Sequential rank cumulative sum charts for location and scale

C van Zyl

orcid.org 0000-0001-5979-7294

Thesis submitted in fulfilment of the requirements for the degree *Doctor of Philosophy in Risk Analysis* at the North-West University

Supervisor: Prof F Lombard

Graduation May 2018

22231609

"All men have stars, but they are not the same things for different people. For some, who are travellers, the stars are guides. For others they are no more than little lights in the sky. For others, who are scholars, they are problems..."
Antoine de Saint-Exupéry, The Little Prince.

Acknowledgements

I would like to acknowledge the following people and institutions who have been invaluable in the endeavour of completing this thesis. First and foremost, my supervisor, Professor Freek, for your dedication, patience, hard work and ample time that you so generously provided in guiding me. Thank you for sharing your wisdom and vast knowledge (of statistics and life in general) as enthusiastically as you do. It is true what Isaac Newton said: "If I have seen further, it is by standing upon the shoulders of giants." Prof, you are a great teacher – an inspiration. I am truly privileged to have been taught by such a renowned scholar. Equally, I would like to thank your wife, Chrystal, for her hospitality and kindness.

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged.[1] I thank SASA for liaising with the NRF in this regard.

To many non-statistician friends for their interest in my studies. Also to statistician colleagues at the Centre for BMI. I also thank Johan Blaauw for translating the abstract to Afrikaans.

My parents-in-law, Lucia and Koos, my siblings-in-law, niece and nephews for their encouragement and support.

My parents, Berna and Gerhard, and my sister, Esmari, who continually enquired after my studies and progress. Thank you for showing a keen interest and for all the investments (in time, love and otherwise) that you made towards my education and in me as person. My grandmother whose encouragement remains a fond memory. Thank you Henda for your support, for believing in me and for sharing my passion in forever gaining new knowledge.

And last, but not least, to Willie (who willingly married me mid-PhD): I am tremendously grateful for your love, support and companionship and that you joined me in this adventure. You listened to my statistics-rambling, you have put up with my frustrations and you have shared in my excitements. You now probably know more about statistics than you would have wanted to. Thank you for brewing me coffee when I needed it most, for having patience with many late nights and for being a PhD-computer-widower. I am glad that you (we) survived this thesis. Thank you also for proofreading my thesis a few (too many, I guess) times. I could not accomplish this feat without you.

God, for granting me opportunities, talents and the privilege to study.

[1] Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

Contents

List of Figures

List of Tables

Abstract

In this thesis we construct CUSUMs based on the signed and unsigned sequential ranks of independent observations for the purpose of detecting either a persistent location or a persistent scale shift. In designing these CUSUMs we consider two scenarios, namely detecting a shift when the in-control distribution is symmetric around a known median and when either the symmetry assumption fails or the in-control median is unknown. We then extend our CUSUM designs to the class of Girschick-Rubin CUSUMs. All of our CUSUMs are distribution free and fully self starting: no parametric specification of the underlying distribution is necessary in order to find correct control limits that guarantee a specified nominal in-control average run length given a reference value. In particular, our sequential rank CUSUMs have zero between-practitioner variation. Furthermore, these CUSUMs are robust against the effect of spurious outliers. The out-of-control average run length properties of the CUSUMs are gauged qualitatively by theory-based calculations and quantitatively by Monte Carlo simulation.

We show that in the case where the underlying distribution is normal with an unknown variance, our sequential rank CUSUMs based on a Van der Waerden-type score can be used to good effect, because the out-of-control average run lengths correspond very well to those of the standard normal distribution CUSUM where the variance is assumed known. For heavier tailed distributions we show that use of the Wilcoxon sequential rank score is indicated. Where transient special causes are apt to occur frequently, use of a Cauchy score is indicated. We illustrate the implementation of our CUSUMs by applying them to data from industrial environments.

Keywords: CUSUM, distribution free, self starting, signed sequential ranks, unsigned sequential ranks.

Uittreksel

In hierdie proefskrif konstrueer ons kumulatiewesomprosedures (KUSOM-prosedures) wat gebaseer is op die betekende en onbetekende sekwensiële range van onafhanklike waarnemings. Die doel van hierdie KUSOM's is om 'n volhardende verskuiwing in lokaliteit of spreiding te identifiseer. Ons beskou twee gevalle, naamlik om 'n verskuiwing te identifiseer wanneer die binne-beheerverdeling simmetries is rondom 'n bekende mediaan, en wanneer óf die simmetrieaanname ongeldig is óf die waarde van die binnebeheermediaan onbekend is. Ons brei die konstruksie van hierdie KUSOM's uit na die klas van Girshick-Rubin-KUSOM's. Al hierdie KUSOM's is verdelingsvry en ten volle self-inisiërend: geen parametriese spesifisering van die onderliggende verdeling is noodsaaklik om die korrekte kontrolelimiete te vind wat 'n gespesifiseerde binnebeheer- gemiddelde looplengte waarborg vir 'n gegewe verwysingswaarde nie. In die besonder het ons KUSOM's geen tussen-praktisynvariasie nie. Verder is hierdie KUSOM's robuus teen die effek wat sporadiese uitskieters op die data mag hê. Die eienskappe van die buitebeheer- gemiddelde looplengte word kwalitatief deur teoriegebaseerde berekeninge gemeet, en kwantitatief deur Monte Carlo-simulasie.

Ons toon aan dat ons sekwensiëlerang-KUSOM's wat gebaseer word op 'n Van der Waerden-tipe-telling met groot sukses gebruik kan wanneer die onderliggende verdeling normaal is met 'n onbekende variansie omdat die buitebeheer- gemiddelde looplengtes goed ooreenstem met dié van die standaardnormaalverdeling-KUSOM waar die variansie as bekend aanvaar word. Vir verdelings met swaar sterte toon ons aan dat die Wilcoxon-sekwensiëlerangtelling gebruik kan word. Waar transiënte spesiale oorsake geneig is om gereeld voor te kom, beveel ons die gebruik van 'n Cauchy-telling aan. Ons illustreer die toepassing van ons KUSOM's deur dit toe te pas op data uit industriële omgewings.

Sleutelwoorde: betekende sekwensiële range, KUSOM, onbetekende sekwensiële range, self-inisiërend, verdelingsvry.

Frequently used notation

Abbreviations

1. SPC abbreviates statistical process control.
2. CUSUM abbreviates cumulative sum.
3. GR CUSUM abbreviates the CUSUM of Girschick and Rubin (1952).
4. NSS CUSUM abbreviates the normal self-starting CUSUM.
5. GSS CUSUM abbreviates the gamma self-starting CUSUM.
6. HD CUSUM abbreviates the CUSUM of Hawkins and Deng (2010).
7. RA CUSUM abbreviates the CUSUM of Ross and Adams (2012).
8. SSR abbreviates signed sequential rank.
9. SRL abbreviates sequential rank location.
10. KSR abbreviates the Klotz sequential rank score used in the scale CUSUM or GR CUSUM.
11. MSR abbreviates the Mood sequential rank score used in the scale CUSUM or GR CUSUM.
12. ARL abbreviates average run length.
13. ARL_0 is the symbol used to indicate the nominal value of the in-control ARL.
14. IC abbreviates in control.
15. OOC abbreviates out of control.
16. LMP abbreviates locally most powerful.
17. IQR abbreviates the inter-quartile range.
18. i.i.d. means independent and identically distributed.

Mathematical symbols

1. (F, f) and (G, g) denote the in-control and out-of-control pair of distribution and density functions, respectively.
2. $f'(x)$ denotes the derivative of f with respect to x, unless stated otherwise.
3. For a number x, $\text{sign}(x) = 1$ if $x > 0$, $= -1$ if $x < 0$ and 0 if $x = 0$.
4. The indicator function is $\mathbb{1}(A) = \begin{cases} 1 & \text{if } A \text{ is true} \\ 0 & \text{if } A \text{ is false.} \end{cases}$
5. r_i is the rank of X_i among $X_1, \dots, X_i$ – known as the sequential rank of X_i.
6. r_i^+ is the rank of $|X_i|$ among $|X_1|, \dots, |X_i|$ – known as the sequential rank of $|X_i|$.
7. $R_{n,i}$ is, for $i \le n$, the rank of X_i among $X_1, \dots, X_n$.
8. $R_{n,i}^+$ is, for $i \le n$, the rank of $|X_i|$ among $|X_1|, \dots, |X_n|$.
9. $X_{n:i}$ is, for $i \le n$, the i^{th} order statistic of the data $X_1, \dots, X_n$.
10. H_0 and H_1 denote a null and alternative hypothesis, respectively.
11. $I_0(f)$ is Fisher's information in a density f belonging to a location parameter family.
12. $I_1(f)$ is Fisher's information in a density f belonging to a scale parameter family.
13. $a := b$ means a is defined by the expression b on the right-hand side.
14. $a \approx b$ means a is approximately equal to b.
15. $N(\mu, \sigma^2)$ denotes a normal distribution (or random variable) with mean μ and variance σ^2.
16. Φ, ϕ and Φ^{-1} denote the standard normal distribution, density and inverse distribution functions, respectively.
17. t_ν denotes a Student t distribution (or random variable) with ν degrees of freedom.
18. $U(a, b)$ denotes the uniform distribution (or random variable) on (a, b).
19. $SN(\alpha)$ denotes a skew-normal distribution (or random variable) with skewness parameter α.
20. $\log x$ denotes the natural logarithm of x.
21. $a_n = o(b_n)$ means that $\lim_{n \to \infty} a_n / b_n = 0$.
22. $a_n = O(b_n)$ means that $\limsup_{n \to \infty} |a_n / b_n| < \infty$.
23. $A_n \stackrel{\mathcal{D}}{\Longrightarrow} B$ means the sequence of random objects A_n converges in distribution to the random object B as $n \to \infty$.
24. $A \stackrel{\mathcal{D}}{=} B$ means the random objects A and B have the same distribution.
25. $A \stackrel{\mathcal{D}}{\approx} B$ means the random objects A and B have approximately the same distribution.
26. $\lfloor a \rfloor$ denotes the largest integer smaller than or equal to a.
27. $\lceil a \rceil$ denotes the smallest integer larger than or equal to a.

1 A review of some relevant CUSUM literature

1.1 Page-type CUSUMs

Cumulative sum procedures are a class of statistical process control (SPC) instruments devised for the purpose of monitoring a process to detect structural shifts in its characteristics. The aim is to identify and signal the onset of a small persistent shift as soon as possible. These control procedures find application in diverse scientific fields, including engineering (Timmer et al., 2001), public health and medicine (see for instance Woodall (2006), Ledolter and Kardon (2012) or Shanmugam et al. (2012)), seismology (Basseville and Nikiforov, 1993) and business and finance (see Kahya and Theodossiou (1999), Yi et al. (2006), Lam and Yam (1997), Golosnoy and Schmid (2007), Mukherjee (2009) or Coleman et al. (2001)). For a review, see Stoumbos et al. (2000). The theory and various applications of CUSUMs are described in the book by Hawkins and Olwell (1998). By far the majority of the CUSUM literature is concerned with procedures requiring specific assumptions about the functional form of the underlying distribution functions. These procedures are typically rather sensitive to deviations from distributional assumptions. In this thesis we concern ourselves with the design and analysis of signed and unsigned sequential rank CUSUMs which are free of overly specific distributional assumptions.

Page (1954) developed the first parametric CUSUM to monitor the mean of a normal distribution with zero mean and known variance. We will refer to this as the standard normal CUSUM. It is well known that the assumption of normality is often violated in practical applications. Then, the control limits of the standard normal CUSUM do not apply and the in-control average run length cannot be guaranteed a priori. Hawkins and Olwell (1998, Section 3.5) illustrate how the misspecification of the underlying distribution affects the in-control average run length of the CUSUM. Furthermore, Hawkins and Olwell (1998, Chapter 7) and Keefe et al. (2015) illustrate how misestimation of the presumed known values of nuisance parameters, such as the variance, can have disastrous effects on the in-control average run length of the CUSUM. This is problematic because we cannot know the out-of-control properties of a CUSUM if the (true) in-control average run length does not equal the nominal value (Quesenberry, 1995). Furthermore, Hawkins and Olwell (1998, Section 3.7) and Harrison and Lai (1999) show that the Page CUSUM can be very sensitive to deviations from normality, especially when the underlying distribution is heavy tailed. Subsequent to Page (1954), there followed a large volume of published work extending the sphere of application to distributions in the exponential family.

However, developing CUSUMs for location-scale families of distributions which, except for the normal and gamma distributions, are not in the exponential family, has received scant attention. In particular, little or no attention has been paid to heavy-tailed distributions in the location-scale family. In Chapter 6, we provide a practical application which occurs frequently in the process industries and in which an assumption of normality is tenuous at best. In light of this application, it is all the more surprising that more attention has not been given in the literature to CUSUMs based on distributions with tails heavier than those of the normal.

A natural approach towards constructing CUSUMs that are free of overly specific distributional assumptions is to replace the observed data by rank-based equivalents which are distribution free. Bakir (2001) presents a survey of various types of distribution-free control charts. Among these are control charts based on signs, ranks and signed ranks, respectively. By "distribution free" is meant that the in-control properties and control limits of the CUSUM do not depend on the functional form of the underlying distribution function or on any parameters. In this thesis we distinguish between two scenarios:

(I) detecting a location or a scale shift in a distribution in which the in-control median is specified; and

(II) detecting a location or a scale shift in a distribution in which the in-control median is unspecified because it is unknown.

Suppose that data accrue from a distribution which is symmetric around a known median, which can be taken to be zero without loss of generality (scenario I). Assuming that rational groups of $k > 1$ observations are available at each time point, Bakir and Reynolds (1979) and Bakir (2006) developed a CUSUM to detect a shift from a zero to a non-zero median based on the Wilcoxon signed rank statistic calculated within each rational group. They use the signed rank statistic $\text{sign}(X_i)R^+_{n,i}$ (see the list of mathematical symbols). For singly accruing data, Lombard and Van Zyl (2018) and Van Zyl (2015) developed CUSUMs for a location or a scale shift. Their signed sequential rank (SSR) CUSUM is based on the signs $s_i = \text{sign}(X_i)$ and on the sequential ranks r^+_i of the observations – see 6 in the list of mathematical symbols. The presence of the signs also enables detection of the onset of asymmetry. A basic property of the sequential ranks r^+_i is that they are statistically mutually independent and also statistically independent of the $\text{sign}(X_i)$ when the underlying distribution is in control, that is, when the underlying distribution is continuous and symmetric around zero (see Reynolds (1975) or Khmaladze (2011)). Because the in-control median is specified, these CUSUMs have the ability to detect when a distribution is already out of control at the onset of monitoring. Because the statistics $\text{sign}(X_i)r^+_i$ are sequential in nature, they are very well suited to the construction of CUSUMs for time ordered data. Thus, in essence these are Page CUSUMs with the i.i.d. values $X_1, X_2, \ldots$ with in-control mean zero replaced by the independent, non-identically distributed sequence $s_1r^+_1, s_2r^+_2, \ldots$ which also have in-control mean zero.

When the symmetry assumption is not justified or when the in-control median is unknown (scenario II), signed statistics cannot be used to good effect and resort must be had to unsigned statistics. When transforming X to $\mu + X$ or to σX, $\sigma > 0$, the unsigned sequential ranks r_i (see 5 in the list of mathematical symbols) remain unchanged. A direct consequence is that unsigned sequential rank CUSUMs cannot usefully incorporate numerical information about a specified in-control location or scale parameter. Bhattacharya and Frierson (1981) and Lombard (1981) developed truncated sequential tests for a location shift based on the unsigned sequential ranks r_i. However, these are not CUSUM procedures in the commonly accepted sense of the term. The first construction of a fully-fledged sequential rank CUSUM for a location shift is due to McDonald (1990). A more recent development is the Hawkins and Deng (2010) CUSUM, which is based upon the ordinary ranks $R_{n,i}$, $i \le n$ for $n \ge 15$. The basis of their CUSUM is a sequential application of the changepoint tests of Pettitt (1979). Even more recently, Liu et al. (2014) introduced a rank-based adaptive CUSUM for location shifts. Their chart consists of using a one-step ahead estimate of the shift size and then incorporating this information into the reference value. The focus of the charts mentioned thus far falls predominantly on location shifts, but neglects other types of structural shifts. In this regard, Ross and Adams (2012) designed a CUSUM to detect arbitrary shifts in a distribution by adapting the changepoint model methodology of Hawkins and Deng (2010) and applying it to the Cramér-von-Mises and Kolmogorov-Smirnov statistics. Other nonparametric approaches were made by Chatterjee and Qiu (2009), Gandy and Kvaløy (2013) and Saleh et al. (2016), who proposed that control limits be estimated by bootstrapping from an in-control Phase I sample. Thus, the control limits used in Phase II depend on the Phase I sample. This implies that control limits must be generated anew whenever a new data set appears. Thus, a table of control limits that can be used universally is out of the question, which is somewhat unsatisfactory from a practical perspective. Furthermore, Chatterjee and Qiu's (2009) procedure requires a large Phase I sample, while the procedures of Gandy and Kvaløy (2013) and Saleh et al. (2016) only guarantee upper and lower limits to the in-control average run length with a specified probability close to one.

1.2 Girschick-Rubin CUSUMs

Another instrument in the SPC toolbox is a CUSUM initially developed by Girschick and Rubin (1952) using a Bayesian argument on the assumption that the changepoint has a geometric prior distribution. A continuous time version of this CUSUM was developed by Shiryaev (1963) in the context of detecting a shift in the drift of a Brownian motion. Roberts (1966) compares a number of SPC procedures, among others the Page (1954) CUSUM and the Girschick and Rubin (1952) CUSUM. In the literature, the originators of this CUSUM, namely Girschick and Rubin (1952), seem to have been forgotten and their CUSUM is now generally referred to as the "Shiryaev-Roberts" CUSUM. In this thesis, however, we will refer to the "Girschick-Rubin" CUSUM, abbreviated "GR CUSUM". Using this nomenclature helps to avoid confusion in the use of acronyms "SR" (sequential rank) CUSUM and "SR" (Shiryaev-Roberts) CUSUM.

The vast majority of literature on GR CUSUMs concentrates on detecting a mean shift in a fully-specified distribution, in particular the normal distribution. Pollak and Siegmund (1991) derive the GR CUSUM for the situation where the in-control mean is unknown, while they assume that the underlying variance is known. This situation (that σ^2 is known and μ is unknown) rarely occurs in practice, hence we will not consider this case in this thesis. For monitoring the normal variance, Lazariv et al. (2013) worked out the GR CUSUM in detail, while Zhang et al. (2011b) proposed a single chart to detect either a mean or a variance shift in the normal distribution. Zhang et al. (2011a) proposed a GR CUSUM for a normal variance shift using grouped observations. Other than for the normal distribution, GR CUSUMs for location or scale shifts have not been treated in the literature. This provides a motivation for designing distribution-free procedures. In fact, Pollak (2009, p.4) pointed out the need for distribution-free GR CUSUMs and mentioned that sequential ranks could provide a solution. Nevertheless, the literature on GR CUSUMs without distributional assumptions is extremely limited. In Chapter 5 we attempt to give the distribution-free GR CUSUM its rightful place as a "leading tool" for changepoint detection as Pollak (2009) had hoped.

1.3 Contributions of the thesis

In this thesis, we propose distribution-free CUSUMs based on signed and unsigned sequential ranks to detect either a location or a scale shift from data arising singly over time. The in-control properties of our CUSUMs are distribution free. The control limits that we provide can be used universally and depend only on the particular rank score function that is used – they are valid no matter what the form of the underlying distribution. The only distributional assumptions we make, for technical convenience alone, is that the underlying distribution is continuous with a strictly increasing distribution function. Furthermore, our CUSUMs are self starting in that no Phase I parameter estimates are required in order to initiate the CUSUMs. However, in practice the availability of a relatively small Phase I sample can aid in the effective design of the CUSUM, for instance, it can be used to determine an appropriate reference value. Since the Phase I data are independent of the sequential ranks of the Phase II data, the Phase II in-control average run length remains guaranteed. This is an important feature that is not present in parametric CUSUMs when parameters have to be estimated. We gauge the out-of-control average run length properties of our CUSUMs qualitatively via theoretical calculations and quantitatively by Monte Carlo simulation studies. Our CUSUMs are, to the best of our knowledge, new.

1.4 Structure of the thesis

In Chapter 2, we discuss some existing parametric and distribution-free CUSUMs in more detail. In Chapter 3, we develop sequential rank CUSUMs for a location shift. We briefly discuss the SSR CUSUM of Lombard and Van Zyl (2018) and Van Zyl (2015), which are applicable in scenario I. We then construct sequential rank CUSUMs for location shifts in scenario II. Chapter 4 develops sequential rank CUSUMs for detecting scale shifts. In these chapters we also make a clear distinction between scenarios I and II given on page 2. In Chapter 5, our focus is on Girschick-Rubin CUSUMs for detecting a location or a scale shift in both scenarios I and II. In Chapter 6, we show the application of our CUSUMs to three sets of data. In Chapter 7, we summarise our main results and provide some pointers to issues requiring further research. The Appendices provide technical details of some calculations referred to in the preceding chapters. We also provide a suite of MATLAB programs that we use throughout the thesis in Monte Carlo simulations and in the applications. The data sets that we use are also also provided in an Excel file.

2 Technical details of some existing CUSUMs

We briefly discuss some of the existing parametric CUSUMs. Hawkins and Olwell (1998, Chapter 3) provide full details on the standard normal CUSUM and illustrate that the normal CUSUM is non-robust against deviations from underlying normality. We also discuss existing distribution-free CUSUMs, which we will use in the thesis for comparative purposes.

2.1 The Page CUSUM

The original formulation of the CUSUM was by Page (1954), who used the partial sums

$$S_i = \sum_{j=1}^{i} (\xi_j - \zeta) \tag{2.1}$$

where ξ_j are independent $N(\mu, 1)$ random variables and where $S_0 = 0$ and $\zeta = 0$ to construct the CUSUM

$$D_i = \begin{cases} 0, & i = 0 \\ S_i - \min\limits_{0 \le k \le i} S_k, & i \ge 1, \end{cases} \tag{2.2}$$

for detecting a shift in the mean of the ξ_j away from zero. The CUSUM signals that a shift away from zero has possibly occurred when D_i first exceeds the control limit h. The run length is

$$N = \min\{i \ge 1 : D_i \ge h\} \tag{2.3}$$

and the in-control average run length (ARL_0) is $\mathrm{E}[N]$ calculated under the assumption that $\mu = 0$. The control limit h is determined in order to make $\mathrm{E}[N|\mu = 0]$ equal to a finite nominal value ARL_0. For the normal distribution one can either obtain h from Monte Carlo simulation or from the Markov Chain approach (see Hawkins (1992) as an extension of Brook and Evans (1972)), or by using freely available software such as the R packages "CUSUMdesign" (Hawkins et al., 2016) or "spc" (Knoth, 2016).

Consider the random variables $\xi_1, \ldots, \xi_\tau, \ldots, \xi_i$. The changepoint τ is defined as the last index before the mean shifts, thus $1 \le \tau \le i - 1$. That is $\xi_1, \ldots, \xi_\tau$ have mean zero while $\xi_{\tau+1}, \xi_{\tau+2}, \ldots$ have non-zero mean.

The performance of the CUSUM is usually judged by the out-of-control average run length (OOC ARL), which is

$$E[N - \tau | N > \tau, \mu \neq 0]. \tag{2.4}$$

The OOC ARL is the expected number of observations after the shift conditional upon there being no signal before the shift. Explicit expressions for the OOC ARL (2.4) when $\tau > 0$ are not available in the literature. However, the computing power available on modern personal computers makes the estimation of the OOC ARL by Monte Carlo simulation, a straightforward matter. In the formulation (2.1), the reference value ζ acts as a tuning parameter for the target shift size. The OOC ARL of the CUSUM can be tuned to a specific target size so that for shifts strictly smaller than the target, the OOC ARL is large, while for shifts larger than the target, the OOC ARL is small. This is useful in ensuring that unproductive tinkering to the process is kept to the minimum.

The recursion (2.2) can be written in the equivalent form

$$D_i = \begin{cases} 0, & i = 0 \\ \max\{0, D_{i-1} + \xi_i - \zeta\}, & i \geq 1, \end{cases} \tag{2.5}$$

(Hawkins and Olwell, 1998, Section 1.9). This is the form that is typically used in practical applications and which also makes it clear that the CUSUM is in fact a Markov chain when in control. Of course, it is also of interest to monitor for a downward shift. Then the CUSUM recursion is

$$D_i^- = \begin{cases} 0, & i = 0 \\ \min\{0, D_{i-1}^- + \xi_i + \zeta\}, & i \geq 1 \end{cases}$$

with the corresponding run length N^-. We can simultaneously control for both upward and downward shifts using the two-sided CUSUM. Then, it is common practice to exhibit both the upper CUSUM sequence and the downward CUSUM sequence D_i^- against i in a single plot together with their corresponding control limits $\pm h$. The two-sided CUSUM signals at time

$$N^{\pm} = \min\{N, N^-\}.$$

It is well known (Van Dobben De Bruyn, 1968) that the relation

$$\frac{1}{\mathrm{E}[N^{\pm}]} = \frac{1}{\mathrm{E}[N]} + \frac{1}{\mathrm{E}[N^-]} \tag{2.6}$$

holds. When $\mu = 0$ the two terms on the right-hand side are equal, hence $\mathrm{E}[N^{\pm}] = \mathrm{E}[N]/2 = \mathrm{E}[N^-]/2$ in this case. Therefore, to attain a given ARL_0 the control limit h is chosen to make both the upper and downward IC ARLs equal to twice the nominal IC ARL of the two-sided CUSUM.

The presence of the zero barrier in D_i given in (2.5) causes the CUSUM to be restarted and to forget its past behaviour. Thus, D_i has the renewal property and using this property it can be shown that $\mathrm{E}[N]$ is finite – see (Siegmund, 1985, Section 2.6). Thus, the CUSUM has a false signal probability $\mathrm{P}(N < \infty | \mu = 0) = 1$ where the term "false signal" means that the CUSUM signals a shift whilst there has not been a true shift.

