Optimal Design of Cumulative Sum Control Charts Under Shift Uncertainty

Wenpo Huang[1] Lianjie Shu[1,3] Wei Jiang[2]
[1]Faculty of Business Administration, University of Macau
[2]Antai College of Economics and Management, Shanghai Jiaotong University
[3]Corresponding author: Lianjie Shu, email: ljshu@umac.mo

Abstract

This paper considers efficient design of the cumulative sum (CUSUM) control chart for detecting process mean shifts with unknown magnitude. A fast and accurate algorithm based on the gradient method is developed for this purpose. Optimal design parameters are obtained and compared with the one obtained through simulations in literature. The gradient-based method is shown to provide more accurate and faster design of the CUSUM chart under uncertainty than using Monte Carlo simulations.
Keywords: Average Run Length; Simulation; Statistical Process Control.

1 Introduction

The cumulative sum (CUSUM) control chart first introduced by Page (1954) has received a great deal of attention due to its simplicity and optimality property. Statistical theory has shown that the CUSUM scheme gives the minimum out-of-control average run length (ARL), often denoted as ARL_1, at a particular level of mean shifts among the alternatives in the worse case scenario, given a fixed in-control ARL (ARL_0), see Pollak (1985), Siegmund (1985) and Ritov (1990). The ARL_0 refers to the average number of observations before a false alarm is triggered while ARL_1 represents the average number of observations for a control chart required to signal an out-of-control situation.

The conventional CUSUM chart is often designed based on this optimality criterion. In particular, the reference value can be selected as $k = \delta/2$ if the CUSUM chart is designed to detect a standardized shift of magnitude δ in process means when the zero-state ARL (corresponding to the worst case) is taken as the performance criterion (Bagshaw and Johnson 1975). However, this ARL-based design has limited applicability as the shift to be occurred in the future is rarely known.

A number of new criteria have been proposed to account for uncertainty in the shift size. Although these new criteria are motivated from different purposes, they generally fall into a category known as expected weighted ARL (EWARL). See, for example, Wu et al. (2004), Chen and Chen (2007), and Reynolds and Lou (2010). The design of control charts under random shifts is more complex than that under known shift sizes. Monte Carlo simulation is routinely used to evaluate and design control charts in this case. Examples include Chen and Chen (2007), Reynolds and Lou (2010) and Ryu et al. (2010). The optimal parameters of control charts were often obtained by using a grid search algorithm based on simulations. In this sense, the optimization is carried out in a somewhat trial-and-error manner. The precision of the optimal parameters obtained greatly depends on the grid size as well as the number of replicates used in the simulations. As a result, extensive computation load is often required to achieve high accuracy.

In order to make the design of CUSUM charts more efficient under random mean shifts, this paper proposes a gradient approach. The gradient information provides valuable information in search of the optimal parameters and thus can be expected to greatly improve the efficiency in the design and sensitivity analysis of control charts. There are very few papers discussing gradient-based algorithms in the design of control charts. Some exceptions include the following Lele (1996) and Fu and Hu (1999). However, most of these paper rely on simulations to estimate the ARL gradients. In contrast, this paper derives the ARL gradients based on the integral equation approach without running a large number of simulations.

2 The Design Criterion under Unknown Shifts

Suppose X_t, $t = 1, 2, \ldots$ is a sequence of independent observations following normal with mean μ and variance σ^2. The process is considered to be in-control when $\mu = \mu_0$ and out-of-control when $\mu = \mu_0 + \delta$ ($\delta \neq 0$), where δ is referred to as the shift size of the mean level. Assume for simplicity, $\sigma^2 = 1$ and μ_0 =0. Otherwise, one can simply standardize X_t using process mean and standard deviation.

When the main objective is to detect upward/positive changes in the process mean, the one-sided upper CUSUM chart can be devised as

$$Y_t = \max\{0, Y_{t-1} + X_t - k\}, \tag{1}$$

where the initial value is set as $Y_0 = z$ ($0 \leq z \leq h$). An alarm is triggered on this CUSUM chart when $Y_t > h$. Let $T(k, h)$ be the run length, defined as the number of observations collected until the first out-of-control signal is triggered, of the upper CUSUM chart in (1) with reference value k and threshold h. Denote $L(k, h; z, \delta)$ the ARL of the CUSUM chart conditioned on $Y_0 = z$ and the underlying shift size of δ. That is,

$$L(k, h; z, \delta) = E(T(k, h)|Y_0 = z, \mu = \delta).$$

In particular, $L(k, h; z, 0)$ refers to the ARL_0 of the upper CUSUM chart with head start z.

