Malaysian Journal of Science **23**(2) : 181 - 187 (2004)

Fitting Simulated Run Length Distribution using Translated Gamma Distribution

Chong Fook Seng and **Pooi Ah Hin***

Institute of Mathematical Sciences, Faculty of Science, University of Malaya, 50603 Kuala Lumpur, Malaysia
* ahpooi@um.edu.my
Received 14 October 2003, accepted in revised form 28 December 2004

ABSTRACT Theoretically iterative formulas can be used to find the run length distribution of the two-sided CUSUM scheme. But in practice, we face the problem of requiring long computing time to get the run length distribution using the iterative formulas. An alternative way of finding the run length distribution is by means of simulation. As the simulated run length distribution exhibits random fluctuation when compared with the actual distribution, we can improve the simulated distribution by fitting it with a smooth curve. It is found that the curve given by the right tail of the translated gamma distribution gives a good fit to the simulated results. The methods based on moments and least squares are used to estimate the parameters of the translated gamma distribution. Numerical examples in the case when the observations have an exponential or normal distribution are presented.

ABSTRAK Dari segi teori, rumus lelaran boleh digunakan untuk mencari taburan larian panjang bagi carta kawalan hasiltambah longgokan (carta 'CUSUM'). Tetapi pada praktisnya, kita menghadapi masalah yang didatangkan oleh masa yang panjang untuk mendapatkan taburan larian panjang dengan menggunakan rumus lelaran. Satu cara alternatif yang boleh digunakan untuk mencari taburan larian panjang ialah melalui simulasi. Oleh kerana taburan larian panjang yang berdasarkan simulasi berbeza secara rawak dari hasil benar, kita memperbaiki taburan yang berdasarkan simulasi dengan menyuaikan taburan dengan suatu lengkung licin. Didapati bahawa lengkung yang diberikan oleh sisi kanan taburan gama yang telah ditranslasikan memberi suaian yang baik bagi hasil simulasi. Kaedah yang berdasarkan momen dan kuasa dua terkecil digunakan untuk menganggar parameter dalam taburan gama yang telah ditranslasikan. Contoh berangka dalam kes bila cerapan mempunyai taburan eksponen atau normal dilaporkan.

(Two-sided CUSUM scheme,run length distribution,translated gamma distribution)

INTRODUCTION

In the Shewhart control charts, such as the $\bar{X}$-chart and R-chart, a plotted point represents information corresponding to that observation only. It does not use information from previous observations. On the other hand a cumulative sum chart, usually called a CUSUM chart, uses information from all of the prior samples by displaying the cumulative sum of the deviation of the sample values from a specified target value. As the CUSUM chart uses information from previous samples, it is more effective than the Shewhart control charts in detecting relatively small shifts in the process mean. Thus, CUSUM procedures are widely used to monitor the quality of products from manufacturing processes. There are two main types of CUSUM procedures, namely the one-sided and two-sided CUSUM procedures. The run length of the CUSUM procedure is the time elapsed before the process is declared to be out of control. The run length distribution and its parameters measure the performance of a CUSUM procedure. The average run length (ARL) is often used as the major criterion for selecting a suitable CUSUM procedure. Much have been written on the evaluation of the run length distribution and the

ARL of the one-sided CUSUM scheme (see for example, Page [1], Ewan and Kemp [2], Goel and Wu [3], Brook and Evans [4], Zacks [5], Reynolds [6], Khan [7], Woodall [8], Lucas and Crosier [9], Fellner [10], Vardeman and Ray [11], Gan [12], Bourke [13], Reynold and Stoumbos [14], Chang and Gan [15], Bowker and Lieberman [16], Hakwins [17]). Theoretically the iterative formulas in Waldmann [18] and Tan and Pooi [19] can be used to find the run length distribution of two-sided CUSUM. But the time required for the computation is usually very long.

An alternative method is to use simulation to estimate the run length distribution of the two-sided CUSUM and then fit the simulated distribution using a smooth curve. In this paper, we propose using the right tail of the translated gamma distribution to fit the simulated distribution. The methods based on moments and least squares can be used to estimate the parameters of the translated gamma distribution.

We demonstrate that when the observations have an exponential or normal distribution, the right tail of the translated gamma distribution gives a good fit to the simulated run length distribution of the two-sided CUSUM.

Using Translated Gamma Distribution to Fit the Simulated Run Length Distribution.