2.2 The likelihood ratio CUSUM

Suppose that $X_1, \ldots, X_\tau$ are independent and have density function $f(x)$ and that $X_{\tau+1}, X_{\tau+2}, \ldots$ are independent with density function $f(x-\mu)$, $\mu \neq 0$. Here, μ is the target shift of specified size μ. The target shift size is defined to be the smallest shift μ which is regarded as "operationally" significant. One interpretation of this is that shifts of size smaller than μ are of little or no consequence, while shifts in excess of μ are important and should be detected as quickly as possible.

Then the likelihood ratio for distinguishing between the hypotheses H_0 : *no change occurs* (the in-control situation) and H_1 : *a change occurs* (the out-of-control situation) is

$$\Lambda_\tau = \frac{f(X_{\tau+1}-\mu)\ldots f(X_i-\mu)}{f(X_{\tau+1})\ldots f(X_i)}. \tag{2.7}$$

Define

$$D_i = \max_{0\le\tau\le i-1} \sum_{j=\tau+1}^{i} \log\frac{f(X_j-\mu)}{f(X_j)}. \tag{2.8}$$

Then, (2.8) can be written as

$$\begin{aligned} D_i &= \max_{0\le\tau\le i-1}\left(\sum_{j=1}^{i}\log\frac{f(X_j-\mu)}{f(X_j)} - \sum_{j=1}^{\tau}\log\frac{f(X_j-\mu)}{f(X_j)}\right) \\ &= \sum_{j=1}^{i}\log\frac{f(X_j-\mu)}{f(X_j)} - \min_{0\le\tau\le i-1}\sum_{j=1}^{\tau}\log\frac{f(X_j-\mu)}{f(X_j)}. \end{aligned} \tag{2.9}$$

Then, (2.9) has the form of (2.2) with $\xi_j - \zeta = \log\frac{f(X_j-\mu)}{f(X_j)}$. In the special case where f is the standard normal density, we find that $\xi_j - \zeta = \mu(X_j - \mu/2)$. Replacing h in (2.3) by μh leads to the Page CUSUM (2.2) with $\zeta = \mu/2$.

Non-normal distributions typically do not have such simple forms of ξ_i and ζ. Nevertheless, if μ is "close" to zero, then (2.9) can be written approximately in the form (2.5) as we now show. This will provide a justification for formulating our distribution-free CUSUMs in the form of (2.5) with ξ_i replaced by a suitable sequential rank score. Set $\xi_j = \log\frac{f(X_j-\mu)}{f(X_j)} - \mathrm{E}\left[\log\frac{f(X-\mu)}{f(X)}\right]$ and $\zeta = -\mathrm{E}\left[\log\frac{f(X-\mu)}{f(X)}\right]$. The Taylor expansion of $f(X-\mu)$, neglecting terms of order $\mu^p,\ p\ge 3$, is

$$f(X-\mu) \approx f(X) - \mu f'(X) + \frac{\mu^2}{2} f''(X)$$

from which we get

$$\begin{aligned}
\log \frac{f(X-\mu)}{f(X)} &\approx \log \frac{f(X) - \mu f'(X) + \mu^2 f''(X)/2}{f(X)} \\
&= \log\left(1 - \mu \frac{f'(X)}{f(X)} + \frac{\mu^2 f''(X)}{2f(X)}\right) \\
&\approx -\mu \frac{f'(X)}{f(X)} + \frac{\mu^2 f''(X)}{2f(X)} - \frac{\mu^2}{2}\left(\frac{f'(X)}{f(X)}\right)^2,
\end{aligned}$$

the last approximation following from the Taylor expansion of $\log(1-x)$ for x "close" to zero. Assuming $f(\pm\infty) = 0$, we have

$$\mathrm{E}\left[\frac{f'(X)}{f(X)}\right] = \int_{-\infty}^{\infty} \left(\frac{f'(v)}{f(v)}\right) f(v)dv = \int_{-\infty}^{\infty} f'(v)dv = 0$$

and, assuming that $f'(\pm\infty) = 0$, we have

$$\mathrm{E}\left[\frac{f''(X)}{f(X)}\right] = \int_{-\infty}^{\infty} \left(\frac{f''(v)}{f(v)}\right) f(v)dv = \int_{-\infty}^{\infty} f''(v)dv = 0.$$

Then,

$$\mathrm{E}\left[\log \frac{f(X-\mu)}{f(X)}\right] \approx -\frac{\mu^2}{2} \mathrm{E}\left[\left(\frac{f'(X)}{f(X)}\right)^2\right] = -\frac{\mu^2}{2} I_0(f)$$

where

$$I_0(f) = \mathrm{E}\left[\left(\frac{f'(X)}{f(X)}\right)^2\right]$$

is the Fisher information in the location parameter family $\{f(x-\mu),\ -\infty < \mu < \infty\}$ (Hájek et al., 1999, Section 2.2.3).

Thus, $\log \frac{f(X_j-\mu)}{f(X_j)}$ in (2.9) can be approximated by $\xi_j - \zeta$ where

$$\xi_j = \log \frac{f(X_j - \mu)}{f(X_j)} + \frac{\mu^2}{2} I_0(f)$$

and

$$\zeta = \frac{\mu^2}{2} I_0(f).$$

Then, (2.9) has the same form as (2.2) with $\mathrm{E}[\xi_j] = 0$ for all $1 \le j \le \tau$.

We can obtain a parallel approximation for the CUSUM recursion (2.5) to detect scale shifts of target size $\rho \neq 1$ in a scale parameter family $\{\rho^{-1} f(X/\rho),\ 0 < \rho < \infty\ \rho \neq 1\}$. Then we can approximate ξ_j and ζ by

$$\xi_j = \log \frac{\rho^{-1} f(X_j/\rho)}{f(X_j)} + \frac{\rho^2}{2} I_1(f)$$

and

$$\zeta = \frac{\rho^2}{2} I_1(f)$$

where

$$I_1(f) = \mathrm{E}\left[\left(-1 - X\frac{f'(X)}{f(X)}\right)^2\right] = \int_{-\infty}^{\infty} \left(-1 - x\frac{f'(x)}{f(x)}\right)^2 f(x)dx,$$

the Fisher information in the scale parameter family $\{\rho^{-1} f(x/\rho),\ 0 < \rho < \infty\ \rho \neq 1\}$ (Hájek et al., 1999, Section 2.2.3).

2.3 Parametric self-starting CUSUMs

The discussion of standard normal CUSUMs stressed the necessity of precise in-control parameter specifications. If the data $X_1, X_2, \ldots$ arise from a $N(\mu, \sigma^2)$ distribution and μ and σ^2 are known, the standard procedure is to apply the standard normal CUSUM to the standardised data $(X - \mu)/\sigma$. There are, however, countless examples of problems where the in-control parameters are, at least to some extent, unknown. The application of the standard normal CUSUM cannot be justified when directly applying estimates of μ and σ from an in-control Phase I sample. Hawkins and Olwell (1998, Chapter 7) give an example where the estimated IC ARL is substantially inflated when the Phase I estimates of σ and μ are used as if they were the true values. For these estimates to be sufficiently close to their true values, an enormous amount of Phase I data would be required. This is clearly problematic in most practical implementations. One way of dealing with this problem is for the standard normal CUSUM to continuously incorporate updated estimates of μ and σ. The self-starting CUSUM of Hawkins (1987) is such a scheme. In this section we will describe the NSS (normal self-starting) CUSUM for a mean shift and the NSS and GSS (gamma self-starting) CUSUMs for a variance shift.

2.3.1 The NSS mean CUSUM

Suppose that the i.i.d. random variables $X_1, X_2, \ldots$ follow a normal $N(\mu, \sigma^2)$ distribution where both μ and σ are unknown. The NSS CUSUM of Hawkins (1987) is based on the sequence of statistics

$$T_i = \frac{X_i - \overline{X}_{i-1}}{s_{i-1}},$$

for $i \geq 3$ and where

$$\overline{X}_{i-1} = \frac{1}{i-1}\sum_{j=1}^{i-1} X_j$$

and

$$s_{i-1} = \sqrt{\frac{1}{i-2}\sum_{j=1}^{i-1}(X_j - \overline{X}_{i-1})^2}.$$

Observe that the distribution of T_i does not depend on σ or μ. Set

$$\xi_i = \Phi^{-1}\left(H\left(\sqrt{\frac{i-1}{i}}T_i, i-2\right)\right), \tag{2.10}$$

where $H(\cdot, \nu)$ denotes the distribution function of a t distribution with ν degrees of freedom. Then the ξ_i sequence are i.i.d. standard normal for $i \geq 3$ (Hawkins, 1987). That the T_i are statistically independent is a consequence of Basu's lemma – see Lehmann and Casella (1998, Theorem 6.2.1, p.42). The NSS CUSUM sequence is given by (2.5) for ξ_i in (2.10). The recommended reference value and control limits are the same as for the standard normal CUSUM. Whether the truly "optimal" reference value for the NSS is in fact $\zeta = \mu/2$, has to date not been assessed in the literature.

2.3.2 The NSS standard deviation CUSUM

Suppose that we wish to detect an increase of size ρ in the standard deviation from σ (unknown) to $\rho\sigma$ in a normal distribution. The NSS CUSUM then consists in substituting ξ_i^2 for ξ_i from (2.10) in the CUSUM recursion (2.5). The recommended reference value is

$$\zeta = \frac{\log \rho^2}{1 - \rho^{-2}} \tag{2.11}$$

and the same control limits as for the standard normal variance CUSUM apply – see Hawkins and Olwell (1998, Section 4.1.3, p.86-87).

2.3.3 The GSS standard deviation CUSUM

Suppose we wish to detect an increase of size ρ in the standard deviation of a Gamma distribution with the density function

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}}\, x^{\alpha-1}\, \exp(-x/\beta)$$

where the shape parameter α is known. The GSS CUSUM consists in using

$$\xi_i = G^{-1}\left(F_{2\alpha,2\alpha(i-1)}\left(\frac{X_i}{\overline{X}_{i-1}}\right), \alpha, 1\right) \tag{2.12}$$

in (2.5). In (2.12) $F_{2\alpha,2\alpha(i-1)}$ denotes the distribution function of the F distribution with 2α and $2\alpha(i-1)$ degrees of freedom and $G^{-1}(\cdot, \alpha, 1)$ denotes the inverse of the Gamma distribution function with parameters α and 1. For the degrees of freedom to be defined, the GSS CUSUM requires at least $m = 3$ observations to initiate. The reference value is

$$\zeta = \frac{\log \rho}{1 - \rho^{-1}}.$$

The special case $\alpha = 1$ is the exponential distribution for which Gan (1992) provides control limits.

Hawkins and Olwell (1998, Chapter 7) use the term "self starting" because they argue that a large Phase I sample is not needed for the CUSUM monitoring to commence. However, Keefe et al. (2015) show that caution should be exercised when applying the scheme of Hawkins (1987). This CUSUM can only initiate at the third observation. Keefe et al. (2015) show that the in-control ARL of this CUSUM depends on $\overline{X}_2$ and s_2 and is not, in general, equal to the nominal value. The implication is that if two practitioners are sampling from the same population and their $\overline{X}_2$ and s_2 differ, then their IC ARLs will, in general, also differ and not equal the nominal IC ARL. Thus, the continual updating of the estimates of μ and σ does not have the desired effect. They refer to this as between-practitioner variation. On the contrary, if the unconditional IC ARL is computed by averaging the conditional IC ARL over all possible $m \geq 2$ initial observations, it equals the nominal value. However, it is a moot point whether the unconditional IC ARL is a relevant quantity. The same issues are present in the NSS and GSS standard deviation CUSUMs.

2.4 The Girschick-Rubin CUSUMs

Let $X_1, \ldots, X_\tau$ be i.i.d. $N(0,1)$ and let $X_{\tau+1}, \ldots, X_n$ be i.i.d. $N(\mu, 1)$ where $\mu \neq 0$ is the target shift size. The GR CUSUM is constructed by summing the likelihood ratio (2.7), rather than maximising it, over possible changepoints $\tau = 0, 1, \ldots, i-1$ (Girschick and Rubin, 1952). Thus, we define

$$D_i = \sum_{\tau=0}^{i-1} \Lambda_\tau = \sum_{\tau=0}^{i-1} \prod_{j=\tau+1}^{i} \frac{\phi(X_j - \mu)}{\phi(X_j)} = \sum_{\tau=0}^{i-1} \exp\left(\mu(S_i - S_\tau) - \frac{(i-\tau)\mu^2}{2}\right) \tag{2.13}$$

where $S_i = \sum_{j=1}^{i} X_j$. Observe that (2.13) can be written in the recursive form

$$\begin{aligned}
D_i &= \sum_{\tau=0}^{i-1} \exp\left(\mu(S_i - S_\tau) - \frac{(i-\tau)\mu^2}{2}\right) \\
&= \sum_{\tau=0}^{i-2} \exp\left(\mu(S_i - S_\tau) - \frac{(i-\tau)\mu^2}{2}\right) + \exp\left(\mu(S_i - S_{i-1}) - \frac{(i-i+1)\mu^2}{2}\right) \\
&= \sum_{\tau=0}^{i-2} \exp\left(\mu(S_{i-1} + X_i - S_\tau) - \frac{(i-1-\tau)\mu^2}{2} - \frac{\mu^2}{2}\right) + \exp\left(\mu X_i - \frac{\mu^2}{2}\right) \\
&= \exp\left(\mu X_i - \frac{\mu^2}{2}\right)\left(\sum_{\tau=0}^{i-2} \exp\left(\mu(S_{i-1} - S_\tau) - \frac{(i-1-\tau)\mu^2}{2}\right) + 1\right) \\
&= (1 + D_{i-1}) \exp\left(\mu\left(X_i - \frac{\mu}{2}\right)\right) \qquad (2.14)
\end{aligned}$$

for $i \geq 1$ and with $D_0 = 0$ (Pollak, 1987, p.752), which is more suited to practical implementation. The GR CUSUM signals at time

$$N = \min\{i \geq 1 : D_i \geq h\} \qquad (2.15)$$

that a shift has possibly occurred, where $h > 0$ is the control limit that guarantees a nominal IC ARL. An approximation to the control limit h is given by Pollak (1987), who uses the facts that D_i has the Markov property and that $(D_i - i)$ is a martingale. The approximation is

$$ARL_0 \approx h/\vartheta(\mu) \qquad (2.16)$$

for a "large" h, where

$$\vartheta(\mu) = 2\mu^{-2} \exp\left(-2\sum_{j=1}^{\infty} \frac{1}{j}\Phi\left(-\frac{\sqrt{\mu^2 j}}{2}\right)\right) \approx \exp(-0.583\mu)$$

for a μ "close" to 0 (Pollak and Siegmund, 1991). The recursion (2.14) shows that the GR CUSUM is a Markov chain and by argumentation similar to that in Section 2.1, we can infer that the GR CUSUM will produce a false alarm with probability 1.

To detect a shift in the standard deviation of a $N(0,1)$ distribution, the GR CUSUM can be constructed as follows. Let $X_1, \ldots, X_\tau$ be i.i.d. $N(0,1)$ and let $X_{\tau+1}, X_{\tau+2}, \ldots$ be i.i.d. $N(0, \rho^2)$ where $\rho > 1$ is the target shift size. Then, $D_0 = 0$ and, for $i \geq 1$,

$$\begin{aligned}
D_i = \sum_{\tau=0}^{i-1} \Lambda_\tau &= \sum_{\tau=0}^{i-1} \prod_{j=\tau+1}^{i} \frac{\rho^{-1}\phi(X_j \rho^{-1})}{\phi(X_j)} \\
&= \sum_{\tau=0}^{i-1} \rho^{-i-\tau} \exp\left(\frac{\rho^2 - 1}{2\rho^2}\left(\sum_{j=1}^{i} X_j^2 - \sum_{j=1}^{\tau} X_j^2\right)\right)
\end{aligned}$$

which can be written in the recursive form

$$D_i = (1 + D_{i-1})\rho^{-1} \exp\left(\frac{1}{2}\left(1 - \rho^2\right) X_i^2\right). \tag{2.17}$$

The run length N is also defined by (2.15). A formula such as (2.16) is not available to obtain appropriate control limits. However, a table of control limits has been generated by Monte Carlo simulation and is shown in Table 9 of Appendix B.

Self-starting GR CUSUMs analogous to those of Hawkins and Olwell (1998, Chapter 7) can be constructed. However, the problem of between-practitioner ARL variation is also present in these CUSUMs.

2.4.1 The two-sided GR CUSUM

It is important to also detect a downward shift in the mean. The downward GR CUSUM recursion D_i^- is given by (2.14) where we replace μ with $-\mu$ everywhere. To detect a decrease in standard deviation of size $\rho < 1$ the recursion is defined by D_i^-, which is (2.17). The downward GR CUSUM signals at time

$$N^- = \min\{i \geq 1 : D_i^- \geq h^-\}$$

where $h^- = h$ for the GR mean CUSUM. The two-sided GR CUSUM signals at time

$$N^{\pm} = \min\{N, N^-\}.$$

Analogous to the relation (2.6) one would expect the following to hold

$$\frac{1}{\mathrm{E}[N^{\pm}]} = \frac{2}{\mathrm{E}[N]} \tag{2.18}$$

in the in-control case if both upper and downward CUSUMs have the same IC ARL. Such a relation has not yet been proven for the GR CUSUM. However, the following numerical evidence indicates that the relation might well be true in this case also. In Monte Carlo simulations with 10 000 independent trials we found that, with a target $\mu = 0.5$, the absolute difference between the left-hand and right-hand sides of (2.18) is approximately 0.0002 when the true shift is 0.

In the GR standard deviation CUSUM, the target shifts are $\rho = 1 + \lambda > 1$ (upper CUSUM) and $\rho = 1 - \lambda < 1$ (downward CUSUM). Then, $h \neq h^-$ if the nominal IC ARLs of the upper and downward CUSUMs are the same. Then (2.18) seems to be true again.

Next, we will discuss the following distribution-free CUSUMs: the McDonald (1990) CUSUM, the Hawkins and Deng (2010) CUSUM and the Ross and Adams (2012) CUSUM.

Since no parameter estimates or Phase I data are required for these CUSUMs to function properly, they are devoid of the problems surrounding unknown parameters in the parametric case. In Chapters 3 and 4, the performance and properties of these CUSUMs will be compared to the sequential rank CUSUMs that we construct.

2.5 Distribution-free CUSUMs

2.5.1 The McDonald CUSUM

McDonald (1990) developed a distribution-free CUSUM to detect a shift away from an unknown current median. His CUSUM is based on the unsigned sequential ranks

$$r_i = 1 + \sum_{j=1}^{i} \mathbb{1}\left(X_j < X_i\right).$$

If $X_1, X_2, \ldots$ are i.i.d., then $\xi_i = r_i/(1+i)$ are, for $i \geq 1$, independent and non-identically distributed and converge to a uniform distribution on (0,1) as $i \to \infty$. McDonald (1990) proceeds to develop a CUSUM for the uniform distribution, the idea being that the properties of this CUSUM would approximate those of a CUSUM based directly on the sequential ranks. In Chapter 3, we construct CUSUMs based directly on the sequential ranks and investigate their in-control and out-of-control properties.

2.5.2 The Hawkins and Deng CUSUM

Recently, Hawkins and Deng (2010) proposed a distribution-free CUSUM to detect a shift away from a current unknown median. Their CUSUM is based on a sequential version of the two-sample Wilcoxon statistic. Henceforth we will refer to this as the HD CUSUM. Define, for $i \geq 15$,

$$\xi_i = \max_{1 \leq \tau \leq i-1} |\xi_{\tau,i}| \tag{2.19}$$

with

$$\xi_{\tau,i} = \left\{ \sum_{j=1}^{\tau} \sqrt{\frac{12(i+1)}{i-1}} \left(\frac{R_{i,j}}{1+i} - \frac{1}{2} \right) \right\} \Big/ \sqrt{\tau(i-\tau)(i-1)}$$

where $R_{i,j}$ is the rank of X_j among $X_1, \ldots, X_i$. The HD CUSUM is defined by the recursion (2.5) with $D_0 = D_1 = \cdots = D_{14} = 0$, $\zeta = 0$ and signals at time

$$N = \min\{i \geq 15 : D_i \geq h_i\}$$

where h_i is the control limit which depends on i. The presence of the absolute value in (2.19) shows that this is in fact a two-sided CUSUM.

The question arises how the sequence of control limits h_i is to be chosen in order to guarantee a nominal IC ARL. Hawkins and Deng (2010) define the h_i by fixing at $\alpha = 1/ARL_0$ the conditional probability of a false signal at i, given no previous false signal. That this yields an IC ARL of size ARL_0 is proved as follows. Let $C_1, C_2, \ldots$ be a sequence of statistics and let $h_1, h_2, \ldots$ be constants chosen such that, for $0 < \alpha < 1$ and all $i \geq 1$,

$$\mathrm{P}(C_i > h_i | C_1 \leq h_1, \ldots, C_{i-1} \leq h_{i-1}) = \alpha. \tag{2.20}$$

Define the run length

$$N = \min\{i \geq 1 : C_i \geq h_i\}.$$

Then,

$$\mathrm{P}(N = 1) = \mathrm{P}(C_1 > h_1) = \alpha \tag{2.21}$$

and for $i \geq 2$

$$\begin{aligned} \mathrm{P}(N = i) &= \mathrm{P}(C_1 \leq h_1, \ldots, C_{i-1} \leq h_{i-1}, C_i > h_i) \\ &= \mathrm{P}(\xi_i > h_i | C_1 \leq h_1, \ldots, C_{i-1} \leq h_{i-1})\ \mathrm{P}(C_1 \leq h_1, \ldots, C_{i-1} \leq h_{i-1}) \\ &= \alpha\ \mathrm{P}(N > i - 1). \end{aligned} \tag{2.22}$$

Thus, for all $i \geq 1$ we have the relation

$$\mathrm{P}(N = i) = \alpha\, \mathrm{P}(N > i - 1).$$

From (2.21) and (2.22) we find

$$1 = \sum_{i \geq 1} \mathrm{P}(N = i) = \alpha \sum_{i \geq 1} \mathrm{P}(N > i - 1) = \alpha\, \mathrm{E}[N],$$

whence $\mathrm{E}[N] = 1/\alpha$.

The distribution in (2.20) is highly discrete for small values of i making it impossible to find h_i which makes the left side of (2.20) exactly equal to α. This ceases to be a problem for $i \geq 15$. Table 1 of Hawkins and Deng (2010) gives control limits h_i for typical IC ARLs, ranging from 50 to 2000.

A substantial computational effort is required to obtain the h_i. Also the HD CUSUM in its present form does not incorporate a reference value ζ, which implies that it cannot be tuned to meet specific OOC ARL objectives.

2.5.3 The Ross and Adams CUSUM

Distribution-free CUSUMs for shifts more general than shifts in the median or scale were developed by Ross and Adams (2012). We will refer to these as RA CUSUMs. These CUSUMs are based on the Kolmogorov-Smirnov (KS) and the Cramér-von-Mises (CvM) statistics. In general, Ross and Adams (2012) conclude that the CUSUM based on the CvM statistic outperforms the one based on the KS statistic and, in addition, that the CvM is simpler to implement. The CvM statistic is

$$\psi_{\tau,i} = \sum_{j=1}^{i} \left(\widehat{F}_{1,i}(X_j) - \widehat{F}_{2,i}(X_j)\right)^2$$

where

$$\widehat{F}_{1,i}(x) = \frac{1}{\tau}\sum_{j=1}^{\tau} \mathbb{1}(X_j \le x)$$

and

$$\widehat{F}_{2,i}(x) = \frac{1}{i-\tau}\sum_{j=\tau+1}^{n} \mathbb{1}(X_j \le x).$$

The standardised CvM statistic is, for $i \ge 1$,

$$\xi_i = \max_{\tau \ge 1} \frac{\psi_{\tau,i} - \mu_\psi}{\sigma_\psi}$$

where

$$\mu_\psi = \frac{i+1}{6i}$$

and

$$\sigma_\psi = \frac{(i+1)\left(\left(1-\frac{3}{4\tau}\right)i^2 + (1-\tau)i - \tau\right)}{45 i^2 (i-\tau)}.$$

The CUSUM signals a shift when $\xi_i > h_i$ $(i \ge 20)$ where the h_i are control limits. The h_i are chosen in the same manner as in the HD CUSUM. Software for the implementation of both the CvM and KS RA CUSUMs is available in the R package called "cpm" (http://cran.r-project.org/web/packages/cpm/index.html).

3 Sequential rank CUSUMs for location

A natural approach towards constructing CUSUMs that are free of overly specific distributional assumptions is to replace the observed data by rank-based equivalents that are distribution free. In this chapter, we introduce distribution-free CUSUMs based on signed and unsigned sequential ranks to detect a location shift. We again distinguish between two scenarios: detecting a location shift in a distribution (I) when the in-control median is specified; and (II) when the in-control median is unspecified. When the distribution is in control, these CUSUMs are distribution free in that the control limits do not depend on the functional form of the underlying distribution. The sequential rank CUSUMs do not require the existence of any moments of the underlying distribution and are robust against the effect of spurious outliers. We briefly discuss the design of the signed sequential rank (SSR) CUSUM of Van Zyl (2015) and Lombard and Van Zyl (2018) which centres attention on scenario I. We then proceed to develop the unsigned sequential rank location (SRL) CUSUM for scenario II. The CUSUMs are defined as in (2.5) with ξ_i and ζ appropriately chosen.

