Title	CUSUM design for detection of event-rate increases for a Poisson process
Author(s)	Bourke, Patrick D.
Publication date	2021-12-03
Original citation	Bourke, P. D. (2021) 'CUSUM design for detection of event-rate increases for a Poisson process', Quality Engineering, 34(1), pp. 36-51. doi: 10.1080/08982112.2021.1978489
Type of publication	Article (peer-reviewed)
Link to publisher's version	http://dx.doi.org/10.1080/08982112.2021.1978489 Access to the full text of the published version may require a subscription.
Rights	**© 2021, Taylor & Francis Group. This is an Accepted Manuscript of an item published by Taylor & Francis in Quality Engineering on 3 December 2021, available online: https://doi.org/10.1080/08982112.2021.1978489**
Embargo information	Access to this article is restricted until 12 months after publication by request of the publisher.
Embargo lift date	2022-12-03
Item downloaded from	http://hdl.handle.net/10468/13876

CUSUM Design for Detection of Event-rate Increases for a Poisson Process

ABSTRACT

In quantifying the performance of a CUSUM chart for detecting upward shifts in event rate, it has been recommended that steady-state evaluation of performance measures such as ARL be used. In this article, the methodology for making such evaluations using the Markov-chain approach is presented, for the case of an exponential CUSUM. This is much more efficient than the alternative of simulation, which is still in use. It is also shown that if one is using steady-state ARL as a measure of detection performance, one can find better choices for the CUSUM parameter *k* than that provided by the SPRT-based formula. Two types of shift in event rate are considered, and corresponding tables of recommended choices of CUSUM parameters (*k*, *h*) are presented for ten levels of in-control ARL, and for nine sizes of shift. These tables can assist quality engineers in the design of CUSUMs for monitoring inter-event times in steady-state operation. It is also shown that these exponential CUSUM tables may be used to find values for the parameters of a geometric CUSUM or a Bernoulli CUSUM chart for monitoring a proportion, provided the in-control value of the proportion is no more than approximately 0.5%.

Key-words: Statistical process control; Process monitoring; Cumulative sum chart; Exponential CUSUM; Geometric CUSUM; Steady-state ARL; Markov chain

by

Patrick D. Bourke

Affiliation: **Department of Statistics, University College Cork**

Ireland.

July 2021

CUSUM Design for Detection of Event-rate Increases for a Poisson Process

by

Patrick D. Bourke

Section 1: Introduction

This paper is concerned with monitoring inter-event times using a CUSUM chart, so that sudden increases (shifts) in event rate can be quickly detected. A landmark paper is Lorden and Eisenberger (1973), where it was assumed that the times between failures have an exponential distribution, and that following an increase, the failure rate remains at the increased level until the CUSUM generates a signal. In that paper, it was also assumed that when a shift occurs, *it happens at the instant of occurrence of a failure*. This assumption avoids the added complexity that arises when a shift can occur at some independent time-point, unconnected with previous event occurrences. In this paper, both types of shift are considered.

The monitoring of inter-event data using CUSUM charts is a topic of continuing interest to practitioners and researchers in Statistical Process Monitoring, with an extensive literature. There is a broader category of monitoring schemes, now known as "time between events" charts (TBE charts), that includes alternative monitoring statistics, and an increasing range of assumed distributions for the inter-event data. The review paper on monitoring the rate of a rare event (Woodall and Driscoll, 2015) provides a valuable perspective. A difficulty facing non-specialist practitioners wishing to keep up-to-date with current developments, is the task of deciding which of the proposed methods are most useful, given the large number of competing monitoring schemes and the sometimes-overstated claims for superior detection performance. In the case of TBE charts, the review (Ali et al. 2016) contains a listing and commentary on 165 articles and provides useful information. Very few articles in this review have dates prior to year 2000, and contributions by renowned CUSUM-chart innovators such as Page, Johnson, Lorden, Lucas and Yashchin are not mentioned.

Among the many topics in recent research on TBE charts, are: the effects on chart performance of estimation errors for required process parameters (Zhang et al. 2014), charts that use distributions such as the Weibull to model inter-event times (Shafae et al. 2015; Sparks and Hazrati-Marangaloo 2020) and a critical review of the use of Average Run Length (ARL) as a detection-performance measure, with suggestions for alternatives (Woodall and Faltin, 2019; Rizzo et al. 2020).

Following the Lorden-Eisenberger seminal paper on the detection of failure-rate increases, the other landmark papers on the exponential CUSUM chart are Lucas (1985), Vardemann and Ray (1985) and Gan (1994). None of these papers are specific about the way in which a shift in event rate can occur, unlike the Lorden-Eisenberger paper.

In this article, the focus is on the development of Markov-chain-based methodology for evaluating steady-state Average Run Length for an exponential CUSUM chart. The importance of the use of steady-state performance metrics has been stressed in Woodall and Driscoll (2015). Many steady-state ARL evaluations in the general CUSUM literature still rely on simulation. In Schuh et al. (2013), simulation was used to provide steady-state ARL evaluations for an exponential CUSUM. The methodology developed in this paper may be used to provide an efficient means of evaluating steady-state performance for an exponential CUSUM, in contrast to the relative slowness of simulation. As part of this methodology, consideration is given to the way in which a shift in event rate can occur, and two types of shift are considered. The Markov-chain-based methodology developed here may also be applicable to other TBE distributions, such as the Weibull, and this requires further research. Tables are presented to assist quality engineers in choosing exponential CUSUM parameters. It is also shown that these tables can be used to choose the (k, h) parameter-values for a geometric CUSUM chart (and a Bernoulli CUSUM chart) for monitoring a proportion, provided that the in-control value of the proportion is no greater than about 0.5%.

The paper is organized as follows. In Section 2, the details of the exponential CUSUM for detecting an upward shift in event rate are briefly reviewed, together with a discussion of various performance metrics such as ARL and Average Time to Signal (ATS). In Section 3, consideration is given to several ARL-based approaches to the choice of the CUSUM parameters *k* and *h*, including the use of steady-state ARL as a measure of detection performance. That Section also provides discussion on the way in which a shift in event rate can occur. For each of two types of shift mechanism, I explain how to use Markov-chain discretization of the exponential CUSUM (Brook and Evans, 1972) to develop the methodology for steady-state ARL evaluation. Simulation was used throughout the paper to provide a confirmatory check on the Markov-chain-based evaluations of ARL.

Approaches to the choice of the parameter *k* are considered in Section 4. Evaluations of the detection performance of exponential CUSUMs using the Sequential Probability Ratio Test (SPRT) formula for *k* are given, and it is shown in Table 1 that, if one is using steady-state ARL to quantify detection performance, one can find better choices for the parameter *k*.

In Section 5, the principal numerical results of the paper are presented. Two tables are given, and these set forth *recommended parameter-values* for detection of each of nine increased levels of event rate. In the first of these tables it is assumed that a shift can occur independently of previous events at an unknown point in time. In the second table, it is assumed that the shift in event rate accords with that used in Lorden and Eisenberger (1973). The benefit of using the recommended parameter-values rather than the corresponding SPRT-based set is presented as a percentage decrease in out-of-control ARL. Two further tables are provided showing these percentage decreases, and in every one of the 90 cases that were considered for each of the two shift-types, there is a benefit to using the recommended parameter-values. The benefit is greatest (up to 10%) when the supposed shift-size is small, and the specified level for in-control ARL has a low value such as 25.

A CUSUM scheme that is closely related to the exponential CUSUM is the geometric CUSUM, used for monitoring a proportion. In Section 6, it is shown that one can use the two main tables (Tables 2 and 3) provided in this article to choose parameter values for a geometric CUSUM, and hence for a Bernoulli CUSUM. In Section 7, an example is presented of an application using Table 2 to design a CUSUM scheme. The final Section provides a summary and mentions some related investigations.

Section 2: The Exponential CUSUM Scheme and Related Performance Measures

We consider a CUSUM scheme for detecting increases in the occurrence rate (μ) of events such as equipment failures, major storms or disease outbreaks over time. (CUSUM schemes for detecting decreases in event rate may be considered of lesser importance, and a corresponding investigation of such schemes is deferred.) The time-interval (X_i) from one event to the next is recorded, and the following CUSUM statistic (C_i) is updated as each X_i observation becomes available.

$$C_i = \max(0, C_{i-1} + k - X_i) \qquad i = 1, 2 \ldots \tag{1}$$

In starting the CUSUM, an initial value (C_0) must be chosen. This is often taken to be zero, but a head-start or FIR (Fast Initial Response) level (Lucas and Crosier, 1982) may also be used. The quantity k is the *reference value* of the CUSUM and is one of the two parameters of the CUSUM scheme. The second parameter of the CUSUM is the *decision interval* value, denoted by h. When the CUSUM first exceeds the value h, this is referred to as a signal from the CUSUM scheme and is taken to indicate that the event rate may have increased. However false signals are possible. The average number of CUSUM evaluations until a signal, when the initial CUSUM value is set at zero, is termed zero-state ARL. In designing the CUSUM, we want to choose values for k and h to achieve an appropriate balance between the following interacting objectives:

If the event rate remains at an acceptable level (μ_0), we want to restrict the frequency of false signals.

If the event rate shifts upward to a minimally unacceptable level (μ_1), we want the CUSUM to signal as quickly as possible.

The relationship between the Sequential Probability Ratio Test (Wald and Wolfowitz, 1948) and a CUSUM scheme (Page, 1954) is discussed in Hawkins and Ollwell (1998), pages 135-138. This relationship leads to the SPRT-based choice for the parameter k

$$k = [\ln(\mu_1) - \ln(\mu_0)]/(\mu_1 - \mu_0) \quad [2]$$

This choice for k is frequently referred to as having an optimality property, related to a result in Moustakides (1986). Extensive numerical investigations on this were reported in Gan (1994) for the case of zero-state ARL as a measure of detection performance (to be considered later).

