

## Sensitivity Analysis of Cusum - Shewhart Control Schemes

Emmanuel Yashchin

IBM Corporation  
Thomas J. Watson Research Center  
Department of Mathematical Sciences  
Yorktown Heights, NY 10598.

### *Abstract*

Sensitivity analysis plays an important role in the problem of design of Cusum-Shewhart control schemes. This problem usually requires performing a repetitive analysis of a sequence of schemes with fixed parameters. Since such analysis is typically associated with an extensive computational effort, availability of procedures for sensitivity analysis of the Run Length characteristics at a given point would enable one to reduce substantially the number of steps needed to complete the design. In the present work we derive such procedures and examine their properties - in the context of matrix analysis (discretized schemes) and Brownian Motion approximation. The results turn out to be especially attractive for the purpose of efficient search for appropriate signal levels.

## *1. Introduction*

In recent years, Cusum - Shewhart control schemes (charts) have become increasingly popular in industrial quality control as means for monitoring the quality of manufactured products. This popularity is primarily based on the fact that performance of this type of schemes is proven to be statistically superior to their classical counterparts - Shewhart schemes ( $\bar{x}$  - charts, p - charts, etc.) in the sense that with the same degree of protection against false alarms, they have a much better sensitivity with respect to out-of-control situations. Further, Cusum - Shewhart schemes are "analyzable"; in other words, it is possible to examine, by analytic means, the RL behavior of a scheme for any given stochastic pattern of incoming (iid) observations (ex. see Brook and Evans (1972), Lucas (1982), Woodall (1983, 1984) or Yashchin (1985 a,b)); approximate results for some non-iid cases are also available (ex. see Bagshaw and Johnson (1974, 1975)). Finally, these schemes are closely associated with the Cumulative Sum technique for graphical representation of serial data, which is useful by itself for purposes of diagnostics, retrospective data analysis, estimation, etc. (ex. see Woodward, R. and Goldsmith (1964), van Dobben de Bruyn (1968) Bissell (1969) and guide by the British Standards Institution (1980-1983).

One of the most attractive properties of Cusum - Shewhart control schemes is their "designability". In other words, once the "good" and "bad" levels of the process as well as corresponding sensitivity requirements are specified, one can come up with a Cusum - Shewhart scheme (and determine the relevant sampling intensity) to meet these requirements. This property of Cusum - Shewhart schemes is especially important in situations where data is collected and/or processed automatically and in situations where several parameters are controlled simultaneously. Several design procedures were recently discussed in Lucas (1985), Woodall (1986) and Yashchin (1985,a). The latter work also introduces a package DARCS (presently called CONTRD) for design, analysis and running of Cusum - Shewhart schemes and gives examples of its application.

In the present work we address the problem of sensitivity analysis of one-sided Cusum-Shewhart schemes. This topic is especially important in connection with the problem of design. Indeed, in order to find a suitable control scheme, one usually has to perform a repetitive analysis of a sequence of

schemes with fixed parameters. Such analysis is typically associated with an extensive computational effort, which is primarily related to manipulations of underlying Markov transition matrices of typical sizes ( $50 \times 50$ ). Therefore, availability of procedures for sensitivity analysis of the Run Length characteristics at a given point enables one to reduce substantially the number of steps needed to complete the design. The most important procedure of this kind is related to sensitivity by the signal level of the scheme, as more or less appropriate values of other parameters can be determined by alternative means (ex. one could use analogy with SPRT to determine a "good" reference value, see Lorden (1971) and Lucas (1985)); we consider it here in more detail. The material of the present work serves as a basis for automated design functions implemented in the mentioned package CONTRD.

## 2. Preliminaries

In this section we provide some information needed for the derivation of our main results and introduce the appropriate notation. We shall assume that the observations  $X_1, X_2, \dots$  form a sequence of iid random variables with distribution function  $F(x)$ . The Cusum scheme (introduced by Page (1954)) is defined in terms of three parameters:  $h \geq 0$  (signal level),  $k$  (reference value) and  $0 \leq s_0 \leq h$  (headstart). Since its primary goal is to detect shifts in the process level upwards, we shall call it an *upper* Page's scheme.

**Definition 2.1.** The upper Page's scheme  $(h, k, s_0)$  is an operator transforming the sequence  $X_1, X_2, \dots$  into a set of random variables  $S_0, S_1, \dots, S_N$  defined by

$$S_0 = s_0; S_n = \max \{ S_{n-1} + (X_n - k), 0 \}, n = 1, 2, \dots, N \quad (2.1)$$

where  $N$  (Run Length) is the first index for which  $S_N \geq h$ . If  $N < \infty$  we shall say that the scheme signals at the epoch  $N$ .

If an additional signal criteria is introduced, namely:

if a single observation  $X_i$  satisfies  $X_i \geq c$ , trigger an out-of-control signal at the epoch  $i$ ,

the above control procedure will be called an upper Page's scheme supplemented by Shewhart's control limit,  $c$ . Clearly, the behaviour of the control scheme will be affected by introduction of this criteria only if  $c \leq h + k$ . Here and in what follows we refer to such (supplemented) Page's schemes as Cusum-Shewhart control schemes.