Traditionally, the CUSUM chart is optimally designed based on the ARL criterion aimed at quickly detecting a pre-specified mean shift. However, future shift is rarely known and it is more reasonable to assume a random distribution for the shift. See, for example, Chen and Chen (2007) and Ryu et al. (2010). In this case, the design criterion is not to minimize the ARL but the EWARL over a range of possible shifts $\delta \in [a, b]$. Namely,

$$\text{EWARL} = E[\omega(\delta)L(k, h; z, \delta)] = \int_a^b \omega(\delta)L(k, h; z, \delta)g(\delta)\mathrm{d}\delta,$$

where $\omega(\delta)$ is a weight function to indicate the importance of the shift magnitude δ, and $g(\delta)$ is the probability density function of δ.

3 Gradient Estimation

Once the form of $g(\delta)$ is specified and $\omega(\delta)$ is selected based on user's preference, the goal is to design a control chart minimizing EWARL. Let $G(k,h)$ be the EWARL function of k and h, the optimal CUSUM chart can be obtained from the following optimization model,

$$\min_{k,h} G(k,h) = \int_a^b \omega(\delta)L(k,h;z,\delta)g(\delta)\mathrm{d}\delta \tag{2a}$$

$$\text{subject to } L(k,h;z,0) = \mathrm{ARL}_0, \tag{2b}$$

where ARL_0 is a pre-specified in-control zero-state ARL. Note that the decision variable h is chosen to maintain a fixed ARL_0 value, so can be viewed as an implicit function of k. Denote $h = \eta(k)$, the above optimization model can also be written as

$$\min_{k} \Theta(k) = G(k,\eta(k)) = \int_a^b \omega(\delta)L(k,\eta(k);z,\delta)g(\delta)\mathrm{d}\delta. \tag{3}$$

For the sake of simplicity, we limit discussion to the case with $z = 0$.

The search of the minimum in the optimization problem (3) is equivalent to finding the root of $\Theta'(k) = 0$, where $\Theta'(k)$ is the first derivative of $\Theta(k)$ w.r.t k. As the calculation of $\Theta'(k)$ depends on the first-order partial derivatives of the ARL of the CUSUM chart w.r.t. the parameters k and h, denoted as $L_k(k,h,z,\delta)$ and $L_h(k,h,z,\delta)$ respectively, we first derive $L_k(k,h;z,\delta)$ and $L_h(k,h;z,\delta)$ in this section.

3.1 Estimate of $L_k(k,h,z,\delta)$

Based on Equation (1), it follows that

$$L(k,h;z,\delta) = 1 + L(k,h;0,\delta)F(k-z;\delta) + \int_0^h L(k,h;x,\delta)f(x+k-z;\delta)\mathrm{d}x, \tag{4}$$

where $f(x;\delta)$ and $F(x;\delta)$ are the PDF and CDF of the normal distribution with mean δ and variance 1 respectively. Differentiating both sides of Equation (4) w.r.t. k yields

$$\begin{aligned} L_k(k,h;z,\delta) = {} & L_k(k,h;0,\delta)F(k-z;\delta) + + \int_0^h L(k,h;x,\delta)f'(x+k-z;\delta)\mathrm{d}x \\ & + L(k,h;0,\delta)f(k-z;\delta) + \int_0^h L_k(k,h;x,\delta)f(x+k-z;\delta)\mathrm{d}x \end{aligned} \tag{5}$$

where $f'(x;\delta)$ is the first-order derivative of $f(x;\delta)$ w.r.t. x. The integrals in Equations (4)-(5) may be computed very accurately and efficiently by using a Gauss-Legendre quadrature method with N points (Press et al. 1992). See also Luceno and Puig-Pey (2000).

Let $u_1,\ldots,u_N$ $(0 < u_1 < \cdots < u_N < h)$ and $\omega_1,\ldots,\omega_N$ $(0 \le w_i \le 1, i = 1,\ldots,N)$ be the corresponding Gauss-Legendre quadrature abscissas and weights respectively over the interval $[0,h]$. Let $u_0 = 0$. By replacing the integral with Gauss-Legendre quadrature abscissas and weights and evaluating at $z = u_0, u_1, \ldots, u_N$

results in an $(N+1)$-dimensional linear system:

$$L(k,h;u_i,\delta) = 1+L(k,h;u_0,\delta)F(k-u_i;\delta)+\sum_{j=1}^{N}\omega_j L(k,h;u_j,\delta)f(u_j+k-u_i;\delta), \quad (6)$$

for $i = 0, 1, \cdots, N$. Let $\mathbf{L}$ and $\mathbf{F}$ as two column vectors whose ith elements are $L(k,h;u_{i-1},\delta)$ and $F(k-u_{i-1};\delta)$ respectively. Define $\mathbf{A}_1$ an $(N+1)\times N$ matrix whose element in row i and column j is $\omega_j f(u_j+k-u_{i-1};\delta)$. Equation (6) can be written as

$$\mathbf{L} = \mathbf{1}_{N+1} + \mathbf{\Omega}_1\mathbf{L},$$

where $\mathbf{1}_{N+1}$ is a $(N+1)\times 1$ unit column vector and $\mathbf{\Omega}_1 = [\mathbf{F}, \mathbf{A}_1]$. That is, $\mathbf{L} = (\mathbf{I}-\mathbf{\Omega}_1)^{-1}\mathbf{1}_{N+1}$.