The use of the above SPRT formula for choosing k for an exponential CUSUM is frequently recommended and is well-established in the literature. For example, in Zhang et al. (2014) the authors write (page 275) that the exponential CUSUM chart is popular "*because it can be optimized for a given shift*", and the cited reference in support of this is Gan (1994). In Shafae et al. (2015) as part of a summary of the exponential CUSUM chart, the authors write (page 840) "*the reference parameter* k_e *is given by*" and this is followed by the SPRT formula for k. In Schuh et al. (2013), SPRT-based values of k were used in investigations (using simulation) of steady-state ARL performance for exponential CUSUMs. One of the aims of this paper is to demonstrate by counter-example that the optimality associated with the SPRT value for k is not applicable when detection performance is measured using steady-state ARL or FIR ARL. It is explained (pages 46,47) in Yashchin (1993) that this optimality relates to "worst-case ARL". Before further discussion on this matter, we need to consider some performance measures for an exponential CUSUM scheme.

Performance Measures for the Exponential CUSUM: ARL, ANES and ATS

The speed of detecting a signal can be quantified in terms of Average Run Length (ARL), which is the average number of CUSUM evaluations *following some condition* until a signal is generated. Given that the CUSUM is evaluated following each event, an alternative term (in the present context) for ARL could also be used: Average Number of Events to Signal (ANES). We shall consider two types of ARL measure: initial-state ARL and steady-state ARL.

Initial-state ARL (denoted $ARL_{IS}(\mu)$) is the average number of CUSUM evaluations up to the occurrence of a signal, where the initial CUSUM value is set at IS, with the event rate continuing at level μ. Frequently-used values for IS are zero or $h/2$, the latter being the FIR level for the initial CUSUM value, proposed in Lucas and Crosier (1982). We denote the corresponding ARLs by $\text{ARL}_{\text{ZERO}}(\mu)$ and $\text{ARL}_{\text{FIR}}(\mu)$, and point out that, for a specified CUSUM, $\text{ARL}_{\text{ZERO}}(\mu) > \text{ARL}_{\text{FIR}}(\mu)$.

Steady-state ARL (denoted $\text{ARL}_{\text{SS}}(\mu)$) is defined here for an out-of-control condition. We assume that the process has been in stable in-control operation for a considerable period, with the CUSUM being reset to an initial value whenever a false signal occurs. The distribution of CUSUM values will have reached steady state prior to a sudden upward shift in event rate from μ_0 to μ. Then, $\text{ARL}_{\text{SS}}(\mu)$ is the average number of CUSUM evaluations from the moment of shift up to the occurrence of a signal.

A closely-related performance measure is Average Time to Signal (ATS). In the present case, ATS = ARL/μ, bearing in mind that the average value for inter-event time is $(1/\mu)$. The ATS measure is relevant when one needs to compare the detection performance of monitoring schemes which have quite different conditions for generating a signal. This is the case in Schuh et al. (2013) which uses the ATS metric to compare the performances of exponential CUSUMs and Poisson CUSUMs. In the present paper, comparisons of alternative monitoring schemes are confined to TBE CUSUMs, so that it is not necessary to use ATS, although some users of CUSUMs may prefer to work with ATS. The results on detection performance presented below are given in terms of ARL.

It is important to be aware that, for exponential inter-event times with event rate μ, the ARL for a CUSUM scheme with parameters k and h, with initial CUSUM value C_0, is equal to the ARL for a CUSUM with parameters $k\mu$ and $h\mu$, with initial CUSUM value $(C_0)\mu$, when the event rate is 1. Because of this property of an exponential CUSUM scheme, in drawing up tables of recommended parameter values for k and h, we can set the value of the in-control event rate at 1, without loss of generality.

Section 3: Approaches to the Choice of Parameters *k, h*

In the early literature on the design of CUSUM schemes, the recommended approach was to search for (k, h) values that would minimize $\mathrm{ARL_{ZERO}}(\mu_1)$, subject to the requirement that $\mathrm{ARL_{ZERO}}(\mu_0)$ be at least some specified level (100, 150, 200, etc). This was the approach used in Gan (1994). There are some other approaches:

B Find k, h so that $\mathrm{ARL_{FIR}}(\mu_1)$ is minimized, while achieving a required level for $\mathrm{ARL_{FIR}}(\mu_0)$.

C Find k, h so that $\mathrm{ARL_{SS}}(\mu_1)$ is minimized, while achieving a required level for $\mathrm{ARL_{ZERO}}(\mu_0)$.

D Find k, h so that $\mathrm{ARL_{SS}}(\mu_1)$ is minimized, while achieving a required level for $\mathrm{ARL_{FIR}}(\mu_0)$.

It is shown later that for approach B, the SPRT-based choice for k is not optimal. In this paper, we use approach D, because of the benefit of using Fast Initial Response in the operation of a CUSUM. In the case where the observations are intervals in a continuum (such as time intervals, as in this paper), there is another factor that affects the evaluation of ARL: the type of shift mechanism.

Two Types of Shift Mechanism and Evaluation of Steady-state ARL

As mentioned in the Introduction, it was assumed in Lorden and Eisenberger (1973) that a shift in the failure rate could only happen at (or immediately after) the instant when a failure occurred. The wording (page 168) is as follows: "*Note that (2) specifies that the increase in failure rate from λ to $\lambda(1 + \theta)$ occurs instantaneously after X_{m-1} is observed*." The consequence of this assumption is that the stream of inter-event data (X_i) has the following structure:

{*a time-ordered sequence of* $\{X_i\}$ *with event rate* μ_0,

immediately followed by a sequence of $\{X_i\}$ *with event rate* μ_1}

In this paper we shall refer to this type of shift as a Lorden-Eisenberger shift (or L-E shift).

In many of the papers on the monitoring of inter-event data published after Lorden and Eisenberger (1973), there is no discussion of the timing of the shift in event rate. In Gan (1994) there appears to be an implicit assumption that the shift is an L-E shift.

A second type of shift is considered in this paper: the shift could occur independently of previous event occurrences, and we can think of the time-point for a shift as chosen independently and randomly in some specified interval. We refer to this type of shift as an independent random shift. A similar assumption, leading to equivalent post-shift ARL performance, is that the shift occurs at some fixed but unknown future time-point. This gives rise to the following three-component structure for the stream of inter-event data:

A time-ordered sequence $\{X_i\}$ *of inter-event times, with event rate* μ_0.

A single inter-event time Y, the first part of which is due to the rate μ_0 *and the second part due to the rate* μ_1. (*A shift in event rate has given rise to Y.*)

The third component is a time-ordered sequence $\{X_i\}$ *with event rate* μ_1.

The probability distribution of the second component (the composite random variable *Y*) needs to be determined, and the derivation is given in Appendix A. The cumulative distribution function (CDF) of *Y* was found to be as follows:

$$F(y) = 1 + \frac{\mu_0}{\mu_1 - \mu_0} e^{-\mu_1 y} + \frac{\mu_1}{\mu_0 - \mu_1} e^{-\mu_0 y} \qquad [3]$$

subject to $\mu_0 \neq \mu_1$. Other assumed forms for the TBE distribution will require corresponding derivations of the distribution of *Y*, which may be in closed form or in discretized form. The mean value of *Y* is $(1/\mu_0) + (1/\mu_1)$, and thus the presence of *Y* can slow the detection of an increase in event rate. The form of the distribution of *Y* is a consequence of the independent-random-shift assumption, and of course the assumed TBE distribution (exponential). If one considers the incidence of a shift occurring independently and randomly in time, one can realize that this shift is more likely to fall into one of the longer time-intervals between successive events. In the present case, the duration from the event immediately preceding the shift up to the shift itself has an exponential distribution with mean $1/\mu_0$. This has been confirmed by simulation, and a proof is given in Appendix A.

The evaluation of steady-state ARL using Markov-chain analysis for a Poisson process with the L-E shift is relatively straightforward but does not appear in the research literature. First it is necessary to evaluate the steady-state distribution of the discretized non-signalling CUSUM values prior to the shift. This probability distribution is then applied to the appropriate set of ARL values (for various CUSUM initial states) following the shift. The details are given in Appendix B.

When the shift is an independent random shift, there is an additional step in the evaluation of steady-state ARL. As for the case of an L-E shift, we first evaluate the steady-state distribution of discretized CUSUM values for the *completed* inter-event times prior to the shift. Next, we need to accommodate the influence of the random variable Y (see equation [3]) associated with the shift. We construct a transition matrix similar in form to that used in evaluating the steady-state distribution of CUSUM values prior to the shift, but now in constructing this matrix, we replace the CDF of the exponential distribution with the CDF of the random variable Y. This transition matrix is then used to do a one-step revision of the (pre-shift) steady-state distribution of the CUSUM values. This revised distribution is the distribution of non-signalling CUSUM values immediately following the observation of the random variable Y. The revised distribution is then applied to the relevant vector of initial-state ARL values (using the event rate μ_1) to give the steady-state ARL value. The details are given in Appendix C.