Let us clarify the roles of the parameters in a Cusum-Shewhart scheme. The reference value  $k$  is usually chosen to be close to the midpoint between the acceptable and unacceptable levels of the process; therefore, it acts as an "anchor" preventing the scheme from drifting in on-target situations. On the other hand, if the process level is unacceptable, the successive differences  $(X_n - k)$  become typically positive, they accumulate in (2.1) causing the scheme to eventually "float up" and signal.

The signal level  $h$  characterizes the degree of accumulation of information allowed in the control scheme. If  $h = 0$ , we do not allow any accumulation of evidence against the on-target hypothesis and are prepared to signal on the basis of a single observation - in other words our Cusum scheme turns into a pure Shewhart scheme with upper control limit  $k$ .

The headstart  $s_0^+$  implements the Fast Initial Response feature, i.e. it provides an instrument for detecting *initially present* out-of-control conditions earlier than similar conditions occurring later. The rationale for using a headstart is as follows: if the process is on target, the Page's scheme will be (most likely) brought to zero by the reference value, so that in this case the expected effect of the headstart is minimal; otherwise, however, the out-of-control will be triggered much sooner (ex. see Lucas and Crossier (1982)). Finally, supplementing the scheme by a Shewhart's limit improves the sensitivity of the scheme with respect to substantial increases in the process level - in other words, it removes some of the "inertia" of a Cusum scheme when facing a sharp change of the process (ex. see Lucas (1982)).

Note that schemes based on only two parameters, signal level and reference value, are frequently found quite satisfactory for practical purposes.

At present, one of the most efficient methods for analysis of run length distribution (Brook and Evans (1972)) is based on discretization of the values of Page's scheme, and then treating it as a Markov Chain. It is clear that  $S_0, S_1, \dots$  form a Markov Chain which is discrete in time, but may be contin-

uous in space. The levels 0 and  $h$  are reflecting and absorbing barriers of the chain, respectively.

For computational purposes we discretize the values of  $S_0, S_1, \dots$  as shown in Fig. 2.1.

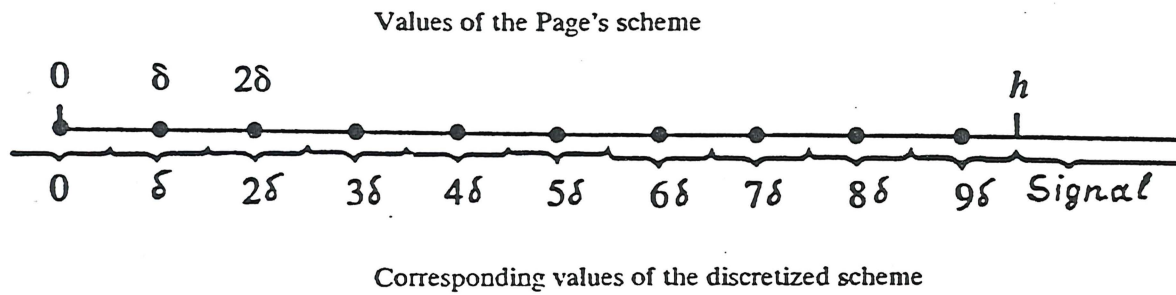


Fig. 2.1. Discretization of the values of one-sided Page's scheme.

In other words, the values of  $S_0, S_1, \dots$ , will be rounded to the center of a corresponding group. The number of groups will be termed the *level of discretization* of the scheme and denoted by  $d$ ; for example, in the case represented by Fig. 2.1, the level of discretization is  $d = 10$ . The length of an interval corresponding to a single group,  $\delta$ , will be called the *discretization interval*; it is always related to the level of discretization by means of the formula  $\delta = h/(d - 0.5)$ . Such a method of discretization usually gives approximations of good quality and is recommended in many sources (ex. Brook and Evans (1972)). Our studies show (ex. see Yashchin (1985, Table 1)) that levels of discretization of order  $d \approx 30$  are satisfactory for most practical purposes. The reason for that is related to the fact that we discretize *the states* of the Page's schemes but *not* the observations themselves. Thus, relatively low sensitivity with respect to level of discretization is explained by compensation of roundoff errors when computing subsequent values of the scheme.

The transition matrix  $P$  of the corresponding Markov chain can be expressed in terms of  $F(x)$ :

$$P_{d+1} = \begin{pmatrix} R & (I_d - R)\mathbf{1} \\ \mathbf{0}^T & 1 \end{pmatrix}, \quad (2.2)$$

where the elements  $r_{ij}$  ( $i, j = 0, 1, \dots, d - 1$ ) of  $R$  are given by

$$r_{ij} = \begin{cases} F(k + (-i + 0.5)\delta), & j = 0 \\ F(k + (j - i + 0.5)\delta) - F(k + (j - i - 0.5)\delta), & j > 0, \end{cases} \quad (2.3)$$

$\mathbf{1}$  is a vector of unities and  $I_d$  is a  $(d \times d)$  identify matrix. In what follows, bold letters will always correspond to column-vectors. Thus, columns and rows of  $R$  will be denoted by  $\mathbf{c}_0, \mathbf{c}_1, \dots, \mathbf{c}_{d-1}$  and  $\mathbf{r}_0^T, \mathbf{r}_1^T, \dots, \mathbf{r}_{d-1}^T$ , respectively.

