South African Statistical Journal
2018, Vol. 52, No. 1, 93–113

SEQUENTIAL RANK CUSUMS FOR LOCATION AND DISPERSION

C. van Zyl
Unit for Business Mathematics and Informatics
North-West University, Potchefstroom, South Africa
e-mail: *vanzylcorli@gmail.com*

F. Lombard [1]
Unit for Business Mathematics and Informatics
North-West University, Potchefstroom, South Africa
e-mail: *fred.lombard@gmail.com*

We develop CUSUMs based on sequential ranks of the observations to detect changes over time in the location and dispersion of a distribution. The CUSUMs are distribution free in the sense that the appropriate control limits do not depend on the form or any parameters of the unknown underlying distribution. As such the CUSUMs are fully self starting. The in- and out-of-control average run length properties of the CUSUMs are gauged qualitatively via theory-based calculations and quantitatively by Monte Carlo simulation. The CUSUMS are shown to perform very well when compared to some existing parametric and nonparametric CUSUMS. Implementation of the CUSUMs is illustrated in an application based on real data from an industrial environment.

Key words: CUSUM, Distribution-free, Self starting, Sequential ranks.

1. Introduction

Let the sequence of independent observations $X_1, X_2, \ldots$ be the numerical values of a product quality characteristic measured over time. In the application treated in Section 4, for instance, the quality characteristic is the ash content of coal, measured in real time by an online x-ray fluorescent gauge. A change in product quality manifests itself as a change in the distribution of the X-values. Sequential CUSUM procedures – see, for instance, Hawkins and Olwell (1998) – are often used to detect such a change. However, their successful implementation generally requires specific assumptions about the form of the underlying distributions. Even minor misspecification of the functional form of the distribution and of the presumed known values of nuisance parameters can have a disastrous effect on their performance – see Hawkins and Olwell (1998, Sections 3.5 and 3.7.1), Keefe, Woodall and Jones-Farmer (2015) and Saleh, Zwetsloot, Mahmoud and Woodall (2016), where extensive further references can also be found.

Distribution-free CUSUMs can overcome such difficulties to a large extent. Here, the term “distribution-free” means that the in-control properties of the CUSUM do not depend upon the

[1] Corresponding author.
MSC2010 subject classifications. 62L10, 62G99.

parametric form of the underlying in-control distribution or on any of its parameters. In particular, replacing the data by their sequential ranks leads to distribution-free CUSUMs that are easy to implement. The i^{th} sequential rank, r_i, is the rank of X_i among the first i observations $X_1, \ldots, X_i$, that is

$$r_i = \sum_{j=1}^{i} \mathbf{1}(X_j \leq X_i),$$

where $\mathbf{1}(\cdot)$ denotes the indicator function. Successive sequential ranks are independently distributed and $r_i/(i+1)$, which is uniformly distributed on the set $\{1/(i+1), \ldots, i/(i+1)\}$, has mean $1/2$ as long as the underlying continuous distribution, whatever it may be, remains unchanged – see Barndorff-Nielsen (1963). When the underlying distribution changes, so does the mean of $r_i/(i+1)$ and a sequential rank CUSUM should then be able to detect the change. These facts about sequential ranks, together with their naturally sequential nature, make them attractive building blocks of distribution-free CUSUMs.

Bhattacharya and Frierson (1981) and Lombard (1983) developed distribution-free truncated sequential tests for location change based on sequential ranks. However, these are not CUSUM procedures in the commonly accepted sense of the term because they terminate after a predetermined number of observations. The first fully fledged CUSUM based on sequential ranks is due to McDonald (1990). His approach was based on the convergence to a uniform distribution of $r_i/(i+1)$. In essence McDonald (1990) develops a CUSUM to detect a change from a uniform to a non-uniform distribution. Our approach is more direct and relies on the approximately normal distribution of partial sums of functions $\psi(r_i/(i+1))$ of the sequential ranks, where ψ is a score function defined in Section 2. The introduction of functions of the sequential ranks also provides flexibility in the choice of the most appropriate CUSUM, a point illustrated in the application in Section 4 of the paper.

The main contributions of the present paper are (a) the development of sequential rank CUSUMS to detect location and scale changes in an underlying continuous distribution and (b) the development of a formula of sorts that allows one to gauge prior to implementation the likely OOC ARL (out-of control average run length) behaviour of the CUSUM. The latter feature can be especially useful in designing the CUSUM. Our sequential rank CUSUMs guarantee an IC ARL (in-control average run length) equal to the nominal value, regardless of the form of the continuous underlying distribution. Furthermore, the CUSUMs are devoid of the between-practioner effects noted by Keefe et al. (2015). These effects are present to a greater or lesser extent in all existing CUSUMs that rely for their successful Phase II implementation on parameter estimates made from Phase I data. Finally, since control limits for the sequential rank CUSUMs are not functionally related to the underlying data, they can be generated "once and for all time" by, for instance, Monte Carlo simulation using standard uniform random numbers. This is in contrast to some other non-parametric CUSUMS such as those proposed by Saleh et al. (2016), Gandy and Kvaløy (2013) and Chatterjee and Qiu (2009), which rely on bootstrapping from an in-control Phase I sample to obtain appropriate control limits. As the (unknown) distribution which produces the data changes from application to application, so will the control limits have to change. Furthermore, the bootstrap methods do not guarantee attainment of a nominal in-control ARL, but provide only lower or upper confidence limits for the actually attained ARL.

Two sets of circumstances must be distinguished clearly. In the first, the in-control value of a quality parameter of interest is specified. The original Page (1954) CUSUM was designed to handle such instances, assuming an underlying normal distribution. Generalizations to other distributions also took the in-control parameter value as known. Distribution-free CUSUMs applicable in this situation were developed by Bakir and Reynolds (1979) for observations appearing naturally in groups of size two or more, and by Lombard and Van Zyl (2018) for singly arriving observations. A second set of circumstances, which is the focus of the present paper, is described by Hawkins (1987). There, the process is declared to be in control upon startup, notwithstanding the fact that the current value of the parameter(s) of interest is either unknown or has been estimated from a Phase I sample. The objective is then to detect a change away from the true current value, whatever the latter may be. Such circumstances gave rise to the construction of, for instance, a self-starting CUSUM for monitoring the current mean of a normal distribution (Hawkins, 1987) and subsequent extensions (Hawkins and Olwell, 1998) to monitoring parameter values in distributions from the exponential family (Hawkins and Olwell, 1998, Chapter 7).

The present paper is structured as follows. In Section 2 the CUSUMs for detecting a change in the median are developed. A special case is considered in more detail and a table of control limits is provided. The OOC behaviour of the CUSUMs are also studied. A formula of sorts is provided that allows one to gauge the likely OOC performance prior to implementation. In Sections 2.5 and 2.6 the CUSUMS are compared to some competing parametric and distribution-free procedures. Section 3 provides information and results for monitoring dispersion in the underlying distribution. In Section 4 implementation of the CUSUMS is demonstrated in an application to XRF online monitoring of coal ash levels.

