The Korean Journal of Applied Statistics (2009)
22(6), 1131–1142

Design of Variance CUSUM

Eun-kyung Lee[1] · Sung Hoon Hong[2] · Yoon Dong Lee[3]

[1]Ewha Womans University; [2]Chonbuk National University; [3]Sogang University

(Received July 2009; accepted August 2009)

Abstract

We suggest a fast and accurate algorithm to compute ARLs of CUSUM chart for controling process variance. The algorithm solves the characteristic integral equations of CUSUM chart (for controling variance). The algorithm is directly applicable for the cases of odd sample sizes. When the sample size is even, by using well-known approximation algorithm combinedly with the new algorithm for neighboring odd sample sizes, we can also evaluate the ARLs of CUSUM charts efficiently and accurately. Based on the new algorithm, we consider the optimal design of upward and downward CUSUM charts for controling process variance.

Keywords: CUSUM, process variance, chi-squared distribution.

1. Introduction

Statistical process control is an important tool for quality control. X-bar chart and R chart are well-known methods of statistical process control. X-bar chart is used to control the process mean. R chart and S chart aim to control process variance. Control chart techniques are viewed as an application of statistical hypotheses test. Control chart is used to test whether the process is out-of-control or in-control, repeatedly.

In hypotheses testing, Wald's sequential probability ratio test(SPRT) is known as the best testing method, in the meaning that it achieves the pre-specified type-I error probability and type-II error probability at the smallest sample size (and so at the minimum sample cost). Cumulative sum(CUSUM) chart is an application of SPRT to process control problem. The CUSUM chart suggested by Page (1954) is a sequential version of X-bar chart for controling process mean. The sequential version of S chart is also studied by many authors. Ramirez (1989), Ramirez and Juan (1989), Box and Ramirez (1991a, 1991b, 1991c), Howell (1987), Chang and Gan (1995) considered and tested CUSUM procedures for controling process variance.

Ramirez and his co-authors considered statistics of the form $S_t = \sum_{i=1}^{t}\{(X_i-\mu)^2-k\}$ for normally distributed quality characteristic X_i with mean μ. When each sample X_i has multiple observations $(X_{i,1},\dots,X_{i,n})$, they suggested to use $S_t = \sum_{i=1}^{t}\{(\bar{X}_i-\bar{\bar{X}})^2-k\}$ where $\bar{X}_i$ denotes the mean of

Research of the first author was supported by Priority Research Centers Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology (2009-0093827). Reasearch of the third author was supported by Sogang University Reasearch Grant of 2009.

[3]Corresponding author: Associate professor, Sogang Business School, 1st street Sinsoo-Dong, Mapo-Goo, Seoul 121-742, Korea. E-mail: widylee@gmail.com

i^{th} sample, *i.e.* $\bar{X}_i = (1/n)\sum_{j=1}^{n} X_{i,j}$, and $\bar{\bar{X}}$ denotes the total mean. The value k is the reference value. It is pointed by authors that the statistic S_t suggested by Ramirez and his co-authors is more adequate to control process mean rather than to process variance (Chang and Gan, 1995).

Chang and Gan (1995) considered two kinds of CUSUM charts for controling process variance. The one is variance CUSUM, using the sample variance $Q_i = 1/(n-1)\sum_{j=1}^{n}(X_{i,j} - \bar{X}_i)^2$ of each sample, and the other is log-variance CUSUM, using logarithmic-transformed sample variance $\log(Q_i)$. Chang and Gan (1995) studied the properties of these two versions and compared with others. They showed variance CUSUM has better performance than log-variance CUSUM. However, they suggested to use log-variance CUSUM, in the reason that the average run length(ARL) of log-variance CUSUM is easily approximated by using the well-known results of CUSUM chart for controling mean of normally distributed quality characteristic. They mentioned, computing ARLs of variance CUSUM is 'intractable' and needs 'extensive computing work'. To evaluate the performance of variance CUSUM, for comparison with other procedures, they relied on simulation method. Also they mentioned, they used trial and error approach for selecting the appropriate values of design parameters of log-variance CUSUM, because the log-variance CUSUM does not follow the exact distribution assumed in the model.

The sample variance of normally distributed observations follows the chi-squared distribution, which is a case of gamma distribution. When each sample has n observations and the observations are normally distributed with mean μ and variance σ^2, the sample variance follows gamma distribution of shape parameter $\nu = (n-1)/2$ and scale parameter σ^2/ν; say $G(\nu, \sigma^2/\nu)$. The SPRT and CUSUM procedures on gamma distribution has studied by many authors. Regula (1975) tried a simple approximation method in a very restrictive case. Vardeman and Ray (1985) studied ARLs of CUSUM chart for the quantities distributed by exponential distribution, *i.e.* $G(1, 1)$. Stadje (1987) obtained the exact solutions of the integral equations describing the characteristics of SPRT for exponential distribution. Also, Gan (1992), Gan and Choi (1994), Gan (1994) studied more specific topics related to CUSUM procedures on exponentially distributed quantities. Kohlruss (1994) studied to extend the results to Erlang distribution, *i.e.* $G(n, 1)$ with integer shape parameter n. Lee (2004) suggested a method unifying all related previous researches, by extending the method used by Vardeman and Ray (1985).

In this paper, we propose a fast and accurate algorithm to compute ARLs of variance CUSUM chart. The algorithm is developed based on the solutions shown by Lee (2004). We derived simpler forms of the solutions than the ones in there, and devised an algorithm to avoid numerical difficulty in handling ill-conditioned matrix. The solution is only applicable for the cases of odd sample sizes. When the sample size n is even, we use the well-known approximation algorithm. Easy-to-compute information for neighboring odd integers can be used to increase the speed of approximation algorithm for even integers. Based on the new much faster algorithm, we consider the optimal design of variance CUSUM charts in both of upward and downward direction.

