The innovation cumulative sum chart

Alex J. Koning

Econometric Institute
Erasmus University Rotterdam
P.O. Box 1738
NL-3000 DR Rotterdam
The Netherlands

January 16, 1996

Abstract

This paper investigates the acceptable quality level under which quality characteristics have a normal probability distribution with an unknown location parameter. The traditional cumulative sum depends on the unknown parameter, and hence can not be used. Replacing the unknown parameter by its estimator yields the estimated cumulative sum, but the replacement clearly affects the behavior under the acceptable quality level. To overcome this problem, the innovation cumulative sum is proposed.

The innovation cumulative sum performs at least as good as the estimated cumulative sum under a linear trend rejectable quality level, and clearly performs better for a rejectable quality level with the change point occurring in the first 70 percent of the sample.

Finally, the definition of the innovation cumulative sum is extended for acceptable quality levels involving quality characteristics having a nonnormal distribution.

Keywords Quality control; Cumulative sum chart; Innovation approach.

Contents

1 Introduction

Control charts, first developed by Walter A. Shewhart at Bell Telephone Laboratories [see Shewhart (1931)], have become fundamental tools for detecting, diagnosing and correcting industrial production problems [see Ishikawa (1982)].

In paragraph 5.1.3 in Wetherill and Brown (1991) the use of Shewhart control charts is divided into two phases: the set-up phase and the operational phase. In the operational phase, the production process is already in control, and the objective is to keep it in control. Usually, extensive process knowledge and specific process requirements are available, leading to a simple acceptable quality level: an acceptable quality level which completely specifies the distribution of the quality characteristics.

In the set-up phase the production process is not in control, and the objective is to bring it in control. Typically, process knowledge and requirements are still being developed, leading to a composite acceptable quality level: an acceptable quality level which only partially specifies the distribution of the quality characteristics; that is, some parameters of the distribution of the quality characteristics remain unknown.

Although developed for simple acceptable quality levels, Shewhart control charts are easily modified to accomodate composite acceptable quality levels as well. The only modification needed is to replace the unknown parameters by estimators. The effects of this modification are limited, and tend to disappear as the sample size grows large.

Cumulative sum charts, originally proposed in Page (1954) [see also Dobben de Bruyn (1968)], are more effective than Shewhart control charts in detecting small departures from a simple acceptable quality level. Unfortunately, their straightforward adaption to composite acceptable quality levels can be hazardous. In this paper we give some examples in which the replacement of unknown parameters by estimators drastically changes the behavior under the acceptable quality level of the cumulative sum chart, and propose a composite acceptable quality level adaption of the cumulative sum chart which is based on the so-called innovation parts of the quality characteristics rather than on the quality characteristics themselves. The behavior of the innovation cumulative sum chart under a composite acceptable quality level resembles the behavior of the traditional cumulative sum chart under a simple acceptable quality level.

The outline of the paper is as follows. In section 2 we develop the innovation cumulative sum for an acceptable quality level under which the quality characteristics are independent random variables having a normal distribution with unknown expectation θ and variance 1. The behavior of this innovation cumulative sum under a rejectable quality level is investigated in section 3. In section 4 an extended definition of the innovation cumulative sum is presented, which also applies to non-normal distributions.

2 The acceptable quality level

Consider random variables $X_1, X_2, \ldots, X_n$, which under the acceptable quality level are independent and normally distributed with expectation θ and variance 1. If θ is known, then the acceptable quality level is simple, and we are able to use the traditional cumulative sum

$$C_j = \sum_{i=1}^{j} (X_i - \theta) .$$

The decision whether the cumulative sum is out of control at time j is often taken with the aid of a V-mask [see Figure 1]. The cumulative sum is out of control at time j if one or more of the earlier values $C_1, \ldots, C_{j-1}$ fall outside the V-mask with origin C_j. In Lucas (1982) an equivalent control chart is proposed, which employs control lines rather than a V-mask. Although the equivalent control chart has certainly practical advantages, we prefer to use the original V-mask cumulative sum procedure for ease of exposition.

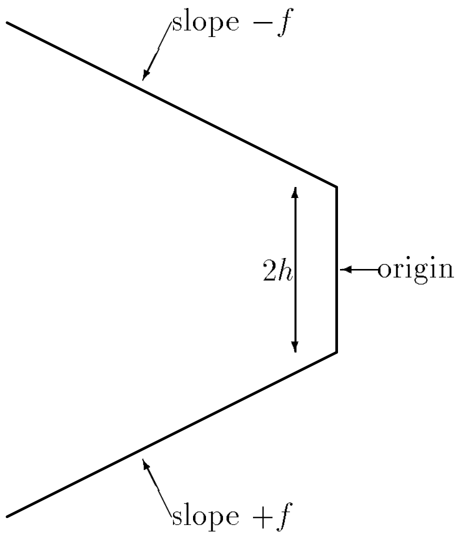

Figure 1: The V-mask $V_{f,h}$ consists of two backwards pointing arms with slopes $-f$ and f attached to a vertical line of length $2h$; the midpoint of the vertical line is the origin of this mask.

If θ remains unknown, then the acceptable quality level is composite. Obviously, the traditional cumulative sum can be adapted by replacing θ by the sample mean $\bar{X}_n = n^{-1} \sum_{i=1}^{n} X_j$, which yields the estimated cumulative sum

$$\hat{C}_j = \sum_{i=1}^{j} \left(X_i - \bar{X}_n \right) .$$

Note that the use of the sample mean to estimate θ implies that the estimated cumulative sum can only be applied to past data.

Unfortunately, the replacement of θ by the sample mean changes the behavior of the cumulative sum drastically. This is most easily seen by comparing C_n to $\hat{C}_n$. Under the acceptable quality C_n is a random variable with expectation zero and variance n, whereas $\hat{C}_n$ is degenerate in zero [that is, takes the value zero with probability 1].