Analysis of the run length distribution can be performed as follows. The vector  $\boldsymbol{\mu}$  containing ARL's corresponding to headstarts  $0, \delta, \dots, (d-1)\delta$  is given by

$$\boldsymbol{\mu} = (I_d - R)^{-1} \mathbf{1} \quad (2.4)$$

(Brook and Evans (1972)). The d.f. of the run length (for all values of the headstart) is

$$P\{R.L. > n\} = R^n \mathbf{1}.$$

Under the assumption that all the eigenvalues of  $R$  have the same algebraic and geometric multiplicities, there exist a spectral representation of the form

$$R = U \cdot [\text{diag}(\lambda_0, \lambda_1, \dots, \lambda_{d-1})] U^{-1}, \quad (2.5)$$

where  $\lambda_0 \geq |\lambda_1| \geq \dots \geq |\lambda_{d-1}|$  are the eigenvalues of  $R$  ( $\lambda_0$  is the Perron-Frobenius eigenvalue); the columns of  $U$  are the corresponding right eigenvectors and the rows of  $U^{-1}$  are the corresponding left eigenvectors. Denoting by  $\mathbf{u}_0, \mathbf{u}_1, \dots, \mathbf{u}_{d-1}$  the columns of  $U$ , we obtain that

$$P\{R.L. > n\} = \{w_0 \mathbf{u}_0, w_1 \mathbf{u}_1, \dots, w_{d-1} \mathbf{u}_{d-1}\} \cdot (\lambda_0^n, \lambda_1^n, \dots, \lambda_{d-1}^n)^T, \quad n = 0, 1, \dots \quad (2.6)$$

where the weights  $w_0, w_1, \dots$  chosen so that  $\sum_0^{d-1} w_i \mathbf{u}_i = \mathbf{1}$ . In what follows, we shall always assume that the right eigenvectors are scaled so that their sum is  $\mathbf{1}$ , which is equivalent to requirement that  $U^{-1} \mathbf{1} = \mathbf{1}$ .

The assumption we made about the same geometric and algebraic multiplicity of the eigenvectors of  $R$  holds in almost all practical situations; otherwise, use of a canonical Jordan matrix representation

instead of (2.5) leads to results analogous to those obtained in the present work. We do not consider this possibility in the context of sensitivity analysis, since in these rare cases where it is relevant, it is always possible to re-compute the quantities associated with a control scheme without invoking the sensitivity analysis.

### 3. Sensitivity by the signal level

Let us suppose that, after performing analysis of a CUSUM-Shewhart scheme, we would like to examine the effect of increase of  $h$  by  $\delta$ . The matrix  $\tilde{R}$  corresponding to these conditions is as follows:

$$\tilde{R} = \begin{pmatrix} R & c_d \\ r_d^T & r_{dd} \end{pmatrix}, \quad (3.1)$$

where elements of  $r_d$  and  $c_d$  are computed in accordance with formulas (2.3). It is not difficult to see that only upper-right and lower-left elements of  $\tilde{R}$  need to be computed; all the other elements come from the matrix  $R$ .

In this section we derive the basic quantities associated with the new scheme in terms of those corresponding to the original one. First, we denote by  $p^*$  the vector  $(I - R)^{-1}c_d$ . The  $j^{\text{th}}$  component of this vector has the following probabilistic meaning:

$$p_j^* = P \left\{ \begin{array}{l} \text{At the moment of signal} \\ \text{the state of the chain is } d \end{array} \mid s_0 = j\delta \right\}. \quad (3.2)$$

This fact follows easily by considering  $d$  as a *separate* absorbing state in the original scheme and analyzing the resulting transition matrix. Our first result is related to derivation of  $\tilde{\mu}$ , the new set of ARL's.

**Lemma 3.1.** Let the modified scheme start from  $s_0 = \delta d$ . Then its average run length,  $\ell$ , is given by

$$\ell = (1 + r_d^T \mu) / (1 - r_{dd} - r_d^T p^*) \quad (3.3)$$

**Proof.** By taking into account probabilistic meaning of  $\mathbf{p}^*$  and using the formula for conditional expectation, we can write

$$\ell = 1 + r_{dd} \cdot \ell + \Gamma_d^T (\mu + \mathbf{p}^* \cdot \ell), \quad (3.4)$$

from which (3.3) follows immediately. ■

A direct consequence of the above lemma is the following:

**Theorem 3.1.** The set of ARL's,  $\tilde{\mu}$  of the modified scheme is given by

$$\tilde{\mu} = \begin{pmatrix} \mu + \mathbf{p}^* \cdot \ell \\ \ell \end{pmatrix} \quad (3.5)$$
■

As one can see, derivation of  $\tilde{\mu}$  does not require additional computational effort as the matrix  $(I - R)^{-1}$  is available from the analysis of original scheme.

In the procedure of search for the signal level for which a quantile of a run length distribution has a fixed value, it is also desirable to efficiently obtain the spectral representation (2.5) of  $\tilde{R}$ . To obtain the new set of eigenvalues  $\tilde{\lambda}_0, \dots, \tilde{\lambda}_d$  and right eigenvectors  $\tilde{\mathbf{u}}_0, \dots, \tilde{\mathbf{u}}_d$ , we can use the following approach. First of all, let us find the weights  $v_0, v_1, \dots, v_{d-1}$  so that