2. Variance CUSUM Charts

When the quality characteristic is assumed to be normally distributed with a process mean μ and a process variance σ^2, the process mean is controlled by monitoring the sample mean $\bar{X}_i$. The sample variance $Q_i = 1/(n-1)\sum_{j=1}^{n}(X_{i,j} - \bar{X}_i)^2$ is a statistic to monitor the process variance σ^2 and it follows gamma distribution with shape parameter $\nu = (n-1)/2$ and scale parameter σ^2/ν.

Controlling the process variance σ^2 is performed by monitoring whether the sample variance is far from the pre-specified value σ_0^2, which is referred to 'in-control variance'. In-control variance σ_0^2 can be estimated from the past experience of the process performance. For more details about how to estimate σ_0^2, refer to Lucas (1976). In designing control schemes, we need a value σ_1^2 for referencing the variance of process in out-of-control state. We call σ_1^2 'out-of-control variance'. The statistical hypotheses test with the statistic Q_i, for the hypotheses $H_1 : \sigma^2 = \sigma_1^2$ against $H_0 : \sigma^2 = \sigma_0^2$ ($\sigma_0^2 \neq \sigma_1^2$) is equivalent to the test for $H_1 : \sigma^2/\sigma_0^2 = \sigma_1^2/\sigma_0^2$ against $H_0 : \sigma^2/\sigma_0^2 = 1$ with the statistic Q_i/σ_0^2. We call the property invariance of scale transformation for testing variance. From the property, we assume $\sigma_0^2 = 1$ without loss of generality.

The SPRT for the hypotheses $H_1 : \sigma^2 = \sigma_1^2$ ($\sigma_1^2 \neq 1$) against $H_0 : \sigma^2 = 1$ is performed by following steps; For each $t = 1, 2, \ldots$ and appropriately defined l and h ($l \leq 0 \leq h$) and $R_0 = s \in [l, h]$,

S1) set $R_t = R_{t-1} + (Q_t - k)$, where $k = (\sigma_1^2 \log \sigma_1^2)/(\sigma_1^2 - 1)$,

S2) if $R_t > h$, then conclude H_1 is right, if $R_t < l$, then conclude H_0 is right,

S3) otherwise take a more sample, *i.e.* set t to be $t + 1$ and go on.

The CUSUM procedure to control the process variance is a variant of the SPRT. The CUSUM procedure issues the out-of-control signal when $R_t^U > h$ for upward CUSUM to detect upward change ($\sigma_1^2 > 1$), and when $R_t^D < l$ for downward CUSUM to detect downward change ($\sigma_1^2 < 1$). The CUSUM statistics R_t^U and R_t^D are updated by the rule,

$$R_t^U = \max(0, R_{t-1}^U) + (Q_t - k) \quad \text{and} \quad R_t^D = \min(0, R_{t-1}^D) + (Q_t - k)$$

for each $t = 1, 2, \ldots$, where $R_0^U = s \in [0, h]$ and $R_0^D = s \in [0, h]$. After issuing the signal, the same procedure is repeated with the same initial value s.

Important characteristics of CUSUM charts are summarized by average run length(ARL). We will use $H(s)$ and $L(s)$ to denote the ARLs of upward CUSUM and downward CUSUM as functions of initial value s. The functions $H(s)$ and $L(s)$ are written in the form of integral equations, as follows;

$$H(s) = 1 + H(0)F_\nu(k - s) + \int_0^h H(x) f_\nu(x - s + k)dx, \quad s \in [0, h], \tag{2.1}$$

$$L(s) = 1 + L(0)(1 - F_\nu(k - s)) + \int_l^0 L(x) f_\nu(x - s + k)dx, \quad s \in [l, 0], \tag{2.2}$$

where $F_\nu(x)$ and $f_\nu(x)$ are the cumulative distribution function(CDF) and the probability density function(PDF) of each Q_i. That is, $F_\nu(x)$ and $f_\nu(x)$ are CDF and PDF of gamma distribution $G(\nu, \sigma^2/\nu)$, respectively.

When we need to distinguish the variables and the parameters between upward procedure and downward procedure, indexed notations will be used; R_t^U, h_U, k_U, s_U for upward scheme, R_t^D, l_D, k_D, s_D for downward scheme. Since the boundary value l is negative, we define a new positive value $h_D = -l$. As long as there is no confusion, we will drop the indices. The ARL function $L(s)$ for downward CUSUM is defined on negative value $s \in [-h_D, 0]$. When we need parallel comparison with $H(s)$ and $L(s)$, we use the function $L_p(s) = L(s + h)$ shifted to the region $s \in [0, h]$.

To detect both of upward and downward changes of variance, two-sided CUSUM procedure is used. Two-sided CUSUM chart is a combined procedure of upward CUSUM and downward CUSUM as component procedures. As long as the out-of-control signals are not issued simultaneously from the component procedures, the ARL of two-sided CUSUM is given as the half of harmonic mean of

ARLs of two component one-sided procedures. More generally, when $HL(s_U, s_D)$ denotes the ARL function of two-sided CUSUM procedure w.r.t. the initial values s_U and s_D, Lucas (1985) showed

$$HL(s_U, s_D) = \frac{H(s_U)L(0) + H(0)L(s_D) - H(0)L(0)}{H(0) + L(0)}.$$

For more details, refer Lucas (1985).