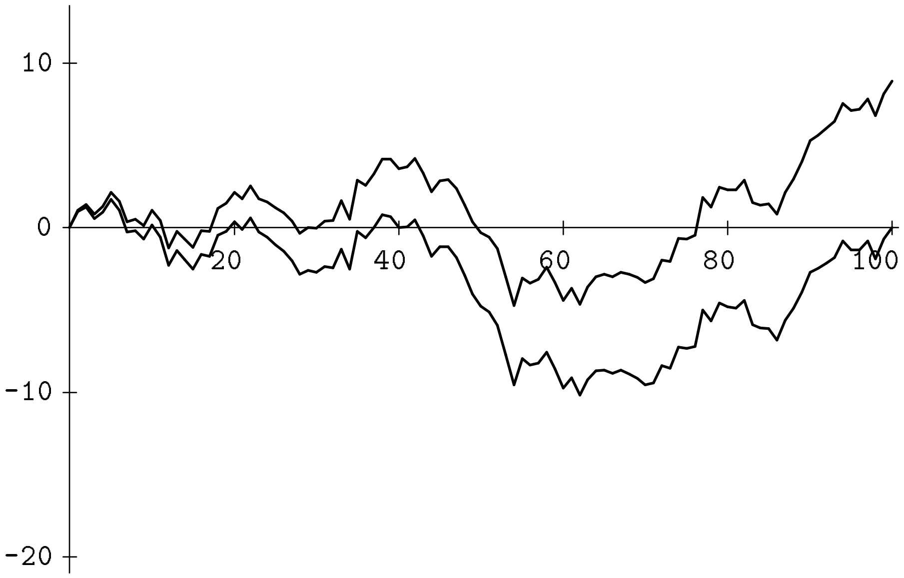

Figure 2: Traditional and estimated cumulative sum for a sample of size 100 from the standard normal distribution.

Figure 2 gives a further demonstration of the change in behavior. In this figure the traditional cumulative sum C_j [with $\theta = 0$] and the estimated cumulative sum $\hat{C}_j$ are plotted versus j. Both cumulative sums are based on the same random sample $X_1, X_2, \ldots, X_{100}$ from the standard normal distribution. Observe that the estimated cumulative sum returns to zero eventually, whereas the traditional cumulative sum does not. In Bissel (1994) a similar figure is found on p. 198, but the issue of the change in behavior is not addressed.

The effects of estimation do not vanish as n increases. Mathematically, the change in behavior has the following asymptotic consequence. Let $[nt]$ denote the largest integer not exceeding nt. Then the stochastic proces $\left(n^{-1/2}C_{[nt]}\right)_{t\in[0,1]}$ converges weakly to a standard Wiener process under the simple acceptable quality level, whereas the stochastic process $\left(n^{-1/2}\hat{C}_{[nt]}\right)_{t\in[0,1]}$ converges weakly to a standard Brownian bridge under the composite acceptable quality level [the process $\left(-n^{-1/2}\hat{C}_{[nt]}\right)_{t\in[0,1]}$ is called a standardized time series in Schruben (1982, 1983)].

The source of this problem is that knowledge of $\bar{X}_n$ and $X_1, X_2, \ldots, X_{j-1}$ yields information about the random variable X_j, since $\sum_{\ell=i}^{n} X_\ell = n\bar{X}_n - \sum_{\ell=1}^{i-1} X_i$ and X_j are not independent.

A similar problem occurs in the field of goodness of fit. The empirical process, the rescaled difference between the empirical distribution function and the cumulative distribution function, is often used for investigating a simple null hypothesis. A straightforward application of the empirical process for investigating a composite null hypothesis involves replacing the unknown parameter in the cumulative distribution function by an estimator. Unfortunately, the replacement changes the behavior of the empirical process and makes its distribution intractable [see Durbin (1973)]. This problem can be avoided by using the innovation approach proposed in Khmaladze (1981) [see also Khmaladze (1993), and paragraph VI.3.3.4 in Andersen et al (1993)], in which a goodness of fit process is transformed by subtracting the conditional expectation given the combination of its past and the value taken by the estimator. The innovation cumulative sum proposed in this paper is obtained by applying the innovation approach to the traditional cumulative sum.

At time instance i the random variable $\sum_{\ell=i}^{n} X_\ell$ summarizes the relevant information contained in the past $X_1, \ldots, X_{i-1}$ and the estimator $\bar{X}_n$. To describe the dependence between X_i and $\sum_{\ell=i}^{n} X_\ell$ under the acceptable quality level, we introduce the random variable

$$\breve{X}_i = X_i - \frac{1}{n-i+1} \sum_{\ell=i}^{n} X_\ell.$$

Observe that

$$\breve{X}_i = X_i - \frac{\operatorname{covar}\left(X_i, \sum_{\ell=i}^{n} X_\ell\right)}{\operatorname{var}\left(\sum_{\ell=i}^{n} X_\ell\right)} \sum_{\ell=i}^{n} X_\ell = X_i - \mathcal{E}\left(X_i \mid \sum_{\ell=i}^{n} X_\ell\right).$$

Observe that $\breve{X}_i$ is in fact the residual obtained by regressing X_i on $\sum_{\ell=i}^{n} X_\ell$. It follows that $\breve{X}_i$ and $\sum_{\ell=i}^{n} X_\ell$ are uncorrelated, and hence independent [this in turn implies that $\breve{X}_i$ and $\bar{X}_n$ are independent].

Furthermore, observe that X_i can be written as the sum of $\breve{X}_i$ and $(n-i+1)^{-1} \sum_{\ell=i}^{n} X_\ell$. This representation decomposes X_i in a part which is independent of $\sum_{\ell=i}^{n} X_\ell$ and a part which is completely determined by $\sum_{\ell=i}^{n} X_\ell$.