2. Sequential rank location CUSUMs

Sequential rank CUSUMs for detecting a change in the unknown median of a distribution can be constructed by standardizing the score functions that are typically used in two sample rank tests for location. These score functions $\psi(u)$, $0 < u < 1$, are usually symmetric around zero and satisfy the relations

$$\int_0^1 \psi(u)du = 0 \quad \text{and} \quad \int_0^1 \psi^2(u)du = 1.$$

Some well-known examples are the Wilcoxon score

$$\psi(u) = \sqrt{12}(u - 1/2) \tag{1}$$

and the Van der Waerden score

$$\psi(u) = \Phi^{-1}(u),$$

where Φ^{-1} denotes the inverse of the standard normal CDF. If $X_1, \ldots, X_i$ are i.i.d. (the in-control situation) the sequential rank r_i is uniformly distributed on the set $\{1, \ldots, i\}$, whence

$$E\left[\psi\left(\frac{r_i}{i+1}\right)\right] = \frac{1}{i}\sum_{j=1}^{i}\psi\left(\frac{j}{i+1}\right) = 0$$

and

$$var\left[\psi\left(\frac{r_i}{i+1}\right)\right] = \frac{1}{i}\sum_{j=1}^{i}\psi^2\left(\frac{j}{i+1}\right).$$

Writing

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

it then follows that the statistic

$$\xi_i = \psi\left(\frac{r_i}{i+1}\right)/\sqrt{\eta_i}$$

has zero mean and unit variance. In particular, for the Wilcoxon score (1), $\eta_i = (i-1)/(i+1)$ and

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

Furthermore, if $X_1, X_2, \ldots$ are independent with a fixed continuous distribution, the ξ_i are independently, but not identically, distributed and converge to a random variable with a uniform distribution on the interval $(-\sqrt{3}, \sqrt{3})$.

Since the mean of ξ_i becomes positive when the mean of the data increases, we define a CUSUM to detect an upward shift by setting $D_1^+ = 0$ and then computing recursively

$$D_1^+ = 0, \quad D_n^+ = \max\left\{0, D_{n-1}^+ + \xi_n - \zeta\right\} \tag{3}$$

for $n \geq 2$, where the reference value ζ is a positive constant. An alarm is raised as soon as D_n exceeds the control limit $h > 0$, that is, at the index

$$N^+ = \min\{n > 1 : D_n^+ \geq h\}, \tag{4}$$

which is known as the run length. An alarm is interpreted as a signal that the median of the distribution may has increased. The control limit h is chosen to ensure that the IC ARL equals a pre-specified finite value, ARL_0 – see Section 2.1. To detect a possible decrease in the median, a second CUSUM is run concurrently. This is defined by replacing all the plus signs in (3) and (4) by minus signs. The run length is then $N = \min\{N^+, N^-\}$ and the IC ARL is $ARL_0/2$. It is commonplace in applications to plot D_n^+ and $-D_n^-$ together on the vertical axis against n on the horizontal axis. In the sequel we will frequently use the CUSUM based upon the standardized Wilcoxon score and henceforth refer to it as the WSR (Wilcoxon sequential rank) CUSUM.

2.1 In-control properties

Given a reference constant ζ, the control limit h is chosen to ensure a pre-specified IC ARL. The distribution-free character of the CUSUMs when the X-process is in control allows fairly precise estimation via Monte Carlo simulation of h for any given IC ARL, reference constant ζ and score function ψ. Table 1 gives control limits h that guarantee a range of IC ARL values for the upper WSR CUSUMs D^+ at a range of reference constants. Details of the computations are given in Appendix A. Tables of control limits for sequential rank CUSUMs based on the Van der Waerden and Cauchy scores are available from the authors upon request.

The control limits in Table 1 at reference constants $\zeta \leq 0.25$ are quite close to those of a standard normal Page-type CUSUM (see Page, 1954). This is not entirely an unexpected result in view of the asymptotic normality of partial sums of the ξ_i. Regardless of the reference constant used, no

Table 1. Control limits for the WSR CUSUM.

	IC ARL						
ζ	100	200	300	400	500	1 000	2 000
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	5.25	5.96
0.50	2.73	3.31	3.68	3.93	4.13	4.74	5.34

parameter estimates are required to initiate the CUSUM. Given a reference constant ζ, any specified ARL_0 is guaranteed upon use of the appropriate h from Table 1. Thus, the CUSUM is fully self starting: no parameter estimates are required to initiate the CUSUM.

2.2 Out-of-control properties

An out-of-control situation occurs at index $\tau \geq 1$ if the distribution of X shifts upward or downward immediately after τ, that is, if the common CDF G of $X_{\tau+1}, X_{\tau+2}, \ldots$ is related to the common CDF F of $X_1, \ldots, X_\tau$ by $G(x) = F(x - \mu)$, where $\mu \neq 0$. The efficacy of the CUSUM is then often judged by the out-of-control average run length (OOC ARL)

$$ARL_\tau = E[N - \tau | N > \tau], \tag{5}$$

the expected time to an alarm after onset of an out-of-control situation, conditional upon there having been no alarm prior to the change. The scale invariance of sequential ranks implies that a shift μ in the median may be expressed in units of an unknown underlying scale parameter σ, which typically functions as a measure of dispersion of the underlying distribution. This could be the standard deviation, if it exists, or the inter-quartile range (IQR). Thus, the shift size in units of this scale parameter is $\delta = \mu/\sigma$. Accordingly, we may assume without loss of generality in the theoretical development of the properties of the CUSUM that $\sigma = 1$ (i.e. that the data are expressed in units of σ) and that the shift size is δ.

Some general insights into the OOC ARL behaviour of the CUSUMs can be gained by restricting attention to the behaviour of the upward CUSUM D^+ when a nominally "small" positive shift δ occurs after a "large" changepoint τ. Such an approach is in line with the primary objective of CUSUM methodology, which is to detect quickly a relatively small persistent shift. An informal calculation, shown in Appendix B, indicates that then

$$E[\xi_{\tau+k}] \approx \delta\theta_0\tau \log\left(\frac{\tau + k}{\tau + k - 1}\right), \tag{6}$$

where

$$\theta_0 = \int_{-\infty}^{\infty} \psi'(F(x)) f^2(x) dx \tag{7}$$

and where F and f denote the CDF and PDF respectively of X_1. For fixed τ the function $\log((\tau + k)/(\tau + k - 1))$, $k \geq 1$ is positive and decreases to zero as k increases. This implies that after a shift the CUSUM will show a drift which initially is upward and then tends to zero as more out-of-control data arrive. Thus, the CUSUM will not increase indefinitely after a shift but will return eventually to seemingly "in-control" behaviour. This is also clear on intuitive grounds: as observations accrue after the change, the effect of the pre-change data on the post-change sequential ranks gradually diminishes until the latter effectively become uniformly distributed again. The implication for statistical practice is that an alarm from the sequential rank CUSUM should be acted upon quickly.

2.3 Choice of reference constant

The initial upward drift rate of the CUSUM after a change is approximately $\delta\theta_0$. This follows from (6) upon making τ large and keeping k fixed, in which case $\tau \log((\tau + k)/(\tau + k - 1)) \approx 1$. In a standard normal CUSUM which targets a shift size δ_1 for detection, the optimal reference constant is $\delta_1/2$. This fact might be interpreted as suggesting $\zeta_1 = \theta_0\delta_1/2$ as an appropriate reference constant in a sequential rank CUSUM. However, the analogy is flawed because a standard normal CUSUM increases indefinitely after a change while a sequential rank CUSUM does not exhibit this feature. If a too large reference constant is used the consequent large downward drift imposed on the CUSUM impairs its ability to detect relatively small changes. Thus, ζ_1 should perhaps be regarded as an upper bound, a reference constant somewhat less than ζ_1 being more appropriate.

