BULLETIN OF THE
INSTITUTE OF MATHEMATICS
ACADEMIA SINICA
Volume 33, Number 2, June 2005

RUN LENGTH DISTRIBUTION OF TWO-SIDED CUSUM PROCEDURES FOR CONTINUOUS RANDOM VARIABLES

BY

Y. F. TAN (陳怡霏) AND A. H. POOI(貝仁錫)

Abstract. Cumulative sum (CUSUM) control charts are very effective in detecting special causes. In judging the performance of a CUSUM procedure, it is important to know its run length distribution. Presently iterative formulas are derived for finding the run length distribution of two-sided CUSUM. The application of the iterative formulas is illustrated in the normal two-sided CUSUM.

1. Introduction. Cumulative sum (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 one-sided CUSUM scheme is further classified into two types, namely the lower-sided CUSUM and upper-sided CUSUM. The lower-sided CUSUM is intended to detect an upward shift in the process mean while the upper-sided CUSUM is intended to detect a downward shift in the process mean. The two-sided CUSUM procedure is intended to detect a shift which could be downward or upward in the process mean.

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

Received by the editors December 10, 2002 and in revised form August 4, 2004.
Key words and phrases: Average run length (ARL), cumulative sum, iterative formulas.

run length (ARL) is often used as the major criterion for selecting a suitable CUSUM procedure. However the used of ARL has been widely criticized and the use of percentage points has been recommended (see, for example Barnard [1] and Bissell [2]). Much have been written on the determination of run length distribution and its moments for the one-sided CUSUM procedure (for example Page [3], Ewan and Kemp [4], Goel and Wu [5], Brook and Evans [6], Zacks [7], Woodall [8], Fellner [9], Hawkins [10], Vardeman and Ray [11], Gan [12], Luceño and Puig-Pey [13] and Luceño and Puig-Pey [14]).

In Kemp [15], the ARL of the two-sided CUSUM procedures is given in terms of the ARL of the lower-sided and upper-sided CUSUM procedures.

In Woodall [16], the Markov chain representation of the one-sided CUSUM procedure proposed by Brook and Evans [6] is extended to the two-sided CUSUM procedure. This extension is similar to that described by Lucas and Crosier [17] but makes a number of improvements.

A recent review of the literature of CUSUM charts may be found in Hawkins and Olwell [18].

In Section Two of this paper, iterative formulas are developed for finding the run length distribution of two-sided CUSUM for continuous random variables.

Section Three is devoted to the comparison of the iterative formulas for computing run length distributions of one-sided and two-sided CUSUM.

In Section Four, we use the iterative formulas for two-sided CUSUM to compute the probability of getting an out of control signal at time t $(1 \leq t \leq 7)$.

2. Run length distribution of two-sided CUSUM.

2.1. V-mask and definition of run length distribution. Suppose the quality characteristic is a random variable x with mean μ and variance

σ^2. Let $x_{i1}, x_{i2}, \ldots, x_{in}$ be the i^{th} sample of size n of quality characteristic. Then the sample mean $\bar{x}_i = 1/n \sum_{j=1}^{n} x_{ij}$ of the i^{th} sample will have mean μ and variance σ^2/n. Let μ_0 be the target value of the process mean μ. The cumulative sums may be defined as

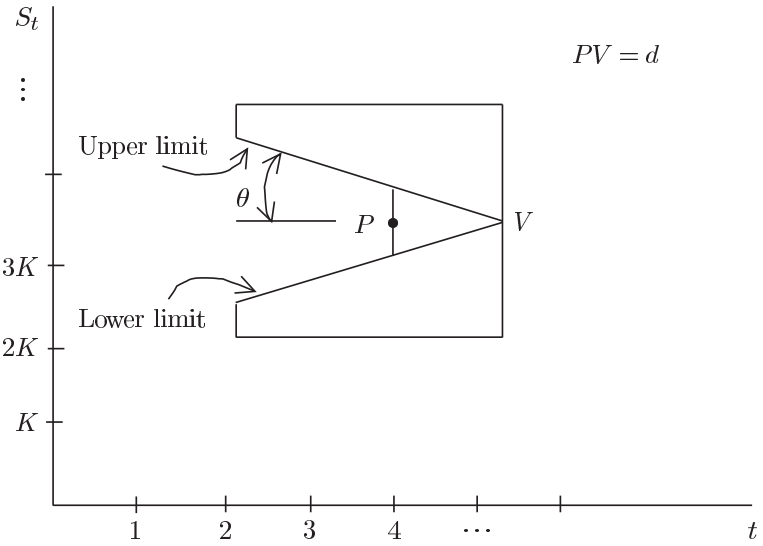
$$S_t = \sum_{i=0}^{t} y_i, \quad t = 0, 1, 2, \ldots$$

where $y_i = c(\bar{x}_i - \mu_0)$, c is a positive constant and $S_0 = 0$.

If the process mean μ is in control at the target value μ_0, then the values of S_t will fluctuate on a horizontal straight line when t increases.

If the process mean μ is not equal to the target value μ_0 (i.e. the process mean μ is out of control), then S_t will tend to increase when $\mu > \mu_0$ and it will tend to decrease when $\mu < \mu_0$.

To determine whether $\mu = \mu_0$, $\mu > \mu_0$ or $\mu < \mu_0$, we may use the decision rule based on the following V-mask.



K-scalar factor (usually we choose $K = 2\sigma/\sqrt{n}$); V-vertice; P-location point; t-time; d-lead distance; θ-half angle; (d and θ are the parameters of

the V-mask).

We place the mask such that the location point P coincides with the most recently plotted CUSUM point with PV in a horizontal position. The decision rule based on V-mask is given as follows.

If all the points plotted previously are within the two lines labeled respectively as the lower limit and upper limit of the V-mask then we conclude that the process is in control.

If there is a point below the lower limit then we conclude that the process is out of control and a positive shift of the process mean is detected.

If there is a point above the upper limit then we conclude that the process is out of control and a negative shift of the process mean is detected.

Let T be the value of t when an out of control signal is detected for the first time. T is known as the run length of the CUSUM. The expected value of T is called the average run length (ARL) of the CUSUM.

Let $k = K \tan\theta$ and $h = dK \tan\theta$. Furthermore let $S_t^+ = \text{maximum}\{0, S_{t-1}^+ + \bar{x}_t - \mu_0 - k\}$ and $S_t^- = \text{minimum}\{0, S_{t-1}^- + \bar{x}_t - \mu_0 + k\}$ where $S_0^+ = S_0^- = 0$. It can be shown that an out of control signal is detected at t by the V-mask with parameter d and θ if and only if $S_t^+ > h$ or $S_t^- < h$. The procedure based on S_t^+ and S_t^- is known as the tabular two-sided CUSUM procedure.

If (j, S_j), $0 \leq j \leq m-1$ are within the lower and upper limits of the V-mask when P coincides with (m, S_m) for $m = 1, 2, \ldots, t$, then the process is said to be still in control at time t.