Finally, observe that

$$\operatorname{cov}\left(\breve{X}_i, \breve{X}_k\right) = \begin{cases} \dfrac{n-i}{n-i+1} & \text{if } i = k, \\ 0 & \text{if } i \neq k. \end{cases}$$

Hence, the random variables

$$\breve{X}_1 \sqrt{\frac{n}{n-1}}, \breve{X}_2 \sqrt{\frac{n-1}{n-2}}, \ldots, \breve{X}_{n-1} \sqrt{\frac{2}{1}}$$

are independent and standard normal. Basing a traditional cumulative sum on these random variables yields the innovation cumulative sum

$$\breve{C}_j = \sum_{i=1}^{j} \breve{X}_i \sqrt{\frac{n-i+1}{n-i}}.$$

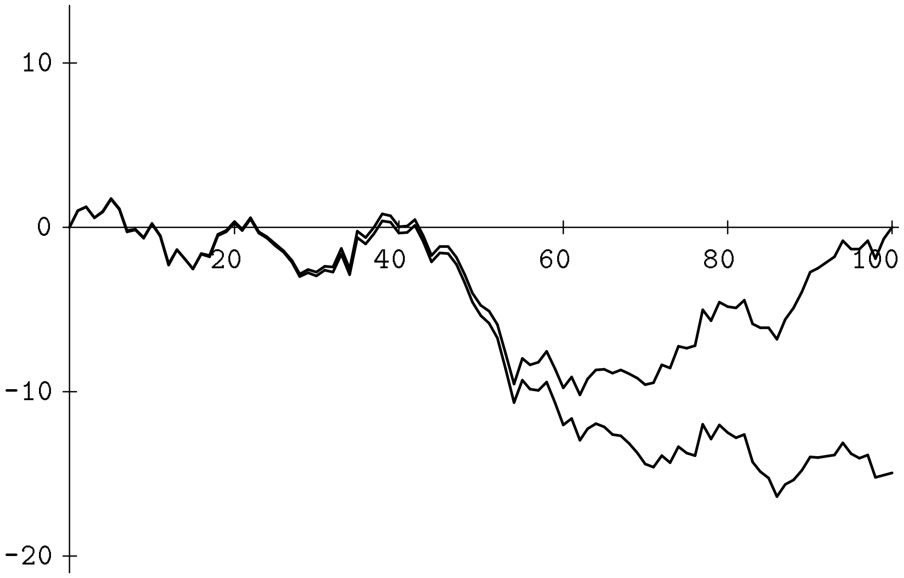

Figure 3: Estimated and innovation cumulative sum for a sample of size 100 from the standard normal distribution.

Figure 3 plots the estimated cumulative sum $\hat{C}_j$ and the innovation cumulative sum versus j. Both cumulative sums are based on the same random sample $X_1, X_2, \ldots, X_{100}$ from the standard normal distribution.

Initially, the difference between the estimated and the innovation cumulative sum is only slight. However, as the ultimate return to zero of the estimated cumulative sum comes closer, the different behavior of the innovation cumulative sum starts to materialize.

Mathematically, the stochastic process $\left(n^{-1/2}\breve{C}_{[nt]}\right)_{t\in[0,1]}$ converges weakly to a standard Wiener process under the composite acceptable quality level.

Observe that the availability of the random variables $\breve{X}_i\sqrt{(n-i+1)/(n-i)}$ opens up the possibility of implementing an equivalent control chart along the lines of Lucas (1982). We may even add the fast initial response feature proposed in Lucas and Crosier (1982).

To investigate the behavior of the traditional, the estimated and the innovation cumulative sum, we simulated 10,000 random samples of size 1000 from the standard normal distribution using S-Plus [see section 5.2 in Venables and

h	*upper* *trad.*	*estim.*	*innov.*	*lower* *trad.*	*estim.*	*innov.*	*two-sided* *trad.*	*estim.*	*innov.*
4	9480	9553	9434	9526	9562	9467	9977	9977	9974
5	6578	6629	6471	6626	6606	6551	8857	8738	8785
6	3159	3124	3130	3258	3160	3183	5377	5181	5281
7	1367	1251	1249	1384	1293	1323	2541	2358	2393
8	466	424	453	559	507	533	998	914	966
9	191	153	179	220	192	206	406	342	381
10	66	49	69	62	53	62	128	102	131

Table 1: Number of out of control simulations [out of 10,000] for the acceptable quality level.

Ripley (1994)]. Table 1 reports for relevant values of h the number of simulations in which the V-mask $V_{0.5,h}$ yielded an out of control signal [since the availability of the full sample estimator $\bar{X}_n$ implies that we are dealing with past data, the usual evaluation of the behavior of the cumulative sums by means of the average run length is not appropriate]. The one-sided versions consider exceedance of the upper or the lower part of the V-mask only.

The numbers of out of control signals for the two-sided estimated cumulative sum are slightly less than the numbers of out of control signals for either the two-sided traditional or the two-sided innovation cumulative sum. This can be explained from the fact that the Brownian bridge is in a global sense flatter than the Wiener process. However, the flatness of the Brownian bridge shows up only marginally. Since the Brownian bridge and Wiener process have comparable local behavior [that is, the same modulus of continuity, see section 14.1 in Shorack and Wellner (1986)], this suggests that the out of control signals detected by the V-mask procedure are mainly triggered by local fluctuations.

In section 3 some simulation results for rejectable quality levels are presented. A proper interpretation of these results should take into account the slightly different performance of the estimated cumulative sum under the acceptable quality level.

3 The rejectable quality level

In this section we investigate the behavior of the estimated and innovation cumulative sums under a rejectable quality level in which the random variables $X_1, \ldots, X_n$ are still independent and normally distributed with variance 1, but do not have a common expectation θ anymore. Denote the expectation of X_i by

μ_i, and introduce

$$Z_i = X_i - \mu_i, \quad \breve{Z}_i = Z_i - \frac{1}{n+i-1}\sum_{\ell=i}^{n} Z_\ell.$$

Since we may write

$$\hat{C}_j = \sum_{i=1}^{j}\left(Z_i - \bar{Z}_n\right) + \sum_{i=1}^{j}\hat{\mu}_i$$

with

$$\hat{\mu}_i = \mu_i - n^{-1}\sum_{\ell=1}^{n}\mu_\ell.$$

it follows that the estimated cumulative sum based on the random variables $X_1, \ldots, X_n$ can be in fact decomposed into a estimated cumulative sum based on independent standard normal random variables $Z_1, \ldots, Z_n$ and a deterministic drift $\sum_{i=1}^{j}\hat{\mu}_i$. Likewise, since we may write

$$\breve{C}_j = \sum_{i=1}^{j}\breve{Z}_i\sqrt{\frac{n-i+1}{n-i}} + \sum_{i=1}^{j}\breve{\mu}_i$$

with

$$\breve{\mu}_i = \left(\mu_i - \frac{1}{n+i-1}\sum_{\ell=i}^{n}\mu_\ell\right)\sqrt{\frac{n-i+1}{n-i}},$$

it follows that the innovation cumulative sum based on the random variables $X_1, \ldots, X_n$ can be in fact decomposed into a innovation cumulative sum based on independent standard normal random variables $Z_1, \ldots, Z_n$ and a deterministic drift $\sum_{i=1}^{j}\breve{\mu}_i$.