3. Solution and Numerical Algorithm

Out of many ways to evaluate ARL of CUSUM chart, the most reliable way is to solve the characteristic integral equation directly. Goel and Wu (1971) solved the characteristic equation for CUSUM chart for controling process mean, numerically by using Gaussian Hermite quadrature. In this section, we show the algorithms to solve the characteristic integral Equations (2.1) and (2.2) for variance CUSUM.

The integral Equations (2.1) and (2.2) can be solved analytically when the sample size $n = 2\nu + 1$ is odd. As pointed by Kohlruss (1994), the kernel $f_\nu(x-s+k)$ of the Equations (2.1) and (2.2) for odd integer n is separable and the equation can be solved analytically. On the other hand, the kernel for even integer n is not separable and we have not found the method to solve the equations analytically for even integer n. Only numerical methods were tested for even integer n. The numerical method is relatively simple to implement and applicable to both of even and odd integers, but the method is less accurate and needs more time for computing. In the followings we will review the numerical approximation method first, and then provide a fast analytic algorithm applicable for odd integer n.

3.1. Review of numerical methods

Lee (2004) showed that the analytic solutions of Equations (2.1) and (2.2) for odd sample size have the form of piecewise polynomial. The solutions have different form on each interval $[ik, (i+1)k]$, $i = 0, 1, \ldots$. Gaussian quadrature method applied to polynomial is known to give highly precise results, but the accuracy of the results to piecewise polynomials are not guaranteed. When we tested to our cases of odd sample size, which is expected to have monotone solution, the results obtained by Gaussian quadratures are even oscillating up and down at their tail parts. The form of the kernel $f_\nu(x-s+k)$, which is differently defined before and after of $x = s-k$, is attributable to the inaccuracy of the Gaussian quadrature methods for this problem.

Numerically approximated solutions of Equations (2.1) and (2.2) for a positive integer n are tested by Ramirez and Juan (1989). Instead of using abscissa of Gaussian quadrature, the method used by Ramirez and Juan (1989) is based on simple assumption that the solutions are nearly constant in a very short interval. The approximated algorithm is much slower than Gaussian quadrature method, but gives relatively better results. The method used by Ramirez and Juan (1989) is basically equivalent to the discrete Markov chain approximation method suggested by Brook and Evans (1972) in evaluating ARLs of CUSUM. In the followings, we explain the method to approximate the functions $H(s)$. The function $L(s)$ is approximated in the same way.

Let $\{a_0, \ldots, a_q\}$ be an evenly spaced partition of $[0, h]$, satisfying $a_0 = 0$ and $a_q = h$. Take sequences of x_j and s_j by $x_j = s_j = (a_{j-1} + a_j)/2$ for $j = 1, 2, \ldots, q$. The positive integer q can be chosen arbitrary in the consideration of numerical precision. For notational convenience, assume also $x_0 = s_0 = a_0$ and $a_{-1} = -\infty$. Then, the values $H(s_j)$, $j = 0, 1, \ldots, q$ are approximated by the values $\hat{H}(s_j)$ satisfying

$$\hat{H}(s_i) = 1 + \hat{H}(x_0)F_\nu(k - s_i) + \sum_{j=1}^{q} \hat{H}(x_j) \int_{a_{j-1}}^{a_j} f_\nu(x_j - s_i + k)dx, \quad j = 0, 1, \ldots, q. \tag{3.1}$$

The Equation (3.1) is simply converted to matrix form $(I - B)\hat{H} = \mathbf{1}_{q+1}$, where the vector $\hat{H}$ is $(\hat{H}(s_0), \ldots, \hat{H}(s_q))'$, and the matrix B is the $(q+1) \times (q+1)$ matrix of which $(i,j)^{th}$ elements are $b_{i,j} = F_\nu(a_j - s_i + k) - F_\nu(a_{j-1} - s_i + k)$, for $i, j = 0, \ldots, q$. The vector $\mathbf{1}_{q+1}$ is the column vector of size $q + 1$, all of which elements are 1.

The numerical approximation method has advantage that the algorithm applicable for any positive integer n, but the method is much slower than the algorithm based on analytic solution. To approximate precisely, q must be quite large. However, the larger q is used, the more times are required for computing and the more numerical errors come into the result cumulatively. And even, we can not distinguish, only from the approximated results, whether the used integer q is sufficiently large and the result is sufficiently accurate. The approximation algorithm is practically less useful because of the questionable accuracy of the results as well as its slow speed. In the next we suggest faster and more reliable algorithm.

3.2. Algorithm for odd sample size

When the sample size n is odd, the kernel $f_\nu(x - s + k)$ of the Equations (2.1) and (2.2) is separable and the equation can be solved analytically. Following the approach to obtain the solutions of the characteristic integral equations, we suggest an efficient algorithm in the followings.