Let P_t' be the probability that no out of control signal has occurred before or at time t. Let $P_t = P(T = t)$. Then $\{P_t : t = 1, 2, \ldots\}$ is the run length distribution. The relationship between P_t' and the P_t is as follows

$$P_t = P_{t-1}' - P_t', \quad t = 1, 2, 3, \ldots \text{where } P_0' = 1.$$

Therefore we can find P_t once the P'_t are known.

2.2. Derivation of the run length distribution. To find P'_t, we may first draw a V-mask with parameters d and θ on a transparency and then move the point P of the V-mask up and down along the vertical line which passes through the point $(t, 0)$ in the t-axis to determine the range of values of y_t such that no out of control signal has occurred. In this way we can express P'_t in terms of a multiple integral.

For example to find P'_t, we may move the transparency such that the point P of the V-mask moves along the vertical line at time $t = 1$.

We observe that as we move the point P of V-mask down the vertical line at $t = 1$, the upper limit of the V-mask touches the point $(0, 0)$ when $S_1 = -k-h$ (i.e. $y_1 = -k-h$) (See Figure 2.2.1). And as we move the point P of the V-mask up the vertical line at $t = 1$, the lower limit of the V-mask touches the point $(0, 0)$ when $S_1 = k+h$ (i.e. $y_1 = k+h$) (see Figure 2.2.2). Therefore the process is still in control as long as $y_1 \in (-k-h, k+h)$. This means

$$P'_1 = \int_{-k-h}^{k+h} f(y_1)dy_1 \tag{2.1}$$

where f is the probability density function (pdf) of y_i and $i = 1, 2, \ldots$.

We may express P'_1 in terms of the function $F_s(m_1, m_2, h_1, h_2, h_3)$ which is defined as follows. First let m be a positive integer. Next let E_m be the event that $S_m = h_3$, no out of control signal has yet occurred at time $t = m$, the upper limit of the V-mask will touch the point $S_{m+1-m_1} = h_1$ first as we move the point P down the vertical line at time $t = m + 1$ and the lower limit of the V-mask will touch the point $S_{m+1-m_2} = h_2$ first as we move the point P up the vertical line at time $t = m + 1$.

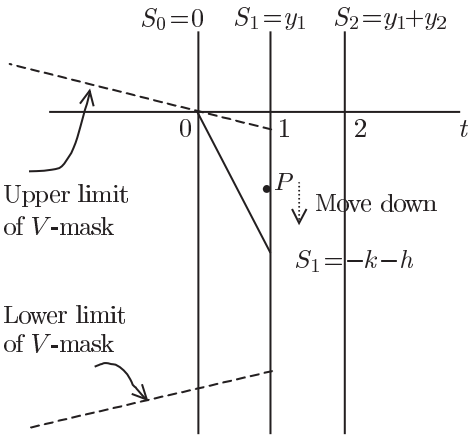


Figure 2.2.1. Lower limit of the integration in (2.1).

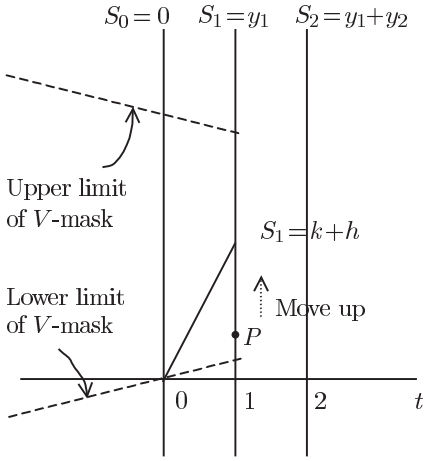


Figure 2.2.2. Upper limit of the integration in (2.1).

(Note: In obtaininng the upper limit indicated in the figure, a V-mask with $\theta = 14^\circ$ and $d = 6$cm has been used. The remaining figures in Section 2 are also drawn with reference to the V-mask with the same parameters.)

Then we define $F_s^{(m)}(m_1, m_2, h_1, h_2, h_3)$ as

$$F_s^{(m)}(m_1, m_2, h_1, h_2, h_3) = P(\text{The process is still in control at time } m+s \mid E_m), \quad m \geq \max(m_1, m_2).$$

The value of $\underset{\sim}{v} = (m_1, m_2, h_1, h_2, h_3)$ is determined by the observations $y_1, y_2, \ldots, y_m$. Suppose m is changed to a larger value m^* say, and the additional observations $y_{m+1}, y_{m+2}, \ldots, y_{m^*}$ together with the previous observations $y_1, y_2, \ldots, y_m$ yield a value $\underset{\sim}{v}^*$ say of $\underset{\sim}{v}$. The value of $\underset{\sim}{v}^*$ may or may not be equal to that of $\underset{\sim}{v}$. In the hypothetical case where $\underset{\sim}{v}^* = \underset{\sim}{v}$,

$$F_s^{(m)}(m_1, m_2, h_1, h_2, h_3) = F_s^{(m^*)}(m_1, m_2, h_1, h_2, h_3)$$

because $y_1, y_2, \ldots$ are all identically distributed. Thus for simplicity, we drop

the letter m appearing in $F_s^{(m)}(m_1, m_2, h_1, h_2, h_3)$ and define

$$F_s(m_1, m_2, h_1, h_2, h_3) = P(\text{The process is still in control at time } m+s \mid E_m), \quad m \geq \max(m_1, m_2).$$

We notice that for

$$-m_1k - h + h_1 - h_3 < m_2k + h + h_2 - h_3 \quad \text{(i.e. } h_1 - h_2 < (m_1k + h) + (m_2k + h)\text{)} \tag{2.2}$$

F_s can be written as

$$F_s(m_1, m_2, h_1, h_2, h_3) = \int_{-m_1k-h+h_1-h_3}^{m_2k+h+h_2-h_3} f(y_{m+1}) \times \left[\int_{\substack{y_{m+2} \\ (y_{m+2},\ldots,y_{m+s})\in R^*}} \cdots \int_{y_{m+s}} f(y_{m+2}) \cdots f(y_{m+s}) dy_{m+s} \cdots dy_{m+2} \right] dy_{m+1} \tag{2.3}$$

where

$R^* = \{(y_{m+2}, \ldots, y_{m+s})$: The points $S_{m+1-m_1}, S_{m+1-m_2}, S_m, S_{m+1}, \ldots, S_{m+v-1}$ are still within the lower and upper limits of the V-mask when P coincides with S_{m+v} for $v = 2, 3, \ldots, s$, in the case when E_m has occurred and the value of y_{m+1} is given$\}$

and the limits for y_{m+1} are obtained as follows.

The lower limit $y_{m+1} = -m_1k - h + h_1 - h_3$ of the range of y_{m+1} is obtained by moving the point P down the vertical line at the time $t = m+1$ until the upper limit of the V-mask touches the point $S_{m+1-m_1} = h_1$ (see Figure 2.2.3). The point $S_{m+1-m_1} = h_1$ shall be called the lower reference point for the range of integration with respect to y_{m+1}.