3.1 The signed sequential rank CUSUM

3.1.1 In-control properties

Suppose $X_1, X_2, \ldots$ are in control if the X_i are continuously and symmetrically distributed around zero with density function $\sigma f(x\sigma)$. Define the sequential rank of $|X_i|$ among $|X_1|, \ldots, |X_i|$, for $i \geq 1$, as

$$r_i^+ = 1 + \sum_{j=1}^{i} \mathbb{1}(|X_j| < |X_i|)$$

and let $s_i r_i^+ = \text{sign}(X_i) r_i^+$ denote the signed sequential rank of X_i. As long as the distribution remains in control, $s_i = \pm 1$ with equal probability 1/2 and the r_i^+ are uniformly distributed on the numbers $\{1, 2, \ldots, i\}$. The r_i^+ are statistically mutually independent and also independent of the signs s_i (Reynolds, 1975). Then, $\text{E}[s_i r_i^+] = 0$. Notice that $s_i r_i^+$ is scale invariant, that is, its value remains unchanged if X_i is replaced by aX_i, $a > 0$. This implies that we may assume, without loss of generality, that $\sigma = 1$. This comes down to expressing the data and the target shift size μ in units of the underlying scale parameter.

Let $\psi(u), u \in (-1,1)$ be an odd and square-integrable function with $\int_{-1}^{1} \psi(u)du = 0$. Set, for $i \geq 1$,

$$\xi_i = \psi\left(\frac{s_i r_i^+}{1+i}\right) \Big/ \sqrt{\eta_i} = s_i \psi\left(\frac{r_i^+}{1+i}\right) \Big/ \sqrt{\eta_i} \tag{3.1}$$

where

$$\eta_i = \frac{1}{i} \sum_{j=1}^{i} \psi^2\left(\frac{j}{1+i}\right) \tag{3.2}$$

so that ξ_i has unit variance. The Wilcoxon statistics, using the score $\psi(u) = u$, are

$$\xi_i = \sqrt{\frac{6(1+i)}{2i+1}} \left(\frac{s_i r_i^+}{1+i}\right) \tag{3.3}$$

which are, for $i \geq 1$, statistically independent with zero means and unit variances. Another popular score is $\psi(u) = \Phi^{-1}(u)$, leading to the Van der Waerden statistics

$$\xi_i = s_i \ \Phi^{-1}\left(\frac{1}{2}\left(1 + \frac{r_i^+}{1+i}\right)\right) \Big/ \sqrt{\eta_i} \tag{3.4}$$

where η_i is given in (3.2). The corresponding CUSUMs will be referred to as the Wilcoxon SSR and Van der Waerden SSR CUSUMs. If the median of the X_i increases away from zero, or even if the distribution of X is asymmetric, $\mathrm{E}[\xi_i]$ ceases to be zero. Therefore, the CUSUM can be expected to be useful in detecting either a shift away from a zero median in a symmetric distribution or in detecting the presence of asymmetry. Because the summand in (3.3) is bounded, the resulting CUSUM (2.5) can be expected to be robust against outliers.

Lombard and Van Zyl (2018) provide control limits for a range of reference values and nominal IC ARLs for the Wilcoxon and the Van der Waerden CUSUMs. For completeness we include these here in Tables 3.1 and 3.2.

Table 3.1: Control limits for the Wilcoxon SSR CUSUM.

	Nominal IC ARL						
ζ	100	200	300	400	500	1000	2000
0.00	8.92	13.07	16.24	18.90	21.30	30.24	43.95
0.10	6.45	8.62	10.05	11.12	12.01	14.79	17.93
0.15	5.65	7.34	8.42	9.21	9.86	11.88	14.06
0.20	5.00	6.37	7.24	7.87	8.37	9.96	11.57
0.25	4.46	5.61	6.33	6.85	7.25	8.52	9.84
0.30	4.01	5.00	5.60	6.03	6.37	7.45	8.53
0.35	3.62	4.48	5.00	5.37	5.66	6.58	7.51
0.40	3.29	4.04	4.49	4.81	5.06	5.87	6.66
0.45	2.99	3.66	4.05	4.34	4.56	5.24	5.96
0.50	2.73	3.31	3.68	3.93	4.13	4.74	5.34

Table 3.2: Control limits for the Van der Waerden SSR CUSUM.

	Nominal IC ARL						
ζ	100	200	300	400	500	1000	2000
0.00	8.808	13.055	16.192	19.048	21.283	30.519	43.599
0.05	7.322	10.317	12.333	13.929	15.210	19.835	24.942
0.10	6.362	8.520	9.945	11.019	11.893	14.787	17.832
0.15	5.532	7.171	8.344	9.173	9.825	11.875	13.987
0.20	4.929	6.352	7.198	7.836	8.321	9.945	11.629
0.25	4.456	5.668	6.320	6.862	7.245	8.578	9.950
0.30	3.997	5.015	5.604	6.099	6.427	7.550	8.654
0.35	3.633	4.503	5.066	5.423	5.756	6.720	7.704
0.40	3.340	4.108	4.588	4.930	5.201	6.062	6.918
0.50	2.800	3.452	3.845	4.135	4.350	5.039	5.732

A comparison between the control limits in Tables 3.1 and 3.2 with those of a standard normal CUSUM at the same reference values $\zeta \le 0.5$ reveals a close correspondence at ARL_0 values in excess of 500. This is a result of the fact that the partial sums $\sum_{i=1}^{n} \xi_i/\sqrt{n}$ converge in distribution to a standard normal distribution as $n \to \infty$.

3.1.2 Out-of-control behaviour

Here, we provide a summary of the main result, namely that the out-of-control properties of the SSR CUSUM are very similar to those of a standard normal CUSUM. More extensive details are given in Lombard and Van Zyl (2018).

Let $X_1, \ldots, X_\tau$ have the common distribution function $F(x)$ and let $X_{\tau+1}, X_{\tau+2}, \ldots$ have the common distribution function $G(x) = F(x - \mu)$ where μ is "small". Let ξ_i be as defined in (3.1) and define the partial sums

$$S_n = \sum_{i=1}^{n} (\xi_i - \zeta).$$

Lombard and Van Zyl (2018) show that the joint distributions of $S_n,\ n \ge 1$ can be approximated by the joint distributions of

$$S_n^* = \sum_{i=1}^{n} (X_i^* - \zeta) + \mu\theta \max(0, n - \tau) = X_1^* + \cdots + X_n^* - n\zeta + \mu\theta \max(0, n - \tau)$$

where $X_1^*, X_2^*, \ldots$ are i.i.d. $N(0,1)$ quantities and where

$$\theta = \frac{1}{\sqrt{\eta}} \int_{-\infty}^{\infty} \psi'(2F(x) - 1) f^2(x) dx \tag{3.5}$$

with $\eta = \int_0^1 \psi^2(u) du$. This can be summarised as the following heuristic.

Heuristic 3.1. *Let ζ be "small" and let a persistent shift of "small" size μ_1 occur at a "large" changepoint τ. Then, the SSR CUSUM behaves approximately as would a standard normal CUSUM with the same μ_1, ζ and h when a shift of size $\mu_1\theta$ commences after τ.* ∎

3.1.3 Design of the CUSUM

Heuristic 3.1 implies in particular that the OOC ARL will be a monotone function of μ_1 if the score ψ is monotone. Furthermore, the OOC ARL of the SSR CUSUM can be estimated given any μ_1, ζ and τ by Monte Carlo simulation that uses only normal random numbers. Further details are given in Lombard and Van Zyl (2018).

From Heuristic 3.1 it follows that

$$\mathrm{E}[\xi_{\tau+i}] \approx \mu\theta \neq 0, \tag{3.6}$$

for a large τ and a fixed i. Suppose we target a shift in the median from zero to $\mu > 0$. In a standard normal CUSUM, the optimal reference value is $\zeta = \mu/2$. Analogously, it seems sensible to use the reference value

$$\zeta = \mu\theta/2$$

in an SSR CUSUM.

The values of θ in Table 3 of Lombard and Van Zyl (2018), reproduced below as Table 3.3, over a range of standardised distributions can be used as a guideline to make an informed choice of ζ. We standardise the normal, t_4 and t_3 distributions to unit standard deviation and the t_2 and t_1 distributions to unit inter-quartile range, since the standard deviation is either infinite or not defined in the latter two cases.

Table 3.3: Values of θ for a range of t_ν distributions.

	Distribution				
Score	normal	t_4	t_3	t_2	t_1
Wilcoxon	0.98	1.18	1.38	1.18	1.10
Van der Waerden	1.00	1.12	1.29	1.06	0.93

If some in-control Phase I data $V_1, \ldots, V_m$ are available, then the appropriate value of θ can be estimated as follows. We estimate f by

$$\widehat{f}(v) = \frac{1}{m}\sum_{j=1}^{m}\frac{1}{b}\phi\left(\frac{v-V_j}{b}\right), \tag{3.7}$$

where b denotes the default (Silverman, 1986, p.45) bandwidth

$$b = 1.06m^{-1/5} \times \min\left\{s, \frac{IQR}{1.35}\right\}$$

and where s and IQR denote the sample standard deviation and the inter-quartile range of $V_1, \ldots, V_m$. Denote by $\widehat{F}(v)$ the empirical distribution function of $V_1, \ldots, V_m$. Observe that

$$\theta = \frac{1}{\sqrt{\eta}} \mathrm{E}\left[\psi'(2F(V) - 1) f(V)\right].$$

This suggests the consistent estimator

$$\widehat{\theta} \quad = \quad \frac{1}{m\sqrt{\eta}} \sum_{j=1}^{m} \psi'\left(2\frac{j}{m+1} - 1\right) \widehat{f}(V_{m:j})$$

of θ, where $V_{m:j}$ is the j^{th} order statistic of $V_1, \ldots, V_m$. Then, the estimated reference value becomes $\widehat{\zeta} = \widehat{\theta}\mu/2$ where μ is expressed in units of $\widehat{\sigma}$, an estimator of the chosen scale parameter σ. In applying the SSR CUSUM with reference value $\widehat{\zeta}$ we use the corresponding h from the appropriate table which guarantees that the Phase II IC ARL is equal to the nominal value. This guarantee comes as a consequence of the distribution-free nature of the SSR CUSUM and the independence of the Phase I and Phase II data. This behaviour stands in stark contrast to the behaviour of the normal CUSUM when the unknown σ is estimated from the data. In that case, the control limit h corresponding to $\widehat{\sigma}$ does not guarantee the correct IC ARL value.

3.2 The sequential rank location CUSUM

3.2.1 In-control properties

A vital assumption of the SSR CUSUM is that the in-control median is specified or known. If the in-control median is unspecified, or if the symmetry assumption is untenable, a useful CUSUM can be constructed using the unsigned sequential ranks

$$r_i = 1 + \sum_{j=1}^{i} \mathbb{1}(X_j < X_i) \tag{3.8}$$

of X_i among $X_1, \ldots, X_i$, $i \geq 1$. When $X_1, X_2, \ldots$ are i.i.d., the r_i are independent with mean $(1+i)/2$ and variance $(i^2-1)/12$ regardless of the distribution underlying the X_i – see Barndorff-Nielsen (1963). The unsigned sequential rank statistics for application in the CUSUM recursion (2.5) are, for $i \geq 2$,

$$\xi_i = \left(\psi\left(\frac{r_i}{1+i}\right) - \bar{\psi}_i\right) \Big/ \sqrt{\eta_i}, \tag{3.9}$$

where $\psi(u)$ is a square-integrable function on the interval (0,1) and where

$$\bar{\psi}_i = \frac{1}{i} \sum_{j=1}^{i} \psi\left(\frac{j}{1+i}\right)$$

and

$$\eta_i = \frac{1}{i}\sum_{j=1}^{i}\psi^2\left(\frac{j}{1+i}\right) - \bar{\psi}_i^2. \tag{3.10}$$

In the special cases where $\psi(u) = u - \frac{1}{2}$ (Wilcoxon) and $\psi(u) = \Phi^{-1}(u)$ (Van der Waerden), $\bar{\psi}_i = 0$ for all $i \geq 1$. Then, $\mathrm{E}[\xi_i] = 0$ and the ξ_i are, for $i \geq 2$, statistically independent with unit variance. Since $\eta_1 = 0$ the CUSUM starts at $i = 2$, that is, $D_0 = D_1 = 0$. Consider now the special case $\psi(u) = u - \frac{1}{2}$. Standardising $r_i/(1+i)$, which are asymptotically uniform (0,1) random variables as $i \to \infty$, gives, for $i \geq 2$,

$$\xi_i = \sqrt{\frac{12(i+1)}{(i-1)}}\left(\frac{r_i}{1+i} - \frac{1}{2}\right). \tag{3.11}$$

We call the CUSUM based on (3.11) the Wilcoxon SRL CUSUM. This CUSUM is reminiscent of the one proposed by McDonald (1990). Another popular score function is $\psi(u) = \Phi^{-1}(u)$, leading to

$$\xi_i = \Phi^{-1}\left(\frac{r_i}{1+i}\right)\Big/\sqrt{\eta_i} \tag{3.12}$$

for $i \geq 2$ and where η_i tends to 1 as $i \to \infty$. To avoid confusion between the score (3.12) and that in (3.4), we will use the term "normal score" for the former rather than "Van der Waerden score", which was used for (3.4).

The following calculations lead us to focus in particular on the Wilcoxon score (3.11). For a distribution with density function f and distribution function F the efficient score for location changes is defined as (Hájek et al., 1999, Section 2.2.4)

$$J(u) = -\frac{f'\left(F^{-1}(u)\right)}{f\left(F^{-1}(u)\right)},\ 0 \leq u \leq 1 \tag{3.13}$$

which has finite variance $I_0(f) = \mathrm{Var}\left[J(U)\right]$, the Fisher information. We compare the Wilcoxon and the normal scores with the efficient scores in a range of distributions. The comparison entails computing the correlation coefficient between the Wilcoxon score and the efficient score, and that between the normal score and the efficient score in a given distribution. We consider the t_ν, the Gumbel and the skew-normal distributions. The correlation coefficients give an indication of how well the Wilcoxon or normal SRL CUSUMs would fare against parametric CUSUMs constructed for data from these distributions when there are no unknown nuisance parameters. The closer the correlation is to 1, the better is the expected performance of the SRL CUSUM.

The density function of Student's t_ν distribution with zero mean is

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}. \tag{3.14}$$

The efficient score (3.13) is

$$J(u) = \frac{(\nu+1)F_\nu^{-1}(u)}{\nu + \left(F_\nu^{-1}(u)\right)^2}$$

where F_ν^{-1} denotes the inverse distribution function of a t_ν distribution. The t_ν distribution exhibits a range of tail heaviness from moderate ($\nu = \infty$) to extremely heavy ($\nu = 1$). Therefore, the t_ν distribution provides a useful basis for our comparison. The correlation coefficients between the efficient scores in five t distributions with the Wilcoxon and normal scores are shown in Table 3.4. Overall, the Wilcoxon score seems to be preferred.

Turning to skew distributions, we consider the Gumbel and skew-normal distributions (Azzalini, 2005). The density of the Gumbel distribution is

$$f(x) = \exp(-x - \exp(-x)) \tag{3.15}$$

and the efficient score (3.13) is

$$J(u) = 1 - \exp\left(-F^{-1}(u)\right).$$

The skew-normal (SN(α)) distribution has density function

$$f(x) = 2\phi(x)\Phi(\alpha x) \tag{3.16}$$

where α is the skewness parameter and the efficient score (3.13) is

$$J(u) = -\left(\frac{\alpha\phi(\alpha F^{-1}(u))}{\Phi(\alpha F^{-1}(u))} - F^{-1}(u)\right).$$

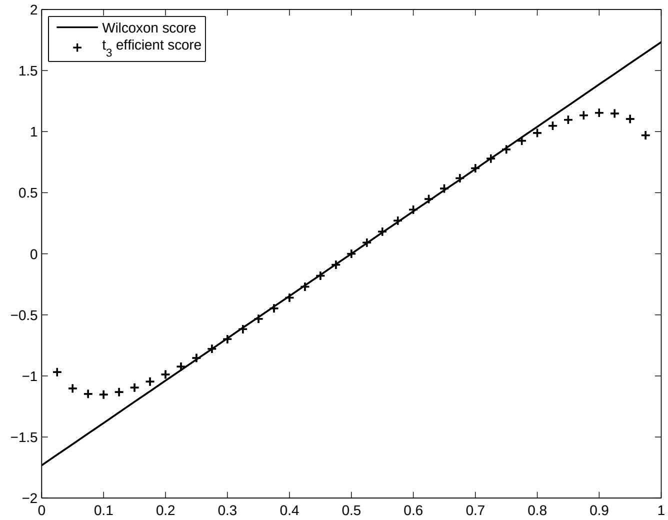
Table 3.4 also gives the correlation coefficients of the efficient score with the Wilcoxon and normal scores in these skew distributions. Again, overall, the Wilcoxon seems to be preferred.

Table 3.4: Correlation of the Wilcoxon and normal scores with the efficient scores (3.13) in various distributions.

	Distribution								
Score	normal	t_4	t_3	t_2	t_1	Gumbel	SN(±1)	SN(±2)	SN(±4)
Wilcoxon	0.98	0.99	0.97	0.93	0.78	0.87	0.91	0.86	0.75
normal	1.00	0.94	0.91	0.84	0.66	0.90	0.89	0.85	0.75

In contrast to the efficient scores in Student t distributions, both the Wilcoxon and normal scores are monotone functions of the OOC median. Figure 3.1 shows a plot of the Wilcoxon score and the efficient score in a t_3 distribution. The non-monotone nature of the t_3 efficient score seen in the figure explains why some larger shifts $|\mu_2| > |\mu_1| > 0$ would be harder to detect than smaller shifts $|\mu_1|$ if a CUSUM for a t_3 distribution were used – see MacEachern et al. (2007) and Liu et al. (2015). In contrast, the monotonicity of the Wilcoxon score shows that this issue will not appear when it is used.

Figure 3.1: The Wilcoxon and t_3 efficient scores where the horizontal axis is $0 \le u \le 1$.



3.2.1.1 Finding the control limit

Given a set of reference values ζ and specified nominal IC ARLs, it is our first priority to find control limits. The distribution-free character of the sequential ranks enables us to obtain control limits for the SRL CUSUM by Monte Carlo simulation using $U(0,1)$ random variables. The partial sums of the ξ_i converge to a normal distribution when the X sequence is in control. Therefore, it is not difficult to imagine that the control limits h of the SRL CUSUM will correspond closely to those of a standard normal CUSUM, especially when ζ is small and h is large. Denote by h_1 the control limits from a standard normal CUSUM, given (ζ, ARL_0). The first step of an iterative algorithm was to estimate the IC ARL (denote the estimate by $\widehat{A}(\zeta, h_1)$) of the SRL CUSUM on the (ζ, h) grid using, for instance, 10 000 independent Monte Carlo-generated realisations with $U(0,1)$ as the in-control distribution. Cubic spline interpolation from $(\zeta, \widehat{A}(\zeta, h_1))$ to (ζ, h) then yields new estimates, h_2, of the correct control limits. A further 10 000 independent Monte Carlo-generated realisations using h_2 produce a new estimated IC ARL $(\zeta, \widehat{A}(\zeta, h_2))$. These steps were repeated until all of the absolute differences $|\widehat{A}(\zeta, h) - ARL_0|$ were less than 3. For $\zeta \le 0.25$, no more that three iterations were required, while for $\zeta > 0.25$, six iterations sufficed. We can expect the number of iterations required to increase as ζ becomes larger.

We find the control limits for the Wilcoxon SRL CUSUM to be approximately the same as those of the Wilcoxon SSR CUSUM. Therefore, we use the control limits in Table 3.1 also for the Wilcoxon SRL CUSUM. For the same reason, we use the control limits given in Table 3.2 for the normal SRL CUSUM.

3.2.2 Out-of-control behaviour

The SRL CUSUM exhibits an out-of-control feature that is different from the SSR CUSUM. The latter CUSUM increases indefinitely after a shift occurred. In contrast, the SRL CUSUM will continue to increase for a while after a shift, but will then start to decrease again to what seems to be an in-control state. The following calculations demonstrate this feature in the special case of the Wilcoxon score $\psi(u) = u - \frac{1}{2}$.

Define $Y_1, Y_2, \ldots$ to be i.i.d. quantities with distribution function $F(y)$ and set $X_{\tau+i} = Y_{\tau+i} + \mu$ which has distribution function $G(x) = F(x - \mu)$. Then we have, for $i \geq 1$,

$$r_{\tau+i} = 1 + \sum_{j=1}^{\tau} \mathbb{1}(Y_j < Y_{\tau+i} + \mu) + \sum_{j=1}^{i-1} \mathbb{1}(Y_{\tau+j} + \mu < Y_{\tau+i} + \mu). \tag{3.17}$$

Let F_τ denote the empirical distribution function of $Y_1, \ldots, Y_\tau$ and let G_i denote the empirical distribution function of $Y_{\tau+1}, \ldots, Y_{\tau+i}$. Then,

$$\frac{r_{\tau+i}}{\tau+i+1} = \frac{1}{\tau+i+1} + \frac{\tau}{\tau+i+1} F_\tau(Y_{\tau+i} + \mu) + \frac{i-1}{\tau+i+1} G_i(Y_{\tau+i} + \mu).$$

Conditional upon $Y_{\tau+i} = y$,

$$\begin{aligned} \mathrm{E}\left[\frac{r_{\tau+i}}{\tau+i+1} \middle| Y_{\tau+i} = y\right] &= \frac{1}{\tau+i+1} + \frac{\tau}{\tau+i+1} \mathrm{E}\left[F_\tau(y+\mu)\right] + \frac{i-1}{\tau+i+1} \mathrm{E}\left[F_i(y)\right] \\ &= \frac{1}{\tau+i+1} + \frac{\tau}{\tau+i+1} F(y+\mu) + \frac{i-1}{\tau+i+1} F(y). \end{aligned}$$

Then,

$$\begin{aligned} \mathrm{E}\left[\frac{r_{\tau+i}}{\tau+i+1}\right] &= \mathrm{E}\left[\mathrm{E}\left[\frac{r_{\tau+i}}{\tau+i+1} \middle| Y_{\tau+i}\right]\right] \\ &= \frac{1}{\tau+i+1} + \frac{\tau}{\tau+i+1} \mathrm{E}\left[F(Y_{\tau+i}+\mu)\right] + \frac{i-1}{\tau+i+1} \mathrm{E}\left[F_i(Y_{\tau+i})\right] \\ &= \frac{1}{\tau+i+1} + \frac{\tau}{\tau+i+1} \mathrm{E}\left[F(Y_{\tau+i}+\mu)\right] + \frac{i-1}{\tau+i+1} \left(\frac{1}{2}\right) \end{aligned}$$

which converges to 1/2 as $i \to \infty$, because $0 \leq \mathrm{E}\left[F(\cdot)\right] \leq 1$. Thus, for the Wilcoxon SRL summand (3.11)

$$\lim_{i \to \infty} \mathrm{E}\left[\xi_{\tau+i}\right] = 0$$

for a fixed $\tau > 0$. Intuitively, this result seems reasonable as well: The effect of the pre-shift median on the post-shift sequential ranks diminishes as observations continue to accrue after the shift. The sequential ranks then again become independent and uniformly distributed with the result that $\mathrm{E}[\xi_{\tau+i}]$ tends to zero so that the new median μ now functions as the in-control value. The implication for statistical practice is that the signal from the SRL CUSUM should be acted upon quickly.

To examine the behaviour of the SRL CUSUM when a shift occurs at a "large" changepoint, we will suppose that $\tau \to \infty$ and $i \to \infty$, but that $i/\tau = O(1)$. Continuing from (3.17), we find that for both τ and i large and a "small" μ

$$\begin{aligned} \mathrm{E}\left[\frac{r_{\tau+i}}{\tau+i+1}\right] &\approx \frac{\tau}{\tau+i+1}\mathrm{E}\left[F(Y_{\tau+i}+\mu)\right] + \frac{i-1}{\tau+i+1}\mathrm{E}\left[F(Y_{\tau+i})\right] \\ &\approx \frac{\tau}{\tau+i+1}\mathrm{E}\left[F(Y_{\tau+i}) + \mu f(Y_{\tau+i})\right] + \frac{i-1}{\tau+i+1}\left(\frac{1}{2}\right) \\ &\approx \frac{1}{2} + \mathrm{E}\left[\frac{\tau}{\tau+i+1}\mu f(Y_{\tau+i})\right]. \end{aligned}$$

Then we find that, in the special case $\psi(u) = u - \frac{1}{2}$,

$$\mathrm{E}\left[\xi_{\tau+i}\right] \approx \frac{\tau}{\tau+i+1}\mu\sqrt{12}\,\mathrm{E}[f(Y_{\tau+i})] \approx \tau\log\left(\frac{\tau+i+2}{\tau+i+1}\right)\mu\sqrt{12}\int_{-\infty}^{\infty} f^2(y)dy$$

which is non-zero, implying that the out-of-control median of ξ will differ from the in-control median, hence the change should be detectable.

Previously, we showed that the SSR CUSUM behaves in an out-of-control situation more or less like the standard normal CUSUM with a linear drift. We now formulate an analogous heuristic for the SRL CUSUM. Define θ by

$$\theta = \frac{1}{\sqrt{\eta}}\int_{-\infty}^{\infty}\psi'(F(x))f^2(x)dx \tag{3.18}$$

with $\eta = \int_0^1 \psi^2(u)du - \bar{\psi}^2$ and $\bar{\psi} = \lim_{i\to\infty}\bar{\psi}_i$.