Let x be a random variable which has the gamma distribution with parameters $\alpha > 0$ and $\beta > 0$ (i.e. $x \sim gamma(\alpha, \beta)$). The mean, variance and third moment of x are given respectively by

$$E(x) = \frac{\alpha}{\beta}, \quad Var(x) = \frac{\alpha}{\beta^2}, \quad E\left(x - \frac{\alpha}{\beta}\right)^3 = \frac{2\alpha}{\beta^3}$$

and the probability density function (p.d.f.) of x is

$$f_x(x;\alpha,\beta) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} & 0 \le x < \infty \\ 0 & otherwise \end{cases}$$

where $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy$

Now let $y = x + x_0$ where x_0 is a constant. Then y is said to have a translated gamma distribution. The mean, variance and third moment of y are given by $E(y) = x_0 + \frac{\alpha}{\beta}$, $Var(y) = \frac{\alpha}{\beta^2}$, $E\left(y - (x_0 + \frac{\alpha}{\beta})\right)^3 = \frac{2\alpha}{\beta^3}$ and the p.d.f. of y is

$$f_y(y;\alpha,\beta,x_0) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} (y-x_0)^{\alpha-1} e^{-\beta(y-x_0)} & x_0 \le y < \infty \\ 0 & otherwise \end{cases}$$

From the simulated run length distribution $(\tilde{P}_1, \tilde{P}_2, \tilde{P}_3, \ldots)$, we can find the corresponding mean m_1, variance m_2 and third moment m_3. We then equate these moments to those of the normalized version of the distribution given by the right tail (starting from time $t = 0.5$) of the translated gamma distribution. In this way we can obtain the following three equations

$$\int_{y=0.5}^{\infty} \frac{y\, f_y(y;\alpha,\beta,x_0)}{A} dy = m_1 \quad (1)$$

$$\int_{y=0.5}^{\infty} (y - m_1)^2 \frac{f_y(y;\alpha,\beta,x_0)}{A} dy = m_2 \quad (2)$$

$$\int_{y=0.5}^{\infty} (y - m_1)^3 \frac{f_y(y;\alpha,\beta,x_0)}{A} dy = m_3 \quad (3)$$

where $A = \int_{y=0.5}^{\infty} f_y(y;\alpha,\beta,x_0)\, dy$

Let $\tilde{\alpha}$, $\tilde{\beta}$ and $\tilde{x}_0$ be respectively the values of α, β and x_0 which satisfactory (1), (2) and (3). The simulated run length distribution is then fitted with $\frac{1}{\tilde{A}} f_y(y;\tilde{\alpha},\tilde{\beta},\tilde{x}_0)$

where $\tilde{A} = \int_{y=0.5}^{\infty} f_y(y;\tilde{\alpha},\tilde{\beta},\tilde{x}_0) dy$.

The method given by equations (1), (2) and (3) may be illustrated by the following examples. Consider the case when the quality characteristic $x \sim \exp(\beta^*)$, that is x is exponentially distributed with mean β^*. When the target value of β^* is β_0, the cumulative sums are given by

$$S_t = \sum_{i=0}^{t} y_i, \qquad t = 0, 1, 2, \ldots$$

where $y_i = \frac{(x_i - \beta_0)}{\beta_0}$ and x_i is the i-th observed value of x.

When $\beta^* = \beta_0$, we can express y_i as $y_i = v_i - 1$ where $v_i = x_i/\beta_0 \sim \exp(1)$. But, if the process mean is shifted from the target value of β_0 to $\beta\beta_0$, then $y_i = v_i - 1$ where $v_i \sim \exp(\beta)$ and the probability density function (pdf) of y_i is given by

$$f(y_i) = \begin{cases} \frac{1}{\beta} e^{-\frac{1}{\beta}(y_i+1)} & y_i \geq -1 \\ 0 & y_i < -1 \end{cases}$$

Let $S_t^+ =$ maximum$\{0, S_{t-1}^+ + y_t - k\}$ and $S_t^- =$ minimum$\{0, S_{t-1}^- + y_t + k\}$ where $S_0^+ = S_0^- = 0$. An out of control signal is detected at t by the CUSUM chart if and only if $S_t^+ > h$ or $S_t^- < -h$.