The upper limit $y_{m+1} = m_2k + h + h_2 - h_3$ of the range of y_{m+1} is obtained by moving the point P up the vertical line at the time $t = m + 1$

until the lower limit of the V-mask touches the point $S_{m+1-m_2} = h_2$ (see Figure 2.2.4). The point $S_{m+1-m_2} = h_2$ shall be called the upper reference point for the range of integration with respect to y_{m+1}.

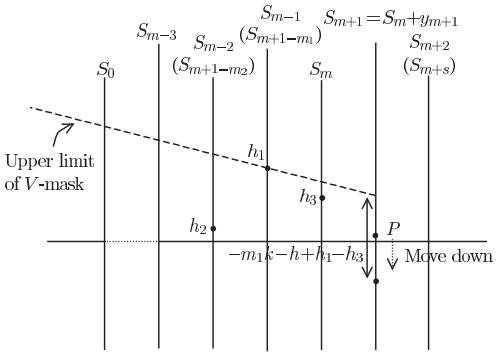


Figure 2.2.3. Lower limit of the integration with respect to y_{m+1} in (2.3), (eg. $s = 2$, $m_1 = 2$, $m_2 = 3$).

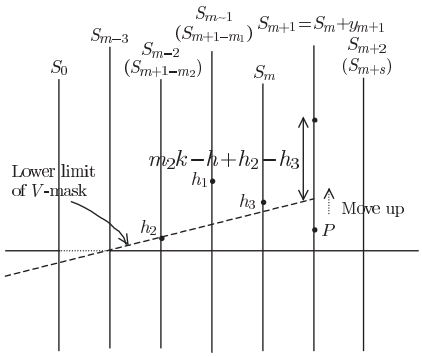


Figure 2.2.4. Upper limit of the integration with respect to y_{m+1} in (2.3), (eg. $s = 2$, $m_1 = 2$, $m_2 = 3$).

Therefore P_1' (see (2.1)) may be expressed as

$$P_1' = F_1(1, 1, 0, 0, 0)$$

We note that for $s \geq 1$,

$$F_s(m_1, m_2, h_1 + c, h_2 + c, h_3 + c) = F_s(m_1, m_2, h_1, h_2, h_3) \tag{2.4}$$

because in the multiple integral representing $F_s(m_1, m_2, h_1+c, h_2+c, h_3+c)$ the ranges of $y_{m+1}, y_{m+2}, \ldots, y_{m+s}$ do not depend on c.

We next try to express P_s' in terms of the function F_{s-1}. By moving the point P of the V-mask along the vertical line at time $t = 2$, we see that for $S_1 \in (k, k+h)$ and $S_2 \in (S_1 - k - h, 2k + h)$ (i.e. $y_2 \in (-k-h, 2k+h-y_1)$) no out of control signal has yet occurred at time $t = 2$ (see Figures 2.2.5 and 2.2.6). Furthermore the lower reference point for the range of y_2 is $S_1 = y_1$ and the upper reference point for the range of y_2 is $S_0 = 0$.

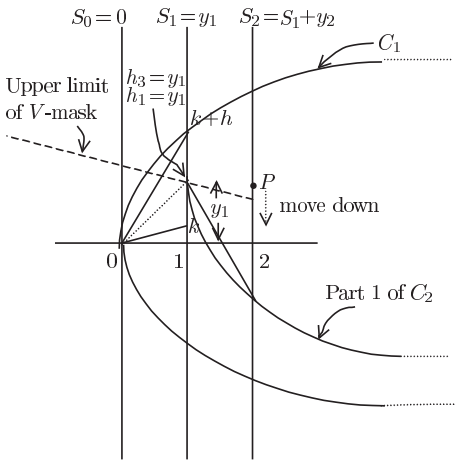


Figure 2.2.5. Lower limit for y_2 when $S_1 \in (k, k+h)$.

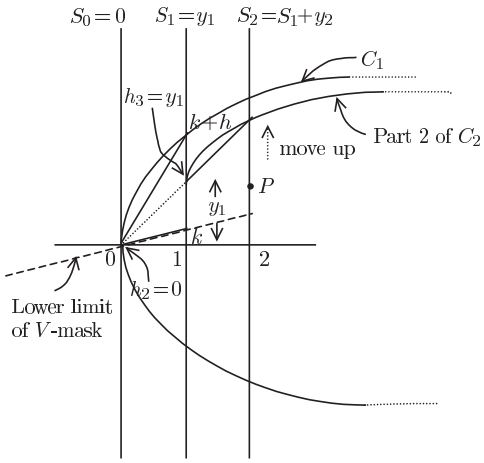


Figure 2.2.6. Upper limit for y_2 when $S_1 \in (k, k+h)$.

(Note: The region bounded within the curve C_1 represents the paths which correspond to P'_s. The region bounded within the curve C_2 represents the paths which correspond to $F_{s-1}(1, 2, y_1, 0, y_1)$).

Next for $S_1 \in (-k, k)$ and $S_2 \in (S_1 - k - h, S_1 + k + h)$ (i.e. $y_2 \in (-k-h, k+h)$) no out of control signal has yet occurred at time $t = 2$ (see Figures 2.2.7 and 2.2.8). Furthermore the lower reference point for the range of y_2 is $S_1 = y_1$ and the upper reference point for the range of y_2 is also $S_1 = y_1$.

Finally for $S_1 \in (-k-h, -k)$ and $S_2 \in (-2k-h, y_1+k+h)$ (i.e. $y_2 \in (-2k-h-y_1, k+h)$) no out of control signal has yet occurred at time $t = 2$. Furthermore the lower reference point for the range of y_2 is $S_0 = 0$ and the upper reference point for the range of y_2 is $S_1 = y_1$. Therefore

$$
\begin{aligned}
P'_s \;=\; & \int_k^{k+h} f(y_1) F_{s-1}(1, 2, y_1, 0, y_1) dy_1 + \int_{-k}^{k} f(y_1) F_{s-1}(1, 1, 0, 0, 0) dy_1 \\
& + \int_{-k-h}^{-k} f(y_1) F_{s-1}(2, 1, 0, y_1, y_1) dy_1 \qquad s = 2, 3, \ldots
\end{aligned} \tag{2.5}
$$

The right side of (2.5) shows that we can find for P'_s for $s \geq 2$ if we know $F_1, F_2, \ldots$.

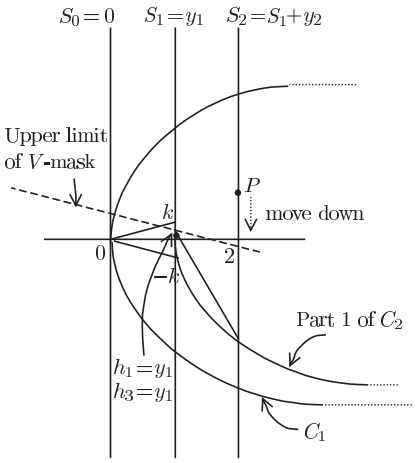


Figure 2.2.7. Lower limit for y_2 when $S_1 \in (-k, k)$.

