Biometrika (1972), **59**, 3, *p.* 539
With 1 *text-figure*
Printed in Great Britain

An approach to the probability distribution of cusum run length

By D. BROOK and D. A. EVANS
University of Newcastle upon Tyne

Summary

The classical method of studying a cumulative sum control scheme of the decision interval type has been to regard the scheme as a sequence of sequential tests, to determine the average sample number for these component tests and hence to study the average run length for the scheme. A different approach in which the operation of the scheme is regarded as forming a Markov chain is set out. The transition probability matrix for this chain is obtained and then the properties of this matrix used to determine not only the average run lengths for the scheme, but also moments and percentage points of the run-length distribution and exact probabilities of run length. The method may be used with any discrete distribution and also, as an accurate approximation, with any continuous distribution for the random variable which is to be controlled. Examples are given for the cases of a Poisson random variable and a normal random variable.

Some key words: Cumulative sum charts; Decision interval schemes; Markov chain applications; Convergence to limiting distributions; Maximum eigenvalue of stochastic matrix; Run-length distribution; Robustness of average run length; Industrial inspection.

1. Introduction

The cumulative sum, cusum, type of sampling inspection scheme, used to assess whether or not a continuing manufacturing process is running satisfactorily, has been described by many authors; see, for example, van Dobben de Bruyn (1968), Bissell (1969) and Kemp (1971). In the discrete case we shall suppose that samples of N items are taken at regular intervals and the number of defectives, D, is observed. In the decision interval type of scheme the process is regarded as in control so long as D is less than some reference value K, but as soon as D exceeds K, we commence plotting the cumulative sums

$$S_n = \sum_{i=1}^{n} (D_i - K)$$

against the number of samples, n. If S_n reverts to zero the process is once again in control, but if S_n reaches or exceeds a value H, known as the decision interval, then appropriate action must be taken. In the 'continuous' case we can use a similar scheme to control the mean or variability of some aspect of a production process.

The problem is to choose suitable values for K and H so that the scheme will have desired properties. Almost all work in this area has concentrated on the average run length of the scheme to be chosen. Thus, in practice, one might select a suitable sampling scheme by fixing the average run length for two situations, one being when the quality level is acceptable and one where it is rejectable. A nomogram, such as those provided by Goel & Wu (1971) or Ewan & Kemp (1960), might be used for this purpose.

The Markov chain approach described in this paper provides a simple means of examining more detailed properties of the particular scheme selected, for example

(1) the actual probability distribution of run length, including upper and lower percentage points;

(2) the effect of departures from the assumed probability distribution;

(3) the operating characteristics of the scheme as the population mean changes;

(4) the average run length for an arbitrary population for which nomograms do not exist.

The simplicity of the Markov chain method arises from the fact that the calculations involve no more than the use of standard computer library subroutines such as matrix multiplication and inversion, and the determination of eigenvalues and eigenvectors.

Previous approaches have been to regard the cumulative sum scheme as a sequence of sequential probability ratio tests (Page, 1954; Kemp, 1958). The resulting integral-difference equations have to be solved by numerical methods to obtain the average sample number of a single sequential test (Barraclough & Page, 1959; Goel & Wu, 1971). Finally, the average run length for the cusum scheme is obtained by compounding these sequential tests. The same approach has been used by Ewan & Kemp (1960) to find approximate results for the higher moments of run length and run-length probabilities.

Our approach to decision interval schemes is different.

2. The Markov chain approach

2·1. *Introduction*

Consider first the discrete case. Let D be an integer random variable and let K and H have positive integer values; fractional values can be covered by suitable rescaling. Then the random variable S_n can only take one of the integer values $0, 1, \ldots, H$. If $S_n = i$, then we will say that the decision interval scheme is in state E_i. Each realization of the scheme can then be regarded as a random walk over the states $E_0, E_1, \ldots, E_H$, where E_H is an absorbing state. We assume that the process is initially in state E_0.

The transition probabilities from state E_i $(i = 0, 1, \ldots, H-1)$ are determined only by the probability distribution of D and are as follows:

$$
\begin{aligned}
p_{i0} &= \operatorname{pr}(E_i \to E_0) \;= \operatorname{pr}(D \leqslant K-i),\\
p_{ij} &= \operatorname{pr}(E_i \to E_j) \;= \operatorname{pr}(D = K+j-i) \quad (j = 1, \ldots, H-1),\\
p_{iH} &= \operatorname{pr}(E_i \to E_H) = \operatorname{pr}(D \geqslant K+H-i).
\end{aligned}
$$

The system forms a Markov chain whose transition probability matrix, $\mathbf{P}$ say, is easily constructed from the probability distribution of D, given H and K. Hence, the exact probability distribution of run length and its moments can be determined.

