NISTIR 5158

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

NIST

NISTIR 5158

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

U.S. DEPARTMENT OF COMMERCE
Ronald H. Brown, Secretary

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
Raymond G. Kammer, Acting Director

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

List of Figures

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 3σ-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

†Key words and subject class. : control charts, Shiryayev-Roberts, Cusum, mass standards

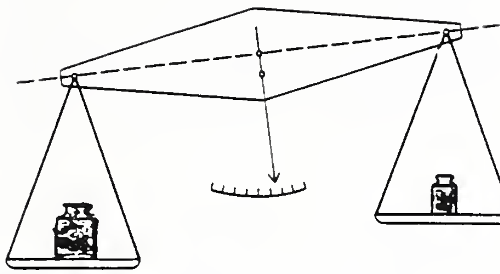
2 Introduction

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 uncertainity. 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).



	NIST	NIST	CLIENT	CLIENT	
	$\boxed{1000}$	$\boxed{1000}$	$\boxed{1000}$	$\boxed{500}$ + $\boxed{300}$ + $\boxed{200}$	
$y_1 =$	$+$	$-$			$+\epsilon_1$
$y_2 =$	$+$		$-$		$+\epsilon_2$
$y_3 =$	$+$			$-$	$+\epsilon_3$
$y_4 =$		$+$	$-$		$+\epsilon_4$
$y_5 =$		$+$		$-$	$+\epsilon_5$
$y_6 =$			$+$	$-$	$+\epsilon_6$

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 ϵ_j are independent, and identically distributed and each μ_i is a mass difference. The standard assumption is that the ϵ_j have a $N(0, \sigma_\epsilon^2)$ distribution, σ_ϵ unknown. The least squares estimates of μ_1 (calculated separately at each point in time) are depicted in Figure 2.

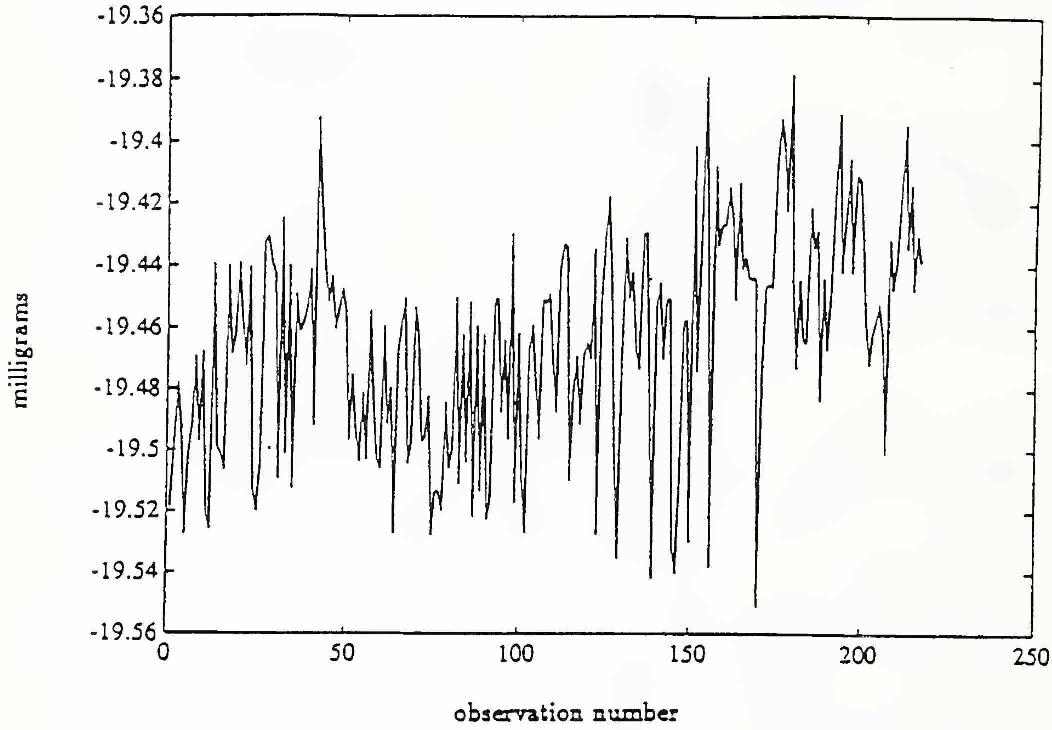


Figure 2: 217 estimates in milligrams of mass differences of two standard weights of 1 kilogram each, made at the NIST between 1975 and 1988.

The mass differences are measured in milligrams. Appendix 1 contains the numerical data. (We will henceforth refer to these estimates of μ_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

Almer 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 Shiryayev-Roberts approach to surveillance, leaving a theoretical justification to the later sections. The paper is therefore organized in the following way. After presenting the Shiryayev-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 Shiryayev-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 ν, 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 survelliance 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_ν. Expectation will be denoted by E_ν. Probability and expectation when there is no change throughout the sequence will be denoted by P_∞ and E_∞, 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

$$E_\infty N \geq B \tag{1}$$

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 3σ-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

$$\sup_{1 \leq \nu < \infty} E_\nu(N - \nu \mid N \geq \nu).$$

A basic statistic when conducting a surveilllance 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

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

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 μ_0 nor σ are known. Therefore one cannot compute Λ_k^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 propositon. 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 $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ and construct the sequence of standardized recursive residuals (Brown, Durbin and Evans, 1975)

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

The distribution of the sequence of Y_i's is independent of μ_0. Now construct the sequence

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

The distribution of the sequence of Z_i's is independent of both μ_0 and σ. 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 Λ_k^n (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 Λ_k^n, 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 $\Lambda_1^n \equiv 1$. 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 $\Lambda_1^n \equiv 1$.

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

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

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

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

As detailed in Appendix 5, the ratio of the integrals in Λ_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 Shiryayev-Roberts procedure would be a 2−sided 3σ-limit Shewhart chart. As mentioned above, the ARL to false alarm of that procedure is ≈ 370. Hence we should specify $B = 370$ for the Shiryayev-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

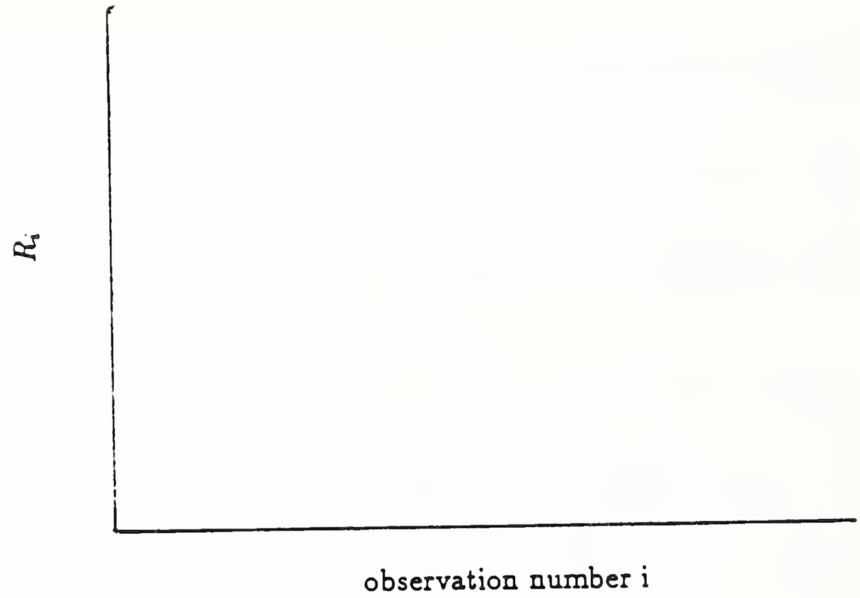


Figure 3: The Shiryayev-Roberts control chart with stopping threshold A.

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

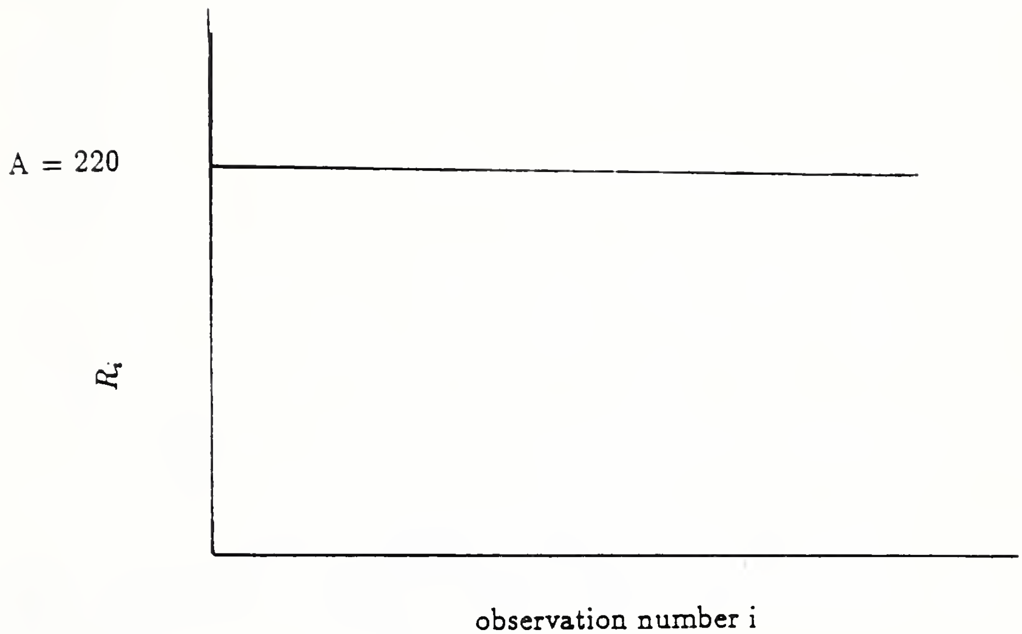


Figure 4: The Shiryayev-Roberts control chart with stopping threshold $A = 220$.

unreasonable. (See Figure 14.)

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_i 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.

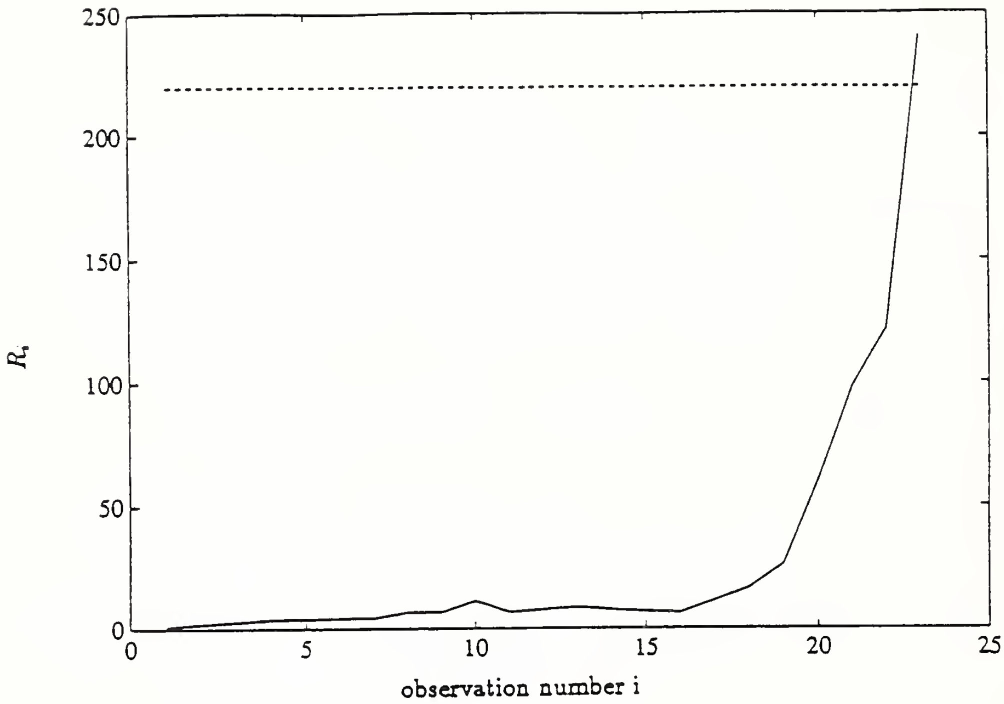


Figure 5: Parametric surveillance for a change of the mean of the mass difference data: $R_i, 1 \leq N_{220} = 23$.

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 fluctations which should not set off an alarm, then one would fix $B > 10000$. To fix B more precisely, one can regard the expected delay $E_\nu(N - \nu \mid N \geq \nu)$. 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 $N(\mu_0 - \delta\sigma, \sigma^2)$)

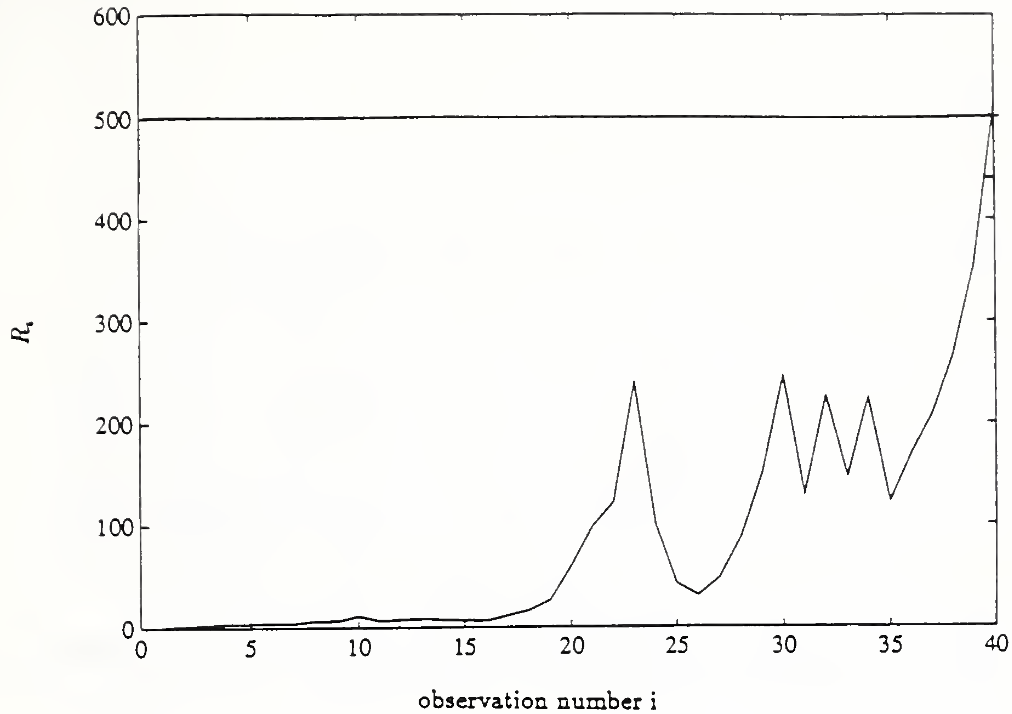


Figure 6: Parametric surveillance for a change of the mean of the mass difference data: $R_i, 1 \leq N_{500} = 40$.

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 ν not close to 1 and $\mu > \delta/2$) is $(logA)/[\delta(\mu - \delta/2)]$. 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 $2logA$. 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.