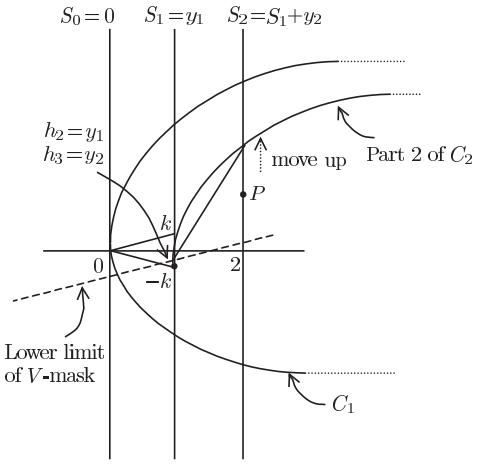


Figure 2.2.8. Upper limit for y_2 when $S_1 \in (-k, k)$.

(Note: The region bounded within the curve C_1 represents the paths which correspond to P'_s. The region bounded within the curve C_2 represents the paths which correspond to $F_{s-1}(1, 1, y_1, y_1, y_1) = F_{s-1}(1, 1, 0, 0, 0)$).

2.3. Iterative formula for computing F_s. To find $F_1, F_2, \ldots$, we may first derive a relation between F_s and F_{s-1} as follows.

First we move the point P of V-mask down the vertical line at time $t = m + 2$. We notice that among the points $S_{m+1-m_1}, S_{m+1-m_2}, S_m$ and $S_{m+1} \in (h_1 - m_1 k, h_2 + m_2 k + h)$, the point S_{m+1} will touch the upper limit of the V-mask first (see Figures 2.3.1 and 2.3.2). But when $S_{m+1} < h_1 - m_1 k$ the point S_{m+1-m_1} instead will touch the upper limit of the V-mask first, and furthermore the upper limit of the V-mask then will cut the vertical line at time $t = m + 1$ at $S_{m+1} = S^{(U)}_{m+1} = h_1 - m_1 k$ and the vertical line at time $t = m$ at $S_m = S^{(U)}_m = h_1 - (m_1 - 1)k$ (see Figures 2.3.1 and 2.3.2).

Next we move the point P of V-mask up the vertical line at time $t = m+2$. We notice that among the points $S_{m+1-m_1}, S_{m+1-m_2}, S_m$ and $S_{m+1} \in (h_1 - m_1 k - h, h_2 + m_2 k)$ the point S_{m+1} will touch the lower limit of the V-mask first (see Figures 2.3.1 and 2.3.2). But when $S_{m+1} > h_2 + m_2 k$ the point S_{m+1-m_2} instead will touch the lower limit of the V-mask first, and

furthermore the lower limit of the V-mask then will cut the vertical line at time $t = m + 1$ at $S_{m+1} = S_{m+1}^{(L)} = h_2 + m_2k$ and the vertical line at time $t = m$ at $S_m = S_m^{(L)} = h_2 + (m_2 - 1)k$ (see Figures 2.3.1 and 2.3.2).

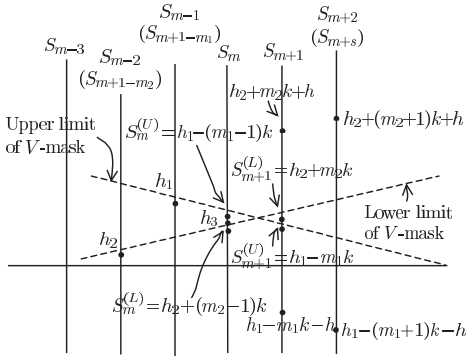


Figure 2.3.1. Lower limit and upper limit of S_{m+2} when $S_{m+1} \in (h_1 - m_1k - h, h_2 + m_2k + h)$ (eg. $s = 2$, $m_1 = 2$, $m_2 = 3$). This is the case when $S_{m+1}^{(L)} \geq S_{m+1}^{(U)}$.

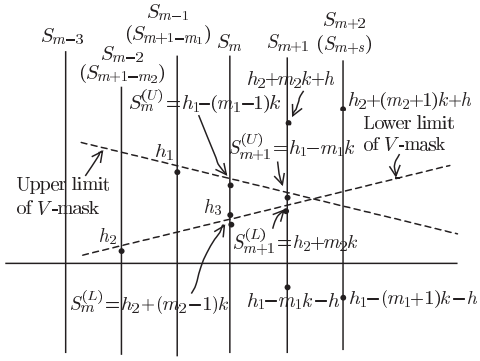


Figure 2.3.2. Lower limit and upper limit of S_{m+2} when $S_{m+1} \in (h_1 - m_1k - h, h_2 + m_2k + h)$ and $S_{m+1} < h_1 - m_1k$ (eg. $s = 2$, $m_1 = 2$, $m_2 = 3$). This is the case when $S_{m+1}^{(L)} < S_{m+1}^{(U)}$.

There are now two cases that we may consider. The first case is one in which $S_{m+1}^{(L)} \geq S_{m+1}^{(U)}$, i.e. $h_2 + m_2k \geq h_1 - m_1k$ (see Figure 2.3.1). The second case is one in which $S_{m+1}^{(L)} < S_{m+1}^{(U)}$, i.e. $h_2 + m_2k < h_1 - m_1k$ (see Figure 2.3.2).

Consider case 1 when

$$S_{m+1}^{(L)} \geq S_{m+1}^{(U)} \quad (\text{i.e. } h_2 + m_2k \geq h_1 - m_1k). \tag{2.6}$$

We notice that for $S_{m+1} \in (h_1 - m_1k - h, h_2 + m_2k + h)$, no out of control signal has yet occurred at time $t = m + 1$ (see Figure 2.3.1). This means that

$$S_m^{(L)} < S_m^{(U)}, \quad (\text{i.e. } h_2 + (m_2 - 1)k < h_1 - (m_1 - 1)k). \tag{2.7}$$

and

$$h_2 + (m_2 - 1)k < h_3 < h_1 - (m_1 - 1)k. \tag{2.8}$$

Combining (2.6), (2.7) and (2.8) we see that for case 1, h_1, h_2, h_3 are subjected to the following restrictions

$$(m_1-1)k+(m_2-1)k<h_1-h_2\leq(m_1+m_2)k$$

and

$$h_2+(m_2-1)k<h_3<h_1-(m_1-1)k \tag{2.9}$$

Furthermore for case 1 we notice that for $S_{m+1}\in(h_2+m_2k,h_2+m_2k+h)$ the lower reference point for the range of integration with respect to y_{m+2} is S_{m+1} (see Figure 2.3.3) and the upper reference point for the range of integration with respect to y_{m+2} is S_{m+1-m_2} (see Figure 2.3.4). For $S_{m+1}\in(h_1-m_1k,h_2+m_2k)$ the lower reference point for the range of integration with respect to y_{m+2} is S_{m+1} (see Figure 2.3.5) and the upper reference point for the range of integration with respect to y_{m+2} is S_{m+1} (see Figure 2.3.6) also. Finally for $S_{m+1}\in(h_1-m_1k-h,h_1-m_1k)$, the lower reference point for the range of integration with respect to y_{m+2} is S_{m+1-m_1} and the upper reference point for the range of integration with respect to y_{m+2} is S_{m+1}. Therefore

$$\begin{aligned}
&F_s(m_1,m_2,h_1,h_2,h_3)\\
&=\int_{h_2+m_2k-h_3}^{h_2+m_2k+h-h_3}f(y_{m+1})F_{s-1}(1,m_2+1,h_3+y_{m+1},h_2,h_3+y_{m+1})dy_{m+1}\\
&+\int_{h_1-m_1k-h_3}^{h_2+m_2k-h_3}f(y_{m+1})F_{s-1}(1,1,0,0,0)dy_{m+1}\\
&+\int_{h_1-m_1k-h-h_3}^{h_1+m_1k-h_3}f(y_{m+1})F_{s-1}(m_1+1,1,h_1,h_3+y_{m+1},h_3+y_{m+1})dy_{m+1}
\end{aligned} \tag{2.10}$$

