Journal of Applied Statistics, Vol. 23, No. 1, 1996, 125–137

CARFAX

Data-analytic aspects of the Shiryayev–Roberts control chart: surveillance of a non-homogeneous Poisson process

RON S. KENETT[1] **& MOSHE POLLAK**[2], [1]*Kenett–Preminger Associates Ltd, Raanana, Israel and* [2]*Department of Statistics, The Hebrew University of Jerusalem, Israel*

SUMMARY *The Shiryayev–Roberts control chart has been proposed as a powerful competitor of the Shewhart control chart and the CUSUM procedure, on theoretical grounds. We demonstrate here the application of a Shiryayev–Roberts control chart to a non-homogeneous Poisson process. We show that, from a data-analytic point of view, the Shiryayev–Roberts surveillance scheme has several advantages over classical CUSUM charts. A case study of power failure times in a computer centre is used to illustrate our main points.*

1 Introduction

The data in Table 1 are times (measured in days) between computer crashes that resulted from power failures, as experienced by a computer centre over a period of 2.3 years. After a crash, the computers are made operational with an 'initial program load', and so we will refer to the data as the IPL data set.

TABLE 1. Number of days between computer crashes that result from power failure at a computer centre

32	10	74	20	5	5	3	4	83	27	11	175	16	11
15	15	121	32	1	22	5	4	53	16	37	3	1	5
11	1	1	11										

The table is read from left to right.

Power failures are potentially very harmful. A computer centre might be able to

Correspondence: Ron S. Kenett, Kenett–Preminger Associates Ltd, PO Box 408, 43101 Raanana, Israel.

0266-4763/96/010125-13 $6.00

tolerate them when they are far enough apart. However, if they become too frequent, one might complain to the power company or decide to invest in an uninterruptable power supply (UPS). The IPL data set was originally analyzed in a UPS cost–benefit study (Kenett, 1981).

Post facto, it seems intuitively clear that, at the end of the period described in Table 1, computer crashes resulting from power failures are more frequent than during the first part of the observation period. Is the variability in failure rates a result of chance alone (common causes) or can it be attributed to special causes that should be investigated? Let us suppose that what the computer centre can tolerate is, at most, an average of one power failure in 3 weeks (21 days). Should the whole sequence described in Table 1 have been observed before deciding that the intensity of power failures has become intolerable? Could this decision have been made earlier? Alternatively, should one wait a little longer before declaring that a change has occurred and see how things develop?

We propose to regard the sequence of power failures as being generated by a non-homogeneous Poisson process with unknown intensity $\theta(t)$. As long as $\theta(t) \leq 1/21$, things are considered to be normal. If $\theta(t)$ exceeds 1/21, then the process is considered to be out of control, with special causes affecting the failure intensity level. One may then demand action to bring the process back in control.

Several previous studies handled the problem of identifying changes in failure intensities by postulating a model on $\theta(t)$ (see Cox & Lewis, 1966; Jelinski & Moranda, 1972; Littlewood & Verrall, 1973; Zacks, 1991). Here, we approach the problem differently. We regard the problem in the context of classical surveillance, amenable to control chart technology. No assumptions are made about the structure of $\theta(t)$.

We monitor the level of $\theta(t)$ by employing a Shiryayev–Roberts (SR) control chart. In the following sections, we briefly review classical surveillance methods and present in detail the less well-known SR scheme. We survey published results that suggest that the SR scheme is at least as efficient as the better-known classical procedures. Moreover, we show that the SR control charts have certain technical advantages over CUSUM and Shewhart charts. We then apply the SR approach to the IPL data of Table 1, pointing out additional, data-analytic advantages of the SR control chart. The terms 'surveillance scheme', 'statistical process control' and 'process monitoring' are used here interchangeably.

2 Basic concepts and notation

The classical change-point problem deals with monitoring a process $\{X_n\}$, $n = 1, 2, \ldots$, for a change in distribution. The process $\{X_n\}$ is observed sequentially; the initial observations have a certain distribution, which may change to another distribution at an unknown point in time, say v. One is interested in detecting such a change quickly, subject to a constraint on false alarms. The problem arises naturally in an industrial context, where the object being monitored is a machine, the quality of whose output is being controlled. The products of the machine yield observations X_n, for which the distribution changes if quality deteriorates or improves. Applications are abundant in a variety of other fields, such as public health (surveillance of congenital malformations, for example), finance (monitoring the stock market), environment (monitoring for changes in pollution level, rainfall, etc.)—wherever surveillance is called for.