In the remainder of this section we discuss two important rejectable quality levels in detail. Both levels have in common that $\hat{\mu}_i$ and $\breve{\mu}_i$ are respectively approximated by $\hat{\delta}(i/n)$ and $\breve{\delta}(i/n)$ for large n. It follows that the deterministic drifts $\sum_{i=1}^{j}\hat{\mu}_i$ and $\sum_{i=1}^{j}\breve{\mu}_i$ are in turn approximated by $n\hat{\Delta}(j/n)$ and $n\breve{\Delta}(j/n)$ respectively, where

$$\hat{\Delta}(t) = \int_0^t \hat{\delta}(s)ds \quad \text{and} \quad \breve{\Delta}(t) = \int_0^t \breve{\delta}(s)ds.$$

In this case the magnitude of the deterministic drifts is of the order n, whereas the magnitude of random fluctuations of the cumulative sums based on $Z_1, \ldots, Z_n$ is only of the order $\sqrt{n}$. Since the random fluctuations become neglegible with respect to the deterministic drifts as n increases, plots of $\hat{\Delta}(t)$ and $\breve{\Delta}(t)$ versus t give an impression of the patterns showing up in graphical displays of the estimated and innovation cumulative sums under a certain rejectable quality level. Let us for simplicity ignore the random fluctuations for a moment: then a V-mask $V_{f,h}$ placed at $\hat{C}_j$ or $\breve{C}_j$ essentially corresponds to a V-mask $V_{f/n,h/n}$ placed at $\hat{\Delta}(j/n)$ or $\breve{\Delta}(j/n)$. We infer that V-masks are more likely to detect patterns with sustained steep increases or decreases.

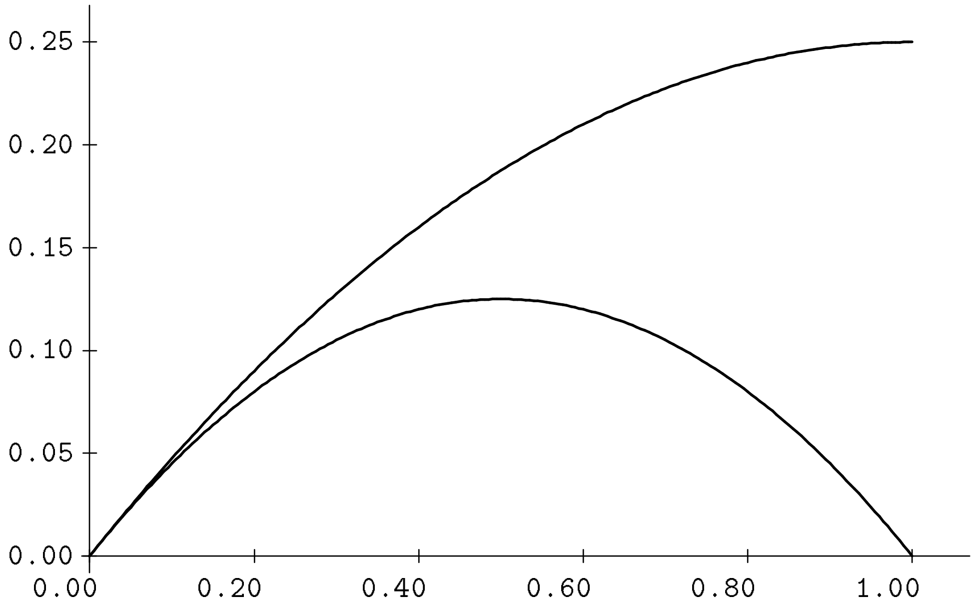

Figure 4: Plot of $\hat{\Delta}(t)$ and $\breve{\Delta}(t)$ versus t for the linear trend rejectable quality level with $\theta_0 - \theta_1 = 1$.

A linear trend rejectable quality level Under the first rejectable quality level the parameter of the quality characteristic distribution changes linearly from θ_0 to θ_1:

$$\mu_j = \theta_0 + (\theta_1 - \theta_0)\,\frac{j}{n}$$

[see also Bissel (1984, 1986)]. One may show that for large n we have $\hat{\mu}_j \approx \hat{\delta}(j/n)$ and $\breve{\mu}_j \approx \breve{\delta}(j/n)$, where

$$\hat{\delta}(t) = \frac{\theta_0 - \theta_1}{2}\,(1 - 2t)\,, \quad \text{and} \quad \breve{\delta}(t) = \frac{\theta_0 - \theta_1}{2}\,(1 - t)\,.$$

Figure 4 plots $\hat{\Delta}(t)$ and $\breve{\Delta}(t)$ versus t for $\theta_0 - \theta_1 = 1$. The slope of $\hat{\Delta}(t)$ decreases from 0.5 to -0.5, and hence the highest absolute slopes are found at both ends of the interval $[0, 1]$. The slope of $\breve{\Delta}(t)$ decreases from 0.5 to 0; although the highest absolute slopes are found only at the beginning of the interval, this is compensated by a smaller decay of absolute slope. Thus, figure 4 does not suggest a clear advantage of one cumulative sum over the other.