$$\sum_{i=0}^{d-1} v_i \mathbf{u}_i = \mathbf{p}^* \quad (3.6)$$

and denote

$$\begin{aligned} C &= \{ v_0 \mathbf{u}_0, v_1 \mathbf{u}_1, \dots, v_{d-1} \mathbf{u}_{d-1} \}; \\ \mathcal{D}(\lambda) &= \text{diag} \{ (\lambda - \lambda_0)^{-1}, (\lambda - \lambda_1)^{-1}, \dots, (\lambda - \lambda_{d-1})^{-1} \}; \\ l(\lambda) &= \left\{ \frac{1 - \lambda_0}{\lambda - \lambda_0}, \frac{1 - \lambda_1}{\lambda - \lambda_1}, \dots, \frac{1 - \lambda_{d-1}}{\lambda - \lambda_{d-1}} \right\}^T; \\ l_1(\lambda) &= \mathcal{D}(\lambda) \cdot \mathbf{1}; \quad l_2(\lambda) = \mathcal{D}(\lambda) \cdot l(\lambda). \end{aligned}$$



Then one can prove the following

**Theorem 3.2.** Suppose that none of the eigenvalues  $\lambda_0, \dots, \lambda_{d-1}$  is also an eigenvalue of  $\tilde{R}$ . Then the set  $\tilde{\lambda}_0, \tilde{\lambda}_1, \dots, \tilde{\lambda}_d$  of eigenvalues of  $\tilde{R}$  correspond to solutions of the equation

$$\lambda - r_{dd} - r_d^T C(\lambda) = 0. \quad (3.7)$$

The right eigenvector  $\tilde{\mathbf{u}}_i$  of  $\tilde{R}$  corresponding to the single eigenvalue  $\tilde{\lambda}_i$  is given by

$$\tilde{\mathbf{u}}_i = \begin{pmatrix} C(\tilde{\lambda}_i) \\ 1 \end{pmatrix} \quad (3.8)$$

If all the eigenvalues of  $\tilde{R}$  are distinct and weights  $w_i$  are chosen as

$$\tilde{w}_i = \left[ 1 + r_d^T U \cdot t_1(\tilde{\lambda}_i) \right] / \left[ 1 + r_d^T C \cdot t_2(\tilde{\lambda}_i) \right], \quad (3.9)$$

then

$$\sum_0^d \tilde{w}_i \tilde{\mathbf{u}}_i = \mathbf{1}. \quad (3.10)$$

**Proof.** By a formula for determinant of a partitioned matrix and our assumption, the characteristic equation for  $\tilde{R}$  is

$$\| (\tilde{R} - \lambda I_{d+1}) \| = \| (R - \lambda I_d) \| \cdot \left[ r_{dd} - \lambda - r_d^T \cdot (R - \lambda I_d)^{-1} \mathbf{c}_d \right] = 0 \quad (3.11)$$

Further, by (2.5),

$$\begin{aligned} - (R - \lambda I_d)^{-1} \mathbf{c}_d &\equiv U \cdot \mathcal{D}(\lambda) \cdot U^{-1} \mathbf{c}_d \equiv \\ &\equiv (v_0 \mathbf{u}_0, \dots, v_{d-1} \mathbf{u}_{d-1}) / (\lambda) \equiv C(\lambda), \end{aligned} \quad (3.12)$$

since for  $\lambda = 1$  the LHS of (3.10) is  $\mathbf{p}^*$ . Therefore, the equation for eigenvalues of  $\tilde{R}$  reduces to (3.11).

To find the right eigenvector corresponding to a detected eigenvalue  $\tilde{\lambda}_i$  of  $\tilde{R}$ , we can set its last component to 1 and then identify the remaining components as  $-(R - \tilde{\lambda}_i I_d)^{-1} \mathbf{c}_d = C \ell(\tilde{\lambda}_i)$ .

Now let us find the weights  $\tilde{w}_i$  so that (3.10) holds. First of all, the left eigenvector corresponding to a detected eigenvalue  $\tilde{\lambda}_i$  of  $\tilde{R}$  can be identified as  $(-r_d^T (R - \tilde{\lambda}_i I_d)^{-1}, 1)$ . Multiplying (3.10) by this eigenvector and using bi-orthogonality property of systems of left and right eigenvectors results in  $\tilde{w}_i = (1 + r_d^T U \mathcal{D}(\tilde{\lambda}_i) U^{-1} \mathbf{1}) / (1 + r_d^T U \mathcal{D}(\tilde{\lambda}_i) U^{-1} C \ell(\tilde{\lambda}_i))$ . Finally, (3.9) follows from the fact that  $U^{-1} \mathbf{1} = \mathbf{1}$  and  $C = U \cdot \text{diag} \{v_0, v_1, \dots, v_{d-1}\}$ .

■

The assumption that none of the eigenvalues of  $R$  are also eigenvalues of  $\tilde{R}$  made in the above theorem is by no means crucial. One can see that such situation occurs if and only if  $r_d^T$  is orthogonal to appropriate right eigenvector of  $R$ . The "new" eigenvectors associated with such eigenvalues are obtained by attaching a trailing zero component to respective "old" eigenvectors. Furthermore, (3.7) - (3.8) still enables one to find the remaining eigenvalues and eigenvectors of  $\tilde{R}$ .