Heuristic 3.2. *Let ζ be "small" and let a persistent shift of "small" size μ_1 occur at a "large" changepoint τ. Then, the SRL CUSUM behaves approximately as would a standard normal CUSUM with the same μ_1, ζ and h when shifts of size $\mu_1\theta\tau\log\left(\frac{n}{n-1}\right)$ commence after $n = \tau$.* ■

In the same manner in which Heuristic 3.1 was derived, the following calculations lead to Heuristic 3.2. Based on Theorem 5.1 of Lombard (1983), in the special case $s = -1$, see Remark 1 on page 103 there, the heuristic is derived in the following manner. Define the partial sums $S_n = \sum_{i=1}^n(\xi_i - \zeta)$. Suppose that $\mu_1 = \beta/h$ and $\zeta = \Delta/h$, so that μ_1 and ζ are "small" when h is "large". Then

$$\left\{\frac{S_{\lfloor h^2 t\rfloor}}{h}, 0 \le t \le 1\right\} \overset{\mathcal{D}}{\Longrightarrow} \left\{W(t) - \Delta t + \beta\tau^*\theta\log\left(\max\left(\frac{t}{\tau^*},1\right)\right), 0 \le t \le 1\right\} \tag{3.19}$$

as $h \to \infty$ where $W(t)$ denotes the standard Brownian motion. Transforming (3.19) back to $t = n/h^2$, $\tau^* = \tau/h^2$, $\zeta = \Delta/h$ and $\mu_1 = \beta/h$, we find that for a "large" enough h

$$\begin{aligned} \left\{\frac{S_n}{h}, 1 \le n \le \lfloor h^2\rfloor\right\} &\overset{\mathcal{D}}{\approx} \left\{W\left(\frac{n}{h^2}\right) - n\frac{\zeta}{h} + \frac{\mu_1}{h}\theta\tau\log\left(\max\left(\frac{n}{\tau},1\right)\right), 1 \le n \le \lfloor h^2\rfloor\right\} \\ &= \frac{1}{h}\left\{W(n) - \zeta n + \mu_1\theta\tau\log\left(\max\left(\frac{n}{\tau},1\right)\right), 1 \le n \le \lfloor h^2\rfloor\right\}, \end{aligned} \tag{3.20}$$

because the Brownian motion $W\left(n/h^2\right)$ has the same distribution as $W(n)/h$, that is,

$$\left\{S_n, 1 \le n \le \lfloor h^2 \rfloor\right\} \overset{\mathcal{D}}{\approx} \left\{W(n) - \zeta n + \mu_1 \theta \tau \log\left(\max\left(\frac{n}{\tau}, 1\right)\right), 1 \le n \le \lfloor h^2 \rfloor\right\}.$$

Note that $W(n) - \zeta n$ equals $\sum_{k=1}^{n}(X_k^* - \zeta)$, which are the partial sums figuring in the standard normal CUSUM and where $X_1^*, X_2^* \dots$ are i.i.d. $N(0,1)$ quantities. Bhattacharya and Frierson (1981) derive this result for the Wilcoxon score as a special case of (3.19).

3.2.3 Design of the CUSUM

Suppose we target a median shift of size μ. Analogous to the standard normal CUSUM, it seems sensible to use the reference value $\zeta = \mu\theta/2$. For the Wilcoxon SRL score $\theta = \sqrt{12}\,\mathrm{E}[f(X)]$ as a special case of (3.18). To gain an idea of the reference values that we are likely to encounter, we calculate these (shown in Table 3.5) for standardised distributions and typical choices of the target μ. Where the variance is infinite or undefined (the t_2 and t_1 distributions), we standardise the distribution to have unit IQR. Thus, while the precise functional form of the underlying distribution may be unknown, the values in Table 3.5 enable one to make a somewhat rational choice of reference value after taking into account the likely tail thickness and skewness of the in-control distribution. Observe that there is not substantial variation in the "optimal" reference values at the various distributions. Since the out-of-control behaviour is determined by the single parameter θ, rather than by the form of the distribution, a minor "misspecification" of the "optimal" reference value should not have a great impact on the OOC ARL of the CUSUM. In the case of the Wilcoxon SRL CUSUM, the reference value must not exceed $\sqrt{12}$, otherwise the CUSUM will be and remain identically zero. However, it is unlikely in practice that interest would centre on target shifts of sizes larger than $2 \times \sqrt{12} = 6.928$ standard deviations.

Table 3.5: Values of θ and reference values ζ for the Wilcoxon SRL CUSUM with target shift μ.

		Distribution								
		normal	t_4	t_3	t_2	t_1	Gumbel	SN(±1)	SN(±2)	SN(±4)
θ		0.98	1.18	1.38	1.18	1.10	1.11	0.98	1.00	1.05
	0.25	0.12	0.15	0.17	0.15	0.14	0.14	0.12	0.13	0.13
μ	0.50	0.25	0.30	0.35	0.30	0.28	0.28	0.25	0.25	0.26
	1.00	0.49	0.59	0.69	0.59	0.55	0.56	0.49	0.50	0.53

Heuristic 3.2 provides us with a useful method of estimating the OOC behaviour of a CUSUM prior to implementation. Suppose we fix a reference value ζ and a control limit h to guarantee a nominal IC ARL value. We wish to estimate the OOC ARL of the CUSUM at various choices of the changepoint and the true shift. Denote by $\mathcal{W}(\mu_1)$ the OOC ARL of the SRL CUSUM (based on either the Wilcoxon or normal score) when a median shift of size $\mu_1 > 0$ occurs and by $\mathcal{N}(\mu_1\theta)$ the OOC ARL of a standard normal CUSUM with the same ζ and h when there are mean shifts of size $\mu_1\theta\tau\log\left(\frac{\tau+i}{\tau+i-1}\right)$, $i \ge 1$. Then, Heuristic 3.2 says that, for

"small" ζ and "large" τ,

$$\mathcal{W}(\mu_1) \approx \mathcal{N}(\mu_1\theta). \tag{3.21}$$

There exist currently no explicit expressions or analytic approximations to either $\mathcal{W}(\mu_1)$ or $\mathcal{N}(\mu_1\theta)$ in the literature. Nevertheless, we can easily estimate the numerical value of $\mathcal{W}(\mu_1)$ by Monte Carlo simulation from the true underlying distribution and the approximation $\mathcal{N}(\mu_1\theta)$ by Monte Carlo simulation using only normal random numbers for a given μ_1, ζ and τ.

With a view to practical design of the CUSUM, we now assess the degree to which the approximation (3.21) may be useful. Towards this, we use Monte Carlo simulation in four distributions – the normal and t_3, which are symmetric, and the Gumbel and skew-normal(4), which are asymmetric, all standardised to unit standard deviation. For $\tau = 100$, we compare $\mathcal{N}(\mu_1\theta)$ with $\mathcal{W}(\mu_1)$ (based on the Wilcoxon score) in one-sided upper CUSUMs with nominal IC ARL 500 for target shift sizes $\mu = 0.25$ and 0.5. We take $\zeta \approx \theta\mu/2$ as the reference value (we are, for the moment, assuming that θ is known). Table 3.6 shows estimated OOC ARLs from 10 000 Monte Carlo trials per design. The control limits for the SRL CUSUM come from linear interpolation in Table 3.1. In the table below, the triple $(\mu;\ \zeta;\ h)$ denotes the target shift size, the reference value and the control limit h that guarantees the nominal IC ARL. The true shift sizes are μ_1=0.125, 0.25, 0.375, 0.50, 0.75 and 1.00.

We are particularly interested in the approximation at shifts $\mu_1 \geq \mu$ where the OOC ARL is indicated in boldface in Table 3.6. In the normal and t_3 distributions, the approximation (3.21) fares acceptably well at $\mu = 0.25$ and 0.5 and would most certainly be useful for the purpose of designing the CUSUM. If, however, the distribution is skew, the results indicate that (3.21) is less useful for design purposes.

Table 3.6: OOC ARL approximations for (3.21) if $\mathcal{W}$ is based on the Wilcoxon score.

Distribution	θ	$(\mu;\ \zeta;\ h)$	Approx.	μ_1					
				0.125	0.25	0.375	0.50	0.75	1.00
normal	0.98	(0.25; 0.12; 11.08)	$\mathcal{W}(\mu_1)$	259	**124**	**56**	**33**	**19**	**14**
			$\mathcal{N}(\mu_1\theta)$	276	**126**	**55**	**32**	**17**	**12**
		(0.50; 0.25; 7.24)	$\mathcal{W}(\mu_1)$	303	161	76	**38**	**17**	**12**
			$\mathcal{N}(\mu_1\theta)$	305	164	74	**35**	**15**	**10**
t_3	1.38	(0.25; 0.17; 9.22)	$\mathcal{W}(\mu_1)$	218	**73**	**31**	**20**	**12**	**10**
			$\mathcal{N}(\mu_1\theta)$	221	**73**	**30**	**18**	**10**	**8**
		(0.50; 0.35; 5.66)	$\mathcal{W}(\mu_1)$	278	118	44	**21**	**11**	**8**
			$\mathcal{N}(\mu_1\theta)$	239	101	37	**19**	**9**	**6**
Gumbel	1.11	(0.25; 0.14; 10.26)	$\mathcal{W}(\mu_1)$	265	**114**	**51**	**28**	**16**	**12**
			$\mathcal{N}(\mu_1\theta)$	259	**103**	**44**	**25**	**14**	**10**
		(0.50; 0.28; 6.68)	$\mathcal{W}(\mu_1)$	310	168	76	**35**	**15**	**10**
			$\mathcal{N}(\mu_1\theta)$	284	136	58	**27**	**12**	**8**
SN(4)	1.05	(0.25; 0.13; 10.64)	$\mathcal{W}(\mu_1)$	267	**123**	**54**	**31**	**17**	**13**
			$\mathcal{N}(\mu_1\theta)$	266	**112**	**48**	**28**	**15**	**11**
		(0.50; 0.26; 7.07)	$\mathcal{W}(\mu_1)$	317	178	85	**39**	**16**	**11**
			$\mathcal{N}(\mu_1\theta)$	298	148	63	**31**	**14**	**9**

In practical applications, θ is unknown of course. However, the methodology in Section 3.1.3 can be applied to obtain a consistent estimator of θ. Corresponding to $\widehat{\theta}$, there exists a known control limit $\widehat{h}$ (given $\widehat{\zeta} = \widehat{\theta}\mu/2$) that will guarantee that the Phase II IC ARL is equal to the nominal value. This comes as a consequence of the independence of the Phase I and II data. Table 3.6 was generated under the assumption that the true value of θ is known. If θ is estimated from a sufficient amount of Phase I data, one can expect some degree of degeneration in the quality of the approximation.

3.2.4 Comparison with the Hawkins and Deng and the Ross and Adams CUSUMs

In this section, we compare the two-sided SRL CUSUM with the distribution-free HD and RA CUSUMs discussed in Sections 2.5.2 and 2.5.3. In particular, we use the Wilcoxon SRL CUSUM because it serves usefully as an omnibus CUSUM. We use data from a standardised normal, t_3 and skew-normal ($\alpha = 4$) distribution. Table 3.7 shows estimated OOC ARLs (from 20 000 Monte Carlo trials) of the CUSUMs at a range of true shifts μ_1. These shifts were induced at the changepoints $\tau = 50$ and $\tau = 250$. To initiate the HD CUSUM, denoted by HD in the table, we use the recommended 14 initial observations (Hawkins and Deng, 2010), while for the RA CUSUM (denoted by RA in the table) we use the recommended 19 (Ross and Adams, 2012). Thus, when we say that $\tau = 50$ for the HD CUSUM, we mean that observations $X_1, \ldots, X_{14}$ are used to initiate the CUSUM, $X_{15}, \ldots, X_{50}$ are in control and $X_{51}, X_{52}, \ldots$ are out of control. For all the CUSUMs, the two-sided nominal IC ARL was specified at 500. For the SRL CUSUM we use target shift sizes $\mu = 0.25$, 0.5 and 1.0. Recall that neither the HD CUSUM nor the RA CUSUM uses a reference value. In Table 3.7 the heading SRL_μ denotes the SRL CUSUM given the target μ, with the reference value $\zeta = \mu\theta/2$ and control limit h indicated in the last two rows of the table. For the normal distribution $\theta = 0.98$, for the t_3 distribution $\theta = 1.38$ and for the SN(4) distribution $\theta = 1.05$.

From Table 3.7 we observe the following: At small target shift $\mu \leq 0.5$, the SRL CUSUM performs better (smaller OOC ARL) than both the HD and the RA CUSUMs given that the true shift μ_1 is small. We also see clearly the ability of the SRL CUSUM to be tuned to a specific target. Observe in all three distributions that the HD CUSUM gives much the same OOC ARL as the SRL CUSUM at the target $\mu = 0.5$. When the target shift is large (e.g. $\mu = 1.00$), the HD and the RA CUSUMs perform better than the SRL CUSUM, which we expect because the SRL CUSUM is designed specifically to detect small persistent shifts.

Table 3.7: OOC ARL comparison of the Wilcoxon SRL CUSUM with the HD and RA CUSUMs in the standardised normal, t_3 and skew-normal($\alpha = 4$) distributions.

normal data										
	$\tau = 50$					$\tau = 250$				
μ_1	$SRL_{0.25}$	$SRL_{0.5}$	$SRL_{1.0}$	HD	RA	$SRL_{0.25}$	$SRL_{0.5}$	$SRL_{1.0}$	HD	RA
0.10	435	462	481	489	474	354	406	437	417	423
0.25	292	358	413	384	388	118	176	271	169	182
0.50	84	127	249	142	153	35	35	62	38	41
0.75	29	31	78	33	36	21	19	20	18	19
1.00	19	16	21	15	16	16	13	13	11	12
ζ	0.12	0.245	0.49							
h	13.517	8.664	4.830							
t_3 data										
	$\tau = 50$					$\tau = 250$				
μ_1	$SRL_{0.25}$	$SRL_{0.5}$	$SRL_{1.0}$	HD	RA	$SRL_{0.25}$	$SRL_{0.5}$	$SRL_{1.0}$	HD	RA
0.10	417	460	480	463	455	303	384	434	365	355
0.25	215	312	413	286	282	66	113	249	80	80
0.50	37	65	215	47	44	22	21	38	21	20
0.75	17	16	57	14	14	14	12	13	11	11
1.00	12	10	13	9	9	11	9	8	8	7
ζ	0.17	0.345	0.69							
h	11.050	6.643	3.253							
skew-normal($\alpha = 4$) data										
	$\tau = 50$					$\tau = 250$				
μ_1	$SRL_{0.25}$	$SRL_{0.5}$	$SRL_{1.0}$	HD	RA	$SRL_{0.25}$	$SRL_{0.5}$	$SRL_{1.0}$	HD	RA
0.10	445	479	498	490	492	356	433	499	441	457
0.25	297	371	468	393	407	113	190	351	173	190
0.50	81	145	305	142	158	32	34	87	36	38
0.75	28	34	122	34	39	20	17	22	18	18
1.00	18	16	31	15	15	15	12	12	11	11
ζ	0.13	0.263	0.525							
h	12.906	8.238	4.507							

In the table above we assumed that we know the "optimal" reference value ζ for use in the CUSUM. However, in practice one rarely knows a priori what the truly "optimal" reference value should be. To evaluate the OOC ARL performance of the SRL CUSUM (based on the Wilcoxon score) at "misspecified" ζs, we will conduct the following simulation. Suppose that the data arise from a t_3 distribution, but we choose the reference values appropriate for the normal distribution; and that data arise from a normal distribution, but we choose the reference value appropriate for the t_3 distribution. We specify an overall ARL_0 of 500. We induce median shifts of various sizes $\mu_1 = 0.1$, 0.25, 0.5, 0.75 and 1.00 at the changepoints $\tau = 50$ or $\tau = 250$. The OOC ARL estimates are shown in Table 3.8 where the two last rows give the "wrong" reference values and their corresponding control limits. We see that when a smaller reference value is used, the OOC ARL at small shifts are slightly smaller, while the reverse is true when a slightly larger reference value is used. When a too large reference value is used, the OOC ARL estimates are much larger (e.g. $\zeta = 0.69$ instead of $\zeta = 0.49$). This is to be expected, because the SRL CUSUM is designed with the aim of detecting small persistent shifts – which a target of $\mu = 1.0$ is no longer. Clearly, if tight control is required, a smaller rather than a larger ζ is preferred.

Table 3.8: OOC ARL estimates of the SRL CUSUM at "misspecified" reference values ζ.

normal data						
	$\tau = 50$			$\tau = 250$		
μ_1	$SRL_{0.25}$	$SRL_{0.5}$	$SRL_{1.0}$	$SRL_{0.25}$	$SRL_{0.5}$	$SRL_{1.0}$
0.10	453	466	489	377	435	465
0.25	316	373	438	139	217	336
0.50	96	178	315	34	41	105
0.75	28	46	153	19	18	28
1.00	17	16	50	14	12	13
ζ	0.17	0.345	0.69			
h	11.05	6.643	3.253			

t_3 data						
	$\tau = 50$			$\tau = 250$		
μ_1	$SRL_{0.25}$	$SRL_{0.5}$	$SRL_{1.0}$	$SRL_{0.25}$	$SRL_{0.5}$	$SRL_{1.0}$
0.10	432	461	484	362	406	452
0.25	297	359	432	118	178	279
0.50	89	140	255	35	37	68
0.75	30	37	98	22	19	22
1.00	19	16	28	16	13	13
ζ	0.12	0.245	0.49			
h	13.517	8.664	4.83			

4 Sequential rank CUSUMs for scale

In this chapter, we construct sequential rank CUSUMs to detect a shift in scale. Let the in-control observations $X_1, \ldots, X_\tau$ have the density function $f(x - \mu)$, while the out-of-control observations $X_{\tau+1}, X_{\tau+2}, \ldots$ have the density function $g(x) = f((x-\mu)/\sigma_1)/\sigma_1$. A scale parameter typically functions as an indicator of dispersion in the underlying distribution. The heading of this chapter could therefore just as well read "Sequential rank CUSUMs for dispersion". When the distribution is in control, these CUSUMs are distribution free. In agreement with the location CUSUMs, we distinguish between two scenarios: detecting a scale shift when (I) the in-control median μ is specified; and (II) the in-control median μ is unspecified. We will not assume any value of the in-control scale parameter σ. If, however, σ is known this information cannot be usefully incorporated. This is because sequential ranks and ordinary ranks are scale invariant. Thus, the properties of the rank-based procedures will be the same, whether σ is known or not and, therefore, we can take $\sigma = 1$ without loss of generality.

While the location CUSUMs are scale invariant, their application nevertheless requires the scale to remain constant. Running a scale CUSUM together with a location CUSUM enables one to monitor the validity of this fundamental assumption. We develop unsigned sequential rank scale CUSUMs for scenario II. A discussion of the sequential rank scale CUSUM for scenario I can be found in Lombard and Van Zyl (2018).

4.1 The sequential rank scale CUSUM (median unknown)

4.1.1 In-control properties

When, at the onset of monitoring, either the in-control median is unknown or the symmetry assumption is untenable, a useful CUSUM to detect a persistent shift in scale can be based on the sequential ranks r_i in (3.8). We assume that both the in-control median and scale are unknown. The CUSUM recursion (2.5) can be based on the summands

$$\xi_i = \left(\psi\left(\frac{r_i}{1+i}\right) - \bar{\psi}_i\right)^2 \Big/ \eta_i - 1, \tag{4.1}$$

which are, for $i \geq 2$, independent with zero mean and where ψ is the score used in the SRL CUSUM and η_i is given in (3.10) – see Section 3.2.1. Two possible choices are the squares of the Wilcoxon summand (3.11),

$$\xi_i = \frac{12(i+1)}{i-1}\left(\frac{r_i}{1+i} - \frac{1}{2}\right)^2 - 1, \tag{4.2}$$

centred to have zero mean and which is reminiscent of the score used in the Mood test in Hájek et al. (1999, Section 4.2); and the squares of the normal summand (3.12),

$$\xi_i = \left\{\Phi^{-1}\left(\frac{r_i}{1+i}\right)\right\}^2 \Big/ \eta_i - 1, \tag{4.3}$$

where η_i is given in (3.10). Observe that (4.3) is based on the score used in the two-sample Klotz test (Hájek et al., 1999, Section 4.2). Thus, we will refer to the resulting CUSUMs as the "Mood" (MSR) and "Klotz" (KSR) CUSUMs, respectively.

Tables 4.1 and 4.2 show control limits for the MSR and KSR CUSUMs, respectively, at a range of reference values ζ. These were obtained by Monte Carlo simulation and cubic spline interpolation in the manner set out in Section 3.2.1.1.

Table 4.1: Control limits for the MSR CUSUM.

	IC ARL						
ζ	100	200	300	400	500	1000	2000
0.000	7.991	11.676	14.528	16.972	19.050	27.363	39.112
0.100	5.747	7.638	8.875	9.764	10.529	12.976	15.605
0.150	5.044	6.557	7.479	8.197	8.717	10.545	12.382
0.200	4.472	5.715	6.492	7.034	7.501	8.910	10.363
0.250	4.038	5.117	5.735	6.207	6.582	7.717	8.910
0.300	3.675	4.598	5.138	5.553	5.850	6.815	7.835
0.400	3.078	3.830	4.237	4.560	4.789	5.537	6.312
0.500	2.638	3.236	3.592	3.831	4.019	4.633	5.235

Table 4.2: Control limits for the KSR CUSUM.

	IC ARL						
ζ	100	200	300	400	500	1000	2000
0.000	10.704	16.263	20.650	24.346	27.753	41.161	61.566
0.100	8.562	12.340	14.855	16.903	18.631	24.678	31.721
0.200	7.319	10.285	12.087	13.597	14.762	18.753	23.227
0.250	6.811	9.374	11.158	12.495	13.411	17.085	20.892
0.375	5.954	8.116	9.477	10.537	11.410	14.205	17.239
0.500	5.317	7.168	8.445	9.348	10.070	12.485	14.997
0.625	4.774	6.489	7.582	8.425	9.120	11.282	13.578
0.750	4.406	5.963	7.000	7.719	8.365	10.371	12.472

In order to gain an impression of how the MSR and KSR CUSUMs will perform compared to CUSUMs based on efficient scores in a range of distributions, we compute the correlation coefficient of the MSR and KSR scores with the efficient scores. The efficient score for a distribution in a scale parameter family with density function f and distribution function F is (Hájek et al., 1999, p.18)

$$J(u) = -1 + F^{-1}(u)\left(-\frac{f'(F^{-1}(u))}{f(F^{-1}(u))}\right), \tag{4.4}$$

which has finite variance $I_1(f) = \mathrm{Var}[J(U)]$, the Fisher information. Analogous to the comparison that we made in Section 3.2.1, we calculate the correlation coefficients between the MSR score and the efficient score and that between the KSR score and the efficient score for the distributions (3.14), (3.15) and (3.16) and show these values in Table 4.3. The MSR score tends to have higher correlation coefficients in the heavy-tailed t distributions. Since we seek robustness against outliers, this provides some justification for focusing our attention on the MSR CUSUM.

Table 4.3: Correlation of the Mood and Klotz scores with the efficient scores (4.4).

	Distribution								
Score	normal	t_4	t_3	t_2	t_1	Gumbel	SN(±1)	SN(±2)	SN(±4)
Mood	0.87	0.98	0.99	1.00	0.96	0.84	0.77	0.76	0.75
Klotz	1.00	0.95	0.92	0.87	0.74	0.97	0.86	0.81	0.76

4.1.2 Out-of-control behaviour

In Section 3.2.2 we showed that the SRL CUSUM behaves in an out-of-control situation more or less like a standard normal CUSUM with a logarithmic drift. We now formulate a similar heuristic for the sequential rank scale CUSUM. Define

$$\theta_1 = 2\int_{-\infty}^{\infty} \frac{1}{\eta}\psi(F(x))\psi'(F(x))xf^2(x)dx. \tag{4.5}$$

Heuristic 4.1. *Let ζ be "small" and let a persistent scale increase of "small" size σ_1 occur at a "large" changepoint τ. Then, the sequential rank scale CUSUM behaves approximately as would a standard normal CUSUM for the mean with the same ζ and h when shifts of size $\theta_1(\log\sigma_1)\tau\log\left(\frac{n}{n-1}\right)$ commence after $n = \tau$.* ∎

Heuristic 4.1 can be formulated from Theorem 5.1 of Lombard (1983) along the same lines as Heuristic 3.2, except that we replace μ_1 and θ there by $\log\sigma_1$ and θ_1 here.

The sequential rank scale CUSUM has OOC properties similar to those of the SRL CUSUM. The aforementioned CUSUM will only increase for a limited time after a shift before the effect of the pre-shift scale parameter on the post-shift sequential ranks will diminish and the CUSUM will return to seemingly in control behaviour. To see this, consider the right-hand side of (3.20) with τ fixed. Then,

$$\mathrm{E}[\xi_{\tau+i}] \approx (\log\sigma_1)\theta_1\tau\log\left(\frac{\tau+i}{\tau+i-1}\right)$$

and

$$\tau \log\left(\frac{\tau+i}{\tau+i-1}\right) = \tau \log\left(1+\frac{1}{\tau+i-1}\right) \approx \frac{\tau}{\tau+i-1},$$

which is positive and for every fixed $\tau > 0$, tends to zero as $i \to \infty$.

