

The Changepoint Model for Statistical Process Control

DOUGLAS M. HAWKINS and PEIHUA QIU

University of Minnesota, Minneapolis, MN 55455

CHANG WOOK KANG

Hanyang University, Seoul, Korea

Statistical process control (SPC) requires statistical methodologies that detect changes in the pattern of data over time. The common methodologies, such as Shewhart, cumulative sum (cusum), and exponentially weighted moving average (EWMA) charting, require the in-control values of the process parameters, but these are rarely known accurately. Using estimated parameters, the run length behavior changes randomly from one realization to another, making it impossible to control the run length behavior of any particular chart. A suitable methodology for detecting and diagnosing step changes based on imperfect process knowledge is the unknown-parameter changepoint formulation. Long recognized as a Phase I analysis tool, we argue that it is also highly effective in allowing the user to progress seamlessly from the start of Phase I data gathering through Phase II SPC monitoring. Despite not requiring specification of the post-change process parameter values, its performance is never far short of that of the optimal cusum chart which requires this knowledge, and it is far superior for shifts away from the cusum shift for which the cusum chart is optimal. As another benefit, while changepoint methods are designed for step changes that persist, they are also competitive with the Shewhart chart, the chart of choice for isolated non-sustained special causes.

Introduction

WITH statistical process control (SPC), one aims to detect and diagnose situations in which a process has gone out of statistical control. This problem has both process and statistical aspects to it—for an outline of some of its statistical modeling aspects, see Crowder et al. (1997). Operationally, the state of statistical control may be described as one in which the process readings appear to follow a common statistical model. One model is that while the process is in statistical control, the successive process readings X_i are independent and sampled from the same

distribution. When the process goes out of control, it can do so in several ways. A distinction can be made between *isolated* special causes, those that affect a single process reading and then disappear, and *sustained* special causes, those that continue until they are identified and fixed. In statistical terms, an isolated special cause results in a single process reading (or a single rational subgroup) that appears to come from some distribution other than the in-control distribution. It is analogous to an outlier. The Shewhart \bar{X} with an R - or S -chart is an excellent tool for detecting special causes that lead to large changes, whether sustained or isolated. It is less effective for diagnosing small changes in the process.

A sustained special cause changes the statistical distribution from the in-control distribution to something else, and the distribution will remain in this out-of-control state until some corrective action is taken. For example, the mean of the process readings could change to a different value, or the variability could change. Standard tools for detecting sustained changes are the cumulative sum (cusum)

Dr. Hawkins is a Professor in the School of Statistics. He is a Fellow of ASQ. His email address is doug@stat.umn.edu.

Dr. Peihua Qiu is an Associate Professor in the School of Statistics. His email address is qiu@stat.umn.edu.

Dr. Chang Wook Kang is an Associate Professor in the Department of Industrial Engineering. His email address is cwkang57@hotmail.com.

chart and the exponentially weighted moving average (EWMA) chart. This paper focuses on another, less familiar, method aimed at detecting sustained changes, the change-point formulation.

Making the model more specific, we suppose that the process readings can be modeled by two normal distributions,

$$\begin{aligned} X_i &\sim N(\mu_1, \sigma_1^2) \text{ for } i = 1, 2, \dots, \tau; \\ X_i &\sim N(\mu_2, \sigma_2^2) \text{ for } i = \tau + 1, \dots, n. \end{aligned}$$

The series length n is fixed in traditional statistical settings, but increases without limit in Phase II SPC settings. Both settings will be discussed, with context indicating which of the two applies. The in-control distribution is $N(\mu_1, \sigma_1^2)$. The readings follow this distribution up to some instant τ , the *change-point*, at which point they switch to another normal distribution differing in mean ($\mu_1 \neq \mu_2$), in variability ($\sigma_1 \neq \sigma_2$), or in both mean and variability. Using this changepoint model to describe process readings leads to two statistical tasks; a testing task and an estimation task. The testing task is to decide whether there has indeed been a change. If so, the estimation task is to estimate τ , the time at which it occurred, and perhaps also to estimate some or all of the parameters μ_1 , μ_2 , σ_1 , and σ_2 .

With few exceptions, work on changepoints has focused on shifts in mean only, i.e., $\mu_1 \neq \mu_2$ but $\sigma_1 = \sigma_2 = \sigma$, and this framework is used here also.

Within this changepoint model are three scenarios based on the amount of process knowledge: (1) all parameters μ_1 , μ_2 , and σ are known exactly *a priori*; (2) the in-control parameters μ_1 and σ are known exactly, but μ_2 is not; and (3) all parameters μ_1 , μ_2 , and σ are unknown.

First Scenario—the Cusum Chart

The cusum chart for location is appropriate for the first of these settings, where all the parameters except the changepoint are known. We assume that $\mu_2 > \mu_1$, the case $\mu_2 < \mu_1$ being handled analogously. The cusum chart is then defined by

$$\begin{aligned} S_0 &= 0 \\ S_i &= \max(0, S_{i-1} + X_i - k), \end{aligned}$$

where the ‘reference value’ $k = (\mu_1 + \mu_2)/2$.

The cusum chart signals a shift if $S_i > h$, where h is the ‘decision interval,’ which is chosen to set the in-control average run length (ARL) at some acceptable level. If a shift is signaled, then the most recent

epoch j at which $S_j = 0$ gives the maximum likelihood estimate of the last instant before the mean shift (Lai, 2001).

The cusum method has attractive theoretical properties in that, in a certain sense, it is the optimal test for a shift in mean from μ_1 to μ_2 . See Hawkins and Olwell (1998) for a more detailed discussion of cusum methods, their properties and further references.

Note that we need to know all three parameters to set up the cusum chart. We need to know μ_1 and μ_2 to calculate the reference value, and μ_1 and σ to calculate the decision interval h .