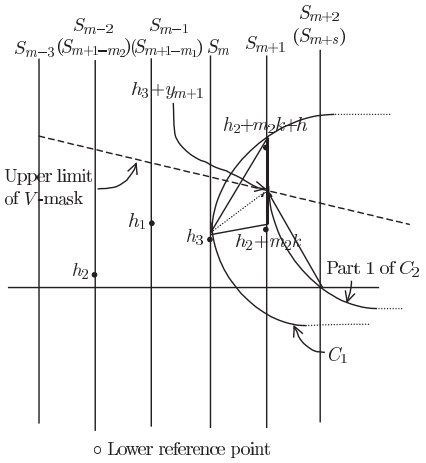


Figure 2.3.3. Lower reference point for the range of integration with respect to y_{m+2} for $S_{m+1} \in (h_2 + m_2k, h_2 + m_2k + h)$.

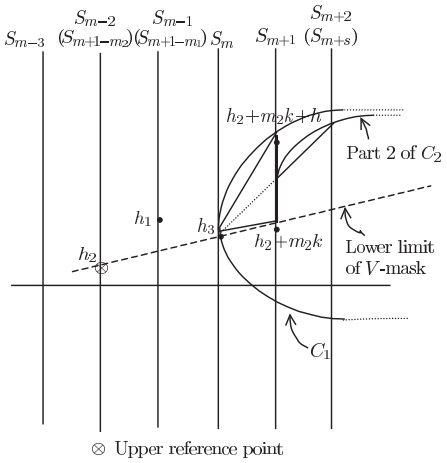


Figure 2.3.4. Upper reference point for the range of integration with respect to y_{m+2} for $S_{m+1} \in (h_2 + m_2k, h_2 + m_2k + h)$.

(Note: The region bounded within the curve C_1 represents the paths which correspond to $F_s(m_1, m_2, h_1, h_2, h_3)$. The region bounded within the curve C_2 represents the paths which correspond to $F_{s-1}(1, m_2 + 1, h_3 + y_{m+1}, h_2, h_3 + y_{m+1})$.)

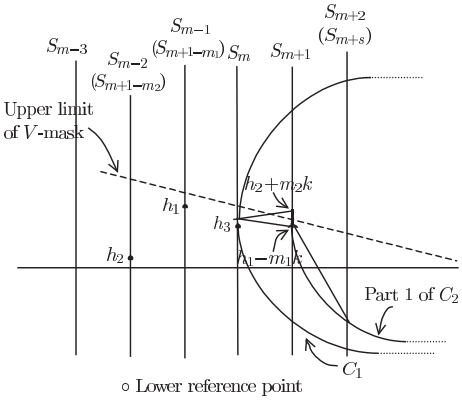


Figure 2.3.5. Lower reference point for the range of integration with respect to y_{m+2} for $S_{m+1} \in (h_1 - m_1k, h_2 + m_2k)$.

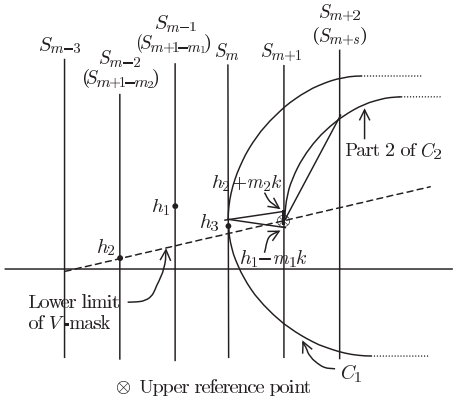


Figure 2.3.6. Upper reference point for the range of integration with respect to y_{m+2} for $S_{m+1} \in (h_1 - m_1k, h_2 + m_2k)$.

(Note: The region bounded within the curve C_1 represents the paths which correspond to $F_s(m_1, m_2, h_1, h_2, h_3)$. The region bounded within the curve C_2 represents the paths which correspond to $F_{s-1}(1, 1, 0, 0, 0)$.)

Case 2 can be subdivided further into 2 cases (case 2.1 and case 2.2). Case 2.1 corresponds to the case when $S_{m+1}^{(U)} - S_{m+1}^{(L)} < h$. Case 2.2 corresponds to the case when $S_{m+1}^{(U)} - S_{m+1}^{(L)} \geq h$.

Consider case 2.1. We have

$$S_{m+1}^{(L)} < S_{m+1}^{(U)} \quad \text{(i.e. } h_2 + m_2 k < h_1 - m_1 k\text{)} \tag{2.11}$$

and

$$S_{m+1}^{(U)} - S_{m+1}^{(L)} < h \quad \text{(i.e. } h_1 - h_2 < (m_1 + m_2)k + h\text{)} \tag{2.12}$$

We notice that for $S_{m+1} \in (h_1 - m_1 k - h, h_2 + m_2 k + h)$, no out of control signal has yet occurred at time $t = m + 1$ (see Figure 2.3.2). This means that

$$S_m^{(L)} < S_m^{(U)} \quad \text{(i.e. } h_2 + (m_2 - 1)k < h_1 - (m_1 - 1)k\text{)} \tag{2.13}$$

and

$$h_2 + (m_2 - 1)k < h_3 < h_1 - (m_1 - 1)k \tag{2.14}$$

Combining (2.11), (2.12), (2.13) and (2.14) we see that for case 2.1, h_1, h_2, h_3 are subjected to the following restrictions

$$(m_1 + m_2)k < h_1 - h_2 < (m_1 + m_2)k + h$$

and

$$h_2 + (m_2 - 1)k < h_3 < h_1 - (m_1 - 1)k \tag{2.15}$$

Furthermore for case 2.1 we notice that for $S_{m+1} \in (h_1 - m_1 k, h_2 + m_2 k + h)$ the lower reference point for the range of integration with respect to y_{m+2} is S_{m+1} (see Figure 2.3.7) and the upper reference point for the range of integration with respect to y_{m+2} is S_{m+1-m_2} (see Figure 2.3.8). For $S_{m+1} \in (h_2 + m_2 k, h_1 - m_1 k)$, the lower reference point for the range of integration with respect to y_{m+2} is S_{m+1-m_1} (see Figure 2.3.9), and the

