MARCEL DEKKER, INC. • 270 MADISON AVENUE • NEW YORK, NY 10016

COMMUNICATIONS IN STATISTICS
Theory and Methods
Vol. 32, No. 2, pp. 475–495, 2003

Run-Length Distribution for Control Charts with Runs and Scans Rules

Galit Shmueli[1,*] and Ayala Cohen[2]

[1]Department of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA
[2]Faculty of Industrial Engineering and Management, Technion-Israel Institute of Technology, Haifa, Israel

ABSTRACT

In order to increase the power of the classical Shewhart control charts for detecting small shift, several supplementary rules based on runs and scans were introduced by the Western Electric Company in 1956. In this article we introduce a new method for computing the run-length distribution for a Shewhart chart with runs and scans rules. Our method yields an exact expression for the run-length generating function. We can then use either one of two techniques for extracting the probability function. One leads to recursive formulas and the other to non-recursive formulas. We investigate the performance of some

*Correspondence: Galit Shmueli, Department of Decision and Information Technologies, Robert H. Smith School of Business, University of Maryland, College Park, MD 20742-1815, USA; E-mail: gshmueli@rhsmith.umd.edu.

DOI: 10.1081/STA-120018196 0361-0926 (Print); 1532-415X (Online)

popular runs and scans rules and show that the run-length distribution is highly skewed. Comparing the entire distributions of different rules, rather than simply the widely-used expectations (ARLs), leads to important new conclusions on the advantages of applying each of these rules vs. using a simple chart. Finally, we introduce a Web application that incorporates these theoretical results into a simple and practical tool that can be used by practitioners.

Key Words: Control charts; Runs rules; Scans rules; Western Electric; Run length distribution.

1. INTRODUCTION

A popular method for monitoring a process, originated by Shewhart in 1924, is by using statistical control charts. In their simplest form, control charts consist of upper and lower control limits along with the target value. Small samples are then drawn from the process at regular intervals and the appropriate value is calculated and plotted on the chart (e.g., the sample average is plotted on a chart that monitors the process mean).

When the variation in process quality is due to random causes alone, the process is said to be in-control. In this case, the chart should not signal (all points should fall within the control limits). If the process variation includes both random and special causes of variation, the process is said to be out-of-control. Control charts should then signal an alarm by exceeding the control limits. An alarm signaled by a control chart indicates that special causes of variation are present, and some action should be taken, ranging from taking a re-check sample to the stopping of a production line in order to trace and eliminate these causes. On the other hand, an alarm may be a false one, when in practice no change has occurred.

The design of control charts is a compromise between the risks of not detecting real changes and false alarms. The American Standard is based on "three-sigma" control limits (corresponding to 0.27% of false alarms), while the British Standard uses "3.09 sigma" limits (corresponding to 0.2% of false alarms). In both cases it is assumed that a normal distribution underlies the relevant estimators. It is then easy to compute the probability of detecting a shift of a given size, using normal probabilities. When other statistics, rather than the mean are used for monitoring (e.g., the sample standard deviation for monitoring the process variation), the normality assumption no longer holds. In these cases "three-sigma" limits are usually improper, and probability limits are preferred.

The probability of exceeding the limits can then be calculated according to the distribution of the sample statistic.

It has been shown that Shewhart-type charts are efficient in detecting medium to large shifts, but are insensitive to small shifts. Consequently, other types of control charts have evolved for the detection of small shifts (e.g., Cusum and EWMA charts). In comparison with the simplicity of Shewhart charts, the latter are usually more complicated in the sense of plotting the charts, interpreting them, and assessing their properties.