4.1.3 Design of the CUSUM

Suppose we target a scale shift of size $\rho > 1$. We can specify a reference value at the hand of Heuristic 4.1. The fact that the drift is logarithmic, and not linear, implies that the optimal reference value is unknown. Nevertheless, we will continue to propose

$$\zeta = \frac{\theta_1 \log \rho}{2}$$

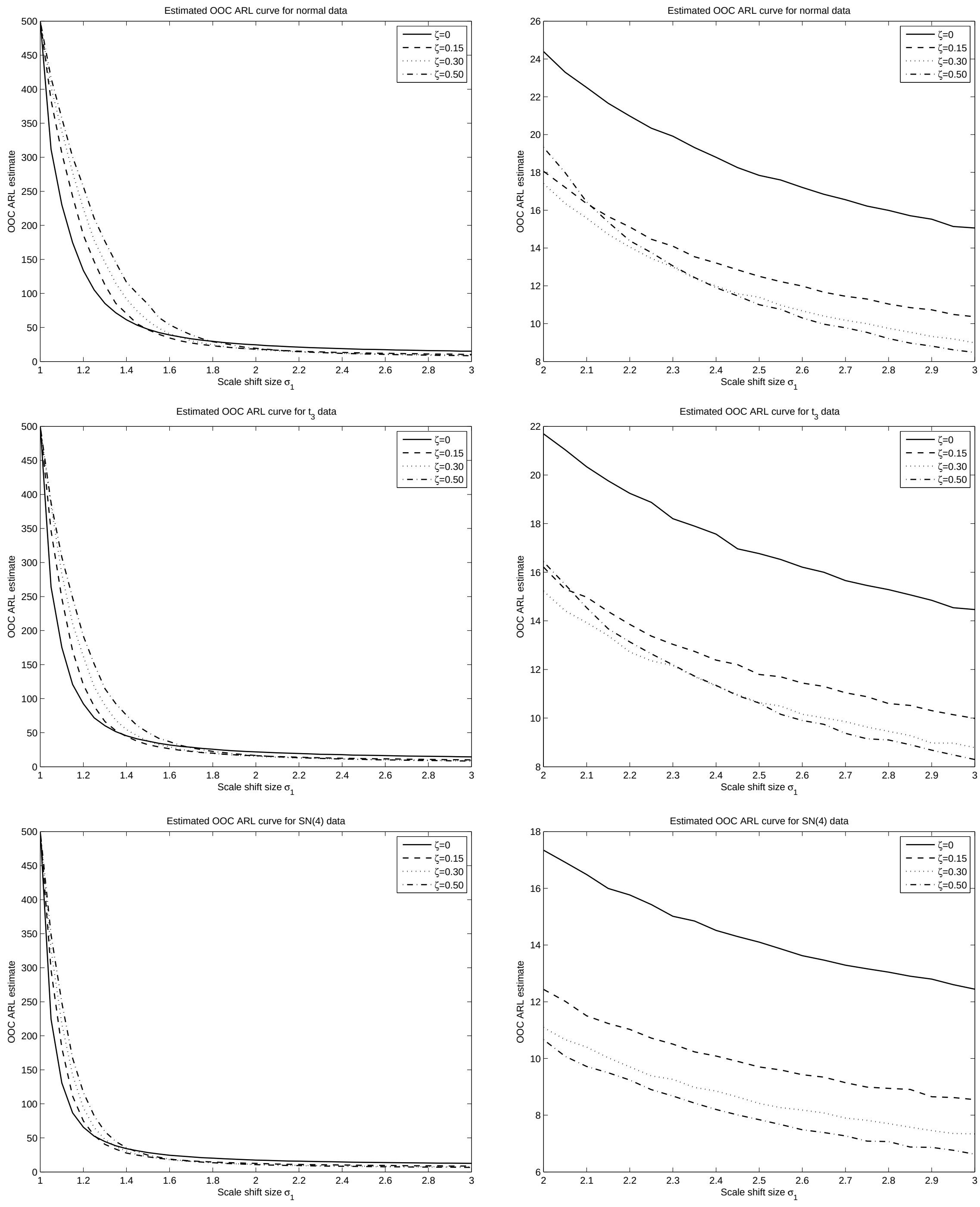
as an appropriate reference value, where θ_1 is defined in (4.5). Table 4.4, an analog of Table 3.5, shows a range of reference values for the MSR and KSR CUSUMs. To detect a downward shift of size $\rho < 1$, the reference value is $\zeta^- = \frac{-\theta_1 \log \rho}{2}$. Observe again, as in Table 3.5, that the "optimal" reference values do not vary extremely over the various distributions and, therefore, we do not expect a major impact on the OOC ARL when there are minor "misspecifications" in the reference value.

Table 4.4: Values of θ_1 and reference values ζ for the MSR and KSR CUSUMs with target shift ρ.

		Distribution								
		normal	t_4	t_3	t_2	t_1	Gumbel	SN(±1)	SN(±2)	SN(±4)
Mood score										
θ_1		1.1027	0.9351	0.8865	0.7999	0.6079	1.0118	1.0377	1.0306	0.9945
ρ	1.25	0.1230	0.1043	0.0989	0.0892	0.0678	0.1129	0.1158	0.1150	0.1110
	1.50	0.2236	0.1896	0.1797	0.1622	0.1232	0.2051	0.2104	0.2089	0.2016
	2.00	0.3822	0.3241	0.3072	0.2772	0.2107	0.3507	0.3596	0.3572	0.3447
Klotz score										
θ_1		2.0000	1.4377	1.3048	1.1026	0.7427	1.8470	1.8325	1.7368	1.5935
ρ	1.25	0.2231	0.1640	0.1456	0.0114	0.0829	0.2061	0.2045	0.1938	0.1778
	1.50	0.4055	0.2980	0.2645	0.0208	0.1506	0.3744	0.3715	0.3521	0.3231
	2.00	0.6931	0.5095	0.4522	0.0356	0.2574	0.6401	0.6351	0.6019	0.5523

To graphically illustrate that minor "misspecifications" in ζ do not have a disastrous effect on the OOC ARL, we resort to Monte Carlo simulation. For illustrative purposes, we only apply the MSR CUSUM, but we find the same results to be true when applying the KSR CUSUM. We specify an ARL_0 of 500 and obtain control limits for a range of reference values $\zeta = 0$, 0.15, 0.30 and 0.50 from Table 4.1. We simulate data from a normal, t_3 and $SN(4)$ distribution and induce various scale shifts $1 \leq \sigma_1 \leq 3$ at the changepoint $\tau = 250$. Figure 4.1 shows plots of the OOC ARL estimates from 20 000 Monte Carlo trials per design. The top two, middle two and bottom two figures pertain to the normal, t_3 and $SN(4)$ distributions, respectively. The right panels are magnified versions of the curves in the left panels for $2 \leq \sigma_1 \leq 3$. In the left panels we see that the smaller ζ is, the better the CUSUM is able to detect small increases away from σ. In the right panels the situation is reversed, but the differences, except at $\zeta = 0$, are small enough not to be of much practical significance. Thus, if tight control is required, a small positive reference value, say $0.1 \leq \zeta \leq 0.25$, should be used. Then, if a large shift occurs, the OOC ARL will be much the same as that from a larger reference value. Overall, it is clear that smaller reference values, rather than larger ones, are to be preferred. However, reference values too close to 0 should be avoided.

Figure 4.1: OOC ARL estimates of the MSR CUSUM at four ζs and a changepoint of 250 for normal, t_3 and skew-normal ($\alpha = 4$) data.



If some in-control Phase I data $V_1, \ldots, V_m$ are available, then we can estimate θ_1 by the methodology in Section 3.1.3. Observe that θ_1 in (4.5) can be expressed as

$$\theta_1 = 2\frac{1}{\eta}\,\mathrm{E}[\psi(F(V))\psi'(F(V))Vf(V)].$$

Then a consistent estimator of θ_1 is

$$\widehat{\theta_1} = \frac{2}{m\eta}\sum_{j=1}^{m}\frac{1}{\eta}\psi\left(\frac{j}{m+1}\right)\psi'\left(\frac{j}{m+1}\right)V_{m:j}\widehat{f}(V_{m:j})$$

where $\widehat{f}$ is the Gaussian kernel estimator in (3.7) and where $V_{m:j}$ is the j^{th} order statistic of $V_1, \ldots, V_m$, $1 \le j \le m$. Then, the estimated reference value is $\widehat{\zeta} = \widehat{\theta_1}(\log\rho)/2$. Again, because the Phase I and II data are independent and the CUSUM is distribution free, there exists a known control limit h, given $\widehat{\zeta}$, that will guarantee a Phase II IC ARL equal to the nominal value.

Just as in the case of the SRL CUSUM, Heuristic 4.1 provides us with a useful method of estimating the OOC behaviour of a CUSUM prior to implementation. Suppose we fix a reference value ζ and a control limit h to guarantee an ARL_0. Denote by $\mathcal{M}(\sigma_1)$ the OOC ARL of the CUSUM (either MSR or KSR) when a scale shift of size σ_1 occurs and by $\mathcal{N}(\theta_1 \log\sigma_1)$ the OOC ARL of a standard normal CUSUM with the same ζ and h when there are mean shifts of size $(\log\sigma_1)\theta_1\tau\log\left(\frac{\tau+i}{\tau+i-1}\right)$, $i \ge 1$. Then, Heuristic 4.1 says that, for "small" ζ and "large" τ,

$$\mathcal{M}(\sigma_1) \approx \mathcal{N}(\theta_1 \log\sigma_1). \tag{4.6}$$

Since there exist currently no explicit expressions or analytic approximations to either $\mathcal{N}(\theta_1 \log\sigma_1)$ or $\mathcal{M}(\sigma_1)$, we again use Monte Carlo simulation to estimate these quantities using only normal random numbers.

To show that the approximation (4.6) may be useful in designing the CUSUM, we generate Table 4.5, an analog of Table 3.6. Observe that the approximation fares well when the distribution is symmetric or moderately skew, but is less useful when the distribution is heavily skewed. To obtain the OOC ARLs in the table, we use Monte Carlo simulation (10 000 trials per design). We apply the upper MSR CUSUM to data from four standardised distributions – the normal and t_3, which are symmetric, and the Gumbel and skew-normal(4), which are asymmetric, all standardised to unit standard deviation. We induce a scale shift σ_1 = 1.10, 1.25, 1.375, 1.50, 1.75 or 2.00 at the changepoint $\tau = 250$. The control limits (to guarantee an ARL_0 of 500) come from linear interpolation in Table 4.1. The triple $(\rho;\ \zeta;\ h)$ denotes the target shift size, the reference value and the control limit. As in Table 3.6, and in some of the other tables to follow, the entries in boldface denote the OOC ARL approximations at shifts $\sigma_1 \ge \sigma$.

Table 4.5: OOC ARL approximations for (4.6) if $\mathcal{M}$ is based on the Mood score.

Distribution	θ_1	(ρ; ζ; h)	Approx.	σ_1					
				1.1	1.25	1.375	1.50	1.75	2.00
normal	1.10	(1.25; 0.12; 9.77)	$\mathcal{M}(\sigma_1)$	201	**62**	**36**	**26**	**18**	**14**
			$\mathcal{N}(\theta_1 \log \sigma_1)$	162	**61**	**37**	**26**	**17**	**13**
		(1.50; 0.22; 7.11)	$\mathcal{M}(\sigma_1)$	231	71	37	**26**	**17**	**13**
			$\mathcal{N}(\theta_1 \log \sigma_1)$	186	71	40	**27**	**16**	**12**
t_3	0.89	(1.25; 0.10; 10.57)	$\mathcal{M}(\sigma_1)$	235	**86**	**50**	**35**	**23**	**18**
			$\mathcal{N}(\theta_1 \log \sigma_1)$	181	**80**	**49**	**35**	**23**	**18**
		(1.50; 0.18; 7.97)	$\mathcal{M}(\sigma_1)$	263	105	56	**37**	**22**	**17**
			$\mathcal{N}(\theta_1 \log \sigma_1)$	195	91	54	**37**	**22**	**16**
Gumbel	1.01	(1.25; 0.11; 10.17)	$\mathcal{M}(\sigma_1)$	216	**72**	**42**	**30**	**20**	**16**
			$\mathcal{N}(\theta_1 \log \sigma_1)$	173	**72**	**43**	**31**	**20**	**15**
		(1.50; 0.20; 7.50)	$\mathcal{M}(\sigma_1)$	251	84	45	**31**	**19**	**14**
			$\mathcal{N}(\theta_1 \log \sigma_1)$	191	83	47	**32**	**19**	**14**
SN(4)	0.99	(1.25; 0.11; 10.17)	$\mathcal{M}(\sigma_1)$	175	**52**	**31**	**23**	**16**	**13**
			$\mathcal{N}(\theta_1 \log \sigma_1)$	173	**72**	**43**	**31**	**20**	**15**
		(1.50; 0.20; 7.50)	$\mathcal{M}(\sigma_1)$	194	56	31	**22**	**15**	**12**
			$\mathcal{N}(\theta_1 \log \sigma_1)$	191	83	47	**32**	**19**	**14**

The above table was generated on the assumption that θ_1 is known, which is rarely the case in practice. When using an estimate $\widehat{\theta_1}$, we can use $\mathcal{N}(\widehat{\theta_1} \log \sigma_1)$ as an approximation to $\mathcal{M}(\sigma_1)$. Then it should be kept in mind that there may be some degeneration in the quality of the approximation. The construction of an estimate $\widehat{\theta_1}$ was discussed above.

4.1.4 Comparing the KSR CUSUM with the NSS CUSUM

Suppose that data arise from a $N(\mu, \sigma^2)$ distribution where both μ and σ are unknown. Since both procedures are scale and location invariant, we may assume that $\mu = 0$ and $\sigma = 1$ without loss of generality. However, the construction of the CUSUMs treat μ and σ as if they were unknown. The NSS standard deviation CUSUM can be used to detect a scale shift (see Section 2.3.2). However, it should be kept in mind that the NSS CUSUM suffers from between-practitioner variation in that the IC ARL conditional upon the estimates based on the initial observations, is not guaranteed to equal the nominal IC ARL. Currently there exists no parametric CUSUM which does not suffer from this problem. Notwithstanding this, we show here a comparison between the unconditional ARLs of the NSS CUSUM and the ARL of the KSR CUSUM (which has zero between-practitioner variation). From Table 4.3 we saw that the correlation between the Klotz score and the efficient score in a normal distribution is 1.00. This suggests that, when the data are in fact normal, we can expect the KSR CUSUM to perform well.

We generate data from a $N(0, 1)$ distribution and introduce various scale shifts σ_1 at three changepoints $\tau = 50$, 100 and 250. The target shift sizes are $\rho = 1.25$ and 1.5. The nominal IC ARL is set at 500. The NSS CUSUM is initiated after 3 observations, these being used to obtain initial estimates of μ and σ^2. The reference values and control limits for the KSR CUSUM come from Tables 4.4 and 4.2. For the NSS CUSUM, the reference value is given in (2.11) and the control limits come from anygeth.exe of Hawkins and Olwell (1998, p.229-230). Table 4.6 shows OOC ARLs of the two upper CUSUMs at the range of shifts indicated in the first column. Throughout small scale shifts σ_1, the KSR performs equally well or better than the NSS CUSUM and for large τ ($\tau > 100$) it performs as well as the NSS CUSUM over the full range of σ_1 values. It is only at the small changepoint $\tau = 50$ where the KSR CUSUM is clearly not competitive with the NSS CUSUM. This is explained by the fact that the conditional IC ARL of the NSS CUSUM at $\tau = 50$ is consistently less than the nominal 500.

Table 4.6: OOC ARL comparison of the KSR and NSS CUSUMs for normal data.

		$\tau = 50$				$\tau = 100$				$\tau = 250$			
		KSR		NSS		KSR		NSS		KSR		NSS	
ρ		1.25	1.50	1.25	1.50	1.25	1.50	1.25	1.50	1.25	1.50	1.25	1.50
	1.10	321	350	315	348	**250**	**274**	259	289	**168**	**191**	201	229
	1.25	166	214	140	174	88	116	76	103	**50**	**59**	52	67
σ_1	1.50	50	77	32	42	26	32	21	23	20	20	18	19
	1.75	23	31	15	15	16	16	12	12	13	12	11	11
	2.00	16	17	10	9	12	12	9	9	10	9	8	8
ζ		0.22	0.41	1.24	1.46								
h		14.20	10.96	15.44	12.17								

4.2 The sequential rank CUSUM for survival data

Survival data is often modeled by an exponential-type distribution, such as the Weibull distribution. The densities of these distributions typically have modes at or close to zero and decrease monotonically toward the right tail and are parameterised to have a scale parameter only.

The following sequential rank scale CUSUM can be applied in such situations. If we transform $\{X_1, \dots, X_\tau, \sigma X_{\tau+1}, \dots\}$ to

$$\{\log X_1, \dots, \log X_\tau, \log \sigma X_{\tau+1}, \dots\} = \{\log X_1, \dots, \log X_\tau, \log \sigma + \log X_{\tau+1}, \dots\},$$

the scale shift in X translates to a location shift in $\log X$. Because the log function is one-to-one and monotone increasing, the $\log X$ sequence has the same sequential ranks as the X sequence. Thus, the SRL CUSUM can be applied (see Section 3.2) with

$$\xi_i = \sqrt{\frac{12(i+1)}{i-1}} \left(\frac{r_i}{i+1} - \frac{1}{2} \right). \tag{4.7}$$

In the special case where the X_i values come from an exponential-type, such as the Weibull distribution, the Savage score is optimal when testing for a scale shift (Hájek et al., 1999, p.106). The summand of the Savage score is

$$\begin{aligned}\xi_i = \sum_{j=i-r_i+1}^{i} \frac{1}{j} &\approx \log\left(\frac{i}{i-r_i+1}\right) \\ &\approx -\log\left(1-\frac{r_i}{1+i}\right) \\ &\approx \frac{r_i}{1+i},\end{aligned}$$

which, when standardised, is the same as (4.7).

The GSS CUSUM (see Section 2.3.3) is the only such parametric CUSUM available that can be applied to such data. However, the GSS CUSUM also suffers from in-control between-practitioner variation.

The HD CUSUM can also be applied directly. We will now compare for an exponential distribution the OOC ARLs of the SRL and the HD CUSUM. We specify an ARL_0 of 500 and the target shift sizes as $\log\rho = \log 1.5$ and $\log 2.0$ in the SRL CUSUM. We induce various shifts σ_1 after the changepoints $\tau = 50$ or $\tau = 200$ and we show the OOC ARL estimates (from 20 000 Monte Carlo trials) in Table 4.7. The control limits and reference values for the SRL CUSUM are shown in the last two rows of the table. Clearly, the SRL CUSUM performs better than the HD, especially when the true shift σ_1 is small.

Table 4.7: OOC ARL estimates of the HD and SRL CUSUMs for standardised exponential data.

	$\tau = 50$			$\tau = 200$		
σ_1	$\text{SRL}_{\log 1.5}$	$\text{SRL}_{\log 2.0}$	HD	$\text{SRL}_{\log 1.5}$	$\text{SRL}_{\log 2.0}$	HD
1.10	391	412	475	294	329	425
1.25	251	297	398	139	176	238
1.50	118	158	235	48	62	72
1.75	54	76	115	28	30	36
2.00	30	45	55	21	20	24
ζ	0.22	0.38				
h	7.899	5.309				

5 Girschick-Rubin (GR) CUSUMs based on sequential ranks

In this chapter, we construct distribution-free GR CUSUMs for a location and a scale shift along the same lines as the CUSUMs we constructed in Chapters 3 and 4. We again distinguish between the scenarios set out in Chapter 1, namely detecting a shift when (I) the in-control median is specified and (II) when it is unspecified. The GR CUSUMs are defined as in Section 2.4. To avoid confusing the GR CUSUMs in this chapter with the CUSUMs that we introduced in Chapters 3 and 4, we will refer to those in Chapters 3 and 4 as Page-type CUSUMs.

5.1 The signed sequential rank (SSR) GR CUSUM

5.1.1 In-control properties

Suppose $X_1, X_2, \ldots, X_\tau$ are symmetrically distributed around zero and that $X_{\tau+1}, X_{\tau+2}, \ldots$ each have the same distribution as $X_1 + \mu$. Our aim is to detect this median shift as soon as possible after the changepoint τ. We make no assumption regarding the numerical value of the in-control scale parameter σ of X, except that it remains constant.

An appropriate choice of ξ_i is the SSR summand (3.1) for application in the GR CUSUM (2.14), replacing the X_i with ξ_i. The proposed SSR GR CUSUM recursion is $D_0 = 0$ and

$$D_i = (1 + D_{i-1}) \exp\left(2\zeta \left\{\xi_i - \zeta\right\}\right) \tag{5.1}$$

where ζ is a reference value.

The control limits for the Wilcoxon and Van der Waerden SSR GR CUSUMs, based on the scores (3.3) and (3.4), respectively, are given in Tables 5.1 and 5.2, respectively. These control limits were generated by Monte Carlo simulation and cubic spline interpolation in the manner set out in Section 3.2.1.1. However, here we choose the standard normal GR CUSUM control limits (from (2.16)) as starting values in the algorithm. These control limits are approximately the same as those of the standard normal GR CUSUM for $\zeta \leq 0.25$ and a large enough nominal IC ARL.

Table 5.1: Control limits for the Wilcoxon SSR GR CUSUM.

	IC ARL						
ζ	100	200	300	400	500	1000	2000
0.05	94.340	188.680	283.020	377.860	471.700	940.655	1893.367
0.10	89.000	178.510	270.891	356.020	446.020	896.559	1778.575
0.15	83.970	170.351	251.920	339.934	425.357	838.649	1675.962
0.20	79.230	158.460	237.690	316.920	395.956	792.953	1596.642
0.25	74.760	149.520	224.550	299.050	373.600	724.589	1431.821
0.375	62.950	125.890	184.044	238.265	298.568	573.107	1085.053
0.50	51.702	97.749	141.514	189.194	227.826	417.194	800.985

Table 5.2: Control limits for the Van der Waerden SSR GR CUSUM.

	IC ARL						
ζ	100	200	300	400	500	1000	2000
0.05	94.416	190.806	282.670	378.195	474.576	935.923	1876.796
0.10	89.488	175.766	267.354	354.032	445.081	884.219	1774.917
0.15	83.140	167.845	253.443	335.063	421.524	844.982	1670.371
0.20	79.667	160.263	240.673	317.766	395.560	788.146	1594.134
0.25	75.427	150.978	224.917	302.088	373.034	744.495	1490.629
0.375	63.991	128.590	189.882	254.517	318.599	639.878	1283.644
0.50	56.283	108.704	161.695	218.773	273.193	546.388	1489.709

5.1.2 Out-of-control behaviour

We will show in this section that the SSR GR CUSUM exhibits similar out-of-control behaviour as the SSR Page-type CUSUM (see Section 3.1), namely that the SSR GR CUSUM behaves out of control approximately as would a standard normal GR CUSUM.

Define $S_n = \sum_{i=1}^{n} \xi_i$. The SSR GR CUSUM written in the form (2.13) is

$$D_i = \sum_{j=0}^{i-1} \exp\left(\mu(S_i - S_j) - \frac{(i-j)\mu^2}{2}\right).$$

Using the Heuristic 3.1, we infer the following heuristic.

Heuristic 5.1. *Let ζ be "small" and let a persistent shift of "small" size μ_1 occur at a "large" changepoint τ. Then, the SSR GR CUSUM behaves approximately as would a normal GR CUSUM with the same μ_1, ζ and h when a shift of size $\mu_1\theta$ commences after τ.* ■

5.1.3 Design of the GR CUSUM

Suppose we target a median shift of size μ. Then, for a fixed i and a "large" τ we have shown in (3.6) that $\text{E}[\xi_{\tau+j}] \approx \mu\theta$ where θ is defined in (3.5).

Heuristic 5.1 says that, for a "small" ζ and μ_1 and a "large" τ the SSR GR CUSUM should be approximated well by a normal GR CUSUM, that is, (3.21) should hold where $\mathcal{W}(\mu_1)$ denotes the OOC ARL of an SSR GR CUSUM at a true shift μ_1 and where $\mathcal{N}(\mu_1\theta)$ denotes the OOC ARL of the corresponding standard normal GR CUSUM. In order to assess the usefulness of this approximation, Table 5 (shown in Appendix B) was generated using the Wilcoxon SSR GR CUSUM. The results in the table indicate that the approximation is indeed very well suited for designing the CUSUM, even at small shifts μ_1 (in contrast to the approximation for the Page-type SSR CUSUM, which is useful only for $\mu_1 \geq \mu$). Recall that to obtain an estimate $\widehat{\theta}$ of θ, the methodology in Section 3.1.3 can be followed. Then we can use $\mathcal{N}(\mu_1\widehat{\theta})$ as an approximation to $\mathcal{W}(\mu_1)$.

5.1.4 Comparison with the SSR CUSUM

In this section we compare the SSR GR and Page-type (see Section 3.1) CUSUMs. We choose the Wilcoxon score because of its high overall correlation with the efficient scores in a range of t_ν distributions which, therefore, serves usefully as an omnibus score. Here we have two CUSUMs that are precisely comparable in that neither one has between-practitioner variation and both are based on the same assumptions. Table 5.3 shows estimated OOC ARLs (from 20 000 independent Monte Carlo trials per case) of both CUSUMs for data from a standardised normal ($\theta = 0.98$) and t_3 ($\theta = 1.38$) distribution shifted by an amount μ_1 after the changepoints $\tau = 50$ or $\tau = 250$. The target shifts are $\mu = 0.25, 0.5$ and 1.0 and we use $\zeta = \mu\theta/2$ as the reference value (see Table 3.5). In the table we abbreviate the Page-type CUSUM by "SSR" and the GR CUSUM by "GR". The pair of target and reference value (μ; ζ) and the corresponding control limit that guarantees an upper ARL_0 of 500 are shown in the bottom rows of each table.

Clearly, the GR CUSUM performs much better than the Page-type CUSUM when the true shift μ_1 is small, while there is little or no difference between the OOC ARLs at large shifts μ_1. Therefore, we can conclude that the GR has much to recommend it over the Page-type CUSUM. This corresponds to the conclusion that Moustakides et al. (2009) arrived at when they compared the standard normal GR with the standard normal Page (1954) CUSUM, namely that the only significant difference in performance is at small shifts μ_1.