When the sample size $n = 2\nu + 1$ is odd, the sample variance Q_i follows the gamma distribution $G(\nu, \sigma^2/\nu)$. From the property of invariance of scale transformation, we can assume that Q_i follows the gamma distribution $G(\nu, 1)$ with integer shape parameter ν and scale parameter 1. We use the notations $H_\nu(s)$ and $H_{\nu,i}(s)$ to denote $H(s)$ for given ν and $H(s)$ on the interval $[ik,\, (i+1)k]$, $i = 0, 1, \ldots, m$, respectively. Here and hereafter m denotes the smallest integer greater than or equal to h/k. The solution of the integral Equation (2.1) is

$$\begin{aligned} H_{\nu,i}(s) &= 1 + i + H(0) - e^{s-ik} \sum_{j=0}^{i} \sum_{l=0}^{\nu-1} a_{l,j} e_{l+\nu(i-j)}((l + i - j)k - s), \\ H(0) &= -m + \sum_{j=0}^{m} \sum_{l=0}^{\nu-1} \alpha_{l,j} p_{0,l}^{j}(h), \end{aligned}$$

where $e_\nu(x) = \sum_{k=0}^{\nu} x^k/k!$ and $p_{u,l}^{j}(s) = e^{h-mk+k} e_{l+\nu(m-j)-u}((l + m - j - 1)k - s)$. The constants $\alpha_{l,j}$, $j = 0, 1, \ldots, m$, $l = 0, 1, \ldots, \nu - 1$ are determined by

$$\sum_{j=0}^{i} \sum_{l=0}^{\nu-1} q_{u,l}^{i,j} \alpha_{l,j} = 1 \quad \text{and} \quad \sum_{j=0}^{m} \sum_{l=0}^{\nu-1} q_{u,l}^{m+1,j} \alpha_{l,j} = \delta_0(u),$$

for $i = 1, 2, \ldots, m$ and $u = 0, 1, \ldots, \nu - 1$, where

$$\begin{aligned} q_{u,l}^{i,j} &= e_{l+\nu(i-j)-u}((l - j)k) - e^k e_{l+\nu(i-j-1)-u}((l - j - 1)k), \\ q_{0,l}^{m+1,j} &= e_l(lk) \cdot \delta_0(j), \\ q_{u,l}^{m+1,j} &= p_{u,l}^{j}(h) - p_{u-1,l}^{j}(h). \end{aligned}$$

The function $\delta_c(j)$ has values 1 when $j = c$ and 0 when $j \neq c$. For the downward CUSUM, we have

$$\begin{aligned} L_{\nu,i}(s) &= 1 + i - e^{s-ik} \sum_{j=0}^{i} \sum_{l=0}^{\nu-1} a_{l,j} e_{l+\nu(i-j)}((l+i-j)k - s), \\ L(0) &= m - e^{-h} \sum_{j=0}^{m-1} \sum_{l=0}^{\nu-1} a_{l,j} p_{\nu,l}^{j}(0). \end{aligned}$$

The constants $\alpha_{l,j}$, $j = 0, 1, \ldots, m$, $l = 0, 1, \ldots, \nu - 1$ for downward CUSUM are determined by the conditions;

$$\sum_{j=0}^{i} \sum_{l=0}^{\nu-1} q_{u,l}^{i,j} \alpha_{l,j} = 1 \quad \text{and} \quad \sum_{j=0}^{m} \sum_{l=0}^{\nu-1} q_{u,l}^{m+1,j} \alpha_{l,j} = d(u),$$

for $i = 1, \ldots, m$ and $u = 0, \ldots, \nu - 1$, where

$$\begin{aligned} d(u) &= m \left[e_{\nu-u-1}(-h) e^{h} - 1 \right], \\ q_{u,l}^{m+1,j} &= (1 - \delta_m(j)) e_{\nu-u-1}(-h) p_{\nu,l}^{j}(0) - p_{u,l}^{j}(h). \end{aligned}$$

The conditions are shown in simpler matrix form. Assume $\boldsymbol{\alpha}_j$'s are column vectors of size ν, of which elements are $\alpha_{l,j}$, $l = 0, \ldots, \nu - 1$, and $\boldsymbol{q}^{i,j}$'s are matrices $\{q_{u,l}^{i,j}\}_{u,l}$ of size $\nu \times \nu$. Also, $\boldsymbol{d}$ is a column vector with elements $\delta_0(u)$, $u = 0, 1, \ldots, \nu - 1$ for upward CUSUM or $d(u)$, $u = 0, 1, \ldots, \nu - 1$ for downward CUSUM. The conditions are re-expressed in linear equation $Q\boldsymbol{\alpha} = \boldsymbol{\delta}$, where $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_0^t, \boldsymbol{\alpha}_1^t, \ldots, \boldsymbol{\alpha}_m^t)^t$ is a column vector of size $\nu(m+1)$, and

$$Q = \begin{pmatrix} \boldsymbol{q}^{1,0} & \boldsymbol{q}^{1,1} & \mathbf{0} & \cdots & \mathbf{0} \\ \boldsymbol{q}^{2,0} & \boldsymbol{q}^{2,1} & \boldsymbol{q}^{2,2} & \cdots & \mathbf{0} \\ \cdots & \cdots & \cdots & \cdots & \mathbf{0} \\ \boldsymbol{q}^{m,0} & \boldsymbol{q}^{m,1} & \boldsymbol{q}^{m,2} & \cdots & \boldsymbol{q}^{m,m} \\ \boldsymbol{q}^{(m+1),0} & \boldsymbol{q}^{(m+1),1} & \boldsymbol{q}^{(m+1),2} & \cdots & \boldsymbol{q}^{(m+1),m} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\delta} = \begin{pmatrix} \mathbf{1}_\nu \\ \mathbf{1}_\nu \\ \cdots \\ \mathbf{1}_\nu \\ \boldsymbol{d} \end{pmatrix}$$

are block matrix of size $\nu(m+1) \times \nu(m+1)$ and column vector of size $\nu(m+1)$, respectively. The constants $\alpha_{l,j}$'s are obtained by $\boldsymbol{\alpha} = Q^{-1}\boldsymbol{\delta}$. Since the matrix Q goes to be ill-conditioned for large ν and m, we need to devise an algorithm to avoid the difficulty of ill-conditioned property in inverting the matrix Q for large ν and m. We used the block structure of the matrix Q and applied block-wise forward substitution method, as follows.