upper reference point for the range of integration with respect to y_{m+2} is S_{m+1-m_2} (see Figure 2.3.10). Finally for $S_{m+1} \in (h_1 - m_1k - h, h_2 + m_2k)$, the lower reference point for the range of integration with respect to y_{m+2} is S_{m+1-m_1}, and the upper reference point for the range of integration with respect to y_{m+2} is S_{m+1}. Therefore

$$
\begin{aligned}
& F_s(m_1, m_2, h_1, h_2, h_3) \\
= & \int_{h_1 - m_1k - h_3}^{h_2 + m_2k + h - h_3} f(y_{m+1}) F_{s-1}(1, m_2+1, h_3+y_{m+1}, h_2, h_3+y_{m+1}) dy_{m+1} \\
& + \int_{h_2 + m_2k - h_3}^{h_1 - m_1k - h_3} f(y_{m+1}) F_{s-1}(m_1+1, m_2+1, h_1, h_2, h_3+y_{m+1}) dy_{m+1} \\
& + \int_{h_1 - m_1k - h - h_3}^{h_2 + m_2k - h_3} f(y_{m+1}) F_{s-1}(m_1+1, 1, h_1, h_3+y_{m+1}, h_3+y_{m+1}) dy_{m+1}
\end{aligned}
\tag{2.16}
$$

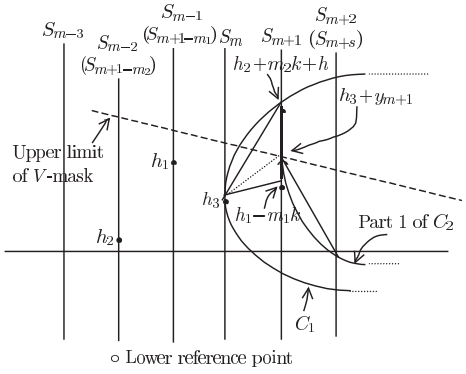


Figure 2.3.7. Lower reference point for the range of integration with respect to y_{m+2} for $S_{m+1} \in (h_1 - m_1k, h_2 + m_2k + h)$ in case 2.1.

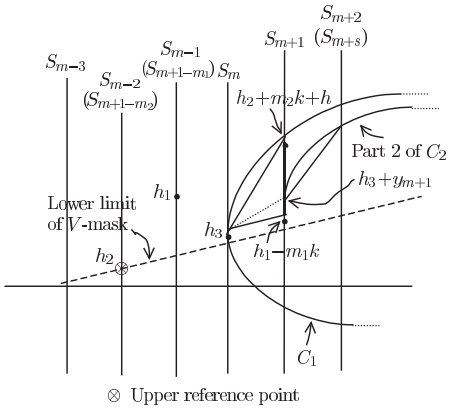


Figure 2.3.8. Upper reference point for the range of integration with respect to y_{m+2} for $S_{m+1} \in (h_1 - m_1k, h_2 + m_2k + h)$ in case 2.1.

(Note: The region bounded within the curve C_1 represents the paths which correspond to $F_s(m_1, m_2, h_1, h_2, h_3)$. The region bounded within the curve C_2 represents the paths which correspond to $F_{s-1}(1, m_2 + 1, h_3 + y_{m+1}, h_2, h_3 + y_{m+1})$.)

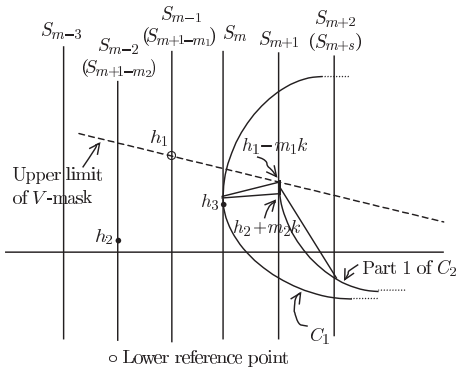


Figure 2.3.9. Lower reference point for the range of integration with respect to y_{m+2} for $S_{m+1} \in (h_2 + m_2k, h_1 - m_1k)$ in case 2.1.

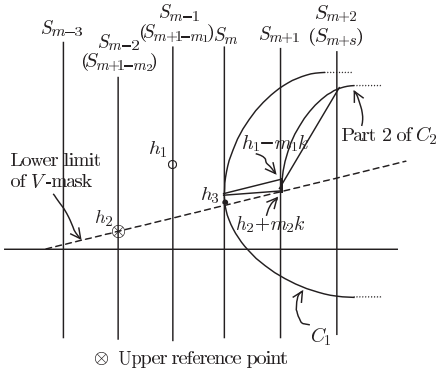


Figure 2.3.10. Upper reference point for the range of integration with respect to y_{m+2} for $S_{m+1} \in (h_2 + m_2k, h_1 - m_1k)$ in case 2.1.

(Note: The region bounded within the curve C_1 represents the paths which correspond to $F_s(m_1, m_2, h_1, h_2, h_3)$. The region bounded within the curve C_2 represents the paths which correspond to $F_{s-1}(m_1 + 1, m_2 + 1, h_1, h_2, h_3 + y_{m+1})$.)

Consider case 2.2. We have

$$S_{m+1}^{(L)} < S_{m+1}^{(U)} \quad \text{(i.e. } h_2 + m_2k < h_1 - m_1k) \tag{2.17}$$

and

$$S_{m+1}^{(U)} - S_{m+1}^{(L)} \geq h \quad \text{(i.e. } h_1 - h_2 \geq (m_1 + m_2)k + h) \tag{2.18}$$

We notice that for $S_{m+1} \in (h_1 - m_1k - h, h_2 + m_2k + h)$, no out of control signal has yet occurred at time $t = m + 1$ (see Figure 2.3.2). This means that

$$S_m^{(L)} - S_m^{(U)} \quad \text{(i.e. } h_2 + (m_2 - 1)k < h_1 - (m_1 - 1)k) \tag{2.19}$$

and

$$h_2 + (m_2 - 1)k < h_3 < h_1 - (m_1 - 1)k \tag{2.20}$$

Next we notice that

$$h_2 + m_2k + h > h_1 - m_1k - h \quad \text{(see (2.2))} \tag{2.21}$$

Combining (2.17), (2.18), (2.19), (2.20) and (2.21) we see that for case 2.2, h_1, h_2, h_3 are subjected to the following restrictions

$$(m_1 + m_2)k + h \leq h_1 - h_2 < (m_1 + m_2)k + 2h$$

and

$$h_2 + (m_2 - 1)k < h_3 < h_1 - (m_1 - 1)k \tag{2.22}$$

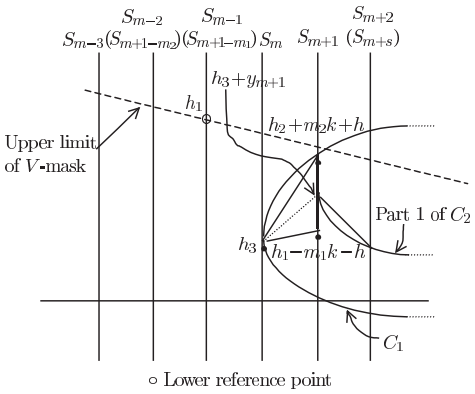


Figure 2.3.11. Lower reference point for the range of integration with respect to y_{m+2} for $S_{m+1} \in (h_1 - m_1k - h, h_2 + m_2k + h)$ in case 2.2.

