Center for Quality and Productivity Improvement
University of Wisconsin
610 Walnut Street
Madison, Wisconsin 53705

(608) 263-2520
(608) 263-1425 FAX
quality@engr.wisc.edu

Report No. 169

Influence of the Sampling Interval, Decision Limit, and Autocorrelation on the Average Run Length in Cusum Charts

Alberto Luceño and George E.P. Box

September 1998

CQPI
University of Wisconsin

Report No. 169

Report Series
in Quality and Productivity

Influence of the Sampling Interval, Decision Limit, and Autocorrelation on the Average Run Length in Cusum Charts

Alberto Luceño

E.T.S. de Ingenieros de Caminos

University of Cantabria–Santander, Spain

George E.P. Box

Center for Quality
& Productivity Improvement

University of Wisconsin–Madison, USA

ABSTRACT

This paper shows how the average run length (ARL) for a one-sided Cusum chart varies as a function of the length of the sampling interval between consecutive observations, the decision limit for the Cusum statistic, and the amount of autocorrelation between successive observations. It is shown that the rate of false alarms can be decreased considerably, without modifying the rate of valid alarms, by decreasing the sampling interval and appropriately increasing the decision interval. Also that this can be done even when the shorter sampling interval induces moderate autocorrelation between successive observations.

KEYWORDS: Change point problem, Detection, False alarm, Robustness, Statistical process control.

CQPI Report No. 169, September 1998

Influence of the Sampling Interval, Decision Limit, and Autocorrelation on the Average Run Length in Cusum Charts

Alberto Luceño & George Box

This paper shows how the average run length (ARL) for a one-sided Cusum chart varies as a function of the length of the sampling interval between consecutive observations, the decision limit for the Cusum statistic, and the amount of autocorrelation between successive observations. It is shown that the rate of false alarms can be decreased considerably, without modifying the rate of valid alarms, by decreasing the sampling interval and appropriately increasing the decision interval. Also that this can be done even when the shorter sampling interval induces moderate autocorrelation between successive observations.

1. INTRODUCTION

By process monitoring we mean sequential surveillance of a system to detect a change of state that can characterize a malfunction. The general problem can also be stated as that of serially collecting and analyzing data to quickly detect a signal buried in noise. We suppose that observations $y_1, y_2, \ldots$ of some characteristic of interest are made at equispaced times τS, where $\tau = 1, 2, \ldots$ and S is the *sampling interval*, that is, the time interval between successive observations, and that monitoring is performed by continually updating the value of some quantity Q calculated from the incoming data and declaring detection of a malfunction as soon as Q changes in some specifically defined manner. Correct detection of a malfunction will be called a *valid alarm*; incorrect detection a *false alarm*. Because of the system noise, some detection errors will occur. An efficient monitoring scheme is one that produces alarms; the largest proportion of which is valid and the smallest proportion false.

Such "change point" problems have been considered by a number of authors (see, for example, Basseville and Nikiforov, 1993; Crowder, Hawkins, Reynolds, and Yashchin, 1997; Ferger, 1995; Gordon and Pollak, 1995; Lai, 1995; Lorden, 1971; and Roberts, 1966). In particular, a cumulative Fisher score statistic (see Box and Ramírez, 1992, and Box and Luceño, 1997) or Cuscore has been proposed for the quantity Q.

In this paper we suppose that the system is in a satisfactory state when

$$y_t = \mu_0 + a_t, \qquad (1.1)$$

where $\{a_t\}$ is a series of normally, independently, and identically distributed random variables with mean zero and variance $\sigma_a^2 = \sigma_y^2$, often referred to as white noise. We suppose also that a malfunction is characterized by an increase $d = D\,\sigma_a$ in the mean from μ_0 to $\mu_0 + d = \mu_1$, where the quantity D is called the *standardized increase in the mean.* The monitoring problem is then one of detecting a step change signal buried in white noise and the appropriate detection statistic (Barnard, 1959, and Page, 1954, 1957, 1961) is the Cusum

$$Q = \Sigma(y_i - \mu_0).$$

Comprehensive descriptions and applications of Cusums may be found, for example, in Box and Luceño (1997, chapter 3), Duncan (1974, chapter 22), or Montgomery (1991, chapter 7).