Formally, we will let P_v denote the distribution of the process $\{X_n\}$ when the first observation after a change is at time v. The distribution of $\{X_n\}$ when there is no change will be denoted by P_∞, and E_v will denote the expectation under P_v. A surveillance scheme is essentially a stopping time N with respect to the sequence $\{X_n\}$, at which one stops and declares a change to have taken place. N is a random variable, dependent only on observations past to present (inclusive). The standard constraint on false alarms is

$$E_\infty N \geq B \tag{1}$$

where $E_\infty N$ is the average run length (ARL) to false alarm and B reflects the user acceptable ARL. The speed of detection of a surveillance scheme is usually characterized by the average delay from the time of change v to the time of detection N. A formal expression for the average delay (ADEL) is

$$\text{ADEL} = \sup_{1 \leq v < \infty} E_v[N - (v-1) \mid N \geq v] \tag{2}$$

For example, consider the classical two-sided 3σ-limit Shewhart scheme for detecting a change in the mean of a normal distribution. Here, $N = \min\{n \mid |X_n - \mu_0| \geq 3\sigma\}$, where μ_0 is the in-control mean and σ is the standard deviation. Clearly, if there never is a change, then N is a geometrically distributed random variable with parameter $p = 2[1 - \Phi(3)]$. Hence, $E_\infty N = 1/[2(1 - \Phi(3))] \approx 370$. Analogously, for the one-sided 3σ-limit Shewhart scheme for detecting an increase in the mean of a normal distribution, $N = \min\{n \mid X_n - \mu_0 \geq 3\sigma\}$ and N is geometrically distributed with $p = 1 - \Phi(3)$ (when $v = \infty$), so that $E_\infty N = 1/[1 - \Phi(3)] \approx 740$. The average delay, as in equation (2), can be calculated in a similar manner. For example, for the two-sided 3σ-limit Shewhart scheme, if the true change is from a mean μ_0 to a mean $(\mu_0 + \sigma)$, then the average delay is $\text{ADEL} = 1/[\Phi(-4) + 1 - \Phi(2)] \approx 44$.

The classical change-point problem deals with the case where the X_n terms are independent, having density f_0 before a change and density f_1 afterwards. The assumption is that f_0 is known. If f_1 is unknown (as it is in almost all applications), then classical change-point methodology essentially calls for choosing an f_1 value that is representative of possible or likely post-change distributions. The celebrated CUSUM scheme can be shown (Lorden, 1971) to have representation as a maximum likelihood technique for appropriate choice of f_1 as follows.

Let $f_{(v)}(x_1, \ldots, x_n)$ be the joint density of $X_1, \ldots, X_n$ under P_v, i.e.

$$f_{(v)}(x_1, \ldots, x_n) = f_0(x_1) f_0(x_2) \ldots f_0(x_{v-1}) f_1(x_v) \ldots f_1(x_n)$$

With this notation, the simple CUSUM scheme uses the statistic

$$\begin{aligned} &\max_{1 \leq k \leq n} f_{(k)}(X_1, \ldots, X_n)/f_{(\infty)}(X_1, \ldots, X_n) \\ &= \max_{1 \leq k \leq n} f_1(X_k) f_1(X_{k+1}) \ldots f_1(X_n)/[f_0(X_{k+1}) \ldots f_0(X_n)] \end{aligned} \tag{3}$$

and declares at

$$M_A = \min\{n \mid \max_{1 \leq k \leq n} f_{(k)}(X_1, \ldots, X_n)/f_{(\infty)}(X_1, \ldots, X_n) \geq A\} \tag{4}$$

that a change has occurred.

Recently, there has been renewed interest in the SR approach. This method uses the statistic

$$R_n = \sum_{k=1}^{n} \frac{f_{(k)}(X_1, \ldots, X_n)}{f_{(\infty)}(X_1, \ldots, X_n)} \tag{5}$$

and declares at

$$N_A = \min\{n \mid R_n \geq A\} \tag{6}$$

that a change is in effect (cf. Shiryayev, 1963a; Roberts, 1966).

3 Comparison of surveillance schemes—a review

The most popular and, apparently, the easiest surveillance scheme to apply is the Shewhart control chart. The disadvantage of the Shewhart control scheme is that it is not sensitive to small or medium-sized changes. CUSUM schemes were invented to overcome this difficulty (Page, 1954).

CUSUM schemes possess optimality properties (Lorden, 1971; Moustakides, 1986; Ritov, 1990). It can be shown that, if B is large and the post-change distribution is f_1, then

$$\text{ADEL} \approx (\log \text{B})/E_\infty \log\left[\frac{f_1(X)}{f_0(X)}\right]$$

when the observations are independent. For example, if $B = 370$, the observations are normal and the change is a shift in the mean by one standard deviation, then this expression for ADEL is approximately equal to 10. (It should be recalled that, for the corresponding Shewhart scheme, ADEL $\approx$ 44.)

SR schemes have also been shown to possess optimality properties (Pollak, 1985; Yakir, 1994). In terms of speed of detection, the CUSUM and SR schemes are generally comparable (Shiryayev, 1963b; Roberts, 1966; Pollak & Siegmund, 1985, 1991; Mevorach & Pollak, 1991).

The main advantage of the SR over the CUSUM method is in the relative ease of its application under minimal asumptions. To demonstrate this advantage, we compare the properties of M_A and N_A that correspond to the stopping times under the CUSUM and SR schemes respectively. To obtain the levels of A in both schemes set to satisfy equation (1), the expectations $E_\infty M_A$ and $E_\infty N_A$ must be computed, or at least approximated. Evaluation of $E_\infty M_A$ is possible when the observations are independent, so that the sequence of CUSUM statistics forms a Markov chain. The Markov property is lost when the observations are not independent. In contrast, the SR approach does not require the observations to be independent.