One attempt to increase the power of Shewhart-type charts is by adding supplementary stopping rules based on runs and scans. The most popular stopping rules were suggested by the "Western Electric Company." Although runs and scans rules have been widely used, only relatively few studies were done to determine the exact properties of these charts. Based on Markov chains, Champ and Woodall (1987) derived the exact run-length distribution and applied it to compute the time until signaling an alarm for Shewhart chart with one or more runs/scans rules. Palm (1990) used their method to construct tables of run-length percentiles, when one or more of 3 additional rules are used with a Shewhart chart for monitoring the mean (X-bar chart). Approximations for false alarm rates when applying supplemental runs/scans rules to an X-bar chart were given by Walker et al. (1991), who used simulations. Formulas and tables for long-run probabilities of signaling were given by Champ and Woodall (1997) for several rules applied to an X-bar chart. Lowry et al. (1995) used the Markov chain approach to study the Western Electric run rules for S- and R-charts, and suggested alternative rules, also based on runs.

In this article, we suggest a new method for computing the exact run-length distribution for various types of runs and scans rules. Our method is based on obtaining the run-length generating function, from which moments can be easily derived. These generating functions turn out to be rational functions, but differentiating them becomes too cumbersome after only a few derivatives. We therefore use their special form to derive both recursive as well as non-recursive exact expressions for the probability function. After introducing various types of runs and scans rules and some notation we describe our method. We then apply this method to some popular runs and scans rules.

Although our method leads to the same results as the method by Champ and Woodall (1987), it has several advantage over the Markov chain method. We discuss these advantages and compare the two methods.

This article has two contributions to the study of Shewhart control charts. Firstly, we introduce an alternative method for calculating the run-length distribution. Secondly, we compare the different runs and scans rules and demonstrate the uninformative nature of the ARL. A web application is presented that enables to select a runs or scans rule and the size of the shift in the process, outputting a plot of the corresponding run-length distribution. A summary concludes this article.

2. STOPPING RULES: TYPES AND NOTATION

The Western Electric Handbook (1956) suggests a set of decision rules for detecting non-random patterns on control charts. Specifically, it suggests concluding that the process is out of control if either

1. A single point exceeds the 3-sigma control limits;
2. Two out of three consecutive points exceed the 2-sigma warning limits;
3. Four out of five consecutive points exceed the 1-sigma limits;

or

4. Eight consecutive points fall on one side of the center line.

According to the Western Electric Handbook (1956) these rules should be applied separately to each side of the center line, and should be used either with an X-bar chart, or with an R-chart for samples size of 4 or more. Similar runs and scans rules were suggested for use with an R-chart, when the sample size is 2.

We use a general and compact notation for a runs/scans rule, similar to the one used by Champ and Woodall (1987): The rule signals if k of the last m *standardized* points fall in zone Z, and is denoted by $T(k, m, Z)$. When $k = m$ we call this a *run* and when $k < n$ we call this a *scan* (Balakrishnan and Koutras, 2002). The seven zones which are considered are:

Zone A_1 = the interval $(2, 3)$
Zone B_1 = the interval $(1, 2)$
Zone C_1 = the interval $(0, 1)$
Zone A_2 = the interval $(-3, -2)$
Zone B_2 = the interval $(-2, -1)$
Zone C_2 = the interval $(-1, 0)$
Zone S (for "Signal") = the interval $(-\infty, -3) \cup (3, \infty)$

(see Fig. 1). Thus the usual Shewhart chart is denoted by $T(1, 1, S)$.

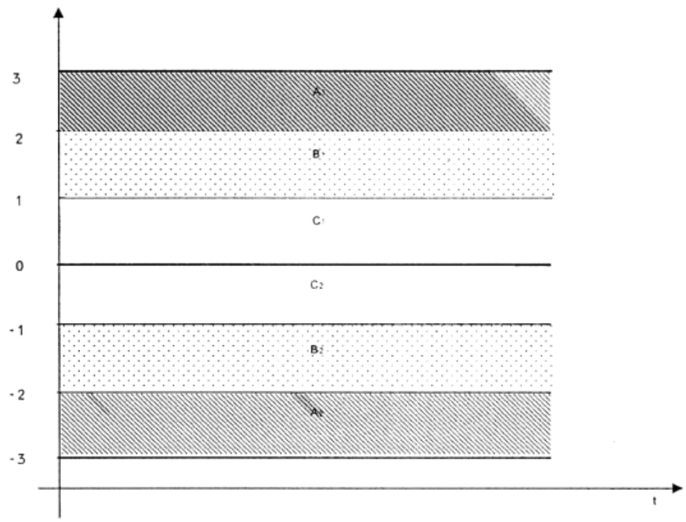

Figure 1. A control chart divided into the seven zones.