For formal detection, it is convenient to use the centered Cusum

$$Q^* = \Sigma(y_i - \bar{\mu})$$

where $\bar{\mu} = (\mu_0 + \mu_1)/2 = \mu_0 + d/2$. This centered Cusum is essentially a handicapped score. When the mean is equal to μ_0, the quantity $d/2$ will be subtracted from each deviation, producing a downward trend in the centered Cusum. But, if the mean level is greater than $\mu_0 + d/2$, the centered Cusum will tend to increase. The Cusum detection procedure requires that we accept that a genuine increase in the mean above μ_0 has occurred as soon as the centered Cusum has risen from its previous minimum by more than some quantity h called the *decision interval.*

For given standardized increase in the mean D and the sampling interval S, the standardized decision interval $H = h/\sigma_a$ will determine the characteristics of the Cusum chart. A quantity describing the performance of such schemes is the *run length* (RL), that is, the elapsed time before an alarm is triggered. This is a random variable varying from one run to another, and its mean is called the *average run length* (ARL). When the ARL is large, the probability distribution for the RL is close to a geometric distribution (Ewan and Kemp, 1960, and Brook and Evans, 1972) so that it can be completely determined from its ARL. On this basis, it has become customary to define the Cusum scheme in terms of the ARL for false alarms (L_a) when the process is in the acceptable condition $\mu = \mu_0$, and the ARL for valid alarms (L_r) when the process is in the rejectable condition $\mu = \mu_1$.

As a specific example, suppose the data were measurements of the temperature of a fluid in a reactor taken every $S = 6$ minutes. Then if μ_0 were the acceptable temperature, our monitoring system would be required to indicate an alarm if the temperature rose to the unacceptable level μ_1. If $D = 3$ and $H = 2.64$, the methods in Goel and Wu (1971) give $L_r = 2.435$ and $L_a = 16{,}666$. Thus, if the temperature stayed at the satisfactory level μ_0, the average elapsed time before a false alarm was triggered would be $T_a = L_a S = 100{,}000$ minutes (69.4 days). However, with the temperature at the unsatisfactory level μ_1, an alarm will occur on the average after an elapsed time of $T_r = L_r S = 14.61$ minutes.

If we characterize the Cusum in terms of these averages, the above scheme seems very satisfactory. Unfortunately however, as is well known, RLs are subject to very large variations about their ARLs. This may be particularly disturbing in the case of the RL distribution for satisfactory performance, when the ARL is very large, because the standard deviation of the RL is then almost equal to its ARL and hence is very large. For illustration, Table 1 shows nine quantiles of the RL distribution for the acceptable state with $\mu = \mu_0$ in the example above. Although it would be true that when the reactor is in the acceptable state, the RL distribution would have its mean at 100,000 minutes (69.4 days), we see that more than 5% of false alarms would occur in the first 4 days and about 25% in the first 20 days. It is then necessary

(1) to find ways of increasing the smaller quantiles of the RL distribution (expressed as elapsed time) corresponding to satisfactory performance

(2) to characterize potential performance in such a way that a fairer picture of the behavior of the monitoring scheme is gained.

Section 2 considers the effect on the ARL and quantiles of the RL distribution of reducing the sampling interval and simultaneously increasing the decision interval. Section 3 analyzes how the resulting ARL and quantiles change when there is a moderate autocorrelation between consecutive observations. Concluding remarks are in Section 4.