To demonstrate the importance of this, consider surveillance for a change of a normal mean when the initial mean and the standard deviation are unknown. Since f_0 is unknown, none of the classical schemes—neither Shewhart, CUSUM nor SR—can be applied directly to the sequence $\{X_n\}$. In practice, one may try to obtain a learning sample to estimate the nuisance parameters. However, because of the extreme sensitivity of these schemes to misspecification of parameters (see van Dobben de Bruyn, 1968, Section 2.3), the learning sample must be very large for the schemes to be valid and/or efficient. To circumvent this, define $Y_i = (X_i - X_1)/(X_2 - X_1)$ for $i = 3, 4, 5, \ldots$, and base the surveillance scheme on the sequence $\{Y_n\}$. Subtracting X_1 from X_i eliminates the problem of the unknown initial mean, and dividing by $(X_2 - X_1)$ gets rid of the unknown standard deviation (assuming $v \geq 3$). Hence, $f_{(\infty)}$ of the $\{Y_n\}$ process is known. In principle, this should be enough to evaluate $E_\infty M_A$ and $E_\infty N_A$. However, there is a price paid for getting rid of the nuisance parameters: the Y_i terms are dependent. For the CUSUM approach, this causes a serious difficulty. However, for the SR method, as we shall now show, it does not. A complete working out of the normal mean problem is presented in Pollak and Siegmund (1991) and Croarkin *et al.* (1993). For non-parametric surveillance problems handled by a similar approach, see Gordon and Pollak (1990, 1994, 1995) and Bell *et al.* (1994).

Evaluation of $E_\infty N_A$ relies on a martingale argument that is valid even when the observations are dependent. Let

$$\Lambda_{k,n} = f_{(k)}(X_1, \ldots, X_n)/f_{(\infty)}(X_1, \ldots, X_n)$$

Under P_∞, every likelihood ratio sequence $\{\Lambda_{k,n}\}$, $n = 1, 2, \ldots$, is a martingale with unit expectation. Therefore, we have that

$$R_n - n = \sum_{k=1}^{n} (\Lambda_{k,n} - 1)$$

is a martingale with zero expectation. By the optional sampling theorem, $E_\infty(R_{N_A} - N_A) = 0$. Hence, $E_\infty N_A = E_\infty R_{N_A}$. By its definition, $R_{N_A} \geq A$, so that $E_\infty N_A \geq A$. Since N_A is the first time that R_n exceeds A, the excess is typically not great, so that setting $A = B$ yields a moderately conservative procedure that satisfies equation (1). Often, the existence of a constant $C > 1$ such that $E_\infty N_{\mathrm{A}} \approx CA$ can be shown and its value computed (Pollak, 1987), so that $A = B/C$. (None the less, there is a way to handle the CUSUM problem with nuisance parameters (see Siegmund & Venkatraman, 1995).)

To summarize, the SR approach generally detects a change as fast as the CUSUM method and, in problems with a complicated parameter structure, it is often easier to apply. For this reason, data-analytic aspects of the SR methodology merit attention.

In the next sections, we describe aspects of the SR approach in the context of the surveillance of a non-homogeneous Poisson process. We find that, from a data-analytic perspective, the SR control scheme enjoys certain advantages over the CUSUM approach. Such advantages, although detailed here for the surveillance of a non-homogeneous Poisson process, hold for surveillance problems in general.

4 SR surveillance of a non-homogeneous Poisson process: theoretical details

Consider a situation where one observes a non-homogeneous Poisson process with unknown intensity $\theta(t)$. While the precise value of $\theta(t)$ and even its functional form are not always of direct interest, it may be meaningful to know whether or not $\theta(t)$ has surpassed a certain level. For example, the chronological sequence of defective items in the output of an industrial process or time between failures (such as the IPL data of Table 1) can often be viewed as a non-homogeneous Poisson process. If the intensity exceeds a certain level (1/21 in Section 1), then the process is considered to be out of control. In another example, the series of software failures (or 'bugs') encountered in developing a software program can be regarded as a non-homogeneous Poisson process (see, for instance, Kenett & Pollak, 1986). The intensity of software failures changes as a result of debugging attempts. One is interested in knowing when the program is 'reasonably good', so that it can be released for operational use—something which can be formulated as a low enough level of $\theta(t)$. Thus, the fact that $\theta(t)$ is too large or too small has significance. One may want to stop the process and bring it under control or, when this is not possible—such as when conducting a surveillance of congenital malformations in newborn infants—one may want to start an investigation or take other action.

Suppose that $0 < w_0 < \infty$ is a value of particular interest, in the sense that, after $\theta(t)$ crosses w_0, one should be made aware of this as quickly as possible. Suppose that this occurs at an unknown time $0 \le v$. We will consider two cases:

Case (i) : $\theta(t) \le w_0$, for $0 \le t < v$

Case (ii): $\theta(t) \ge w_0$, for $0 \le t < v$

In both cases, w_0 is assumed to be known and v unknown. Case (i) corresponds to surveillance for an increase in the output of defectives in an industrial process. Case (ii) corresponds to a reliability growth problem, such as that associated with improving a software program.