If the process control variable is theoretically continuous, the Markov representation will still apply if we group the variable's possible values into discrete class intervals corresponding, for example, to the rounding-off interval used when making practical observations. As the rounding-off interval gets finer, we approach a limiting form. We find that for the normal distribution about 10 class intervals enables us to estimate the average run length to within $\pm 5\%$ of its limiting value. We return to the continuous case in §4 after first giving detailed results for the discrete case.

2.2. *Construction of the transition probability matrix*

Let $p_r = \mathrm{pr}\,(D-K = r)$ and let $F_r = \mathrm{pr}\,(D-K \leqslant r)$; then the transition probability matrix $\mathbf{P}$ defined in §2·1 has the following form:

$$\mathbf{P} = \begin{bmatrix} F_0 & p_1 & p_2 & \cdots & p_j & \cdots & p_{H-1} & 1-F_{H-1} \\ F_{-1} & p_0 & p_1 & \cdots & p_{j-1} & \cdots & p_{H-2} & 1-F_{H-2} \\ \vdots & \vdots & \vdots & & \vdots & & \vdots & \vdots \\ F_{-i} & p_{1-i} & p_{2-i} & \cdots & p_{j-i} & \cdots & p_{H-1-i} & 1-F_{H-1-i} \\ \vdots & \vdots & \vdots & & \vdots & & \vdots & \vdots \\ F_{1-H} & p_{2-H} & p_{3-H} & \cdots & p_{j-(H-1)} & \cdots & p_0 & 1-F_0 \\ 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 1 \end{bmatrix}.$$

For example, the probability that we pass from state E_i to state E_j has the value $p_{j-i}\,(i,j \neq 0, H)$. Note that all row sums are unity and that, since state E_H is absorbing, the last line consists of zeros except for the last element. When D is a nonnegative random variable, there will be a triangular block of zeros in the lower left hand corner corresponding to $r < -K$. For the central $H-1$ columns, all elements along a line parallel to the main diagonal have the same value.

Many of the results we require can be obtained by working with the matrix $\mathbf{R}$ obtained from $\mathbf{P}$ by deleting the final row and column. In particular, all the eigenvalues of $\mathbf{R}$ are eigenvalues of $\mathbf{P}$.

3. Sampling properties of cusum run length

3·1. *Average run length and higher order moments*

Let X_i be the number of steps taken starting from E_i to reach the absorbing state E_H for the first time and let $\mu_i^{(s)}$ be the sth factorial moment of X_i, where

$$\mu_i^{(s)} = E\{X_i^{(s)}\} = E\{X_i(X_i-1)\ldots(X_i-s+1)\}.$$

By considering the Markov chain one step later, we find that

$$\begin{aligned}\mu_i^{(s)} &= \sum_{r=s}^{\infty} r^{(s)}\,\mathrm{pr}\,(X_i = r) = \sum_{r=s}^{\infty} r^{(s)} \sum_{j=0}^{H-1} p_{ij}\,\mathrm{pr}\,(X_j = r-1) \\ &= \sum_{j=0}^{H-1} p_{ij}\{\mu_j^{(s)} + s\mu_j^{(s-1)}\} \quad (s = 2, 3, \ldots; i = 0, 1, \ldots, H-1).\end{aligned}$$

In matrix form this becomes

$$(\mathbf{I}-\mathbf{R})\,\boldsymbol{\mu}^{(s)} = s\mathbf{R}\boldsymbol{\mu}^{(s-1)} \quad (s = 2, 3, \ldots), \tag{3·1}$$

where $\mathbf{R}$ is the matrix obtained from the transition probability matrix $\mathbf{P}$ by deleting the last row and column, i.e. those referring to the absorbing state E_H, $\mathbf{I}$ is the $H \times H$ identity matrix and $\boldsymbol{\mu}^{(s)}$ is the vector of sth factorial moments for the random variables

$$X_0, X_1, \ldots, X_{H-1}.$$

The special case $s = 1$ is easily seen to lead to the equation

$$(\mathbf{I}-\mathbf{R})\,\boldsymbol{\mu} = \mathbf{1}, \tag{3·2}$$

where the vector $\mathbf{1}$ has each of its H elements equal to unity. The first element of the vector

$\boldsymbol{\mu}$ gives the average run length for a cusum chart starting from zero, usually referred to as its average run length whilst, in general, the ith element gives the mean of the run-length distribution when starting from state E_i ($i = 0, 1, \ldots, H-1$).

If the average run length is all that is required, then equation (3·2) for $\boldsymbol{\mu}$ can be very simply solved for the average run length by using pivotal condensations, starting with the last diagonal element.

