed ranks. (However, there is a difference when regarding the speed of detection: transforming the observations transforms their post-change density, too. The density $f_1$ should therefore be seen as a representation of the post-change density of the $F_0^{-1}G_0$-transformed observations.)

The choice of $f_0$ and $f_1$ as proposed by Gordon and Pollak (1991) is $f_0(x) = (1/2)\exp\{-|x|/2\}$ and $f_1(x) = p\alpha\exp\{-\alpha x\}1(x \geq 0) + (1 - p)\beta\exp\{\beta x\}1(x < 0)$, where $1/2 \leq p \leq 1, 0 < \alpha \leq 1 \leq \beta < \infty$. (Thus the post-change distribution is stochastically larger.) This choice of $f_0$ and $f_1$
enables computation of $\Lambda_k^n$ via the following lemma (Savage, 1956):

**Lemma:** Let $Y_1, Y_2, \ldots, Y_n$ be i.i.d. $\exp(1)$-distributed random variables and let $x_1, x_2, \ldots, x_n$ be positive constants. Then

$$P(\frac{Y_1}{x_1} < \frac{Y_2}{x_2} < \ldots < \frac{Y_n}{x_n}) = \prod_{k=1}^{n} \left( \frac{x_k}{\sum_{i=k}^{n} x_i} \right).$$

(This can be proven either directly, or by induction on $n$.)

To see how this can be used to compute $f_{\nu=k}(r_1, r_2, \ldots, r_n)$ (the numerator of $\Lambda_k^n$), consider the first five mass difference observations (the first five values of Column 4 of Appendix 1). Clearly, $r_1 = 1, r_2 = 2, r_3 = 3, r_4 = 3, r_5 = 1$; i.e. $X_3 < X_1 < X_2 < X_4 < X_3$. Therefore, $f_{\nu=k}(r_1, r_2, \ldots, r_5)$ is equal to $P_{\nu=k}(X_3 < X_1 < X_2 < X_4 < X_3)$. Without loss of generality, imagine that we made the transformation $F_0^{-1}G_0$, so the transformed observations are $X_1^*, \ldots, X_5^*$, and the ranks are unchanged, and $X_i^* \sim f_0$ pre-change, $X_i^* \sim f_1$, post-change. For example's sake, consider the case $k = 3$.

Let $B = \sum_{i=1}^{5} 1(X_i^* < 0)$; that is, $B$ is the number of observations below the pre-change median. Of course, we don't observe $B$ (because we only know the ranks). Nonetheless, $B$ can be of technical help in computing the probability $P_{\nu=3}(X_5^* < X_1^* < X_2^* < X_4^* < X_3^*) = P_{\nu=3}(X_5^* < X_1^* < X_2^* < X_4^* < X_3^*)$, as we shall now show. Clearly

$$P_{\nu=3}(X_5^* < X_1^* < X_2^* < X_4^* < X_3^*) = \sum_{i=0}^{5} P_{\nu=3}(X_5^* < X_1^* < X_2^* < X_4^* < X_3^*, B = i).$$

Consider, for example, the element in this sum corresponding to $i = 2$. If $B = 2$, for the event $(X_5^* < X_1^* < X_2^* < X_4^* < X_3^*)$ to occur, necessarily $X_3^* > X_4^* > X_5^* > 0$ and $0 > X_1^* > X_2^*$. Therefore

$$P_{\nu=3}(X_5^* < X_1^* < X_2^* < X_4^* < X_3^*, B = 2) = P_{\nu=3}(\{X_3^*, X_4^*, X_5^* \text{ are positive and } X_3^* > X_4^* > X_5^*\} \cap \{X_1^*, X_5^* \text{ are negative and } |X_5^*| > |X_1^*|\}) = P_{\nu=3}(X_3^*, X_4^*, X_5^* \text{ are positive and } X_3^* > X_4^* > X_2^*) \times P_{\nu=3}(\{X_1^*, X_5^* \text{ are negative and } |X_5^*| > |X_1^*|\}).$$
Note that if \( X_i^* \) is positive, then \( X_i^* \sim \exp(1) \) if \( i < 3 \) and \( X_i^* \sim \exp(\alpha) \) if \( i \geq 3 \). Likewise, if \( X_i^* \) is negative, \( \mid X_i^* \mid \sim \exp(1) \) if \( i < 3 \) and \( \mid X_i^* \mid \sim \exp(\beta) \) if \( i \geq 3 \). Also, \( P_{\nu=3}(X_i^* > 0) = 1/2 \) or \( p \) corresponding to whether \( i < 3 \) or \( i > 3 \). Finally, note that \( Y_1/x \sim \exp(x) \) if \( Y_1 \sim \exp(1) \). Putting this together one obtains by the Lemma that