We do not consider the EWMA procedure. It also requires known in-control parameters, also involves a tuning constant and a control limit, and if properly tuned gives performance close to that of the cusum chart. Our later comments on cusum charts largely carry over to EWMA methods.

Second Scenario—Cusum and GLR Changepoint

It is rare indeed that there is only one value of μ_2 to which the mean might shift after a change, and so the optimality properties of the cusum chart are less attractive than they might seem. However, although strictly optimal only for the particular shift corresponding to its reference value, the cusum chart is close to optimal for the range of shifts close to that for which it is optimal. Common practice is therefore to design the cusum chart for some shift just large enough to be thought detectable and practically significant, and to rely on this near-optimality argument to believe that this cusum chart will be a good choice for all μ_2 values likely to occur.

An alternative is the generalized likelihood ratio (GLR) approach discussed in Lai (2001) for the situation μ_1 , σ known, but μ_2 unknown. In this case, the unknowns μ_2 and τ are estimated by maximum likelihood, and then the maximized likelihood gives a GLR test for the presence of a changepoint against the null hypothesis of a single unbroken sequence. This same modeling framework was used by Pignatiello and Samuel (2001). This latter work differs from the GLR approach in being an ‘add-on’ to conventional charting methods such as Shewhart, cusum, or EWMA. The conventional method is used to decide that a change has occurred, and then the changepoint likelihood is used for the followup problem of estimating τ and μ_2 . This differs from Lai’s

self-contained GLR in which the maximized likelihood and the likelihood ratio are used for both detection and estimation.

Gombay (2000) extended this framework to one allowing for additional unknown ‘nuisance’ parameters. This leads to a procedure in which σ can be unknown. However, the in-control mean μ_1 still needs to be known *a priori* in order to have known statistical properties.

The assumption of known in-control mean and standard deviation underlies both the standard charting methods (Shewhart, cusum, and EWMA charts), and the GLR changepoint approach set out by Lai (2001), Pignatiello and Samuel (2001), and Nedumaran and Pignatiello (2001). Where do these known parameter values come from? The values used are generally not exact parameter values, but estimates obtained in a Phase I study. Any errors in these quantities, including the random errors that are inevitable if the μ_1 and σ values are estimated, lead to an inability to fix the in-control run length properties of the more sensitive methods such as the cusum, EWMA, and GLR.

To illustrate this point, we suppose that a Phase I sample of size 100 is used, and its mean and standard deviation substituted for μ_1 and σ . Let \bar{X} take one of three values: the true value μ_1 ; a ‘low’ value one standard error below μ_1 ; or a ‘high’ value one standard error above μ_1 . We do the same for the sample standard deviation. We consider a cusum chart designed for a one-sigma upward shift in mean and an in-control ARL 500. Then, the actual ARL values for these possible estimates of μ_1 and σ (calculated using the public-domain software on www.stat.umn.edu/users/cusum) are given in Table 1.

The top left and bottom right corners of this table show ARLs a factor of three different from the nominal 500. Since a random error of at least one standard error occurs in about a third of the samples, this means that the unavoidable random errors in the estimates make even reasonably accurate targeting of the in-control ARL impossible. While it is not essential to get the ARL exactly right for a chart to be useful, it is surely a cause of concern that the false alarm rate could easily, uncontrollably, and undetectably vary by an order of magnitude.

The framework of ‘exactly known in-control parameters’ is then more of a convenient fiction than fact. When only rather insensitive methods such as

TABLE 1. The Actual ARL of the Cusum Procedure With Estimated Mean and Standard Deviation

		\bar{X}		
		low	true	high
s	low	190	300	500
	true	290	500	900
	high	450	850	1680

the Shewhart chart were used, having only estimates was not a source of much concern, but this is no longer true when we add sensitive procedures such as cusum charts, EWMA charts, and the GLR method to the mix.

Recognition that estimating parameters affects run length behavior is not new. Quesenberry (1993) noted that the run length behavior of Shewhart \bar{X} charts using estimates of the in-control mean and standard deviation could differ substantially from the known-parameter case. Nedamaram and Pignatiello (2001) suggested modifications to the control limits to account for the change in false alarm probability, and Jones and Champ (2001) gave some discussion in the context of the EWMA chart.

Third Scenario—None of the Parameters Known

Finally, consider the model with none of the parameters known. We can test for the presence of a changepoint with another generalized likelihood ratio test. This test is a ‘two-sample *t* test’ between the left and right sections of the sequence, maximized across all possible changepoints (Sen and Srivastava (1975), Hawkins (1977), and Worsley (1979)). For a given putative changepoint j where $1 \leq j \leq n-1$, we let $\bar{X}_{jn} = \sum_{i=1}^j X_i/j$ be the mean of the first j observations, $\bar{X}_{jn}^* = \sum_{i=j+1}^n X_i/(n-j)$ be the mean of the remaining $n-j$ observations, and $V_{jn} = \sum_{i=1}^j (X_i - \bar{X}_j)^2 + \sum_{i=j+1}^n (X_i - \bar{X}_j^*)^2$ be the residual sum of squares.

Based on the assumption that there is a single changepoint at instant j , \bar{X}_{jn} and \bar{X}_{jn}^* are the maximum likelihood estimators (MLEs) of μ_1 and μ_2 , and $\hat{\sigma}_{jn}^2 = V_{jn}/(n-2)$ is the usual pooled estimator of σ^2 . A conventional two-sample *t*-statistic for comparing the two segments would be

$$T_{jn} = \sqrt{\frac{j(n-j)}{n}} \frac{\bar{X}_{jn} - \bar{X}_{jn}^*}{\hat{\sigma}_{jn}}$$

In the null case that there was no changepoint and j was chosen arbitrarily, T_{jn} would follow a Student's t -distribution with $n - 2$ degrees of freedom.