The problem being considered is one of continuous time. To set up SR schemes, we follow the discrete time analog of the problem, as described in Sections 2 and 3. We construct schemes where a value w is set to represent the intensity of the process after a change occurs. Practical methods for choosing w will be discussed in the next section. Meanwhile, it suffices to say that $w > w_0$ in Case (i) and $w < w_0$ in Case (ii).

Intuitively, discrimination between pre-change and post-change situations will be most difficult if $\theta(t) \equiv w_0$ before the change, instead of $\theta(t) \le w_0$ in Case (i) and $\theta(t) \ge w_0$ in Case (ii). Therefore, we construct the surveillance schemes with this more difficult case in mind. In other words, the SR schemes will be designed with an equality in equation (1) if $\theta(t) \equiv w_0$ before the change. Thus, if $\theta(t)$ is not identically equal to w_0 before the change, then equation (1) will be satisfied with an inequality. A formal proof of this is given in Appendix B. To summarize, the (simple) SR scheme will be constructed for the case where $\theta(t) \equiv w_0$ for $0 \le t < v$ and $\theta(t) \equiv w$ for $v \le t < \infty$. However, the scheme may be legitimately applied even if this is not the exact structure of $\theta(t)$.

Let $X_1, X_1 + X_2, X_1 + X_2 + X_3, \ldots$ be the jump times of the process. Then, let

$$n(t) = \max\left\{n \mid n \ge 0, \sum_{i=1}^{n} X_i \le t\right\}, \qquad \text{where } \sum_{i=1}^{0} X_i = 0$$

To set up a SR procedure, one first has to compute $\Lambda_{s,t}$, i.e. the ratio of the likelihood of the process observed until time t when $v = s$, to the likelihood when $v = \infty$.

Consider first the denominator of $\Lambda_{s,t}$, which is the likelihood of $X_1, X_1 + X_2, \ldots, \sum_{i=1}^{n(t)} X_i, n(t)$ when $v = \infty$. This is the joint density of $X_1, X_2, \ldots, X_{n(t)}$ multiplied by the probability that the last $(t - \sum_{i=1}^{n(t)} X_i)$ time units elapsed without a failure, which is

$$w_0 \exp(-w_0 X_1)\, w_0 \exp(-w_0 X_2) \ldots w_0 \exp(-w_0 X_{n(t)}) \exp\left[-w_0\left(t - \sum_{i=1}^{n(t)} X_i\right)\right]$$
$$= w_0^{n(t)} \exp(-w_0 t)$$

Likewise, the numerator of $\Lambda_{s,t}$ is the product of the P_∞ (likelihood of the process observed in $[0, s)$) multiplied by the P_s (likelihood of the process observed in $[s, t)$), which is (by similar considerations) given by

$$w_0^{n(s)} \exp(-w_0 s)\, w^{n(t)-n(s)} \exp[-w(t-s)]$$

Hence, we have

$$\Lambda_{s,t} = \left(\frac{w}{w_0}\right)^{n(t)-n(s)} \exp[(w_0 - w)(t-s)]$$

In discrete time, R_n of equation (5) was obtained by summing $\Lambda_{k,n}$ over k. The analog in continuous time is obtained by integrating $\Lambda_{s,t}$ over s. We denote the analogous statistic by $R_w(t) = {}_0\int^t \Lambda_{s,t}\,ds$. A straightforward calculation yields (see Appendix A)

$$R_w(t) = \frac{1}{w_0}\left\{\frac{w_0}{w_0 - w}\left(\frac{w}{w_0}\right)^{n(t)} \exp\left[(w_0 - w)t\right]\right.$$

$$\left. + \sum_{j=1}^{n(t)} \left(\frac{w}{w_0}\right)^{n(t)-j} \exp\left[(w_0 - w)\left(t - \sum_{i=1}^{j} X_i\right)\right] - \frac{w_0}{w_0 - w}\right\}$$

By analogy to equation (6), we set

$$T_{A,w} = \min\{t \mid R_w(t) \geq A\}$$

To implement these procedures, an evaluation of $E_\infty T_{A,w}$ is required. This will provide a choice of A that satisfies equation (1). The numerical algorithms for this procedure are supplied in Theorem 1 below, the proof of which is outlined in Appendix B.

Theorem 1

(i) If $\theta(t) \leq w_0$ for all $t \geq 0$, then
$E_\infty T_{A,w} \geq E_\infty(T_{A,w} \mid \theta(t) = w_0$ for all $t \geq 0)$
For all $A > 0$, we have
$E_\infty T_{A,w} > A$
$E_\infty(T_{A,w} \mid \theta(t) = w_0$ for all $t \geq 0) = AC_w[1 + o(1)]$
where $o(1) \to 0$ as $A \to \infty$ and

$$C_w = \frac{w \log(w/w_0) - w + w_0}{w - w_0 - w_0 \log(w/w_0)}$$

(ii) If $\theta(t) \geq w_0$ for all $t \geq 0$, then
$E_\infty T_{A,w} \geq A$

with equality holding if (and essentially only if) $\theta(t) = w_0$ for all $t \geq 0$.

Basically, Theorem 1 states that setting $A = B$ always causes equation (1) to be satisfied, although the inequality may be sharp in Case (i), making the procedure somewhat conservative. One can be less conservative when B is not small, by setting $A = B/C$, where C is a constant given in Theorem 1 ($C = C_w$ in Case (i) and $C = 1$ in Case (ii)).