In fact, the non-constant nature of the post-change drift suggests an approach in which the reference constant is dispensed with and instead the control limit is allowed to vary with n. Such an approach, based upon a sequence of two-sample Mann-Whitney statistics, has been implemented by Hawkins and Deng (2010). Ross and Adams (2012) applied the same method to sequences of Kolmogorov-Smirnov and Cramér-von-Mises statistics. Some comparisons between the performances of the latter CUSUMs and the WSR CUSUM are given in Section 2.5 of the paper.

In any event, no matter what reference value is used, a sequential rank CUSUM will remain distribution free and will guarantee a nominal IC ARL. The effect of the reference constant on the performance of the CUSUM will only become apparent in an out-of-control situation. In this respect, the results shown in the next section of the paper are of some use in deciding which reference value to use.

2.4 Out-of-control ARL

An argument outlined in Appendix B leads to a heuristic tool that is useful in estimating a priori the OOC ARL of a sequential rank CUSUM:

If a small shift in the median of size δ occurs at a large $n = \tau$, then a sequential rank CUSUM with small reference value ζ and large control limit h behaves approximately like a standard normal CUSUM with the same ζ and h when shifts of size $\theta_0\delta\tau\log(n/(n-1))$ occur at $n = \tau + k$, $k \geq 0$.	(8)

The heuristic (8) suggests that approximations to the OOC ARL (5) of a sequential rank CUSUM can be found by pretending that the underlying distribution is normal. Such an approximation can

Table 2. Values of θ_0 for the WSR CUSUM in five standardized distributions.

	Normal	t_3	t_2	t_1	Gumbel
θ_0	0.98	1.37	1.18	1.10	1.12

be useful in CUSUM design since it obviates to a large extent the need to simulate data from a range of putative underlying distributions in order to gauge a priori the likely out-of-control behaviour of the CUSUM . Given a reference constant ζ, a nominal in-control ARL value ARL_0 (hence the appropriate control limit h from Table 1) and a changepoint τ, denote by $\mathcal{W}(\delta)$ the OOC ARL of a WSR CUSUM when a median shift of size $\delta > 0$ is introduced at $n = \tau + 1$. Also, denote by $\mathcal{N}(\theta_0\delta)$ the OOC ARL of a standard normal CUSUM with these same values of $\zeta,\ \tau$ and h when mean shifts of sizes $\theta_0\delta\tau\log(n/(n-1))$ occur at $n = \tau + k,\ k \geq 0$. Then (8) says that

$$\mathcal{W}(\delta) \approx \mathcal{N}(\theta_0\delta) \tag{9}$$

whenever δ is "small" and τ is "large". There is no analytical expression available for $\mathcal{N}(\theta_0\delta)$ in the literature. Nevertheless, given $\zeta, h,\ \tau$ and a nominal value of θ_0 (or a Phase I estimate of it, if such data are available), the numerical value of $\mathcal{N}(\theta_0\delta)$ can be estimated by simulation *using standard normal random numbers only*. Thus, we have what amounts to an approximation "formula" for the OOC ARL $\mathcal{W}(\delta)$ of the WSR CUSUM.

To facilitate practical implementation of (9), Table 2 shows the numerical value of θ_0 in five standardized distributions: a normal distribution, Student t-distributions with 3, 2 and 1 degrees of freedom and a Gumbel distribution (the distribution of the logarithm of a standard exponential random variable). The first four of these are symmetric while the Gumbel is moderately skew. The t_2 and t_1 distributions were standardized to unit IQR. The other three distributions were standardized to unit variance. While the precise functional form of the underlying distribution may be unknown, the values in Table 3 enable one to make a somewhat rational choice of θ_0 after taking into account the likely tail thickness and skewness of the in-control distribution and the nature of the scale parameter of choice.

If some in-control Phase I data $X_1^*, \ldots, X_m^*$ (in the original units of measurement) were available, an estimate of θ_0 can be made as follows. Observe that equation (7) can be written in the form

$$\theta_0 = \sqrt{12}E[f(X^*/\sigma)],$$

with f denoting the density function of X^*/σ. This suggests the estimator

$$\hat{\theta}_0 = \frac{\sqrt{12}}{m}\sum_{i=1}^{m}\hat{f}(X_i^*/\hat{\sigma}),$$

where $\hat{f}$ and $\hat{\sigma}$ are consistent estimators of f and σ respectively. For instance, if $\hat{f}$ is a kernel estimator

$$\hat{f}(x) = \frac{1}{mb}\sum_{i=1}^{m}\phi\left(\frac{x - X_i^*/\hat{\sigma}}{b}\right)$$

with bandwidth b and kernel ϕ, then

$$\hat{\theta}_0 = \frac{\sqrt{12}}{m^2 b} \sum_{i=1}^{m} \sum_{j=1}^{m} \phi\left(\frac{X_i^* - X_j^*}{\hat{\sigma} b}\right).$$

To assess the finite sample validity of (9), some Monte Carlo simulations involving upper WSR and standard normal CUSUMs were run. The following numerical example is representative of a general pattern observed in these simulations. Data were generated from four distributions, each standardized to unit standard deviation. These were the normal and t_3 distributions which are symmetric, and two asymmetric distributions, the Gumbel distribution and a heavily skewed normal distribution with skewness parameter $\alpha = -4$ (see Azzalini, 2005). Mean shifts of sizes $\delta_1 = 0.25$ and 0.5 were targeted and actual mean shifts δ of various sizes, shown in the first column of Table 3, were induced at $\tau = 100$. In each instance, the reference constant was taken to be (the sub-optimal) $\zeta = \theta_0\delta_0/2$ and h was taken from Table 1 to give an IC ARL of 500. The (ζ, h) pairs used are shown in the third row of Table 3. The true values $\mathcal{W}(\delta)$ were estimated in each instance from 20 000 Monte Carlo trials using random numbers from the true underlying distribution while $\mathcal{N}(\theta_0\delta)$ was obtained in a simulation which used only standard normal random numbers.

In the two symmetric distributions the approximations are excellent, except for small shifts at the $(\zeta, h) = (0.35,\ 5.66)$ combination (shown in the last two columns). In that case, ζ is arguably not small and h is not large, so that the assumptions upon which the heuristic rests are not met. In the moderately skew Gumbel distribution the approximation is acceptable, except perhaps at small shifts $\delta \leq 0.25$. However, in the skew-normal distribution the approximation would be useful only at shifts $\delta \geq 0.5$. This is no doubt due to the fact that a symmetric score function is being used on data from a distinctly asymmetric distribution.

2.5 Comparison with other distribution-free CUSUMs

We now compare the two-sided WSR CUSUM with the distribution-free CUSUM of Hawkins and Deng (2010) (the HD CUSUM) and the Kolmogorov-Smirnov CUSUM of Ross and Adams (2012) (the RA CUSUM). Data from standardised normal and t_3 distributions are used. Table 5 shows estimated (from 20 000 Monte Carlo trials) OOC ARLs of the CUSUMs at a range of shifts δ induced at the changepoints $\tau = 50$ and $\tau = 250$. To initiate the HD CUSUM, the recommended 14 initial observations (Hawkins and Deng, 2010, page 168) were used while 19 observations were used to initiate the RA CUSUM (Ross and Adams, 2012, page 106). In all three CUSUMs the two-sided nominal IC ARL was specified as 500. In the WSR CUSUM, feigning ignorance of θ_0, we use reference values $\delta_1 = 0.125$, 0.25 and 0.5 throughout. The corresponding control limits are $h = 13.34,\ 8.52$ and 4.74. In Table 4 the subscript on the WSR heading denotes the reference value. Neither the HD CUSUM nor the RA CUSUM in their present form use a reference value.