Section 4: Some ARL-based Comparisons of Alternative Choices for *k*

In this Section we consider the choice of the parameter k for detecting a specified size of upward shift in event rate. If we had chosen to use zero-state ARL as the performance measure of the CUSUM following the shift, *and if the shift is of type L-E,* the best choice of k would be that given by the SPRT formula, as demonstrated comprehensively in Gan (1994). At the start of the previous Section it was mentioned that for approach B (where $ARL_{FIR}(\mu_1)$ is used as the detection-performance measure) the SPRT-based choice for k is not optimal. This will now be shown, using two CUSUM schemes. For the first scheme, suppose the shift to be detected is from $\mu_0 = 1$ to $\mu_1 = 1.5$, and we require that $ARL_{FIR}(\mu_0 = 1)$ be 50. (The initial state of the CUSUM in all the FIR ARL evaluations in this paper is $h/2$.) A search for the best choice of k (in the sense that $ARL_{FIR}(\mu_1)$ is at a minimum) was conducted, and yielded k = 0.882, h = 4.3594, for which $ARL_{FIR}(\mu_1)$ = 10.814. The SPRT-based values of the CUSUM parameters are k = 0.811, h = 3.3494, for which $ARL_{FIR}(\mu_1)$ = 11.053, which is 2.2% higher than that for k = 0.882. For the second CUSUM scheme, suppose the shift to be detected is from $\mu_0 = 1$ to $\mu_1 = 2$, and we require that $ARL_{FIR}(\mu_0 = 1)$ be 100. A parameter search was again conducted, and yielded k = 0.755, h = 3.5027, for which $ARL_{FIR}(\mu_1)$ = 7.754. The corresponding SPRT-based parameter values are k = 0.693, h = 2.7708, for which $ARL_{FIR}(\mu_1)$ = 7.932, which is 2.3% higher.

Next, it will be shown that, if one is using steady-state ARL as the detection-performance measure, the SPRT-based choice for k can be improved upon. The search methods described later in this Section were used to find "*recommended values*" for k, which depend on the specified size of shift, and on the chosen rate of false alarms. It is of interest to see how the detection performance of a CUSUM with a recommended value for k compares with that of a corresponding CUSUM that uses the SPRT choice for k. Some initial comparisons are presented in Table 1, where the shift-type is "independent random".

Table 1: Six Examples of Comparisons of $ARL_{SS}(\mu_1)$ for Competing CUSUM Schemes.
In each comparison, two matching CUSUMs are given, and steady-state ARLs are evaluated.
The in-control ARLs are FIR, with initial-state = h/2

Detection of a shift from μ = 1 to μ = 1.5

Levels of $ARL_{FIR}(1)$	Recommended choices for k, h			SPRT choice for k			% decrease in ARL for	average % decr. in $ARL_{SS}(\mu)$ for
	k	h	$ARL_{SS}(1.5)$	k	h	$ARL_{SS}(1.5)$	μ = 1.5	μ in [1.25, 1.75]
ARL = 25	1.406	19.3350	10.184	0.811	2.4692	11.377	10.5	11.1
ARL = 100	0.898	6.2618	21.085	0.811	4.3531	21.601	2.4	2.4
ARL = 300	0.859	7.6855	31.935	0.811	6.1425	32.408	1.5	1.7

Detection of a shift from μ = 1 to μ = 2.5

Levels of $ARL_{FIR}(1)$	Recommended choices for k, h			SPRT choice for k			% decrease in ARL for	average % decr. in $ARL_{SS}(\mu)$ for
	k	h	$ARL_{SS}(2.5)$	k	h	$ARL_{SS}(2.5)$	μ = 2.5	μ in [2.25, 2.75]
ARL = 25	0.717	1.8057	6.092	0.611	1.2433	6.159	1.1	1.1
ARL = 100	0.671	2.5511	9.476	0.611	2.0369	9.573	1.0	1.1
ARL = 300	0.650	3.1605	12.532	0.611	2.7087	12.607	0.59	0.65

In Table 1, two sizes of upward shift in event rate, *viz*. from 1 to 1.5, and from 1 to 2.5, were used, together with three levels for initial-state ARL(μ_0), with a FIR setting of $h/2$. Table 1 shows the percentage reduction in $\text{ARL}_{\text{SS}}(\mu_1)$ when the recommended (k, h) is used rather than the SPRT-based choice for k. One can see that the percentage advantage of using the recommended (k, h) is greatest (at 10.5%) for detecting the smaller shift, when the in-control ARL is at 25. The results of 90 such comparisons are provided in Table 4 in the next Section.

The search procedure used to find the recommended choices for (k, h) in Table 1 will now be described for the case of an independent random shift. The search was conducted in two stages (to reduce the total search time). Suppose that the specified value of in-control ARL is 100, and we wish to detect an upward shift from event rate $\mu_0 = 1$ to event rate $\mu_1 = 1.5$. In the first stage of search, I chose a series of "two-decimal-place" values for k (e.g. 0.10, 0.11, 0.12, …,1.2) and for each value of k, I searched for the value of h such that the in-control ARL_{FIR} marginally exceeds 100. For each one of these one hundred and eleven (k, h) pairs that meet the in-control ARL specification, I used the methodology in Appendix C to evaluate the steady-state ARL where the shift to $\mu_1 = 1.5$ had occurred independently at a random point in time. In the first stage of search, the recommended (k, h) choice for the case of ARL = 100 and $\mu_1 = 1.5$ was indicated by the minimum of these steady-state ARL values, and this was $k = 0.90$, $h = 6.318$. In the second stage of search, a series of 21 three-decimal-place values for k (including, and on either side of the k-value arising from the first stage) was investigated in a similar way to yield the recommended (k, h) pair. In Table 1 we see that the recommended choice is $k = 0.898$, $h = 6.2618$, and the value for $\text{ARL}_{\text{SS}}(\mu_1{=}1.5)$ is 21.085. This two-stage search procedure was used in searching for the 90 recommended (k, h) values listed in each of Tables 2 and 3 in a later Section.

The Effect on the Comparisons of Approximate Specification of μ_1

The question may be raised as to the effect on the relative performance of a CUSUM using the recommended choice for k and a CUSUM with the SPRT choice for k, when the shifted event-rate (μ) differs somewhat from the level (μ_1) used in setting up the CUSUMs. To investigate this, the percentage decrease in ARL (when using the recommended choice for k, h rather than the SPRT-based choice) was found for a range of event-rate (μ) levels from 1.25 to 1.75 for the CUSUM schemes in Table 1 with in-control ARL = 100, and μ_1 specified as 1.5. The average decrease was found to be 2.4%. Similar investigations were conducted for the other 5 pairs of matched CUSUMs in Table 1, and the average decreases are reported in the right-hand column of Table 1

To explore this further, the percentage decreases in ARL when using the recommended choices for k, h rather than the SPRT-based choices are presented in two graphs. In Figure 1, for the three comparisons in the top half of Table 1, the percentage decreases in ARL value are displayed for levels of μ ranging from 1.25 to 1.75. For ARL = 100, the improvement is almost 7% when the actual event-rate is smaller at 1.25 than the specified level of 1.5. The improvement drops towards 0 (and below) if the actual event-rate is about 1.63 (or above) rather than 1.5. For ARL = 300, the improvement is about 6.5% when the actual shift is 1.25, and the improvement drops toward 0 if the actual shift is just below 1.57.

For the comparisons in the lower half of Table 1, the improvements in ARL for each shifted event-rate in the range from 2.25 to 2.75 are illustrated in Figure 2. There is a similar pattern to that in Figure 1, but the percentage improvements in ARL are smaller. The two Figures demonstrate the fact that ARLs for event rates that are *below* the specified rate (such as 1.5 in Figure 1, and 2.5 in Figure 2) are even lower for the recommended choices of k than for the SPRT-based choices. However, the SPRT-based choices of k begin to out-perform the recommended choices for k, for values of μ sufficiently *above* the specified rate (e.g above 1.63 in Figure 1 for in-control ARL = 100, and above 2.71 in Figure 2 for in-control ARL = 100).

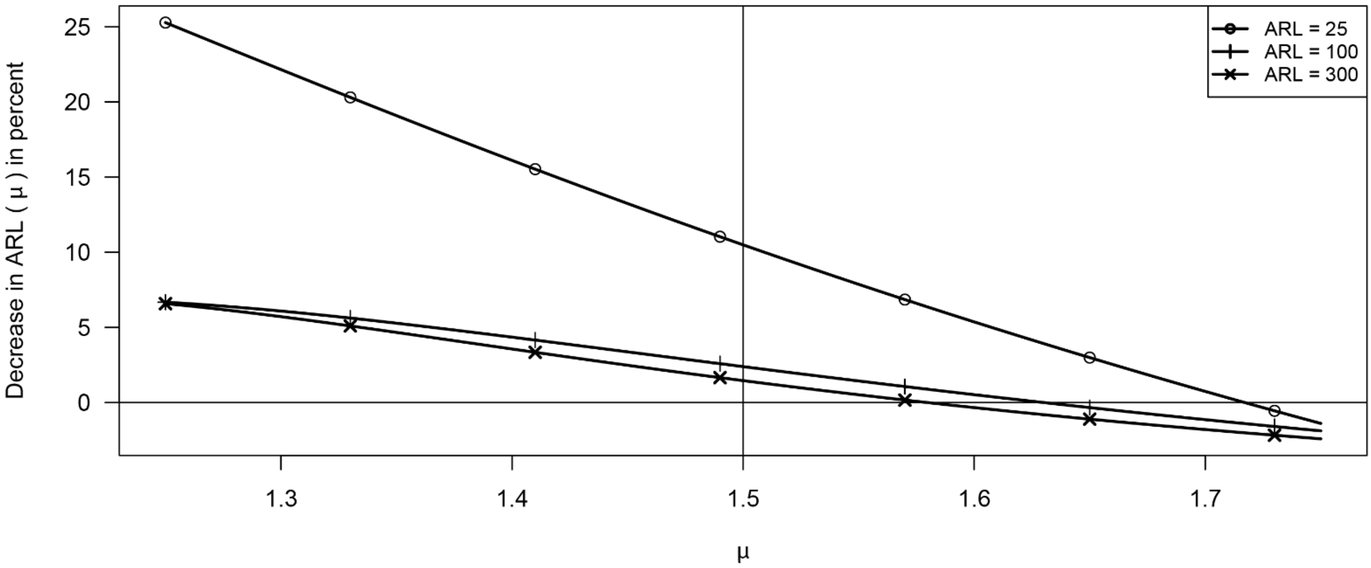

Figure 1: Percentage decrease in ARL(μ) for recommended CUSUM schemes compared to SPRT-based schemes, for a range of values for μ on either side of the target event-rate of 1.5. The schemes being compared are given in Table 1 (top part).