Suppose we use the CUSUM scheme with $k = 0.8386$ and $h = 1.2437$ to detect a shift given by $\beta = 1.0$. The run length distribution is first obtained by means of simulation which uses $N_s = 60000$ values of the observation vector$(y_1, y_2, \ldots)$. The values of the moments based on the simulated distribution are $m_1 = 2.0058E+01$, $m_2 = 7.8194E+02$ and $m_3 = 4.5620E+04$. With these values of m_1, m_2 and m_3, equations (1), (2) and (3) then produce the values $\tilde{\alpha} = 1.1656$, $\tilde{\beta} = 0.0570$ and $\tilde{x}_0 = -8.2329$. The normalized version of the translated gamma distribution and the simulated distribution are shown in Figure 1.

The figure shows that the simulated distribution exhibits random fluctuation and the p.d.f. based on $\tilde{\alpha}, \tilde{\beta}, \tilde{x}_0$ can fit the simulated distribution well.
Figure 2 gives the results when we use the CUSUM scheme with $k = 1.0034$ and $h = 1.1149$ for exponential distribution with $\beta = 1.5$. We observe that the right tail of the translated gamma distribution again fit the simulated distribution well.

Next consider the case when the quality characteristic is a random variable $x \sim N(\mu, \sigma^2)$, that is x is normally distributed with mean μ and variance σ^2. When the target value of μ is μ_0, and the standard deviation is known to be σ, the cumulative sums are given by

$$S_t = \sum_{i=0}^{t} y_i, \qquad t = 0, 1, 2, \ldots$$

where $$y_i = \frac{(\bar{x}_i - \mu_0)}{\sigma/\sqrt{n}}$$

If $\mu = \mu_0$, then $y_i \sim N(0,1)$. But if μ is shifted from the target value μ_0 by an amount $\Delta\sigma/\sqrt{n}$, then $y_i \sim N(\Delta, 1)$ and the probability density function of y_i is given by

$$f(y_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y_i - \Delta)^2}, \qquad -\infty < y_i < \infty$$

Figure 3 gives the results when we use the CUSUM for the normal distribution with $k = 4.4633$, $h = 0.5494$ and $\Delta = 3.75$. The figure again shows that the fit given by the right tail of the translated gamma distribution is satisfactory

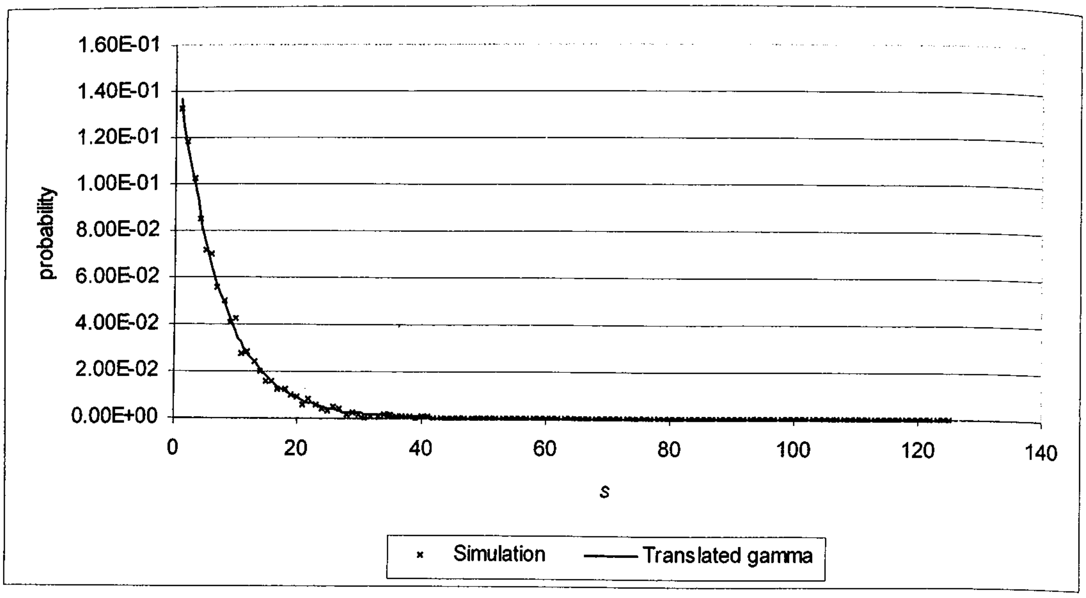

Figure 1. Probability of getting out of control signal at time s for the exponential distribution ($k = 0.8386$, $h = 1.2437$ and $\beta = 1.00$; $\tilde{\alpha} = 1.1656$, $\tilde{\beta} = 0.0570$, $\tilde{x}_0 = -8.2329$)