\[
P_{\nu=3}(X_3^*, X_4^*, X_2^* \text{ are positive and } X_3^* > X_4^* > X_2^*) \]
\[
= P_{\nu=3}(X_3^* > X_4^* > X_2^* \mid X_3^*, X_4^*, X_2^* \text{ are positive}) \cdot P_{\nu=3}(X_3^*, X_4^*, X_2^* \text{ are positive})
\]
\[
= P\left(\frac{Y_3}{\alpha} > \frac{Y_4}{\alpha} > \frac{Y_1}{1}\right) \cdot p \cdot \frac{1}{2}.
\]
\[
= \frac{\alpha}{\alpha + \alpha} \cdot \frac{1}{1 + \alpha + \alpha} \cdot p \cdot \frac{1}{2}
\]
\[
= \frac{1}{2} \cdot \frac{p^2}{4 \alpha + 1}.
\]

In a similar fashion,

\[
P_{\nu=3}(X_1^*, X_5^* \text{ are negative and } \mid X_1^* \mid > \mid X_5^* \mid) = \frac{1}{2} (1-p) \frac{\beta}{\beta \cdot 1 + \beta} = \frac{11 - p}{21 + \beta}.
\]

The other elements on the right side of (11) can be computed analogously. The same considerations yield a calculation of \( \Lambda_n^* \) for general \( n \geq k \geq 1 \). A computer program for the sequence of statistics \( R_n \) is given in Figure 25. The program is written in MATLAB language. (See The Math Works, 1989, for a description.)

In Section 5, the nonparametric scheme used is two-sided. \( R_n \) of this scheme is obtained by running the program in Figure 25 once on the data (in Column 4 of Appendix 1) and once on the same data multiplied by \(-1\), and averaging the two outputs (in a manner analogous to the analysis in Section 6).

As for the choice of parameters \( p, \alpha, \beta \) we follow the lines of Gordon and Pollak (1989). If \( G_0 \) and \( G_1 \) are the true pre-and post-change distributions, one can show (for \( \nu \) not too close to the beginning of surveillance, and for large \( A \)) that

\[
E_{\nu}(N_A - \nu \mid N_A \geq \nu) \approx \frac{\log A}{D} \quad (12)
\]
% Inputs: data row of en entries
% procparm row of procedure parameters: (alpha, beta, p)
% Outputs: ren row of en entries, one for each Shiryayev-Roberts statistic computed