Table 5.3: OOC ARL comparison of the SSR GR CUSUM with the SSR Page-type CUSUM.

normal data												
	$\tau = 50$						$\tau = 250$					
μ_1	SSR	GR	SSR	GR	SSR	GR	SSR	GR	SSR	GR	SSR	GR
0.10	161	**149**	195	**176**	242	**230**	163	**142**	190	**175**	242	**228**
0.25	**59**	**59**	70	**63**	100	**89**	58	**55**	69	**63**	96	**87**
0.50	26	28	**25**	**25**	32	**30**	26	25	25	**24**	31	**29**
0.75	17	19	15	16	16	**15**	17	17	15	15	**15**	**15**
1.00	13	15	11	12	11	**10**	13	13	11	11	10	**7**
(μ; ζ)	(0.25; 0.1225)		(0.50; 0.245)		(1.00; 0.49)							
h	10.92	433.60	7.35	374.13	4.20	232.90						

t_3 data												
	$\tau = 50$						$\tau = 250$					
μ_1	SSR	GR	SSR	GR	SSR	GR	SSR	GR	SSR	GR	SSR	GR
0.10	127	**114**	164	**150**	234	**213**	124	**113**	162	**147**	227	**209**
0.25	40	**38**	49	**44**	85	**72**	40	**37**	48	**43**	79	**68**
0.50	17	18	17	**16**	23	**20**	17	18	**16**	**16**	21	**19**
0.75	12	13	10	11	11	**10**	12	12	10	10	**10**	**10**
1.00	10	10	8	8	**7**	**7**	10	10	8	8	**7**	**7**
(μ; ζ)	(0.25; 0.1725)		(0.50; 0.345)		(1.00; 0.69)							
h	9.16	409.94	5.73	316.99	2.84	129.91						

5.1.5 An efficient self-starting GR CUSUM for a normal distribution

Suppose that data $X_1, X_2, \ldots, X_\tau$ are $N(0,1)$ quantities and $X_{\tau+1}, X_{\tau+2}, \ldots$ are $N(\mu_1, 1)$ quantities. Then the standard normal GR CUSUM can be applied. The VdW SSR GR CUSUM would be the "optimal" sequential rank CUSUM in this situation. However, the SSR GR CUSUM cannot use the information that the standard deviation is one, since it it scale invariant. Nevertheless, we can expect the VdW SSR CUSUM to perform well in this situation. The following Monte Carlo simulation results tend to bear out this conclusion. We estimate OOC ARLs (10 000 Monte Carlo trials) for both upper CUSUMs with an ARL_0 of 500 and target shifts $\mu = 0.25$ and 0.5 at a range of mean shifts μ_1 and at two changepoints $\tau = 50$ and 250. The estimated OOC ARLs at the shifts shown in the first column are shown in Table 5.4 where the subscripts on the NGR_μ and VdW_μ denote the target shifts. Observe that in all instances it is true that the OOC ARLs of the two CUSUM are virtually equal. Overall, we conclude that the VdW SSR GR CUSUM is in no way inferior in its out-of-control performance to the normal GR CUSUM.

Next, suppose that the standard deviation σ of the normal distribution is unknown. Then the standard normal GR CUSUM cannot be applied, while the VdW SSR GR can still be applied since no assumption of the value of σ was necessary to implement it. The VdW SSR GR CUSUM is therefore the only existing GR CUSUM which can be applied successfully if it is known that the underlying distribution is normal. Furthermore, the in-control properties of the VdW CUSUM remain valid even if there is deviation from the normality assumption.

Table 5.4: OOC ARL comparison of the VdW SSR GR CUSUM with the normal GR CUSUM.

	$\tau = 50$				$\tau = 250$			
μ	$NGR_{0.25}$	$VdW_{0.25}$	$NGR_{0.5}$	$VdW_{0.5}$	$NGR_{0.25}$	$VdW_{0.25}$	$NGR_{0.5}$	$VdW_{0.5}$
0.10	146	147	173	173	140	141	176	176
0.25	58	59	61	62	53	53	60	60
0.50	27	27	23	24	25	25	23	24
0.75	17	18	14	15	16	17	14	14
1.00	13	14	10	11	12	12	10	11
ζ		0.125		0.25				
h	432.187	433.303	373.571	373.034				

5.2 The sequential rank location GR CUSUM

5.2.1 In-control properties

When the symmetry assumption is untenable or when the in-control median is unknown, a useful GR CUSUM can be constructed using the unsigned sequential ranks, based on the same assumptions as the SRL CUSUM in Section 3.2. The SRL summand ξ_i given in (3.9) can be applied to the recursion (5.1). Special cases of ξ_i are the Wilcoxon SRL summand given in (3.11) and the normal SRL summand given in (3.12). The resulting CUSUM is abbreviated to the SRL GR CUSUM. Because the SRL ξ_i summand has the same in-control asymptotic distribution as the SSR ξ_i summand, the control limits in Tables 5.1 and 5.2 apply to the Wilcoxon and the normal SRL GR CUSUMs, respectively.

5.2.2 Out-of-control behaviour

The Heuristic 3.2 now leads directly to the following.

Heuristic 5.2. *Let ζ be "small" and let a persistent shift of "small" size μ_1 occur at a "large" changepoint τ. Then, the SRL GR CUSUM behaves approximately as would a standard normal GR CUSUM for the mean with the same μ_1, ζ and h when shifts of size $\mu_1\theta\tau\log\left(\frac{n}{n-1}\right)$ commence after $n = \tau$.* ■

We can compare quantitatively the OOC behaviour of the SRL GR CUSUM with that of the Page-type SRL CUSUM (see Section 3.2) for different distributions in a similar way in which we conducted the comparison in Section 5.1.4. Some simulation results are shown in Table 7 of Appendix B to support the conclusion that there is a practical difference between the OOC ARLs of the two CUSUMs at small shifts μ_1. The indication is that the GR CUSUM shows better performance than the Page-type CUSUM.

5.2.3 Design of the GR CUSUM

Suppose we target a location shift of size μ. Then, $\mathrm{E}[\xi_{\tau+i}] \approx \mu\theta$ is non-zero for a "large" τ and a fixed i, but becomes zero if we fix τ and let $i \to \infty$ (see Section 3.2.2). The definition of θ is given in (3.18). Thus, just like the Page-type CUSUM, the SRL GR CUSUM will only increase for a while after a shift and will then revert back to a seemingly in-control state as it takes on the new median as the current value.

Heuristic 5.2 says that, for a "small" ζ and a "large" τ, the approximation (3.21) should hold where $\mathcal{W}(\mu_1)$ denotes the OOC ARL of a SRL GR CUSUM given a median shift size $\mu_1 > 0$ and where $\mathcal{N}(\mu_1\theta)$ denotes the OOC ARL of the GR CUSUM for a standard normal mean with the same ζ and h when shifts of size $\mu_1\theta\tau\log\left(\frac{\tau+i}{\tau+i-1}\right)$, $i \geq 1$ occur after the changepoint τ. To illustrate that this approximation indeed fares very well and would most certainly be useful for designing the CUSUM, we construct Table 6 (an analogue of Table 5) given in Appendix B. Again, in contrast to the Page-type CUSUM, the approximation seems to be useful also at small values of μ_1. Recall that an estimate $\widehat{\theta}$ of θ can be obtained by the methodology explained in Section 3.1.3. Then we can use $\mathcal{N}(\mu_1\widehat{\theta})$ as an approximation to $\mathcal{W}(\mu_1)$.

5.3 Sequential rank scale GR CUSUMs

In this section, we construct sequential rank GR CUSUMs to detect a scale shift. Let $X_1, \ldots, X_\tau$ have the density function $f(x-\mu)$, while $X_{\tau+1}, X_{\tau+2}, \ldots$ have the density function $f((x-\mu)/\sigma_1)/\sigma_1$ and where τ denotes the changepoint. We construct sequential rank GR CUSUMs for scenario II where μ is unknown. A consequence of the scale invariance of sequential ranks is that these GR CUSUMs cannot usefully incorporate any numerical information about the value of the in-control scale parameter σ. The sequential rank scale GR CUSUM recursion is given by (2.17) where we replace X_i with a sequential rank score function ξ_i. For the case where μ is known or when the distribution is concentrated on the positive axis, analogous constructions of the GR CUSUMs can be made. An application will be shown in Chapter 6.

5.3.1 The sequential rank scale GR CUSUM (median unknown)

GR CUSUMs for scale shifts can be constructed using the recursion (2.17) with X_i there replaced by sequential rank score functions ξ_i, such as (3.11) and (3.12). Tables 5.5 and 5.6 show control limits for the MSR and KSR GR CUSUMs, respectively.

Table 5.5: Control limits for the MSR GR CUSUM.

	Nominal IC ARL						
ρ	100	200	300	400	500	1000	2000
1.05	91.033	171.102	244.800	310.088	371.186	621.416	994.452
1.10	76.289	130.105	174.909	210.426	244.639	368.753	531.892
1.15	62.846	99.319	127.033	149.470	170.816	245.657	339.060
1.20	51.664	78.448	96.936	113.354	126.212	174.287	236.300
1.25	42.587	62.683	77.602	89.115	98.501	132.909	176.040
1.30	36.028	51.765	62.958	71.568	79.086	105.021	137.350
1.40	27.203	37.425	44.481	49.725	54.645	70.918	90.221
1.50	21.126	28.265	33.162	36.793	39.981	50.900	64.048
1.75	12.828	16.465	18.847	20.767	22.215	27.217	32.834

Table 5.6: Control limits for the KSR GR CUSUM.

	Nominal IC ARL						
ζ	100	200	300	400	500	1000	2000
1.05	92.723	184.174	271.080	360.106	449.637	898.639	1799.358
1.10	83.381	162.491	239.800	325.938	398.048	797.846	1581.211
1.15	75.092	146.142	217.388	286.714	356.182	714.673	1413.447
1.20	67.780	130.696	196.838	258.300	320.574	642.608	1287.126
1.25	61.064	118.163	173.981	229.121	286.125	574.951	1158.767
1.30	55.836	108.483	160.442	210.916	262.007	537.729	1055.231
1.40	47.028	89.005	132.914	176.885	217.634	441.020	906.368
1.50	40.030	76.701	112.904	150.689	187.838	378.919	769.071
1.75	28.907	53.603	79.874	105.633	130.735	258.573	538.108

We can compare the out-of-control behaviour of the MSR GR CUSUM with that of the Page-type CUSUM for different distributions similar to the comparison that we drew in Section 5.2.2. Some simulation results are shown in Table 8 of Appendix B. The conclusion is once again that the GR CUSUM performs better than the Page-type CUSUM in that the OOC ARLs at small shifts σ_1 are substantially smaller than those of the Page-type CUSUM.

6 Applications

We implement our sequential rank CUSUMs in three applications, namely monitoring shifts in the ash content of coal, monitoring the mutual consistency between replicates of laboratory measurements on coal ash; and monitoring the time intervals between successive coal mining disasters.

6.1 Monitoring ash content

Coal ash is a waste product which forms when coal is burned by industrial plants to produce electricity. The percentage ash present is an indication of the quality of the coal, because more ash means more waste product, and vice versa. We have available measurements $X_1, X_2, \ldots, X_{75}$ of the percentage ash per unit mass present in batches of coal. These observations were generated by an X-ray fluorescent (XRF) gauge that measures the amount of coal present in a batch of coal on a conveyor belt, which transports the coal from the mine to a blending site. These measurements are used by the blending site to create a blend of coal with an approximately constant ash content. It is a known fact that the XRF gauge cannot provide accurate readings when the mass of the of coal on the conveyor is too small. Then, the gauge is apt to generate either excessively large or small measurements. These events are known in SPC literature as transient special causes. Ideally, the SPC protocol should be robust against such causes in order to prevent frequent false signals.

In this application, the protocol is to monitor for shifts in the median ash level away from the current value. The effect of the transient causes mentioned above, is for the data to exhibit spurious outliers. Thus, use of a robust CUSUM is indicated.

The designs of the Page-type CUSUMs that we apply are as follows. We run the two-sided Wilcoxon SRL each with an individual ARL_0 of 2000 (an overall ARL_0 of 1000). The target median shift away from the current value is specified as $\mu = 1.00$ standard deviation and we choose $\theta = 1.00$, $\zeta = 0.5$ and $h = 5.34$, which should be appropriate for a distribution with tails that are moderately heavier than those of the normal distribution. If the CUSUM signals at $N = n$, we restart the CUSUM at observation $n + 1$. Table 6.1 shows the run lengths N, the changepoint estimates $\widehat{\tau}$ and the direction of the putative shift for these data. The changepoint estimate is the usual one, namely the last observation at which the CUSUM sequence was at zero prior to signalling (Page, 1954). The CUSUM plots are shown in the left panel of Figure 6.1

with the estimated changepoints indicated by the vertical dashed lines. Figure 6.2 shows the full data set together with a fitted robust Loess regression curve. It seems clear that a substantial shift in the median has occurred somewhere after observation 40 and that special causes were in effect prior to that. The changepoint at 30 detected by the Wilcoxon SRL CUSUM is no doubt due to the presence of these special causes. Observe from the Wilcoxon CUSUM sequence in Figure 6.1 that the sequence reverts back to the control limit after the signal. This shows clearly what we meant in previous chapters by "the CUSUM returns to what seems to be an in-control state after some time".

Table 6.1: Run lengths and changepoint estimates of the Wilcoxon SRL CUSUM and the direction of the putative shift.

	N	$\hat{\tau}$	Direction
Wilcoxon SRL CUSUM	40	30	upward
	51	43	downward

The right panel in Figure 6.1 shows an SRL CUSUM based on the efficient score in a Cauchy distribution (see Appendix A), namely

$$\xi_i = \sqrt{2}\sin\left(2\pi\left(\frac{r_i}{1+i} - \frac{1}{2}\right)\right).$$

With the same reference value $\zeta = 0.5$, a control limit guaranteeing a two-sided ARL_0 of 1000 is $h = 5.25$ (see Table 2 of Appendix A). This CUSUM signals only once at $N = 53$ and indicates a changepoint $\hat{\tau} = 41$. If a smaller reference value $\zeta = 0.25$ is used, there is again only one changepoint indicated at $\hat{\tau} = 41$. Thus, from an operational point of view, this CUSUM should be preferred when it is known, a priori, that the process is prone to the frequent occurrence of special transient causes.

Figure 6.1: In the left panel is the two-sided SRL Wilcoxon CUSUM and in the right panel the Cauchy CUSUM for the XRF ash data. The axes are the time index (horizontal) and the CUSUM sequence (vertical). The dashed horizontal barriers are the control limits and the dashed vertical lines are the changepoint estimates of the respective CUSUMs.

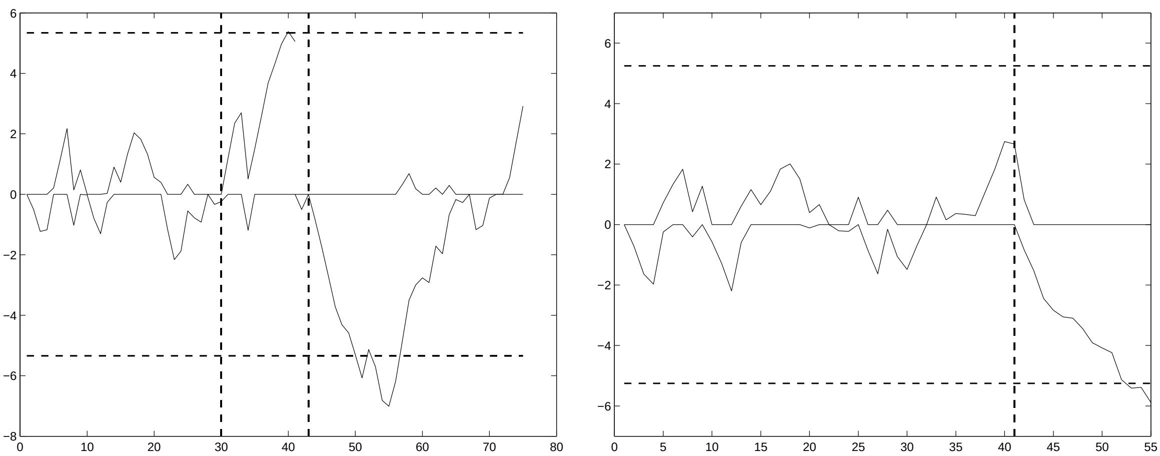
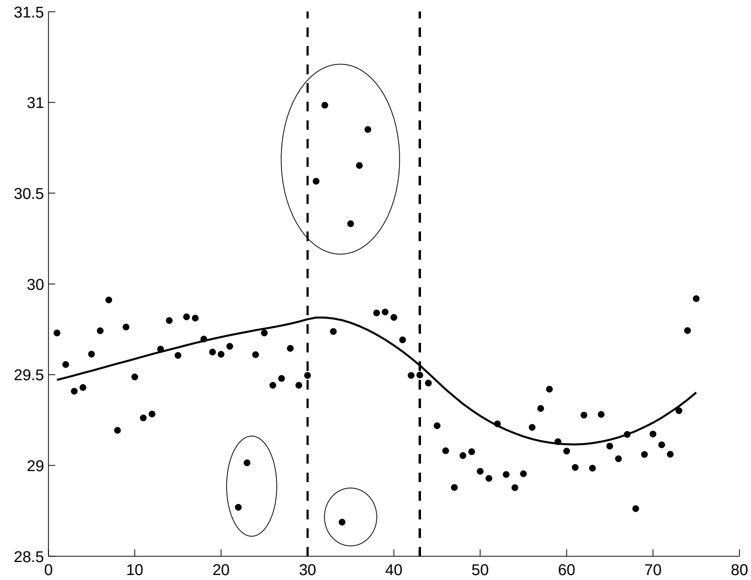


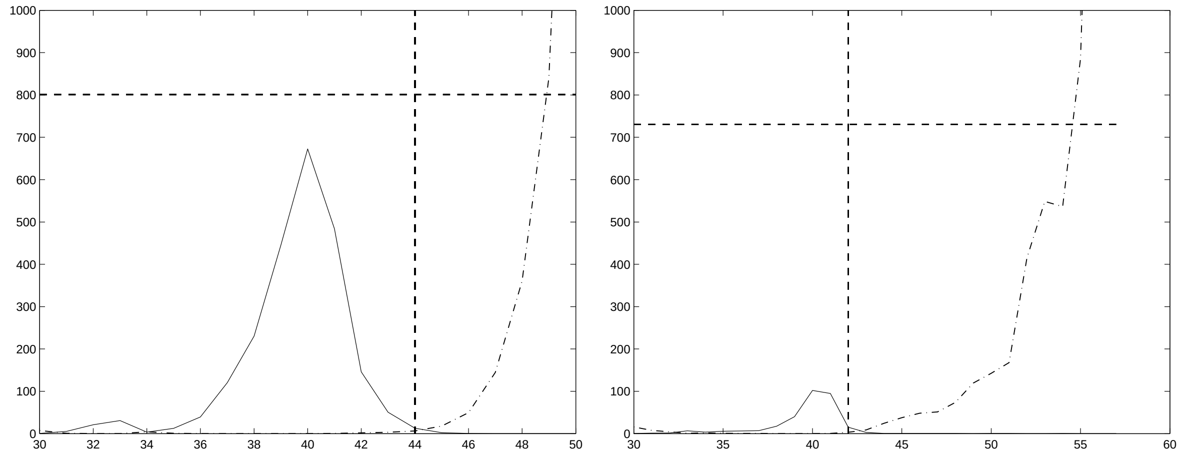
Figure 6.2: A time series plot of the XRF ash data $\{X_1, \ldots, X_{75}\}$ where the solid line is a Loess regression with smoothing span 0.8. The vertical dashed lines are the changepoint estimates from Table 6.1. The encircled observations are those caused by transient special causes.



We proceed to run the GR CUSUMs on the data to see how the results compare. We fix the individual ARL_0s at 2000 and the same target shift sizes as above. Then to detect a median shift we apply the two-sided Wilcoxon and Cauchy SRL GR CUSUMs, both using $\zeta = 0.50$, with respective control limits $h = 800.985$ and $h = 730.520$ from Table 5.1 and Table 3 of Appendix A. The Wilcoxon GR CUSUM signals a decrease at observation 49, while the Cauchy GR CUSUM signals a decrease at time 55. Both CUSUMs only signal once. The respective CUSUM sequences are shown in Figure 6.3. Since both CUSUMs are always positive, the standard changepoint estimate, namely the last time that the hitting CUSUM was at zero, is not applicable. Instead, as an ad hoc measure we propose to estimate the changepoint as the last index at which the hitting CUSUM is less than the non-hitting CUSUM. While this seems like a reasonable estimate, its statistical properties still need to be investigated. This, however, is outside the scope of this thesis. Using the proposed estimate, the Wilcoxon and Cauchy CUSUMs indicate $\widehat{\tau} = 44$ and $\widehat{\tau} = 42$.

The only substantial difference between the two analyses, is that the Wilcoxon GR CUSUM seems to be more robust against the effects of the transient special causes than the Wilcoxon Page-type CUSUM. This is certainly a feature that warrants further investigation.

Figure 6.3: In the left and right panels are the two-sided Wilcoxon and Cauchy GR CUSUMs, respectively, for the XRF data. The solid lines are the upward sequences, while the dash-dotted lines are the downward sequences. The horizontal dashed barriers are the control limits and the vertical dashed lines are the changepoint estimates. We only show the sequences after time 30 to prevent distorting the figures.



6.2 Monitoring inter-laboratory consistency

In this application we have available independent pairs of assay values $(V_{1,i}, V_{2,i}), i \geq 1$ which are the estimated percentages of the ash content of a batch of coal randomly split into two identical subsamples. Ideally, the V_1 and V_2 should be true replicates which would imply that the laboratories produce mutually consistent results. In particular, $\mathrm{E}[X_i] = 0$ where $X_i = V_{1,i} - V_{2,i}$. The focus here is on the variability of the X values since an increase in variability would imply that one or both of the laboratories are not following the prescribed laboratory analysis protocols.

Since well-calibrated laboratory equipment is typically involved, we would not expect either V_1 or V_2 to have distributions which are substantially non-normal. Thus, we choose $\theta_1 = 1$. To detect an increase in the scale of a target size $\rho = 1.5$ away from the current level, we choose $\zeta = 0.20$ from Table 4.4 with the corresponding control limit $h = 10.363$ which gives the upper MSR Page-type CUSUM a nominal ARL_0 of 2000. The left panel in Figure 6.4 shows the MSR CUSUM. This CUSUM signals at time $N = 217$ with a changepoint estimate $\widehat{\tau} = 129$. The right panel in Figure 6.4 shows the upper MSR GR CUSUM, which uses the recursion

$$D_i = (1 + D_{i-1})\rho_1^{-1} \exp\left(\frac{1}{2}\left(1 - \rho_1^2\right)\xi_i^2\right)$$

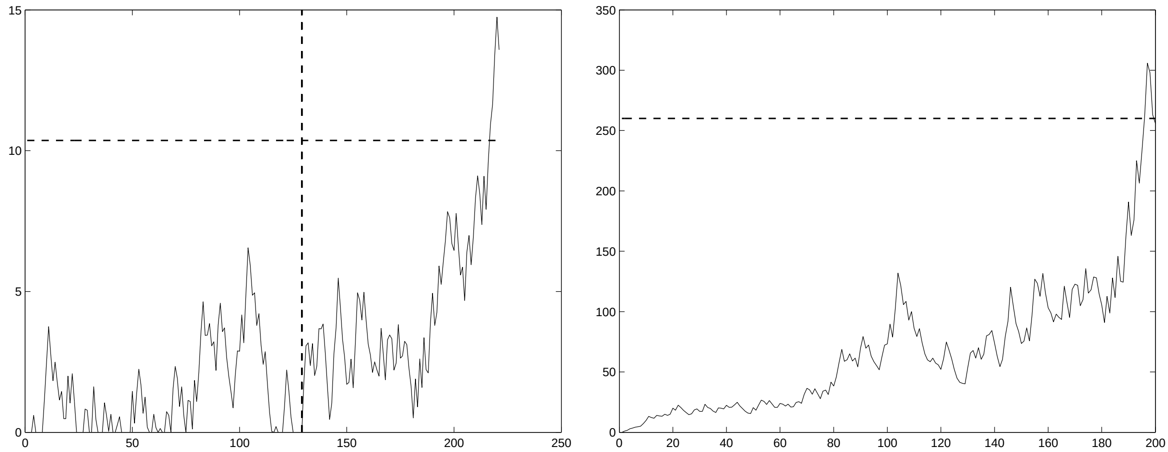
where

$$\xi_i^2 = 12\left(\frac{r_i}{1+i} - \frac{1}{2}\right)^2$$

and where $\rho_1^2 = \theta_1 \log \rho + 1 = 1.405$. The control limit that guarantees an ARL_0 of 2000 is $h = 260.05$. This GR CUSUM signals an increase at time $N = 196$, which is 21 observations earlier than the Page-type CUSUM.

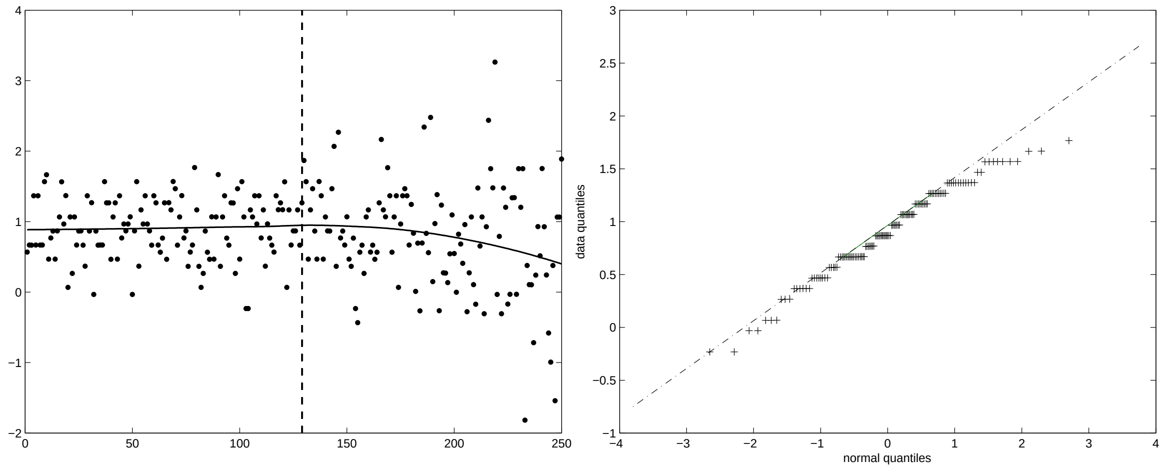
The changepoint estimate for the SRL GR CUSUM proposed in Section 6.1 is not applicable to the scale GR CUSUM. Nevertheless, looking at the CUSUM plot it would seem that the last substantial increase towards the control limit starts at around observation number 125, which is very close to the changepoint estimate from the MSR CUSUM. Obtaining a changepoint estimate for the scale GR CUSUM is a matter for further research.