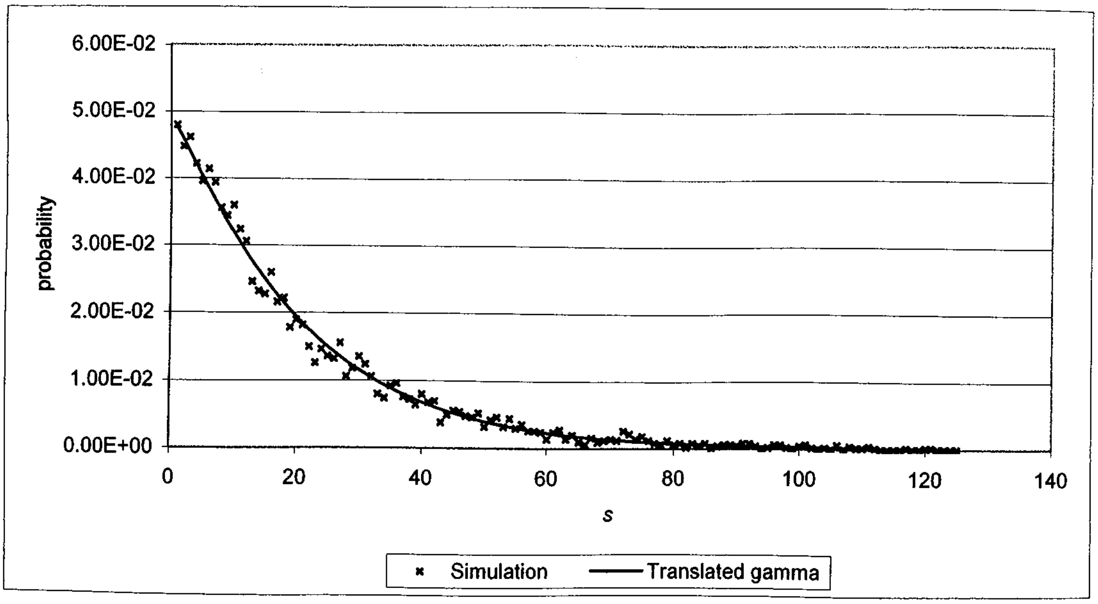

Figure 2. Probability of getting out of control signal at time s for the exponential distribution ($k = 1.0034$, $h = 1.1149$ and $\beta = 1.50$; $\tilde{\alpha} = 0.9869$, $\tilde{\beta} = 0.1404$, $\tilde{x}_0 = 0.6000$

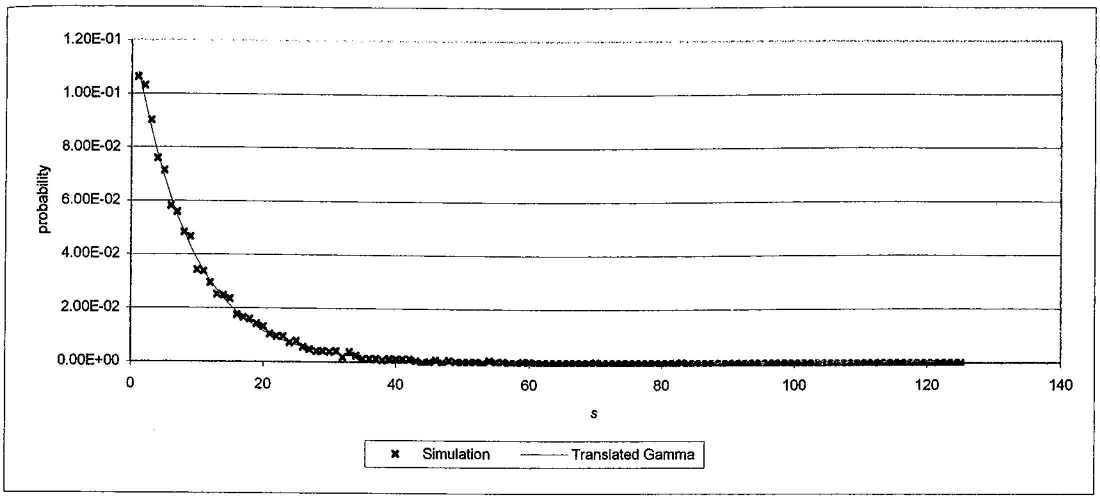

Figure 3. Probability of getting out of control signal at time s for the normal distribution ($k = 4.4633, h = 0.5494$ and Δ 3.75; $\tilde{\alpha} = 1.0141$, $\tilde{\beta} = 0.1204, \tilde{x}_0 = 0.5419$