-----get parameters-----
en = length(data);
alpha = procparm(1); beta = procparm(2); p = procparm(3); q = 1-p;
ln2alpha = log(2*p*alpha); ln2qbeta = log(2*q*beta);
lnratio = ln2alpha - ln2qbeta;
% initialize ren
ren = zeros(1,en);
for n = 2:en, %------compute ren(n)------
    lnbnml = zeros(1,n-1); \ computation of log of binomial(n,1/2) probs
    y = 1:1:n;
yn = log(y);
culny = cumsum(lny);
    for i=1:(n-1),
        lnbnml(i) = culny(n)-culny(i)-culny(n-i)-n*(log(2));
    end
end
lnbnml = [-n*(log(2)) lnbnml -n*(log(2))]; \ row of logs of bin(n,1/2) probs
datan = data(:,1:n);
% initialize and allocate vectors
incr = 1:1:n; \ 1 to n row vector
decr = [(n-1:-1:1) 1]; \ n-1 to 1 then 1
reverse = (n+1:-1:1); \ row vector to reverse order
lambdakn = zeros(1,n); \ to hold little lambda sub kn's
(dummy,invrank)= sort (datan'); \ index of smallest in invrank(1)
invrank = invrank'; \ row of inverse ranks
% compute vector of lambda sub nk's
for k = 1:n,
    timegek = (invrank>=k); \ time not less than k
    vsubk = cumsum(timegek);
    % for i= 1 to n care for i=0
        usubk = (n+1-k) - vsubk;
        lnvdenom = log([ 1 \ (1*(vsubk./incr)... *((beta-1)))]);
        lnudenom = log([(1*(n-k+1)*(alpha-1)/n) \ (1*(usubk./decr)... *((alpha-1)))]);
        \lnvdenom and lnudenom are n+1 vectors containing
        \ index of vectors is one plus number of putative
        \ log-denominators
        \ index of vectors is one plus number of putative
        \ negatives m
        lnprodneg = cumsum(lnvdenom);
dummy1 = lnudenom(reverse);
dummy2 = cumsum(dummy1);
lnprodp = dummy2(reverse);
usubkm = [(n+1-k) usubk];
lambdakn = sum(exp(lnbnml + (n+1-k)*ln2qbeta...
            \ usubkm\lnratio \ lnprodneg \ lnprodp));
end
ren(n) = sum(lambdakn);
end
ren(1)=1;

Figure 25: MATLAB - language computer program for computing $R_n$ for the a (one-sided) nonparametric SR scheme for detecting an increase.
where
\[
D = (1 - G_1(0)) \log(2p\alpha) + G_1(0) \log(2(1 - p)\beta) \\
+ (1 - \alpha) \int_0^\infty F_0^{-1}(G_0(x))dG_1(x) \\
+ (\beta - 1) \int_{-\infty}^0 F_0^{-1}(G_0(x))dG_1(x).
\]

Therefore, if \(G_0\) and \(G_1\) are surmised to be (approximately) the pre-and-post change distributions, \(p, \alpha, \beta\) can be chosen to maximize \(D\) (so as to minimize (12)). For the example worked out in Section 4, \(G_0 = N(\mu_0, \sigma^2)\) and \(G_1 = N(\mu_0 + \sigma, \sigma^2)\). Here \(D\) is maximized by \(p = .8413, \alpha = .531, \beta = 1.703\).

The integral in the definition of \(D\) must be evaluated numerically. See Gordon and Pollak (1989, 1991) for details.

15 Appendix 7. Computation of \(\lim_{A\to\infty} E_{\infty}N_A/A\).

Suppose first that the observations are independent, having density \(f_0\) pre-change and \(f_1\) post-change. As in Section 2, \(N_A = \min\{n \mid R_n \geq A\}\) where \(R_n = \sum_{k=1}^n f_{\nu=k}(X_1, \ldots, X_n)/f_{\nu=\infty}(X_1, \ldots, X_n)\). Denote \(x^+ = \max(x, 0)\) and \(S_n = \sum_{k=1}^n \log(f_1(X_k)/f_0(X_k))\). It can be shown (Pollak, 1987, in conjunction with Siegmund, 1985) that
\[
\lim_{A\to\infty} \frac{E_{\nu=1} S_1 e^{\sum_{\nu=1}^\infty \frac{1}{\nu} S_1 e^{-S_1^+}}}{A} = E_{\nu=1} S_1 e^{\sum_{\nu=1}^\infty \frac{1}{\nu} S_1 e^{-S_1^+}}. \tag{13}
\]

If the pre-and-post change distributions are not known, but an invariance structure enables one to construct one-sided schemes as in the problems of Sections 3, 4 and 6 then it can often be shown that (13) still holds (Gordon and Pollak, 1990), with any choice of nuisance parameters (since the procedure is invariant with respect to them). Thus, if the pre-and post-change distributions are \(N(\mu_0, \sigma^2)\) and \(N(\mu_0 + \sigma, \sigma^2)\) respectively, one may compute (13) with \(\mu_0 = 0, \sigma = 1\).

Calculation of (13) usually requires a numerical analysis. An exception is the nonparametric procedure of Figure 25; it can be shown that if \(2p\alpha\) and \(2(1 - p)\beta\) are both less than one then (13)=1/\(\alpha\). A table of (13) for values of \(\delta\) for the parametric procedure of Section 3 can be found in Pollak.