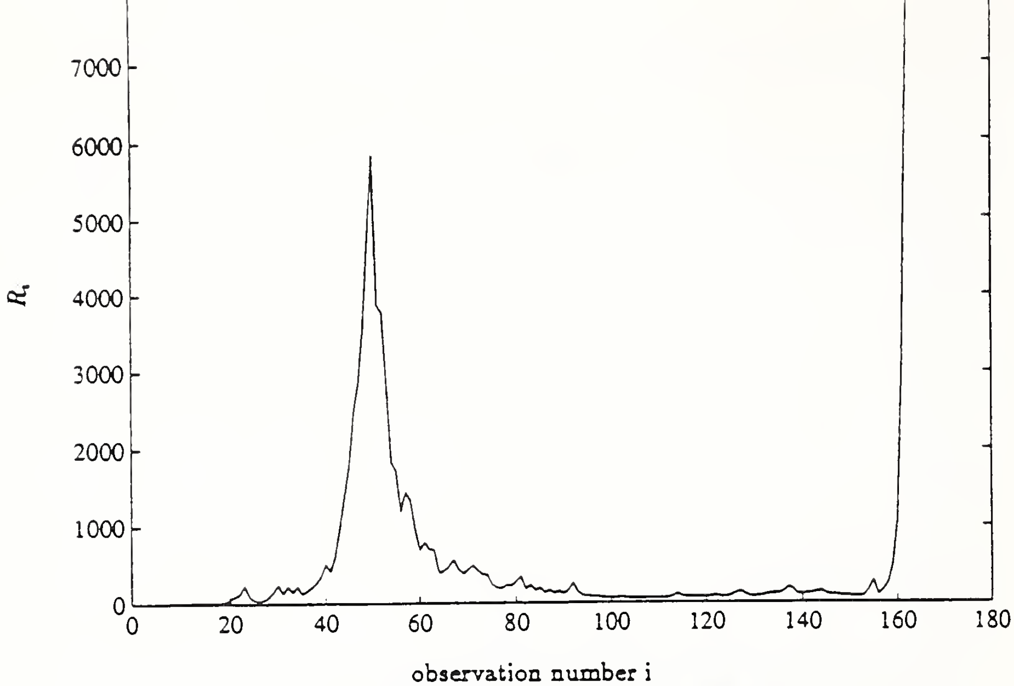


Figure 7: Parametric surveillance for a change of the mean of the mass difference data: $R_i, 1 \leq N_{6000} = 162$.

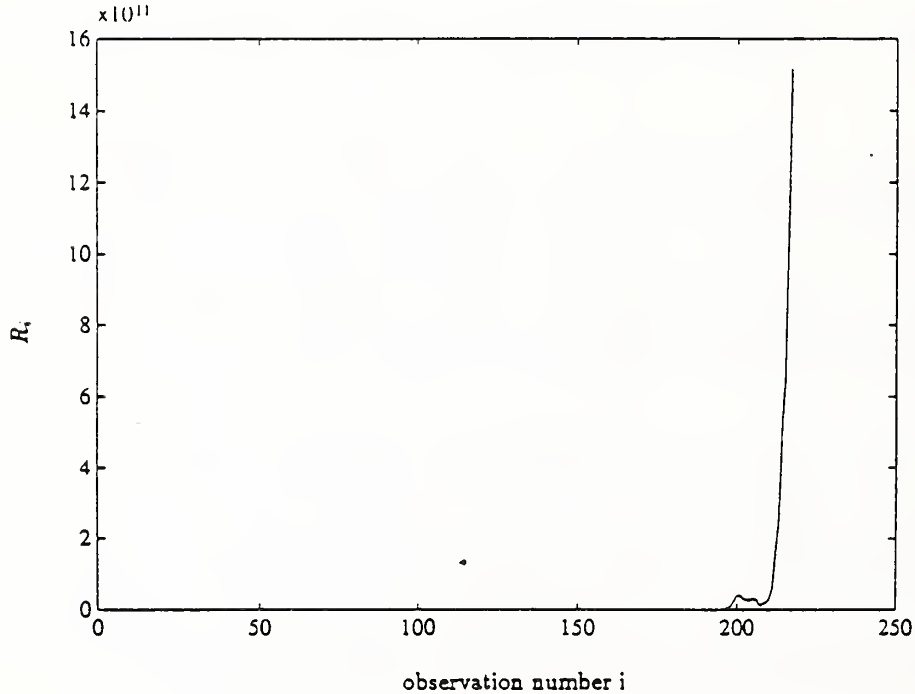


Figure 8: Parametric surveillance for a change of the mean of the mass difference data: $R_i, 1 \leq i \leq 217$.

5 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 ith sequential rank $r_i = \sum_{j=1}^{i} I(X_j \leq X_i)$, which is the rank of the ith 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 Shiryayev-Roberts (NPSR) procedure will be based on the sequence $R_n = \sum_{k=1}^{n} \Lambda_k^n, n = 1, 2, \ldots,$ where Λ_k^n 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_i.

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. Suppose again that one is alert for a change of $\pm\sigma$ in the mean. 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.

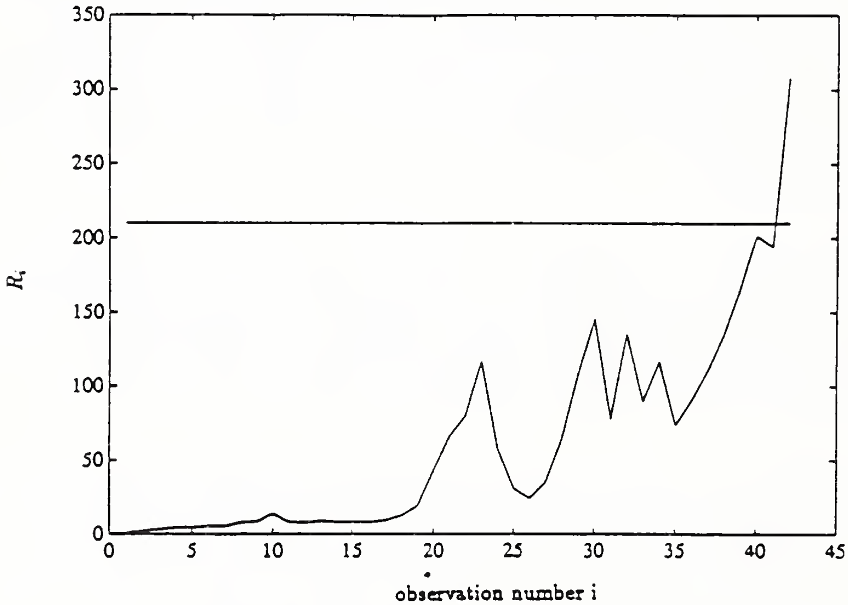


Figure 9: Nonparametric surveillance for a change of the mean of the mass difference data: $R_i, 1 \leq i \leq N_{210} = 42$.

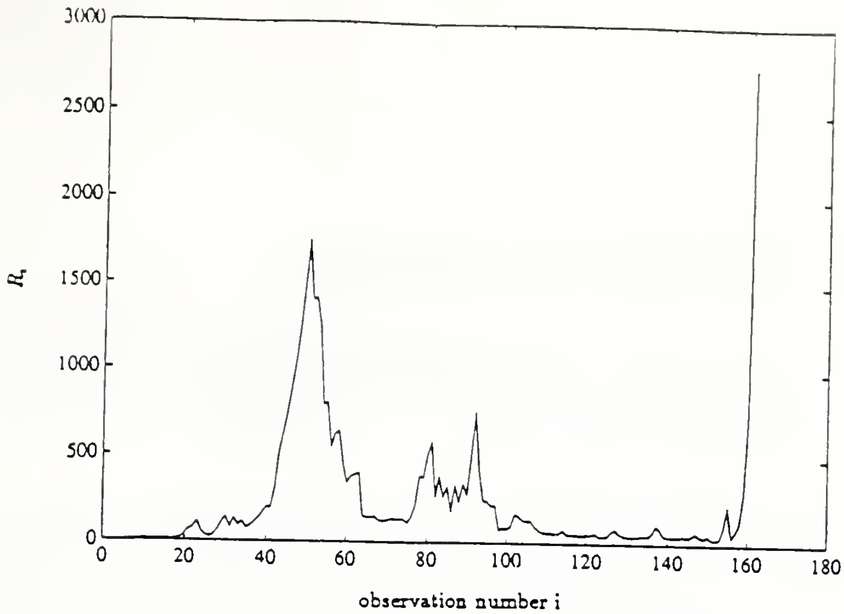


Figure 10: Nonparametric surveillance for a change of the mean of the mass difference data: $R_i, 1 \leq i \leq 161$.

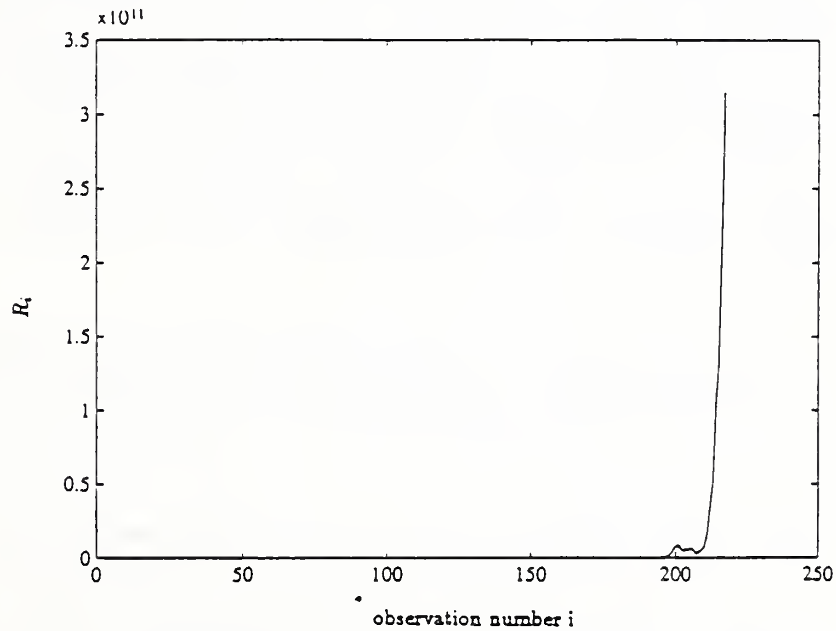


Figure 11: Nonparametric surveillance for a change of the mean of the mass difference data: $R_i, 1 \leq i \leq 217$.

6 Post-detection analysis

Suppose the series R_i 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 uncertainity 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

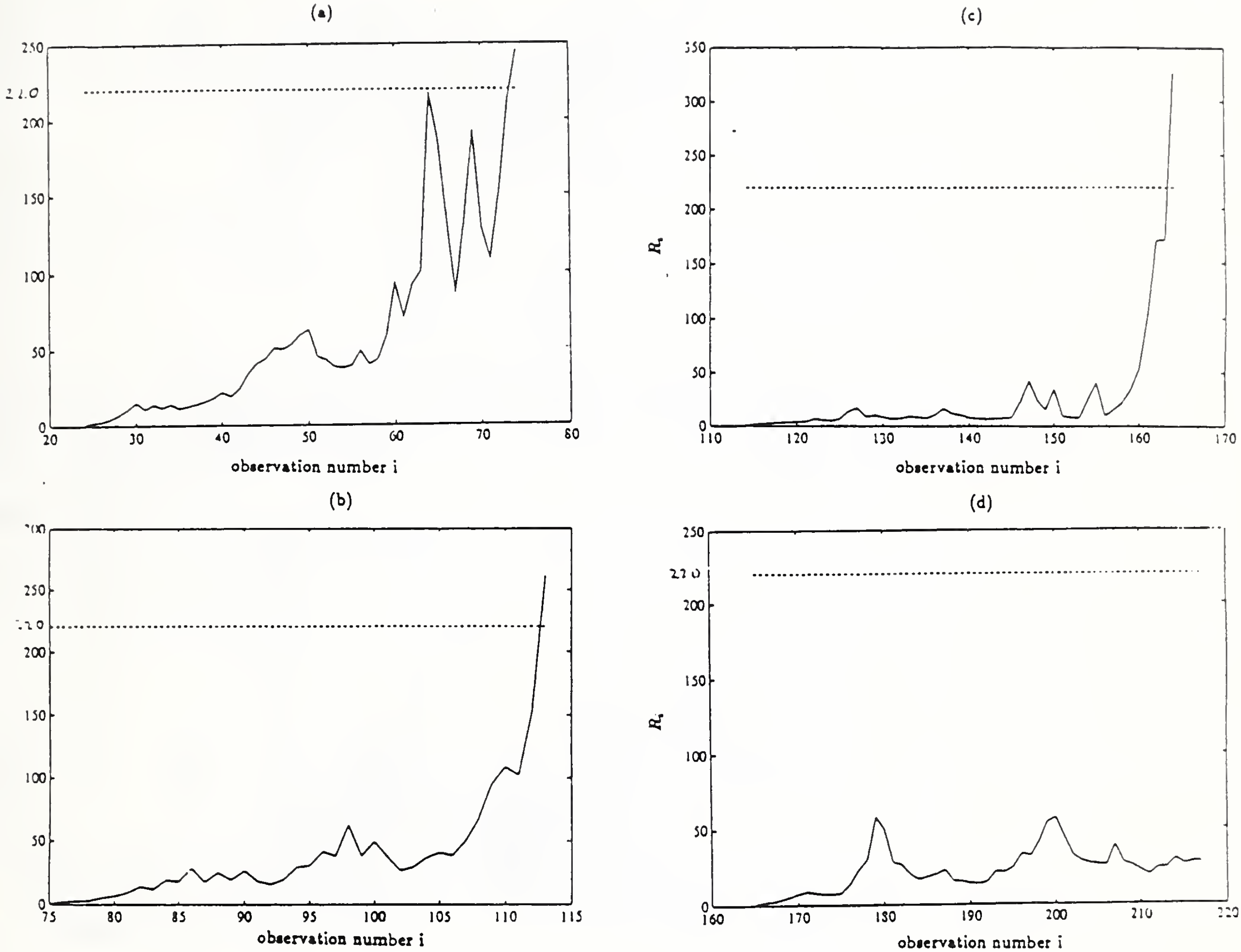


Figure 12: Post-detection parametric analysis of mass difference data: surveillance for a change of mean. ARL to false alarm=370, starting anew after each observation. (a) $R_i, 24 \le i \le N_{220} = 74$. (b) $R_i, 75 \le i \le N_{220} = 113$. (c) $R_i, 114 \le i \le N_{220} = 164$. (d) $R_i, 165 \le i \le 217$.

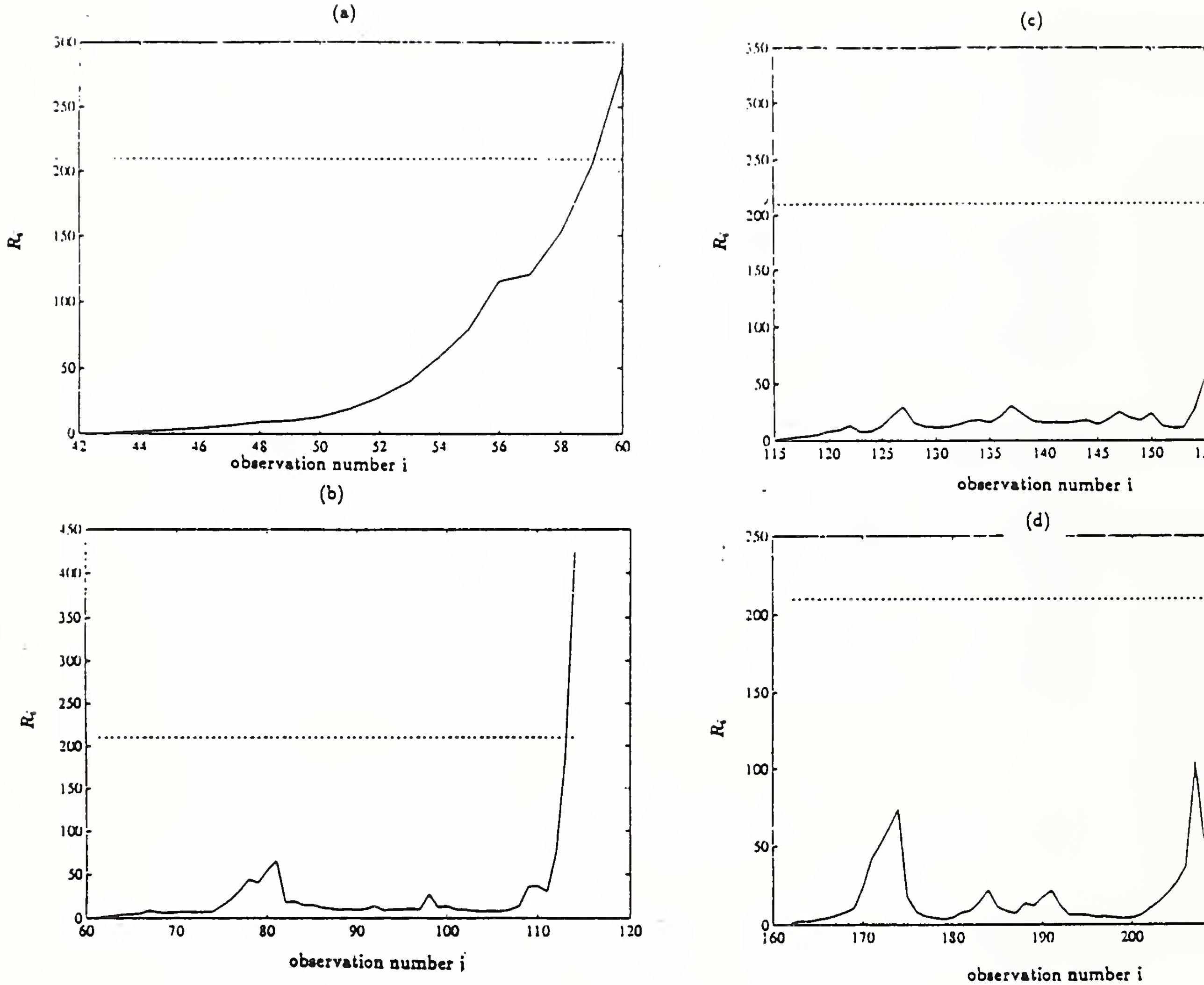


Figure 13: Post-detection nonparametric analysis of mass data: surveillance for a change of mean. ARL to false alarm=370, starting anew after each observation. (a) $R_i, 43 \leq i \leq N_{210} = 60$. (b) $R_i, 61 \leq i \leq N_{210} = 114$. (c) $R_i, 115 \leq i \leq N_{210} = 161$. (d) $R_i, 162 \leq i \leq 217$.