Following Champ and Woodall (1987), we consider the following rules (and some of their combinations) for a standardized control chart:

$R_1 = \{T(1, 1, S)\}$
$R_2 = \{T(2, 3, A_2), T(2, 3, A_1)\}$
$R_3 = \{T(4, 5, B_1 \cup A_1), T(4, 5, A_2 \cup B_2)\}$
$R_4 = \{T(8, 8, C_1 \cup B_1 \cup A_1), T(8, 8, A_2 \cup B_2 \cup C_2)\}$
$R_5 = \{T(2, 2, A_2), T(2, 2, A_1)\}$
$R_6 = \{T(5, 5, B_1 \cup A_1), T(5, 5, A_2 \cup B_2)\}$

In addition, we denote the probability of a point being in zone A_i, B_i, or C_i by α_i, β_i, and γ_i ($i = 1, 2$), respectively. The probability of exceeding the control limits is denoted by δ. These probabilities can be computed for an in-control process or for any given shift in the process mean. For example, if the process is in-control, a set of "3-sigma" control limits and "2-sigma" warning limits in the normal case leads to $\delta = 2\Phi(-3) = 0.0027$ and $\alpha_1 + \alpha_2 = 2(\Phi(3) - \Phi(2)) = 0.0428$. If, on the other hand, there is a shift of "1-sigma", then these probabilities are $\delta = 1 - \Phi(2) + \Phi(-4) = 0.023$ and $\alpha_1 + \alpha_2 = \Phi(2) - \Phi(1) + \Phi(-3) - \Phi(-4) = 0.137$. For sample statistics that are non-normal, these probabilities can be calculated using the sampling distribution of the statistic.

3. RUN-LENGTH GENERATING AND PROBABILITY FUNCTIONS WITH RUNS AND SCANS RULES

We define a "suffix" as a sequence of letters denoting the areas on the plot in which the last points fell. For example, A_1S denotes that the last two points fell in areas A_1 and S. To obtain the run-length generating function for a single rule, or a combination of rules, we follow Feller (1968) and generalize the method by take the following steps:

1. Map all of the Q distinct suffixes that cause an alarm.
2. Define $u_n^{(i)}$ as the probability that the nth point completes the ith suffix, with $u_0^{(i)} = 1$ $(i = 1, \ldots Q)$.
3. Write a recurrence relation for each $u_n^{(i)}$ $(i = 1, \ldots Q)$.
4. Compute the Q power series $U_i(s) = \sum_{n=0}^{\infty} u_n^{(i)} s^n$ $(i = 1, \ldots, Q)$.
5. The run-length generating function is given by $\mathcal{G}(s) = 1 - 1/(\sum_{i=1}^{Q} U_i(s) - Q + 1)$.

The ARL is then easily obtained by differentiating the generating function once.

Using this procedure for run-related variables leads to generating functions of a special form, namely, a rational function. In many cases this ratio of polynomials is very complicated, and the ordinary way of obtaining the probability function by differentiation is laborious. However, two alternative techniques for deriving the probability function from a rational generating function are by partial fraction expansion (Feller, 1968; Shmueli and Cohen, 2000), and by using a combinatorial approach (Stanley, 1990). The latter method leads to a recursive formula for the probability function, which is useful, for example, if one is interested in the entire probability function. The former method, which leads to a non-recursive formula for the probability function, is more efficient for purposes of estimation, or computation of specific probabilities.