Let $\mathbf{X}$ be the random vector of run lengths with elements $X_0, X_1, \ldots, X_{H-1}$; then higher order factorial moments of $\mathbf{X}$ can easily be obtained recursively *via* (3·1). We have that

$$\boldsymbol{\mu} = \boldsymbol{\mu}^{(1)} = (\mathbf{I}-\mathbf{R})^{-1}\mathbf{1}, \tag{3·3}$$

$$\begin{aligned}\boldsymbol{\mu}^{(s)} &= s\{(\mathbf{I}-\mathbf{R})^{-1}\mathbf{R}\}\boldsymbol{\mu}^{(s-1)},\\ &= s\{(\mathbf{I}-\mathbf{R})^{-1}-\mathbf{I}\}\boldsymbol{\mu}^{(s-1)} \quad (s = 2, 3, \ldots).\end{aligned} \tag{3·4}$$

Thus the matrix multiplication in (3·4) can be avoided and what is required for each cycle is the premultiplication of a scalar times the previous solution vector by the matrix $(\mathbf{I}-\mathbf{R})^{-1}-\mathbf{I}$.

Note that $(\mathbf{I}-\mathbf{R})^{-1}\mathbf{R} = \mathbf{R}(\mathbf{I}-\mathbf{R})^{-1}$, so that the solution to (3·4) may be written in either of the following forms:

$$\begin{aligned}\boldsymbol{\mu}^{(s)} &= s!\,(\mathbf{I}-\mathbf{R})^{-s}\mathbf{R}^{s-1}\mathbf{1}\\ &= s!\mathbf{R}^{s-1}(\mathbf{I}-\mathbf{R})^{-s}\mathbf{1} \quad (s = 1, 2, \ldots).\end{aligned} \tag{3·5}$$

Hence the probability distribution of the run-length vector $\mathbf{X}$ can be regarded as a multi-dimensional generalization of a geometric distribution over the positive integers. For, let N be such that $\text{pr}\,(N = r) = pq^{r-1}$ ($r = 1, 2, \ldots$), where $p+q = 1$ ($0 < p < 1$); then its sth factorial moment is $s!\,q^{s-1}p^{-s}$ ($s = 1, 2, \ldots$). If $\mathbf{R}$ was diagonal, then the elements of $\mathbf{X}$ would each be geometric random variables. The scalar parameter p is replaced by the matrix $(\mathbf{I}-\mathbf{R})$ and the elements of the vector $(\mathbf{I}-\mathbf{R})\mathbf{1}$ are the probabilities of jumping to the absorbing state, i.e. the conventional success, the first occurrence of which terminates the process.

Equation (3·5) shows that for large s the higher-order factorial moments are proportional to the corresponding moments of a geometric distribution. The necessary algebraic details are discussed in § 3·3, but briefly, if λ is the largest eigenvalue of $\mathbf{R}$, then λ is real and less than unity, and we find that

$$\boldsymbol{\mu}^{(s)} \simeq s!\,\lambda^{s-1}(1-\lambda)^{-s}\,\mathbf{c},$$

where $\mathbf{c}$ is a vector of positive constants.

We conclude this section with a numerical illustration of the moments of run-length distribution. Suppose that D has a Poisson distribution with mean 3·2, for example, samples of size 80 with the defective level at the unacceptably high value of 4 %. An average run length of about 3 can be arranged by choosing $K = 2$ and $H = 3$. In this case

$$\mathbf{R} = \begin{bmatrix} 0·3799 & 0·2226 & 0·1781 \\ 0·1712 & 0·2087 & 0·2226 \\ 0·0408 & 0·1304 & 0·2087 \end{bmatrix};$$

using (3·3) and (3·4), we find $\boldsymbol{\mu}^{(1)}$ by summing the rows of $(\mathbf{I}-\mathbf{R})^{-1}$, $\boldsymbol{\mu}^{(2)}$ by postmultiplying $\{(\mathbf{I}-\mathbf{R})^{-1}-\mathbf{I}\}$ by $2\boldsymbol{\mu}^{(1)}$, and so on. From these we can determine the central moments and also coefficients of skewness and kurtosis which are shown in Table 1.

Table 1. *Central moments of cusum run length for the case $K = 2$ and $H = 3$, where D has a Poisson distribution with mean* 3·2

State	μ	μ_2	μ_3	μ_4	σ	σ/μ	μ_3/σ^3	$(\mu_4/\sigma^4)-3$
0	3·01	3·95	13·5	120·1	1·99	0·66	1·72	4·71
1	2·43	3·35	12·7	105·3	1·83	0·75	2·08	6·41
2	1·82	2·22	9·6	74·0	1·49	0·82	2·91	12·09

The variance is greater than the mean run length for any of the possible initial states and the skewness and kurtosis are always positive. A basic comparison is with a geometric distribution over the positive integers with mean $1/p$. This has

$$\sigma^2 = q/p^2, \quad \mu_3/\sigma^3 = (1+q)/q^{\frac{1}{2}}, \quad (\mu_4/\sigma^4)-3 = (q^2+4q+1)/q.$$

In particular, if $1/p = 3{\cdot}01$, then $\sigma^2 = 6{\cdot}03$, $\sigma = 2{\cdot}46$, $\mu_3/\sigma^3 = 2{\cdot}04$ and $(\mu_4/\sigma^4)-3 = 6{\cdot}2$. Thus the run length from state E_0 has a smaller variance than a geometric random variable with the same mean and it also has smaller skewness and kurtosis.