2. EFFECT OF REDUCING THE SAMPLING INTERVAL AND AUGMENTING THE DECISION INTERVAL

Efficient design of a Cusum chart may be done by observing that the ARL (expressed as a multiple of the sampling interval S) increases approximately *exponentially* with the decision interval when the process behaves satisfactorily with mean μ_0, whereas this increase is only approximately *linear* when the process mean is μ_1. In addition, the ARL (expressed as elapsed time) is a linear function of S irrespective of whether the process behavior is satisfactory or not. Thus, suppose the length of the sampling interval is divided by 3 so that $S = 2$ minutes and simultaneously the standardized decision limit is multiplied by 3 so that $H = 7.92$. Then assuming that model (1.1) remains essentially valid with the shorter sampling interval, the first two lines of Table 2 show that the ARL and upper quantiles ($q_{0.95}$ and $q_{0.99}$) for the RL distribution corresponding to valid alarms decrease slightly, whereas the ARL and lower quantiles ($q_{0.01}$ and $q_{0.05}$) for the RL distribution of false alarms increase drastically. In particular, the scheme in the first line produces 5% of false alarms by 5,136 minutes (3.6 days), whereas the same percent of false alarms is reached only by 13.1 billions minutes (24,924 years) if the scheme in the second line is used.

An explanation of the differences between the RL distributions in the first two lines of Table 2 is as follows. By dividing the sampling interval by 3, the ARL and quantiles for the RL distribution (expressed in minutes) is divided

percent of false alarms	1	5	10	25	50	75	90	95	99
minutes	1,014	5,136	10,542	28,776	69,318	138,624	230,238	299,544	460,470
days	0.70	3.57	7.32	19.98	48.14	96.27	159.89	208.02	319.77
(These quantiles have been computed numerically using the true RL distribution, rather than its geometric approximation.)									

Table 1: *Some quantiles of the RL distribution for false alarms having T_a = 100,000 minutes (equivalent to 69.4 days) in the example of Section 1.*

scheme	false alarms acceptable conditions			valid alarms rejectable condition		
	$q_{0.01}$	$q_{0.05}$	ARL(T_a)	ARL(T_r)	$q_{0.95}$	$q_{0.99}$
previous scheme $S = 6$ minutes $H = 2.64$, $D = 3$	1,014	5,136	100,000	14.61	24.00	30.00
more frequent sampling $S = 2$ minutes $H = 7.92$, $D = 3$	2.56×10^9	13.1×10^9	255×10^9	11.91	18.00	22.00
more frequent sampling $S = 2$ minutes $H = 7.92$, $D = 3$ autocorrelation $\phi = 0.4$	$.556 \times 10^6$	2.84×10^6	55.3×10^6	17.63 (15.14)	28.00 (26.00)	36.00 (32.00)

***Table 2:** ARLs and quantiles in minutes for different schemes and temperature conditions.*

exactly by 3 no matter whether the temperature in the reactor is in the acceptable or rejectable condition. Now on the one hand, when the temperature is in the rejectable condition, the ARL and upper quantiles for the RL distribution increase approximately *linearly* with the standardized decision interval so that, when this is multiplied by 3, the ARL and upper quantiles for valid alarms are roughly multiplied by 3 and therefore remain approximately unchanged with respect to the initial scheme. On the other hand, when the temperature is in the acceptable condition, the ARL and lower quantiles for false alarms increase approximately *exponentially* with the standardized decision interval and hence are much larger for the new scheme than for the initial scheme. This rationalization is in agreement with Siegmund's approximate formula for evaluating the ARLs for false and valid alarms (see Siegmund, 1985, p. 27; or Woodall and Adams, 1993).