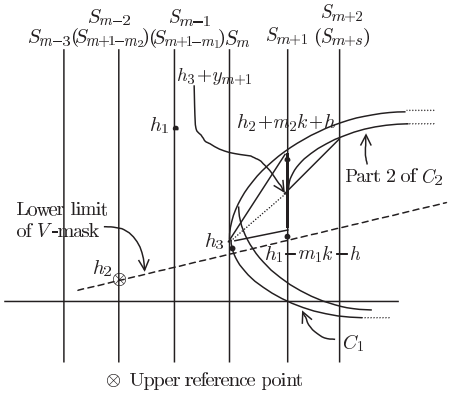


Figure 2.3.12. Upper reference point for the range of integration with respect to y_{m+2} for $S_{m+1} \in (h_1 - m_1k - h, h_2 + m_2k + h)$ in case 2.2.

(Note: The region bounded within the curve C_1 represents the paths which correspond to $F_s(m_1, m_2, h_1, h_2, h_3)$. The region bounded within the curve C_2 represents the paths which correspond to $F_{s-1}(m_1 + 1, m_2 + 1, h_1, h_2, h_3 + y_{m+1})$.)

Furthermore for case 2.2 we notice that for $S_{m+1} \in (h_1 - m_1k - h, h_2 + m_2k + h)$ the lower reference point for the range of integration with respect to y_{m+2} is S_{m+1-m_1} (see Figure 2.3.11), and the upper reference point for

the range of integration with respect to y_{m+2} is S_{m+1-m_2} (see Figure 2.3.12). Therefore

$$\begin{aligned} &F_s(m_1, m_2, h_1, h_2, h_3) \\ &= \int_{h_1-m_1k-h-h_3}^{h_2+m_2k+h-h_3} f(y_{m+1}) F_{s-1}(m_1+1, m_2+1, h_1, h_2, h_3+y_{m+1}) dy_{m+1} \end{aligned} \tag{2.23}$$

3. Comparison of the iterative formulas for computing run length distributions of one-sided and two-sided CUSUM. Let us consider a one-sided CUSUM procedure with cumulative sums

$$S_t = \max\{0, S_{t-1} + y_t - k\}$$

formed by the random variables $y_1, y_2, \ldots$ where $S_0 = w$ for some specified parameter w, $0 \le w < h$. Let $P'_t(w)$ be the probability that no out of control signal has occurred before or at time t.

Fix $w \in I = [0, h)$. Then from the literature for one-sided CUSUM, we know that $P'_t(w)$ can be computed recursively via

$$P'_0(u) = 1, \quad u \in I$$

and for all $t \ge 1$, $u \in I$

$$P'_t(u) = \left[\int_{x=0}^{k-u} f(x)dx\right] P'_{t-1}(0) + \int_0^h f(x-u+k) P'_{t-1}(x)dx \tag{3.1}$$

Equation (3.1) essentially relates $P'_t(\cdot)$ to $P'_{t-1}(\cdot)$.

Now let us examine the recursive formulas in equations (2.5), (2.10), (2.16) and (2.23) for P'_s in the two-sided CUSUM. We note that equations (2.10), (2.16) and (2.23) relate $F_s(m_1, m_2, \cdot, \cdot, \cdot)$ to $[F_{s-1}(1, m_2 + 1, \cdot, \cdot, \cdot)$, $F_{s-1}(m_1 + 1, m_2 + 1, \cdot, \cdot, \cdot)$ and $F_{s-1}(m_1 + 1, 1, \cdot, \cdot, \cdot)]$. Next we see that after

finding the $F_s(m_1, m_2, \cdot, \cdot, \cdot)$ by using (2.10), (2.16) and (2.23) we can find P'_s by using (2.5). Thus basically equations (2.10), (2.16) and (2.23) in the two-sided CUSUM represent the counterpart of equation (3.1) in the upper one-sided CUSUM.

For the one-sided CUSUM, there are many procedures for finding $P'_t(u)$ given by (3.1). They can be classified into two main types, namely the Markov Chain approach and the numerical integration approach. Various procedures may be used to perform the numerical integration. The most common one seems to be the Gauss-Legendre quadrature which has been used by Goel and Wu [5], Vance [19], Gan [20] and Luceño and Puig-Pey ([13] and [14]). The ideas used in these procedures may also be applied to the recursive formulas in equations (2.10), (2.16) and (2.23). In the next section, we use numerical integration to find P'_s, $1 \leq s \leq 7$ in the two-sided CUSUM for normally distributed random variables.

4. Two-sided CUSUM for normally distributed random variables. Consider the case when the quality characteristic $x \sim N(\mu, \sigma^2)$, that is x is normally distributed with mean μ and known variance σ^2. The cumulative sums are

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

where $y_i = c(\bar{x}_i - \mu_0)$ (see Section 2).

We may choose c to be $\sqrt{n}/\sigma$ so that $y_i \sim N(0, 1)$ when $\mu = \mu_0$. If the process mean 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

$$f(y_i) = 1/\sqrt{2\pi}\exp(-(y_i - \Delta)^2/2), \quad -\infty < y_i < \infty.$$

We may find the distribution $P_1, P_2, \ldots$ of the run length of two-sided CUSUM as outlined below.

The values of $F_{s+1}(m_1', m_2', h_1', h_2', h_3')$ can be found based on the values of $F_s(m_1, m_2, h_1, h_2, h_3)$ by means of Romberg integration. In performing the numerical integration involving F_s we have made use of the following approximations.

If h_1, h_2 and h_3 are such that $F_s(m_1, m_2, h_1, h_2, h_3)$ is a function of a single variable (y_1,say), then $(N+1)$ equally spaced values $y_{11} < y_{12} < \cdots < y_{1N+1}$ of y_1 are chosen. Next the values of F_s are approximated by the straight lines joining the consecutive points (y_{1i}, F_s evaluated at y_{1i}) and (y_{1i+1}, F_s evaluated at y_{1i+1}), $i = 1, 2, \ldots, N$.

If h_1, h_2 and h_3 are such that $F_s(m_1, m_2, h_1, h_2, h_3)$ is a function of two variables (y_1 and y_2, say) then $(N+1)$ equally spaced values of y_1 are chosen, and for each chosen value of y_1, $(N+1)$ equally spaced values of y_2 are chosen. Next the values of F_s for fixed value c_1 of y_1 are approximated by the straight lines joining the consecutive points (y_{2j}, F_s evaluated at $y_1 = c$ and $y_2 = y_{2j}$) and (y_{2j+1}, F_s evaluated at $y_1 = c$ and $y_2 = y_{2j+1}$), where y_{2j} is the j^{th} chosen value of y_2, $j = 1, 2, \ldots, N$.