In Table 4 we see that the performances of the HD and RA CUSUMS are more or less the same as that of the WSR CUSUM at reference value 0.25. At the smaller reference value the WSR CUSUM is clearly the better of the three CUSUMs. However, following from the discussion in Section 2.3, the WSR performance at the smaller shifts degenerates badly when a large reference value is used.

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

Table 3. WSR CUSUM OOC ARL approximations in four distributions. IC ARL = 500; changepoint τ = 100.

	Normal (θ_0 = 0.98)				t_3 (θ_0= 1.38)			
(ζ, h)	(0.10, 12.01)		(0.25, 7.25)		(0.15, 9.86)		(0.35, 5.66)	
δ	$W(\delta)$	$N(\theta_0\delta)$	$W(\delta)$	$N(\theta_0\delta)$	$W(\delta)$	$N(\theta_0\delta)$	$W(\delta)$	$N(\theta_0\delta)$
0.125	252	259	306	307	211	214	274	254
0.250	118	113	163	164	70	67	114	104
0.375	52	53	76	78	30	28	44	38
0.500	32	32	37	36	19	18	20	18
0.750	19	18	17	16	12	11	10	9
1.000	14	13	11	10	9	8	8	6
	Gumbel (θ_0 = 1.11)				Skew-normal(−4) (θ_0 = 1.05)			
(ζ, h)	(0.15,9.86)		(0.25,7.25		(0.15,9.86)		(0.25,7.25)	
δ	$W(\delta)$	$N(\theta_0\delta)$	$W(\delta)$	$N(\theta_0\delta)$	$W(\delta)$	$N(\theta_0\delta)$	$W(\delta)$	$N(\theta_0\delta)$
0.125	233	258	265	279	244	270	274	292
0.250	92	102	117	132	103	117	127	146
0.375	42	44	49	55	45	51	54	64
0.500	25	26	25	27	28	28	29	30
0.750	15	14	14	13	16	15	15	14
1.000	11	10	10	9	12	11	10	9

Table 4. OOC ARL comparison between the WSR, HD and RA CUSUMs. IC ARL = 500; changepoints at τ = 50 and τ = 250.

	Normal data									
	τ = 50					τ = 250				
δ_1	$WSR_{0.125}$	$WSR_{0.25}$	$WSR_{0.5}$	HD	RA	$WSR_{0.125}$	$WSR_{0.25}$	$WSR_{0.5}$	HD	RA
0.10	444	462	472	489	474	366	401	453	417	423
0.25	307	357	426	384	388	117	174	285	169	182
0.50	91	133	253	142	153	34	36	63	38	41
0.75	30	35	91	33	36	21	18	20	18	19
1.00	18	15	24	15	16	15	13	12	11	12
	t_3 data									
	τ = 50					τ = 250				
δ_1	$WSR_{0.125}$	$WSR_{0.25}$	$WSR_{0.5}$	HD	RA	$WSR_{0.125}$	$WSR_{0.25}$	$WSR_{0.5}$	HD	RA
0.10	422	436	466	463	455	295	340	426	365	355
0.25	201	261	366	286	282	61	81	175	80	80
0.50	35	47	126	47	44	22	20	24	21	20
0.75	18	15	23	14	14	15	12	11	11	11
1.00	14	11	9	9	9	12	9	8	8	7

Table 5. OOC ARL comparison between the WSR and HD CUSUMs in an exponential distribution. IC ARL = 500; changepoints at $\tau = 50$ and $\tau = 200$.

	$\tau = 50$			$\tau = 200$		
Δ	$\mathrm{WSR}_{0.23}$	$\mathrm{WSR}_{0.39}$	HD	$\mathrm{WSR}_{0.22}$	$\mathrm{WSR}_{0.39}$	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
h	7.899	5.309		7.899	5.309	

will transform a scale change into a location change. Since the sequential ranks of the data remain unchanged under this transformation, the WSR CUSUM can be applied with ξ_i given in (2).

The HD CUSUM can also be applied directly to the X-values. Table 5 shows the estimated OOC ARLs of the WSR and HD CUSUMs in data from an exponential distribution. The IC ARL is 500 and the target scale increases are $\Delta_1 = 1.5$ and $\Delta_1 = 2.0$, which translate to target location shift sizes $\delta_1 = 1.12\log(1.5)/2 = 0.23$ and $\delta_1 = 1.12\log(2.0)/2 = 0.39$ in the WSR CUSUM. The ARLs are estimated for two changepoints $\tau = 50$ and $\tau = 200$ and at the five scale shifts Δ shown in the first column of the table. The subscript on the WSR headings in the table are the reference values and the control limits used in the WSR CUSUM are shown in the last row. Clearly, the WSR CUSUM performs better than the HD CUSUM in this particular instance.

2.6 Comparison with a normal self-starting CUSUM

Consider next the original self-starting CUSUM for a normal mean developed by Hawkins (1987) and described in further detail by Hawkins and Olwell (1998, Section 7.2). We will refer to it here as the NSS CUSUM. The data initially come from a normal distribution with unknown mean and variance. At some point the mean changes away from the initial value. The NSS CUSUM was designed to detect such a change. The NSS CUSUM recursion has the form (3) with

$$\xi_n = \Phi^{-1}(T_{n-2}(a_n V_n)),$$

where T_{n-2} denotes the CDF of the t_{n-2} distribution, $a_n = \sqrt{(1-1/n)}$ and $V_n = (X_n - \bar{X}_{n-1})/s_{n-1}$, for $n \geq 3$, with

$$s_{n-1}^2 = \sum_{i=1}^{n-1}(X_i - \bar{X}_{n-1})^2/(n-2).$$

The reference constant is taken as $\zeta = \delta_1/2$ where δ_1 is the targeted shift and the control limit is that which would be applicable in a standard normal CUSUM with reference constant ζ. While the NSS CUSUM can in principle be started after as few as $n = 3$ observations have been gathered, the deleterious effect of outliers on its performance can be mollified if a "warmup" sample of size $m > 3$ is used to initiate the CUSUM. This means that we set $D_1^+ = \cdots = D_m^+ = 0$. Here we will take $m = 15$. Since the data are known to come from a normal distribution, the VSR CUSUM would seem to be the most appropriate sequential rank procedure. Notwithstanding this, we will use the WSR

Table 6. NSS and WSR CUSUM OOC ARL comparison.