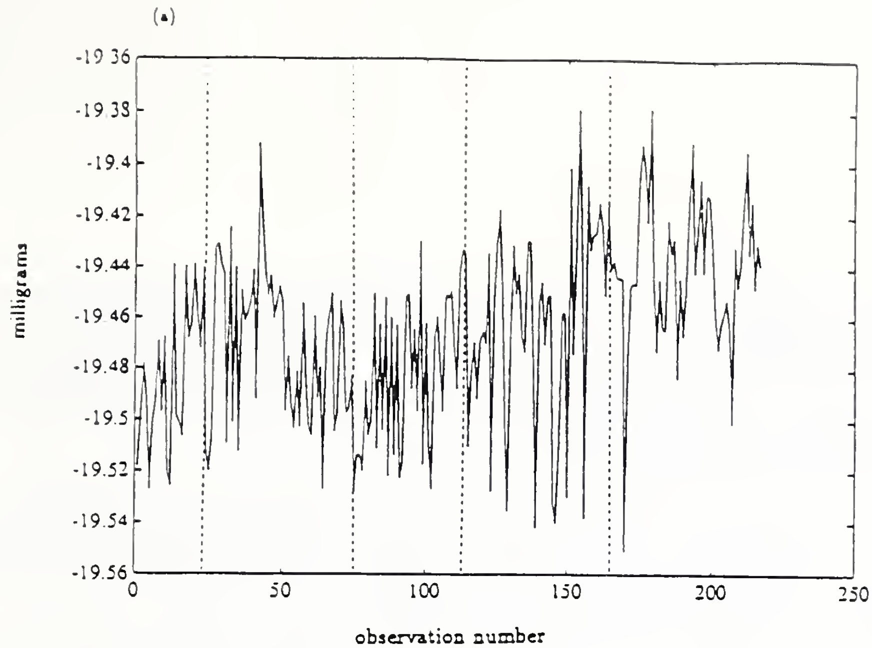


Figure 14: Surveillance segments of mass difference data, based on parametric control for change of mean. ARL to false alarm=370. Putative change= 1 standard deviation.

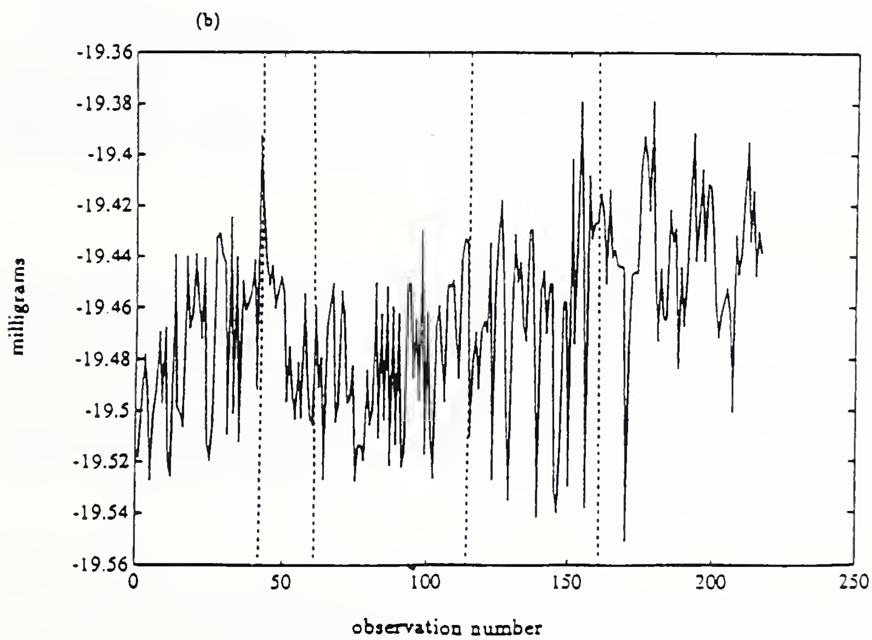


Figure 15: Surveillance segments of mass difference data, based on non-parametric control for change of mean. ARL to false alarm=370. Putative change= 1 standard deviation.

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 Shiryayev-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 Shiryayev-Roberts procedure, as suggested above. We conjecture that (1) is satisfied approximately. The reasoning: Λ_k^n for small k do not play a dominant role in $R_n = \sum_k^n \Lambda_k^n$, 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 ν 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 Λ_k^n are calculated accordingly.

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

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

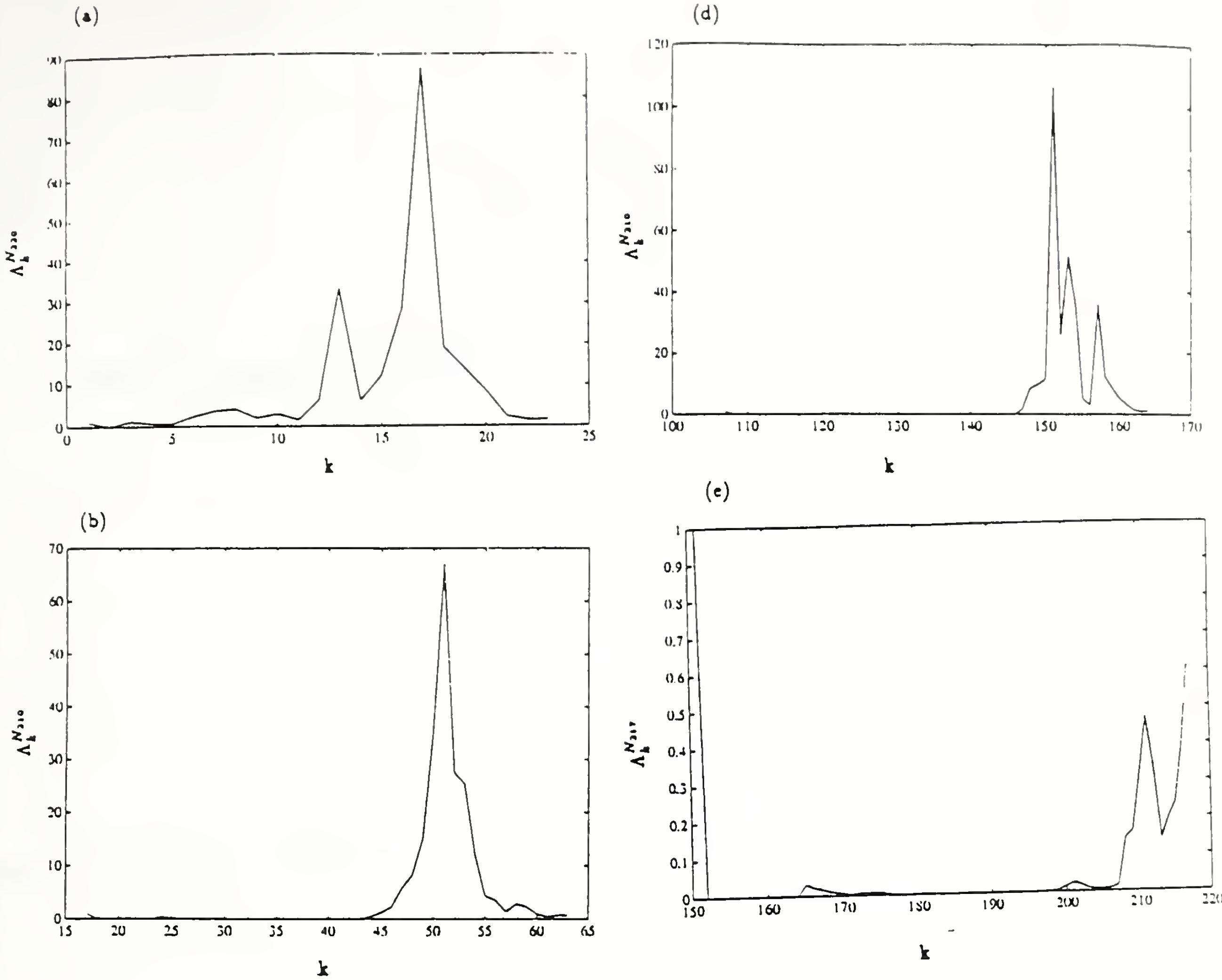
and we will stop and declare that a change is in effect at N_{220}, the first time that R_n exceeds 220. The resulting control chart is given in Figure 17(a). (In order to facilitate reference to Figure 2 and Appendix 1, the index i of the R_i is translated back again to match the original serial numbers 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.

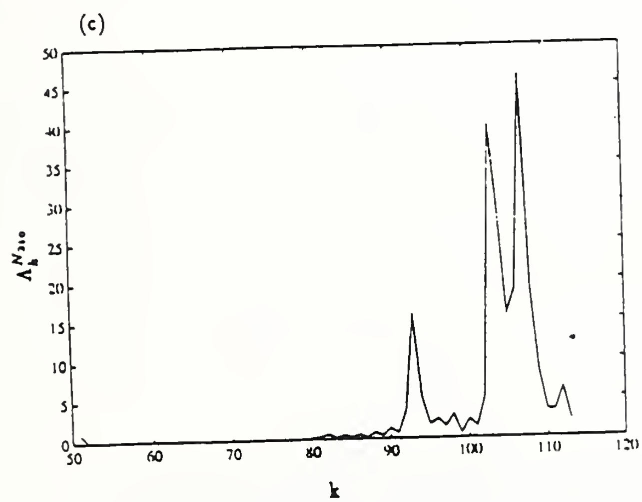
From Figure 16(b), we obtain $\hat{\nu} = 51$. Since the next observation will be the $64th$, it means that to continue we should set $kay = 14$. The resulting analysis is given in Figure 17(b) and Figure 16(b). We stop again after the $113th$ observation, and $\hat{\nu} = 107$.

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 $\hat{\nu} = 151$.

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, Λ_k^{217} 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.





(a) $1 \leq k \leq 23; \hat{\nu} = 17; n = N_{220} = 23.$

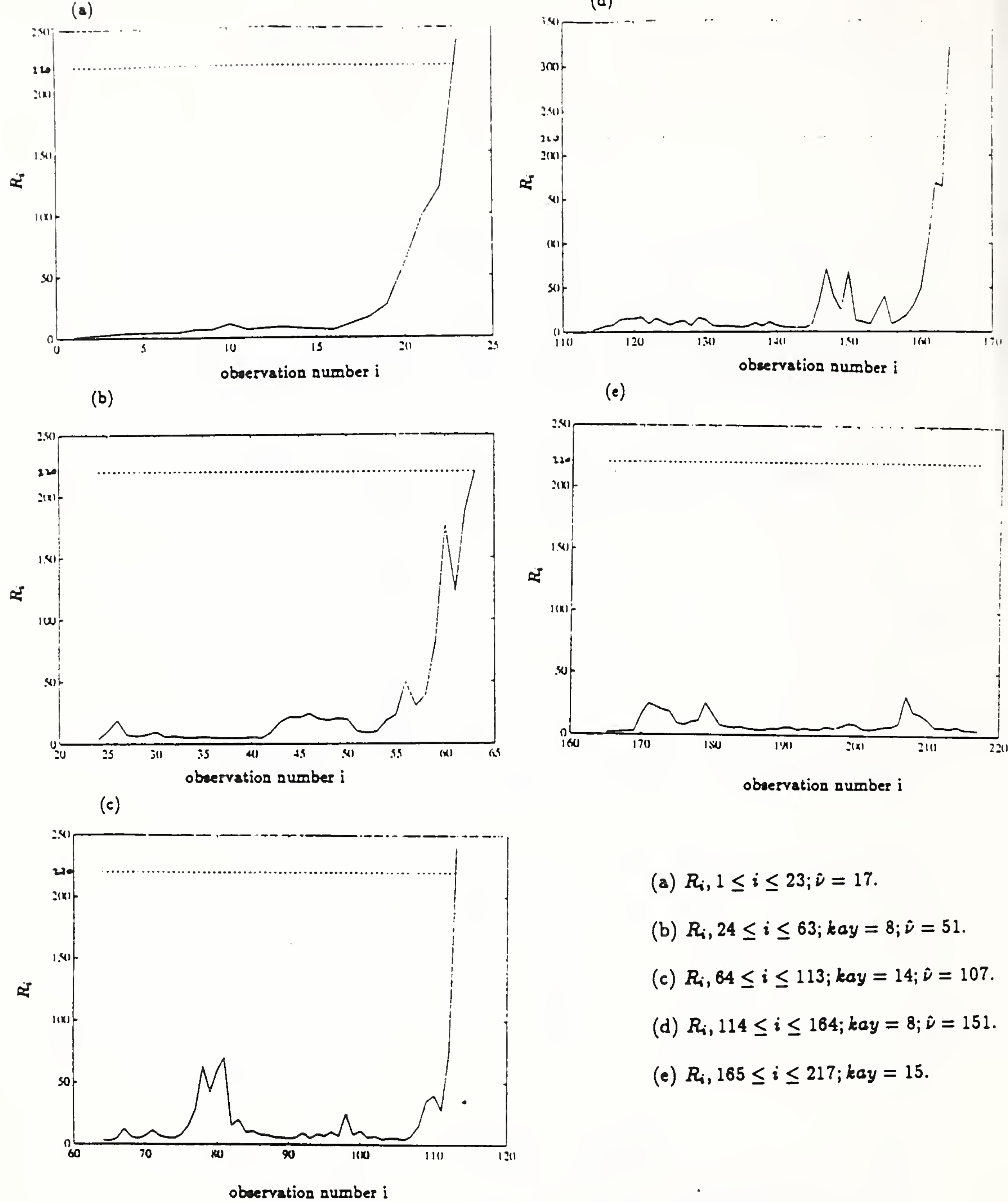
(b) $17 \leq k \leq 63; \hat{\nu} = 51; n = N_{220} = 63.$

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

(d) $107 \leq k \leq 164; \hat{\nu} = 151; n = N_{220} = 164.$

(e) $151 \leq k \leq 217; n = 217.$

Figure 16: Post-detection parametric analysis of mass difference data also using pre-detection data for surveillance for a change in mean: Λ_k^n as a function of k.