The generalized likelihood ratio test for the presence of a changepoint consists of finding $T_{\max,n}$, the maximum of $|T_{jn}|$ over all $1 \leq j \leq n-1$. The j giving the maximum is the MLE of the true changepoint τ , and the corresponding \bar{X}_j and \bar{X}_j^* are MLEs of the unknown means μ_1 and μ_2 . If $T_{\max,n}$ exceeds some critical value h_n , then we conclude that there was indeed a shift. Otherwise, we conclude that there is not sufficient evidence of a shift.

Finding suitable critical values h_n is not a trivial task. A very easy conservative bound for the cutpoint giving a significance level of no more than α comes from the Bonferroni inequality

$$\begin{aligned} Pr[T_{\max,n} > h_n] &\leq \sum_{j=1}^{n-1} Pr[|T_{jn}| > h_n] \\ &= (n-1)Pr[|t_{n-2}| > h_n], \end{aligned}$$

which is $(n-1)$ times the two-sided tail area of a t -distribution, with $n-2$ degrees of freedom. Thus, choosing for h_n the two-sided $\alpha/(n-1)$ fractile of a t -distribution with $n-2$ degrees of freedom will give a test of size at most α . This simple Bonferroni inequality is exact (Worsley (1979)) if

$$h_n^2 \geq \begin{cases} \frac{n-2}{2}[n + \sqrt{(n^2-4)}], & \text{if } n \text{ is even} \\ n(n-2), & \text{if } n \text{ is odd.} \end{cases}$$

This condition is met for small n and small α combinations, but the Bonferroni bound becomes very conservative if n is large. Worsley (1982) gave a much tighter conservative bound that could be used for moderate-size n , and large-sample asymptotics were given by Irvine (1982).

We also mention briefly the more general changepoint formulation in which either or both of the parameters μ, σ may change at the changepoint τ . Sullivan and Woodall (1996) discussed this formulation and the resultant generalized likelihood ratio test. It has the advantage of providing a single diagnostic that can be used to detect shifts in either the mean or the variance, or in both. It has the disadvantage of being much more sensitive to the normality assumption than is $T_{\max,n}$. Furthermore, while bounds, approximations, and extreme value results are known for the null distribution of $T_{\max,n}$, there is hardly any finite-sample theory for Sullivan and Woodall's statistic. For these reasons, we will not consider it

further here. We use the term 'changepoint formulation' to mean the formulation in this third scenario, in which none of the process parameters is considered known exactly.

Phase I and Phase II Problems

We have so far assumed that we had a single sample of size n . Turning to SPC applications, standard practice involves two measurement phases. In Phase I, a set of process data is gathered and analyzed. Any unusual 'patterns' in this data set indicate a lack of statistical control and lead to adjustments and fine tuning. It is common, for example, that the early readings will be more variable than the later as a result of adjustments and fine-tuning. There may be outliers, indicative of isolated special causes; these too will be diagnosed and steps taken to prevent their recurrence. One or more changepoints in mean is another possibility.

Once all such assignable causes are accounted for though, we will be left with a clean set of data, gathered under stable operating conditions and illustrative of the actual process performance. This set is then used to estimate the in-control distribution of X , including its mean μ_1 , standard deviation σ , and form.

The changepoint formulation has not been used extensively in the SPC context, but where it has, has tended to be confined to Phase I problems. This, for example, is the setting of the Sullivan and Woodall (1996) proposal for finding changepoints in the mean and/or variance.

In Phase I, with its static set of data X_1, \dots, X_n , traditional fixed-sample statistical methods are appropriate. So, for example, it is appropriate to calculate $T_{\max,n}$ for the whole data set and test it against a suitable fractile of the null distribution of the test statistic for that value of n . If the analysis indicates a lack of control in the Phase I data set, more data will be gathered after process adjustment until a clean data set is achieved.

Phase II data are the process readings gathered subsequently. Unlike the fixed set of Phase I, they form a never-ending stream. As each new reading accrues, the SPC check is re-applied. For this purpose, fixed significance level control limits are not appropriate; rather, concern is with the run lengths, both in- and out-of-control. A convenient summary of the frequency of false alarms is the in-control average run length (ARL), which should be large, and

a method's performance can be summarized by its out-of-control ARL values.

Traditional methods such as the Shewhart, cusum, and EWMA charts require a Phase I data set to have parameter estimates that can be plugged into the Phase II calculations. These methods require one to draw a conceptual line below the Phase I data, and separate the estimation data (Phase I) from the ongoing SPC data (Phase II). With the changepoint formulation by contrast one does not assume known parameters and so does not require the estimates produced by a Phase I study. Rather, once the preliminaries are complete and the initial process stability has been achieved, the formulation allows us to go seamlessly into SPC in which, at each instant, *all* accumulated process readings are analyzed and all data used to test for the presence of a changepoint. It also provides an ongoing stream of ever-improving estimates of the parameters while the process remains in control.

The schematic of this approach is that as each new observation X_n is added to the data set, the changepoint statistic $T_{\max,n}$ is computed for the sequence X_1, X_2, \dots, X_n . If $T_{\max,n} > h_n$ (where $\{h_n\}$ is a suitably chosen sequence of control limits), then we conclude that there has been a change in mean. The MLEs of the changepoint τ , the before-and-after means μ_1 and μ_2 , and the standard deviation σ follow at once. By looking for something in the process that changed around the time of the estimated τ , we can then take appropriate corrective action.

Using the changepoint approach in ongoing SPC charting leads to a number of questions arising from differences between the changepoint and cusum methodologies and between the fixed-sample and dynamic-sample situations of the changepoint test. The most immediate is the choice of the control limit sequence $\{h_n\}$. While fixed-sample distributional theory makes it possible to specify a value of h_n that will give a *marginal* false alarm probability of α , such a sequence is not suitable for SPC purposes. It is the *conditional* probability of a false alarm, given that there was no false alarm at the previous test, that is relevant. Since the successive $T_{\max,n}$ values are highly correlated and the correlation increases with n , the conditional probability is far below the marginal probability, making marginal α control limits irrelevant.