3·2. *The probability distribution of run length*

If $\mathbf{P}$ is the transition probability matrix as defined in §2·2, then the last column of $\mathbf{P}^r$ gives $\text{pr}(X_i \leqslant r)\,(i = 0, 1, \ldots, H-1)$, together with unity as the last element. If we write $\mathbf{P}$ in the partitioned form

$$\mathbf{P} = \left[\begin{array}{c:c} \mathbf{R} & \mathbf{p}_H \\ \hdashline \mathbf{0}^T & 1 \end{array}\right],$$

where $\mathbf{p}_H$ is a vector of H elements and $\mathbf{R}$ is the $H \times H$ matrix discussed above, then

$$\mathbf{p}_H = (\mathbf{I}-\mathbf{R})\,\mathbf{1},$$

since $\mathbf{P}$ is a stochastic matrix and hence its row sums are equal to unity. It follows that

$$\mathbf{P}^r = \left[\begin{array}{c:c} \mathbf{R}^r & (\mathbf{I}-\mathbf{R}^r)\,\mathbf{1} \\ \hdashline \mathbf{0}^T & 1 \end{array}\right] \quad (r = 1, 2, \ldots).$$

Let $\mathbf{F}_r$ be a vector of length H whose elements are the distribution functions of run length starting from states $E_0, E_1, \ldots, E_{H-1}$, that is

$$\mathbf{F}_r = \{\text{pr}(X_0 \leqslant r), \text{pr}(X_1 \leqslant r), \ldots, \text{pr}(X_{H-1} \leqslant r)\}^T \quad (r = 1, 2, \ldots).$$

Then

$$\mathbf{F}_r = (\mathbf{I}-\mathbf{R}^r)\,\mathbf{1} \quad (r = 1, 2, \ldots),$$

and the first element of $\mathbf{F}_r$ gives the cumulative probability for run length for a cusum scheme starting from zero. Note that the vector $\mathbf{R}^r\mathbf{1}$ gives the corresponding right hand tail probabilities. In the case of short average run length the elements of $\mathbf{R}^r\mathbf{1}$ decrease quite rapidly with increasing r and vanish to 4 decimal places for moderately large r, say $r > 20$.

For our example in §3, for which the average run length is 3·01, the probability distribution of run length starting from state E_0 is shown in Fig. 1.

Let $\mathbf{L}_r$ be a vector of length H whose elements are the values of the probability functions of run length starting from states $E_0, E_1, \ldots, E_{H-1}$, that is,

$$\mathbf{L}_r = \{\text{pr}(X_0 = r), \text{pr}(X_1 = r), \ldots, \text{pr}(X_{H-1} = r)\}^T \quad (r = 1, 2, \ldots).$$

Then

$$\mathbf{L}_1 = \mathbf{p}_H = (\mathbf{I}-\mathbf{R})\,\mathbf{1}, \tag{3·6}$$

$$\mathbf{L}_r = \mathbf{R}\mathbf{L}_{r-1} = \mathbf{R}^{r-1}\mathbf{L}_1 \quad (r = 1, 2, \ldots), \tag{3·7}$$