5 SR surveillance of a non-homogeneous Poisson process: practical considerations

Consider the classical change-point problem described in Section 1. Obviously, if the post-change density is known, then this known value should be the 'representative f_1'. If it is not known, then one can specify a representative f_1 and employ a SR stopping rule. A typical SR control chart form is shown in Fig. 1.

When monitoring a non-homogeneous Poisson process, the statistics $R_w(t)$ are charted as a function of time t. As long as $R_w(t)$ remains below A, no alarm is

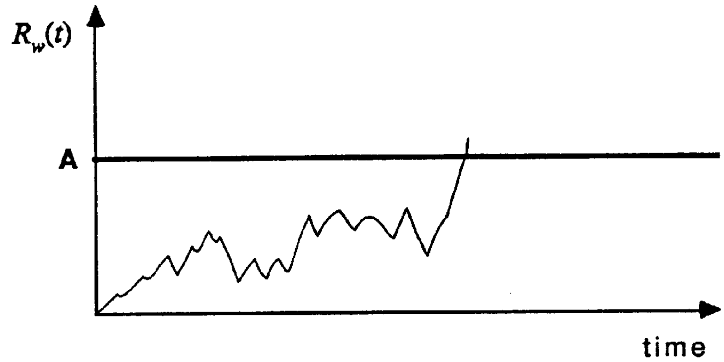

FIG. 1. Typical SR control chart.

raised and monitoring is continued. An alarm is raised the first time that $R_w(t)$ crosses A. The following example will illustrate some of the intricacies involved in the choice of w.

Consider the IPL data of Table 1. Returning to the scenario of Section 1, we take $w_0 = 1/21$. What should w be? (Recall that w is set before viewing the sequence of observations, so that Table 1 cannot be used to choose w.) If there is previous experience with power failures, it may be used to choose w. If not—as in our case—then one can set w as a typical value of the conjectured post-change values of $\theta(t)$. Alternatively, we can set w such that, if $\theta(t)$ reaches w, then we would clearly want to raise an alarm. (For example, $w = (1.1)/21$ may not be a good choice, because the limit of one power failure in 3 weeks is not greatly exceeded.) For a first example, take $w = 2/21$. (This means that the power failure rate is two in 3 weeks—double the maximum acceptable rate.) The graph of $R_{2/21}(t)$ is charted in Fig. 2. By Theorem 1, $C_{2/21} = 1.259$. Therefore, to satisfy equation (1) with $B = 370$, one may take $A = B/1.259 = 294$; for $B = 740$, take $A = 588$. Here, $T_{588,\,2/21} = 154$ and $T_{588,\,2/21} = 823$. Clearly, the relatively frequent failures between $t = 141$ and $t = 158$ are not detected as an out-of-control state when $B = 740$, although an alarm would be raised (at $t = 154$) for $B = 370$.

The value of $R_{2/21}(t)$ at $t = 158$ is $R_{2/21}(158) = 509.1$. This means that the largest value of B that would cause an alarm to be raised at $t = 158$ is $509.1 \times 1.259 = 641$. In other words, a scheme with an ARL to false alarm of 641 days would still have raised an alarm at $t = 158$. Thus, the value 641 plays a role analogous to a P-value in hypothesis testing.

The same can be carried out for any other t. The value of the highest B that would still have caused an alarm to be raised at t is $1.259R_{2/21}(t)$. In other words, the height of $R_{2/21}(t)$ (up to a multiplicative constant) has a meaning analogous to that of a P-value. As another example, let us consider $R_{2/21}(835) = 2080.6$. The B value—the highest value of B that would have caused an alarm to be raised at 835—is $1.259 \times 2080.6 = 2607$. Clearly, at $t = 835$, the evidence that the process has gone out of control by $t = 835$ is considerably stronger than the evidence at $t = 158$ (that the process has gone out of control by $t = 158$). The data-analytic interpretation of the height of the SR statistic is unique to the SR scheme; for the CUSUM method, the value of the threshold A that corresponds to B (such that one has equality in equation (1)) is not linear in B, making a naked-eye interpretation of the present height of the control statistic more difficult.