To investigate the behavior of the cumulative sums under the rejectable quality level, we used the samples simulated in section 2 to construct 10,000 random samples of size 1000 obeying the linear trend rejectable quality level with $\theta_0 - \theta_1 = 1$. Table 2 reports for relevant values of h the number of simulations in which the V-mask $V_{0.5,h}$ yielded an out of control signal. Although not decidedly better, the two-sided innovation cumulative sum seems to perform at least as

h	*upper* *trad.*	*estim.*	*innov.*	*lower* *trad.*	*estim.*	*innov.*	*two-sided* *trad.*	*estim.*	*innov.*
4	9998	9999	10000	9999	10000	5357	10000	10000	10000
5	9945	9965	10000	9959	9977	1949	10000	10000	10000
6	9646	9739	9991	9658	9728	640	9990	9985	9991
7	8891	9077	9887	8890	9041	226	9867	9864	9890
8	7863	8056	9591	7815	7962	65	9521	9458	9593
9	6697	6725	9010	6594	6633	26	8893	8718	9013
10	5505	5505	8240	5490	5424	4	8000	7734	8241

Table 2: Number of out of control simulations [out of 10,000] for the linear trend rejectable quality level with $\theta_0 = 0.5$ and $\theta_1 = -0.5$.

good as the two-sided estimated cumulative sum. Observe that virtually all out of control signals of the two-sided innovation cumulative sum are generated by the upper arm of the V-mask.

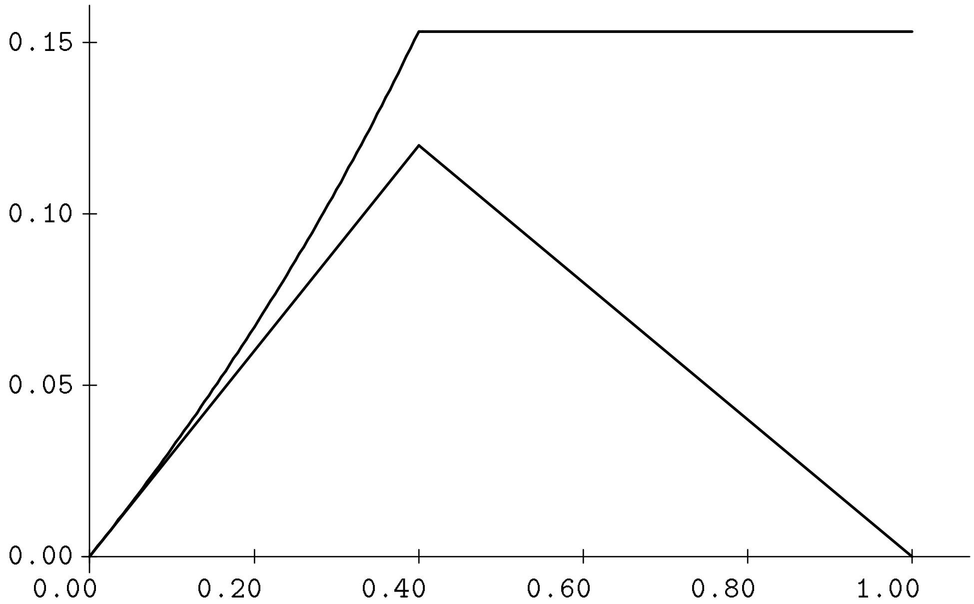

Figure 5: Plot of $\hat{\Delta}(t)$ and $\breve{\Delta}(t)$ versus t for the change point rejectable quality level with $p = 0.4$ and $\theta_0 - \theta_1 = 0.5$.

A change point rejectable quality level Under the second rejectable quality level the parameter of the quality characteristic distribution jumps suddenly from

θ_0 to θ_1 after $[np]$ observations:

$$\mu_j = \begin{cases} \theta_0 & \text{if } j = 1, \ldots, [np], \\ \theta_1 & \text{if } j = [np]+1, \ldots, n \end{cases}$$

[see also Page (1957)]. If the relative position p of the jump remains fixed, then one may show that for large n we have $\hat{\mu}_j \approx \hat{\delta}(j/n)$ and $\breve{\mu}_j \approx \breve{\delta}(j/n)$, where

$$\hat{\delta}(t) = \begin{cases} (1-p)(\theta_0 - \theta_1) & \text{if } t \leq p, \\ p(\theta_1 - \theta_0) & \text{if } t > p, \end{cases} \quad \text{and} \quad \breve{\delta}(t) = \begin{cases} \dfrac{1-p}{1-t}(\theta_0 - \theta_1) & \text{if } t \leq p, \\ 0 & \text{if } t > p. \end{cases}$$

Figure 5, which plots $\hat{\Delta}(t)$ and $\breve{\Delta}(t)$ versus t for $p = 0.4$ and $\theta_0 - \theta_1 = 0.5$, illustrates that $\breve{\Delta}(t)$ reaches its steepest slope $\theta_0 - \theta_1$ just before p. The slope of $\breve{\Delta}(t)$ is either $(1-p)(\theta_0 - \theta_1)$ or $p(\theta_1 - \theta_0)$, and thus less steep. This indicates that for sufficiently large n the innovation cumulative sums should perform better than the estimated cumulative sum in detecting change points. Moreover, the better performance is especially expected to materialize for values of p in the vicinity of 0.5; that is, for change points occurring somewhere near the middle of sample.

To investigate the behavior of the cumulative sums under a change point rejectable quality level, we performed simulation studies for nine different values of p. In each simulation study the samples simulated in section 2 were used to construct 10,000 random samples of size 1000 obeying the change point rejectable quality level with $\theta_0 - \theta_1 = 0.5$. Tables 3–11 report for relevant values of h the number of simulations in which the V-mask $V_{0.5,h}$ yielded an out of control signal.

	upper			*lower*			*two-sided*		
h	*trad.*	*estim.*	*innov.*	*trad.*	*estim.*	*innov.*	*trad.*	*estim.*	*innov.*
4	9964	9976	9986	9762	9818	9289	9999	9999	9999
5	9536	9607	9759	7686	7768	6136	9893	9892	9915
6	8442	8531	8963	4474	4427	2899	9127	9115	9271
7	7169	7252	7850	2113	2030	1196	7791	7747	8118
8	5909	5916	6708	980	898	482	6327	6258	6876
9	4752	4731	5617	404	357	181	4957	4907	5705
10	3783	3737	4618	164	144	56	3887	3833	4655

Table 3: Number of out of control simulations [out of 10,000] for the rejectable quality level with $\theta_0 = 0.5(1-p)$, $\theta_1 = -0.5p$ and $p = 0.1$.