which is closely analogous to the univariate geometric form

$$\mathrm{pr}\,(N = r) = q^{r-1}p \quad (r = 1, 2, \ldots).$$

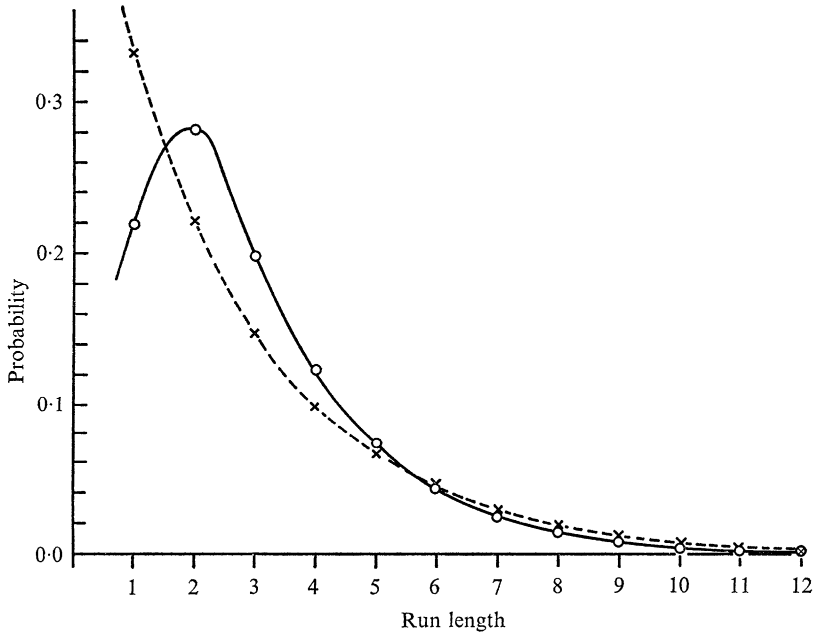

Fig. 1. Probability distribution of run length for the case $H = 3$, $K = 2$, where D has a Poisson distribution with mean 3·2. Actual run-length distribution: —○—. Corresponding probabilities for a geometric distribution with mean 3·01: - - - × - - -.

The corresponding probabilities for a geometric distribution with the same mean, 3·01, are shown, for comparison, in Fig. 1.

3·3. *The limiting form of the probability distribution of run length*

For cusum schemes with short average run length, the probability function of run length can be calculated from the simple recursive formula $\mathbf{L}_{r+1} = \mathbf{R}\mathbf{L}_r$; see (3·6) and (3·7). For long average run length this would be impracticable. For acceptable quality levels the average run length is usually 500 or more and the probability distribution of run length has a very long tail and the upper 5 % point may be as large as 1300. We now show that under these circumstances the run-length probability distribution has a geometric type tail in the limit for large r and we then obtain an approximation for the upper percentage points of the distribution of run length.

The matrix $\mathbf{R}$ is an $H \times H$ matrix whose row sums are all less than unity. In practical applications $\mathbf{R}$ is irreducible, and hence necessarily primitive. By the Perron–Frobenius Theorem we deduce that $\mathbf{R}$ has a simple real eigenvalue, λ say, $\lambda < 1$, which exceeds the moduli of all the other eigenvalues and for which there exist positive right hand and left hand eigenvectors, $\mathbf{x}$ and $\mathbf{y}^T$, respectively; see for example Cox & Miller (1965, p. 120). By expressing $\mathbf{R}$ in its Jordan canonical form, we deduce that

$$\lim_{r\to\infty} (1/\lambda^r)\,\mathbf{R}^r = \mathbf{x}\mathbf{y}^T/(\mathbf{y}^T\mathbf{x}),$$

where $\mathbf{Rx} = \lambda\mathbf{x}$ and $\mathbf{y}^T\mathbf{R} = \lambda\mathbf{y}^T$, with $\mathbf{x}$ and $\mathbf{y} > \mathbf{0}$.

Thus an approximate expression for $\mathbf{L}_r$, the probability that run length equals r, is

$$\mathbf{L}_r \simeq (1-\lambda)\,\lambda^{r-1}\left(\frac{\Sigma y}{\Sigma xy}\right)\mathbf{x}. \tag{3·8}$$

Similarly, an approximate expression for $\mathbf{1}-\mathbf{F}_{r-1}$, the probability that the run length equals r or more, may be found to be

$$\mathbf{1}-\mathbf{F}_{r-1} = \mathbf{R}^{r-1}\mathbf{1} \simeq \lambda^{r-1}\left(\frac{\Sigma y}{\Sigma xy}\right)\mathbf{x}, \tag{3·9}$$

so that the limiting distribution of run length for large r, from any initial state, is of the geometric type with parameter λ, but with a multiplying constant. Note that the approximations in (3·8) and (3·9) do not form a proper probability distribution if evaluated for all values of r. They are intended only to give accurate approximations to the probabilities for large r and do in fact seem to be accurate for r as low as 5 or 6 even for short average run length; see the example below.

The conditional probability that the run length is equal to r, given that it is greater than or equal to r, tends to $(1-\lambda)$ as r tends to infinity, for any initial starting state.

Denote the elements of $\mathbf{x}$ by $x_0, x_1, \ldots, x_{H-1}$ and let

$$c_i = (x_i\,\Sigma y)/(\Sigma xy) \quad (i = 0, 1, \ldots, H-1);$$

then the particular results for the run length from the initial state E_0 are, for large r,

$$\begin{aligned} &\text{pr (run length} \leqslant r|E_0) \simeq 1-c_0\lambda^r,\\ &\text{pr (run length} = r|E_0) \simeq c_0(1-\lambda)\,\lambda^{r-1},\\ &\text{pr (run length} \geqslant r|E_0) \simeq c_0\lambda^{r-1}. \end{aligned} \tag{3·10}$$