Figure 6.4: The upper MSR Page-type (left) and GR (right) CUSUMs for the laboratory ash data against the time index. The horizontal dashed barriers are the control limits, while the vertical dashed line in the left panel is the changepoint estimate.



In retrospect, we plot a time series of the data $X_1, \ldots, X_{250}$ in the left panel of Figure 6.5 where we indicate the changepoint estimate $\widehat{\tau} = 129$ by a dashed line. We also plot a Loess regression on the data. From this plot it seems plausible that a scale shift occurred after the changepoint estimate. Also, we construct a normal Q-Q plot of the presumably in-control data $X_1, \ldots, X_{129}$ shown in the right panel of Figure 6.5. This indicates that the right tail of the distribution may be slightly lighter than that of a normal distribution. Evidently, our assumption that the data would not deviate too much from normality was sound.

Figure 6.5: A time series plot of the ash data $\{X_1, \ldots, X_{250}\}$ is shown in the left panel where the solid line is a Loess regression with smoothing span 0.8. The dashed line indicates the changepoint estimate from the MSR CUSUM. In the right panel is a normal Q-Q plot of $\{X_1, \ldots, X_{129}\}$.

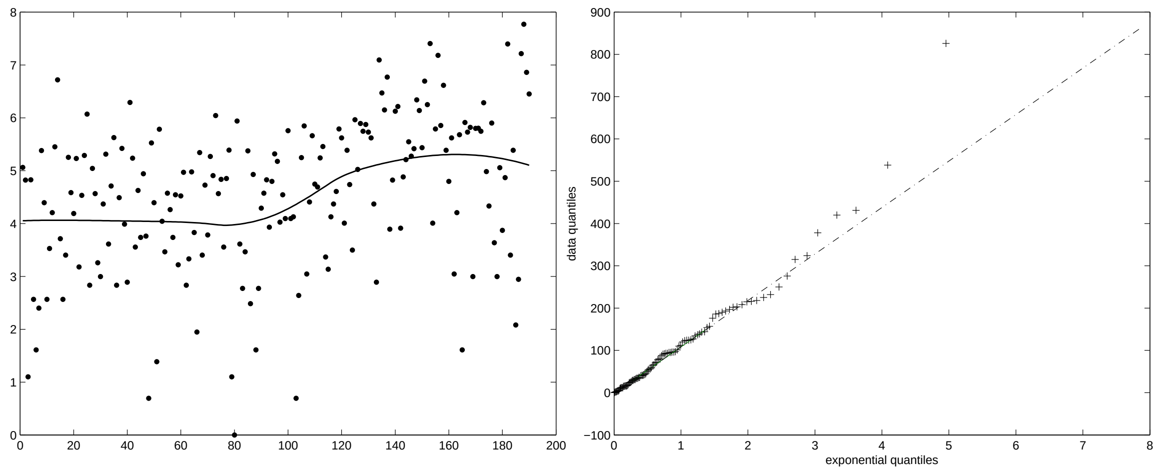


In some standards documents a requirement is that the standardised ratio $Z = X/Y$, where $Y = (V_1 + V_2)/2$, must serve as the measure of agreement between V_1 and V_2 to be monitored. If V_1 and V_2 indeed follow normal distributions, it is shown in Appendix C that the density function of Z exhibits, to a good approximation, tail behaviour similar to that of a t_3 distribution. In light of these applications, it is all the more surprising that control procedures for distributions with heavier tails have not been given proper attention.

6.3 Monitoring intervals between coal mining disasters

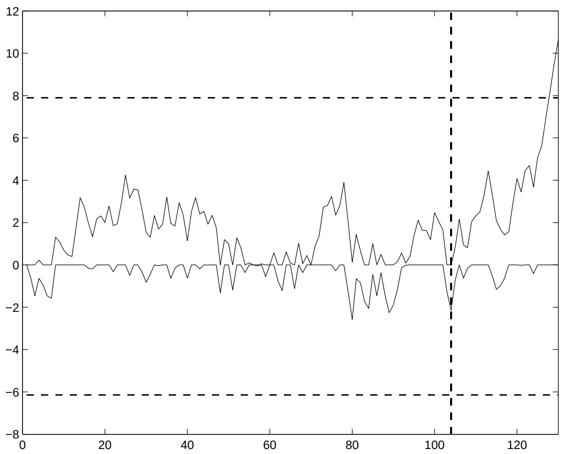
The data analysed here consist of independent measurements $V_1, \ldots, V_{190}$ of the time intervals (measured in days) between successive mining explosions in which 10 or more men were killed (Jarrett, 1979). The left panel in Figure 6.6 shows a scatter plot of the full data sequence $\log(V_i + 1)$ together with a Loess regression curve. From it we see that there seems to be an increase in the median from around time 100. In Figure 6.6 we also show an exponential Q-Q plot of the data $V_1, \ldots, V_{100}$ and see that the distribution of V seems to have a somewhat heavier tail than the exponential distribution. This could be a result of underestimation of the changepoint which would imply that some of the data in the Q-Q plot arose from the out-of-control distribution. In any case, a possible non-exponential distribution does not invalidate the CUSUM which is distribution free. Typically, it is of interest to monitor disaster data as a measure of the extent of control of safety procedures at coal mines. More dispersed time intervals between disasters indicate a lower frequency of accidents, while less dispersed time intervals indicate the contrary. It is now of interest to see what conclusions are reached if a CUSUM had been applied to these data from the outset.

Figure 6.6: A time series plot (left) of the mining disasters data $\log(V_i + 1)$, $i = 1, \ldots, 190$ where the solid line is a Loess regression with smoothing span 0.8. In the right panel is a standard exponential distribution Q-Q plot of the data $\{V_1, \ldots, V_{100}\}$.



We apply the CUSUMs in Section 4.2 to the $Y = \log V$ data in which case a shift in the rate of explosions would present itself as a shift in the median of the Y values. We assume that the tails of the distribution of the V are not substantially different from those of the exponential distribution. Therefore, we do not anticipate that the tails of the distribution of Y will be much heavier than those of a Gumbel distribution. Thus, we choose $\theta = 1.10$. Suppose that a scale shift in the V of 50% of the (unknown) current value is regarded as beneficial (increase) or detrimental (decrease). Then the reference values are $\zeta \approx 1.1\log(1.5)/2 = 0.22$ and $\zeta^{-} \approx -1.1\log(0.5)/2 = 0.38$. To guarantee an overall ARL_0 of 500, the control limits of the Wilcoxon SRL CUSUM are $h = 7.899$ and $h^{-} = 6.141$. The CUSUM signals an increase at observation $N = 128$ with $\widehat{\tau} = 104$, which seems to be valid when looking at the time series plot. We show these CUSUMs in Figure 6.7.

Figure 6.7: The two-sided Wilcoxon SRL CUSUM sequences for the coal mining disaster intervals, given an overall ARL_0 of 500. The vertical dashed line is the changepoint estimate, while the horizontal dashed barriers are the control limits.



Because the data do not arise with a high frequency and because there is little cost when a false alarm is signalled, we may feel that a smaller ARL_0 would be more suitable. We will choose instead an overall ARL_0 of 100 and reapply the CUSUM. The control limits corresponding to ζ and ζ^{-} given above are $h = 6.070$ and $h^{-} = 4.212$. Then the CUSUM signals an increase at time $N = 127$ with $\widehat{\tau} = 104$, leaving the results virtually unchanged.

7 Epilogue

We conclude this thesis with a summary of our main conclusions and we offer a few remarks concerning future research into the field of sequential rank CUSUMs. When referring to "CUSUMs" in the remainder of this section, we do not merely refer to the Page (1954) CUSUMs, but also to those of Girschick and Rubin (1952), unless stated otherwise.

This study had as its objective the construction of Page (1954) and Girschick and Rubin (1952) CUSUMs based on signed and unsigned sequential ranks to detect small shifts in location or scale. We expanded on work done by Van Zyl (2015) and Lombard and Van Zyl (2018) on signed sequential rank Page-type CUSUMs to a class of unsigned sequential rank CUSUMs. The former CUSUMs are tailor made for situations where the in-control median is known and specified, while the latter CUSUMs can be used whenever the in-control distribution is not necessarily symmetric or when the in-control median is unknown. In designing sequential rank scale CUSUMs, one should keep in mind that these CUSUMs cannot usefully incorporate any numerical information regarding the in-control scale parameter. This comes as a consequence of the scale invariant nature of sequential (and non-sequential) ranks. We also constructed sequential rank GR CUSUMs. We believe that this thesis makes a significant contribution to the literature on GR CUSUMs by providing an omnibus GR CUSUM design which may be used in distributions other than the normal. In particular, we showed that these GR CUSUMs have much to recommend them over the Page-type CUSUMs.

All of these sequential rank CUSUMs have the distinguishing features that they are distribution free, are fully self starting and have zero between-practitioner average run length variation. That is, no parametric specification of the underlying distribution is required a priori to find a known control limit that guarantees an in-control average run length equal to a specified nominal value. In particular, it was possible to create once-and-for-all tables of control limits for the various CUSUMs, which is quite satisfactory from a practical stance. The only specification required in the designs of these CUSUMs is a reference value. The only restriction is that the reference value cannot be excessively (and impractically) large, e.g. larger than $\sqrt{12}$ in the case of the unsigned sequential rank location CUSUM, to prevent the CUSUM from being and remaining identically equal to zero. One can also choose a reference value arbitrarily, rather than from a Phase I data sample; the in-control average run length will still be guaranteed. Indications were that minor misspecifications in the "optimal" reference value do not result in disastrously incorrect out-of-control average run lengths and, thus, an arbitrarily chosen reference value is warranted. Therefore, no Phase I sample is required for our CUSUMs to initiate. However, we provided some simulation results indicating that the use of a smaller rather than a larger reference value is preferred. However, the presence of a relatively small Phase I sample can aid the effective design of the CUSUM in tailoring its out-of-control behaviour to the specifications of the data, in which case the in-control average run length is still guaranteed to be equal to the nominal value. This comes as a result of the independence of the sequential ranks of the Phase I and the Phase II data. Finally, a table summarising the design, assumptions, strengths and limitations of all of our sequential rank CUSUMs are given in Table 7.1 at the end of this Epilogue.

We formulated heuristics for the out-of-control behaviour of these CUSUMs, the result being that the out-of-control average run length can be approximated by that of a standard normal CUSUM with an appropriately chosen drift. This result makes it possible to design a CUSUM prior to its implementation. The only CUSUM for which we did not formulate such a heuristic, is the sequential rank scale GR CUSUM. Furthermore, we showed that when the underlying distribution is normal with an unknown variance, our sequential rank CUSUMs using a Van der Waerden-type statistic can be applied, without it being in any way in its performance inferior to the standard normal CUSUM which assumes a known variance. In addition, using the sequential rank CUSUM in this setting completely removes the issue of between-practitioner variation which besets parametric CUSUMs. Furthermore, we indicated use of a Wilcoxon statistic in an omnibus version of the CUSUMs because of its high overall correlation with the efficient scores in various types of distributions. In one of our practical applications, we also used a CUSUM based on a Cauchy score, which we showed by means of an application to be useful when special transient causes are apt to occur frequently.

In addition, we illustrated that the CUSUMs using monotone scores, for example the Wilcoxon or Van der Waerden score, are not subject to the problem that larger shifts may be harder to detect than smaller shifts, as is the case in some parametric CUSUMs designed for heavy-tailed distributions. This is because the sequential rank score functions are monotone functions of the out-of-control shift parameter. Thus, our CUSUMs provide an ideal solution when it is known that the underlying distribution has rather heavy tails.

We compared quantitatively our CUSUMs with their competitors, such as the CUSUMs of Hawkins and Deng (2010) or Ross and Adams (2012) and found that generally the sequential rank CUSUMs performed similarly or better. In addition, the sequential rank CUSUMs involve use of a reference value, making the CUSUMs flexible for tuning to a specified target shift, which the CUSUMs of Hawkins and Deng (2010) or Ross and Adams (2012), in their present form, are unable to do. We were able to relate the out-of-control behaviour of the sequential rank CUSUMs to a specified target shift by using an appropriately chosen reference value. Furthermore, we compared the sequential rank GR CUSUMs to the sequential rank Page-type CUSUMs: not only did we find that, in general, the out-of-control average run lengths of the GR CUSUMs are smaller at small true shifts than those of the Page-type CUSUMs, but we also found that, for a given score function, they seem to be more robust against the effect that spurious outliers or special transient causes may have on the data. This is most certainly a feature that warrants some further research.

In the literature it is typically assumed that the CUSUMs have a unit Type I error, that is that they will signal when in control regardless of whether a shift actually occurred. This “fact” seems to be assumed purely on intuitive grounds. Providing rigorous proofs to substantiate this intuition is a topic for further research.

With regard to changepoint estimators, we offer the following remarks. We have applied the usual changepoint estimator given by Page (1954). There exists currently no generally accepted changepoint estimator for the GR CUSUM even for the case of a standard normal distribution. In our data analysis we saw that a plausible changepoint estimator for the location two-sided GR CUSUM is the last observation where the hitting CUSUM was below the non-hitting CUSUM. The statistical properties of this proposed estimator need further investigation. However, this ad hoc estimator would not be suited to either the one-sided location GR CUSUM or the scale GR CUSUM. Finding changepoint estimators for the class of sequential rank GR CUSUMs is an open problem.

An examiner has remarked on the possibility that “the same factors that cause a location shift would often lead to an increase in volatility as well.” One way of addressing this would be to run location and scale CUSUMs in tandem in all applications. However, some care would have to be exercised in stating the overall IC ARL of such a scheme since the two CUSUMs do not operate independently – it is unlikely that the overall IC ARL would be equal to one half of the nominal IC ARL of the two components. Furthermore, if a shift in location occurs, it is fairly certain that the scale CUSUM will also signal as a result of the shift. An after-the-fact analysis would then reveal the true nature of the shift.

Table 7.1: A summary of the features, assumptions, strengths and limitations of our various sequential rank CUSUMs.

Shift type	Location		Scale	
CUSUM name	SSR CUSUM or SSR GR CUSUM	SRL CUSUM or SRL GR CUSUM	SR scale CUSUM or SR scale GR CUSUM	For survival data
Reference	Page CUSUM: Section 3.1 on page 19.	Page CUSUM: Section 3.2 on page 23.	Page CUSUM: Section 4.1 on page 35.	Page CUSUM: Section 4.2 on page 42.
	Girshick-Rubin CUSUM: Section 5.1 on page 45.	Girshick-Rubin CUSUM: Section 5.2 on page 49.	Girshick-Rubin CUSUM: Section 5.3.1 on page 51.	Girshick-Rubin CUSUM: Analogous to the Page CUSUM.
Score	Equation (3.1): $\xi_i = s_i \psi\left(\frac{r_i^+}{1+i}\right) \Big/ \sqrt{\eta_i}$	Equation (3.9): $\xi_i = \left(\psi\left(\frac{r_i}{1+i}\right) - \bar{\psi}_i\right) \Big/ \sqrt{\eta_i}$	Equation (4.1): $\xi_i = \left(\psi\left(\frac{r_i}{1+i}\right) - \bar{\psi}_i\right)^2 \Big/ \eta_i - 1$	Equation (4.7): $\xi_i = \sqrt{\frac{12(i+1)}{i-1}} \left(\frac{r_i}{i+1} - \frac{1}{2}\right) \Big/ \sqrt{\eta_i}$
Assumptions	• Continuous distributions. • Symmetric in-control distribution. • Specified in-control median. • No assumption regarding in-control scale parameter. • Independence of observations.	• Continuous distributions. • No assumption regarding functional form of in-control distribution. • No assumption regarding in-control scale parameter. • Independence of observations.	• Continuous distributions. • No assumption regarding functional form of in-control distribution. • No assumption regarding in-control scale parameter. • Independence of observations.	• Continuous distributions. • No assumption regarding functional form of in-control distribution. • No assumption regarding in-control scale parameter. • Independence of observations.
Strengths	• Distribution free • Guaranteed IC ARL • OOC ARL can be calculated analytically and verified by simulation. • Fully self starting • Once-and-for-all table of control limits.	• Distribution free • Guaranteed IC ARL • OOC ARL can be calculated analytically and verified by simulation. • Fully self starting • Once-and-for-all table of control limits.	• Distribution free • Guaranteed IC ARL • OOC ARL can be calculated analytically and verified by simulation. • Fully self starting • Once-and-for-all table of control limits.	• Distribution free • Guaranteed IC ARL • OOC ARL can be calculated analytically and verified by simulation. • Fully self starting • Once-and-for-all table of control limits.
Limitations	• Cannot accommodate discrete distributions. • In-control median should be specified. • Limited range of detectable shifts. • Applies only to independent observations (zero serial correlation).	• • Cannot accommodate discrete distributions. • Limited range of detectable shifts. • Applies only to independent observations (zero serial correlation).	• Cannot accommodate discrete distributions. • Limited range of detectable shifts. • Applies only to independent observations (zero serial correlation). • In-control scale value cannot be specified.	• Cannot accommodate discrete distributions. • Limited range of detectable shifts. • Applies only to independent observations (zero serial correlation). • In-control scale value cannot be specified.

Appendices

A Cauchy CUSUMs

Let $X_1, \ldots, X_n$ be i.i.d random variables from the Cauchy distribution with density and distribution function

$$f(x) = \frac{1}{\pi(1+x^2)},$$

$$F(x) = \frac{1}{\pi}\tan^{-1}(x) + \frac{1}{2}.$$

We are interested in testing the hypothesis $H_0 : \mu = 0$ against the alternative $H_1 : \mu > 0$. Then the density of the X_i under the alternative is $f(x-\mu)$. The LMP test for an upward change in location – see for instance Rao (2002, p.455) – is based on the score statistic

$$\sum_{i=1}^{n} \frac{\partial}{\partial \mu} \log f(X_i - \mu) = \sum_{i=1}^{n} \frac{\frac{\partial}{\partial \mu} f(X_i - \mu)}{f(X_i - \mu)}$$

evaluated in the point $\mu = 0$ and where $f(x-\mu)$ is the density of X. Let U be a uniform random variable on $(0,1)$. Then we have

$$\begin{aligned}
\left.\frac{\frac{\partial}{\partial \mu} f(X-\mu)}{f(X-\mu)}\right|_{\mu=0} &= \frac{2X}{1+X^2} \overset{\mathcal{D}}{=} \frac{2F^{-1}(U)}{1+(F^{-1}(U))^2} \\
&= \frac{2\tan\left(\pi\left(U-\frac{1}{2}\right)\right)}{1+\tan^2\left(\pi\left(U-\frac{1}{2}\right)\right)} \\
&= \frac{2\sin\left(\pi\left(U-\frac{1}{2}\right)\right)/\cos\left(\pi\left(U-\frac{1}{2}\right)\right)}{1+\sin^2\left(\pi\left(U-\frac{1}{2}\right)\right)/\cos^2\left(\pi\left(U-\frac{1}{2}\right)\right)} \\
&= \frac{2\sin\left(\pi\left(U-\frac{1}{2}\right)\right)\cos\left(\pi\left(U-\frac{1}{2}\right)\right)}{\cos^2\left(\pi\left(U-\frac{1}{2}\right)\right)+\sin^2\left(\pi\left(U-\frac{1}{2}\right)\right)} \\
&= 2\sin\left(\pi\left(U-\frac{1}{2}\right)\right)\cos\left(\pi\left(U-\frac{1}{2}\right)\right) \\
&= \sin\left(2\pi\left(U-\frac{1}{2}\right)\right) \qquad (1)
\end{aligned}$$

Since the quantities $r_i/(1+i)$ converge in distribution to a $U(0,1)$ random variable, the sequential rank analogue of (1) is

$$\sin\left(2\pi\left(\frac{r_i}{1+i}-\frac{1}{2}\right)\right),$$

which results in the use of the statistics

$$\xi_i = \sqrt{2}\sin\left(2\pi\left(\frac{r_i}{1+i} - \frac{1}{2}\right)\right),$$

which are for $i \geq 2$ independent with zero mean and unit variance.

Tables 2 and 3 give control limits for the resulting Cauchy CUSUMs.

Table 2: Control limits for the Cauchy SRL CUSUM.

	Nominal IC ARL						
ζ	100	200	300	400	500	1000	2000
0.00	9.217	13.352	16.459	19.249	21.393	30.683	43.932
0.05	7.780	10.585	12.615	14.139	15.424	20.024	25.148
0.10	6.722	8.789	10.208	11.232	12.164	14.970	17.994
0.15	5.891	7.510	8.547	9.382	9.990	12.015	14.103
0.20	5.205	6.495	7.338	7.990	8.457	10.011	11.651
0.25	4.632	5.749	6.425	6.960	7.291	8.576	9.865
0.30	4.166	5.118	5.653	6.098	6.412	7.470	8.541
0.40	3.400	4.095	4.530	4.848	5.075	5.839	6.615
0.50	2.801	3.339	3.664	3.899	4.084	4.674	5.259

Table 3: Control limits for the Cauchy SRL GR CUSUM.

	Nominal IC ARL						
ζ	100	200	300	400	500	1000	2000
0.05	95.765	192.439	285.674	381.390	476.601	964.311	1913.501
0.10	93.132	183.361	275.425	367.908	452.787	898.857	1809.996
0.15	88.594	176.599	261.351	350.731	432.588	856.823	1691.013
0.20	84.564	165.496	249.283	319.383	409.970	798.850	1558.596
0.25	80.590	156.995	231.363	305.003	376.361	727.209	1435.899
0.30	75.430	149.052	214.924	279.291	350.728	674.588	1300.577
0.40	66.266	124.975	178.107	235.537	286.417	538.215	1032.902
0.50	55.733	100.700	141.267	181.302	219.092	399.776	731.185

The values of θ defined in (3.18) on page 28 corresponding to the Cauchy score were calculated for a range of distributions. These are shown in the θ row of Table 4. The correlation coefficients between the efficient score (3.13) in a location parameter family and the Cauchy score were also calculated for these distributions and are shown in the “Correlations” row of Table 4.

Table 4: Values of θ for the Cauchy SRL CUSUM and the correlation coefficients with the efficient scores (3.13).

	Distribution								
	normal	t_4	t_3	t_2	t_1	Gumbel	SN(±1)	SN(±2)	SN(±4)
θ	0.65	1.00	1.24	1.19	1.41	0.52	0.67	0.79	1.19
Correlations	0.66	0.84	0.88	0.94	1.00	0.47	0.60	0.63	0.62

B Supplementary tables for the GR CUSUMs

With reference to Section 5.1, we give Table 5 below; and with reference to Section 5.2, we give Tables 6 and 7 below.

Table 5: SSR GR CUSUM OOC ARL approximations if $\mathcal{W}$ is based on the Wilcoxon score.

Distribution	θ	(μ; ζ; h)	Approx.	μ_1					
				0.125	0.25	0.375	0.50	0.75	1.00
normal	0.98	(0.25; 0.12; 436.716)	$\mathcal{W}(\mu_1)$	116	**56**	**36**	**26**	**18**	**14**
			$\mathcal{N}(\mu_1\theta)$	114	**55**	**35**	**26**	**17**	**13**
		(0.50; 0.245; 371.901)	$\mathcal{W}(\mu_1)$	144	63	35	**24**	**15**	**11**
			$\mathcal{N}(\mu_1\theta)$	142	62	35	**24**	**15**	**11**
t_4	1.18	(0.25; 0.148; 422.202)	$\mathcal{W}(\mu_1)$	100	**45**	**28**	**21**	**14**	**11**
			$\mathcal{N}(\mu_1\theta)$	100	**44**	**27**	**20**	**13**	**10**
		(0.50; 0.295; 342.683)	$\mathcal{W}(\mu_1)$	126	50	28	**19**	**12**	**9**
			$\mathcal{N}(\mu_1\theta)$	126	51	28	**19**	**11**	**8**
t_3	1.38	(0.25; 0.173; 416.949)	$\mathcal{W}(\mu_1)$	87	**38**	**24**	**18**	**12**	**10**
			$\mathcal{N}(\mu_1\theta)$	88	**37**	**23**	**17**	**11**	**8**
		(0.50; 0.345; 321.306)	$\mathcal{W}(\mu_1)$	115	43	24	**16**	**10**	**8**
			$\mathcal{N}(\mu_1\theta)$	110	42	23	**15**	**9**	**7**

Table 6: SRL GR CUSUM OOC ARL approximations for (3.21) if $\mathcal{W}$ is based on the Wilcoxon score.