In all cases, we deal with the addition of runs or scans rules to the classical simple rule "stop if a single point exceeds the control limits" (R_1). In this section, we specify the recurrence relations that correspond to the addition of each type of runs or scans rule to the ordinary Shewhart chart, and the run-length generating and probability functions in each case. The formulas that we give for the generating functions, probability functions, and ARLs in this section can be applied to any type of "3-sigma" chart, even when non-normal statistics are used (e.g., R-chart, S-chart). Numerical results and plots are given only for the normal case (e.g., for the X-bar or P-chart), but similar results can be computed for non-normal cases using the general formulas.

3.1. Runs Rules and k Order Distributions

Using the notation described in Sec. 2, the rules presented can be divided into two types:

1. Runs rules of the type $T(k, k, Z)$ ("stop if k consecutive points fall in zone Z"). This includes rules $R_1, R_4, R_5,$ and R_6.
2. Scans rules of the type $T(k, k+1, Z)$ ("stop if k points within the last $k+1$ consecutive points fall in zone Z"). This includes rules R_2 and R_3.

The waiting time until an event of type 1 occurs, belongs to the "order k geometric distribution." This distribution deals with waiting times for the first run of k consecutive successes in Bernoulli trials (Johnson et al., 1992). When $k = 1$ this reduces to the ordinary geometric distribution. The waiting time until an event of type 2 occurs, belongs to the "k-out-of-d ($d > k$) distribution," which deals with the waiting time until k successes occur within a maximum of d consecutive Bernoulli trials. This distribution was studied with relation to system reliability (known as *consecutive-k-out-of-m-from-n:F* systems, see Balakrishnan and Koutras, 2002; Chao et al., 1995; Shmueli, 2003).

These two distributions are discussed in Shmueli and Cohen (2000), and expressions for some of their generating and probability functions are given there. We use these derivations to obtain the run-length distributions for the various cases.

3.2. Runs of Points Within Certain Zones (R_4, R_5, R_6 with R_1)

In this subsection, we consider the distribution of runs within certain zones. These distribution are relevant for rules R_4, R_5, R_6 in Shewhart charts for which an alarm is signaled following a run of k points in one of two zones, which we denote by Z_1 and Z_2, or by a single point in zone S. In other words, the run-length is the waiting time for the first run of either k events of type a, k events of type b, or one event of type c, in i.i.d. trials with four possible outcomes: a, b, c, and d. For example, for the combination of rules R_5 and R_1 there are three possible suffixes: S, A_1A_1, and A_2A_2.

Let the probability of falling in zone Z_i be denoted by p_i $(i = 1, 2)$. The three recurrence relations corresponding to the three possible suffixes S, $Z_1 \cdots Z_1$, and $Z_2 \cdots Z_2$, are given by:

$$\begin{aligned} \delta &= u_n^{(1)} \\ p_1^k &= u_n^{(2)} + u_{n-1}^{(2)} p_1 + u_{n-2}^{(2)} p_1^2 + \cdots + u_{n-k-1}^{(2)} p_1^{k-1} \\ p_2^k &= u_n^{(3)} + u_{n-1}^{(3)} p_2 + u_{n-2}^{(3)} p_2^2 + \cdots + u_{n-k-1}^{(3)} p_2^{k-1} \end{aligned} \tag{3.1}$$

Multiplying by s^n and summing to infinity, we obtain the three corresponding power functions $U_1(s)$, $U_2(s)$, $U_3(s)$, which are then combined to obtain the run-length generating function:

$$\begin{aligned} \mathcal{G}(s) &= 1 - \cfrac{1}{\frac{1-s+\delta s}{1-s} + \frac{1-s+(1-p_1)p_1^k s^{k+1}}{(1-s)(1-(p_1 s)^k)} + \frac{1-s+(1-p_2)p_2^k s^{k+1}}{(1-s)(1-(p_2 s)^k)} - 2} \\ &= \frac{\delta s + (p_1^k + p_2^k)s^k - [p_1^k(\delta+p_1) + p_2^k(\delta+p_2)]s^{k+1} - 2(p_1p_2)^k s^{2k} + (p_1p_2)^k(\delta+p_1+p_2)s^{2k+1}}{1-(1-\delta)s - [p_1^k(\delta+p_1) + p_2^k(\delta+p_2)]s^{k+1} - (p_1p_2)^k s^{2k} - (p_1p_2)^k(1-p_1-p_2-\delta)s^{2k+1}} \end{aligned} \tag{3.2}$$