Equation (3·8) gives $E(X_i) \simeq c_i/(1-\lambda)$, where $c_i \simeq 1$ for long run length and in particular, the average run length is approximately

$$c_0/(1-\lambda). \tag{3·11}$$

Approximation (3·9) above can be used to find the approximate value of the upper percentage point of the probability distribution of run length. Let $r_i(\alpha)$ be the value of r for an upper tail probability of α when starting from state E_i; then

$$\begin{aligned} r_i(\alpha) &\simeq 1+(1/\log\lambda)\log\{(\alpha\Sigma xy)/(x_i\,\Sigma y)\}\\ &= 1+(1/\log\lambda)\log(\alpha/c_i) \quad (i = 0, 1, \ldots, H-1). \end{aligned} \tag{3·12}$$

When λ is close to unity, c_0 is approximately unity and hence the upper percentage point for run length is approximately $1-(\text{ARL})\log\alpha$, where ARL is the average run length; this agrees with the formula anticipated by Ewan & Kemp (1960, equation 16). However, for moderately large average run length, a better approximation is

$$1+(1/\log\lambda)\log(\alpha/c_0).$$

Reverting briefly to the discussion in §3·1 on higher-order moments, we have from the properties of λ in relation to $\mathbf{R}$ that $1/(1-\lambda)$ is the largest eigenvalue of $(\mathbf{I}-\mathbf{R})^{-1}$, so that

$$\boldsymbol{\mu}^{(s)} = s!\mathbf{R}^{s-1}(\mathbf{I}-\mathbf{R})^{-s}\mathbf{1} \simeq \frac{s!\,\lambda^{s-1}(\Sigma y)}{(1-\lambda)^s(\Sigma xy)}\mathbf{x}.$$

For the example introduced in §3·1, we find

$$\lambda = 0\cdot5849,\ c_0/(1-\lambda) = 3\cdot66,$$
$$\mathbf{y}^T = (0\cdot3348,\ 0\cdot3184,\ 0\cdot3469),$$
$$\mathbf{x}^T = (0\cdot5030,\ 0\cdot3286,\ 0\cdot1684),$$

where the elements of $\mathbf{x}$ and of $\mathbf{y}$ have been standardized to sum to unity. Hence

$$c_0 = (x_0\Sigma y)/(\Sigma xy) = 1\cdot5178,$$

and the approximations for large r are as follows:

$$\text{pr}(\text{run length} = r|E_0) \simeq (0\cdot6300)(0\cdot5849)^{r-1},$$
$$\text{pr}(\text{run length} \geqslant r|E_0) \simeq (1\cdot5178)(0\cdot5849)^{r-1}.$$

We obtain agreement with the exact values of (3·6) and (3·7) to four decimal places for $r \geqslant 6$.

The percentage points for this example can be accurately approximated using (3·12). For example, the 5% point given by (3·12) is 7·4, whilst an evaluation of the exact probabilities shows that pr(run length $\geqslant 7$) = 0·0608 and pr(run length $\geqslant 8$) = 0·0356. The approximation of Ewan and Kemp estimates the 5% point at 10·02 which corresponds to a true probability of 0·012.

4. Applications to continuous distributions

4·1. *General distributions*

We now consider the case where the attribute of the production process which is to be controlled is assessed by a theoretically continuous random variable, Z say, so that the inspection scheme is required to detect a change in the mean of Z. In practice Z will usually correspond to an actual measured dimension, but may equally well correspond to some other feature, such as a measure of the spread of a probability distribution. We shall consider the one-sided type of decision interval scheme for which we accumulate the deviations of Z from a reference value K, whilst the cumulative sum is positive, until either we cross the decision boundary H or the cumulative sum reverts back to zero.

The operation of such a scheme forms a Markov process with a continuous state space. Good approximations to the various characteristics of the run-length distribution can be obtained by discretizing the probability distribution of Z so that the cumulative sum is restricted to a finite set of values. The Markov chain representation discussed in the earlier sections then becomes available. This procedure is an extension of what is done in practice anyway when rounding-off is applied to the observations on the random variable Z.

Suppose that we wish to represent the continuous scheme by a Markov chain having $t+1$ states labelled $E_0, E_1, \ldots, E_t$, where E_t is absorbing. Then the probability that the chain remains in the same state at the next step should correspond to the case where the cumulative sum does not change in value by more than a small amount, say $\frac{1}{2}w$; that is, the next value of Z does not differ from the reference value K by more than $\frac{1}{2}w$. The quantity w

determines the width of the grouping interval involved in the discretization of the probability distribution of Z.

Not all choices of value for w are equally suitable though, because properties like average run length and percentage points are quite sensitive to the effective width of the decision interval. So as a further restriction we require that the probability of a jump from E_i to the absorbing state E_t should be equal to the probability that the cumulative sum for $(Z-K)$ jumps beyond the point H from a position in $(0, H)$ which corresponds approximately with the state E_i.