Distribution	θ	(μ; ζ; h)	Approx.	μ_1					
				0.125	0.25	0.375	0.50	0.75	1.00
normal	0.98	(0.25; 0.12; 436.716)	$\mathcal{W}(\mu_1)$	171	**69**	**40**	**28**	**18**	**14**
			$\mathcal{N}(\mu_1\theta)$	164	**70**	**40**	**28**	**18**	**13**
		(0.50; 0.245; 371.901)	$\mathcal{W}(\mu_1)$	215	88	43	**27**	**16**	**12**
			$\mathcal{N}(\mu_1\theta)$	219	85	43	**26**	**15**	**11**
t_3	1.38	(0.25; 0.173; 416.949)	$\mathcal{W}(\mu_1)$	130	**45**	**26**	**19**	**13**	**10**
			$\mathcal{N}(\mu_1\theta)$	164	**50**	**26**	**17**	**11**	**8**
		(0.50; 0.345; 321.306)	$\mathcal{W}(\mu)$	180	58	27	**17**	**11**	**8**
			$\mathcal{N}(\mu_1\theta)$	258	81	33	**18**	**10**	**7**
Gumbel	1.11	(0.25; 0.139; 429.563)	$\mathcal{W}(\mu_1)$	161	**61**	**34**	**24**	**16**	**12**
			$\mathcal{N}(\mu_1\theta)$	163	**62**	**33**	**23**	**15**	**11**
		(0.50; 0.278; 355.301)	$\mathcal{W}(\mu_1)$	218	87	40	**24**	**14**	**10**
			$\mathcal{N}(\mu_1\theta)$	216	79	37	**22**	**13**	**9**
SN(4)	1.05	(0.25; 0.131; 433.278)	$\mathcal{W}(\mu_1)$	165	**66**	**37**	**26**	**17**	**13**
			$\mathcal{N}(\mu_1\theta)$	167	**65**	**37**	**25**	**16**	**12**
		(0.50; 0.263; 361.778)	$\mathcal{W}(\mu_1)$	229	90	41	**26**	**15**	**11**
			$\mathcal{N}(\mu_1\theta)$	216	84	39	**24**	**14**	**10**

Table 7: OOC ARL comparison of the SRL Page-type CUSUM with the SRL GR CUSUM.

normal data												
	$\tau = 50$						$\tau = 250$					
μ_1	SRL	GR	SRL	GR	SRL	GR	SRL	GR	SRL	GR	SRL	GR
0.10	348	**323**	383	**364**	411	**403**	236	**202**	287	**264**	332	**317**
0.25	185	**162**	235	**209**	295	**281**	78	**70**	102	**87**	158	**145**
0.50	52	**48**	69	**58**	129	**104**	27	**27**	29	**27**	40	**36**
0.75	22	24	22	**20**	44	**33**	17	18	16	**16**	17	**16**
1.00	15	17	13	**13**	14	**13**	13	14	11	**11**	11	**10**
(μ; ζ)	(0.25; 0.1225)		(0.50; 0.245)		(1.00; 0.490)							
h	10.92	433.60	7.35	374.13	4.20	232.90						

t_3 data												
	$\tau = 50$						$\tau = 250$					
μ_1	SRL	GR	SRL	GR	SRL	GR	SRL	GR	SRL	GR	SRL	GR
0.10	324	**254**	365	**320**	422	**377**	199	**131**	251	**180**	310	**263**
0.25	131	**110**	190	**167**	279	**265**	49	**45**	68	**60**	132	**116**
0.50	25	25	37	**32**	100	**81**	18	18	17	**17**	25	**23**
0.75	14	14	13	12	26	**20**	12	12	10	11	11	**10**
1.00	11	11	9	9	9	**9**	9	10	8	8	7	**7**
(μ; ζ)	(0.25; 0.173)		(0.50; 0.345)		(1.00; 0.690)							
h	9.16	416.95	5.73	321.31	2.84	131.70						

skew-normal(4) data												
	$\tau = 50$						$\tau = 250$					
μ_1	SRL	GR	SRL	GR	SRL	GR	SRL	GR	SRL	GR	SRL	GR
0.10	352	**323**	399	**379**	429	**413**	240	**205**	300	**264**	358	**343**
0.25	190	**167**	253	**219**	334	**307**	77	**66**	113	**92**	189	**160**
0.50	50	**46**	76	**61**	167	**132**	26	**26**	28	**26**	51	**41**
0.75	20	22	22	**20**	58	**44**	16	17	15	**15**	18	**16**
1.00	14	16	12	13	18	**15**	12	13	11	**11**	10	**10**
(μ; ζ)	(0.25; 0.132)		(0.50; 0.263)		(1.00; 0.525)							
h	10.58	435.86	7.03	367.07	4.93	214.26						

Table 8 below shows a comparison between the MSR Page-type and GR CUSUMs from Section 5.3.

Table 8: OOC ARL comparison of the MSR Page-type CUSUM with the MSR GR CUSUM.

normal data												
	$\tau = 50$						$\tau = 250$					
ρ	1.25		1.50		2.00		1.25		1.50		2.00	
σ_1	MSR	GR	MSR	GR	MSR	GR	MSR	GR	MSR	GR	MSR	GR
1.10	336	306	362	290	363	320	209	138	239	165	253	197
1.25	161	158	195	139	217	156	64	54	76	57	94	62
1.50	46	68	61	47	82	47	26	28	26	26	28	26
1.75	24	42	25	27	31	24	18	20	17	18	17	17
2.00	14	33	17	21	18	17	14	17	13	14	12	12
ζ	0.123		0.223		0.381							
ρ_1^2		1.245		1.446		1.762						
h	9.766	215.382	7.109	125.305	4.950	71.253						

t_3 data												
	$\tau = 50$						$\tau = 250$					
ρ	1.25		1.50		2.00		1.25		1.50		2.00	
σ_1	MSR	GR	MSR	GR	MSR	GR	MSR	GR	MSR	GR	MSR	GR
1.10	347	342	365	329	381	341	238	159	266	192	298	228
1.25	208	222	236	190	271	202	87	69	105	75	133	86
1.50	78	114	95	82	128	80	35	35	36	34	42	35
1.75	38	72	42	44	63	39	23	24	22	23	23	23
2.00	25	54	25	31	32	25	18	20	17	17	16	16
ζ	0.099		0.180		0.308							
ρ_1^2		1.199		1.361		1.617						
h	10.594	253.792	7.929	154.608	5.725	89.591						

skew-normal($\alpha = 4$) data												
	$\tau = 50$						$\tau = 250$					
ρ	1.25		1.50		2.00		1.25		1.50		2.00	
σ_1	MSR	GR	MSR	GR	MSR	GR	MSR	GR	MSR	GR	MSR	GR
1.10	292	292	322	271	346	291	172	121	198	142	222	167
1.25	119	146	149	114	185	125	50	47	56	48	68	51
1.50	34	62	40	40	55	36	23	25	22	23	22	22
1.75	20	41	20	24	22	20	16	18	15	17	14	15
2.00	15	32	14	18	14	15	13	15	12	15	11	12
ζ	0.109		0.199		0.340							
ρ_1^2		1.221		1.401		1.686						
h	10.122	233.836	7.500	138.320	5.375	79.700						

The control limits, obtained by Monte Carlo simulation, for the normal GR CUSUM from Section 2.4 to detect a standard deviation shift are given in Table 9.

Table 9: Control limits for the normal GR CUSUM.

	Nominal IC ARL						
ρ	100	200	300	400	500	1000	2000
1.05	91.20	181.29	275.90	366.82	458.58	926.05	1839.47
1.10	84.41	168.43	252.46	339.44	425.66	854.41	1716.75
1.15	77.14	157.56	236.64	316.37	399.55	780.13	1581.68
1.20	72.75	143.14	217.14	285.02	364.84	724.40	1471.88
1.25	68.24	136.75	203.49	273.53	341.81	676.91	1370.00
1.30	62.63	126.53	191.04	248.45	321.32	637.66	1284.16
1.40	54.90	112.57	169.43	223.86	280.92	549.38	1111.16
1.50	49.49	99.18	150.23	199.84	249.89	500.65	1006.45
0.75	38.87	76.88	115.74	143.04	192.13	389.46	767.06
0.98	96.84	193.00	290.62	390.85	485.10	974.85	1966.23
0.90	94.84	190.82	285.79	382.49	476.40	951.20	1905.31
0.85	92.11	187.02	282.55	374.83	472.09	945.59	1882.31
0.80	91.68	182.62	271.33	363.39	464.28	907.58	1809.25
0.75	87.99	178.54	267.64	355.08	447.56	901.00	1798.33
0.70	86.33	174.45	260.13	348.25	433.01	878.79	1755.48
0.60	80.46	165.26	252.28	321.96	410.83	826.60	1676.67
0.50	75.27	154.98	229.64	307.44	376.46	757.77	1517.37
0.25	59.49	117.18	180.40	241.28	293.47	601.14	1184.60

C Computations for the density of the ratio of two independent normal random variables

Lemma C.1. *Let V_1 and V_2 be two independent $N(\mu, \sigma^2)$ random variables and set*

$$X = V_1 - V_2 \text{ and } Y = (V_1 + V_2)/2.$$

If μ/σ is large, then the tails of the probability density function of $Z = X/Y$ are proportional to the tails of the density of a t distribution with 3 degrees of freedom.

Proof. Observe that $V_1 - V_2$ follows a $N(0, 2\sigma^2)$ distribution and that $V_1 + V_2$ follows a $N(2\mu, 2\sigma^2)$ distribution. We want to find the density function of

$$Z = 2\frac{V_1 - V_2}{V_1 + V_2} \overset{\mathcal{D}}{=} 2\frac{N(0,1)}{N\left(\sqrt{2}\mu/\sigma, 1\right)} := 2\frac{X_1}{X_2}.$$

Set $b = \sqrt{2}\mu/\sigma$. From the independence of V_1 and V_2 follows the independence of X_1 and X_2 and then

$$\mathrm{P}\left(\frac{X_1}{X_2} \in du | X_2 = a\right) = \mathrm{P}(X_1 \in d(ay)) = \frac{|a|}{\sqrt{2\pi}} e^{-\frac{1}{2}a^2y^2} dy$$

and

$$\mathrm{P}(X_2 \in da) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(a-b)^2} da.$$

Then

$$\begin{aligned}
\mathrm{P}\left(\frac{X_1}{X_2} \in dy\right) &= \int_{-\infty}^{\infty} \frac{|a|}{\sqrt{2\pi}} e^{-\frac{1}{2}a^2y^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(a-b)^2} da \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} |a| \exp\left[\frac{a^2y^2}{2} - \frac{(a-b)^2}{2}\right] da \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} |a| \exp\left[-\frac{a^2y^2}{2} - \frac{a^2}{2} - \frac{b^2}{2} + ab\right] da \\
&= \frac{1}{2\pi} e^{-\frac{b^2}{2}} \int_{-\infty}^{\infty} |a| \exp\left[-\frac{a^2y^2 - 2ab + a^2}{2}\right] da \\
&= \frac{1}{2\pi} e^{-\frac{b^2}{2}} \int_{-\infty}^{\infty} |a| \exp\left[-\frac{a^2(1+y^2) - 2ab + b^2/(1+y^2)}{2} + \frac{b^2/(1+y^2)}{2}\right] da \\
&= \frac{1}{2\pi} \exp\left[-\frac{b^2}{2} + \frac{b^2/(1+y^2)}{2}\right] \int_{-\infty}^{\infty} |a| \exp\left[-\frac{\left(a\sqrt{1+y^2} - b/\sqrt{1+y^2}\right)^2}{2}\right] da \\
&= \frac{1}{2\pi} \exp\left[-\frac{b^2}{2} + \frac{b^2/(1+y^2)}{2}\right] \int_{-\infty}^{\infty} |a| \exp\left[-\frac{\left(a - b/(1+y^2)\right)^2 (1+y^2)}{2}\right] da \\
&= \frac{1}{\sqrt{2\pi}\sqrt{1+y^2}} \exp\left[-\frac{b^2}{2} + \frac{b^2/(1+y^2)}{2}\right] \times \int_{-\infty}^{\infty} \frac{|a|\sqrt{1+y^2}}{\sqrt{2\pi}} \exp\left[\frac{(a - b^2/(1+y^2))^2(1+y^2)}{2}\right] da \\
&= \frac{1}{\sqrt{2\pi}\sqrt{1+y^2}} \exp\left[-\frac{b^2}{2} + \frac{(b^2/(1+y^2))}{2}\right] \mathrm{E}\left[\left|N\left(\frac{b}{(1+y^2)}; \frac{1}{(1+y^2)}\right)\right|\right] \\
&= \frac{1}{\sqrt{2\pi}(1+y^2)} \exp\left[-\frac{b^2}{2} + \frac{b^2/(1+y^2)}{2}\right] \mathrm{E}\left[\left|N\left(\frac{b}{(1+y^2)^{1/2}}; 1\right)\right|\right].
\end{aligned}$$

Set $v = b/\sqrt{1+y^2}$. We proceed to evaluate $\mathrm{E}\left[|N(v;1)|\right]$ in the following calculations.

$$\begin{aligned}
\mathrm{E}\left[|N(v;1)|\right] &= \frac{1}{\sqrt{2\pi}} \int_0^{\infty} a\left(e^{\frac{-(a+v)^2}{2}} + e^{\frac{-(a-v)^2}{2}}\right) da \\
&= \frac{1}{\sqrt{2\pi}} \left(\int_0^{\infty} a e^{-\frac{(a+v)^2}{2}} da + \int_0^{\infty} a e^{\frac{-(a-v)^2}{2}} da\right) \\
&= \frac{1}{\sqrt{2\pi}} \left(\int_0^{\infty} (a+v) e^{-\frac{(a+v)^2}{2}} da - \int_0^{\infty} v e^{-\frac{(a+v)^2}{2}} da + \int_0^{\infty} (a-v) e^{\frac{-(a-v)^2}{2}} da - \int_0^{\infty} v e^{\frac{-(a-v)^2}{2}} da\right) \\
&= \frac{1}{\sqrt{2\pi}} \Big(2\exp(-v^2/2) - v\sqrt{2\pi}\,\mathrm{P}(N(-v;1) > 0) + v\sqrt{2\pi}\,\mathrm{P}(N(v;1) > 0)\Big) \\
&= \sqrt{\frac{2}{\pi}} \exp(-v^2/2) + v\left(\mathrm{P}(N(v;1) > 0) - \mathrm{P}(N(-v;1) > 0)\right) \\
&= \sqrt{\frac{2}{\pi}} \exp(-v^2/2) + v\left(1 - 2\,\mathrm{P}(N(v;1) < 0)\right) \\
&= \sqrt{\frac{2}{\pi}} \exp(-v^2/2) + v\left(1 - 2\Phi(-v)\right).
\end{aligned}$$

By substituting v back to $b/\sqrt{1+y^2}$, we obtain

$$\mathrm{E}\left[\left|N\left(\frac{b}{\sqrt{1+y^2}};1\right)\right|\right] = \sqrt{\frac{2}{\pi}}\exp\left(-\frac{b^2}{1+y^2}/2\right) + \frac{b}{\sqrt{1+y^2}}\left(1-2\Phi\left(-\frac{b}{\sqrt{1+y^2}}\right)\right).$$

Then, combining these calculations we find that

$$\begin{aligned} f_{X_1/X_2}(y) &= \frac{1}{\sqrt{2\pi}(1+y^2)}\exp\left(-\frac{b^2y^2}{2(1+y^2)}\right)\times\left\{\sqrt{\frac{2}{\pi}}\exp\left(-\frac{b^2}{2(1+y^2)}\right)+\frac{b}{\sqrt{1+y^2}}\left(1-2\Phi\left(-\frac{b}{\sqrt{1+y^2}}\right)\right)\right\} \\ &= \frac{1}{\pi(1+y^2)}\exp\left(-\frac{b^2}{2}\right)+\frac{b}{\sqrt{2\pi}(1+y^2)^{3/2}}\exp\left(-\frac{b^2y^2}{2(1+y^2)}\right)\left(1-2\Phi\left(-\frac{b}{\sqrt{1+y^2}}\right)\right). \end{aligned}$$

It is probable that the last expression is already available somewhere in the literature. However, we have been unable to find an appropriate reference.

Note that the first term on the right-hand side of this last equation is proportional to the density of a Cauchy distribution, which implies that the ratio X_1/X_2 has no moments. In practice, of course, $(V_1+V_2)/2$ cannot be exactly normally distributed because ash values must always be positive. In fact, in practical situations the ratio $b=\mu/\sigma$ is typically very large so that the first term is negligible for practical purposes. Thus, the density function f_{X_1/X_2} is approximately equal to

$$f_{X_1/X_2}(y) \approx C\frac{b}{\sqrt{2\pi}(1+y^2)^{3/2}}\exp\left(-\frac{b^2y^2}{2(1+y^2)}\right)\left(1-2\Phi\left(-\frac{b}{\sqrt{1+y^2}}\right)\right), \qquad (2)$$

where C is a constant. A Taylor expansion for large $|y|$ now gives $y^2/(1+y^2)\approx 1$ and

$$1-2\Phi\left(-\frac{b}{\sqrt{1+y^2}}\right) \approx 1-2\left(\Phi(0)-\frac{b}{\sqrt{1+y^2}}\phi(0)\right) = \sqrt{\frac{2}{\pi}}\frac{b}{\sqrt{1+y^2}}.$$

Then the right-hand side of (2) is, thus, approximately equal to

$$C\frac{b}{\sqrt{2\pi}(1+y^2)^{3/2}}\exp\left(-\frac{b^2}{2}\right)\sqrt{\frac{2}{\pi}}\frac{b}{\sqrt{1+y^2}} = C\frac{b^2}{\pi(1+y^2)^{(3+1)/2}}\exp\left(-\frac{b^2}{2}\right)$$

which is proportional to the density in the tail of a t_3 distribution. ∎

Bibliography

Azzalini, A. (2005). The skew-normal distribution and related multivariate families, *Scandinavian Journal of Statistics* **32**(2): 159–188.

Bakir, S. T. (2001). Classification of distribution-free quality control charts, *Proceedings of the Annual Meeting of the American Statistical Association, August 5-9, 2001*, pp. 5–9.

Bakir, S. T. (2006). Distribution-free quality control charts based on signed-rank-like statistics, *Communications in Statistics - Theory and Methods* **35**(4): 743–757.

Bakir, S. T. and Reynolds, M. R. (1979). A nonparametric procedure for process control based on within-group ranking, *Technometrics* **21**(2): 175–183.

Barndorff-Nielsen, O. E. (1963). On the limit behaviour of extreme order statistics, *The Annals of Mathematical Statistics* **34**(3): 992–1002.

Basseville, M. and Nikiforov, I. V. (1993). *Detection of abrupt changes: theory and application*, Vol. 104, Englewood Cliffs: Prentice Hall.

Bhattacharya, P. K. and Frierson, D. (1981). A nonparametric control chart for detecting small disorders, *The Annals of Statistics* **9**(3): 544–554.

Brook, D. A. E. D. and Evans, D. A. (1972). An approach to the probability distribution of CUSUM run length, *Biometrika* **59**(3): 539–549.

Chatterjee, S. and Qiu, P. (2009). Distribution-free cumulative sum control charts using bootstrap-based control limits, *The Annals of Applied Statistics* **3**(1): 349–369.

Coleman, S. Y., Arunakumar, G., Foldvary, F. and Feltham, R. (2001). SPC as a tool for creating a successful business measurement framework, *Journal of Applied Statistics* **28**(3-4): 325–334.

Gan, F. F. (1992). Exact run length distribution for one-sided exponential CUSUM schemes, *Statistica Sinica* **2**(1): 297–312.

Gandy, A. and Kvaløy, J. (2013). Guaranteed conditional performance of control charts via bootstrap methods, *Scandinavian Journal of Statistics: Theory and Applications* **40**(4): 647–668.

Girschick, M. A. and Rubin, H. (1952). A Bayes approach to a quality control model, *The Annals of Mathematical Statistics* **23**(1): 114–125.

Golosnoy, V. and Schmid, W. (2007). EWMA control charts for monitoring optimal portfolio weights, *Sequential Analysis: Design Methods and Applications* **26**(2): 195–224.

Hájek, J., Šidák, Z. and Sen, P. K. (1999). *Theory of Rank Tests*, 2nd edn, San Diego: Academic Press.

Harrison, P. J. and Lai, I. C. H. (1999). Statistical process control and model monitoring, *Journal of Applied Statistics* **26**(2): 273–292.

Hawkins, D. M. (1987). Self-starting CUSUM charts for location and scale, *Journal of the Royal Statistical Society - Series D (The Statistician)* **36**(4): 299–316.

Hawkins, D. M. (1992). Evaluation of average run length of cumulative sum charts for an arbitrary data distribution, *Communications in Statistics - Simulation and Computation* **21**(4): 1001–1020.

Hawkins, D. M. and Deng, Q. (2010). A nonparametric change-point control chart, *Journal of Quality Technology* **42**(2): 165–173.

Hawkins, D. M. and Olwell, D. H. (1998). *Cumulative Sum Charts and Charting for Quality Improvement*, New York: Springer.

Hawkins, D. M., Olwell, D. H. and Wang, B. (2016). http://cran.r-project.org/web/packages/cusumdesign/cusumdesign.pdf.

Jarrett, R. (1979). A note on the intervals between coal-mining disasters, *Biometrika* **66**(1): 191–193.

Kahya, E. and Theodossiou, P. (1999). Predicting corporate financial distress: A time-series CUSUM methodology, *Review of Quantitative Finance and Accounting* **13**(4): 323–345.

Keefe, M. J., Woodall, W. H. and Jones-Farmer, L. A. (2015). The conditional in-control performance of self-starting control charts, *Quality Engineering* **27**(4): 488–499.

Khmaladze, E. V. (2011). Sequential ranks, *International Encyclopedia of Statistical Science*, New York: Springer, pp. 1308–1311.

Knoth, S. (2016). http://cran.r-project.org/web/packages/spc/spc.pdf.

Lam, K. and Yam, H. C. (1997). CUSUM techniques for technical trading in financial markets, *Financial Engineering and the Japanese Markets* **4**(3): 257–274.

Lazariv, T., Schmid, W. and Zabolotska, S. (2013). On control charts for monitoring the variance of a time series, *Journal of Statistical Planning and Inference* **143**(9): 1512–1526.

Ledolter, J. and Kardon, R. (2012). Detection of the onset of multiple sclerosis from SD-OCT thickness measurements of the retinal nerve fibre layer, *Quality Engineering* **25**(1): 3–10.

Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation*, 2nd edn, New York: Springer.

Liu, G., Pu, X., Wang, L. and Xiang, D. (2015). CUSUM chart for detecting range shifts when monotonicity of likelihood ratio is invalid, *Journal of Applied Statistics* **42**(8): 1635–1644.

Liu, L., Tsung, F. and Zhang, J. (2014). Adaptive nonparametric CUSUM scheme for detecting unknown shifts in location, *International Journal of Production Research* **52**(6): 1592–1606.

Lombard, F. (1981). An invariance principle for sequential nonparametric test statistics under contiguous alternatives, *South African Statistical Journal* **15**(2): 129–152.

Lombard, F. (1983). Asymptotic distribution of rank statistics in the change-point problem, *South African Statistical Journal* **17**(1): 83–105.

Lombard, F. and Van Zyl, C. (2018). Signed sequential rank CUSUMs, *Computational Statistics and Data Analysis* **118**(1): 30–39.

MacEachern, S. N., Rao, Y. and Wu, C. (2007). A robust-likelihood cumulative sum chart, *Journal of the American Statistical Association* **102**(480): 1440–1447.

McDonald, D. (1990). A CUSUM procedure based on sequential ranks, *Naval Research Logistics* **37**(5): 627–646.

Moustakides, G. V., Polunchenko, A. S. and Tartakovsky, A. G. (2009). Numerical comparison of CUSUM and Shiryaev-Roberts procedures for detecting changes in distribution, *Communications in Statistics - Theory and Methods* **38**(16-17): 3225–3239.

Mukherjee, A. (2009). Some rank-based two-phase procedures in sequential monitoring of exchange rate, *Sequential Analysis* **28**(2): 137–162.

Page, E. S. (1954). Continuous inspection schemes, *Biometrika* **41**(1): 100–115.

Pettitt, A. N. (1979). A non-parametric approach to the change-point problem, *Applied Statistics* **28**(2): 126–135.

Pollak, M. (1987). Average run length of an optimal method of detecting a change in distribution, *The Annals of Statistics* **15**(2): 749–779.

Pollak, M. (2009). The Shiryaev-Roberts detection procedure in retrospect - theory and practice, *Proceedings of the 2nd International Workshop on Sequential Methodologies, University of Technology of Troyes, Troyes, France.*

Pollak, M. and Siegmund, D. (1991). Sequential detection of a change in a normal mean when the initial value is unknown, *The Annals of Statistics* **19**(1): 394–415.

Quesenberry, C. P. (1995). Geometric Q charts for high quality process, *Journal of Quality Technology* **27**(4): 304–315.

Rao, C. R. (2002). *Linear Statistical Inference and Its Applications*, 2nd edn, New York: Wiley.

Reynolds, M. R. (1975). A sequential signed-rank test for symmetry, *The Annals of Mathematical Statistics* **3**(2): 382–400.

Roberts, S. W. (1966). A comparison of some control chart procedures, *Technometrics* **8**(3): 411–430.

Ross, G. J. and Adams, N. M. (2012). Two nonparametric control charts for detecting arbitrary distribution changes, *Journal of Quality Technology* **44**(2): 102–116.

Saleh, N. A., Zwetsloot, I. M., Mahmoud, M. A. and Woodall, W. H. (2016). CUSUM charts with controlled conditional performance under estimated parameters, *Quality Engineering* **28**(4): 402–415.

Shanmugam, G., Clark, L. L., Burton, H. J., Warren, A. E. and O'Blenes, S. B. (2012). Improving and standardizing capture of pediatric cardiac surgical complications, *The Journal of Thoracic and Cardiovascular Surgery* **144**(3): 570–576.

Shiryaev, A. N. (1963). On optimum method in quickest detection problems, *Theory of Probability and its Applications* **8**(1): 22–46.

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

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*, Vol. 26, CRC Press.

Stoumbos, Z. G., Reynolds, M. R., Ryan, T. P. and Woodall, W. H. (2000). The state of statistical process control as we proceed into the 21st century, *Journal of the American Statistical Association* **95**(451): 992–998.

Timmer, D. H., Pignatiello, J. J. and Longnecker, M. T. (2001). Applying an AR(1) CUSUM control chart to data from a chemical process, *Quality Engineering* **13**(1): 107–114.

Van Dobben De Bruyn, C. (1968). *Cumulative Sum Tests: Theory and Practice*, London: Griffin.

Van Zyl, C. (2015). *CUSUM procedures based on sequential ranks*, Master's thesis, North-West University.

Woodall, W. H. (2006). The use of control charts in health-care and public-health surveillance, *Journal of Quality Technology* **38**(2): 98–104.

Yi, G., Coleman, S. and Ren, Q. (2006). CUSUM method in predicting regime shifts and its performance in different stock markets allowing for transaction fees, *Journal of Applied Statistics* **33**(7): 647–664.

Zhang, J., Zou, C. and Wang, Z. (2011a). An adaptive Shiryaev-Roberts procedure for monitoring dispersion, *Computers and Industrial Engineering* **61**(4): 1166–1172.

Zhang, J., Zou, C. and Wang, Z. (2011b). A new chart for detecting the process mean and variability, *Communications in Statistics - Simulation and Computation* **40**(5): 728–743.