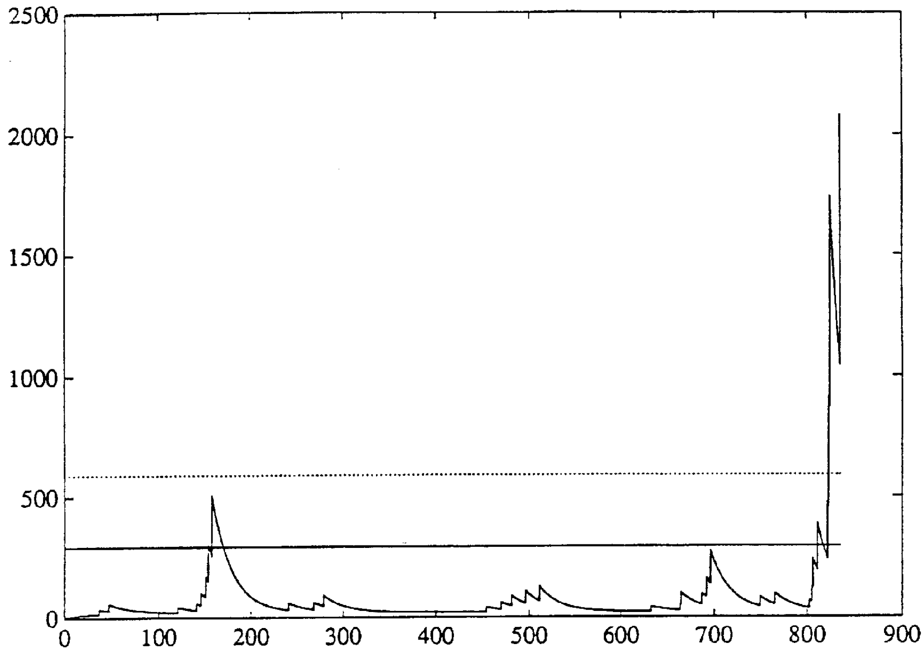

FIG. 2. SR control chart for $R_{2/21}(t)$.

To evaluate our previous choice of $w = 2/21$, we repeat the same type of analysis for other values of w. The results are summarized in Table 2.

TABLE 2. SR stopping times and largest values of B that would have given rise to an alarm, for various values of w

w	$T_{A=370/C_w, w}$	$T_{A=740/C_w, w}$	Largest B for alarm: In first part of series	Largest B for alarm: Anywhere in series
1.5/21	158	823	382	1 467
2/21	154	823	641	2 607
2.5/21	154	158	930	4 036
3/21	154	158	1 209	5 683
3.5/21	154	158	1 442	7 568
4/21	154	158	1 615	9 807
4.5/21	154	158	1 727	11 799
5/21	154	158	1 782	13 326
5.5/21	154	154	1 787	14 270
6/21	154	154	1 751	14 615
6.5/21	154	158	1 682	14 421
7/21	154	158	1 589	13 796
8/21	154	158	1 363	11 764
9/21	154	158	1 125	9 434
10/21	154	158	906	7 380
11/21	154	823	720	5 837
12/21	158	824	570	4 811
13/21	158	824	453	4 207
14/21	823	824	363	3 894

For $B = 370$, similarly to $T_{294, 2/21}$, most stopping times occur at $t = 154$. For $B = 740$, many stopping times catch the rise in intensity of power failures in the period $141 \leq t \leq 158$, while $T_{588, 2/21}$ does not. In fact, when $w = 5.5/21$ and $6/21$, this increase is even detected at $t = 154$. It seems that our choice of $w = 2$ was poor judgement. This becomes even more apparent if we increase B: with $w \approx 5/21$ or $6/21$, we even detect the 141–158 rise with $B = 1750$. The same phenomenon occurs if we further raise B: while nothing would have been detected for $B > 2607$ with $w = 2/21$, with $w \approx 6$, a change would have been detected even for $B = 14\,000$. It seems that $w \approx 5/21$ or $6/21$ would have been the best choice. (Of course, initially we could not have known that.)

6 Concluding remarks

(1) Although the flavour of this paper is not Bayesian, we would like to conclude with a Bayesian remark. In a discrete time problem, for small values of p of a geometric (P) prior on v, the posterior probability of a change is

$$P(v \leq n \mid X_1, \ldots, X_n) \approx R_n/(R_n + 1/p)$$

Likewise, in the continuous time case, if the prior on v is $\exp(\eta)$ and η is small, then the posterior probability of a change is

$$P(v \leq n \mid n(t), X_1, \ldots, X_{n(t)}) \approx R(t)/[R(t) + 1/\eta]$$

Hence, if the representative of the post-change process is correct, then the value of $R(t)$ can be given a direct interpretation in terms of the posterior probability that a change has taken place.

As an example, consider the IPL data with $w_0 = 1/21$ and $w = 6/21$. Suppose that it were reasonable to expect an increase in power failures of about one per year, i.e. $\eta = 1/365$. Since $R_{6/21}(158) = 976.8$, the posterior probability of a change being in effect at $t = 158$ is about $8153.6/(8153.6 + 365) = 96\%$. (If $\theta(t) \leqslant 1/21$ for $t < v$ (instead of $\theta(t) = 1/21$), then these probabilities will be lower bounds.) These posterior probabilities provide excellent diagnostics for decision-making.

(2) Shewhart emphasized the theory of probability as the tool of the statistician. The Shewhart control charts are a prime example of applied statistics technology that combines properties derived from the theory of probability and efficiency in application. None the less, when small-to-medium-sized changes are expected in a process, the standard Shewhart charts will not be quick to flag the change. The SR chart possesses optimal properties in its speed of detection of change. These properties are equivalent to those of the well-known CUSUM charts. However, the SR scheme will work under a broad range of conditions, where the theory for CUSUM charts is wanting.

For example, if the samples drawn from a process are dependent, then one encounters difficulty in implementing the CUSUM technique, whereas the SR charts are readily applicable. Moreover, the interpretation of the detection level A of an SR chart permits direct evaluation of the procedure's performance. One does not need lengthy tables to determine the ARL of the monitoring scheme. The practitioner using an SR chart can immediately identify the impact of changing the detection level (see Section 5).