These requirements lead to

$$w = 2H/(2t-1). \tag{4·1}$$

The transition probabilities for the Markov chain are then as follows, for $i = 0, 1, \ldots, t-1$;

$$P_{i0} = \mathrm{pr}\,(E_i \to E_0) = \mathrm{pr}\,(Z-K \leqslant -iw+\tfrac{1}{2}w),$$

$$P_{ij} = \mathrm{pr}\,(E_i \to E_j) = \mathrm{pr}\{(j-i)\,w-\tfrac{1}{2}w < Z-K \leqslant (j-i)\,w+\tfrac{1}{2}w\} \quad (1 \leqslant j \leqslant t-1),$$

$$P_{it} = \mathrm{pr}\,(E_i \to E_t) = \mathrm{pr}\{(t-i)\,w-\tfrac{1}{2}w < Z-K\}.$$

Note that $\mathrm{pr}\,(E_0 \to E_t) = \mathrm{pr}\,(Z-K > H)$ for any choice of w satisfying (4·1).

If we write $p_r = \mathrm{pr}\,(rw-\tfrac{1}{2}w < Z-K \leqslant rw+\tfrac{1}{2}w)$ and $F_r = \mathrm{pr}\,(Z-K \leqslant rw+\tfrac{1}{2}w)$, then the transition probability matrix, $\mathbf{P}$, for the Markov chain is constructed as before and has the general form shown in §2·2, except that the states are labelled $E_0, E_1, \ldots, E_t$. The number of states is now arbitrary and is not determined directly by H, as in the discrete case. The matrix obtained from $\mathbf{P}$ by deleting the last row and column is denoted by $\mathbf{R}$, as before, and the sampling properties of the corresponding cusum scheme are determined in the manner described above for the discrete case.

For the normal distribution, estimates of characteristics such as average run length or percentage points have been found to converge quite quickly as t is increased. We have obtained agreement to within 5 % of the limiting value when $t = 5$ and agreement to within 1 % of the limiting value when $t = 10$, both for cases of short average run length and cases of long average run length.

4·2. *The normal distribution*

Without loss of generality we can assume that the unit of measurement for Z is chosen so that the standard deviation is unity and we assume that the problem is one of controlling an $N(\gamma, 1)$ random variable using a decision interval scheme with reference value K and decision interval H. But if Z is $N(\gamma, 1)$, then $Z-K$ is $N(\gamma-K, 1)$, so that we simply write $\mu = \gamma-K$ and consider two alternative values for μ, say μ_0 for acceptable quality level and μ_1 for rejectable quality level. For simplicity we will suppose that $\mu_0 < 0 < \mu_1$, corresponding to the one-sided case where we want to detect an increase in the mean value.

Given μ and H, we choose a value for t and find $w = 2H/(2t-1)$. We then compute the unit normal distribution function at the $(2t-1)$ points

$$a-\mu, \quad a+w-\mu, \quad a+2w-\mu, \ldots, a+(2t-2)\,w-\mu = H-\mu,$$

where $a = -(H-w)$, and hence the probabilities for the discretized distribution. In practice rather fewer evaluations will be required since the unit normal density function will be effectively zero over portions of the nominal range, $(-H+w-\mu, H-\mu)$, when $H > 3$.

For computer use we have usually used one of the rational polynomial approximations given by Hastings (1955, p. 187), giving at least 7 or 8 decimal place accuracy.

For example, let $H = 3$ and take $t = 5$, so that the absorbing state is E_5 and the width of the grouping interval is $w = \frac{2}{3}$. The matrix of transition probabilities, $\mathbf{R}$, is 5×5 and the rows and columns correspond to the non-absorbing states $E_0, E_1, \ldots, E_4$. The vector of average run lengths corresponding to each of these five possible starting states is given by $\boldsymbol{\mu} = (\mathbf{I}-\mathbf{R})^{-1}\mathbf{1}$. The first element of $\boldsymbol{\mu}$ is the average run length and for the case $\mu = 1{\cdot}0$ its value is 3·77. The maximum eigenvalue of $\mathbf{R}$ is $\lambda = 0{\cdot}5121$ and $c_0 = 4{\cdot}343$. Hence, using (3·10), we see that the probability that the run length from E_0 does not exceed r is approximately $1-(4{\cdot}343)(0{\cdot}5121)^r$ and the upper 5 % and 1 % points, from (3·12), are approximately 7·67 and 10·08. Alternatively the approximate cumulative probability distribution of run length may be obtained by evaluating $\mathbf{R}^r\mathbf{1}$ $(r = 1, 2, \ldots)$, and hence the probabilities of a run length from E_0 of 8 or more and of 10 or more are found to be 3·9 % and 1·05 %, approximately.