By using the above approximations, we can evaluate P_s' given by (2.5), (2.10), (2.16) and (2.23). To provide a check for the formulas given by (2.5), (2.10), (2.16) and (2.23) we compare the computed values of $P_s = P_{s-1}' - P_s'$, $1 \leq s \leq 7$ with those obtained by simulation which uses 60,000 values of $(y_1, y_2, \ldots, y_s)$.

Tables 4.1 – 4.3 show the computed values of P_s when $N = 33$ is used and the simulated value $\tilde{P}_s$. We see that P_s and $\tilde{P}_s$ in general do not differ much. The tables also show three values of ARL (ARL(A), ARL(B) and ARL(S)). The average run lengths ARL(A), ARL(B) and ARL(S) are based on the formulas

$$\text{ARL(A)} = \sum_{s=1}^{7} sP_s$$

$$1/\text{ARL(B)}=1/\text{ARL}_1+1/\text{ARL}_2$$

$$\text{ARL(S)} = \sum_{s=1}^{7} s\tilde{P}_s$$

where ARL_1 is the ARL of lower one-sided CUSUM and ARL_2 is the ARL of the upper one-sided CUSUM (see Kemp [15]). We see that $\sum_{s=1}^{7} P_s \approx \sum_{s=1}^{7} \tilde{P}_s \approx 1$ and ARL(A)≈ARL(B)≈ARL(S). The above findings indicate that the formulas given by (2.5), (2.10), (2.16) and (2.23) for finding P_s are valid.

Table 4.1. Computed and simulated probabilities of getting an out of control signal at time s ($k = 2.0481$, $h = 1.4337$ and $\triangle = 3.50$).

s	P_s	$\tilde{P}_s$
1	5.07260348685709E-0001	5.07000000000000E-0001
2	3.66788394702659E-0001	3.66066666666667E-0001
3	9.76028582084776E-0002	9.84333333333333E-0002
4	2.21379204793683E-0002	2.11500000000000E-0002
5	4.85815371228221E-0003	4.88333333333333E-0003
6	1.05826689807627E-0003	1.30000000000000E-0003
7	2.30136233283350E-0004	3.16666666666667E-0004
Total	9.99936078919856E-0001	9.99150000000000E-0001
ARL(A) = 1.65444871821679E+0000	ARL(B) = 1.65496333774341E+0000	ARL(S) = 1.65346666666667E+0000

Table 4.2. Computed and simulated probabilities of getting an out of control signal at time s ($k = 2.488$, $h = 2.4876$ and $\triangle = 2.00$).

s	P_s	$\tilde{P}_s$
1	2.30744740067377E-0001	2.30983333333333E-0001
2	5.39872207752866E-0001	5.40983333333333E-0001
3	1.82892001204641E-0001	1.81250000000000E-0001
4	3.82214127102981E-0002	3.68500000000000E-0002
5	6.84949322814084E-0003	7.23333333333333E-0003
6	1.18372967571990E-0003	1.36666666666667E-0003
7	1.97506105412190E-0004	2.83333333333333E-0004
Total	9.99961090744455E-0001	9.98950000000000E-0001
ARL(A) = 2.05478319696113E+0000	ARL(B) = 2.05493582803509E+0000	ARL(S) = 2.05045000000000E+0000

Table 4.3. Computed and simulated probabilities of getting an out of control signal at time s ($k = 0.4852$, $h = 0.1208$ and $\Delta = 0.00$).

s	P_s	$\tilde{P}_s$
1	5.44514753214789E-0001	5.44883333333333E-0001
2	2.49703665324007E-0001	2.45766666666667E-0001
3	1.12820945812990E-0001	1.13633333333333E-0001
4	5.08515075670394E-0002	5.23500000000000E-0002
5	2.30909271051045E-0002	2.25666666666667E-0002
6	1.04268696344777E-0002	1.03166666666667E-0002
7	4.71025137602634E-0003	4.08333333333333E-0003
Total	9.96118920034434E-0001	9.93600000000000E-0001
ARL(A) = 1.79677856453450E+0000	ARL(B) = 1.83082357212641E+0000	ARL(S) = 1.79003333333333E+0000

Theoretically P_s, $s \geq 8$ can be found by using (2.5), (2.10), (2.16) and (2.23). But the efficient evaluation of P_s, $s \geq 8$ poses a problem which is much more challenging than that in solving equation (3.1) for the one-sided CUSUM.

Acknowledgment.

The authors are grateful to the referee for his useful suggestions.

References

1. G. A. Barnard, *Sampling inspection and statistical decisions*, J.R. Statist. Soc. B, **16**(1954), 151-165.

2. A. F. Bissell, *CUSUM techniques for quality control*, Applied Statistics, **18**(1969), 1-30.

3. E. S. Page, *Continuous inspection schemes*, Biometrika, **42**(1954), 100-114.

4. W. D. Ewan and K. W. Kemp, *Sampling inspection of continuous processes with no autocorrelation between successive results*, Biometrika, **47**(1960), 363-380.

5. A. L. Goel and S. M. Wu, *Determination of A.R.L. and a contour nomogram for CUSUM charts to control normal mean*, Technometrics, **13**(1971), 221-230.

6. D. Brook and D. A. Evans, *An approach to the probability distribution of CUSUM run length*, Biometrika, **59**(1972), 539-549.

7. S. Zacks, *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**(1981), 2245-2258.

8. W. H. Woodall, *The distribution of the run length of one-sided CUSUM procedures for continuous random variables*, Technometrics, **24**(1983), 295-301.

9. W. H. Fellner, *Algorithm AS 258, average run length for cumulative sum schemes*, Applied Statistics, **39**(1990), 402-412.

10. D. M. Hawkins, *A fast accurate approximation for average run lengths of CUSUM control charts*, Journal of Quality Technology, **24**(1992), 37-43.

11. S. Vardeman and D. Ray, *Average run lengths for CUSUM schemes when observations are exponentially distributed*, Technometrics, **27**(1985), 145-150.

12. F. F. Gan, *Exact run length distribution for one-sided exponential CUSUM schemes*, Statistica Sinica, **2**(1992), 297-312.

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

14. A. Luceño and J. Puig-Pey, *Computing the run length probability distribution for CUSUM charts*, Journal of Quality Technology, **34**(2002)(2), 209-215.

15. K. W. Kemp, *Formal expressions which can be applied to CUSUM charts* (with discussion), Journal of the Royal Statistical Society, **B33**(1971), 331-360.

16. W. H. Woodall, *On the Markov chain approach to the two-sided CUSUM procedure*, Technometrics, **26**(1984), 41-46.

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

18. D. M. Hawkins and D. H. Olwell, Cumulative Sum Charts and Charting for Quality Improvement, Springer (1998), New York, NY.

19. L. C. Vance, *Average run lengths of cumulative sum control charts for controlling normal means*, Journal of Quality Technology, **18**(1986), 189-193.

20. F. F. Gan, *The run length distribution of a cumulative sum control chart*, Journal of Quality Technology, **25**(1993), 205-215.

Faculty of Engineering, Multimedia University, 63100 Cyberjaya, Selangor, Malaysia.

Institute of Mathematical Sciences, University of Malaya, 50603 Kuala Lumpur, Malaysia.