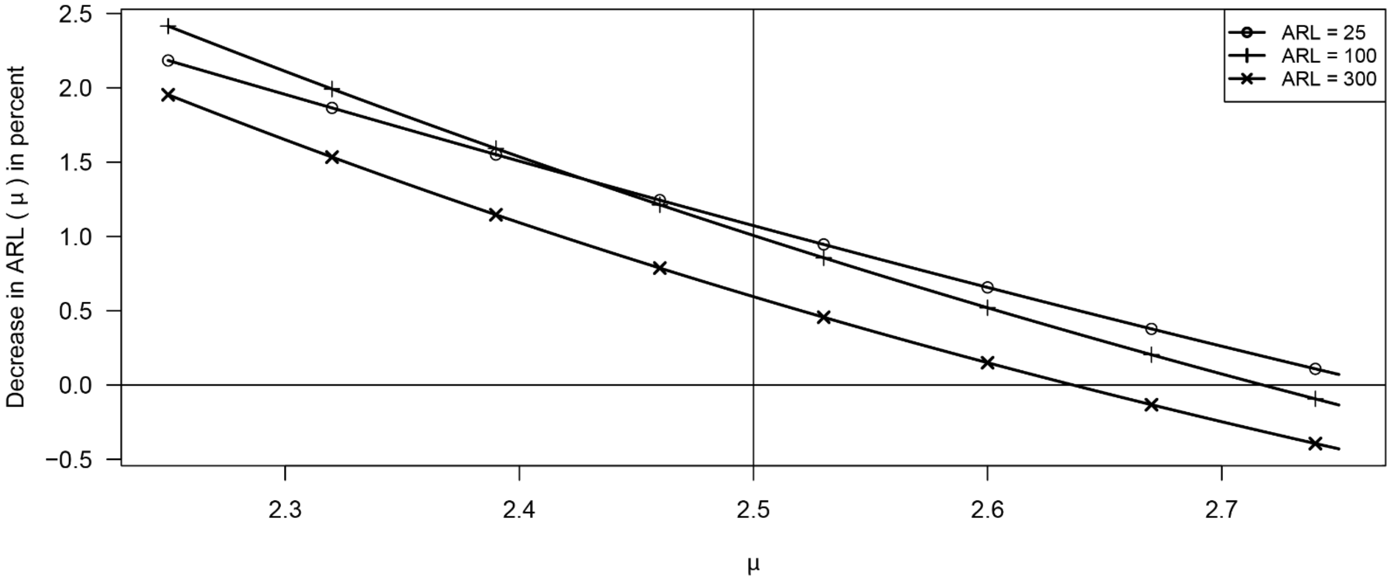

Figure 2: Percentage decrease in ARL(μ) for recommended CUSUM schemes compared to SPRT-based schemes, for a range of values for μ on either side of the target event-rate of 2.5. The schemes being compared are given in Table 1 (bottom part).

Section 5: Tables of Recommended Values for Parameters *k*, *h*

Finding suitable values for the parameters of most CUSUMs is not easy. The designer (or user) of a time-between-events CUSUM needs to have the following information:

The in-control event rate μ_0.

A suitable level for the in-control ARL, so that false signals are acceptably infrequent.

The minimum level (μ_1) of the event rate for which the user would like the CUSUM to signal as quickly as possible.

To provide help to quality engineers (or other users) in choosing values for the parameters of an exponential CUSUM chart, two tables (Table 2 and Table 3) have been developed. Table 2 is relevant if the engineer accepts the assumption that a shift in event rate can occur independently at a random (or unknown) point in time. Table 3 is the corresponding table where the shift can only occur at the instant of an event (i.e. a Lorden-Eisenberger shift).

In developing the Tables, ten levels were chosen for in-control ARL as follows: 25, 50, 100, 150, 200, 300, 500, 700, 1000, 2000. The nine sizes of the upward shift (μ_1) in event rate listed in each of Tables 2 and 3 are 1.5, 2, 2.5, 3, 3.5, 4, 5, 7, 10. (A spreadsheet version of these tables is provided in the online supplemental material.) The search procedure used to find the recommended values for *k* and *h* for each of these 90 scenarios has been described earlier.

The speed of detection for a monitoring scheme is usually quantified in terms of the ARL following a shift, and the values of steady-state ARL(μ_1) are provided in each of Tables 2 and 3. The recommended (k, h) values (for a specified $ARL_{FIR}(\mu_0)$ and shift size) are the same in Tables 2 and 3 for 66 of the 90 cases. Differences mainly occur when $ARL_{FIR}(\mu_0) \leq 50$. In 15 cases, the difference in the values of *k* for corresponding positions in Tables 2 and 3 is 0.001. Thus, in most cases, the same set of (k, h) parameter values is applicable for both types of shift, which is an interesting finding. However, there are important differences for the corresponding $ARL_{SS}(\mu_1)$ entries in Tables 2 and 3. This is because detection is slower for independent random shifts than for L-E shifts.

Table 2: **Recommended Values of CUSUM Parameters (*k*, *h*) for Detection of Each of Nine Upward Shift in Event-rate, for Each of Ten Levels of In-control ARL_{FIR}.**
The Shift is an Independent Random Shift. $ARL_{SS}(\mu_1)$ is included for each shifted event-rate μ_1.

ARL values		Size of shift μ_1								
		1.5	2	2.5	3	3.5	4	5	7	10
25	*k*	1.406	0.862	0.717	0.628	0.603	0.603	0.478	0.376	0.289
	h	19.3350	2.9203	1.8057	1.3245	1.2061	1.2061	0.7710	0.4906	0.2870
	ARL(μ_1)	10.2	7.5	6.1	5.3	4.7	4.4	3.9	3.4	3.1
50	*k*	0.955	0.790	0.696	0.618	0.558	0.511	0.472	0.367	0.285
	h	5.8219	3.1036	2.2218	1.6757	1.3412	1.1149	0.9442	0.6039	0.3902
	ARL(μ_1)	15.4	9.9	7.7	6.5	5.8	5.3	4.7	4.0	3.5
100	*k*	0.898	0.762	0.671	0.605	0.550	0.505	0.437	0.373	0.281
	h	6.2618	3.5977	2.5511	1.9913	1.6104	1.3455	1.0121	0.7453	0.4669
	ARL(μ_1)	21.1	12.6	9.5	7.9	6.9	6.3	5.4	4.6	4.0
150	*k*	0.879	0.750	0.660	0.596	0.544	0.501	0.434	0.347	0.280
	h	6.7135	3.8992	2.7525	2.1511	1.7559	1.4749	1.1092	0.7321	0.5070
	ARL(μ_1)	24.8	14.2	10.6	8.7	7.6	6.9	5.9	4.9	4.2
200	*k*	0.869	0.743	0.656	0.591	0.540	0.498	0.433	0.347	0.280
	h	7.0837	4.1201	2.9267	2.2711	1.8558	1.5647	1.1821	0.7880	0.5376
	ARL(μ_1)	27.7	15.5	11.4	9.3	8.1	7.3	6.2	5.2	4.4
300	*k*	0.859	0.735	0.650	0.585	0.535	0.494	0.430	0.345	0.271
	h	7.6855	4.4436	3.1605	2.4431	1.9975	1.6871	1.2788	0.8531	0.5599
	ARL(μ_1)	31.9	17.3	12.5	10.2	8.8	7.9	6.7	5.5	4.7
500	*k*	0.849	0.728	0.644	0.580	0.529	0.489	0.427	0.343	0.270
	h	8.4672	4.8830	3.4617	2.6753	2.1726	1.8372	1.4016	0.9333	0.6227
	ARL(μ_1)	37.6	19.7	14.0	11.3	9.7	8.6	7.3	6.0	5.0
700	*k*	0.844	0.724	0.640	0.577	0.527	0.486	0.425	0.342	0.270
	h	9.0108	5.1728	3.6524	2.8271	2.3013	1.9360	1.4795	0.9889	0.6628
	ARL(μ_1)	41.6	21.3	15.0	12.0	10.3	9.1	7.7	6.3	5.3
1000	*k*	0.840	0.721	0.637	0.574	0.525	0.484	0.423	0.341	0.269
	h	9.6221	5.4989	3.8672	2.9862	2.4373	2.0480	1.5622	1.0497	0.6994
	ARL(μ_1)	45.9	23.0	16.1	12.8	10.9	9.7	8.1	6.6	5.5
2000	*k*	0.833	0.716	0.633	0.570	0.521	0.481	0.419	0.339	0.267
	h	10.7847	6.1332	4.3030	3.3099	2.6944	2.2694	1.7194	1.1615	0.7708
	ARL(μ_1)	54.5	26.4	18.2	14.4	12.2	10.8	9.0	7.2	6.0

Confirmatory Checks Using Simulation

The $ARL_{SS}(\mu_1)$ values in Table 2 were found using the Markov-chain-based methodology in Appendix C.
For the exponential CUSUM (k = 0.591, h = 2.2711) for detecting a shift from μ = 1 to μ = 3, with $ARL_{FIR}(1)$ = 200 25 million simulation replicates yielded $ARL_{SS}(3)$ = 9.32367, with a standard error (SE) of 0.00071.
The corresponding Markov-chain-based ARL value is 9.32402. The difference in the ARL estimates is less than the SE of the simulation-based estimate. Other confirmatory checks give similar results.
These simulation checks provide reassurance on the correctness of the Markov-chain-based methodology in Appendix C.