	$\tau = 50$				$\tau = 100$			
δ	$NSS_{.25}$	$WSR_{.25}$	$NSS_{.5}$	$WSR_{.5}$	$NSS_{.25}$	$WSR_{.25}$	$NSS_{.5}$	$WSR_{.5}$
0.25	237	234	312	305	166	163	240	235
0.5	70	69	144	132	37	38	76	72
0.75	21	22	51	42	16	16	22	21
1.0	11	12	15	14	10	11	10	11
1.25	8	9	8	8	8	8	7	8
1.5	7	8	6	7	6	7	5	6
1.75	6	7	5	6	5	6	4	5
2.0	5	6	4	5	5	6	4	5

CUSUM because the boundedness of the Wilcoxon score provides more protection against outliers in the data. Furthermore, in fixed size samples, the Wilcoxon two-sample test is known to be 95% efficient relative to the two-sample t-test, which suggests that the WSR CUSUM should fare well against the NSS CUSUM. The WSR CUSUM is initiated without any warmup observations and will use the same reference constants as the NSS CUSUM. Table 6 shows the results of a Monte Carlo comparison of the CUSUMs at a nominal IC ARL of 500 and reference constants $\zeta = 0.25$ and $\zeta = 0.5$ when changes of size δ, indicated in the first column of the table are introduced after $\tau = 50$ and $\tau = 100$ observations.

Inspection of the results in the table reveals that the WSR CUSUM performs as well as, or better than, the NSS CUSUM. This somewhat surprising phenomenon has also been observed in other contexts – see, for instance, Hawkins and Deng (2010, p. 170). In contrast to the NSS CUSUM, the WSR CUSUM has the added advantages that its in-control behaviour is impervious to deviations from the normality assumption and that it requires no “warmup data” in order to initialize.

3. A sequential rank CUSUM for dispersion

Despite the fact that the in-control properties of a sequential rank location CUSUM do not depend upon the dispersion in the underlying distribution, the validity of the CUSUM does require the dispersion to remain unchanged. Running a CUSUM that can detect a change in the unknown numerical value of a dispersion parameter σ, enables one to monitor the validity of this assumption. A CUSUM that immediately suggests itself is the squared value of the summand in the WSR CUSUM, adjusted to have in-control mean zero,

$$\xi_i^* = \xi_i^2 - 1,$$

with ξ_i from (2). Thus,

$$\xi_i^* = \frac{12(i+1)}{i-1}\left(\frac{r_i}{i+1} - \frac{1}{2}\right)^2 - 1. \tag{10}$$

Since the in-control distribution of ξ_i^* is not symmetric around zero, the relationship between the upward and downward CUSUMs is not as straightforward as it is for location CUSUMs. In the present context, the upward and downward CUSUMs are

$$D_1^+ = 0,\ D_n^+ = \max\left\{0, D_{n-1}^+ + \xi_n^* - \zeta^+\right\}, \quad n \geq 2,$$

Table 7. Control limits for the upward MSR CUSUM.

ζ	IC ARL						
	100	200	300	400	500	1 000	2 000
0.00	7.99	11.68	14.53	16.97	19.05	27.36	39.11
0.05	6.64	9.11	10.94	12.36	13.45	17.35	21.71
0.10	5.75	7.64	8.88	9.76	10.53	12.97	15.60
0.15	5.04	6.56	7.48	8.20	8.72	10.55	12.38
0.20	4.47	5.72	6.49	7.03	7.50	8.91	10.36
0.25	4.04	5.12	5.74	6.21	6.58	7.72	8.91
0.30	3.68	4.60	5.14	5.55	5.85	6.82	7.84
0.35	3.36	4.17	4.65	5.01	5.28	6.14	6.98
0.40	3.08	3.83	4.24	4.56	4.79	5.54	6.31
0.45	2.85	3.51	3.90	4.17	4.39	5.04	5.73
0.50	2.64	3.24	3.57	3.83	4.02	4.63	5.24

Table 8. Control limits for the downward MSR CUSUM.

ζ	IC ARL						
	100	200	300	400	500	1 000	2 000
0.00	8.00	11.75	14.57	16.95	19.02	27.25	39.08
0.05	6.51	8.93	10.71	12.02	13.02	16.96	21.04
0.10	5.40	7.15	8.34	9.13	9.86	12.10	14.46
0.15	4.54	5.92	6.73	7.31	7.82	9.40	10.95
0.20	3.89	4.94	5.58	6.03	6.39	7.54	8.72
0.25	3.37	4.19	4.71	5.06	5.35	6.24	7.15
0.30	2.92	3.58	4.00	4.29	4.51	5.25	5.96
0.35	2.51	3.06	3.41	3.63	3.84	4.42	5.02
0.40	2.16	2.62	2.90	3.11	3.26	3.74	4.23
0.45	1.86	2.24	2.47	2.64	2.78	3.17	3.58
0.50	1.58	1.90	2.10	2.23	2.34	2.67	3.00

with control limits h^+, and

$$D_1^- = 0,\ D_n^- = \max\left\{0, D_{n-1}^- - \xi_n^* - \zeta^-\right\}, \quad n \geq 2,$$

with control limits h^-. In general, the reference constants ζ^+ and ζ^- need not have the same numerical value. Even if they are identical, the control limits h^+ and h^- that give the upward and downward CUSUMs the same in-control ARL will differ, especially at the larger reference constants. Thus, separate tables of control limits are required to detect increases and decreases in dispersion at a nominal in-control ARL.

In view of the fact that the summand (10) is reminiscent of the score function in the two-sample scale test of Mood (1954), we will refer to the corresponding CUSUMs as MSR CUSUMs. Table 7 shows control limits h^+ and h^- obtained by Monte Carlo simulation for a range of nominal IC ARLs and reference constants ζ^+ and ζ^- for the upward and downward MSR CUSUMs.

If the unknown σ changes to $\sigma\Delta$ where $0 < \Delta \neq 1$, then the analogue of (6) is (see Appendix B)

$$E[\xi_{\tau+k}^*] \approx \theta^* \tau(\log \Delta) \log\left(\frac{\tau + k}{\tau + k - 1}\right),$$

Table 9. Values of θ^* in the MSR CUSUM for five distributions.

	Normal	t_3	t_2	t_1	Gumbel
θ^*	1.10	0.89	0.80	0.61	1.01

with

$$\theta^* \equiv 24 \int_{-\infty}^{\infty} (G(x) - 1/2) x g^2(x) dx \tag{11}$$

and G and g denoting the CDF and PDF of $X - \nu$ and ν denoting the (unknown) in-control median. This is analogous to the result (6) for the location CUSUM with θ_0 and δ there replaced by θ^* and $\log \Delta$. The result suggests that taking $\zeta^+ = \zeta^- = \log \Delta_0 > 0$ is equivalent to targeting upward and downward scale changes proportional to $\Delta_0 > 1$ and $1/\Delta_0$ respectively. Values of θ^* in the five distributions listed in Table 2 are shown in Table 9.

If some Phase I data $X^*_{(1)} < \cdots < X^*_{(m)}$ are available, θ^* can be estimated by

$$\hat{\theta}^* = \frac{24}{m} \sum_{i=1}^{m} \left(\frac{i}{m+1} - \frac{1}{2} \right) (X^*_{(i)} - \hat{\nu}_m) \hat{g}(X^*_{(i)}), \tag{12}$$

where $\hat{g}$ is the kernel estimate

$$\hat{g}(x) = \frac{1}{mb} \sum_{j=1}^{m} \phi \left(\frac{x - X^*_{(i)}}{b} \right)$$

and $\hat{\nu}_m$ is the median of the Phase I data. When g can be assumed to be symmetric around zero, $\hat{\nu}_m$ in (12) can be set equal to zero.