The ideal would be a sequence of h_n such that the 'hazard' or 'alarm rate' (the conditional probability of a false alarm at any n , given that there was no

previous false alarm) was a constant α , as is the case with the Shewhart chart. With constant hazard, the in-control ARL would be $1/\alpha$, providing the same direct link between the hazard rate and the ARL that the Shewhart chart has when the in-control parameters are known. This approach was used by Margavio et al. (1995) in the context of an EWMA chart in which, even with known parameters, the false alarm rate changes over time. Margavio et al. (1995) derived control limit sequences that would fix the false alarm rate for the EWMA chart to a specified value, just as we wish to do for the changepoint approach. Distributional theory of the $T_{\max,n}$ sequence is far from being able to provide such a sequence, however, so we attacked this problem using simulation.

In principle, since the changepoint method does not rely on parameter estimates from Phase I, it is possible to start testing for a changepoint with the third process reading. Table 2 (obtained by simulation of 16 million sequences of length 200) shows the control limits for α values of 0.05, 0.02, 0.01, 0.005, 0.002, and 0.001, corresponding to in-control ARLs of 20, 50, 100, 200, 500, and 1000, for n values in the range 3 to 200.

As the values in Table 2 illustrate, for each α the control limits $h_{n,\alpha}$ decrease sharply initially, but then stabilize. The last four entries for $\alpha = 0.05$ are omitted because the simulation did not allow for accurate estimation of the control limit, but the value 2.30 can be used.

We believe that starting testing at the third observation is not a good idea, however. Leaping straight into the changepoint formulation with three readings would imply enormous faith in a not-yet-tested assumption of normality, and we assume that no one would be so trusting. Rather, we assume that a practitioner would gather a modest number of observations to get at least an initial verification that the normal distribution was a reasonable fit, and only then start the formal change-point testing. In line with this thinking, our main simulation is based on the assumption of an initial 9 readings without testing, with testing starting at the 10th observation. While 9 readings is also a slender basis for believing in a normal distribution for the quality variable, it perhaps represents a reasonable compromise between the conflicting desires to start control as soon as possible and to check for severe departures from normality.

The same sixteen million sequences of length 200

TABLE 2. Cutoffs $h_{n,\alpha}$ for Sample Size n and Hazard Rate α Starting at Sample 3

n	α					
	0.05	0.02	0.01	0.005	0.002	0.001
3	38.19	95.49	191.0	382.0	954.9	1910
4	7.321	11.84	16.91	24.10	38.30	54.51
5	4.874	6.908	8.902	11.42	15.75	20.02
6	4.057	5.399	6.615	8.047	10.36	12.50
7	3.621	4.697	5.600	6.616	8.169	9.553
8	3.344	4.274	5.024	5.829	7.020	8.031
9	3.158	3.992	4.649	5.340	6.317	7.130
10	3.024	3.790	4.384	4.997	5.847	6.541
11	2.924	3.640	4.186	4.745	5.512	6.124
12	2.845	3.524	4.036	4.552	5.257	5.807
13	2.783	3.433	3.916	4.402	5.058	5.562
14	2.732	3.357	3.821	4.282	4.895	5.368
15	2.691	3.296	3.742	4.181	4.763	5.211
16	2.655	3.244	3.677	4.098	4.655	5.080
17	2.625	3.200	3.620	4.031	4.564	4.968
18	2.598	3.161	3.570	3.968	4.486	4.879
19	2.574	3.128	3.528	3.916	4.418	4.795
20	2.554	3.100	3.491	3.871	4.362	4.727
22	2.521	3.050	3.429	3.794	4.260	4.607
24	2.493	3.011	3.380	3.732	4.184	4.511
26	2.470	2.979	3.338	3.682	4.117	4.439
28	2.452	2.952	3.305	3.641	4.064	4.375
30	2.435	2.929	3.277	3.607	4.022	4.324
35	2.405	2.886	3.221	3.538	3.936	4.222
40	2.383	2.854	3.182	3.491	3.873	4.147
45	2.366	2.830	3.151	3.453	3.827	4.094
50	2.354	2.810	3.127	3.426	3.790	4.053
60	2.334	2.785	3.094	3.383	3.736	3.990
70	2.323	2.765	3.070	3.355	3.702	3.947
80	2.316	2.751	3.053	3.333	3.677	3.918
90	2.308	2.741	3.040	3.318	3.656	3.895
100	2.304	2.734	3.030	3.307	3.640	3.875
125		2.717	3.010	3.281	3.610	3.844
150		2.711	2.997	3.264	3.591	3.822
175		2.705	2.994	3.257	3.579	3.804
200		2.701	2.985	3.248	3.570	3.794

were used to find cutpoints up to $n = 200$. An independent set of five million sequences of length 1000 verified the visual impression that the cutpoints stabilize to constant values in each column. The resulting control limits, presented in Table 3, are our recommendation for implementation of the change-point SPC scheme. As in Table 2, there are blanks where the surviving simulation sample size was too small for reliable estimation of the control limit, but

the missing values can safely be replaced by the value immediately above. Table 2 and 3 are extracts from a larger table that can be downloaded from www.stat.umn.edu/hawkins, which lists cutoffs for all starting points from observation 3 through 21. The entries in this table have standard errors with a median of 0.03% of the table value and with a maximum of 1% of the value.