3. EFFECT OF THE AUTOCORRELATION BETWEEN SUCCESSIVE OBSERVATIONS

We now suppose that when the system stays in a satisfactory state but the length of the sampling interval is reduced to $S = 2$ minutes, model (1.1) does not remain valid but can be replaced by an autoregressive time series model of order one, AR (1), such as

$$(y_t - \mu_0) = \phi(y_{t-1} - \mu_0) + a_t, \qquad (3.1)$$

where ϕ is the autocorrelation coefficient between consecutive observations and b_t is a white noise with variance σ_b^2. Obviously, for $\phi = 0$, model (3.1) coincides with model (1.1). If model (3.1) is assumed to be valid for $S = 2$ minutes, the corresponding model for $S = 6$ minutes would be

$$(y_t - \mu_0) = \phi^3(y_{t-3} - \mu_0) + a_t, \qquad (3.2)$$

and, assuming that σ_y^2 remains constant, we have

$$\sigma_y^2 = \frac{\sigma_a^2}{1-\phi^6} = \frac{\sigma_b^2}{1-\phi^2},$$

which relates σ_y with σ_a and σ_b.

Suppose now that we want to look for the temperature change (step signal) from μ_0 to μ_1 considered in Sections 1 and 2 hidden in the AR (1) noise of Equation (3.1) with $\phi = .4$, rather than in the white noise of Equations (1.1) or (3.2) with $\phi = 0$. Then application of the results in Box and Luceño (1997, Appendix 10B) shows that this problem is equivalent to looking in white noise for a signal that has a first spike of magnitude

$$D = \frac{d}{\sigma_b} = \frac{3\sigma_y}{\sigma_b} = 3(1-\phi^2)^{-1/2} = 3.2733 \qquad (3.3a)$$

and the remaining spikes of magnitude

$$\begin{aligned} D &= \frac{d(1-\phi)}{\sigma_b} = \frac{3(1-\phi)\sigma_y}{\sigma_b} \\ &= 3(1-\phi)(1-\phi^2)^{-1/2} = 1.9640 \end{aligned} \qquad (3.3b)$$

Let us suppose, for simplicity, that all these spikes including the first one may be taken equal to 1.9640 so that looking for a change of the mean temperature of magnitude $d = 3\sigma_y$ in the AR (1) noise of Equation (3.1) with $\phi = .4$ is approximately equivalent to looking for a standardized change in the mean temperature of magnitude $D = 1.9640$ in white noise. Then the third line of Table 2 gives the ARL and upper quantiles for the RL distribution corresponding to the rejectable condition along with the ARL and lower quantiles for the RL distribution for the acceptable condition. We see that the parameters for valid alarms are close to those obtained previously for the initial scheme (in the first line of Table 2), whereas the parameters for the false alarms remain considerably larger than for the original scheme (although not as large as in the second line of Table 2). In particular, whereas the original scheme produces 5% of false alarms by 5,136 minutes (3.6 days), this percent of false alarms is reached only by 2.84 million minutes (5.4 years) for the scheme in the third line of Table 2 when the autocorrelation coefficient is $\phi = .4$.

When computing the third line of Table 2, it has been assumed that the first spike of magnitude 3.2733 may be replaced by a smaller spike of magnitude 1.9640 without introducing important errors. Consequently, the signal used to compute the third line of the table is a step signal of magnitude 1.9640. If this simplification is not made, that is, if our scheme is looking for a step signal of magnitude 1.9640 but the true signal is given by equations 3.3 a and b, then the resulting ARL and upper quantiles are those shown in parentheses in the third line of Table 2. We see that these values are even closer than their approximations to the values for the original scheme. Clearly the ARL and quantiles for the false alarms remain unchanged because these are computed assuming the true signal is null.

4. CONCLUDING REMARKS

The performance of Cusum charts has been usually analyzed in the literature under the assumption that the white noise sequence in model (1.1) is normally distributed. In particular, Goel and Wu (1971) provided a chart to evaluate the ARL for one-sided Cusum charts under the normal assumption. This assumption is often justified in practice because the values y_t in (1.1) are obtained as an average of a few individual observations of the quality characteristic. If the standardized decision interval H is very large, however, the resulting ARLs are somewhat sensitive to the assumption that the tails of the distribution are not very different from those corresponding to the normal assumption. This problem is one of lack of knowledge of the exact tails of the distribution (because data in the far tail are only very rarely observed), so that no alternative to the normal distribution is likely to be considered more appropriate. Consequently, the actual performance of any tentative scheme chosen on the guidelines provided by theoretical studies should always be checked in practice. If the true tails of the distribution of individual observations are known to be heavier to the normal tails, it might be adequate to use a robust estimate of location (see Mosteller and Tukey, 1977, chapter 10) to define y_t in (1.1), rather than using a simple arithmetic average of several individual observations.