(3) This paper presented SR control charts in the context of failure data and a non-homogeneous Poisson model. Other applications of the SR scheme include continuous data and non-parametric implementation. The same advantages of the SR chart over the CUSUM chart carry over to these other applications.

Acknowledgement

Some of this work was supported by a grant from the United States–Israel Bi-National Science Foundation.

REFERENCES

BELL, C., GORDON, L. & POLLAK, M. (1994) An efficient nonparametric detection scheme and its application to surveillance of a Bernoulli process with unknown baseline, *Change Point Problems, IMS Lecture Notes and Monographs*, 23, pp. 7–27.

COX, D. R. & LEWIS, P. A. W. (1966) *The Statistical Analysis of Series of Events* (London, Chapman and Hall).

CROARKIN, M. C., HAGWOOD, C. & POLLAK, M. (1993) Surveillance schemes with application to mass calibration, *NIST Technical Report 5058* (Gaithersburg, MD, Statistical Engineering Division, NIST).

VAN DOBBEN DE BRUYN, C. S. (1968) *Cumulative Sum Tests* (London, Griffin).

GORDON, L. & POLLAK, M. (1990) Average run length to false alarm for surveillance schemes designed with partially specified pre-change distribution, *Annals of Statistics*, in press.

GORDON, L. & POLLAK, M. (1994) An efficient sequential nonparametric scheme for detecting a change of distribution, *Annals of Statistics*, 22, pp. 763–804.

GORDON, L. & POLLAK, M. (1995) A robust surveillance scheme for stochastically ordered alternatives, *Annals of Statistics*, 23, pp. 1350–1375.

JELINSKI, Z. & MORANDA, P. B. (1972) Software reliability research. *In:* W. FREIBERGER (Ed.), *Statistical Computer Performance Evaluation*, pp. 465–484 (New York, Academic Press).

KENETT, R. S. (1981) An analysis of UNIX™ system crashes due to power failures, *Memorandum for File*, Bell Laboratories.

KENETT, R. & POLLAK, M. (1986) A semi-parametric approach to testing for reliability growth with an application to software systems, *IEEE Transactions on Reliability*, R-35, pp. 304–311.

LITTLEWOOD, B. & VERRALL, J. L. (1973) A Bayesian reliability growth model for computer software, *Journal of the Royal Statistical Society, Series C*, 22, pp. 332–346.

LORDEN, G. (1971) Procedures for reacting to a change in distribution, *Annals of Mathematical Statistics*, 42, pp. 1897–1908.

LORDEN, G. & EISENBERGER, I. (1973) Detection of failure rate increases, *Technometrics*, 15, pp. 167–175.

MEVORACH, Y. & POLLAK, M. (1991) A small sample size comparison of the Cusum and the Shiryayev–Roberts approaches to change point detection, *American Journal of Mathematical and Management Sciences*, 11, pp. 277–298.

MOUSTAKIDES, G. V. (1986) Optimal stopping times for detecting changes in distribution, *Annals of Statistics*, 14, pp. 1379–1387.

PAGE, E. S. (1954) Continuous inspection schemes, *Biometrika*, 41, pp. 100–115.

POLLAK, M. (1985) Optimal detection of a change in distribution, *Annals of Statistics*, 13, pp. 206–227.

POLLAK, M. (1987) Average run lengths of an optimal method of detecting a change in distribution, *Annals of Statistics*, 15, pp. 749–779.

POLLAK, M. & SIEGMUND, D. (1991) Sequential detection of a change in a normal mean when the initial value is unknown, *Annals of Statistics*, 19, pp. 394–416.

POLLAK, M. & SIEGMUND, D. (1985) A diffusion process and its application to detecting a change in the drift of Brownian motion, *Biometrika*, 72, pp. 267–280.

RITOV, Y. (1990) Decision theoretic optimality of the Cusum procedure, *Annals of Statistics*, 18, pp. 1464–1469.

ROBERTS, S. W. (1966) A comparison of some control chart procedures, *Technometrics*, 8, pp. 411–430.

SHIRYAYEV, A. N. (1963a) On optimum methods in quickest detection problems, *Theory of Probability and its Applications*, 8, pp. 22–46.

SHIRYAYEV, A. N. (1963b) On the detection of disorder in a manufacturing process, *Theory of Probability and its Applications*, 8, pp. 247–265.

SIEGMUND, D. & VENKATRAMAN, E. S. (1995) Using the generalized likelihood ratio statistics for sequential detection of a change-point, *Annals of Statistics*, 23, pp. 255–271.

YAKIR, B. (1994) Optimal detection of a change in distribution when the observations are independent, *Technical Report*, Department of Statistics, Hebrew University of Jerusalem.

ZACKS, S. (1991) Detection and change point problems. *In:* B. K. GHOSH & P. K. SEN (Eds), *Handbook of Sequential Analysis*, Chapter 3 (New York, Marcel Dekker).