For purposes of implementation, this table can be

TABLE 3. Cutoffs $h_{n,\alpha}$ for Sample Size n and Hazard Rate α Starting at Sample 10

n	α					
	0.05	0.02	0.01	0.005	0.002	0.001
10	3.662	4.371	4.928	5.511	6.340	7.023
11	3.242	3.908	4.424	4.958	5.697	6.284
12	3.037	3.677	4.167	4.664	5.350	5.890
13	2.909	3.530	3.997	4.468	5.110	5.608
14	2.821	3.424	3.875	4.326	4.931	5.397
15	2.756	3.344	3.780	4.211	4.786	5.229
16	2.704	3.281	3.704	4.121	4.671	5.093
17	2.663	3.228	3.642	4.047	4.576	4.977
18	2.628	3.183	3.587	3.981	4.494	4.885
19	2.599	3.146	3.542	3.926	4.425	4.799
20	2.575	3.115	3.503	3.880	4.367	4.730
22	2.535	3.060	3.437	3.800	4.264	4.610
24	2.504	3.019	3.386	3.736	4.187	4.514
26	2.479	2.985	3.343	3.685	4.119	4.440
28	2.459	2.957	3.308	3.643	4.065	4.375
30	2.440	2.933	3.279	3.609	4.024	4.324
35	2.408	2.888	3.223	3.539	3.937	4.223
40	2.385	2.855	3.184	3.492	3.873	4.147
45	2.368	2.832	3.152	3.454	3.828	4.095
50	2.355	2.811	3.128	3.426	3.791	4.053
60	2.335	2.785	3.094	3.383	3.737	3.989
70	2.324	2.765	3.071	3.355	3.702	3.946
80	2.315	2.752	3.052	3.333	3.677	3.918
90	2.310	2.741	3.040	3.318	3.656	3.895
100	2.302	2.735	3.030	3.307	3.640	3.875
125		2.717	3.011	3.281	3.611	3.844
150		2.710	2.997	3.264	3.591	3.821
175		2.703	2.993	3.257	3.579	3.804
200		2.700	2.985	3.248	3.570	3.794

used as is with interpolation for the sample sizes not listed explicitly. Alternatively, it may be replaced by some closed-form approximation. In particular, a simple, but quite accurate, approximation for $n \geq 11$ is given by

$$h_{n,\alpha} \approx h_{10,\alpha} \left(0.677 + 0.019 \ln(\alpha) + \frac{1 - 0.115 \ln(\alpha)}{n - 6} \right),$$

where $\ln(\cdot)$ is the natural log function.

This simple formula fits the table quite well. An independent simulation of 30,000 samples showed that it gives ARLs of 22, 48, 90, 185, 525, and 1180 for the nominal 20, 50, 100, 200, 500, and 1000 levels. Users who find this level of approximation unac-

ceptable have the choice of fitting some other more precise model for Table 3, or using table lookup.

Implementation

Looking at the formulas suggests that calculating the $T_{\max,n}$ values would be computationally burdensome. This is not the case. To implement the method, build up two arrays of values of

$$S_n = \sum_{i=1}^n X_i,$$

and

$$W_n = \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

The running mean $\bar{X}_n = S_n/n$ need not be stored,

but can be calculated ‘on the fly.’ When a new observation is added, its two new table entries can be calculated quickly from the numerically stable recursions

$$S_n = S_{n-1} + X_n$$

and

$$W_n = W_{n-1} + ((n-1)X_n - S_{n-1})^2/[n(n-1)].$$

Finding $T_{\max,n}$ then involves calculating the two-sample statistic T_{jn} for every possible split point $1 \leq j < n$. It is actually more convenient to find T_{jn}^2 . The variance explained by a split at point j can be shown to be

$$E_{jn} = (nS_j - jS_n)^2/[nj(n-j)],$$

and the analysis of variance identity

$$V_{jn} = W_n - E_{jn}$$

leads to $T_{jn}^2 = (n-2)E_{jn}/(W_n - E_{jn})$. Finding the maximum of these statistics across the allowed j values and comparing this maximum with h_n^2 then gives the changepoint test. If $T_{\max,n}^2 > h_n^2$, leading to the signal of a changepoint, then it is a trivial matter to compute the maximum likelihood estimators $\hat{\mu}_1 = S_j/j$, $\hat{\mu}_2 = (S_n - S_j)/(n-j)$, and $\hat{\sigma}^2 = V_{jn}/(n-2)$ (the customary variance estimator), using the value j leading to the maximum. Though maximum likelihood, these estimators are somewhat biased (see Hinkley (1970, 1971) for a discussion).

The maximizing j is that which maximizes E_{jn} , so the searching step need only evaluate E_{jn} for each j , making further T_{jn}^2 calculation necessary only for the maximizing E_{jn} . Thus, while at process reading number n there are $n-1$ calculations to be performed, each involves only about ten floating point operations, so even if n were in the tens of thousands calculating $T_{\max,n}$ would still be a trivial calculation.

The ever-growing storage requirement for the two tables might be more inconvenient. This can be limited (along with the size of the resulting search) if it is acceptable to restrict the search for the changepoint to the most recent w instants. To do this, one must keep a table of only the w most recent S_j and W_j values.

Note that this is different than the ‘window’ approach discussed by Willsky and Jones (1976), in that observations more than w time periods into the past are not lost, since they are summarized in the window’s leftmost S and W entry. All that is lost is the ability to split at these old instants. Suitable choices for the table size w might be in the 500 to

2000 range. This is large enough that no interesting structure is lost, but small enough to keep the computation for each new reading to less than some 20,000 operations.

In-Control Run Length Behavior and the Shewhart Chart

The purpose of a Shewhart I-chart is to compare the latest reading with the in-control distribution, signalling if it deviates by more than 3 standard deviations. Note that using the trial split point $j = n-1$ leads to a comparison of X_n with the mean of all its predecessors (essentially duplicating the Shewhart I-chart, though with a non-constant control limit of h_n standard deviations). From Table 3, we can see that once the sequence has been running for a while, $h_{n,0.002}$ is not much over 3. This suggests that even though the changepoint formulation is aimed at sustained possibly small shifts, it is able to do most of the work of the Shewhart I-chart as well. The same applies to an \bar{X} chart of rational subgroup means.