h	*upper*			*lower*			*two-sided*		
	trad.	*estim.*	*innov.*	*trad.*	*estim.*	*innov.*	*trad.*	*estim.*	*innov.*
4	9985	9996	10000	9892	9936	9036	10000	10000	10000
5	9834	9857	9967	8589	8685	5667	9974	9966	9985
6	9269	9399	9762	5741	5720	2620	9673	9701	9827
7	8239	8355	9253	3155	3040	1068	8787	8784	9338
8	6997	7119	8432	1603	1501	439	7461	7480	8498
9	5824	5882	7574	774	703	161	6146	6133	7618
10	4691	4666	6643	353	306	51	4887	4813	6658

Table 4: Number of out of control simulations [out of 10,000] for the rejectable quality level with $\theta_0 = 0.5(1-p)$, $\theta_1 = -0.5p$ and $p = 0.2$.

h	*upper*			*lower*			*two-sided*		
	trad.	*estim.*	*innov.*	*trad.*	*estim.*	*innov.*	*trad.*	*estim.*	*innov.*
4	9995	9997	10000	9962	9975	8780	10000	10000	10000
5	9903	9923	9993	9214	9349	5218	9990	9991	9998
6	9327	9457	9909	7017	7050	2356	9800	9790	9936
7	8199	8355	9605	4396	4332	929	8971	8951	9649
8	6855	6934	9003	2504	2373	375	7640	7539	9040
9	5480	5471	8204	1400	1219	135	6131	5947	8235
10	4200	4124	7206	683	636	42	4607	4472	7224

Table 5: Number of out of control simulations [out of 10,000] for the rejectable quality level with $\theta_0 = 0.5(1-p)$, $\theta_1 = -0.5p$ and $p = 0.3$.

h	*upper*			*lower*			*two-sided*		
	trad.	*estim.*	*innov.*	*trad.*	*estim.*	*innov.*	*trad.*	*estim.*	*innov.*
4	9996	9997	10000	9984	9990	8378	10000	10000	10000
5	9837	9895	9996	9573	9642	4706	9990	9990	9997
6	9081	9222	9929	7997	8106	2038	9831	9799	9944
7	7719	7838	9659	5700	5698	797	9042	8931	9689
8	6079	6056	9107	3676	3550	310	7530	7280	9131
9	4469	4409	8224	2217	2091	110	5700	5427	8243
10	3160	3078	7208	1284	1134	34	4048	3805	7221

Table 6: Number of out of control simulations [out of 10,000] for the rejectable quality level with $\theta_0 = 0.5(1-p)$, $\theta_1 = -0.5p$ and $p = 0.4$.

h	*upper*			*lower*			*two-sided*		
	trad.	*estim.*	*innov.*	*trad.*	*estim.*	*innov.*	*trad.*	*estim.*	*innov.*
4	9989	9999	10000	9991	9993	7934	10000	10000	10000
5	9728	9817	9996	9755	9808	4188	9995	9993	9999
6	8678	8852	9917	8679	8825	1732	9818	9783	9929
7	6897	6920	9621	6895	6977	673	9048	8903	9653
8	4963	4851	8906	4936	4840	271	7424	7136	8932
9	3292	3181	7942	3324	3211	94	5506	5220	7955
10	2152	1997	6913	2201	2018	28	3883	3522	6920

Table 7: Number of out of control simulations [out of 10,000] for the rejectable quality level with $\theta_0 = 0.5(1-p)$, $\theta_1 = -0.5p$ and $p = 0.5$.

h	*upper*			*lower*			*two-sided*		
	trad.	*estim.*	*innov.*	*trad.*	*estim.*	*innov.*	*trad.*	*estim.*	*innov.*
4	9986	9993	10000	9999	10000	7518	10000	10000	10000
5	9521	9626	9992	9849	9889	3707	9994	9991	9997
6	7950	8058	9895	9167	9265	1488	9802	9787	9918
7	5705	5678	9455	7768	7864	557	9023	8930	9489
8	3661	3540	8603	6098	6116	218	7487	7298	8627
9	2185	2056	7495	4525	4438	76	5715	5453	7509
10	1253	1122	6324	3201	3077	25	4049	3808	6330

Table 8: Number of out of control simulations [out of 10,000] for the rejectable quality level with $\theta_0 = 0.5(1-p)$, $\theta_1 = -0.5p$ and $p = 0.6$.

h	*upper*			*lower*			*two-sided*		
	trad.	*estim.*	*innov.*	*trad.*	*estim.*	*innov.*	*trad.*	*estim.*	*innov.*
4	9940	9966	10000	9993	9994	7113	10000	10000	10000
5	9154	9299	9979	9899	9910	3261	9986	9983	9984
6	6965	7052	9786	9340	9440	1262	9769	9778	9810
7	4408	4309	9118	8189	8320	448	8980	8951	9159
8	2518	2397	8046	6747	6848	164	7547	7509	8079
9	1308	1190	6629	5327	5308	62	5917	5770	6653
10	623	552	5320	4122	4045	19	4475	4325	5328

Table 9: Number of out of control simulations [out of 10,000] for the rejectable quality level with $\theta_0 = 0.5(1-p)$, $\theta_1 = -0.5p$ and $p = 0.7$.

	upper			*lower*			*two-sided*		
h	*trad.*	*estim.*	*innov.*	*trad.*	*estim.*	*innov.*	*trad.*	*estim.*	*innov.*
4	9893	9929	9999	9992	9995	7144	10000	10000	10000
5	8538	8653	9951	9851	9891	3174	9981	9982	9968
6	5764	5708	9491	9277	9335	1193	9692	9672	9553
7	3154	3039	8378	8254	8327	418	8792	8766	8442
8	1506	1415	6894	7030	7098	144	7468	7439	6932
9	714	603	5317	5773	5785	53	6087	6004	5342
10	332	268	4025	4632	4600	15	4823	4722	4034

Table 10: Number of out of control simulations [out of 10,000] for the rejectable quality level with $\theta_0 = 0.5(1-p)$, $\theta_1 = -0.5p$ and $p = 0.8$.

	upper			*lower*			*two-sided*		
h	*trad.*	*estim.*	*innov.*	*trad.*	*estim.*	*innov.*	*trad.*	*estim.*	*innov.*
4	9731	9816	9994	9960	9971	7881	9999	9999	9998
5	7684	7760	9733	9477	9540	3779	9863	9854	9828
6	4434	4393	8402	8405	8466	1439	9100	9047	8620
7	2072	2021	6394	7073	7191	529	7674	7716	6565
8	908	793	4557	5816	5846	170	6211	6147	4646
9	356	319	3080	4739	4660	57	4927	4799	3114
10	158	123	2065	3811	3736	12	3900	3801	2073

Table 11: Number of out of control simulations [out of 10,000] for the rejectable quality level with $\theta_0 = 0.5(1-p)$, $\theta_1 = -0.5p$ and $p = 0.9$.