Appendix A: Computation of $R_w(t)$

Set $a^- = \max(a, 0)$. Then, we have

$$R_w(t) = \int_0^t \Lambda_{s,t}\, ds = \sum_{j=0}^{n(t)} \int_{\sum_{i=1}^{j} X_i}^{t \wedge \sum_{i=1}^{j+1} X_i} \left(\frac{w}{w_0}\right)^{n(t)-j} \exp\left[(w_0 - w)(t-s)\right] ds$$

$$= \sum_{j=0}^{n(t)} \frac{1}{w_0 - w} \left(\frac{w}{w_0}\right)^{n(t)-j} \left\{ \exp\left[(w_0 - w)(t - \sum_{i=1}^{j} X_i)\right] - \exp\left[(w_0 - w)(t - \sum_{i=1}^{j+1} X_i)^-\right] \right\}$$

$$= \frac{1}{w_0 - w} \left\{ \left(\frac{w}{w_0}\right)^{n(t)} \exp\left[(w_0 - w)t\right] + \left(1 - \frac{w}{w_0}\right) \sum_{j=1}^{n(t)} \left(\frac{w}{w_0}\right)^{n(t)-j} \exp\left[(w - w_0)(t - \sum_{i=1} X_i)\right] - 1 \right\}$$

$$= \frac{1}{w_0} \left\{ \frac{w_0}{w_0 - w} \left(\frac{w}{w_0}\right)^{n(t)} \exp\left[(w_0 - w)t\right] + \sum_{j=1}^{n(t)} \left(\frac{w}{w_0}\right)^{n(t)-j} \exp\left[(w_0 - w)(t - \sum_{i=1}^{j} X_i)\right] - \frac{w_0}{w_0 - w} \right\}$$

Appendix B: Proof of Theorem 1

If $P_1(t)$ and $P_2(t)$ are independent non-homogeneous Poisson processes with intensity functions $\theta_1(t)$ and $\theta_2(t)$, respectively, then $P_3(t) = P_1(t) + P_2(t)$ is a non-homogeneous Poisson process with intensity function $\theta_3(t) = \theta_1(t) + \theta_2(t)$.

We first prove part (ii) of Theorem 1. Let $P_1(t)$ have $\theta_1(t) = w_0$ for all $t \geq 0$, and let $P_2(t)$ be an independent process with $\theta_2(t) = \theta(t) - w_0$ so that $P_3(t) = P_1(t) + P_2(t)$ has the same stochastic structure as our process. Since $n_t - n_s$ is larger for P_3 than for P_1, it follows that, if $w < w_0$, then $\Lambda_{s,t}$ is smaller for P_3 than it is for P_1. Hence, $R_w(t)$ is smaller for P_3 than it is for P_1; so $T_{A,w}$ is larger for P_3 than it is for P_1. Hence, $ET_{A,w}$ for P_3 is larger than it is for P_1 (with equality holding if and only if $\theta(t) = w_0$ for all $t \geq 0$).

A straightforward calculation shows that, if $\theta(t) = w_0$ (as is the case for P_1 above), then $E(R_w(t + y) \mid n(t), X_1, \ldots, X_{n(t)}) = R_w(t) + y$, so that $R_w(t) - t; t \geq 0$ is a martingale with zero expectation. It follows that $R_w(t) - t;\ t \geq 0$ is also a martingale with zero expectation. Therefore, by the optional sampling theorem, $E(R_w(T_{A,w}) - T_{A,w}) = 0$, so that $ET_{A,w} = ER_w(T_{A,w})$. However, by definition, $R_w(T_{A,w}) = A$ (since, under the conditions of part (ii), the process can only increase continuously), so that $ET_{A,w} = A$. This completes the proof of part (ii).

The proof of part (i) is analogous. Here, one must choose P_1 and P_2 to have intensities $\theta_1(t) = \theta(t)$ and $\theta_2(t) = w_0 - \theta(t)$ respectively. Our process is now P_1 and, for analogous reasons, $T_{A,w}$ for P_1 is larger than $T_{A,w}$ for $P_3 = P_1 + P_2$, which has itensity $\theta(t) = w_0$ for all $t \geq 0$. This accounts for the first part of (i). In this case, the process can increase in jumps; hence, $R_w(T_{A,w}) \geq A$ and (by the same martingale argument as above) $E(T_{A,w} \mid \theta(t) = w_0 \text{ for all } t \geq 0) \geq A$. If $\theta(t) = w_0$ for all $t \geq 0$, then $(w/w_0)n_t \exp[(w_0 - w)t] \to 0$ a.s. as $t \to \infty$, so that, as $A \to \infty$, $T_{A,w}$ is asymptotically equivalent to

$$T^{\star}_{A,W} = \min \left\{ t \mid \sum_{j=1}^{n(t)} \left(\frac{w}{w_0}\right)^{n(t)-j} \exp\,(w_0 - w) \sum_{i=j+1}^{n(t)} X_i \geq w_0 A \right\}$$

This is an analog of the Shiryayev–Roberts stopping rule employed in the sequence of waiting times between jumps X_1, X_2 and X_3, so that the results of Pollak (1987) are applicable. This accounts for part (i). The value of C_w is a direct application of the formula for G_0 in Lorden and Eisenberger (1973, p. 171).

Remark: An analogous theorem can be formulated for the case of surveillance for either an increase or a decrease, i.e. $\theta(t) = w_0$ for $0 \leq t \leq v$ and $\theta(t) \neq w_0$ for $t \geq v$.