In the fixed-sample null hypothesis case, the distribution of the split point j maximizing $|T_{jn}|$ is known to have a steep ‘bathtub’ form of distribution, usually splitting off a handful of observations from either the extreme left or the right end of the series (Hawkins (1977)). In the repeated-use SPC setting, except for the first few tests, splits near the left end are rare, but splits close to the right end are common. An independent simulation of 10,000 null sequences illustrates this; the right segment length $n-j$ has a median of 3 readings at $\alpha = 0.05$ and 5 at $\alpha = 0.001$. The corresponding upper quartiles of 7 and 12 readings, respectively, show the compression of this distribution.

Multiple Change Points

This brings us to the question of multiple changepoints. We believe it is in the spirit of SPC that a sequence should contain at most one changepoint. Detecting a changepoint means that there has been a process change. If (as would usually be the case) this is for the worse, then immediate corrective action should be taken to restore the *status quo ante*, and the observations from the estimated changepoint up to the instant of process correction should be removed from the sequence included in the changepoint testing. If the change is for the better, then it is the older readings that are to be removed, and the new readings become the equivalent of Phase I data. Multiple change points within a sequence should not hap-

pen, as they imply that process operators are failing to respond to out-of-control situations.

In startup situations, a slightly different situation is encountered. Myriad minor adjustments are made to the process, as a result of which there may be many changepoints corresponding to ever better quality until the operators succeed in stabilizing the process. In this situation, an appropriate analysis would be to apply the changepoint model backwards (starting from the most recent observation and moving toward the earliest). The first changepoint detected would be an indication of when the last adjustment took effect leading to the presumed in-control state of the process. All data prior to that changepoint would then be discarded, and the changepoint series started from that point.

Performance of the Changepoint Approach

The performance of the changepoint approach can be assessed by the ARLs, both in-control and following a shift in mean. This introduces a complication not seen in Shewhart or known-parameter cusum charts; the response to a shift depends on the number of in-control observations preceding it. The reason for this dependence is that the noncentrality parameter of the two-sample t -statistic depends on the sample sizes. A short in-control period leads to a smaller non-centrality parameter and, therefore, a slower reaction than a longer in-control period.

These points are illustrated in Table 4, where we consider four α values: 0.02, 0.01, 0.005, and 0.002. A shift of size $\delta \in \{0, 0.25, 0.5, 1.0, 2.0\}$ was introduced at $\tau \in \{10, 25, 50, 100, 250\}$. The numbers presented in the table are the ARLs of the changepoint procedure. These were calculated by simulating a data series, adding the shift to all $X_i, i > \tau$, and counting the number of readings from the occurrence of the shift until the chart signalled. Any series in which a signal occurred before time τ was discarded. The approximate formula for h_n was used, so the in-control ARLs differ slightly from nominal.

It can be seen that the ARLs are affected by the amount of history gathered before the shift, with a faster response coming with more history. They also depend on the values of α and δ . When α is larger (i.e., the false-alarm rate is higher) or δ is larger (i.e., the shift size is larger), the ARLs tend to be smaller, as one would anticipate.

Comparison with the Cusum Chart

Comparing the changepoint approach with a standard alternative is complicated by the fact that there is no standard alternative. The cusum chart is the diagnostic of choice in circumstances where it applies, but using the conventional cusum chart requires exact knowledge of the in-control mean and standard deviation. Because of this, the cusum chart can clearly not be applied in the setting we are discussing, in which monitoring starts when only 9 observations have been observed.

The self-starting cusum chart was developed as a way of setting up a cusum control without the demand for a large Phase I sample to estimate parameters (see Hawkins (1987) and Hawkins and Olwell (1998)). Like the changepoint formulation, it does not draw a dividing line between Phases I and II, and can be started with as few as three observations. Writing \bar{X}_j and s_j for the mean and standard deviation of the first j readings, the self-starting cusum chart uses as summand

$$U_n = \Phi^{-1} \left[F_{n-2} \left(\sqrt{\frac{n-1}{n}} \frac{X_n - \bar{X}_{n-1}}{s_{n-1}} \right) \right],$$

where Φ is the standard normal cumulative distribution function, and F_{n-2} is the cumulative distribution function of a t -distribution with $n - 2$ degrees of freedom. The sequences U_n follow an *exact* standard normal distribution for all $n > 2$ and are statistically independent. Thus, they can be cumulatively summed using known-parameter settings. As the in-control history grows, \bar{X}_n approaches μ_1 , s_n approaches σ , and U_n approaches the Z -score $(X_n - \mu_1)/\sigma$, so the self-starting cusum chart approaches a known-parameter cusum chart of $N(0,1)$ quantities.

The self-starting cusum chart, therefore, operates in the same realm as our changepoint formulation, but has the optimality properties of the known-parameter cusum chart in the situation that the process happens to stay in control for a long initial period. It thus makes an appropriate benchmark against which to compare the changepoint formulation.

We set up three self-starting cusum charts, with reference values $k = 0.25, 0.5, \text{ and } 1.0$, thereby being tuned to mean shifts of 0.5, 1, and 2 standard deviations, respectively. By using the software from the Web site www.stat.umn.edu/users/cusum, we can find the cusum decision intervals h leading to

TABLE 4. The ARL of the Changepoint Procedure When Shift Occurs at the "Start" Position With Size δ