Algorithm: follow steps 1, 2, 3

Step 1. $\boldsymbol{v}^{(m+1),j} = \boldsymbol{q}^{(m+1),j}$, $j = 0, 1, \ldots, m$ and $\boldsymbol{b}^{m+1} = \boldsymbol{d}$

Step 2. In the order of $i = m, (m-1), \ldots, 1$

$$\begin{aligned} \boldsymbol{v}^{i,j} &= \boldsymbol{q}^{i,j} - \boldsymbol{q}^{i,i} [\boldsymbol{v}^{(i+1),i}]^{-1} \boldsymbol{v}^{(i+1),j}, \quad j = 0, 1, \ldots, (i-1) \\ \boldsymbol{b}^{i} &= \mathbf{1}_\nu - \boldsymbol{q}^{i,i} [\boldsymbol{v}^{(i+1),i}]^{-1} \boldsymbol{b}^{(i+1)} \end{aligned}$$

Step 3. $\boldsymbol{\alpha}_0 = [\boldsymbol{v}^{1,0}]^{-1} \boldsymbol{b}^{1}$ and $\boldsymbol{\alpha}_i = [\boldsymbol{v}^{(i+1),i}]^{-1} [\boldsymbol{b}^{(i+1)} - \sum_{j=0}^{i-1} \boldsymbol{v}^{(i+1),j} \boldsymbol{\alpha}_j]$, $i = 1, 2, \ldots, m$.

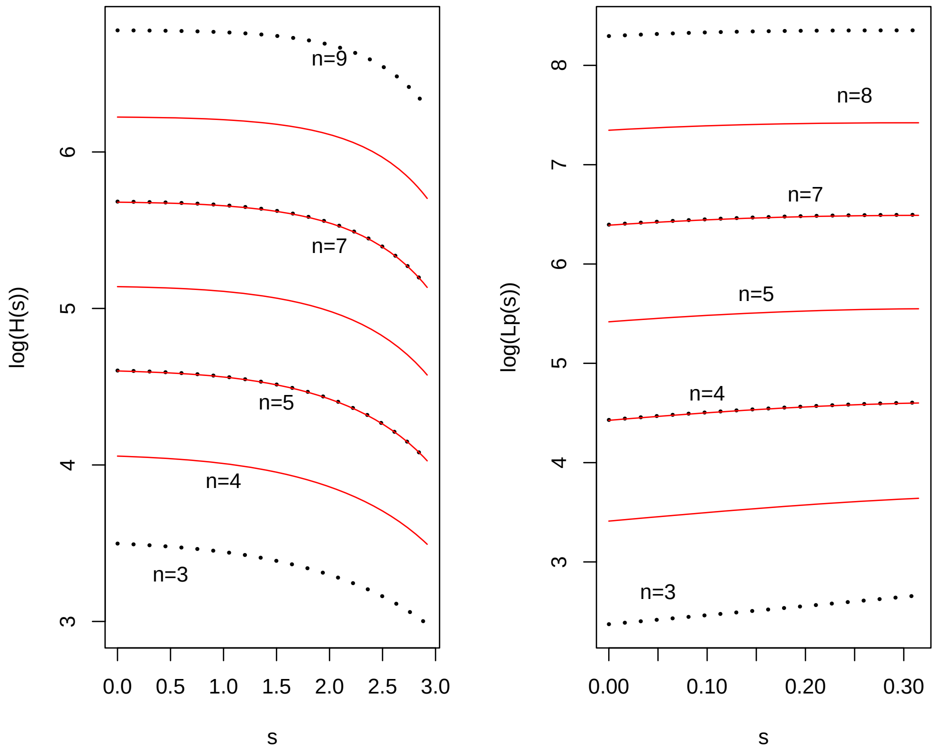

Figure 3.1. Comparison of the numerical results: $\log(H(s))$ and $\log(\hat{H}(s))$ are at left, $\log(L(s))$ and $\log(\hat{L}(s))$ at right. For upward CUSUM, $h = 2.921$, $k = 1.285$ are used, and for downward CUSUM $h = 0.3150$, $k = 0.3491$ are used. The approximated values $\log(\hat{H}(s))$ and $\log(\hat{L}(s))$ are shown in thin solid lines for, $n = 4, 5, 6, 7$ and the analytic solutions $\log(H(s))$ and $\log(L(s))$ are shown in thick dotted lines for $n = 3, 5, 7, 9$. For numerical approximation, $q = 500$ and intervals of equal length, $[a_{j-1}, a_j]$s were used.

3.3. Testing the algorithm and application

To check the accuracy and the consistency of the two approaches (analytic approach and numerical approximation approach), we compared the log-transformed values of the approximated solutions $\hat{H}(s)$ and $\hat{L}(s)$ and the analytic solutions $H(s)$ and $L(s)$. The cases of $h = 2.921$, $k = 1.285$, for upward CUSUM, $h = 0.3150$, $k = 0.3491$, for downward CUSUM, are tested. In Figure 3.1, each dotted line shows the values of $\log(H(s))$ and $\log(L(s))$ when the sample sizes n are 3, 5, 7, 9, and each thin solid line shows the approximated values of $\log(\hat{H}(s))$ and $\log(\hat{L}(s))$ when the sample sizes are 4, 5, 6, 7, 8. For the common sample sizes 5 and 7, the lines obtained by two approaches are matched almost exactly. For the numerical approximation we used $q = 500$ and intervals of equal length were used for $[a_{j-1}, a_j]$, $j = 1, \ldots, q$.