(a) $R_i, 1 \leq i \leq 23; \hat{\nu} = 17$.

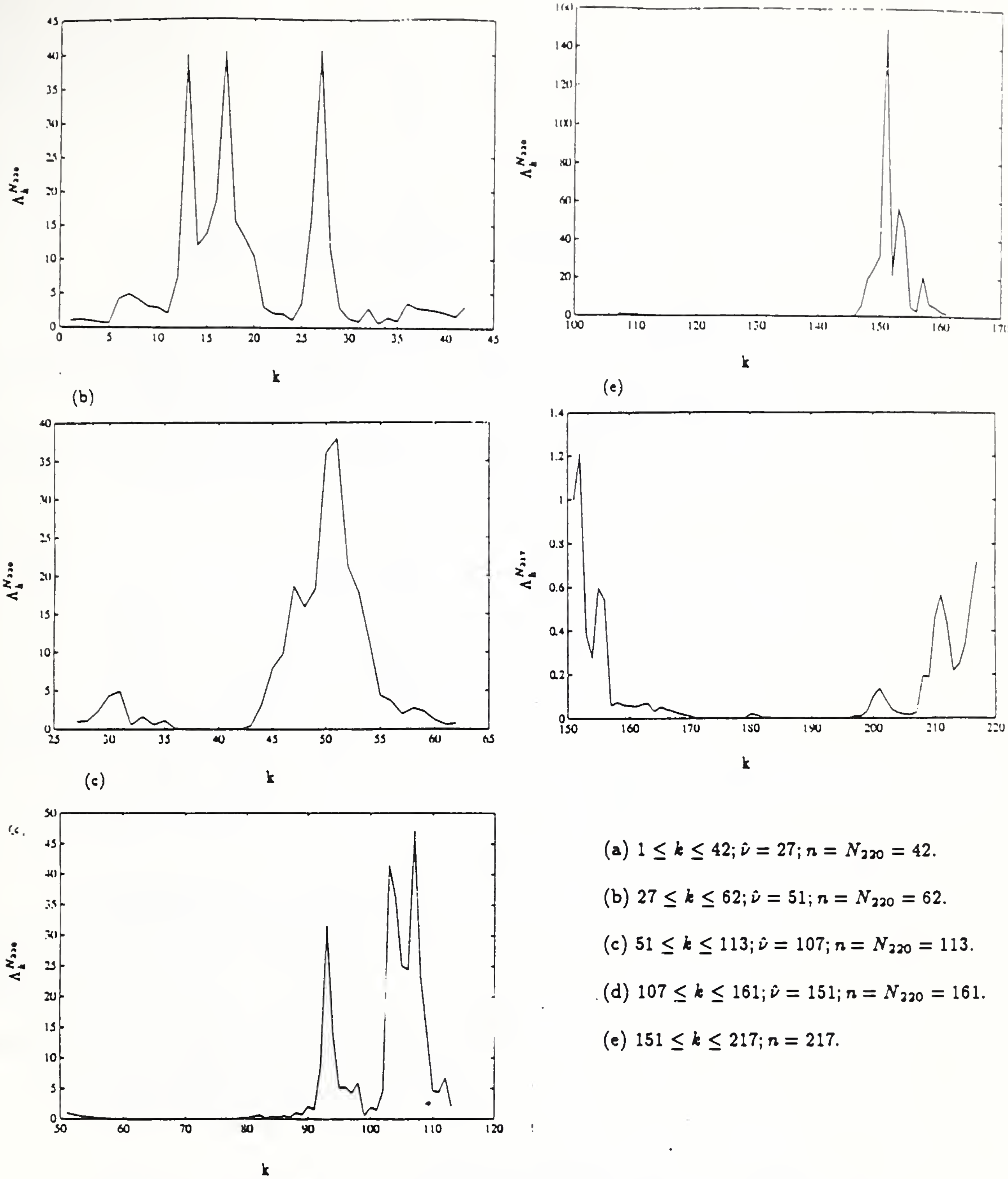
(b) $R_i, 24 \leq i \leq 63; kay = 8; \hat{\nu} = 51$.

(c) $R_i, 64 \leq i \leq 113; kay = 14; \hat{\nu} = 107$.

(d) $R_i, 114 \leq i \leq 164; kay = 8; \hat{\nu} = 151$.

(e) $R_i, 165 \leq i \leq 217; kay = 15$.

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

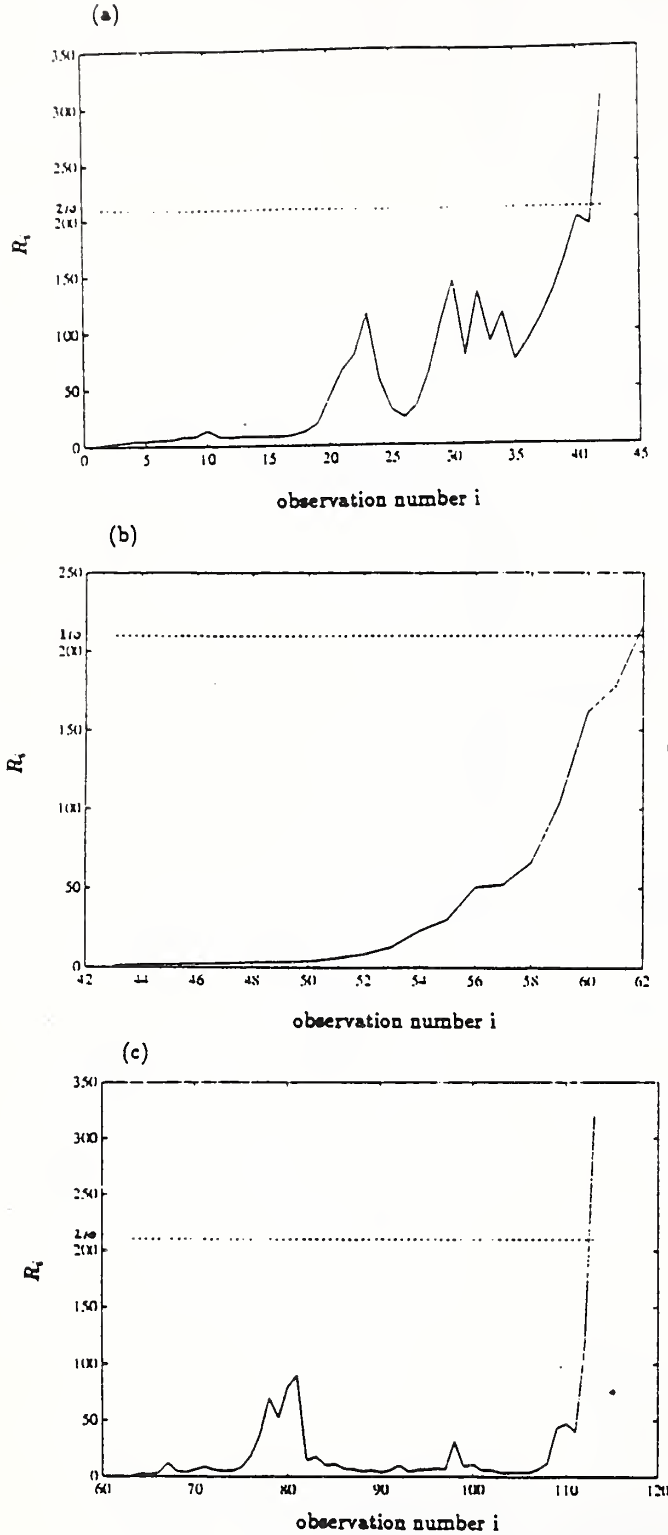
(b) $27 \leq k \leq 62; \hat{\nu} = 51; n = N_{220} = 62.$

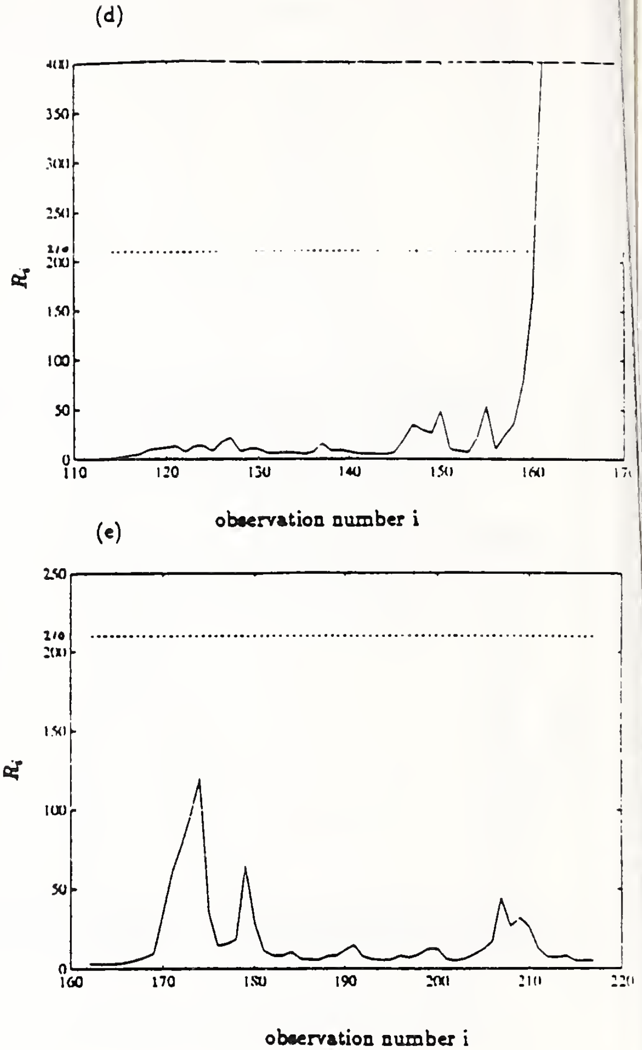
(c) $51 \leq k \leq 113; \hat{\nu} = 107; n = N_{220} = 113.$

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

(e) $151 \leq k \leq 217; n = 217.$

Figure 18: Post-detection nonparametric analysis of mass difference data also using pre-detection data for surveillance for a change in mean: Λ_k^n as a function of k.





(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 nonparametrically (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 σ_ϵ. The residual standard deviation is given by $s = \sqrt{\frac{1}{3}\sum_{j=1}^{6}(y_j - \hat{y}_j)^2}$, where $\hat{y}_1 = \hat{\mu}_1$, $\hat{y}_2 = \hat{\mu}_2$, $\hat{y}_3 = \hat{\mu}_3$, $\hat{y}_4 = \hat{\mu}_2 - \hat{\mu}_1$, $\hat{y}_5 = \hat{\mu}_3 - \hat{\mu}_1$, $\hat{y}_6 = \hat{\mu}_3 - \hat{\mu}_2$. (Note that in the absence of components of error other than ϵ_i, the σ of Sections 3-6 would equal $\sigma_\epsilon/\sqrt{2}$). Estimates of σ_ϵ from each intercomparison design are given in the $6th$ column of Appendex 1 and plotted in Figure 20.

Since three parameters of mean (μ_1, μ_2, μ_3) are estimated and there are six observations, the distribution of each $3s^2/\sigma_\epsilon^2$ is $\chi^2_{(3)} = Gamma(1.5, .5)$. So letting Y_i denote the ith estimate of σ_ϵ, the sequence of observations Y_i with unknown baseline σ_ϵ is being monitored for a change. In order to get rid of the nuisance parameter σ_ϵ, 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 σ_ϵ. (Since our procedure is based on likelihood ratios, any eqivalent invariant set

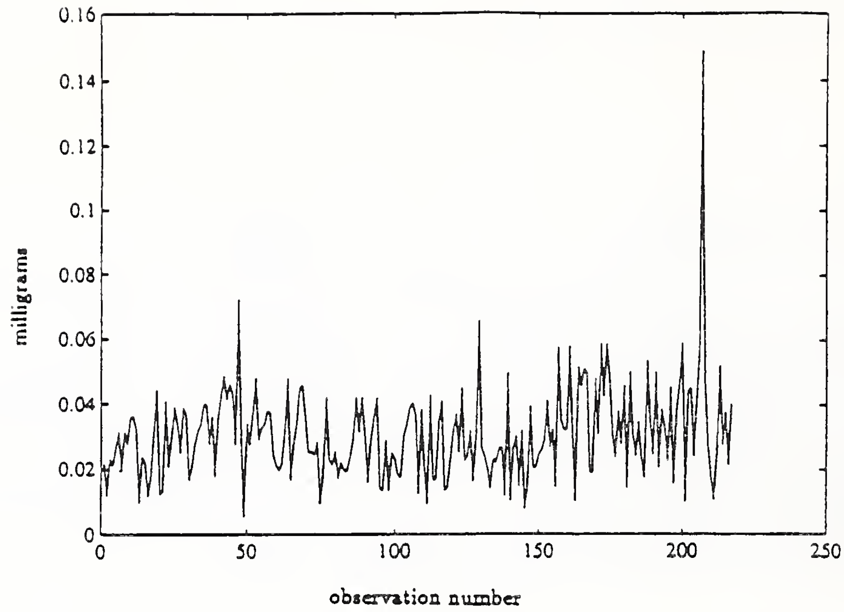


Figure 20: 217 estimates in milligrams of the standard deviation, σ_ϵ of the difference between two 1 kilogram weights, made at NIST between 1975 and 1988.

of statistics - such as $Z_n = Y_n/Y_1$ - will yield the same Λ_k^n'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 $g\sigma_\epsilon$ (or more extreme); that is, if after change the value of the standard deviation becomes $g\sigma_\epsilon$ (or more extreme), it would be of interest to raise an alarm. Then the $\nu = k$ versus $\nu = \infty$ likelihood Λ_k^n of $T_2, \ldots, T_n$ for $n \geq k \geq 1$ (Gordon and Pollak, 1990, Theorem 2) is

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

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 $E_\infty N_A/A \approx 2.6$ (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.