Table 3: **Recommended Values of CUSUM Parameters (*k*, *h*) for Detection of Each of Nine Upward Shifts in Event-rate, for Each of Ten Levels of In-control ARL_{FIR}.**
The Shift is a Lorden-Eisenberger Shift. $ARL_{SS}(\mu_1)$ is included for each shifted event-rate μ_1.

ARL values		Size of shift μ_1								
		1.5	2	2.5	3	3.5	4	5	7	10
25	*k*	1.082	0.848	0.718	0.633	0.603	0.565	0.475	0.375	0.291
	h	6.3046	2.7889	1.8118	1.3490	1.2061	1.0568	0.7621	0.4880	0.2896
	ARL(μ_1)	8.7	5.9	4.7	4.0	3.5	3.2	2.8	2.3	2.0
50	*k*	0.947	0.789	0.695	0.619	0.560	0.513	0.472	0.366	0.285
	h	5.6326	3.0925	2.2139	1.6818	1.3514	1.1240	0.9442	0.6011	0.3902
	ARL(μ_1)	13.1	8.1	6.2	5.2	4.5	4.0	3.5	2.9	2.4
100	*k*	0.897	0.762	0.671	0.604	0.550	0.505	0.438	0.373	0.281
	h	6.2341	3.5977	2.5511	1.9838	1.6104	1.3455	1.0166	0.7453	0.4669
	ARL(μ_1)	18.5	10.7	7.9	6.5	5.6	5.0	4.2	3.4	2.9
150	*k*	0.879	0.750	0.660	0.596	0.544	0.501	0.435	0.348	0.279
	h	6.7135	3.8992	2.7525	2.1511	1.7559	1.4749	1.1141	0.7360	0.5042
	ARL(μ_1)	22.2	12.4	9.0	7.3	6.2	5.6	4.7	3.8	3.1
200	*k*	0.869	0.743	0.656	0.591	0.540	0.498	0.433	0.347	0.279
	h	7.0837	4.1201	2.9267	2.2711	1.8558	1.5647	1.1821	0.7880	0.5345
	ARL(μ_1)	25.0	13.6	9.8	7.9	6.7	6.0	5.0	4.0	3.3
300	*k*	0.859	0.735	0.650	0.585	0.535	0.494	0.430	0.346	0.271
	h	7.6855	4.4436	3.1605	2.4431	1.9975	1.6871	1.2788	0.8574	0.5599
	ARL(μ_1)	29.1	15.4	10.9	8.7	7.4	6.6	5.5	4.4	3.6
500	*k*	0.849	0.728	0.644	0.580	0.529	0.489	0.427	0.344	0.270
	h	8.4672	4.8830	3.4617	2.6753	2.1726	1.8372	1.4016	0.9381	0.6227
	ARL(μ_1)	34.8	17.7	12.4	9.8	8.3	7.3	6.1	4.8	3.9
700	*k*	0.844	0.724	0.640	0.577	0.527	0.486	0.425	0.342	0.270
	h	9.0108	5.1728	3.6524	2.8271	2.3014	1.9360	1.4796	0.9889	0.6628
	ARL(μ_1)	38.7	19.3	13.4	10.6	8.9	7.8	6.5	5.1	4.2
1000	*k*	0.840	0.721	0.637	0.574	0.525	0.484	0.423	0.341	0.269
	h	9.6221	5.4989	3.8672	2.9862	2.4373	2.0480	1.5622	1.0497	0.6994
	ARL(μ_1)	42.9	21.0	14.4	11.3	9.5	8.4	6.9	5.4	4.4
2000	*k*	0.833	0.716	0.633	0.570	0.521	0.481	0.419	0.339	0.267
	h	10.7847	6.1332	4.3030	3.3099	2.6944	2.2695	1.7195	1.1615	0.7708
	ARL(μ_1)	51.6	24.4	16.5	12.9	10.8	9.4	7.7	6.1	4.9

Confirmatory Checks Using Simulation

The $ARL_{SS}(\mu_1)$ values in Table 3 were found using the Markov-chain-based methodology given in Appendix B.
For the exponential CUSUM (k = 0.656, h = 2.9267) for detecting a shift from μ = 1 to μ = 2.5, with $ARL_{FIR}(1)$ = 200 25 million simulation replicates yielded $ARL_{SS}(2.5)$ = 9.76686, with a standard error (SE) of 0.00096.
The corresponding Markov-chain-based ARL value is 9.76566. The difference in the ARL estimates is of the same order of magnitude as the SE of the simulation-based estimate. Other confirmatory checks give similar results.
These simulation checks provide reassurance on the correctness of the Markov-chain-based methodology in Appendix B.

Each of the recommended values for k in Tables 2 and 3 exceeds the corresponding SPRT-based choice, with the greatest difference occurring when both the in-control ARL and the shift size to be detected are at their lowest. This helps to explain the sloping shapes in Figures 1 and 2.

Table 4 contains a listing of the 90 percentage decreases in steady-state ARL(μ_1) for the recommended CUSUM schemes in comparison with the corresponding CUSUMs that use the SPRT-based choice of k, for the case of an independent random shift in event rate.

Table 4: Percentages by Which $\mathrm{ARL}_{\mathrm{SS}}(\mu_1)$ for Each CUSUM Scheme in Table 2 is Less than $\mathrm{ARL}_{\mathrm{SS}}(\mu_1)$ for the Matching CUSUM Based on the SPRT Choice of *k*
(for an independent random shift).

$\mathrm{ARL}_{\mathrm{FIR}}(1)$	Sizes of shift μ_1 above 1								
values	1.5	2	2.5	3	3.5	4	5	7	10
25	10.5	2.4	1.1	1.8	2.1	1.6	0.8	0.3	0.4
50	3.4	1.9	1.2	0.9	0.6	0.6	1.1	0.5	0.2
100	2.4	1.3	1.0	0.8	0.6	0.5	0.3	0.6	0.3
150	2.0	1.1	0.8	0.7	0.5	0.4	0.3	0.2	0.3
200	1.7	1.0	0.7	0.6	0.5	0.4	0.3	0.2	0.3
300	1.5	0.8	0.6	0.5	0.4	0.4	0.3	0.2	0.1
500	1.2	0.7	0.5	0.4	0.3	0.3	0.2	0.2	0.1
700	1.0	0.6	0.4	0.4	0.3	0.3	0.2	0.2	0.1
1000	0.9	0.5	0.4	0.3	0.3	0.2	0.2	0.1	0.1
2000	0.6	0.4	0.3	0.2	0.2	0.2	0.1	0.1	0.1

Table 5: Percentages by Which $\mathrm{ARL}_{\mathrm{SS}}(\mu_1)$ for Each CUSUM Scheme in Table 3 is Less than $\mathrm{ARL}_{\mathrm{SS}}(\mu_1)$ for the Matching CUSUM Based on the SPRT Choice of *k*
(for a Lorden-Eisenberger shift).

$\mathrm{ARL}_{\mathrm{FIR}}(1)$	Sizes of shift μ_1 above 1								
values	1.5	2	2.5	3	3.5	4	5	7	10
25	6.4	3.0	1.6	2.3	2.6	1.9	1.1	0.5	0.8
50	4.0	2.4	1.6	1.2	0.8	0.8	1.4	0.7	0.3
100	2.7	1.5	1.2	0.9	0.8	0.7	0.4	0.7	0.4
150	2.2	1.3	0.9	0.8	0.7	0.6	0.4	0.2	0.4
200	2.0	1.1	0.8	0.7	0.6	0.5	0.4	0.2	0.4
300	1.6	0.9	0.7	0.6	0.5	0.4	0.3	0.2	0.1
500	1.3	0.8	0.6	0.5	0.4	0.4	0.3	0.2	0.1
700	1.1	0.6	0.5	0.4	0.3	0.3	0.3	0.2	0.1
1000	0.9	0.6	0.4	0.4	0.3	0.3	0.2	0.2	0.1
2000	0.7	0.4	0.3	0.3	0.2	0.2	0.2	0.2	0.1

The corresponding percentages for the case of an L-E shift in event rate are given in Table 5. In both tables, the larger percentage decreases in steady-state ARL(μ_1) (for the recommended parameter choices compared with the corresponding SPRT-based choices) are when both the in-control ARL is low and the size of shift is low. The average of these percentages for the 42 CUSUM schemes with in-control ARL at most 500, for detecting shifts to an event rate of at most 4, is 1.25% for Table 4 and 1.36% for Table 5.

The Effect on ARL Values When the Time-Between-Events Distribution is Weibull

It is of interest to investigate the case where the stream of inter-event data is from a Weibull distribution, but one is using the values recommended here for k and h, which are based on the assumed exponential distribution. For more detailed investigations of the use of the Weibull as a TBE distribution, one may consult Shafae et al. (2015) or Sparks and Hazrati-Marangaloo (2020). The CDF for the Weibull is given at the end of Appendix A, and there are two parameters: a scale parameter (λ) and a shape parameter (β). As β is varied from say 0.5 to about 4, the shape of the Weibull probability density function changes from a highly-skewed form to an almost symmetric unimodal form, and this flexibility (with just two parameters) is quite useful.

I chose three of the CUSUM schemes listed in Table 2 for this investigation. Scheme A is for detecting a shift from $\mu_0 = 1$ to $\mu_1 = 2$, with $ARL_{FIR}(\mu_0)$ set at 100, scheme B is for detecting a shift from $\mu_0 = 1$ to $\mu_1 = 3$, with $ARL_{FIR}(\mu_0)$ set at 100, while scheme C is for detecting a shift from $\mu_0 = 1$ to $\mu_1 = 2$, with $ARL_{FIR}(\mu_0)$ set at 300. For each CUSUM scheme I evaluated $ARL_{FIR}(\mu_0)$ and $ARL_{SS}(\mu_1)$ using a Weibull distribution as the TBE distribution, for each of 10 values for the shape parameter. The evaluation of $ARL_{FIR}(\mu_0)$ was done using the Markov-chain approach, and these evaluations were checked using simulation. The evaluation of $ARL_{SS}(\mu_1)$ was done using simulation.