The functions $H(s)$ and $L(s)$ are obtained by assuming the null hypothesis is true; that is, by assuming the true process variance σ^2 is equal to the in-control variance $\sigma_0^2 = 1$. To see the effects of true process variance $\sigma^2 = c$, we use the notations $H_c(s)$ and $L_c(s)$ indexed by c. The curves $H_c(0)$ and $L_c(0)$ are termed to ARL profiles. Let $H(s; h, l, k)$ denote $H(s)$ for given h, l and k. By the property of invariance of scale transformation, $H_c(s)$ is equal to $H(s/c; h/c, l/c, k/c)$. The same

Table 3.1. Comparison of the value of the ARLs: The ARLs of log-variance CUSUM(LV CUSUM) and variance CUSUM(V CUSUM) are compared in two cases (I) and (II). ARLs of variance CUSUM computed by simulation method and analytic algorithm are compared. Case (I) adopts $k = 0.309$, $h = 1.210$ for log-variance CUSUM and $k = 1.285$, $h = 2.921$ for variance CUSUM. Case (II) adopts $k = 0.451$, $h = 0.896$ for log-variance CUSUM and $k = 1.460$, $h = 2.331$ for variance CUSUM.

$\sqrt{c}$	$\sigma_1^2 = (1.3)^2$			$\sigma_1^2 = (1.53)^2$		
	LV-CUSUM*	V CUSUM*	Exact	LV CUSUM*	V CUSUM*	Exact
1.00	100.00	100.00±0.50	99.827	100.00	100.30±0.50	100.257
1.01	86.60	85.50±0.50	85.283	87.30	86.80±0.50	86.934
1.02	75.40	73.40±0.50	73.395	76.50	75.60±0.50	75.798
1.03	66.00	63.60±0.50	63.614	67.50	66.50±0.50	66.443
1.04	58.20	55.50±0.50	55.514	59.70	58.30±0.50	58.545
1.05	51.50	48.70±0.50	48.765	53.20	51.80±0.50	51.844
1.10	30.20	28.10±0.50	27.875	31.60	30.00±0.50	30.256
1.20	13.80	13.00±0.50	12.780	14.40	13.60±0.50	13.648
1.30	8.15	7.75±0.01	7.742	8.31	8.00±0.01	7.970
1.40	5.63	5.46±0.01	5.464	5.61	5.46±0.01	5.455
1.50	4.29	4.22±0.01	4.217	4.19	4.12±0.01	4.122
2.00	2.11	2.08±0.01	2.075	1.96	1.96±0.01	1.969

$*$ marked values were shown in Chang and Gan (1995).

arguments are applied to $L_c(s)$; that is $L_c(s) = L(s/c; h/c, k/c)$.

Table 5 of Chang and Gan (1995) compared upward log-variance CUSUM and upward variance CUSUM. For the comparison, the ARLs of log-variance CUSUM were approximately computed by log-normal approximation, and the ARLs of variance CUSUM were evaluated by simulation. In the study, $n = 5$ was considered. The design parameter k and h of variance CUSUM were selected optimally by the conditions of in-control ARL and the out-of-control variance σ_1^2. They considered the cases when the in-control ARL is 100 and the values of out-of-control variances are $\sigma_1^2 = (1.3)^2$, $\sigma_1^2 = (1.4)^2$ and $\sigma_1^2 = (1.5)^2$. The design parameters of log-variance CUSUM were selected arbitrary with numerical experience of trial and error. In Table 3.1, we compare our exact results and the values Chang and Gan (1995) reported. For compact comparison, we tabulated only two cases, (I): $\sigma_1^2 = (1.3)^2$ and (II): $\sigma_1^2 = (1.5)^2$, out of the three out-of-control variance used in Chang and Gan (1995). The ARL values of variance CUSUM reported by Chang and Gan (1995) were very accurate. In table 1, only the cells typed in boldface show a little difference between the simulation results and the true values obtained by analytic approach.

4. Optimal Design of Variance CUSUM

The key criterion used for designing control chart is ARL. In-control ARL and out-of-control ARL are the ARLs when the process is in-control($\sigma^2 = \sigma_0^2$) state and out-of-control($\sigma^2 = \sigma_1^2$) state, respectively. ARL(0) and ARL(1) denote in-control ARL and out-of-control ARL, respectively. Variance CUSUM has three design parameters n, h and k, and three design criterions, σ_1^2/σ_0^2, ARL(0) and ARL(1). Without loss of generality, we can use σ_1^2, instead of σ_1^2/σ_0^2, as a design criterion, by assuming $\sigma_0^2 = 1$.

From the relationship between SPRT and CUSUM, the optimal reference value $k = (\sigma_1^2 \log \sigma_1^2)/(\sigma_1^2 - 1)$ is determined from out-of-control variance σ_1^2. Refer Lucas (1976) and Moustakides (1986) for more details. After fixing σ_1^2, two criterions ARL(0) and ARL(1), and two parameters n and h are left to determine. For each n, we can find a $h = h_0$ which attains the given ARL(0). Figure 4.1