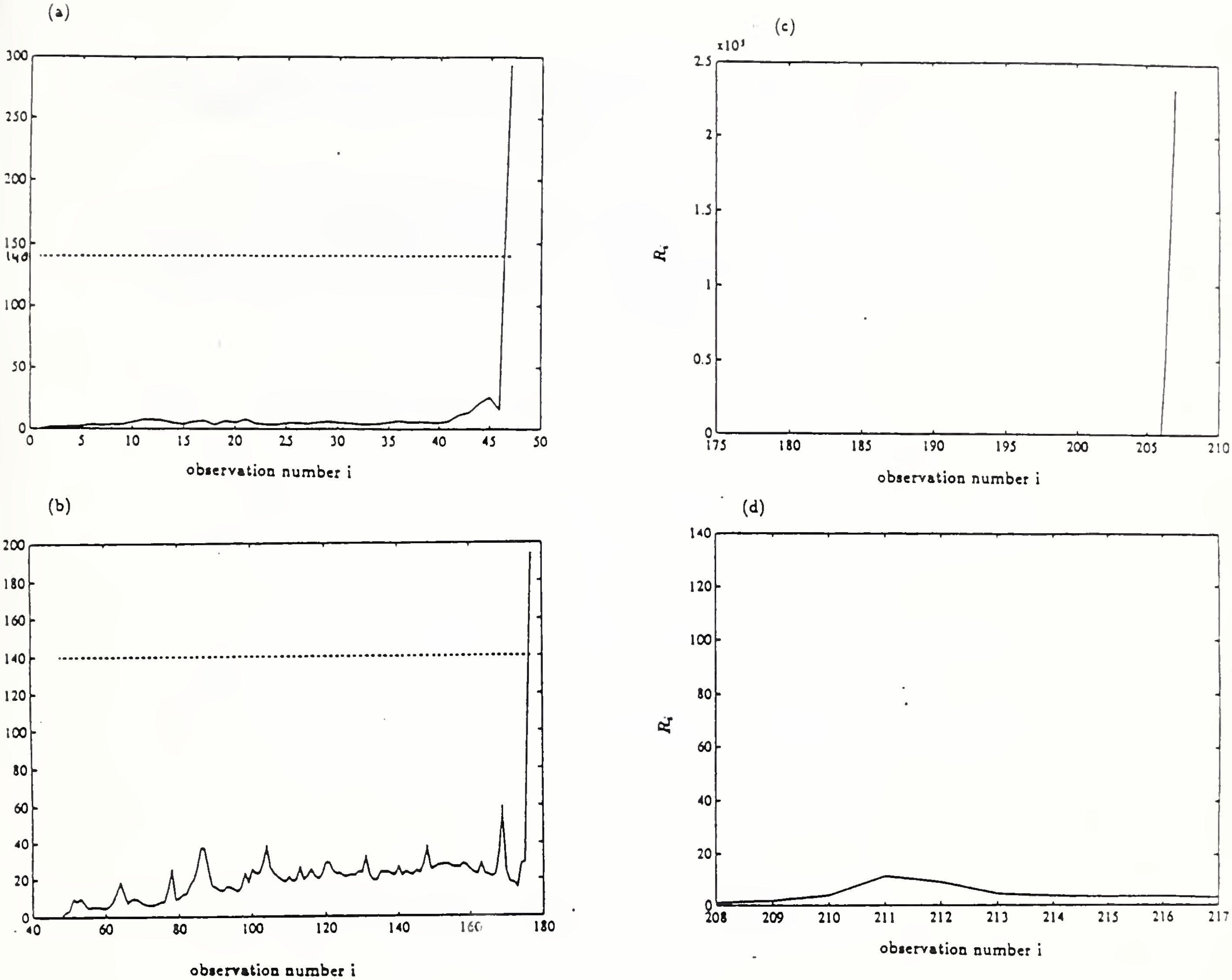


Figure 21: Parametric analysis of mass difference data: surveillance for a change of standard deviation. ARL to false alarm=370, starting anew after each detection. (a) $R_i, 1 \leq i \leq N_{140} = 47$. (b) $R_i, 48 \leq i \leq N_{140} = 177$. (c) $R_i, 178 \leq i \leq N_{140} = 207$. (d) $R_i, 208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

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.

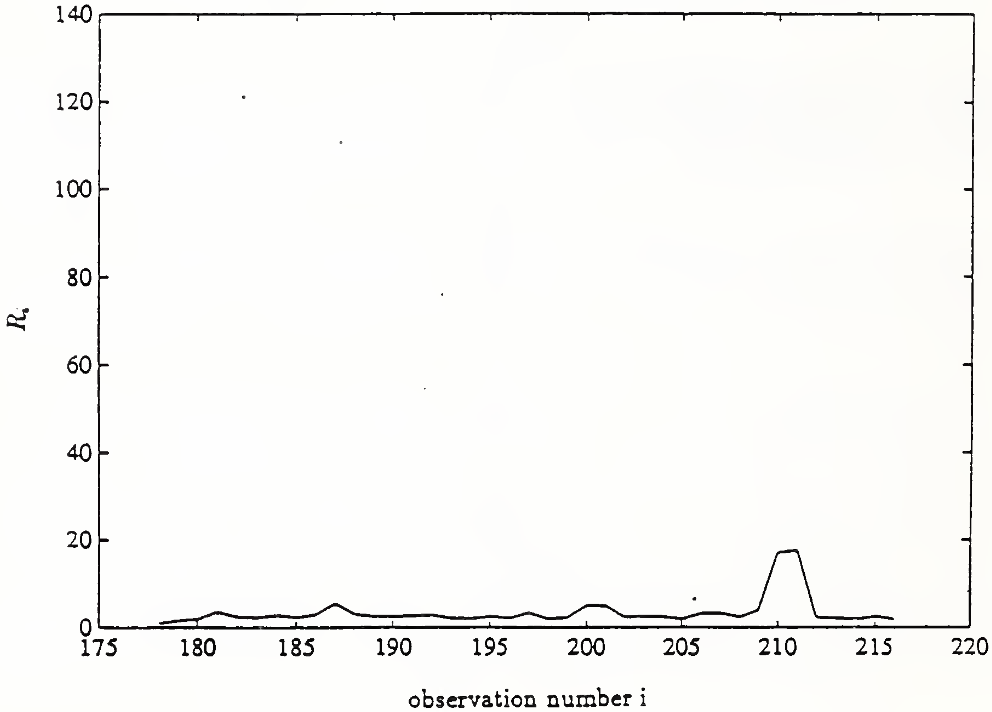


Figure 22: Parametric surveillance for a change of standard deviation, starting with observations #178, deleting observation #207; ARL to false alarm=370, $R_i, 178 \leq i \leq 216$.

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 recognizing 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 $150th$ obervation 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

1	75.884	41.	-19.51836	12.	.0217	41.
2	75.922	41.	-19.49785	12.	.0118	41.
3	76.013	41.	-19.47795	12.	.0232	41.
4	76.129	41.	-19.49223	12.	.0210	41.
5	76.642	41.	-19.52728	12.	.0265	41.
6	76.723	41.	-19.50172	12.	.0317	41.
7	77.220	41.	-19.49191	12.	.0194	41.
8	77.277	41.	-19.46912	12.	.0316	41.
9	77.551	41.	-19.49717	12.	.0274	41.
10	77.742	41.	-19.46766	12.	.0361	41.
11	77.742	41.	-19.52115	12.	.0362	41.
12	77.839	41.	-19.52576	12.	.0320	41.
13	77.954	41.	-19.43922	12.	.0096	41.
14	77.954	41.	-19.49868	12.	.0238	41.
15	78.097	41.	-19.50119	12.	.0224	41.
16	78.218	41.	-19.50662	12.	.0117	41.
17	78.220	41.	-19.43984	12.	.0175	41.
18	78.277	41.	-19.46826	12.	.0314	41.
19	78.398	41.	-19.46286	12.	.0445	41.
20	79.113	41.	-19.43884	12.	.0122	41.
21	79.137	41.	-19.45812	12.	.0132	41.
22	79.140	41.	-19.47232	12.	.0409	41.
23	79.142	41.	-19.44036	12.	.0206	41.
24	79.153	41.	-19.51258	12.	.0295	41.
25	79.183	41.	-19.51992	12.	.0391	41.
26	79.188	41.	-19.50532	12.	.0339	41.
27	79.199	41.	-19.43229	12.	.0250	41.
28	79.242	41.	-19.43039	12.	.0391	41.
29	79.261	41.	-19.43908	12.	.0365	41.
30	79.261	41.	-19.44265	12.	.0164	41.
31	79.312	41.	-19.50940	12.	.0203	41.
32	79.312	41.	-19.42423	12.	.0274	41.
33	79.315	41.	-19.50140	12.	.0317	41.
34	79.315	41.	-19.43997	12.	.0338	41.
35	79.503	41.	-19.51249	12.	.0399	41.
36	79.530	41.	-19.44913	12.	.0399	41.
37	79.546	41.	-19.46149	12.	.0275	41.
38	79.661	41.	-19.45838	12.	.0362	41.
39	79.680	41.	-19.45370	12.	.0177	41.
40	79.731	41.	-19.44111	12.	.0356	41.
41	79.812	41.	-19.49213	12.	.0424	41.
42	79.965	41.	-19.39213	12.	.0487	41.
43	80.199	41.	-19.41918	12.	.0413	41.
44	80.363	41.	-19.44048	12.	.0463	41.
45	80.390	41.	-19.45164	12.	.0431	41.
46	80.390	41.	-19.44348	12.	.0275	41.
47	80.538	41.	-19.46043	12.	.0724	41.
48	81.645	41.	-19.45428	12.	.0297	41.
49	81.653	41.	-19.44764	12.	.0054	41.
50	82.030	41.	-19.45455	12.	.0340	41.
51	82.215	41.	-19.49673	12.	.0275	41.
52	82.304	41.	-19.47514	12.	.0376	41.
53	82.384	41.	-19.49357	12.	.0482	41.
54	82.444	41.	-19.50361	12.	.0290	41.
55	82.481	41.	-19.48108	12.	.0326	41.
56	82.489	41.	-19.50317	12.	.0338	41.
57	82.489	41.	-19.45445	12.	.0381	41.
58	82.737	41.	-19.48382	12.	.0373	41.
59	82.804	41.	-19.50324	12.	.0245	41.
60	82.836	41.	-19.50624	12.	.0212	41.
61	82.839	41.	-19.45943	12.	.0194	41.
62	82.968	41.	-19.49146	12.	.0216	41.
63	83.051	41.	-19.47940	12.	.0322	41.
64	83.124	41.	-19.52737	12.	.0482	41.
65	83.277	41.	-19.46200	12.	.0165	41.

66	83.376	41.	-19.45857	12.	.0272	41.
67	83.570	41.	-19.45041	12.	.0336	41.
68	83.664	41.	-19.50471	12.	.0452	41.
69	83.841	41.	-19.49776	12.	.0459	41.
70	83.852	41.	-19.45337	12.	.0359	41.
71	83.855	41.	-19.46320	12.	.0251	41.
72	83.855	41.	-19.49746	12.	.0254	41.
73	83.855	41.	-19.49578	12.	.0243	41.
74	84.022	41.	-19.48234	12.	.0284	41.
75	84.024	41.	-19.52780	12.	.0094	41.
76	84.070	41.	-19.51346	12.	.0179	41.
77	84.086	41.	-19.51361	12.	.0423	41.
78	84.105	41.	-19.52000	12.	.0228	41.
79	84.140	41.	-19.48437	12.	.0211	41.
80	84.159	41.	-19.50596	12.	.0256	41.
81	84.242	41.	-19.49994	12.	.0170	41.
82	84.258	41.	-19.45018	12.	.0221	41.
83	84.293	41.	-19.51104	12.	.0196	41.
84	84.387	41.	-19.46231	12.	.0192	41.
85	84.425	41.	-19.50418	12.	.0237	41.
86	84.452	41.	-19.45169	12.	.0294	41.
87	84.492	41.	-19.52187	12.	.0422	41.
88	84.522	41.	-19.45963	12.	.0315	41.
89	84.540	41.	-19.51361	12.	.0424	41.
90	84.637	41.	-19.46240	12.	.0303	41.
91	84.774	41.	-19.52267	12.	.0159	41.
92	84.793	41.	-19.51634	12.	.0289	41.
93	84.809	41.	-19.45103	12.	.0363	41.
94	84.828	41.	-19.45056	12.	.0423	41.
95	84.911	41.	-19.48777	12.	.0140	41.
96	84.944	41.	-19.46407	12.	.0132	41.
97	84.949	41.	-19.49652	12.	.0290	41.
98	84.962	41.	-19.42983	12.	.0132	41.
99	84.989	41.	-19.51739	12.	.0251	41.
100	85.005	41.	-19.46179	12.	.0234	41.
101	85.016	41.	-19.50977	12.	.0182	41.
102	85.024	41.	-19.52705	12.	.0173	41.
103	85.027	41.	-19.46708	12.	.0305	41.
104	85.083	41.	-19.45906	12.	.0340	41.
105	85.094	41.	-19.47887	12.	.0392	41.
106	85.613	41.	-19.49657	12.	.0407	41.
107	85.677	41.	-19.45073	12.	.0367	41.
108	85.718	41.	-19.45205	12.	.0124	41.
109	85.755	41.	-19.44903	12.	.0387	41.
110	85.758	41.	-19.47174	12.	.0207	41.
111	85.788	41.	-19.48791	12.	.0094	41.
112	85.887	41.	-19.44142	12.	.0430	41.
113	85.903	41.	-19.43279	12.	.0167	41.
114	85.941	41.	-19.43431	12.	.0167	41.
115	86.070	41.	-19.51034	12.	.0351	41.
116	86.137	41.	-19.48098	12.	.0412	41.
117	86.156	41.	-19.46938	12.	.0134	41.
118	86.172	41.	-19.49175	12.	.0143	41.
119	86.183	41.	-19.46907	12.	.0226	41.
120	86.183	41.	-19.46479	12.	.0320	41.
121	86.199	41.	-19.47025	12.	.0373	41.
122	86.202	41.	-19.43449	12.	.0252	41.
123	86.202	41.	-19.52758	12.	.0454	41.
124	86.312	41.	-19.45366	12.	.0226	41.
125	86.323	41.	-19.43107	12.	.0241	41.
126	86.325	41.	-19.41757	12.	.0321	41.
127	86.328	41.	-19.44877	12.	.0162	41.
128	86.333	41.	-19.50286	12.	.0296	41.
129	86.333	41.	-19.53529	12.	.0663	41.
130	86.333	41.	-19.46749	12.	.0266	41.
131	86.336	41.	-19.43112	12.	.0241	41.

132	86.336	41.	-19.45043	12.	.0204	41.
133	86.344	41.	-19.44215	12.	.0140	41.
134	86.349	41.	-19.46783	12.	.0233	41.
135	86.352	41.	-19.47338	12.	.0222	41.
136	86.352	41.	-19.42965	12.	.0267	41.
137	86.352	41.	-19.42965	12.	.0267	41.
138	86.352	41.	-19.48007	12.	.0117	41.
139	86.355	41.	-19.54186	12.	.0500	41.
140	86.363	41.	-19.45274	12.	.0102	41.
141	86.368	41.	-19.44563	12.	.0267	41.
142	86.384	41.	-19.47038	12.	.0306	41.
143	86.430	41.	-19.45087	12.	.0147	41.
144	86.438	41.	-19.45068	12.	.0324	41.
145	86.460	41.	-19.53248	12.	.0079	41.
146	86.473	41.	-19.54010	12.	.0143	41.
147	86.473	41.	-19.51144	12.	.0398	41.
148	86.478	41.	-19.45766	12.	.0205	41.
149	86.478	41.	-19.45766	12.	.0205	41.
150	86.489	41.	-19.52988	12.	.0247	41.
151	86.505	41.	-19.40143	12.	.0260	41.
152	86.508	41.	-19.47424	12.	.0291	41.
153	86.511	41.	-19.43668	12.	.0415	41.
154	86.511	41.	-19.37900	12.	.0267	41.
155	86.513	41.	-19.42834	12.	.0326	41.
156	86.513	41.	-19.53802	12.	.0147	41.
157	86.516	41.	-19.40812	12.	.0578	41.
158	86.522	41.	-19.43325	12.	.0355	41.
159	86.524	41.	-19.42678	12.	.0323	41.
160	86.524	41.	-19.42678	12.	.0323	41.
161	86.535	41.	-19.41479	12.	.0582	41.
162	86.540	41.	-19.42376	12.	.0309	41.
163	86.608	41.	-19.45083	12.	.0102	41.
164	86.648	41.	-19.41340	12.	.0521	41.
165	86.672	41.	-19.44090	12.	.0459	41.
166	86.677	41.	-19.43733	12.	.0514	41.
167	86.710	41.	-19.44411	12.	.0498	41.
168	86.723	41.	-19.44415	12.	.0191	41.
169	86.723	41.	-19.44415	12.	.0191	41.
170	86.734	41.	-19.55097	12.	.0483	41.
171	86.763	41.	-19.47477	12.	.0309	41.
172	86.801	41.	-19.44680	12.	.0590	41.
173	86.828	41.	-19.44596	12.	.0427	41.
174	86.833	41.	-19.44680	12.	.0590	41.
175	86.849	41.	-19.40648	12.	.0486	41.
176	86.858	41.	-19.39246	12.	.0307	41.
177	86.858	41.	-19.40120	12.	.0237	41.
178	86.995	41.	-19.42234	12.	.0381	41.
179	87.016	41.	-19.37848	12.	.0279	41.
180	87.019	41.	-19.44577	12.	.0459	41.
181	87.019	41.	-19.47299	12.	.0142	41.
182	87.022	41.	-19.44447	12.	.0504	41.
183	87.022	41.	-19.46441	12.	.0310	41.
184	87.134	41.	-19.46499	12.	.0241	41.
185	87.191	41.	-19.42126	12.	.0349	41.
186	87.228	41.	-19.43429	12.	.0233	41.
187	87.255	41.	-19.42879	12.	.0175	41.
188	87.290	41.	-19.48356	12.	.0539	41.
189	87.312	41.	-19.44399	12.	.0349	41.
190	87.333	41.	-19.46733	12.	.0246	41.
191	87.457	41.	-19.44769	12.	.0503	41.
192	87.481	41.	-19.41426	12.	.0206	41.
193	87.492	41.	-19.39112	12.	.0388	41.
194	87.524	41.	-19.44205	12.	.0332	41.
195	87.591	41.	-19.42656	12.	.0226	41.
196	87.694	41.	-19.40541	12.	.0456	41.
197	87.742	41.	-19.44178	12.	.0156	41.