The integral Equation (5) can also be approximated similarly. Define $\mathbf{L}_k$ and $\mathbf{f}_1$ the two $(N+1)\times 1$ column vectors whose ith elements are $L_k(k,h;u_{i-1},\delta)$ and $f(k-u_{i-1};\delta)$, respectively, and $\mathbf{A}_2$ an $(N+1)\times N$ matrix whose element in row i and column j is $\omega_j f'(u_j+k-u_{i-1};\delta)$. Then Equation (5) can be written as

$$\mathbf{L}_k = \mathbf{\Omega}_2\mathbf{L} + \mathbf{\Omega}_1\mathbf{L}_k, \quad (7)$$

where $\mathbf{\Omega}_2 = [\mathbf{f}_1, \mathbf{A}_2]$. Clearly, after the determination of the ARL vector, $\mathbf{L}$, one can compute the ARL gradient w.r.t. k as $\mathbf{L}_k = (\mathbf{I}-\mathbf{\Omega}_1)^{-1}\mathbf{\Omega}_2\mathbf{L}$.

3.2 Estimate of $L_h(k,h;z,\delta)$

Based on the chain rule, differentiating both sides of Equation (4) w.r.t. h yields

$$\begin{aligned} L_h(k,h;z,\delta) &= F(k-z;\delta)L_h(k,h;0,\delta) + L(k,h;h,\delta)f(h+k-z;\delta) \\ &\quad + \int_0^h L_h(k,h;x,\delta)f(x+k-z;\delta)\mathrm{d}x. \end{aligned} \quad (8)$$

Define $\mathbf{L}_h$ and $\mathbf{f}_2$ the two $(N+1)\times 1$ column vectors whose ith elements are $L_h(k,h;u_{i-1},\delta)$ and $f(h+k-u_{i-1};\delta)$, respectively. The above equation can be written in the matrix form as

$$\mathbf{L}_h = \mathbf{f}_2 + \mathbf{\Omega}_3\mathbf{L} + \mathbf{\Omega}_1\mathbf{L}_h, \quad (9)$$

where $\mathbf{\Omega}_3 = \mathbf{f}_2\boldsymbol{\alpha}^T(h)$. After determining the ARL matrix, $\mathbf{L}$, one can obtain the ARL gradient w.r.t. h as $\mathbf{L}_h = (\mathbf{I}-\mathbf{\Omega}_1)^{-1}(\mathbf{f}_2+\mathbf{\Omega}_3\mathbf{L})$.

3.3 Estimate of $\Theta'(k)$

Differentiating both sides of $L(k,h;z,0)$ =ARL$_0$ w.r.t. k yields

$$L_k(k,h;z,0) + \eta'(k)L_h(k,h;z,0) = 0.$$

Therefore, $\eta'(k) = -L_k(k,h;z,0)/L_h(k,h;z,0)$. Based on the chain rule, $\Theta'(k)$ in Equation (3) is given by

$$\Theta'(k) = G_k - G_h L_k(k,h;z,0)/L_h(k,h;z,0),$$

Table 1: The optimal k^* values of CUSUM charts obtained based on the false position method and Monte Carlo simulations under random shifts when $\text{ARL}_0 = 400$

	Distributions of δ			
Method	U[0.5,4]	T[0.5,1.5,4]	T[0.5,3,4]	TN[2.25, 0.5]
False Position	0.8211	0.8439	1.058	0.9771
Simulation	0.79	0.83	1.07	0.97

where G_k and G_h are partial-derivatives of $G(k, h)$ w.r.t. k and h, respectively, i.e., $G_k = \int_a^b \omega(\delta)L_k(k, h; z, \delta)g(\delta)\mathrm{d}\delta$, and $G_h = \int_a^b \omega(\delta)L_h(k, h; z, \delta)g(\delta)\mathrm{d}\delta$. When the partial derivatives of ARL, $L_k(k, h; z, \delta)$ and $L_h(k, h; z, \delta)$, are obtained, G_k and G_h can be similarly evaluated using the Gauss-Legendre quadrature method. For example, G_k can be approximated by

$$G_k = \sum_{j=1}^{M} v_j \omega(\delta_j)L_k(k, h; z, \delta_j)g(\delta_j),$$

where v_j and δ_j are the corresponding Gauss-Legendre quadrature abscissas and weights over $[a, b]$. We used $N = 20$ and $M = 20$ to obtain accurate approximations of the partial derivatives throughout the remaining discussion of this paper,