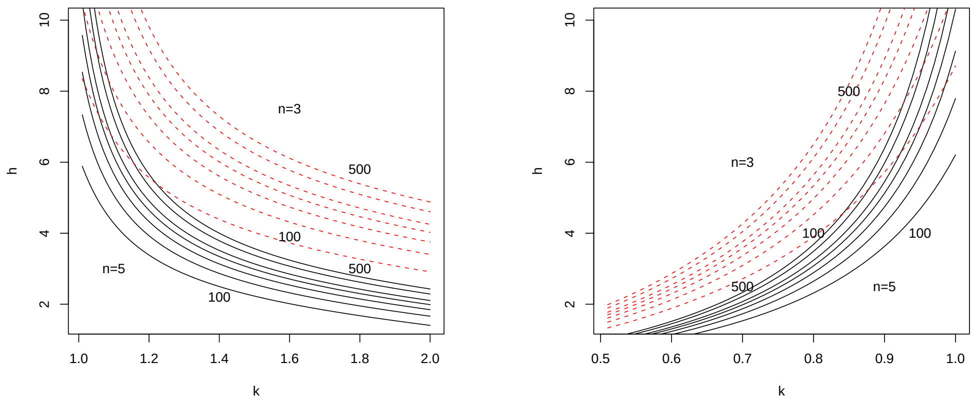

Figure 4.1. The pairs of (k, h) having the same ARL(0) for upward CUSUM (left) and downward CUSUM (right). On each plot, $n = 3$ (dotted lines) and $n = 5$ (solid lines) are tested. Each line correspond to ARL(0) of 100, 150, 200, 250, 300, 400, 500 as in the order shown in each plot.

Table 4.1. Design of upward CUSUM: For given triples of σ_1^2, ARL(0) and n, corresponding values of h_0 and ARL(1) are listed.

$c = \sigma_1^2$	ARL(0)	100		200		500	
	n	h_0	ARL(1)	h_0	ARL(1)	h_0	ARL(1)
$\sigma_1^2 = (1.2)^2$ $k = 1.1934$	2	8.8200	34.04	11.9301	46.63	16.6599	65.78
	3	5.6208	21.17	7.3799	28.23	9.9515	38.63
	4	4.2366	15.69	5.4766	20.59	7.2599	27.72
	5	3.4290	12.60	4.3920	16.32	5.7556	21.71
	7	2.5173	9.23	3.1851	11.68	4.1165	15.24
	9	2.0034	7.43	2.5158	9.22	3.2240	11.83
$\sigma_1^2 = (1.6)^2$ $k = 1.5426$	2	6.6799	7.73	8.6799	9.71	11.4600	12.44
	3	3.8888	5.01	4.9437	6.04	6.3856	7.46
	4	2.7666	3.91	3.4866	4.60	4.4600	5.56
	5	2.1329	3.31	2.6812	3.83	3.4181	4.55
	7	1.4515	2.68	1.8253	3.02	2.3226	3.49
	9	1.0836	2.35	1.3694	2.60	1.7468	2.94
$\sigma_1^2 = (2.2)^2$ $k = 1.9876$	2	5.4200	2.96	7.0400	3.53	9.2401	4.30
	3	2.9322	2.12	3.7749	2.11	4.9072	2.80
	4	1.9600	1.79	2.5400	1.98	3.3066	2.24
	5	1.4201	1.64	1.8632	1.78	2.4486	1.96
	7	0.8455	1.47	1.1550	1.56	1.5590	1.68
	9	0.5353	1.38	0.7781	1.45	1.0927	1.54

shows the lines of pairs (k, h) which have the same ARL(0). On each picture solid lines are the cases of $n = 5$, dotted lines are the cases of $n = 3$. All points passed by each line have the same ARL(0). For each line, ARL(0) is given as 100, 150, 200, 250, 300, 400, 500 in the order from bottom to top. ARL(1) is fully determined by the pair of (n, h_0). It means, for each n, ARL(1) is determined by ARL(0). When σ_1^2 and ARL(0) are given, ARL(1) varies with n. The larger n, the smaller ARL(1) is. The best policy to choose n is to take the one at which ARL(1) gets to be smaller than the required value, at the consideration of practical use. Table 4.1 and 4.2 show the values of h_0 and ARL(1) corresponding to given triples of σ_1^2, ARL(0) and n for upward CUSUM and downward CUSUM.

Table 4.2. Design of downward CUSUM: For given triples of σ_1^2, ARL(0) and n, corresponding values of h_0 and ARL(1) are listed.

$c = \sigma_1^2$	ARL(0)	100		200		500	
	n	h_0	ARL(1)	h_0	ARL(1)	h_0	ARL(1)
$\sigma_1^2 = (0.8)^2$, $k = 0.7934$	2	6.2200	38.59	8.1200	51.00	10.8801	69.10
	3	3.8118	23.07	4.8456	29.79	6.3184	39.38
	4	2.8267	16.68	3.5400	21.28	4.5533	27.75
	5	2.2521	13.08	2.8042	16.58	3.5708	21.51
	7	1.6235	9.19	2.0018	11.52	2.5210	14.78
	9	1.2753	7.13	1.5638	8.83	1.9567	11.24
$\sigma_1^2 = (0.6)^2$, $k = 0.5747$	2	3.0599	14.70	3.7600	17.96	4.7600	22.63
	3	1.7121	8.46	2.0826	10.19	2.5849	12.53
	4	1.2067	6.10	1.4600	7.28	1.7934	8.83
	5	0.9198	4.78	1.1091	5.66	1.3630	6.84
	7	0.6231	3.43	0.7523	3.99	0.9194	4.78
	9	0.4623	2.68	0.5604	3.19	0.6917	3.72
$\sigma_1^2 = (0.4)^2$, $k = 0.3491$	2	1.2601	7.25	1.5199	8.62	1.8600	10.41
	3	0.6497	4.03	0.7857	4.74	0.9550	5.64
	4	0.4466	2.92	0.5334	3.40	0.6401	3.98
	5	0.3150	2.32	0.3817	2.63	0.4782	3.09
	7	0.2162	1.64	0.2554	1.89	0.3003	2.22
	9	0.1474	1.34	0.1878	1.50	0.2307	1.73