In many practical cases we are interested not only in the set of new ARL's or new spectral representation of the transition matrix, but in the matrix  $(I_{d+1} - \tilde{R})^{-1}$ . This matrix can be used not only for purposes of sensitivity analysis of higher order moments, but also for solving the following problem: with all the other parameters fixed, find the maximal  $h$  for which  $\text{ARL} \leq m$ , where  $m$  is some prescribed number (we assume that  $s_0$  is a multiple of  $\delta$  and  $h$  increases in steps of size  $\delta$ ). As we shall see, there exist a simple relation between  $(I_d - R)^{-1}$  and  $(I_{d+1} - \tilde{R})^{-1}$ . By using this relation and repeating the procedure (3.5) one can efficiently solve the above problem.

Let us denote

$$\mathbf{a} = \ell / (1 + r_d^T \mu); \quad \mathbf{b}^T = \mathbf{a} \cdot r_d^T (I_d - R)^{-1}. \quad (3.13)$$

Then the mentioned relation is

$$\left( I_{d+1} - \tilde{R} \right)^{-1} = \begin{pmatrix} (I_d - R)^{-1} + \mathbf{p}^* \mathbf{b}^T & \mathbf{p}^* a \\ \mathbf{b}^T & a \end{pmatrix}. \quad (3.14)$$

The relation (3.14) can be proved by applying an inversion formula for partitioned matrices (ex. see Anderson (1984, p.18)) to

$$\left( I_{d+1} - \tilde{R} \right) = \begin{pmatrix} I_d - R & -\mathbf{c}_d \\ -\mathbf{r}_d^T & 1 - r_{dd} \end{pmatrix} \quad (3.15)$$

Note that (3.14) can be used to prove (3.4) and (3.5), however, the proof given earlier is more interesting as it is based on probabilistic argument only. It is also clear how one can make a "step down" i.e. to find  $(I_d - R)^{-1}$  once  $\left( I_{d+1} - \tilde{R} \right)^{-1}$  is given: determine  $\mathbf{p}^*$  and  $\mathbf{b}^T$  from the last column and row of  $\left( I_{d+1} - \tilde{R} \right)^{-1}$ , respectively, and then subtract  $\mathbf{p}^* \mathbf{b}^T$  from its  $(d \times d)$  principal minor. Therefore, (3.5) enables one to efficiently compute  $\mu$  once  $\tilde{\mu}$  and  $\left( I_{d+1} - \tilde{R} \right)^{-1}$  are given.

#### 4. Sensitivity by $c$ and $k$

In this section we consider the situation when the basic quantities associated with the scheme have been computed and one is interested to examine the effect of increasing  $c$  (or  $k$ ) by  $\delta$ . The transition matrix of the modified scheme can be represented as  $R + E$ , where  $R$  corresponds to the basic scheme and  $E$  introduces changes depending on the parameter being varied. To avoid trivialities, we assume that upper-left elements of both  $R$  and  $R + E$  are less than 1; this implies that ARL's of both schemes are finite. We start by outlining the approach for finding the set of ARL's corresponding to the modified scheme. Denote by  $K$  the matrix  $(I_d - R)^{-1}$  and by  $\tilde{K}$  the matrix  $(I_d - R - E)^{-1}$ . The new set of ARL's is given by

$$\begin{aligned} \bar{\mu} &= \tilde{K} \cdot \mathbf{1} = (I - KE)^{-1} \cdot K\mathbf{1} \equiv \\ &\equiv \left[ I_d + KE + \cdots + (KE)^n + (I - KE)^{-1} (KE)^{n+1} \right] \cdot \mu \end{aligned} \quad (4.1)$$

for every integer  $n$ ; note that under our assumptions the matrices  $I_d - R$ ,  $I_d - R - E$  and, consequently,  $I_d - KE$  are invertible. Since the case of Cusum-Shewhart control schemes the matrix  $E$  is usually "small" (especially for high levels of discretization), (4.1) enables one to evaluate  $\bar{\mu}$  iteratively, starting from  $\mu$ . Denote  $t_n = (KE)^n \mu$ ,  $n = 0, 1, \dots$ , and  $\mu_n = \sum_{i=0}^n t_i$ . Then, at any stage of the iterative procedure, one is able to assess the relative error by using the following

**Statement 4.1.** Suppose that for some  $n$  and  $\varepsilon > 0$ ,  $|Et_n| \leq \varepsilon \cdot \mathbf{1}$  in a componentwise sense.

Then

$$|\bar{\mu} - \mu_n| \leq \varepsilon \bar{\mu} \quad (4.2)$$

**Proof.** Since  $\tilde{K}$  is a fundamental matrix of an absorbing Markov chain (Seneta (1981, p.122)) all its elements are non-negative and, therefore, the function  $\psi(x) = \tilde{K} \cdot x$  is non-decreasing in every component of  $x$ . Thus, by (4.1),

$$|\bar{\mu} - \mu_n| = \|\tilde{K}Et_n\| \leq \varepsilon \|\tilde{K} \cdot \mathbf{1}\| = \varepsilon \cdot \bar{\mu}. \quad (4.3)$$

■

Clearly, (4.2) implies that the absolute error of approximation of  $\bar{\mu}$  by  $\mu_n$  cannot exceed  $|\mu_n \cdot \varepsilon / (1 - \varepsilon^2)|$ .