For the two-sided estimated cumulative sum the number of out of control simulations shows a symmetric picture: it reaches a maximum both before and after $p = 0.5$, and at $p = 0.5$ a local minimum; the locations of the maxima divert from $p = 0.5$ as h increases. For the two-sided innovation cumulative sum the number of out of control simulations shows a less symmetric picture: it increases until a maximum is reached somewhere between $p = 0.3$ and $p = 0.4$, and then steadily decreases. For $p \leq 0.7$ the innovation cumulative sum clearly performs better, for $p \geq 0.8$ the estimated cumulative sum. Sensitivity to early jumps is in some applications an especially attractive property, for instance in detecting initialization bias in simulation output [see Schruben (1982)].

4 Extension to general distributions

In this section we extend the normal distribution theory of the previous section to general distributions. Consider random variables $X_1, X_2, \ldots, X_n$, which under the acceptable quality level are independent and have common continuous probability density function $f_\theta(x)$. Assume that $f_\theta(x)$ satisfies certain regularity conditions guaranteeing consistency and asymptotic normality of the maximum likelihood estimators [for instance, conditions (R1), (R2) and (R3) in paragraph 4.2.2 of Serfling (1980)].

If θ is known, then we are able to generalize the traditional cumulative sum given by

$$C_j = \sum_{i=1}^{j} \rho_\theta(X_i),$$

where $\rho_\theta(t)$ is the classical score function, defined by

$$\rho_\theta(t) = \left.\frac{\partial}{\partial\vartheta}\right|_{\vartheta=\theta} \log f_\vartheta(t).$$

This generalization of the traditional cumulative sum is inspired by Johnson (1961), in which the connection between the sequential probability ratio test and cumulative sum charts is observed.

Martingale theory implies that the stochastic proces $\left(n^{-1/2}C_{[nt]}\right)_{t\in[0,1]}$ converges weakly to a standard Wiener process under the simple acceptable quality level.

If θ is unknown, the maximum likelihood estimator $\hat{\theta}^n$ of θ is found by solving the maximum likelihood equation

$$\left.\frac{\partial}{\partial\vartheta}\right|_{\vartheta=\hat{\theta}^n} \log f_\theta(X_i) = 0. \tag{1}$$

Replacing θ in the traditional cumulative sum by the maximum likelihood estimator $\hat{\theta}^n$ yields the estimated cumulative sum

$$\hat{C}_j = \sum_{i=1}^{j} \rho_{\hat{\theta}^n}(X_i)$$

The replacement of θ by the maximum likelihood estimator $\hat{\theta}^n$ changes the behavior of the cumulative sum in a similar way as we have seen before, since the maximum likelihood equations imply that $\hat{C}_n$ is degenerate in zero.

Mathematically, the stochastic process $\left(n^{-1/2}\hat{C}_{[nt]}\right)_{t\in[0,1]}$ converges weakly to a standard Brownian bridge under the composite acceptable quality level.

Again, the source of the difference in behavior of the traditional and the estimated cumulative sum is that knowledge of $\hat{\theta}^n$ and $X_1, X_2, \ldots, X_{j-1}$ yields information about the random variable X_j.

To make this more explicit, introduce the random variables

$$Y_i = \Sigma^{1/2}\left(\theta - \Sigma^{-1}\rho_\theta(X_i)\right),$$

where

$$\Sigma = \int_{-\infty}^{\infty} \rho_\theta(x)\rho_\theta(x)^T f_\theta(x)dx$$

is the Fisher information matrix. Observe that $Y_1, Y_2, \ldots, Y_n$ are independent random variables. Standard likelihood theory implies that each of these random variables has expectation zero and covariance matrix I. Moreover, we typically have

$$Y_i \approx \Sigma^{1/2}\left(\hat{\theta}^n - \Sigma^{-1}\rho_{\hat{\theta}^n}(X_i)\right) \tag{2}$$

[for instance, under the regularity conditions in paragraph 4.2.2 of Serfling (1980) the maximum likelihood equation (1) implies

$$\left|Y_i - \Sigma^{1/2}\left(\hat{\theta}^n - \Sigma^{-1}\rho_{\hat{\theta}^n}(X_i)\right)\right| \leq H(X_i)\left(\hat{\theta}^n - \theta\right)^2$$

for $\hat{\theta}^n$ in $N(\theta)$, where $H(X_i)$ is a random variable with finite expectation; now Markov's inequality yields (2)].

It follows from (2) and the maximum likelihood equation (1) that the maximum likelihood estimator $\hat{\theta}^n$ may be approximated by $\Sigma^{-1/2}\bar{Y}_n$, where $\bar{Y}_n$ denotes the sample mean of $Y_1, Y_2, \ldots, Y_n$.