4. Application: Monitoring coal quality

The amount of gas produced in a coal gasification facility depends crucially on the ash content of the received coal. Therefore, it is important for coal suppliers to control and manage the coal quality sent to the downstream users. It is equally important for the coal users to know the expected quality of the coal in order to implement pro-active operational changes. For this purpose, an XRF (X-ray fluorescent) coal analyzer which collects real time information on the ash content is mounted over a conveyor belt which transports the coal from the stockyard to the downstream facility. Figure 1 shows a time series plot of 175 successive aggregated ash measurements (standardized to zero mean and unit variance in order to comply with a confidentiality agreement with the provider of the data). The data shown are 15 minute aggregated data which is sufficient for detecting changes in coal quality in time for the downstream units to take action. The coal is reclaimed from two to three heaps of between 20 000 to 40 000 tons each, with varying coal ash yield within and between the heaps. From Figure 1, the short and long term variation in ash content of the reclaimed coal is clearly evident. The variation may be due to different coal sources with varying ash yield being stacked and reclaimed. CUSUMs can now be used to detect real time changes in the median ash yield and in the dispersion of the data.

While the data show substantial variability, differing average levels are clearly discernible in retrospect. Also, there are apparent outliers and groups of outliers among the data. These are shown

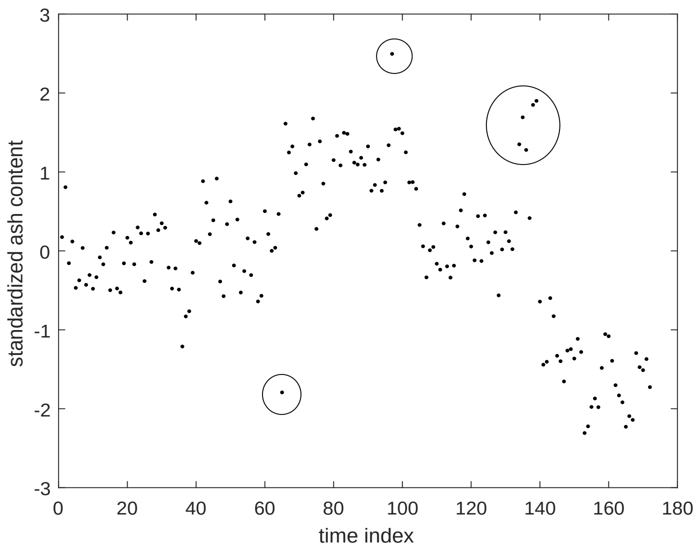

Figure 1. Time series plot of standardized XRF ash content measurements.

enclosed in circles in Figure 1. In the present context, such values typically appear when the mass of coal scanned by the gauge per unit of time is lower than the minimum amount required for accurate determination of ash content. Such outliers do not indicate real changes in coal quality and constitute what are referred to as transient local effects in the CUSUM literature. In this particular application the CUSUM should be robust against such effects in order to prevent frequent false alarms that would compromise effective management of the coal blending system.

To see how a CUSUM would react when such data are observed sequentially, we implement a two-sided WSR CUSUM with a reference constant $\zeta = 0.5$ and control limit $h = 3.68$, which results in an IC ARL of 150. The changepoint is estimated in the usual manner, namely as the last index at which the CUSUM, upper or lower, producing the alarm was last at zero. *If an alarm sounds at index n, the CUSUM is restarted at observation number n.* This rule provides some insurance against restarting at a substantial underestimate of the true changepoint which is then likely to lead to a false alarm shortly thereafter. The results are shown in Table 10, in which the third column indicates the direction, up or down, of the putative change.

Looking at the retrospective time series plot in Figure 1, it seems that some of the substantial number of alarms result from the transient local effects mentioned earlier rather than from sustained changes in the mean. From a coal quality management perspective, a somewhat more acceptable result is obtained upon use of the Cauchy score

$$\psi(u) = \sqrt{2} sin(2\pi(u - 1/2)),$$

which is plotted in Figure 2.

The monotone increasing nature of the function on the interval $1/4 < u < 3/4$ implies an ability of the corresponding CUSUM to detect moderately sized changes. On the other hand, since the function

Table 10. Results of WSR CUSUM implementation on the XRF data.

Alarm at	Changepoint	Shift
$n = 30$	$\tau = 19$	up
$n = 36$	$\tau = 30$	down
$n = 42$	$\tau = 36$	up
$n = 69$	$\tau = 65$	up
$n = 100$	$\tau = 95$	up
$n = 106$	$\tau = 100$	down
$n = 136$	$\tau = 132$	down
$n = 147$	$\tau = 139$	down

Table 11. Results of Cauchy CUSUM implementation on the XRF data.

Alarm at	Changepoint	Median Shift	Sd.	Shift/sd.
$n = 87$	$\tau = 74$	1.06	0.62	1.70
$n = 112$	$\tau = 101$	−1.12	0.44	−2.50
$n = 151$	$\tau = 141$	−1.49	0.73	−2.04

decreases on the intervals $3/4 < u < 1$ and $0 < u < 1/4$ the CUSUM will not be able to detect large changes. In the present context this inability is actually desirable because the CUSUM will then not be affected by the large spurious outliers among the data. In implementing the Cauchy CUSUM, the summand $\xi_i = \psi(r_i/i)$ is used because then $var(\xi_i) \equiv 1$. For the two sided CUSUM with a reference constant $\zeta = 0.5$ and nominal IC ARL 150, the control limit is $h = 3.59$. The CUSUM results, shown in Table 11 and in Figure 3, indicate three median changes, hence four segments of constancy, namely the intervals from 1 to 74, 75 to 101, 102 to 141 and 142 to 172. These also appear to define the only substantial sustained changes visible in Figure 1. The next to last column in Table 11 shows the standard deviations in the first three segments and the last column shows the median shift in units of the estimated standard deviation.

We now consider the monitoring of dispersion in the data. A comparison of ξ_i^* from (10) with ξ_i from (2) indicates that if the dispersion is constant and the median then changes substantially, the MSR CUSUM will tend to raise alarms at more or less the same times as the WSR CUSUM. Indeed, this turns out to be the case for the data shown in Figure 1. Running the MSR CUSUM at $\zeta^+ = \zeta^- = 0.4$ with $h^+ = 5.54$ and $h^- = 3.74$ gives the upper and lower CUSUM an IC ARL of 1 000 each, for an overall IC ARL of 500. Table 12 shows the results. Clearly, the dispersion alarms coincide almost exactly with median shifts identified by the WSR CUSUM and therefore do not correspond to intrinsic changes in dispersion.

The implication of these results is that the dispersion is most likely constant *within* each of the location-change segments identified by the WSR CUSUM. However, this says nothing about the constancy, or otherwise, of the dispersions *between* segments. The detection of an increase in dispersion between segments *in real time* is important in the present application because an increase could point to either a malfunctioning of the XRF equipment or to a fundamental change in the

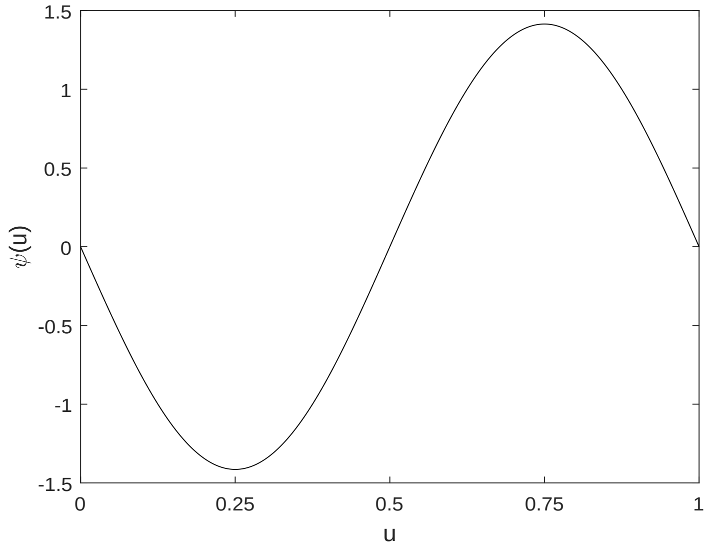

Figure 2. The Cauchy score function.