One should take into consideration that the iterative procedure does not need, in general, to converge. There are several sets of sufficient conditions for convergence; for example, it takes place whenever the norm of  $(KE)$  is less than 1, the absolute value of the dominant eigenvalue of  $(KE)$  is less than 1, etc. Define the norm of a matrix  $A = (a_{ij})$  by  $\|A\| = \max_i \sum_j |a_{ij}|$ . Then a simple criterion for convergence of the iterative procedure can be based on the inequalities  $\|KE\| \leq \|K\| \cdot \|E\|$  and

$$\|K\| = \text{ARL}(0) \leq \left(1 - \sum_j r_{0j}\right)^{-1}, \quad (4.4)$$

where  $ARL(0)$  is the average run length of the basic scheme with headstart 0. To prove (4.4) we first note that since all the elements of  $k$  are non-negative, its norm is equal to the maximal component of  $K \cdot \mathbf{1}$ , i.e. to  $ARL(0)$ . Further, consider the following modification of the Markov chain corresponding to the basic scheme; if the process starts from the headstart 0, it also stays there until absorption occurs. It is clear that average time to absorption of this chain,  $\left(1 - \sum_j r_{0j}\right)^{-1}$ , is greater or equal to  $ARL(0)$ . It is also not hard to show that strict inequality holds in (4.4) provided at least one of the elements  $r_{01}, r_{02}, \dots, r_{0,d-1}$  is positive.

Next we consider two special cases.

a) Sensitivity by the Shewhart's limit,  $c$

In this case the right upper triangular part of the matrix  $R$  consist of zeros (we denote the number of subsequent zeros in its first row by  $i$ ) and  $E$  is of the following form: its right upper  $(i \times i)$  submatrix is  $e \cdot I_i$ , where  $e$  is the probability of one step passage from state 0 to  $i$ ; other elements of  $E$  are zeros. First, we prove that in this case  $\mu_n \rightarrow \bar{\mu}$  as  $n \rightarrow \infty$ . Indeed, if at least one of the elements  $r_{01}, r_{02}, \dots, r_{0,j-1}$  is positive or if  $e < 1 - r_{00}$ , then

$$\|KE\| < \left(1 - \sum_j r_{0j}\right)^{-1} \cdot e \leq 1, \quad (4.5)$$

therefore  $\mu_n \rightarrow \bar{\mu}$ . Otherwise (i.e. when  $R$  is a lower triangular matrix and  $e = 1 - r_{00}$ ), it might happen that  $\|KE\| = 1$ . However, by using inductive argument, it is not difficult to prove that  $(KE)^n \rightarrow 0$  as  $n \rightarrow \infty$ , so that convergence still takes place; the details will be omitted.

Next we discuss a simple direct procedure for finding  $\tilde{K}$ . Denote by  $K_1$  and  $K_2$  the upper left  $(d-i) \times i$  and lower left  $(i \times i)$  minors of  $K$ , respectively. Then direct verification shows that

$$(KE)^n = e^n \begin{pmatrix} 0 & K_1 K_2^{n-1} \\ K_2^n & \end{pmatrix}, \quad (4.6)$$

and, therefore,

$$\tilde{K} = \begin{pmatrix} I_{(d-i) \times (d-i)} & e \cdot K_1 (I_i - e \cdot K_2)^{-1} \\ 0 & (I_i - e \cdot K_2)^{-1} \end{pmatrix} \cdot K; \quad (4.7)$$

note that since  $(e \cdot K_2)^n \rightarrow 0$  as  $n \rightarrow \infty$ , the matrix  $I_i - e \cdot K_2$  is invertible. The formula (4.7) enables one to find  $\bar{\mu}$  directly; it is especially useful in cases where  $i < d$  (in fact, this is true in vast majority of practical cases). One can also see that (4.7) provides an easy way to compute  $K$  on the basis of  $\tilde{K}$  and, consequently, an easy way to evaluate the effect of decreasing  $c$  by  $\delta$ .

**b) Sensitivity by the reference value,  $k$**

In the case when the reference value is increased by  $\delta$ , the matrix  $E$  is given by

$$E = (c_1, c_2 - c_1, \dots, c_d - c_{d-1}), \quad (4.8)$$

where  $c_i$  is the  $i^{\text{th}}$  column of  $R$  ( $i = 0, 1, \dots, d-1$ ). For any vector  $x = (x_0, x_1, \dots, x_{d-1})^T$ , by using summation by parts we obtain

$$Ex = R \cdot (0, x_0 - x_1, \dots, x_{d-2} - x_{d-1})^T + c_d x_{d-1} \quad (4.9)$$

and consequently,

$$KEx = (K - I) \cdot (0, x_0 - x_1, \dots, x_{d-2} - x_{d-1})^T + p^* \cdot x_{d-1} \quad (4.10)$$

As sensitivity analysis by  $h$  is usually performed before analysis by  $k$  (i.e.  $p^*$  is available from the previous computations), the computational effort needed to perform a single step of iterative procedure is essentially equivalent to that needed to multiply a  $(d \times d)$  matrix by a vector. Clearly, the iterative procedure of computing the ARL's corresponding to  $\tilde{R}$  is initiated by assigning the value of  $\mu$  to  $x$ .