198	87.788	41.	-19.41108	12.	.0377	41.
199	87.798	41.	-19.41254	12.	.0471	41.
200	87.801	41.	-19.43138	12.	.0592	41.
201	87.820	41.	-19.45783	12.	.0102	41.
202	87.847	41.	-19.47197	12.	.0445	41.
203	87.866	41.	-19.46213	12.	.0454	41.
204	87.922	41.	-19.45791	12.	.0240	41.
205	87.925	41.	-19.45214	12.	.0409	41.
206	87.957	41.	-19.46248	12.	.0570	41.
207	88.027	41.	-19.50108	12.	.1492	41.
208	88.027	41.	-19.43155	12.	.0469	41.
209	88.094	41.	-19.44742	12.	.0275	41.
210	88.151	41.	-19.43895	12.	.0180	41.
211	88.204	41.	-19.41520	12.	.0106	41.
212	88.207	41.	-19.39467	12.	.0252	41.
213	88.277	41.	-19.43458	12.	.0523	41.
214	88.315	41.	-19.41401	12.	.0275	41.
215	88.339	41.	-19.44789	12.	.0376	41.
216	88.398	41.	-19.43033	12.	.0215	41.
217	88.433	41.	-19.43883	12.	.0403	41.

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 $\mid \frac{X_n - \bar{X}_{n'}}{s_{n'}} \mid$ 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.

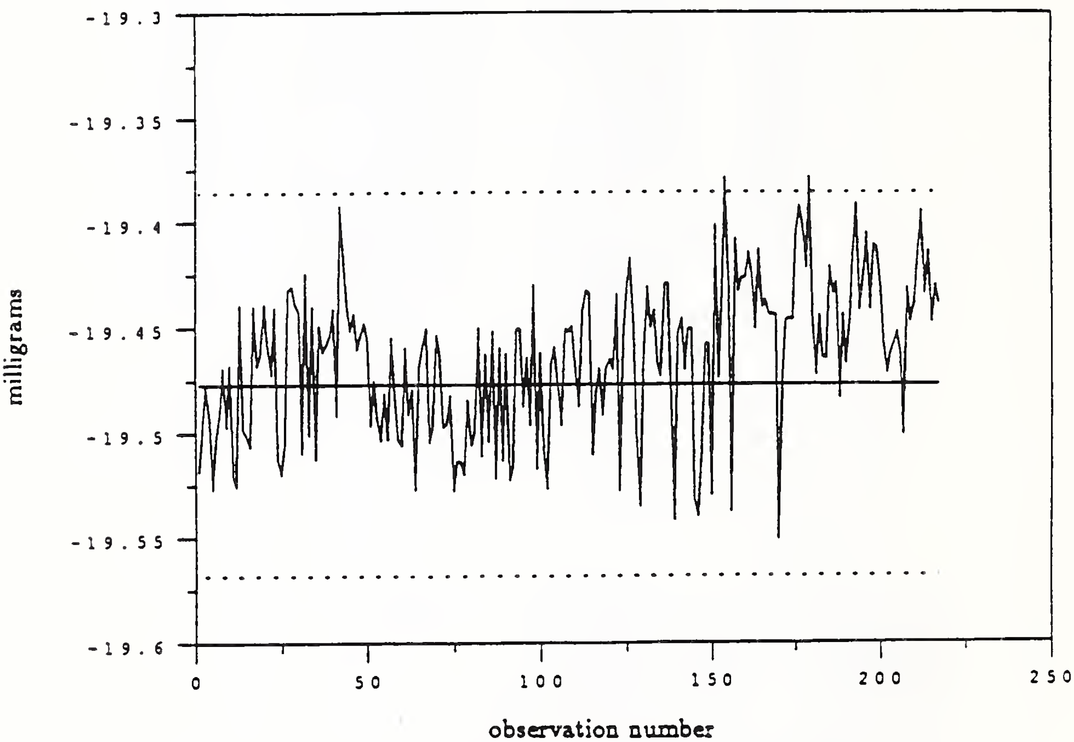


Figure 23: A Shewhart chart for the data of Figure 2.

The Shewart 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 3σ - limit until a half year later when #179 crossed the 3σ - 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 mgl. 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 ν 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, $P(\nu \leq n \mid X_1, \ldots, X_n)$. 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 Shiryayev, 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

$$\begin{aligned} P(\nu = k \mid X_1, \ldots, X_n) &= \frac{f_{\nu=k}(X_1, \ldots, X_n)pq^{k-1}}{\sum_{j=1}^{\infty} f_{\nu=j}(X_1, \ldots, X_n)pq^{j-1}} \\ &= \frac{f_{\nu=k}(X_1, \ldots, X_n)pq^{k-1}}{\sum_{j=1}^{n} f_{\nu=j}(X_1, \ldots, X_n)pq^{j-1} + f_{\nu=\infty}(X_1, \ldots, X_n)q^n}. \end{aligned}$$

Therefore the posterior probability that a change is in effect is

$$\begin{aligned} P(\nu \leq n \mid X_1, \ldots, X_n) &= \frac{\sum_{k=1}^{n} f_{\nu=k}(X_1, \ldots, X_n) p q^{k-1}}{\sum_{j=1}^{n} f_{\nu=j}(X_1, \ldots, X_n) p q^{j-1} + f_{\nu=\infty}(X_1, \ldots, X_n) q^n} \\ &= \frac{\sum_{k=1}^{n} \frac{f_{\nu=k}(X_1, \ldots, X_n)}{f_{\nu=\infty}(X_1, \ldots, X_n)} (\frac{1}{q})^{n-k+1}}{\sum_{k=1}^{n} \frac{f_{\nu=k}(X_1, \ldots, X_n)}{f_{\nu=\infty}(X_1, \ldots, X_n)} (\frac{1}{q})^{n-k+1} + \frac{1}{p}}. \end{aligned}$$

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

$$\sum_{k=1}^{n} \frac{f_{\nu=k}(X_1, \ldots, X_n)}{f_{\nu=\infty}(X_1, \ldots, X_n)} (\frac{1}{q})^{n-k+1} \tag{2}$$

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 Pollak (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. Pollak 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_∞ when k is fixed, the sequence $\Lambda_k^n, n \geq 1$ is a martingale with unit expectation:

$$\begin{aligned}
E_\infty(\Lambda_k^{n+1} \mid X_1,\ldots,X_n) &= E_\infty\left(\frac{f_{\nu=k}(X_1,\ldots,X_{n+1})}{f_{\nu=\infty}(X_1,\ldots,X_n)} \mid X_1,\ldots,X_{n+1}\right) \\
&= E_\infty\left(\frac{f_{\nu=k}(X_{n+1} \mid X_1,\ldots,X_n)f_{\nu=k}(X_1,\ldots,X_n)}{f_{\nu=\infty}(X_{n+1} \mid X_1,\ldots,X_n)f_{\nu=\infty}(X_1,\ldots,X_n)} \mid X_1,\ldots,X_n\right) \\
&= \frac{f_{\nu=k}(X_1,\ldots,X_n)}{f_{\nu=\infty}(X_1,\ldots,X_n)} E_\infty\left[\frac{f_{\nu=k}(X_{n+1} \mid X_1,\ldots,X_n)}{f_{\nu=\infty}(X_{n+1} \mid X_1,\ldots,X_n} \mid X_1,\ldots,X_n)\right] \\
&= \Lambda_k^n \int \frac{f_{X_{n+1}|X_1,\ldots,X_n;\nu=k}(x)}{f_{X_{n+1}|X_1,\ldots,X_n;\nu=\infty}(x)} f_{X_{n+1}|X_1,\ldots,X_n;\nu=\infty}(x)dx \\
&= \Lambda_k^n \int f_{X_{n+1}|X_1,\ldots,X_n;\nu=k}(x)dx \\
&= \Lambda_k^n
\end{aligned}$$

and

$$\begin{aligned}
E_\infty\Lambda_k^n &= \int\cdots\int \frac{f_{\nu=k}(x_1,\ldots,x_n)}{f_{\nu=\infty}(x_1,\ldots,x_n)} f_{\nu=\infty}(x_1,\ldots,x_n)dx_1\cdots dx_n \\
&= \int\cdots\int f_{\nu=\infty}(x_1,\ldots,x_n)dx_1\cdots dx_n \\
&= 1.
\end{aligned}$$

Hence, $R_n - n$ is a P_∞-martingale with zero expectation; for

$$\begin{aligned}
E_\infty(R_{n+1} - (n+1) \mid X_1,\ldots,X_n) &= E_\infty\left(\sum_{k=1}^{n+1} \Lambda_k^{n+1} \mid X_1,\ldots,X_n\right) - (n+1) \\
&= \sum_{k=1}^{n} E_\infty(\Lambda_k^{n+1} \mid X_1,\ldots,X_n) \\
&\quad + E_\infty(\Lambda_{n+1}^{n+1} \mid X_1,\ldots,X_n) - (n+1) \\
&= \sum_{k=1}^{n} \Lambda_k^n + \Lambda_{n+1}^n - (n+1) \\
&= R_n + 1 - (n+1) \\
&= R_n - n
\end{aligned}$$

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

$$E_\infty N_A = E_\infty R_{N_A}. \tag{3}$$

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.

$$\begin{aligned} X_1, X_2, \ldots, X_{\nu-1} &\sim N(\mu_0, \sigma^2) \\ X_\nu, \ldots &\sim N(\mu_0 + \delta\sigma, \sigma^2) \end{aligned}$$

are independent, ν, μ_0, σ^2 are unknown, δ is known (viewed as a representative of the post-change parameter); $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$, $Z_1 = 0$,

$$\begin{aligned} Y_i &= (X_i - \bar{X}_{i-1})\sqrt{\frac{i-1}{i}}; \quad i = 2, 3, \ldots \\ Z_i &= \frac{Y_i}{Y_2}; \quad i = 2, 3, \ldots. \end{aligned}$$

Note that $Cov_{\nu=k}(Y_i, Y_j) = 0$ 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

$$\begin{aligned} E_{\nu=k}Y_i &= [\mu_0 + \delta\sigma - \frac{(i-k)(\mu_0+\delta\sigma) + (k-1)\mu_0}{i-1}]\sqrt{\frac{i-1}{i}} 1(i \geq k) \\ &= \frac{(k-1)\delta\sigma}{i-1}\sqrt{\frac{i-1}{i}} 1(i \geq k) \\ &= \delta\sigma\frac{k-1}{\sqrt{(i-1)i}} 1(i \geq k). \end{aligned}$$

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

$$\begin{aligned} & f_{\nu=k;Z_3,\ldots,Z_n}(z_3, \ldots, z_n) \\ = & \frac{\partial^{n-2}}{\prod_{i=3}^n \partial z_i} E_{\nu=k}[\prod_{i=3}^n \Phi(z_i Y_2 - \frac{\delta(k-1)}{\sqrt{(i-1)i}} 1(i \geq k)) 1(Y_2 > 0) \end{aligned}$$

$$+\prod_{i=3}^{n}[1-\Phi(z_iY_2-\frac{\delta(k-1)}{\sqrt{(i-1)i}}1(i\geq k)]1(Y_2<0)$$

$$= E_{\nu=k} \mid Y_2 \mid^{n-2} (\frac{1}{\sqrt{2\pi}})^{n-2} e^{-\frac{1}{2}[Y_2^2\sum_{i=3}^n z_i^2 - 2Y_2\delta(k-1)\sum_{i=k}^n \frac{z_i}{\sqrt{i(i-1)}} + \delta^2(k-1)^2\sum_{i=k}^n \frac{1}{i(i-1)}]}$$

$$= \int_{-\infty}^{\infty}(\frac{1}{\sqrt{2\pi}})^{n-1} \mid y \mid^{n-2} e^{-\frac{1}{2}[y^2\sum_{i=2}^n z_i^2 - 2y\delta(k-1)\sum_{i=k}^n \frac{z_i}{\sqrt{i(i-1)}} + \delta^2(k-1)^2(\frac{1}{k-1}-\frac{1}{n})]} dy$$

$$= (\frac{1}{\sqrt{2\pi}})^{n-1} e^{-\frac{\delta^2(k-1)^2}{2}[\frac{1}{k-1}-\frac{1}{n}-\frac{(\sum_{i=k}^n \frac{z_i}{\sqrt{i(i-1)}})^2}{\sum_{i=2}^n z_i^2}]} (\frac{1}{\sum_{i=2}^n z_i^2})^{\frac{n-1}{2}}$$

$$\times \int_{-\infty}^{\infty} \mid v - (-\frac{\delta(k-1)\sum_{i=k}^n \frac{z_i}{\sqrt{i(i-1)}}}{\sqrt{\sum_{i=2}^n z_i^2}}) \mid^{n-2} e^{-\frac{1}{2}v^2} dv$$

where $z_2 = 1$, and similarly

$$f_{\nu=\infty; Z_3,\ldots,Z_n}(z_3,\ldots,z_n) = (\frac{1}{\sqrt{2\pi}})^{n-1}(\frac{1}{\sum_{i=2}^n z_i^2})^{\frac{n-1}{2}} \int_{-\infty}^{\infty} \mid v \mid^{n-2} e^{-\frac{1}{2}v^2} dv.$$

Letting

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

obtain for $n \geq k > 2$ that

$$\Lambda_k^n = \frac{\int_{-\infty}^{\infty} \mid v - a_{k,n} \mid^{n-2} e^{-\frac{1}{2}v^2} dv}{\int_{-\infty}^{\infty} \mid v \mid^{n-2} e^{-\frac{1}{2}v^2} dv} e^{-\frac{1}{2}\delta^2(k-1)^2[\frac{1}{k-1}-\frac{1}{n}]+\frac{1}{2}a_{k,n}^2}. \tag{4}$$

Note that Λ_k^n doesn't change if we insert $-\delta$ instead of δ; i.e. $+a_{k,n}$ and $-a_{k,n}$ will give the same value of Λ_k^n.