To confirm the correctness of various exponential-CUSUM ARL evaluations in this paper, it has been possible to use both the Markov-chain approach and simulation. However, only simulation was used to find $ARL_{SS}(\mu_1)$ when using the Weibull distribution. (Nevertheless, some checking of the correctness of the Weibull simulation-programming for steady-state evaluations was possible, by setting β = 1 and comparing with the corresponding Markov-chain-based evaluations for the exponential CUSUM.)

Table 6: ARL Evaluations for Three CUSUM Schemes Using Weibull Data for a Range of Settings of the Weibull Shape Parameter β.

Weibull	CUSUM Scheme A		CUSUM Scheme B		CUSUM Scheme C	
Parameter	k = 0.762	h = 3.5977	k = 0.605	h = 1.9913	k = 0.735	h = 4.4436
Beta	$ARL_{FIR}(1)$	$ARL_{SS}(2)$	$ARL_{FIR}(1)$	$ARL_{SS}(3)$	$ARL_{FIR}(1)$	$ARL_{SS}(2)$
0.6	19.7	11.3	15.2	6.8	35.1	13.6
0.8	43.7	11.6	37.8	7.4	98.2	15.8
0.9	65.8	12.1	60.9	7.7	169.5	16.6
0.95	81.0	12.3	77.9	7.8	224.8	17.1
1.0	***100.0***	***12.6***	***100.0***	***7.9***	***300.0***	***17.3***
1.1	154.3	12.9	167.4	8.1	546.6	17.8
1.2	241.3	13.3	285.1	8.2	1024.8	18.3
1.4	621.1	13.8	872.9	8.4	3933.4	18.9
1.6	1716.0	14.1	2855.7	8.5	16898.8	19.4
1.8	5090.1	14.4	9900.3	8.6	80479.9	19.7

The results are presented in Table 6, which includes the exponential CUSUM values (when the shape-parameter β = 1). It may be seen that for Weibull TBE distributions where β > 1, the in-control ARL_{FIR} values for each CUSUM can considerably exceed the corresponding value (100 or 300) specified for the exponential CUSUMs. The out-of-control steady-state ARL values differ from the corresponding values (12.6, 7.9 and 17.3) for the three exponential CUSUMs to a much lesser extent. Thus, it appears that the CUSUM schemes tabulated in Table 2 may be useful even when the TBE distribution is Weibull, provided that the Weibull parameter β is not too far above 1. However, if values for β less than 1 are appropriate (where the TBE distribution is well-described by the Weibull), one can see in Table 6 that the in-control ARL values can be considerably smaller than the specified levels 100 and 300, and this is undesirable. If a modelling investigation of the TBE distribution suggested that a Weibull distribution with $\beta < 1$ was appropriate, then the tabulated *k*, *h* values provided in Table 2 should not be used in designing a CUSUM.

Section 6: Using Table 2 to Find the Parameters of a Geometric CUSUM or Bernoulli CUSUM

The geometric CUSUM chart was proposed in Bourke (1991) for detection of shifts in a proportion such as fraction nonconforming. In Bourke (2018) tables are provided to assist quality engineers in finding suitable values for the parameters (*k*, *h*) of a geometric CUSUM (or an equivalent Bernoulli CUSUM) to detect upward shifts that can occur independently in time. These tables refer to 18 in-control values of a proportion in the range 4% to 0.1%, and it was shown that interpolation is quite effective, as well as extrapolation for in-control values of a proportion below 0.1%. However, these tables are limited to five levels of in-control ARL (25, 50, 100, 200, 300) and five sizes of upward shift in the proportion (1.5, 2, 3, 5 and 7 which are multiples of the in-control value of the proportion).

A procedure is next described whereby one can use Table 2 in this article to find parameters for a geometric CUSUM chart for any one of ten levels of in-control ARL, and for nine sizes of shift. For sufficiently small in-control values of the proportion (0.5% or less), the procedure works quite well. An example is presented to describe the procedure.

First, a short description of a geometric CUSUM chart is given. There are in fact two closely-related versions of a geometric CUSUM. In the first version (Bourke, 1991) the observations for the CUSUM are the lengths of successive Conforming Run-lengths (CRLs), *excluding* the nonconforming item that marks the end of each CRL. In the second version, the observation of each CRL-value *includes* the nonconforming item that marks the end of each CRL. For either version of a geometric CUSUM, these CRL-values correspond to the inter-event times for an exponential CUSUM. The expression for a geometric CUSUM has the same form (see equation [1]) as that for an exponential CUSUM. The parameters (*k*, *h*) for a geometric CUSUM are usually taken to be integers, because the process data are counts. For a geometric CUSUM, a signal is given when the

CUSUM value reaches or exceeds h. Denoting the value of the proportion that is to be monitored by p, the average value of a CRL-length is $(1/p) – 1$ for version 1 of a geometric CUSUM, and $1/p$ for version 2. However, for our event-rate CUSUMs with parameter values listed in Table 2, the average value of the in-control inter-event time is 1. In the following example, the presentation is given for version 1 of the geometric CUSUM.

Example: Suppose that the in-control value of the proportion (p_0) is 0.5%, and we wish to design a geometric CUSUM for which the in-control $ANNS_{FIR}$ = 100, to detect a shift from 0.5% to 1%. (It is noted that the terminology ANNS (Average Number of Nonconformities to a Signal) is used in Bourke (2018) and corresponds to ANES or ARL as in this article).

Here, the average value for CRL-length = 199 for a process in control. Referring to Table 2, to achieve in-control ARL_{FIR} = 100, for detecting a shift from event rate 1 to 2, we find the k, h values 0.762, 3.5977. We multiply these values by the mean of the in-control CRL data (199), giving 151.6 and 715.9. We round these to integer values for a geometric CUSUM. (I point out that, in the case of parameter k, it works better to round *down* to the nearest integer for version 1 of the CUSUM). For k = 151 and h = 716, with p = 0.005, it was found that the in-control ANNS (with FIR) is 102.1 which is acceptably close to the specified value of 100. The steady-state ANNS evaluation when p has shifted to 1% is 12.76, while the corresponding entry in Table 2 for steady-state ARL (for the exponential CUSUM) is 12.6. The steady-state ANNS values for the geometric CUSUM were evaluated using the methodology presented in Bourke (2018) and earlier in Bourke (2001).

It is of interest to see how well this method works for a range of levels for p_0. The results of an investigation, using the CUSUM specifications in the example above, are presented in Table 7 for p_0 ranging from 0.02 down to 0.002. This Table also contains the results for corresponding investigations for two further exponential CUSUMs from Table 2:

Table 7: Finding parameter values for a geometric CUSUM chart using an exponential CUSUM. Three exponential CUSUMs, with in-control ARLs at 100, 200, 300, are used to show the influence of the level of p_0 on the value of in-control $\mathrm{ANNS}_{\mathrm{FIR}}$.

	Exponential CUSUM with $\mathrm{ARL}_{\mathrm{FIR}}(1) = 100$, $k = 0.762$, $h = 3.5977$			Exponential CUSUM with $\mathrm{ARL}_{\mathrm{FIR}}(1) = 200$, $k = 0.591$, $h = 2.2711$			Exponential CUSUM with $\mathrm{ARL}_{\mathrm{FIR}}(1) = 300$, $k = 0.735$, $h = 4.4436$		
p_0	k	h	$\mathrm{ANNS}_{\mathrm{FIR}}(1)$	k	h	$\mathrm{ANNS}_{\mathrm{FIR}}(1)$	k	h	$\mathrm{ANNS}_{\mathrm{FIR}}(1)$
0.02	37	176	100.7	28	111	238.2	36	218	276.6
0.01	75	356	101.6	58	225	204.2	72	440	324.9
0.009	83	396	106.7	65	250	191.1	80	489	329.2
0.008	94	446	101.8	73	282	198.4	91	551	294.4
0.007	108	510	98.2	83	322	209.8	104	630	298.8
0.006	126	596	99.7	97	376	209.6	121	736	314.2
0.005	151	716	102.1	117	452	203.6	146	884	299.2
0.004	189	896	102.2	147	566	197.8	183	1106	294.4
0.003	253	1196	99.9	196	755	200.4	244	1477	299.9
0.002	380	1795	99.8	294	1133	203.1	366	2217	304.5

$k = 0.591$, $h = 2.2711$, for detecting a shift from 1 to 3, while achieving in-control FIR ARL at 200.

$k = 0.735$, $h = 4.4436$, for detecting a shift from 1 to 2, while achieving in-control FIR ARL at 300.

The ANNS columns in Table 7 show that, for the range of p_0 levels considered, there is some lack of smoothness in the ANNS evaluations across the p_0 values, and this is more pronounced for p_0 levels above 0.5%. A likely explanation is the fact that the (k,h) parameters for a geometric CUSUM are integers. Referring to Table 7, it may be concluded that the proposed method of finding the parameters of a geometric CUSUM works quite well when p_0 is at most 0.5%, and this is intended as approximate guidance.

In some applications of the geometric CUSUM it is assumed that a shift in the value of the proportion can only occur when a nonconforming item is being produced. This type of shift has been termed a “type 1 shift” (Bourke, 2001) or a “fixed shift”, and this corresponds to a Lorden-Eisenberger shift in event-rate monitoring. For such fixed-shifts processes, Table 3 in this article may be used to find the parameters of a geometric CUSUM, with the proviso that p_0 is at most 0.5%.

Section 7: An Example of the Design of an Exponential CUSUM Using Table 2.

This example is intended to show how Table 2 might be used by a quality engineer in choosing the parameters of a CUSUM to detect a sudden increase in the rate of breakdowns for a machine.