We now temporarily ignore the fact that $\hat{\theta}^n$ and $\Sigma^{-1/2}\bar{Y}_n$ differ slightly. Observe that knowledge of $\Sigma^{-1/2}\bar{Y}_n$ and $Y_1, Y_2, \ldots, Y_{i-1}$ yields information about the random variable Y_i, since $\sum_{\ell=i}^n Y_\ell = n\bar{Y}_n - \sum_{\ell=1}^{i-1} Y_i$ and Y_j are not independent. Inspired by the fact that this resembles the situation in section 2, we introduce the random variable

$$\breve{Y}_i = Y_i - \mathcal{E}\left(Y_i \mid \sum_{\ell=i}^n Y_\ell\right) = Y_i - \frac{1}{n-i+1}\sum_{\ell=i}^n Y_\ell$$

[the second equality follows from paragraph 9.11 in Williams (1992)]. Observe that $\breve{Y}_i$ is in fact the residual obtained by regressing Y_i on $\sum_{\ell=i}^n Y_\ell$. It follows that $\breve{Y}_i$ and $\sum_{\ell=i}^n Y_\ell$ are uncorrelated [but not necessarily independent]. Hence, we may decompose Y_i in a part $\breve{Y}_i$ which is uncorrelated with $\sum_{\ell=i}^n Y_\ell$ and a part $(n-i+1)^{-1}\sum_{\ell=i}^n Y_\ell$ which is completely determined by $\sum_{\ell=i}^n Y_\ell$. Finally, observe that

$$\operatorname{cov}\left(\breve{Y}_i, \breve{Y}_k\right) = \begin{cases} \dfrac{n-i}{n-i+1} I & \text{if } i = k, \\ 0 & \text{if } i \neq k. \end{cases}$$

Hence, the random variables

$$\breve{Y}_1\sqrt{\frac{n}{n-1}}, \breve{Y}_2\sqrt{\frac{n-1}{n-2}}, \ldots, \breve{Y}_{n-1}\sqrt{\frac{2}{1}}$$

are uncorrelated, and have expectation zero and covariance matrix I. The traditional cumulative sum based on these random variables is given by

$$\sum_{i=1}^{j} \breve{Y}_i \sqrt{\frac{n-i+1}{n-i}}.$$

Due to the fact that Y_i is an approximation to $\Sigma^{1/2}\left(\hat{\theta}^n - \Sigma^{-1}\rho_{\hat{\theta}^n}(X_i)\right)$, we have that

$$\Sigma^{-1/2}\left(\left\{\frac{1}{n-i+1}\sum_{\ell=i}^{n}\rho_{\hat{\theta}^n}(X_\ell)\right\} - \rho_{\hat{\theta}^n}(X_i)\right)$$

approximates $\breve{Y}_i$. The innovation cumulative sum is defined by

$$\breve{C}_j = \sum_{i=1}^{j}\Sigma^{-1/2}\left(\left\{\frac{1}{n-i+1}\sum_{\ell=i}^{n}\rho_{\hat{\theta}^n}(X_\ell)\right\} - \rho_{\hat{\theta}^n}(X_i)\right)\sqrt{\frac{n-i+1}{n-i}}.$$

Mathematically, the stochastic process $\left(n^{-1/2}\breve{C}_{[nt]}\right)_{t\in[0,1]}$ converges weakly to a standard Wiener process under the composite acceptable quality level.

If $f_\theta(t)$ is the density belonging to a normal distribution with expectation θ and variance 1, then the classical score function equals $\theta - t$, Σ is the one-dimensional identity matrix, and hence

$$\Sigma^{1/2}\left(\hat{\theta}^n - \Sigma^{-1}\rho_{\hat{\theta}^n}(X_i)\right) = X_i.$$

Thus, the definition of $\breve{C}_j$ given in this section coincides with the one given in section 2.

In some applications, it may be necessary to replace Σ in the definition of $\breve{C}_j$ by its estimator

$$\int_{-\infty}^{\infty}\rho_{\hat{\theta}^n}(x)\rho_{\hat{\theta}^n}(x)^T f_{\hat{\theta}^n}(x)dx.$$

The effects of this replacement vanish as n grows large.

References

[1] Andersen, P.K., Borgan, Ø., Gill, R.D., Keiding, N. (1993). *Statistical Models Based on Counting Processes.* Springer-Verlag, Berlin.

[2] Bissell, A.F. (1984). The performance of control charts and cusums under linear trend. *Applied Statistics* **33**, 145–151.

[3] Bissell, A.F. (1986). "Corrigendum" (to Bissell, 1984). *Applied Statistics* **35**, 214.

[4] Bissell, A.F. (1994). *Statistical Methods for SPC and TQM.* Texts in Statistical Science. Chapman and Hall, London.

[5] Dobben de Bruyn, C.S. (1968). *Cumulative Sum Tests: Theory and Practice.* Griffin's Statistical Monographs and Courses **24**. Hafner, New York.

[6] Durbin, J. (1973). *Distribution Theory for Tests based on the Sample Distribution Function.* SIAM, Philadelphia.

[7] Ishikawa, K. (1982). *Guide to Quality Control.* Asian Productivity Organization, Tokyo.

[8] Johnson, N.L. (1961). A simple theoretical approach to cumulative sum charts. *J. Amer. Stat. Assoc.* **56**, 835–840.

[9] Khmaladze, E.V. (1981). Martingale approach in the theory of goodness of fit tests. *Theory Probab. Appl.* **26**, 246–265.

[10] Khmaladze, E.V. (1993). Goodness of fit problems and scanning innovation martingales. *Ann. Statist.* **21**, 798–829.

[11] Lucas (1982). Combined Shewhart-CUSUM quality control schemes. *Journal of Quality Technology* **14**, 51–59.

[12] Lucas, J.M., Crosier, R.B. (1982). Fast initial response for CUSUM quality control schemes: Give your CUSUM a head start. *Technometrics* **24**, 199–205.

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

[14] Page, E.S. (1957). On problems in which a change in a parameter occurs at an unknown point. *Biometrika* **44**, 248–252.

[15] Schruben, L.W. (1982). Detecting initialization bias in simulation output. *Oper. Res.* **30**, 569–590.

[16] Schruben, L.W. (1983). Confidence interval estimation using standardized time series. *Oper. Res.* **31**, 1090–1108.

[17] Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

[18] Shewhart, W.A. (1931). *Economic Control of Quality Manufactured Product.* Van Nostrand, New York.

[19] Shorack, G.R., Wellner, J.A. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.

[20] Venables, W.N., Ripley, B.D. (1994). *Modern Applied Statistics with S-Plus.* Springer-Verlag, New York.

[21] Wetherill, G.B, Brown, D.W. (1991). *Statistical Process Control. Theory and practice.* Chapman and Hall, London.

[22] Williams, D. (1991). *Probability with Martingales.* Cambridge Mathematical Textbooks. Cambridge University Press.