The analogous calculation when $n > k = 2$ yields

$$\Lambda_k^n = \frac{\int_{-\infty}^{\infty} \mid v - a_{2,n} \mid^{n-2} e^{-\frac{1}{2}v^2} dv}{\int_{-\infty}^{\infty} \mid v \mid^{n-2} e^{-\frac{1}{2}v^2} dv} e^{-\frac{1}{2}\delta^2(\frac{3}{2}-\frac{1}{n})+\frac{1}{2}a_{2,n}^2}. \tag{5}$$

Clearly, $\Lambda_1^n \equiv 1$, 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

$$\begin{aligned}
g_m(a) &= \int_a^\infty (v-a)^m e^{-\frac{1}{2}v^2}\,dv \\
f_m(a) &= \int_{-\infty}^a (a-v)^m e^{-\frac{1}{2}v^2}\,dv \\
h_m &= \int_{-\infty}^\infty \mid v \mid^m e^{-\frac{1}{2}v^2}\,dv \\
u_m(a) &= \frac{g_m(a)}{h_m} \\
v_m(a) &= \frac{f_m(a)}{h_m}.
\end{aligned}$$

Thus

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

Now for $m \geq 2$

$$\begin{aligned}
h_m &= 2\int_0^\infty v^{m-1} v e^{-\frac{1}{2}v^2}\,dv \\
&= 2[v^{m-1}(-e^{-\frac{1}{2}v^2}) \mid_0^\infty + (m-1)\int_0^\infty v^{m-2}e^{-\frac{1}{2}v^2}\,dv \\
&= (m-1)h_{m-2}
\end{aligned}$$

where

$$\begin{aligned}
h_0 &= \sqrt{2\pi} \\
h_1 &= 2\int_0^\infty v e^{-\frac{1}{2}v^2}\,dv = 2.
\end{aligned}$$

Also, for $m \geq 2$

$$g_m(a) = \int_a^\infty (v-a)^{m-1}(v-a)e^{-\frac{1}{2}v^2}\,dv$$

$$
\begin{aligned}
&= \int_a^\infty (v-a)^{m-1} v e^{-\frac{1}{2}v^2} dv - a g_{m-1}(a) \\
&= (v-a)^{m-1}(-e^{-\frac{1}{2}v^2}) \,|_a^\infty + (m-1)\int_a^\infty (v-a)^{m-2} e^{-\frac{1}{2}v^2} dv - a g_{m-1}(a) \\
&= (m-1) g_{m-2}(a) - a g_{m-1}(a)
\end{aligned}
$$

where

$$
\begin{aligned}
g_0(a) &= \sqrt{2\pi}(1-\Phi(a)) \\
g_1(a) &= e^{-\frac{1}{2}a^2} - a\sqrt{2\pi}(1-\Phi(a)).
\end{aligned}
$$

Similarly, for $m \geq 2$

$$
f_m(a) = (m-1) f_{m-2}(a) + a f_{m-1}(a)
$$

where

$$
\begin{aligned}
f_0(a) &= \sqrt{2\pi}\Phi(a) \\
f_1(a) &= e^{-\frac{1}{2}a^2} + a\sqrt{2\pi}\Phi(a).
\end{aligned}
$$

Now

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

Denote

$$
\begin{aligned}
w_m(a) &= \frac{g_m(a)}{h_{m-1}} \\
y_m(a) &= \frac{f_m(a)}{h_{m-1}}.
\end{aligned}
$$

For $m \geq 3$

$$
w_m(a) = \frac{(m-1) g_{m-2}(a) - a g_{m-1}(a)}{(m-2) h_{m-3}} = \frac{m-1}{m-2} w_{m-2}(a) - a u_{m-1}(a) \quad (7)
$$

and

$$y_m(a) = \frac{(m-1)f_{m-2}(a) + af_{m-1}(a)}{(m-2)h_{m-3}} = \frac{m-1}{m-2}y_{m-2}(a) + av_{m-1}(a) \qquad (8)$$

where

$$\begin{aligned}
w_1(a) &= \frac{g_1(a)}{h_0} = \frac{e^{-\frac{1}{2}a^2} - a\sqrt{2\pi}(1-\Phi(a))}{\sqrt{2\pi}} \\
w_2(a) &= \frac{g_2(a)}{h_1} = \frac{g_0(a) - ag_1(a)}{h_1} = \frac{\sqrt{2\pi}(1-\Phi(a)) - a[e^{-\frac{1}{2}a^2} - a\sqrt{2\pi}(1-\Phi(a))]}{2} \\
y_1(a) &= \frac{f_1(a)}{h_0} = \frac{e^{-\frac{1}{2}a^2} + a\sqrt{2\pi}\Phi(a)}{\sqrt{2\pi}} \\
y_2(a) &= \frac{f_2(a)}{h_1} = \frac{f_0(a) + af_1(a)}{h_1} = \frac{\sqrt{2\pi}\Phi(a) + a[e^{-\frac{1}{2}a^2} + a\sqrt{2\pi}\Phi(a)]}{2}.
\end{aligned}$$

Finally, note that for $m \geq 2$

$$\begin{aligned}
u_m(a) &= \frac{g_m(a)}{h_m} = \frac{(m-1)g_{m-2} - ag_{m-1}}{(m-1)h_{m-2}} = u_{m-2}(a) - \frac{a}{m-1}w_{m-1}. \quad (9) \\
v_m(a) &= \frac{f_m(a)}{h_m} = \frac{(m-1)f_{m-2} + af_{m-1}}{(m-1)h_{m-2}} = v_{m-2}(a) + \frac{a}{m-1}y_{m-1}. (10)
\end{aligned}$$

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 Λ_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 Λ_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 $\Lambda_k^{en}, 1 \leq k \leq en$ (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 Shiryayev-Roberts parametric control chart
%based on the likelihood ratios of ratios of
%recursive residuals.


%Input Parameters: data (row vector of size en)
%                  d (=delta, the reresentation of
%                     the change in the mean)

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


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);
        W=W(:,n:-1:1);
        x=ones(1:n);
        x=cumsum(x)-1;
        a=(d/s)*x.*W;
        u=zeros(1:n);
        v=u;
        w=u;
        y=u;
        for k=2:n
                p=(exp((-a(k)^2)/2))/2;
                q=(erf(a(k)/sqrt(2)))/2+0.5;
                u(1)=p-a(k)*(1-q)*sqrt(pi/2);
                v(1)=p+a(k)*q*sqrt(pi/2);
                w(1)=u(1)*sqrt(2/pi);
                y(1)=v(1)*sqrt(2/pi);
                u(2)=1-q-a(k)*w(1);
                v(2)=q+a(k)*y(1);
                w(2)=(1-q)*sqrt(pi/2)-a(k)*u(1);
                y(2)=q*sqrt(pi/2)+a(k)*v(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/(k-1))-...
                        (1/n)-(W(k)/s)^2));
                t(2)=t(2)*exp(-0.25*(d^2));
        end
        R(n)=sum(t);
end
```

Figure 24: MATLAB-language computer program for computing R_n for the parametric SR scheme of section 2.

14 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 Λ_k^n 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 Λ_k^n's (hence the R_n's) behavior under P_∞: after all, if the true pre-change distribution is, say, G_0, transforming the observations by $F_0^{-1}(G_0())$ (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 obsevations 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 $f_0(x) = (1/2)\exp\{- \mid x \mid /2\}$ 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 Λ_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

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

(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 Λ_k^n), consider the first five mass difference obsevations (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. $X_5 < X_1 < X_2 < X_4 < X_3$. Therefore, $f_{\nu=k}(r_1, r_2, \ldots, r_5)$ is equal to $P_{\nu=k}(X_5 < X_1 < X_2 < X_4 < X_3)$. 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

$$P_{\nu=3}(X_5^* < X_1^* < X_2^* < X_4^* < X_3^*) = \sum_{i=0}^{5} P_{\nu=3}(X_5^* < X_1^* < X_2^* < X_4^* < X_3^*, B = i). \tag{11}$$

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 $X_3^* > X_4^* > X_2^* > 0$ and $0 > X_1^* > X_5^*$. Therefore

$$\begin{aligned}
& P_{\nu=3}(X_5^* < X_1^* < X_2^* < X_4^* < X_3^*, B = 2) \\
= \; & P_{\nu=3}(\{X_3^*, X_4^*, X_2^* \quad \text{are positive and} \quad X_3^* > X_4^* > X_2^*\} \\
& \cap\{X_1^*, X_5^* \quad \text{are negative and} \quad | X_5^* |>| X_1^* |\}) \\
= \; & P_{\nu=3}(X_3^*, X_4^*, X_2^* \quad \text{are positive and} \quad X_3^* > X_4^* > X_2^*) \times \\
& P_{\nu=3}(\{X_1^*, X_5^* \quad \text{are negative and} \quad | X_5^* |>| X_1^* |).
\end{aligned}$$

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, $| X_i^* | \sim \exp(1)$ if $i < 3$ and $| X_i^* | \sim \exp(\beta)$ 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

$$\begin{aligned}
& P_{\nu=3}(X_3^*, X_4^*, X_2^* \quad \text{are positive and} \quad X_3^* > X_4^* > X_2^*) \\
= \; & P_{\nu=3}(X_3^* > X_4^* > X_2^* \mid X_3^*, X_4^*, X_2^* \text{ are positive }) \cdot P_{\nu=3}(X_3^*, X_4^*, X_2^* \text{ are positive }) \\
= \; & P(\frac{Y_3}{\alpha} > \frac{Y_2}{\alpha} > \frac{Y_1}{1}) \cdot p \cdot p \cdot \frac{1}{2}. \\
= \; & \frac{\alpha}{\alpha} \cdot \frac{\alpha}{\alpha+\alpha} \cdot \frac{1}{1+\alpha+\alpha} \cdot p \cdot p \cdot \frac{1}{2} \\
= \; & \frac{1}{4} \frac{p^2}{2\alpha+1}.
\end{aligned}$$

In a similar fashion,

$$P_{\nu=3}(X_1^*, X_5^* \quad \text{are negative and} \quad | X_5^* | > | X_1^* |) = \frac{1}{2}(1-p)\frac{\beta}{\beta}\frac{1}{1+\beta} = \frac{1}{2}\frac{1-p}{1+\beta}.$$

The other elements on the right side of (11) can be computed analogously. The same considerations yield a calculation of Λ_k^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, α, β 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 ν not too close to the beginning of surveillance, and for large A) that

$$E_\nu(N_A - \nu \mid N_A \geq \nu) \approx \frac{\log A}{D} \tag{12}$$

```
%       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;
ln2palpha = log(2*p*alpha); ln2qbeta = log(2*q*beta);
lnratio = ln2palpha - ln2qbeta;
%       initialize ren
ren = zeros(1,en);
for n = 2:en,  %-----compute ren(n)-------
        lnbnmlt = zeros(1,n-1);   % computation of log of binomial(n,1/2) probs
        y = 1:1:n;
        lny = log(y);
        culny = cumsum(lny);
        for i=1:(n-1),
              lnbnmlt(i) = culny(n)-culny(i)-culny(n-i)-n*(log(2));
        end
        lnbnml = [-n*(log(2)) lnbnmlt -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 su kn's
        [dummy,invrankt]= sort (datan');     % index of smallest in invrankt(1)
        invrank          = invrankt';         % row of inverse ranks
%       compute vector of lambda sub nk's
        for k = 1:n,           % compute little lambda sub kn's
                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
                        %                                    log-denominators
                        %index of vectors is one plus number of putative
                        %                                    negatives m
                lnprodneg    = cumsum(lnvdenom);
                dummy1       = lnudenom(reverse);
                dummy2       = cumsum(dummy1);
                lnprodpos    = dummy2(reverse);
                usubkmod     = [(n+1-k) usubk];
                lambdakn(k)  = sum(exp(lnbnml + (n+1-k)*ln2qbeta...
                                 + usubkmod*lnratio - lnprodneg - lnprodpos));
        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

$$
\begin{aligned}
D &= (1-G_1(0))\log(2p\alpha) + G_1(0)\log(2(1-p)\beta) \\
&+ (1-\alpha)\int_0^\infty F_0^{-1}(G_0(x))dG_1(x) \\
&+ (\beta-1)\int_{-\infty}^0 F_0^{-1}(G_0(x))dG_1(x).
\end{aligned}
$$

Therefore, if G_0 and G_1 are surmised to be (approximately) the pre-and-post change distibutions, p, α, β 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

$$
\lim_{A\to\infty} \frac{E_\infty N_A}{A} = E_{\nu=1} S_1 e^{\sum_{n=1}^\infty \frac{1}{n} E_{\nu=1} e^{-S_n^+}}. \tag{13}
$$

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 δ for the parametric procedure of Section 3 can be found in Pollak

(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 $\sigma_\epsilon^2\chi^2_{(3)}$ 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 $.8735/\sqrt{44}$ 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 $.8211/\sqrt{100}$.) So, at least approximate normality of these observations is reasonably well established.

As for independence, two-sided runs tests produce $p-$values .1056 and .1272 for the sets of observations $51 - 106$ and $107 - 150$ when taken separately, and .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 $.7389/\sqrt{51}$. 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 .00027. 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.)

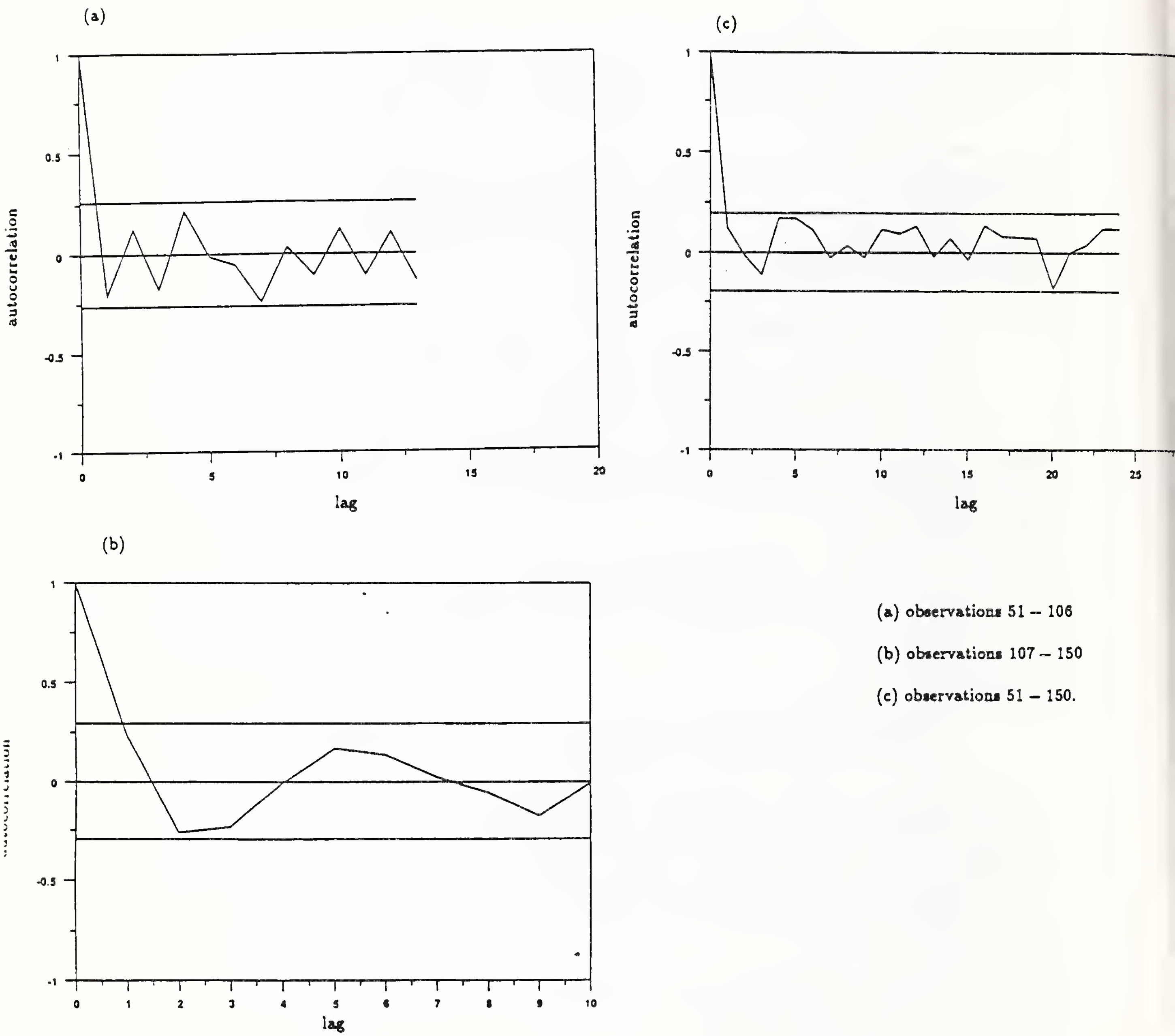


Figure 26: Autocorrelation plots for mass difference data:

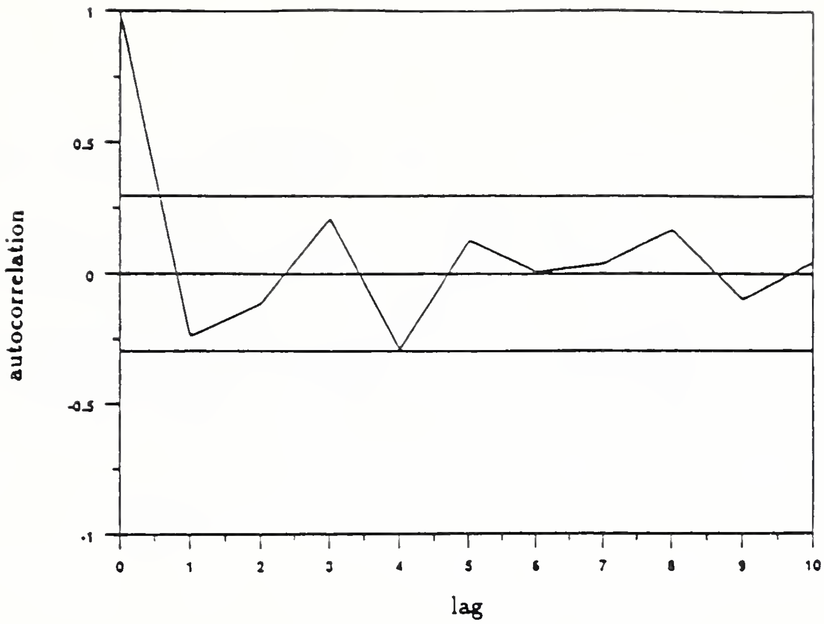


Figure 27: Autocorrelaton plot for mass difference data: observations 166 – 217, with observation 207 deleted.