An expensive machine is used in manufacturing a high-volume single product. Based on experience in the past year, it is estimated that, for normal operation, there is an average time of 50 hours between machine breakdowns. When a breakdown occurs, a quick check is conducted on the condition of the machine. If component X has not worn out, the machine is restarted following a basic reset procedure. One could do a full machine-reset (with replacement of some components, including component X) every time the machine breaks down, but the cost of this might exceed the benefit.

The quality engineer wishes to set up a CUSUM scheme, using the times between breakdowns, to indicate whether the breakdown-rate may have increased to an unacceptable level, in which case a full machine-reset should be conducted. The engineer will use a signal from the CUSUM to decide whether to conduct a full machine-reset. The currently-acceptable breakdown-rate is $\mu_0 = 0.02$ breakdowns per hour and let us suppose (for the moment) that the engineer considers that the unacceptable rate is $\mu_1 = 0.04$ per hour. The engineer wishes to find the parameters of a suitable exponential CUSUM scheme. There are ten CUSUM schemes listed in Table 2 for detecting a rate that is double the in-control rate. Throughout Table 2, the in-control rate is 1, but we can switch the CUSUM parameters to an in-control rate of 0.02 per hour by multiplying the tabled (k, h) values by ($1/\mu_0$), which is 50. The engineer may also prefer to use Average Time to Signal (ATS), rather than ARL (using ATS = ARL/μ).

In deciding which of the ten CUSUM schemes might be suitable, the engineer will consider, for each scheme, the in-control value of ARL(μ_0) and the out-of-control value of ARL(μ_1), together with the cost of a full machine-reset, and the cost of delays in detecting a shift, arising from a higher rate of defective product. In quantifying the cost of such delays, the engineer will use the value of

ARL(μ_1). If the engineer chooses a low value for ARL(μ_0), shifts are more likely to be quickly detected, but the frequency of false signals will be higher, so that unnecessary full machine-resets will be more frequent. Correspondingly, if the engineer chooses a large value for ARL(μ_0), such as 1000, there will be fewer false signals, but the average time to signal (ATS) for a real shift will be greater, incurring more cost. The engineer can choose among the ten CUSUM schemes, by quantifying the cost of operating each scheme. The engineer could also consider other possible shift sizes, and for each shift size in Table 2 there are ten possible CUSUM schemes.

Section 8: Summary and Conclusion

The detection of increases in event rate using an exponential CUSUM scheme has been considered for the case of a process that has been operating in-control for a period long enough for the influence of initial CUSUM settings to have faded. A procedure for evaluating steady-state ARL for this CUSUM, using Markov-chain-based discretization, is described. Two types of shift from an in-control event rate to an out-of-control rate are considered: the shift may occur independently of earlier event-occurrences at some random point in time, or the shift may occur randomly in the stream of events, immediately following an event, as assumed in Lorden and Eisenberger (1973).

Numerical searches have been conducted to find recommended values of parameters (k, h) which achieve, for a specified shift-size, the lowest level of out-of-control steady-state ARL, while meeting a specified level for in-control FIR ARL. These investigations have been carried out for ten levels of in-control FIR ARL ranging from 25 to 2000, and for nine sizes of upward shift in event rate. The resulting sets of CUSUM parameter values are here termed “the recommended parameter-values” and are organized into two tables, one for each type of shift. While it is assumed in the article that there is 100% inspection of the event process, the case of sampling inspection can be investigated by adapting the sampling-inspection approach proposed in Bourke (2001) for monitoring a proportion.

Comparisons of the detection performance (using steady-state ARL) of CUSUM schemes that use the SPRT-based choice of k, and the corresponding recommended parameter-values given here, show that the recommended values give an improvement in performance in all 90 cases considered for each type of shift. This improvement is largest for detection of smaller shift-sizes and for lower levels of in-control ARL specification.

A short investigation was conducted to see how the recommended CUSUM schemes perform when the TBE distribution is Weibull rather than exponential, and a range of levels for the shape parameter of the Weibull distribution has been considered. Simulation was used in these steady-state ARL evaluations. One could seek to develop the corresponding Markov-chain-based methodology for evaluation of steady-state ARL when the TBE distribution is Weibull.

The two tables of recommended exponential CUSUM parameter-values may also be used to find the parameters of a geometric CUSUM chart (or a Bernoulli CUSUM chart) for monitoring a proportion. The procedure works quite well provided that the in-control value of the proportion is no greater than approximately 0.5%.

Another area of investigation is the performance of "double-CUSUM" schemes formed by combining two of the listed schemes in either of Tables 2 and 3, with the intention of broadening the range of shift sizes that would be detected quickly. For example, one could operate together the following two exponential CUSUMs from Table 2, each with in-control FIR ARL at 100:

$k = 0.898$, $h = 6.2618$, a CUSUM for detecting a shift from 1 to 1.5

$k = 0.671$, $h = 2.5511$, a CUSUM for detecting a shift from 1 to 2.5

A signal from either CUSUM is taken as a signal from the double-CUSUM scheme.

Simulation was used to evaluate the in-control ARL of this double-CUSUM scheme, and this was found to be 71.2. From other evaluations where the component CUSUMs have in-control ARL at 100, it appears that the in-control ARLs of such double-CUSUM schemes range from about 60 (where the two component-CUSUMs are aimed at detecting shifts to 1.5 and 5) to about 83 (where the two component-CUSUMs are aimed at detecting shifts to 3 and 4). There are clearly many possible double-CUSUM schemes that could be investigated, and several such schemes are presented in Table 8, which may be useful to practitioners. However, the operational complexity of a double-CUSUM scheme (or a triple-CUSUM scheme) is likely to limit their suitability.

Table 8: Simulation estimates of in-control ARL_{FIR} for various double CUSUM schemes.

Single CUSUM Schemes with in-control ARL = 100			Double CUSUM Schemes: Estimates of $ARL_{FIR}(1)$ Each component-CUSUM is indicated by shift-size		
Shift	*k*	*h*	Component CUSUMs		ARL(1)
1.5	0.898	6.2618	1.5	2.5	71.2
2	0.762	3.5977	1.5	3	67.1
2.5	0.671	2.5511	1.5	4	62.5
3	0.605	1.9913	1.5	5	60.2
4	0.505	1.3455	2	3	77.1
5	0.437	1.0121	2	4	69.0
			2	5	64.9
			3	4	83.3
			3	5	75.4

Single CUSUM Schemes with in-control ARL = 500			Double CUSUM Schemes: Estimates of $ARL_{FIR}(1)$ Each component-CUSUM is indicated by shift-size		
Shift	*k*	*h*	Component CUSUMs		ARL(1)
1.5	0.849	8.4672	1.5	2.5	329.3
2	0.728	4.8830	1.5	3	309.0
2.5	0.644	3.4617	1.5	4	289.3
3	0.580	2.6753	1.5	5	279.4
4	0.489	1.8372	2	3	361.6
5	0.427	1.4016	2	4	321.3
			2	5	302.3
			3	4	402.1
			3	5	358.9

Single CUSUM Schemes with in-control ARL = 1000			Double CUSUM Schemes: Estimates of $ARL_{FIR}(1)$ Each component-CUSUM is indicated by shift-size		
Shift	*k*	*h*	Component CUSUMs		ARL(1)
1.5	0.840	9.6221	1.5	2.5	636.3
2	0.721	5.4989	1.5	3	598.1
2.5	0.637	3.8672	1.5	4	561.8
3	0.574	2.9862	1.5	5	544.6
4	0.484	2.0480	2	3	703.8
5	0.423	1.5622	2	4	623.9
			2	5	588.7
			3	4	789.9
			3	5	701.9

REFERENCES

Ali, S., A. Pievatolo, and R. Göb. (2016). An Overview of Control Charts for High-Quality Processes. *Quality and Reliability Engineering International* 32, 2171-2189.

Bourke, P.D. (1991). Detecting a Shift in Fraction Nonconforming Using Run-Length Control Charts with 100% Inspection. *Journal of Quality Technology* 23:3, 225-238, DOI: 10.1080/00224065.1991.11979328

Bourke P.D. (2001). The geometric CUSUM chart with sampling inspection for monitoring fraction defective. *Journal of Applied Statistics* 28(8): 951-972.

Bourke, P.D. (2018). Choosing parameter values for a geometric CUSUM chart for detecting an upward shift in a proportion. *Quality Engineering* 30:4, 621-634, DOI: 10.1080/08982112.2018.1443341

Brook, D., and D.A. Evans. (1972). An Approach to the Probability Distribution of CUSUM Run-Length. *Biometrika* 59, 539-549.

Crozier, R.B. (1986). A New Two-sided Cumulative Sum Quality Control Scheme. *Technometrics* 28, 187-194.

Gan, F.F. (1994). Design of Optimal Exponential CUSUM Quality Control Charts. *Journal of Quality Technology* 26, 109 – 124.

Hawkins, D.M., and D.H. Olwell. (1998). *Cumulative Sum Charts and Charting for Quality Improvement*. New York: Springer-Verlag.

Lorden, G., and I. Eisenberger. (1973). Detection of Failure Rate Increases. *Technometrics* 15, 167-175.

Lucas, J.M., and R.B. Crosier. (1982). Fast Initial Response for Quality-Control Schemes: Give Your CUSUM a Head Start. *Technometrics* 24, 199-205.

Lucas, J.M. (1985). Counted Data CUSUMs. *Technometrics* 27, 129-144.

Lucas, J.M., and M.S. Saccucci. (1990). Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements. *Technometrics* 32, 1-12.

Moustakides, G.V. (1986). Optimal Stopping Times for Detecting Changes in Distributions, *Annals of Statistics* 14, 1379-1387.

Page, E.S. (1954). Continuous Inspection Schemes. *Biometrika* 41, 100-115.