The exact Average Run Length (ARL) follows by differentiating Eq. (3.2) once:

$$\mathrm{ARL} = \frac{(1-p_1^k)(1-p_2^k)}{\delta[(1-p_1^k)(1-p_2^k)] + (1-p_1)(1+p_2^k)p_1^k + (1-p_2)(1+p_1^k)p_2^k} \tag{3.3}$$

Aki (1992) introduced a different method for obtaining the waiting-time generating function and moments for such cases.

Deriving the pdf by differentiation from these types of generating functions is laborious, as the derivatives become very complicated. Nevertheless, we can use the fact that the generating function is a ratio of two polynomials to obtain the complete probability function more elegantly by using one of the two methods mentioned earlier.

Figure 2 illustrates the run-length distribution when applying rule R_4 to a Shewhart control chart (with R_1), for shifts of different magnitude. For brevity we do not present figures for rules R_5, R_6, but a similar pattern

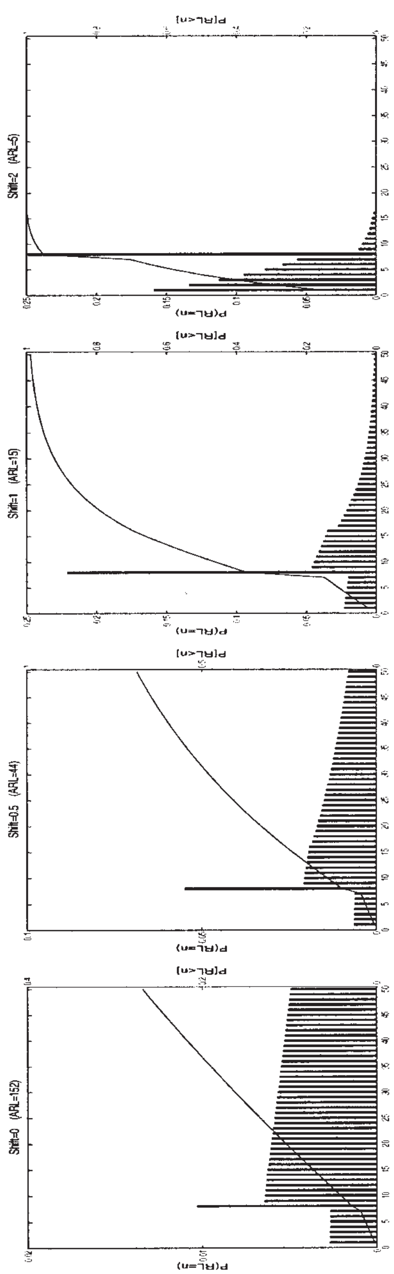

Figure 2. Run-length probability and cumulative distribution, when rule R_4 is added to rule R_1 (the left and right scaling correspond to the pdf and the cdf, respectively).

is observed. For each of the three rules, a spike occurs at $n = 8, 2$, and 5, respectively. This spike describes the point where the runs-rule kicks in. When the process is in-control (shift = 0), the distribution is highly skewed, with a right tail which decreases slowly. As the magnitude of the shift increases, the high probabilities tend to be concentrated at the small values, and the mode at the value k is very pronounced. In these cases, the use of percentiles from the exact distribution seems more informative than relying solely on the ARL. Table 1 gives the run-length median and quartiles for each of the three rules, along with the corresponding ARL. For small shifts, the ARL is considerably smaller than the median, indicating the skewness of the corresponding run-length distribution. This fact was noticed by Palm (1990) who suggested that a practitioner would be more interested in percentiles than in the ARL. For example, applying rule R_6 with R_1, we can say that a false alarm occurs on average every 350 samples. This might be misleading due to the highly skewed run-length distribution. Using the median, we can say that a false alarm will occur within the first 242 samples with probability 0.5. Alternatively, using the first quartile, we can say that a false alarm will not occur within the first 101 samples, with probability 0.75.