## 5. An Example

As an example, let us consider the situation in which the observations are generated by the following mechanism:

$$x_i = \begin{cases} Z_{i1} & \text{with probability 0.5} \\ Z_{i2} & \text{with probability 0.5,} \end{cases}$$

where  $Z_{i1} \sim N(-1.5, 1)$ ,  $Z_{i2} \sim N(1.5, 1)$  are independent random variables. Let the basic scheme be  $h = 3.5$ ,  $k = 1$ ,  $c = 3.5$  and let the level of discretization be  $d = 4$  (thus, in accordance with..., the length of the discretization interval is  $\delta = 1$ ). The matrices  $\hat{R}$  and  $K$  corresponding to this scheme are given by

$$R = \begin{bmatrix} 0.749 & 0.171 & 0.068 & 0 \\ 0.568 & 0.181 & 0.171 & 0.068 \\ 0.432 & 0.136 & 0.181 & 0.171 \\ 0.251 & 0.181 & 0.136 & 0.181 \end{bmatrix}$$

$$K = (I_4 - R)^{-1} = \begin{bmatrix} 26.315 & 6.422 & 3.748 & 1.318 \\ 24.156 & 7.205 & 3.743 & 1.382 \\ 21.451 & 5.523 & 4.387 & 1.377 \\ 16.971 & 4.479 & 2.705 & 2.160 \end{bmatrix};$$

thus, the vector of ARL's corresponding to the headstarts 0, 1, 2, 3 is  $\mu = \{ 37.802, 36.484, 32.737, 26.315 \}^T$ . Further, the eigenvalues of  $R$  are

$$\{ \lambda_0, \dots, \lambda_3 \} = \{ 0.973, 0.010 + 0.053i, 0.010 - 0.053i, 0.301 \}$$

and the corresponding matrix  $U = \{ u_0, \dots, u_3 \}$  is

$$U = \begin{bmatrix} 1.034 & 0.012 - 0.004i & 0.012 + 0.004i & -0.059 \\ 0.995 & -0.040 + 0.039i & -0.040 - 0.039i & 0.085 \\ 0.889 & -0.031 - 0.043i & -0.031 + 0.043i & 0.173 \\ 0.708 & 0.045 + 0.013i & 0.045 - 0.013i & 0.202 \end{bmatrix}$$

First, let us analyze the sensitivity of our control scheme with respect to an increase of  $h$  by  $\delta$ . Clearly, to determine the corresponding matrix  $\tilde{R}$  (see 3.1) we need to compute a single probability (the first component of  $\tilde{r}_4$ ) only. This results in

$$\begin{aligned} r_4^T &= \{ 0.079, 0.171, 0.181, 0.136 \}; \\ c_4^T &= \{ 0, 0, 0.068, 0.171 \}; \quad r_{44} = 0.181. \end{aligned}$$

Subsequently, we find

$$\begin{aligned} (p^*)^T &= K \cdot c_4^T = \{ 0.480, 0.491, 0.534, 0.554 \}, \\ a &= (1 - r_{44} - r_4^T p^*)^{-1} = 0.908, \quad l = 37.701 \end{aligned}$$

and the set of ARL's of the modified scheme:  $\tilde{\mu} = \{ 55.915, 54.999, 52.871, 47.197, 37.701 \}$ . Now, if planned to examine the effect of further increase in  $h$ , the matrix  $(I_5 - \tilde{R})^{-1}$  could be found by first computing  $b^T = \{ 23.698, 6.399, 4.009, 1.687 \}$  (see 3.13) and then applying (3.14).

To compute the new set of eigenvalues by using the equation (3.7), we first find the coefficients  $v_i$  in the representation of  $p^*$  (3.6):

$$\{v_0, \dots, v_3\} = \{ 0.491, 0.650 + 0.107i, 0.650 - 0.107i, 0.746 \},$$

and, subsequently

$$r_d^T C = \{ 0.250, -0.004 - 0.0004i, -0.004 + 0.0004i, 0.051 \}.$$

Now we can solve (3.7) to obtain the eigenvalues of  $\tilde{R}$ .



$$\{\tilde{\lambda}_0, \dots, \tilde{\lambda}_4\} = \{0.982, 0.020 + 0.086i, 0.020 - 0.086i, 0.430, 0.022\}$$

Substitution of these values into (3.8) and normalization by means of the coefficients (3.9) leads to an updated matrix of right eigenvectors (not shown).

Next, let us examine the effect of increasing the Shewhart's limit  $c$  by  $\delta$ . In our case we have  $i = 1$ ,  $e = 0.011$ ,  $K_2 = 16.971$ ,  $(1 - eK_2)^{-1} - 1 = 0.222$ ,  $K\bar{f} = \{26.315, 24.156, 21.451\}$ ; thus, the set of ARL's corresponding to the modified scheme can be found from the relationship  $\bar{\mu} - \mu = \{0.344, 0.316, 0.280, 0.222\}^T \cdot 26.315 = \{9.054, 8.311, 7.380, 5.839\}^T$  (see 4.7). Note that the multiplier 26.315 is the ARL corresponding to the highest headstart of the Page's scheme; one can prove that for the particular type of transition matrices corresponding to such schemes, it is identical to the upper-left element of  $K$ .

Finally, let us perform a sensitivity analysis by the reference value,  $k$ . Computations show that the set of ARL's corresponding to  $k = 2$  is  $\bar{\mu}^T = \{87.9, 87.8, 86.8, 79.6\}$ . After performing the first iteration of the sensitivity analysis procedure (4.8) - (4.10), we obtain  $\mu + \{43.6, 44.0, 42.9, 38.1\} = \{81.4, 80.5, 75.6, 64.4\}$ , which provides a good indication about the consequences of changing  $k$  by 1.