α	start	δ				
		0	0.25	0.5	1.0	2.0
0.02	10	54.5	51.8	43.3	21.3	4.1
	25	54.2	46.7	30.8	9.4	2.4
	50	53.8	42.8	23.3	7.0	2.1
	100	54.7	41.4	19.4	6.0	1.9
	250	58.7	33.4	16.6	6.3	1.8
0.01	10	99.3	95.9	82.4	39.1	5.3
	25	99.4	86.9	55.2	13.4	2.8
	50	99.2	76.3	38.2	8.8	2.3
	100	99.7	69.3	28.5	7.6	2.2
	250	96.9	55.4	23.9	7.0	2.1
0.005	10	195.9	186.2	168.3	84.9	7.0
	25	196.6	174.7	113.9	20.4	3.3
	50	196.3	154.9	68.7	11.2	2.7
	100	196.9	131.3	42.2	9.4	2.5
	250	194.8	99.0	29.7	8.2	2.3
0.002	10	535.8	531.1	492.7	280.9	10.8
	25	538.7	504.3	357.5	43.7	4.3
	50	539.5	457.8	195.4	15.7	3.4
	100	542.8	373.1	77.3	12.3	3.0
	250	546.2	222.9	43.3	10.8	2.8

in-control ARLs of 100 and 500. These intervals are

for ARL = 100 : $k = 0.25, h = 5.69; k = 0.5,$
 $h = 3.51; k = 1.0, h = 1.87;$

for ARL = 500 : $k = 0.25, h = 8.76; k = 0.5,$
 $h = 5.14; k = 1.0, h = 2.71.$

These choices provide benchmarks for the changepoint model with $\alpha = 0.01$ and 0.002 , respectively.

The cusum chart and changepoint tests were started following 9 in-control observations. Two mean-shift settings were simulated: a mean shift starting from the 10th observation, and a mean shift starting from the 100th observation. The first of these illustrates the extreme of using a very short in-control learning period before the shift. This tests not only the quick startup of the schemes but also the potential of the two procedures to monitor short-run settings. Shifting after 100 in-control observations is intended to approximate using the conventional cusum chart in the known-parameter setting (though

imperfectly, as Table 1 shows that a calibration sample of size 100 is not enough to truly control the cusum chart's run length behavior.) We discarded simulated runs in which either scheme signalled a shift before time τ .

Figures 1(a)-1(d) present the resulting ARLs for δ values in the range $[0,3]$. These are on a log scale for clearer comparison. Looking at the right panel, we see that the changepoint formulation is generally not the optimal diagnostic. For small shifts, it is slightly worse than the $k = 0.25$ cusum chart; for medium-size shifts slightly worse than the $k = 0.5$ cusum chart; and for large shifts slightly worse than the $k = 1$ cusum chart. This is no surprise, given the theoretical optimality of a properly-constructed cusum chart. However, for all shifts the cusum chart that beats the changepoint diagnostic does so by only a small margin, and at least one of the wrongly-centered cusum charts is far inferior. In other words, while it is never the best, the changepoint method is always nearly best, something that is not true of any of the choices for the cusum chart.

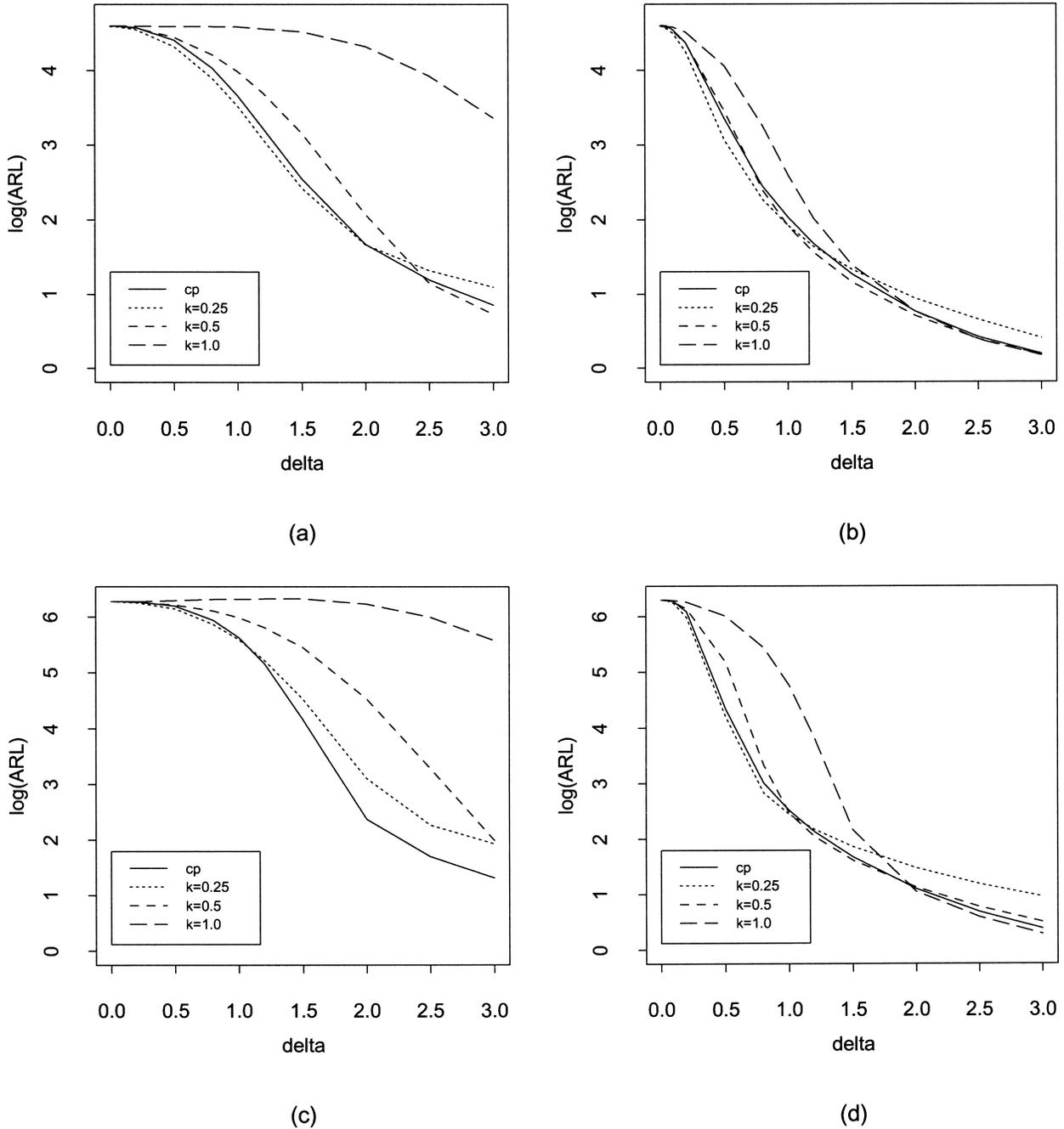


FIGURE 1. The ARLs of the Self-Starting Cusum Chart Procedures and the Changepoint Procedure When a Mean Shift of Size δ is Introduced. (a) $\alpha = 0.01$ and the Shift Starts at the 10th Observation; (b) $\alpha = 0.01$ and the Shift Starts at the 100th Observation; (c) $\alpha = 0.002$ and the Shift Starts at the 10th Observation; (d) $\alpha = 0.002$ and the Shift Starts at the 100th Observation.