Fitting the Right Tail of the Simulated Run Length Distribution

In the case when the simulated run length distribution has a mode, the left tail of the simulated run length distribution is usually quite close to the actual results but the right tail exhibits random fluctuation. We may try to improve the results for the right tail by fitting it with the right tail of the translated gamma distribution.

Let $[x',\infty)$ be the right tail of the run length of which the simulated distribution exhibits random fluctuation. Furthermore let

$$P_s^* = \frac{1}{A} f_y(y;\alpha,\beta,x_0) \tag{4}$$

where $$A = \int_{y=x'-0.5}^{\infty} f_y(y;\alpha,\beta,x_0)\,dy$$

and $f_y(y;\alpha,\beta,x_0)$ is the p.d.f. of the translated gamma distribution. We may use the method of least squares to find the values $\alpha = \hat{\alpha}$, $\beta = \hat{\beta}$ and $x_0 = \hat{x}_0$ such that P_s^* given by equation (4) gives a good fit to the normalized version of the right tail of the simulated distribution given by

$$\tilde{P}_s^{(n)} = \frac{\tilde{P}_s}{\sum_{s'=x'}^{\infty} \tilde{P}_{s'}}, \quad x' \le s < \infty$$

This involves the minimization of the sum of squares

$$G = \sum_{s=x'}^{\infty} \left(\tilde{P}_s^{(n)} - P_s^*\right)^2$$

The fit based on least squares for the right tail of the simulated distribution is then given by $\{P_s^*\hat{A}, x' \le s < \infty\}$

where $$\hat{A} = \int_{y=x'-0.5}^{\infty} f_y(y;\hat{\alpha},\hat{\beta},\hat{x}_0)\,dy.$$

Consider the case when we use the normal CUSUM scheme with $k = 0.0650$, $h = 6.4997$ and $\Delta = 0.25$. The run length distribution is first obtained by means of simulation which uses $N_s = 60\,000$ values of the observation vector $(y_1, y_2, \ldots)$. We choose $x' = 16$. The least squares estimates are found to be $\hat{\alpha} = 1.2840$, $\hat{\beta} = 0.0679$, and $\hat{x}_0 = 7.4256$. The simulated distribution ($\tilde{P}_s$) and the fitted distribution ($P_s^*\hat{A}$) are shown in Figure 4. Figure 4 shows that the fit given by $P_s^*\hat{A}$ is satisfactory.