Rizzo, C., S. Chin, S., E. van den Heuvel, and A. Di Bucchianico. (2020). Performance Measures of Discrete and Continuous Time-Between-Events Control Charts. *Quality and Reliability Engineering International* 36, 2754-2768.

Schuh, A., W.H. Woodall, and J.A. Camelio. (2013). The Effect of Aggregating Data When Monitoring a Poisson Process. *Journal of Quality Technology* 45:3, 260-272, DOI: 10.1080/00224065.2013.11917937

Shafae, M.S., R.M. Dickinson, and W.H. Woodall. (2015). Cumulative Sum Control Charts for Monitoring Weibull-Distributed Time Between Events. *Quality and Reliability Engineering International* 31, 839-849.

Sparks, R. S. and H. Hazrati-Marangaloo. (2020). Exponentially Weighted Moving Averages of Counting Processes When the Time between Events Is Weibull Distributed. Book-chapter in *Quality Control - Intelligent Manufacturing, Robust Design and Charts.* IntechOpen. DOI: 10.5772/intechopen.90873

Vardemann, S. and D. Ray. (1985). Average Run Lengths for CUSUM Schemes When Observations Are Exponentially Distributed. *Technometrics* 27, 145-150.

Wald, A. and J. Wolfowitz. (1948). Optimum Character of the Sequential Probability Ratio Test. *The Annals of Mathematical Statistics* 19, 326-339.

Woodall W.H., and A.R. Driscoll. (2015). Some Recent Results on Monitoring the Rate of a Rare Event. In: Knoth S., Schmid W. (eds) Frontiers in Statistical Quality Control 11. Springer. https://doi.org/10.1007/978-3-319-12355-4_2

Woodall, W.H. and F.W. Faltin. (2019). Rethinking Control Chart Design and Evaluation. *Quality Engineering* 31, 596-605.

Yashchin, E. (1993). Statistical Control Schemes: Methods, Applications and Generalizations. *International Statistical Review* 61, 1, 41-66.

Zhang, M., F.M. Megahed, and W.H. Woodall. (2014). Exponential CUSUM Charts with Estimated Control Limits. *Quality and Reliability Engineering International* 30, 275–286.

Appendix A: Derivation of the Distribution of Random Variable *Y* (equation [3])

The random variable Y is a sum of two parts. The first part (U, say) is the time from the event immediately before the shift, up to the time of the shift. The second part (V, say) is the time from the shift up to the first event following the shift, and V has an exponential distribution with mean μ_1. It may not be so obvious that U also has an exponential distribution, and this will now be shown.

Initially, let us assume that the time of shift is fixed at Ts, so that U is the time from the final random event before the shift up to the fixed shift-time Ts. A Poisson process has no direction in time, so that the distribution of U is the same as the distribution from a fixed time-point up to the next event in the process. This latter distribution is well-known to be exponential (in this case with mean μ_0). Next, we remove the assumption that the time of shift (Ts) is fixed and allow it to be random. We consider

$$Prob[U > t] = \int Prob[U > t \mid Ts = s]\, f(s) ds \qquad \text{(A-1)}$$

where $f(s)$ denotes the probability density function (pdf) of the random variable Ts. The conditional probability in equation (A-1) is $\exp(-\mu_0 t)$, which does not depend on s, and thus can be moved outside the integral. Thus, we have

$$Prob[U > t] = \exp(-\mu_0 t) \int f(s) ds = \exp(-\mu_0 t) \qquad \text{(A-2)}$$

This result that U has an exponential distribution has been checked using simulation for both the case of fixed Ts and random Ts.

The probability distribution of Y is found by forming the convolution of two independent exponential random variables (U and V) and is given in equation [3]. The procedure is straightforward, and the steps are as follows. Denote the pdf of V by $g(v)$. By considering

$$Prob[(U+V) \le y] = \int_0^y Prob[U \le (y-v)] g(v) dv \qquad \text{(A-3)}$$

we arrive at

$$Prob[Y \le y] = \int_0^y \{1 - exp[-\mu_0(y-v)]\} g(v) dv \qquad \text{(A-4)}$$

and then

$$= [1 - \exp(-\mu_1 y)] - \{\mu_1 \exp(-\mu_0 y)\} \int_0^y \exp[-(\mu_1 - \mu_o)v] dv \qquad \text{(A-5)}$$

After a few more steps, we find

$$Prob[Y \le y] = 1 + \left[\frac{\mu_0}{\mu_1 - \mu_0}\right] exp(-\mu_1 y) + \left[\frac{\mu_1}{\mu_0 - \mu_1}\right] exp(-\mu_0 y)$$

as given in equation [3]. In the case of $\mu_0 = \mu_1$, the random variable Y has a Gamma distribution.

CDF and Mean of the Weibull Distribution

$$\text{Prob}[X \leq x] = 1 - \exp[\,-(\lambda x)^{\beta}] \quad \text{and} \quad E[X] = (1/\lambda)\Gamma(1+1/\beta) \qquad \text{(A-6)}$$

Appendix B: Evaluation of Steady-state ARL(μ_1) for an Exponential CUSUM when the possible shift in event rate is a Lorden-Eisenberger shift.

Two types of shift in the event rate are considered in this paper. In the landmark paper (Lorden and Eisenberger, 1973), it was assumed that the shift could only occur at the time of an event, and as mentioned in the earlier sub-section in Section 3 (***Two Types of Shift***), this type is termed a Lorden-Eisenberger shift.

Taking account of the methodology in Brook and Evans (1972), we can find the column vector of initial-state ARL values using

$$\mathbf{m} = (\mathbf{I}\text{-}\mathbf{R})^{-1}\mathbf{1}, \qquad \text{(B-1)}$$

where **I** is an identity matrix, **R** is the transition matrix (of size $t \times t$)) for the discretized CUSUM transient states, and **1** is a column vector of ones. (Later we need to refer to the *full transition matrix* **P**, which is the matrix **R** augmented by a final row and column for the signal state.) The zero-state ARL value is given by the first element of **m**, while the middle element provides a FIR ARL value. The procedures described in Crosier (1986) and Lucas and Saccucci (1990) are then adapted to evaluate steady-state ARL, as follows. Using $\mu = \mu_0$, one adds a column vector (**I**-**R**)**1** to the right of matrix **R**, and then a (t+1)-element row-vector **b**

$$\mathbf{b} = (0, 0, 0,.1..,0) \qquad \text{(B-2)}$$

is added to the base of the resulting matrix. The position of “1” in the vector **b** relates to the resetting of the CUSUM following a false signal. This gives a matrix (**P***) that describes the discretized CUSUM-transitions, including the resetting of the CUSUM to h/2 whenever a false signal is produced. (One could also reset the CUSUM to other initial-state levels by placing the 1 in vector **b** in other positions: this has a minor effect on the steady-state ARL evaluation). We can find the steady-state distribution ($\mathbf{p}$) of CUSUM values associated with the matrix $\mathbf{P}^*$, by solving (subject to $\mathbf{p}^T\mathbf{1} = 1$)

$$\mathbf{P}^{*T}\mathbf{p} = \mathbf{p} \tag{B-3}$$

In the next step, the element in $\mathbf{p}$ corresponding to the signal state is removed, and the other elements in $\mathbf{p}$ are inflated so that they add to 1. We now have the steady-state distribution of discretized (transient) CUSUM values prior to the shift, and this is denoted by the column-vector $\mathbf{p}_{ss}$. We can then evaluate the out-of-control steady-state ARL(μ_1) as follows

$$\text{ARL}(\mu_1) = \mathbf{p}_{ss}^T\mathbf{m} \tag{B-4}$$

where $\mathbf{m}$ is given by equation (B-1). Note that in the evaluation of $\mathbf{m}$ to be used in equation (B-4), we use μ_1 the event rate *following* the shift.

Appendix C: Evaluation of Steady-state ARL(μ_1) for an Independent Random Shift

A description of the structure of the stream of inter-event times is given earlier in the sub-section of Section 3 entitled ***Two Types of Shift Mechanism and Evaluation of Steady-state ARL.***

The stream of inter-event times associated with an independent random shift differs from the corresponding stream for an L-E shift because of the presence of the random variable Y (arising from the occurrence of the shift). The procedure for evaluating the steady-state ARL for an independent random shift is the same as for an L-E shift up to the determination of the steady-state distribution $\mathbf{p}_{ss}$. This is the distribution of discretized CUSUM values *just before* the observation of random variable Y. As mentioned earlier, it is then necessary to revise the $\mathbf{p}_{ss}$ distribution to take account of the observation of Y. This is done using a transition matrix $\mathbf{Q}$ associated with Y. The structure of $\mathbf{Q}$ is the same as the full transition matrix $\mathbf{P}$ (explained in Appendix B), except that instead of using the CDF for an exponential distribution, we use the CDF of the random variable Y, given in equation [3]. The revised distribution of CUSUM values following the observation of Y is given by the row-vector

$$\mathbf{r}_{ss} = \mathbf{p}_{ss}^{T}\mathbf{Q} \qquad \text{(C-1)}$$

The steady-state ARL(μ_1) following an independent random shift is evaluated as follows

$$\text{ARL}(\mu_1) = [1 - \textstyle\sum(\mathbf{r}_{ss})_i] + \mathbf{r}_{ss}(\mathbf{m}+1) \qquad \text{(C-2)}$$

where **m** is given by equation (B-1) and in evaluating the **m** to be used in equation (C-2), we use μ_1 the event rate *following* a shift. To clarify the two terms in equation (C-2), we consider the evaluation of the average value of the Number of Events to a Signal (NES). NES can take the value 1 (when *Y* produces a signal) and the probability of this is the term in square brackets in equation (C- 2). The presence of 1 in the term (**m**+1) of equation (C-2) is because of the need to count the event occurring immediately after the observation of random variable *Y*, when *Y* has not produced a signal.

A Note on the size of the matrix R for transient states

The size used for the matrix R was 800×800