Table 12. Results of Cauchy CUSUM implementation on the XRF data.

Alarm at	Changepoint	Shift
n = 50	τ = 27	up
n = 69	τ = 64	up
n = 110	τ = 104	up
n = 140	τ = 133	up

intrinsic coal characteristics. Retrospective comparison of within-segment dispersions using, for instance, the data in the fourth column of Table 11, is therefore of little value. A method that eliminates the impact of changes in the mean ash content on the monitoring of dispersion is required.

One approach is to monitor the differenced data $Y_i = X_i - X_{i-1},\ i \geq 2$, shown in Figure 4. In such a plot changes in the mean X-value tend to manifest themselves as spurious outliers which, because they are few in number and are also well separated in time, should not affect the functioning of the MSR CUSUM. However, the Y_i are not statistically independent, so that the distribution theory underlying the MSR CUSUM is invalidated. Further research is therefore required into the behaviour of the CUSUM applied to such data .

The fact that the successive Y_i are highly *negatively* correlated implies that the successive sequential ranks will also be negatively correlated. Therefore, the IC ARL of the MSR CUSUM based on the Y_i will, if anything, tend to be much smaller than the nominal value. Indeed, Monte Carlo simulations at a range of ζ^+ values and nominal IC ARL values of 200 and 500 using data from normal and t_2 distributions produced estimated true IC ARLs substantially smaller than the nominal ones.

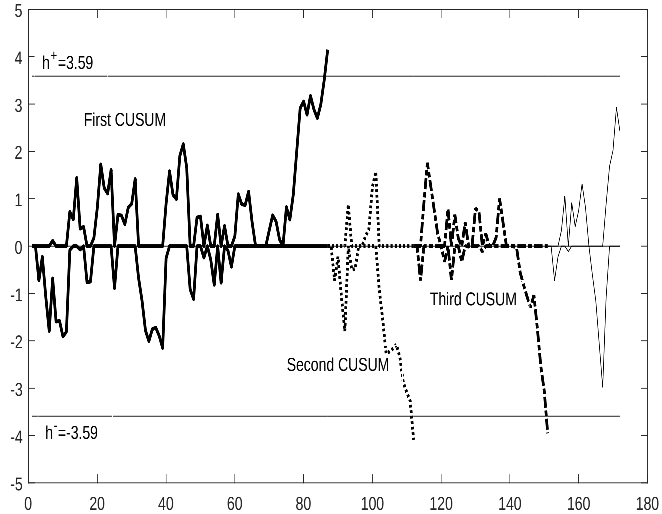

Figure 3. The successive Cauchy CUSUMS, restarted at each alarm time.

Table 13. Simulation results for the MSR CUSUM applied to first differences of i.i.d. data from normal and t_2 distributions.

	IC ARL	Normal		t_2	
	Nominal	250	500	250	500
$\zeta^+ = 0.4$	Estimated	107	176	106	176
$\zeta^+ = 0.1$	Estimated	143	246	136	238

The results are shown in Table 13. It is of some interest to note that the estimated IC ARLs, while substantially less than the nominal values, are more or less the same in data from the two distributions.

As could be expected from an inspection of Figure 4, the MSR CUSUM ($\zeta^+ = \zeta^- = 0.1$, $h^+ = 12.97$, $h^- = 12.10$) with IC ARL substantially less than the nominal 500) raises no alarm. The conclusion is that the dispersion has not changed over any extended period in the full stretch of data.

5. Summary

We develop distribution-free CUSUMs based on sequential ranks to detect changes away from a current, but unknown, median and dispersion parameter of an unknown continuous distribution. The CUSUMs are distribution free in that the in-control average run length does not depend on the functional form of the underlying distribution function. Furthermore, the CUSUMs are fully self starting, are free of between-practitioner effects and can be constructed to accommodate heavy-tailed distributions and occasional extreme outliers. Special attention is focused on CUSUMs based on the Wilcoxon (for location shifts) and Mood (for scale shifts) statistics. Tables of control limits guar-

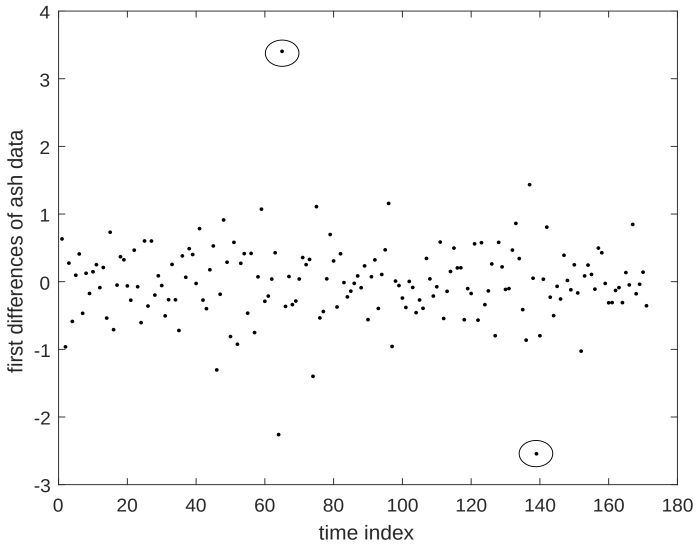

Figure 4. First differences $Y_i = X_i - X_{i-1}$ of the standardized XRF data.

anteeing a specified in-control average run length are provided. Implementation of the CUSUMs is illustrated in an application to data from a process industry.

Acknowledgement. The financial assistance of the National Research Foundation (NRF) towards this research is acknowledged. Opinions expressed and conclusions arrived at, are those of the authors and are not necessarily to be attributed to the NRF.

Appendix

A. Computations for control limits

Because the partial sums of the ξ_i are approximately normally distributed, it is not hard to imagine that the control limits h of the CUSUM will correspond closely to those of a standard normal CUSUM, especially when ζ is small and h is large. Given a large set of reference constants ζ and nominal IC ARL values ARL_0, denote by h_1 the corresponding control limits from a standard normal distribution CUSUM. The first step in an iterative process was to estimate the IC ARL of the CUSUM at each pair on the (ζ, h_1) grid. For this, 10 000 independent Monte Carlo estimates of the IC ARL were generated using uniform random numbers on $[0, 1]$ as the underlying distribution. (Since the sequential ranks are distribution free, *any* distribution will do.) Denote the set of means of the 10 000 estimates at each grid point by $\hat{A}(\zeta, h_1)$. Cubic spline interpolation from $(\zeta, \hat{A}(\zeta, h_1))$ to (ζ, h) then yields new estimates, h_2, of the correct control limits. A further 10 000 Monte Carlo estimates using these new control limits produced a further set of estimated IC ARLs $\hat{A}(\zeta, h_2)$. This process was repeated until all the differences $|\hat{A}(\zeta, h) - ARL_0|$ were less than 3. For $\zeta \leq 0.25$, no more that three iterations were required, while for $\zeta > 0.25$, six iterations sufficed. Finally, the control limits were each checked independently in 100 000 Monte Carlo runs. The largest difference found between nominal and simulation estimated IC ARLs was 3.