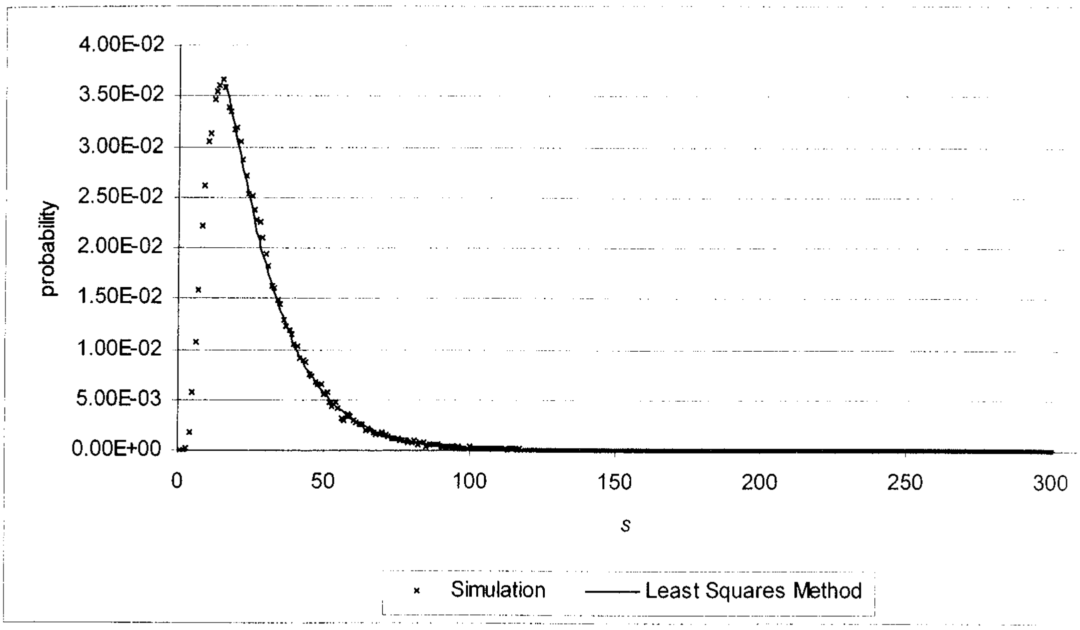

Figure 4. Probability of getting out of control signal at time s for the normal distribution ($k = 0.0650$, $h = 6.4997$ and $\Delta = 0.25$; $\hat{\alpha} = 1.2840$, $\hat{\beta} = 0.0679$, $\hat{x}_0 = 7.4256$

CONCLUSION

Although iterative formulas can yield accurate results for the run length distribution of the two-sided CUSUM, the computation time is very long. However it takes only a few minutes to find the same distribution by means of simulation done on the Pentium IV personal computer. Translated gamma distribution can be used to improve the simulated distribution which exhibits random fluctuation. The improvement is significant and the computation involved is easy.

REFERENCES

1. Page, E.S. (1954). Continuous inspection schemes. *Biometrika* **41**, 100-114.
2. Ewan, W.D. and Kemp, K.W. (1960). Sampling inspection of continuous processes with no autocorrelation between successive results. *Biometrika* **47**, 363-380.
3. 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.
4. Brooks, D. and Evans, D.A. (1972). An approach to the probability distribution of CUSUM runs length. *Biometrika* **59**, 539-549.
5. Zacks, S. (1981). The probability distribution and the expected value of a stopping variable associated with one-sided CUSUM procedures for non-negative integer valued random variables. *Communications in Statistics* **A10**, 2245-2258.
6. Reynolds, M.R., Jr. (1975). Approximations to the average run length in cumulative control charts. *Technometrics* **17**, 65-71.
7. Khan, R.A. (1978). Wald's approximations to the average run length in CUSUM procedures. *Journal of Statistical Planning and Inference* **2**, 63-77.
8. Woodall, W.H. (1983). The distribution of the run length of one-sided CUSUM procedures for continuous random variables. *Technometrics* **24**, 295-301.
9. 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.
10. Fellner, W.H. (1991). Algorithm AS 258, average run length for cumulative Sum schemes. *Applied Statistics* **39**, 402-412.
11. Vardeman, S. and Ray, D. (1985). Average run lengths for CUSUM schemes when observations are exponentially distributed. *Technometrics* **27**, 145-150.
12. Gan, F.F. (1992). Exact run length distributions for one-sided exponential

CUSUM schemes. *Statistica Sinica* **2**, 297-312.

13. Bourke, P.D. (1991). Detecting a shift in the fraction of nonconforming items using run-length control charts with 100% inspection. *J. Qual. Tech.* **23**, 225-238.
14. Reynolds, M.R., Jr. and Stoumbos, Z.G. (1999). A CUSUM chart for monitoring a proportion when inspecting continuously. *J. Qual. Tech.* **31**, 87-108.
15. Chang, T.C. and Gan, F.F. (2001). Cumulative sum charts for high yield processes. *Statistica Sinica* **11**, 791-805.
16. Bowker and Lieberman (1972). Engineering Statistics, 2nd ed., Prentice-Hall, Englewood Cliffs, N.J.
17. Hawkins, D.M. (1992). A fast accurate approximation for average run lengths of CUSUM control charts. *Journal of Quality Technology* **24**, 37-43.
18. Waldman, K.H. (1986). Bounds for the distribution of the run length of one-sided and two-sided CUSUM quality control schemes. *Technometrics* **28**, 61-67.
19. Tan, Y.F. and Pooi, A.H. (2002). Run length distribution of two-sided CUSUM procedures for continuous random variables (Manuscript which has been sent out for consideration of publication).