When $\mu = -0{\cdot}5$, $H = 3$ and $t = 5$, we find that the average run length is 113·47, $\lambda = 0{\cdot}99098$, $c_0 = 1{\cdot}024$ and the approximate value for the mean using (3·11) is

$$c_0/(1-\lambda) = 113{\cdot}53.$$

The approximate probability of a run length of r or more is $c_0\lambda^{r-1} = (1{\cdot}024)\,(0{\cdot}99098)^{r-1}$. For $r = 6$, this expression yields 0·979 compared with the value 0·977 obtained by evaluating $\mathbf{R}^5\mathbf{1}$. The values agree to four significant figures at $r = 12$, both methods yielding 0·9268 for the approximate probability of a run length of 12 or more. Thus the lower 5 % region can be investigated by using a few matrix by vector products and the remainder of the probability distribution by using the modified geometric distribution, $c_0\lambda^{r-1}$. We find that the lower 5 % point is approximately 9·3, the median is approximately 80 and the upper 5 and 1 % points are approximately 334 and 512, respectively.

4·3. *The effect of changing the width of the grouping interval*

The values given in the previous section were obtained using $t = 5$, i.e. 9 grouping intervals for the normal distribution, and are therefore only approximate. The effect of increasing t is to narrow the width of the grouping interval for the discretization and to increase the size of the matrix $\mathbf{R}$.

For the short case, $\mu = 1{\cdot}0$, the average run length at $t = 5$ is about 0·7 % higher than the limiting value of 3·750. For the long case, $\mu = -0{\cdot}5$, the value at $t = 5$ is about 3·3 % lower than the limiting value of 117·59. The values at $t = 5$, 10 and 15 are 113·47, 116·63 and 117·18, respectively. To obtain a simple extrapolation to the limiting value we may assume that for large t, the average run length is

$$\text{ARL} = A + B/t + C/t^2,$$

so that the asymptotic value, A, will equal the second divided difference of $t^2(\text{ARL})$. Using the values given above for $t = 5$, 10 and 15, we obtain 117·52, compared with 117·59 using values at $t = 20$, 25 and 30.

Similarly, the maximum eigenvalue of $\mathbf{R}$ approaches a limit as t is increased. For the long run case, $\mu = -0{\cdot}5$, the value of $1-\lambda$ at $t = 5$ is about 4 % lower than the limiting value for $1-\lambda$ of 0·0087.

Thus the results at $t = 5$ seem quite reasonable for most practical purposes, but if more accuracy is required, then $t = 10$ gives values correct to about three significant figures.

5. Conclusion

We have presented in this paper a method for studying cusum decision interval schemes. Our approach differs from those of previous authors in that it essentially provides an analogue of the process, thus enabling all aspects to be studied. More accurate values for the average run length can be attained by using sophisticated numerical techniques on the integral-difference equations of the traditional approach. For example, Goel & Wu (1971) claim a six significant figure accuracy for the average run length using Gaussian quadrature involving the use of 15 points and the inversion of a 15×15 matrix. Accurate values for the normal distribution are also given by van Dobben de Bruyn (1968, Table 5*c*).

However, the robustness of the average run length to changes in population distribution has recently been questioned, and the possibility of using other run-length statistics such as percentiles has been suggested (Kemp, 1971, §9 and the discussion). The traditional approach is not easily extended to cover such aspects, whereas the Markov approach does provide a practicable method for their investigation.

References

Barraclough, E. D. & Page, E. S. (1959). Tables for Wald tests for the mean of a normal distribution. *Biometrika* **46**, 169–77.

Bissell, A. F. (1969). Cusum techniques for quality control (with discussion). *Appl. Statist.* **18**, 1–30.

Cox, D. R. & Miller, H. D. (1965). *The Theory of Stochastic Processes*. London: Methuen.

Ewan, W. D. & Kemp, K. W. (1960). Sampling inspection of continuous processes with no autocorrelation between successive results. *Biometrika* **47**, 363–80.

Goel, A. L. & Wu, S. M. (1971). Determination of A.R.L. and a contour nomogram for cusum charts to control normal mean. *Technometrics* **13**, 221–30.

Hastings, C. (1955). *Approximations for Digital Computers*. Princeton University Press.

Kemp, K. W. (1958). Formulae for calculating the operating characteristics and the average sample number of some sequential tests. *J. R. Statist. Soc.* B **20**, 379–86.

Kemp, K. W. (1971). Formal expressions which can be applied to cusum charts (with discussion). *J. R. Statist. Soc.* B **33**, 331–60.

Page, E. S. (1954). Continuous inspection schemes. *Biometrika* **41**, 100–15.

van Dobben de Bruyn, C. S. (1968). *Cumulative Sum Tests: Theory and Practice*. London: Griffin.

[*Received September* 1971. *Revised May* 1972]