B. A justification for the heuristic (8)

It is well known – see Page (1954) – that D_n in the CUSUM recursion

$$D_n = \max(0, D_{n-1} + \xi_n - \zeta)$$

can also be expressed in the form

$$D_n = S_n - \min_{0 \le k \le n} S_k, \tag{13}$$

where

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

for $n \ge 1$ and $S_0 := 0$. This equivalence shows that the properties of the CUSUM are determined by the properties of the partial sums S_n. Proposition 1 deals with the properties of these partial sums. The proposition is a special case of Theorem 5.1 in Lombard (1983) and forms the basis for the heuristic. The proof, which is omitted, consists in making some straightforward identifications between the notation used in this paper and that used in Lombard (1983).

(a) Suppose the independent observations $X_1, \ldots, X_\tau$ have common PDF $f(x)$ and $X_{\tau+1}, X_{\tau+2}, \ldots$ have PDF $f(x - \beta/h)$, $\beta \neq 0$. Set $\zeta = \Delta/h$. Then the continuous time process

$$S_{\lfloor h^2 t \rfloor}/h = \left(\textstyle\sum_{i=1}^{\lfloor h^2 t \rfloor} (\xi_i - \zeta)\right)/h, \quad t \ge 0, \tag{14}$$

converges in distribution as $h \longrightarrow \infty$ to the continuous time process

$$Y(t) = W(t) - \Delta t + \beta\, \theta_0 \; \max\{0, \tau^* log(t/\tau^*)\}, \quad t \ge 0, \tag{15}$$

where W denotes a standard Brownian motion and where θ_0 is given in (7).

(b) Suppose the independent observations $X_1, \ldots, X_\tau$ have common PDF $g(x-\nu)$ and $X_{\tau+1}, X_{\tau+2}, \ldots$ have PDF $\exp(-\beta/h) g((x-\nu)\exp(-\beta/h))$, where g has a zero median. Then the continuous time process (14) converges in distribution as $h \longrightarrow \infty$ to the continuous time process Y in (15) with θ_0 replaced by

$$\theta_1 = \int_{-\infty}^{\infty} \psi'(G(x)) x g^2(x) dx. \tag{16}$$

Here, convergence in distribution is meant in the sense of weak convergence of probability measures on the space $D[0, \infty)$ – see Billingsley (1999).

Upon evaluating (14) and (15) at $t = n/h^2$, $1 \le n \le \tau$, and at $t = (\tau + k)/h^2$, $k \ge 1$, Proposition 1(a) suggests that if the location changes by an amount β/h, then the joint distributions of the partial sums S_n/h, $n \ge 1$, can be approximated by those of the sequence

$$Y(n/h^2) = W(n/h^2) - \Delta n/h + (\beta/h)\, \theta_0 \max\{0, (\tau/h) log(n/\tau)\}.$$

Now let $\xi_1^*, \ldots \xi_\tau^*$ be i.i.d. normal(0, 1) and, for $k \geq 1$, let $\xi_{\tau+k}^*$, be normal(μ_k, 1) and independent with

$$\mu_k = (\beta/h)\, \theta_0\, (\tau/h)\, log\, ((\tau + k)/(\tau + k - 1))\,.$$

Set $S_n^* = \xi_1^* + \cdots + \xi_n^* - n\Delta$. Since the sequences $Y(n/h^2)$, $n \geq 1$, and S_n^*/h, $n \geq 1$, are identically distributed and since the sequence S_n/h, $n \geq 1$, is approximated in distribution for large h by the sequence $Y(n/h^2)$, $n \geq 1$, it follows that the joint distributions of S_n^*/h, $n \geq 1$, provide an approximation to the joint distributions of S_n/h, $n \geq 1$, when h is large. Thus, – see (13) – the joint distributions of the standard normal CUSUMs based on the S^* sequence provide an approximation to the joint distributions of the sequential rank CUSUMs based on the S sequence.

For a scale change from σ to $\sigma \exp(-\delta)$, a similar argument holds with θ_0 replaced by θ_1 from (16). Then the following modification of the heuristic (8) is applicable:

If the scale parameter changes from σ to $\sigma\Delta$ at a large $n = \tau$, then a sequential rank dispersion CUSUM with small reference value ζ and large control limit h behaves approximately like a standard normal CUSUM with the same ζ and h when location changes of size $\theta_1 log(\Delta)\tau \log(n/(n-1))$ occur at $n = \tau + k, k \geq 0$.

The Mood CUSUM is the special case where $\psi(u) = 12(u - 1/2)^2 - 1$. Substitution into (16) then shows that θ_1 equals θ^* from (11).

References

Azzalini, A. (2005). The skew-normal distribution and related multivariate families. *Scand. J. Stat.*, **32**, 159–188.

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

Barndorff-Nielsen, O. E. (1963). On the limit behaviour of extreme order statistics. *Ann. Math. Stat.*, **34**, 992–1002.

Bhattacharya, P. K. and Frierson, D. (1981). A nonparametric control chart for detecting small disorders. *Ann. Stat.*, **9**, 544–554.

Billingsley, P. (1999). *Convergence of Probability Measures*. Wiley, New York.

Chatterjee, S. and Qiu, P. (2009). Distribution-free cumulative sum control charts using bootstrap-based control limits. *Ann. Appl. Stat.*, **3**, 349–369.

Gandy, A. and Kvaløy, J. (2013). Guaranteed conditional performance of control charts via bootstrap methods. *Scand. J. Stat.*, **40**, 647–668.

Hawkins, D. M. (1987). Self-starting CUSUM charts for location and scale. *J. Roy. Stat. Soc. D*, **36**, 299–316.

Hawkins, D. M. and Deng, Q. (2010). A nonparametric change-point control chart. *J. Qual. Technol.*, **42**, 165–173.

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

KEEFE, M. J., WOODALL, W. H., AND JONES-FARMER, L. A. (2015). The conditional in-control performance of self-starting control charts. *Qual. Eng.*, **27**, 488–499.

LOMBARD, F. (1983). Asymptotic distribution of rank statistics in the change-point problem. *S. Afr. Stat. J.*, **17**, 83–105.

LOMBARD, F. AND VAN ZYL, C. (2018). Signed sequential rank CUSUMs. *Comput. Stat. Data An.*, **118**, 30–39.

MCDONALD, D. (1990). A CUSUM procedure based on sequential ranks. *Nav. Res. Log.*, **37**, 627–646.

MOOD, A. M. (1954). On the asymptotic efficiency of certain nonparametric two-sample tests. *Ann. Math. Stat.*, **25**, 514–522.

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

ROSS, G. J. AND ADAMS, N. M. (2012). Two nonparametric control charts for detecting arbitrary distribution changes. *J. Qual. Technol.*, **44**, 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. *Qual. Eng.*, **28**, 402–415.

Manuscript received 2018-01-06, revised 2018-03-05, accepted 2018-03-06.