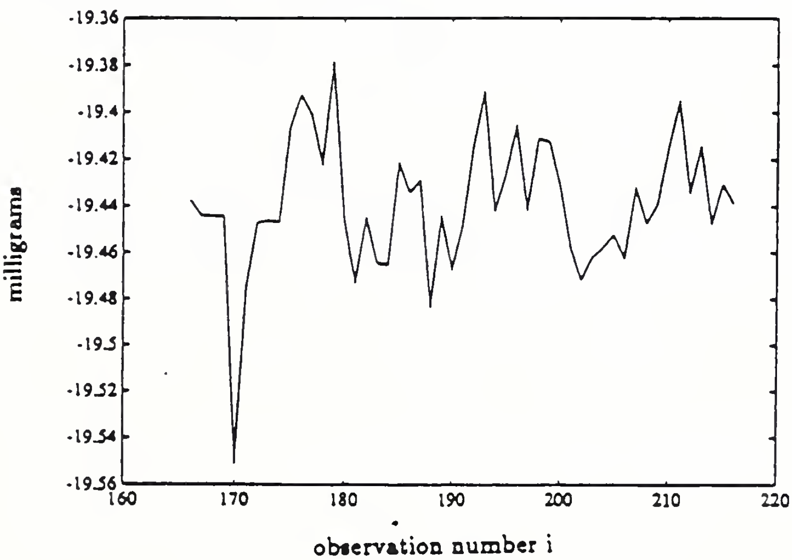


Figure 28: Detail of the mass difference data: observations 166 – 217, with observation 217 deleted.

(b) Diagnostics for the residual standard deviation data

Based on the essential normality of the observations, the distribution of $3s_i^2/\sigma_\epsilon^2$ should be $\chi^2_{(3)} = gamma(1.5, .5)$. The Kolmogorov-Smirnov two-sided statistic computed with an estimate for σ_ϵ 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 $3s_i^2/\sigma_\epsilon^2$ are $\chi^2_{(3)}$, $\Phi^{-1}(F_{\chi^2_{(3)}}(3s_i^2/\sigma_\epsilon^2))$ 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.

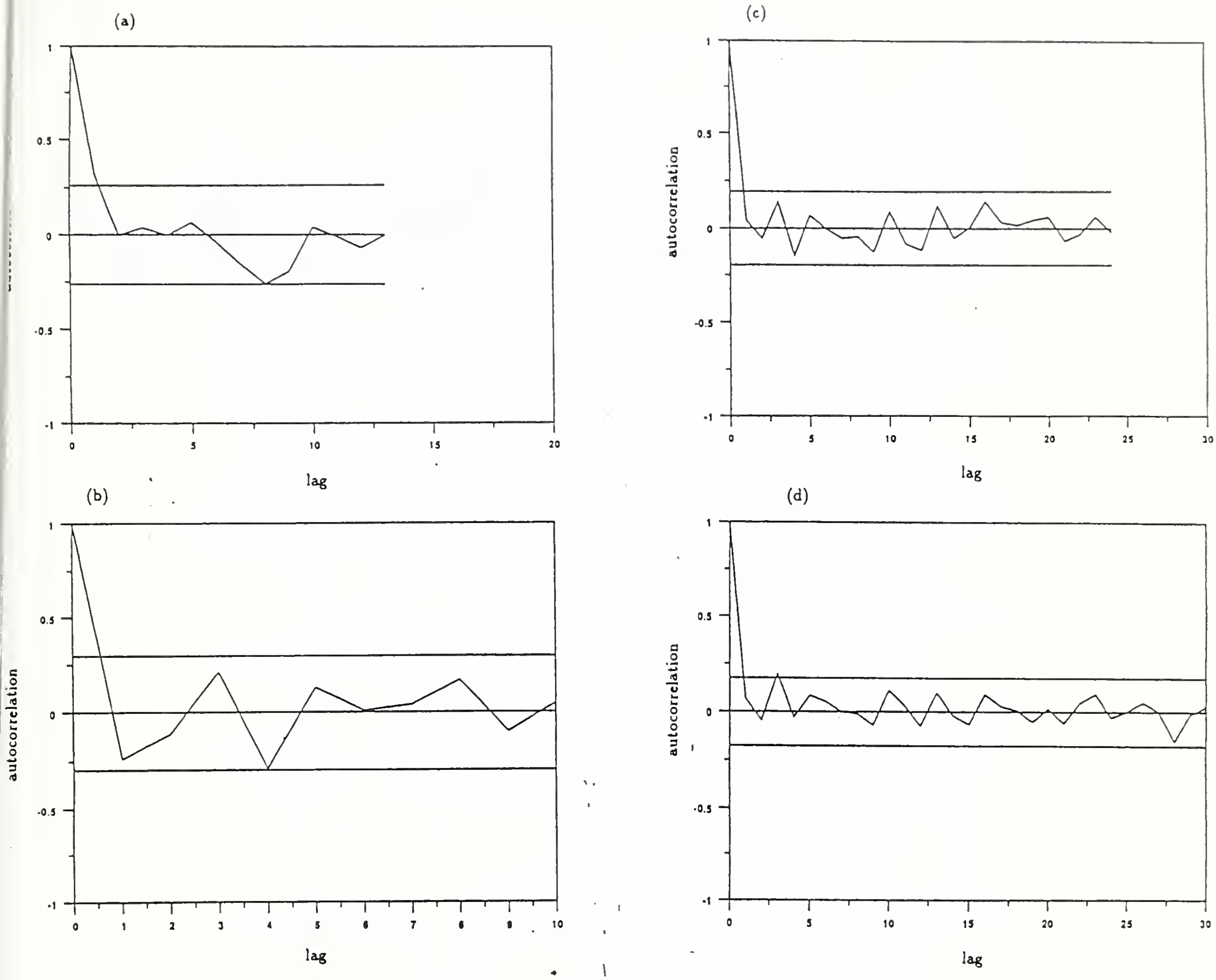


Figure 29: Autocorrelation plots for residual standard deviations transformed to $N(0,1)$ variables: (a) observations $51-106$, (b) observations $107-150$, (c) observations $51-150$ (d) observations $41-165$.

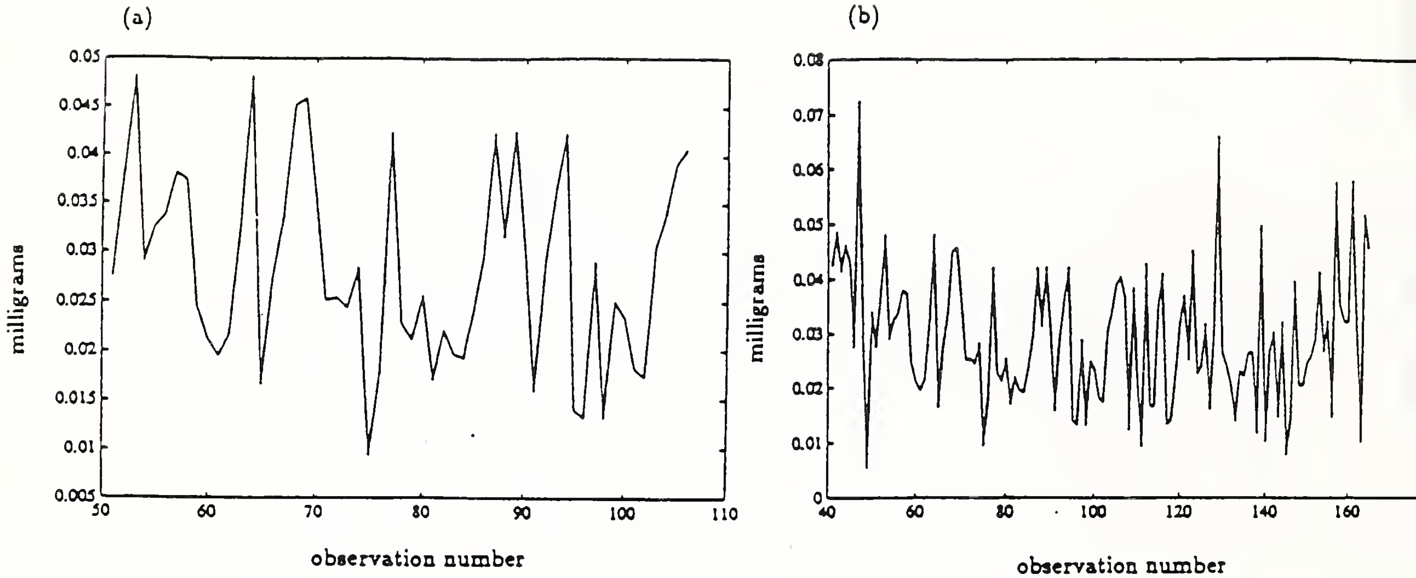


Figure 30: Detail of the mass difference data: (a) observations 51 – 106 (b) observations 41 – 165.

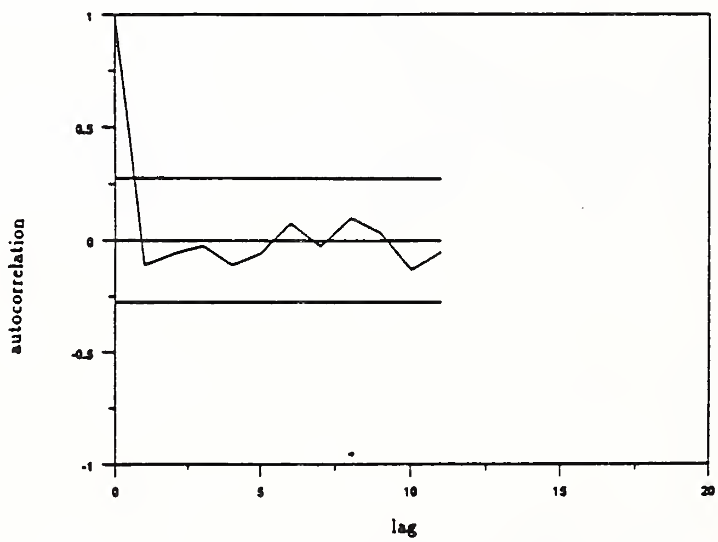


Figure 31: Autocorrelation plot for residual standard deviatons transformed to $N(0,1)$ variables: observations 166 – 217 with observation 207 deleted.

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, $sign(Y_2)$ should be added to the data on which surveillance is based, and Λ_k^n 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 bave been used were f_i 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

[1] Almer, H.E., Keller, J., ed. (1977). Surveillance Test Procedures. Nat. Bur. Stand. (U.S.) *NBSIR* 76-999, 1977 May, 77p, *NTIS* PB268130.

[2] Bell, C., Gordon, L. and Pollak, M. (1992). An efficient nonparametric detection scheme with applications to surveillance of a Bernoulli process with unknown baseline. *Technical Report.*

[3] Brown, R.L., Durbin, J. and Evans, J.M. (1975). Techniques for testing the constancy of regression relationships over time. *J. Royal Statist. Soc. Ser. B 37*, 149-192.

[4] Chow, Y.S., Robbins, H. and Siegmund, D. (1967). *Great Expectations: the Theory of Optimal Stopping.* Houghton-Mifflin.

[5] van Dobben de Bruyn, C.S. (1968). *Cumulative Sum Tests.* Griffin, London.

[6] Filliben, J.J. (1981). Dataplot - an interactive high-level language for graphics, nonlinear fitting, data analysis and mathematics. *Computer Graphics 15*, 199-213.

[7] Gordon, L. and Pollak, M. (1989). An efficient sequental nonparametric scheme for detecting a change of distribution. *Technical Report.* To appear in Ann. Statist.

[8] Gordon, L. and Pollak, M. (1990). Average run length to false alarm for surveillance schemes designed with partially specified pre-change distribution. *Technical Report.* To appear in Ann. Statist.

[9] Gordon L. and Pollak, M. (1991). A robust surveillance scheme for stochastically ordered alternatives. *Technical report.* To appear in Ann. Statist.

[10] Gordon, L. and Pollak, M. (1992). A robust surveillance scheme for detecting a change of dispersion. *Manuscript.*

[11] Jaeger, K.B. and Davis, R.S. (1984). A Primer for Mass Metrology. *NBS Special Publication 700-1, Industrial Management Series*, U.S. Printing Office, Washington, D.C. 20402.

[12] James, B., James, K.L. and Siegmund, D. (1987). Tests for a change-point. *Biometrika 74* 71-83.

[13] Kenett, R.S. and Pollak, M. (1992). Data analytic aspects of the Shiryayev-Roberts control chart: surveillance of a non-homogeneous Poisson process. *Technical Report.* To appear in Techniques.

[14] Kenett, R.S. and Zacks S. (1992). Tracking algorithms for processes with change points. *Technical Report.*

[15] Lorden, G. (1971). Procedures for reacting to a change in distribution. *Ann. Mathem. Statist. 42* 1897-1908.

[16] The MathWorks (1989). Pro-Matlab User's Guide. South Natick, Mass.

[17] Moustakides, G.V. (1986). Optimal stopping times for detecting changes in distributions. *Ann. Statist. 14* 1379-1387.

[18] Page, E.S. (1954). Continuous inspection schemes. *Biometrika 41* 100-115.

[19] Pollak, M. (1985). Optimal detection of a change in distribution. *Ann. Statist. 13* 206-227.

[20] Pollak, M. (1987). Average run lengths of an optimal method of detecting a change in distribution. *Ann. Statist. 15* 749-779.

[21] Pollak, M. and Siegmund, D. (1985). A diffusion process and its application to detecting a change in the drift of Brownian motion. *Biometrika 72* 267-280.

[22] Pollak, M. and Siegmund, D. (1991). Sequential detection of a change in a normal mean when the initial value is unknown. *Ann. Statist. 19* 394-416.

[23] Ritov, Y. (1990). Decision theoretic optimality of the cusum procedure. *Ann. Statist. 18* 1464-1469.

[24] Roberts, S.W. (1966). A comparison of some control chart procedures. *Technometrics 8* 411-430.

[25] Shewhart, W.A. (1931). *The Economic Control of the Quality of Manufactured Product.* Macmillan, New York.

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

[27] Shiryayev, A. N. (1978). *Optimal Stopping Rules.* Springer, New York.

[28] Siegmund, D. (1985). *Sequential Analysis: Tests and Confidence Intervals.* Springer, New York.

[29] Siegmund, D. and Venkatraman, E.S. (1992). Using the generalized likelihood ratio statistics for sequential detection of a change-point. *Stanford University, Dept. of Statistics, Technical Report No. 22.*