4 Searching for the Optimal k Value

The optimal k value, k^*, can be obtained by finding the solution of $\Theta'(k) = 0$ in Equation (3). Root searching algorithms can be used to find the root of $\Theta'(k) = 0$. In this paper, we consider the false position method. The Illinois algorithm proposed by Dowell and Jarratt (1971) can be used to speed up the convergence of the false position method.

To assess the accuracy and efficiency of the false position method, we consider the examples discussed in Ryu et al.(2010). Following Ryu et al.(2010), four different types of random shift distributions $g(\delta)$ over the range $[a = 0.5, b = 4]$ were considered, including (1) the uniform distribution $U[0.5, 4]$; (2) the triangle distribution with mode 1.5, $T[0.5, 1.5, 4]$; (3) the triangle distribution with mode 3, $T[0.5, 3, 4]$, and (4) the truncated normal distribution with mean 2.25 and variance 0.5, $TN[2.25, 0.5]$. The weight function considered is $\omega(\delta) = 1 + \delta^2$.

Table 1 compares the optimal k value obtained by using the false position method and simulations. Clearly, the k^* value attained using the false position method tends to more accurate than that obtained by using simulations. Our results can be approximated with 4 decimal places or even more while the simulated results are often approximated with 2 decimal places (based on 1,000 replicates) with the practical consideration of computation load.

We also compare the efficiency of the false position method with simulations in terms of execution time. From Table 2, it is clear that the CPU time of Monte Carlo simulation method is around 7-8 times as long as that based on the false position method. Note that only a small number of replicates (i.e., 1,000 random δ's) was used in the simulation of Ryu et.al. (2010), the CPU time will increase dramatically if a large number of replicates were used.

Table 2: Comparison of CPU times (in seconds) for searching the optimal k^* value between the simulation and false position method when ARL_0=400

	$\delta \sim$U[0.5,4]	T[0.5,1.5,4]	T[0.5,3,4]	TN[2.25,0.5]
False position	2.5	2.4	2.2	2.4
Simulation	16.1	15.8	15.7	15.8

5 Conclusions

The focus of this paper is to develop an efficient design approach for the CUSUM control chart under an unknown mean shift by using gradient method. Our experiments demonstrate that the gradient method is more efficient and provides more accurate approximation results than exhaustive search based on Monte Carlo simulation.

References

Chen A. and Chen Y.K., (2007). Design of EWMA and CUSUM control charts subject to random shift sizes and quality impacts, *IIE Transactions*, 39, pp. 1127-1141.

Dowell M., Jarratt, P. (1971) A modified regula falsi method for computing the root of an equation, BIT Numerical Mathematics, 11, 168–174.

Fu M.C. and Hu J.Q., (1999). Efficient Design and Sensitivity Analysis of Control charts Using Monte Carlo Simulation, *Management Science*, 45(3), pp. 395-413.

Lele S., (1996). *Steady state analysis of three process monitoring procedures in quality control.* Ph.D. dissertation, University of Michigan, Ann Arbor.

Luceno A. and Puig-Pey, J. (2000). Evaluation of the run-length probability distribution for CUSUM charts: assessing chart performance, *Technometrics*, 42, 411-416.

Page E.S. (1954) Continuous inspection schemes. *Biometrika*, **42**, 243-254.

Pollak M. (1985). Optimal detection of a change in distribution. *Annals of Statistics*, 13, 206-227.

Press W.H., Teukolsky S.A., Vetterling W.T., Flannery B.P.: *Numerical recipes.* Cambridge, U.K. Cambridge University Press (1992)

Reynolds Jr., M.R. and Lou J., (2010). An evaluation of a GLR control chart for monitoring the process mean, *Journal of Quality Technology*, 42(3), 287-310.

Ritov Y. (1990). Decision theoretic optimality of the cusum procedure. *The Annals of Statistics*, 18(3), 1464-1469.

Ryu J. Wan H. and Kim S., (2010). Optimal design of a CUSUM chart for a mean shift of unknown size, *Journal of Quality Technology*, 42(3), 311-326.

Siegmund D. (1985) *Sequential analysis: tests and confidence intervals*, Spring-Verlag, New York.

Wu Z., Shamsuzzaman M. and Pan E.S., (2004). Optimal design of control charts based on Taguchi's loss function and random process shifts, *International Journal of Production Research*, 42(2), 379-390.