53
(1987). For Section 6, the values of (13) have not been tabled anywhere, and the constants were calculated by a computer program specially written to evaluate (13) for the problem of Section 6.

When the problem is 2-sided, when \( f_0 \) is known when \( R_n \) is an average of the statistic for each side, and when \( c_1, c_2 \) are the values of (13) for each of the two sides, then (Pollak, 1987) for the 2-sided procedure (13) = \( 1/[0.5(1/c_1) + 0.5/(1/c_2)] \). We used this for the 2-sided procedures used in (3), (4) and (6), though it must be admitted that we did not formally prove that this is valid also when the pre-change distribution is not (completely) known.

### 16 Appendix 8. Diagnostics

The basic assumption made in the analysis of the data appearing in Appendix 1 is independence of the observations. For the parametric procedures, normality was assumed to underlie the observations (causing the observations in Column 4 to be normally distributed and those of Column 6 to have a \( \sigma^2 X^2(3) \) distribution after being squared). Here we will try to check these assumptions.

These checks are perforce done after the other analyses: if other data are not available for comparison, care must be taken in the diagnostic analysis so that change points do not influence the results in a wrong way.

In the parametric analysis, points of change of the mass differences were estimated to be at observations 17, 51, 107, 151 (Figure 16). The nonparametric analyses estimated them to be at observations 27, 51, 107, 151 (Figure 18). Also, the parametric analyses of the residual standard deviations estimated observations 41 and 166 (and 207) to be points of change. (This is derived from figures analogous to Figures 16 and 18; they are not produced here.) It seems therefore that the sets of observations 51 – 106 and 107 – 150 are homogeneous sets of observations. Therefore we ran diagnostics on these sets. (The first changepoint is not clear cut; observations 14, 17 and 27 are almost equally likely candidates (Figure 18(a)), and the analysis of the residual standard deviations indicates another change at observation 41. Therefore we didn’t run diagnostics on the first 50 observations.)
(a) Diagnostics for the mass difference data

The Kolmogorov-Smirnov two-sided statistic for normality (with estimation of the mean and standard deviation) is $1.0169/\sqrt{56}$ and $0.8735/\sqrt{44}$ for the sets of observations $51 - 106$ and $107 - 150$, respectively. (The critical value for $\alpha = 20\%$ without parameter estimation is approximately $1.07/\sqrt{n}$.) Even if one combines the two sets and looks at the single set $51 - 150$, the statistic's value is $0.8211/\sqrt{100}$.) So, at least approximate normality of these observations is reasonably well established.

As for independence, two-sided runs tests produce $p$-values $0.1056$ and $0.1272$ for the sets of observations $51 - 106$ and $107 - 150$ when taken separately, and $0.1984$ when taken together (i.e. $51 - 150$). Autocorrelation plots (made by Dataplot; see Filliben, 1981) are given in Figure 26. (The horizontal lines are critical values for a 2-sided test, for each lag separately, at a 5% level of significance.) So, independence isn't something to worry about, either.

If one were to do diagnostics for the set $166 - 217$ with the $207th$ observation deleted (see Appendix 6 for a rationale), the two-sided Kolomogorov-Smirnov statistic is $0.7389/\sqrt{51}$. The autocorrelation plot of the data is given in Figure 27.

Things there also look pretty good so far. But, the runs test gives cause to worry: the $p$-value of the (two-sided) run statistic is $0.0027$. Perhaps a closer look at the observations (see Figure 28) may offer a possible explanation: there seem to be a few medium-sized changes, each of small duration, so that in all there is a normal distribution, there is no serial correlation, but there are clumps of observations at different levels, accounting for too few runs. (The duration of each run is too short for the control chart to catch.)
Figure 26: Autocorrelation plots for mass difference data:

(a) observations $51 - 106$
(b) observations $107 - 150$
(c) observations $51 - 150$. 