ACKNOWLEDGMENT

George Box acknowledges support by grants from the Low Emission Partnership and the National Science Foundation Grant No. DMI-9812839. Alberto Luceño acknowledges the support of the Spanish DGICYT Grant PB95-0583 and DGES Grant PB97-0555.

REFERENCES

Barnard, G. A. (1959), "Control Charts and Stochastic Processes," *Journal of the Royal Statistical Society*, Series B, **21**, 239–271.

Basseville, M. and Nikiforov, I. V. (1993), *Detection of Abrupt Changes: Theory and Application*, Englewood Cliffs, New Jersey: Prentice Hall.

Box, G. E. P. and Luceño, A. (1997), *Statistical Control by Monitoring and Feedback Adjustment*, New York: Wiley.

Box, G. E. P. and Ramírez, J. (1992), "Cumulative Score Charts," *Quality and Reliability Engineering International*, **8**, 17–27.

Brook, D. and Evans, D. A. (1972), "An Approach to the Probability Distribution of Cusum Run Length," *Biometrika*, **59**, 539–549.

Crowder, S. V., Hawkins, D. M., Reynolds, M. R. Jr., and Yashchin, E. (1997), "Process Control and Statistical Inference," *Journal of Quality Technology*, **29**, 134-139.

Duncan, A.J., (1974), *Quality Control and Industrial Statistics*, Homewood, Illinois: R.D. Irwin.

Ewan, W. D. and Kemp, K. W. (1960), "Sampling Inspection of Continuous Processes with No Autocorrelation Between Successive Results," *Biometrika*, **47**, 363–380.

Ferger, D. (1995), "Nonparametric Tests for Nonstationary Change-point Problems," *The Annals of Statistics*, **23**, 1848-1861.

Goel, A. L. and Wu, S. M. (1971), "Determination of A. R. L. and a Contour Nomogram for Cusum Charts to Control Normal Mean," *Technometrics*, **13**, 221–230.

Gordon, L. and Pollak, M. (1995), "A Robust Surveillance Scheme for Stochastically Ordered Alternatives," *The Annals of Statistics*, **23**, 1350-1375.

Lai, T. L. (1995), "Sequential Changepoint Detection in Quality Control and Dynamical Systems," *Journal of Royal Statistical Society*, Series B, **57**, 613-658.

Lorden, G. (1971), "Procedures for Reacting to a Change in Distribution," *The Annals of Mathematical Statistics*, **6**, 1897-1908.

Montgomery, D. C.. (1991), *Introduction to Statistical Quality Control*, 2nd ed. New York: Wiley.

Mosteller, F. and Tukey, J. W. (1977), *Data Analysis and Regression, a Second Course in Statistics*, Reading, Massachusetts: Addison-Wesley.

Page, E. S. (1954), "Continuous Inspection Schemes," *Biometrika*, **41**, 100–114.

Page, E. S. (1957), "On Problems in Which a Change in a Parameter Occurs at an Unknown Point," *Biometrika*, **44**, 248–252.

Page, E. S. (1961), "Cumulative Sum Charts," *Technometrics*, **3**, 1–9.

Roberts, S. W. (1966), "A Comparison of Some Control Chart Procedures," *Technometrics*, **8**, 411–430.

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

Woodall, W. H. and Adams, B. M. (1993), "The Statistical Design of Cusum Charts," *Quality Engineering*, **5**, 559–570.