Turning to the left two panels, the changepoint formulation looks even better in comparison with the cusum chart. Only the $k = 0.25$, $\text{ARL}=100$ cusum chart comes close to matching it, while the $k = 1$ cusum chart performs very poorly.

Thus, these figures illustrate that the changepoint formulation is preferable to the cusum chart overall. It involves a slight disadvantage in performance where one knows in advance what the out-of-control mean will be and one has a substantial in-control his-

tory. For short in-control histories, as would be seen in startup and short-run problems, it is slightly inferior to the small- k cusum chart provided a short in-control ARL is used. However these small worst-case disadvantages are surely more than offset by potential large gains in performance elsewhere.

Conclusion

The unknown-parameter changepoint formulation has long been thought of and used as a Phase I analysis tool. It is perhaps even more attractive in ongoing (Phase II) SPC settings in that it does not require knowledge of the in-control parameters, and can therefore dispense with the need for a long Phase I calibration sequence. Not only is it flexible, it is also powerful. While it can not quite match the performance of the cusum chart tuned to the particular shift that happened to occur, it is close to optimal for a wide range of shifts, and so has a much greater robustness of good performance.

Implementation requires control limit values. A table found by simulation leads to a constant hazard function. This table turns out to be quite well approximated by a simple analytical function, leading to a straightforward implementation route.

Acknowledgments

We thank the Editor and referees for many constructive suggestions. This research was supported in part by a Grant-in-Aid of Research, Artistry and Scholarship of University of Minnesota and a NSF grant DMS 9806584.

References

- CROWDER, S. V.; HAWKINS, D. M.; REYNOLDS, M. R., JR.; and YASHCHIN, E. (1997). "Process Control and Statistical Inference". *Journal of Quality Technology* 29, pp. 134–139.
- GOMBAY, E. (2000). "Sequential Change-point Detection with Likelihood Ratios". *Statistics and Probability Letters* 49, pp. 195–204.
- HAWKINS, D. M. and OLWELL, D. H. (1998). *Cumulative Sum Charts and Charting for Quality Improvement*. Springer Verlag, New York, NY.
- HAWKINS, D. M. (1977). "Testing a Sequence of Observations for a Shift in Location". *Journal of the American Statistical Association* 72, pp. 180–186.
- HAWKINS, D. M. (1987). "Self-Starting Cusum Charts for Location and Scale," *The Statistician* 36, pp. 299–315.
- HINKLEY, D. V. (1970). "Inference about the Change-point in a Sequence of Random Variables". *Biometrika* 57, pp. 1–17.
- HINKLEY, D. V. (1971). "Inference about the Change-point from Cumulative Sum Tests". *Biometrika* 58, pp. 509–523.
- IRVINE, J. M. (1982). "Changes in Regime in Regression Models". Ph.D. Dissertation, Yale University.
- JONES, L. A. and CHAMP, C. W. (2001). "The Performance of Exponentially Weighted Moving Average Charts with Estimated Parameters". *Technometrics* 43, pp. 156–167.
- LAI, T. L. (2001). "Sequential Analysis: Some Classical Problems and New Challenges". *Statistica Sinica* 11, pp. 303–350.
- MARGAVIO, T. M.; CONERLY, M. D.; WOODALL, W. H.; and DRAKE, L. G. (1995). "Alarm Rates for Quality Control Charts". *Statistics and Probability Letters* 24, pp. 219–224.
- NEDUMARAN, G. and PIGNATIELLO, J. J., JR. (2001). "On Estimating \bar{X} Control Chart Limits". *Journal of Quality Technology* 33, pp. 206–212.
- PIGNATIELLO, J. J., JR. and SAMUEL, T. R. (2001). "Estimation of the Change Point of a Normal Process Mean in SPC Applications". *Journal of Quality Technology* 33, pp. 82–95.
- QUESENBERRY, C. P. (1993). "The Effect of Sample Size on Estimated Limits for \bar{X} and X Control Charts". *Journal of Quality Technology* 25, pp. 237–247.
- SEN, A. and SRIVASTAVA, M. S. (1975). "On Tests for Detecting Change in Mean When Variance is Unknown". *Annals of the Institute of Statistical Mathematics* 27, pp. 479–486.
- SULLIVAN, J. H. and WOODALL, W. H. (1996). "A Control Chart for Preliminary Analysis of Individual Observations". *Journal of Quality Technology* 28, pp. 265–278.
- WILLSKY, A. S. and JONES, H. G. (1976). "A Generalized Likelihood Ratio Approach to Detection and Estimation of Jumps in Linear Systems". *IEEE Transactions on Automatic Control* 21, pp. 108–112.
- WORSLEY, K. J. (1979). "On the Likelihood Ratio Test for a Shift in Location of Normal Populations". *Journal of the American Statistical Association* 74, pp. 365–367.
- WORSLEY, K. J. (1982). "An Improved Bonferroni Inequality and Applications". *Biometrika* 69, pp. 297–302.

Key Words: *Cumulative Sum Control Charts, Exponentially Weighted Moving Average Control Charts, Shewhart Control Charts.*