### *6. Sensitivity analysis in terms of the Brownian Motion Approximation*

It is well known (ex. see Bagshaw and Johnson (1975)) that the run length distribution of a Page's scheme can be approximated by the distribution of the time to absorption of a Brownian Motion with absorbing barrier at  $h$  and reflected barrier at 0. In particular, the ARL of such approximation is given by

$$ARL = (h^2/\sigma^2) \times \psi(a), \quad (6.1)$$

where  $a = h\mu/\sigma^2$  ( $\mu$  is the drift of the motion and  $\sigma$  is its standard deviation per unit time; clearly,  $\mu$  approximates the mean of the incoming observations minus reference value,  $k$ , while  $\sigma$  approximates their standard deviation) and

$$\psi(a) = \frac{1}{2a^2} (\exp(-2a) - \exp(-2aw_0) + 2a(1 - w_0)); \quad (6.2)$$

here  $w_0$  is the relative headstart, ex.  $w_0 = 1$  corresponds to the signal level.

Direct differentiation shows that the log-derivatives of the ARL by  $h$  and  $\mu$  are given by

$$\begin{aligned} \frac{ARL'(h)}{ARL} &= \frac{1}{h} \left\{ 2 + a \frac{\psi'(a)}{\psi(a)} \right\} \\ \frac{ARL'(\mu)}{ARL} &= \frac{1}{\mu} a \frac{\psi'(a)}{\psi(a)} \end{aligned} \quad (6.3)$$

The above formulas enable one to assess the effect of varying  $h$  or  $k$  from some fixed values for which the ARL is available. However, this type of sensitivity analysis has several drawbacks. First, the approximation (6.1) usually requires a correction (upwards) in the value of  $h$ , to compensate for the "overshoot" of our discrete scheme over  $h$  at the moment of a signal. In our case, the simplest way to introduce such correction is by equating our ARL to (6.1) and then solve for  $h$ . The second problem is related to our need to know the first two moments of the distribution of incoming observations in order to use (6.3). This problem could be bypassed by equating ARL's corresponding to three different headstarts to (6.1) and solving for the moments. However, it may require more computing than the methods discussed in the previous sections (which are free from the above drawbacks). In general, methods based on Brownian Motion approximations can be of use only when we are dealing with relatively high levels of discretization.

As an example, let us consider the scheme  $h = 3$ ,  $k = 1$ , applied to a sequence of standard normal variables. In this case we have  $ARL = 1958$ ,  $\mu = 0 - 1 = -1$ , and solving (6.1) leads to a different signal level,  $h = 4.14$ . Further, (6.3) result in  $ARL'(h) = 2.004 \cdot ARL = 3924$  and  $ARL'(\mu) = -6.297 \cdot ARL = -12330$ . These results are very close to the actual derivatives of the discrete scheme which are 4000 and -12220, respectively.

## References:

1. Anderson, T. W. (1984), *An Introduction to Multivariate Statistical Analysis*, Wiley, New York.
2. Bagshaw, M. and Johnson, R.A. (1975), *The Effect of Serial Correlation on the Performance of CUSUM TESTS II*, *Technometrics*, 17, pp. 73-80.
3. Bissell, A. (1969), *Cusum Techniques for Quality Control*, *Appl. Stat.*, 18, pp. 1-30.
4. British Standards Institution (1980-1983), *Guide to Data Analysis and Quality Control Using Cusum Techniques. Parts 1 - 4*.
5. Brook, D. and Evans, D.A. (1972), *An Approach to the Probability Distribution of Cusum Run Length*, *Biometrika*, 59, pp. 539-549.
6. van Dobben de Bruyn, D.S. (1968), *Cumulative Sum Tests Theory and Practice*, Hafner Publishing Co., N.Y.
7. Johnson, R.A. and Bagshaw, M. (1974), *The Effect of Serial Correlation on the Performance of CUSUM Tests*, *Technometrics*, 16, pp. 103-122.
8. Lorden, G. (1971), *Procedures for Reacting to a Change in Distribution*, *Ann. Math. Stat.*, 42, pp. 1897-1908.
9. Lucas, J. M. (1982), *Combined Shewhart-Cusum Quality Control Schemes*, *J. Qual. Technol.*, 14 (2), pp. 51-59.
10. Lucas, J. M. and Crosier, R. B. (1982), *Fast Initial Response for Cusum Quality Control Schemes: Give your Cusum a Head Start*, *Technometrics*, 24, pp. 199-205.
11. Lucas, J. M. (1985), *Counted Data Cusum's*, *Technometrics*, 27, pp. 129-144.
12. Page, E. (1954), *Continuous Inspection Schemes*, *Biometrika*, 41, pp. 100-115.
13. Seneta, E. (1981), *Non-negative Matrices and Markov Chains*, Springer - Verlag, New York.
14. Woodall, W. H. (1983), *The Distribution of the Run Length of One-Sided Cusum Procedures for Continuous Random Variables*, *Technometrics*, 25, pp. 295-300.
15. Woodall, W. H. (1984) *On the Markov Chain Approach to the Two-sided CUSUM Procedure*, *Technometrics*, 26, pp. 41-46.
16. Woodall, W. H. (1986), *The Design of Cusum Quality Control Charts*, *J. Qual. Technol.*, 18, No. 2, pp. 99 - 102.
17. Woodward, R. and Goldsmith, P.L. (1964) *Cumulative Sum Techniques*. ICI Monograph No. 3, Oliver and Boyd, London.