5. Final Comments

The properties of SPRT and CUSUM procedures for the scale parameter of gamma distribution or of its sub-family have been studied by many authors. To reveal the characteristics of the test and the procedures they used various methods. Sometimes approximation methods and sometimes simulation methods were used. In this paper, we showed the solutions of the characteristic integral equations of variance CUSUM are accurately evaluated by combining analytic method and numerical approximation method. Numerical approximation method can be applied to both of odd and even sample size cases, but the approximation method is very slow for computing and the accuracy of the results is unreliable. By applying approximation algorithm combinedly with analytic solution, we considered the optimal design of CUSUM procedures. With the help of fast analytic algorithm, speed of steps of approximation algorithm was also increased by efficiently taking testing parameter values. To find an initial values of the parameters for even sample size, we used empirical approximation method taking logarithmic average of the ARLs of CUSUM for two adjacent odd sample sizes; that is, for an integer ν,

$$\log H_{\nu+0.5}(s) = \left(\frac{1}{2}\right)\left[\log H_\nu(s) + \log H_{\nu+1}(s)\right].$$

Fast initial response(FIR) feature, suggested by Lucas and Crosier (1982), is frequently considered, in spite of its restrictive advantage (*cf.* Lorden, 1971), in studies of CUSUM chart to improve the nominal performance. Our algorithm is also applicable to compute ARL of CUSUM with FIR feature, by using $H_c(s)$ and $L_c(s)$ instead of $H_c(0)$ and $L_c(0)$. Variance CUSUMs with FIR feature were compared with other control schemes in Chang and Gan (1995).

While CUSUM chart has advantage that it detects small change of mean and/or variance in short time, it has disadvantage that it takes longer time to detect large change than its non-sequential versions. To avoid the disadvantage, combined use of CUSUM chart and X-bar chart was suggested by Lucas (1982). In the same way, combining variance CUSUM and S chart is expected to improve the performance of variance CUSUM. Study on the combining variance CUSUM with X-bar chart or ordinary CUSUM chart is also expected for further research.

References

Box, G. E. P. and Ramirez, J. G. (1991a). Sequential method in statistical process monitoring. Report No. 65, Center for Quality and Productivity Improvement, University of Wisconsin-Madison, Madison, WI.

Box, G. E. P. and Ramirez, J. G. (1991b). Sequential method in statistical process monitoring. Report No. 66, Center for Quality and Productivity Improvement, University of Wisconsin-Madison, Madison, WI.

Box, G. E. P. and Ramirez, J. G. (1991c). Sequential method in statistical process monitoring. Report No. 67, Center for Quality and Productivity Improvement, University of Wisconsin-Madison, Madison, WI.

Brook, D. and Evans, D. A. (1972). An approach to the probability distribution of CUSUM run length, *Biometrika*, **52**, 539–549.

Chang, T. and Gan, F. (1995). A cumulative sum control chart for monitoring process variance, *Journal of Quality Technology*, **25**, 109–119.

Gan, F. F. (1992). Exact run length distributions for one-sided exponential CUSUM scheme, *Statistica Sinica*, **26**, 297–312.

Gan, F. F. (1994). Design of optimal exponential CUSUM control charts, *Journal of Quality Technology*, **26**, 109–124.

Gan, F. and Choi, K. (1994). Computing average run lengths for exponential CUSUM schemes, *Journal of Quality Technology*, **26**, 134–139.

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.

Howell, D. K. (1987). *A CUSUM Scheme for the Control of Process Variance*, Presented at the Joint Statistical Meetings, San Fransisco, CA.

Kohlruss, D. (1994). Exact formulas for the OC and the ASN functions of the SPRT for Erlang distributions, *Sequential Analysis*, **13**, 53–62.

Lee, Y. D. (2004). Unified solutions of integral equations of SPRT for exponential random variables, *Communications in Statistics, Series A, Theory and Method*, **33**, 65–74.

Lorden, G. (1971). Procedures for reacting to a change in distribution, *Annals of Mathematical Statistics*, **42**, 538–551.

Lucas, J. M. (1976). The design and use of cumulative sum quality control schemes, *Journal of Quality Technology*, **8**, 1–12.

Lucas, J. M. (1982). Combined Shewart-CUSUM quality control schemes, *Journal of Quality Technology*, **14**, 51–59.

Lucas, J. M. (1985). Counted data CUSUMs, *Technometrics*, **27**, 129–144.

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

Moustakides, G. V. (1986). Optimal stopping times for detecting changes in distributions, *The Annals of Statistics*, **14**, 1379–1387.

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

Ramirez, J. G. (1989). Sequential method in statistical process monitoring. Ph. D. thesis, University of Wisconsin-Madison, Madison, WI.

Ramirez, J. G. and Juan, J. (1989). A contour nomogram for designing cusum charts for variance, Report No. 33, Center for Quality and Productivity Improvement, University of Wisconsin-Madison, Madison, WI.

Regula, G. A. (1975). Optimal CUSUM procedure to detect a change in distribution for the gamma family, Ph. D. thesis, Case Western Reserve University, Ohio.

Stadje, W. (1987). On the SPRT for the mean of an exponential distribution, *Statistics & Probability Letters*, **5**, 389–395.

Vardeman, S. and Ray, D. (1985). Average run lengths for CUSUM schemes when observations are exponentially distributed, *Technometrics*, **7**, 145–150.