---

56
Figure 27: Autocorrelation plot for mass difference data: observations 166 – 217, with observation 207 deleted.

Figure 28: Detail of the mass difference data: observations 166 – 217, with observation 217 deleted.
(b) Diagnostics for the residual standard deviation data

Based on the essential normality of the observations, the distribution of $3s^2_i/\sigma^2$ should be $\chi^2_3 = \text{gamma}(1.5, .5)$. The Kolmogorov-Smirnov two-sided statistic computed with an estimate for $\sigma$ is .5225/\sqrt{56}, .8939/\sqrt{44}, .343/\sqrt{100}$ and .5186/\sqrt{125} for the sets of observations 51 – 106, 107 – 150, 51 – 150 and 41 – 165, respectively. The autocorrelation plots for these sets are given in Figure 29. There may be some $1-lag$ autocorrelation in the 51 – 106 data, though it disappears when the set is enlarged.

(The autocorrelations are computed for the data transformed to $N(0,1)$ observations; i.e., since $3s^2_i/\sigma^2$ are $\chi^2_3, \Phi^{-1}(F_{\chi^2_3}(3s^2_i/\sigma^2))$ are $N(0,1)$. The transformation was made in order for the critical values (the horizontal lines in Figure 29) to have valid meaning.)

The $p-$values of two-sided run tests are .003, .7603, .1594 and .0064 for the sets of observations 51 – 106, 107 – 150, 51 – 150 and 41 – 165, respectively. Again, a closer look at the observations may offer a possible explanation (see Figure 30). There seems to be a small trend in observations 51 – 106, and observations 41 – 165 seem somewhat $U-$shaped.

Finally, if one wanted to do diagnostics for the set 166 – 217 with the 201st observation deleted, the two-sided Kolmogorov-Smirnov statistic has the value .609/\sqrt{51}; the two-sided runs test has $p-$value .2431; and the autocorrelation plot (for the observations transformed to $N(0,1)$) is given in Figure 31.
Figure 29: Autocorrelation plots for residual standard deviations transformed to $N(0,1)$ variables: (a) observations 51–106, (b) observations 107–150, (c) observations 51–150 (d) observations 41–165.
Figure 30: Detail of the mass difference data: (a) observations 51 - 106 (b) observations 41 - 165.

Figure 31: Autocorrelation plot for residual standard deviations transformed to $N(0, 1)$ variables: observations 166 - 217 with observation 207 deleted.
17 Appendix 9. Remarks

1. We reiterate the comment made in the last paragraph of Appendix 7: there is some theoretical work left to do in order to prove existence of the limit in (13) and its calculation when there are nuisance parameters. Heuristically, because of the considerations of Gordon and Pollak (1991), it's hard to believe that the method is wrong.

2. In Section 3 we presented a two-sided parametric detection scheme. (The scheme is obviously two-sided; multiplying all the observations by \(-1\) doesn't change the \(R_n\) sequence.) If one wants a one-sided scheme, \(\text{sign}(Y_t)\) should be added to the data on which surveillance is based, and \(\Lambda^p\) calculated accordingly.

3. If the pre-change observations can be assumed to have a distribution which is symmetric about zero, then a somewhat stronger nonparametric procedure can be proposed in Section 4. (Its relative strength lies most in detection of an early change.) See Gordon and Pollak (1989).

4. A nonparametric procedure for Section 6 can be obtained from Bell, Gordon and Pollak (1992).

5. The relative efficiency of the nonparametric schemes used here are above 90%. (In other words, for changes not occurring early, one will be detecting the change with a less than 10% added average delay than the parametric schemes which would have been used were \(f_t\) known. See Gordon and Pollak, 1992. Comparison of Figures 12 and 13 corroborates this.)

6. Running the program of the parametric scheme (Figure 24) generally takes longer than the nonparametric one (Figure 25). As a matter of fact, running Figure 24 on the 217 observations in Column 4 of Appendix 1 took 3.25 hours on a Sun Sparc station 2, whereas running Figure 25 (twice - once for each side) took (altogether) only 20 minutes. However it should be noted that when actually doing surveillance on line, one does not have to recompute the whole \(R_n\) sequence; should we have now obtained an observation 218, one would have only had to compute \(R_{218}\). A small change in the program will do this. The computation of \(R_{218}\) only would be at most a matter of minutes.
References