Table 1. Run-length quartiles and ARL when rules R_4, R_5, or R_6 are applied with R_1.

	Rule R_4				Rule R_5				Rule R_6			
Shift	Q_1	Med	Q_3	ARL	Q_1	Med	Q_3	ARL	Q_1	Med	Q_3	ARL
0.0	47	107	210	152.04	81	193	385	277.98	101	242	484	349.38
0.2	35	78	152	110.22	65	155	308	222.55	81	194	387	279.54
0.4	20	43	81	59.71	39	93	186	134.16	48	115	229	165.48
0.6	13	25	45	33.63	22	52	104	75.27	27	62	123	89.07
0.6	9	16	28	21.07	13	30	59	42.96	15	34	66	48.40
1.0	8	11	19	14.58	8	18	35	25.61	9	20	38	27.74
1.2	8	9	14	10.90	5	11	22	16.06	6	12	23	17.05
1.4	6	8	10	8.60	4	8	14	10.60	5	8	15	11.28
1.6	4	8	9	7.03	3	5	10	7.36	4	6	10	7.98
1.8	3	6	8	5.85	2	4	7	5.36	3	5	8	5.97
2.0	2	5	8	4.89	2	3	5	4.07	2	5	6	4.68
2.2	2	3	6	4.08	2	3	4	3.22	2	3	5	3.78
2.4	1	3	5	3.38	1	2	3	2.64	1	3	5	3.14
2.6	1	2	4	2.81	1	2	3	2.22	1	2	4	2.64
2.8	1	2	3	2.35	1	2	3	1.93	1	2	3	2.26
3.0	1	1	2	1.99	1	1	2	1.70	1	1	2	1.95

MARCEL DEKKER, INC. • 270 MADISON AVENUE • NEW YORK, NY 10016

In summary, either the median, or some other quantile can be used to quantify the effectiveness of adding a certain rule for the detection of a shift in the mean, rather than relying solely on the ARL.

3.3. Scans of Type k of the Last $k+1$ Points in Certain Zones (R_2, R_3 with R_1)

The second type of rules is of the form "stop if k of the last $k+1$ consecutive points are in zone Z_1 or Z_2" (in general, $k+1$ can be replaced by any value larger than k). Distributions resulting from such rules has been researched by many authors (see the books by Balakrishnan and Koutras (2002) and by Glaz et al. (2001) for an extensive description of the various methods and results for scan related distributions). Applying such a scans rule, the run-length is the waiting time for the first occurrence of k points within the last k or $k+1$ points in zones Z_1 or Z_2, or one point outside the control limits.

For purposes of simplicity, we illustrate our method for $k=2$. However, it can apply to any k. Seven suffixes will cause an alarm: S, $\{Z_1Z_1\}$, $\{Z_1Z_2Z_1\}$, $\{Z_1\overline{(Z_1 \cup Z_2 \cup S)}Z_1\}$, and three additional suffixes symmetric to the last three, with Z_1 and Z_2 exchanged. The recurrence relations for the first suffix is given in the first equation in (3.1). The ones corresponding to the next three suffixes are:

$$\begin{aligned} p_1^2 &= u_n^{(2)} + p_1\left[u_{n-1}^{(2)} + u_{n-1}^{(3)} + u_{n-1}^{(4)}\right] \\ p_1^2 p_2 &= u_n^{(3)} + u_{n-1}^{(6)} p_1 + p_1 p_2\left[u_{n-2}^{(2)} + u_{n-2}^{(3)} + u_{n-2}^{(4)}\right] \\ p_1^2(1-p_1-p_2-\delta) &= u_n^{(4)} + p_1(1-p_1-p_2-\delta)\left[u_{n-2}^{(2)} + u_{n-2}^{(3)} + u_{n-2}^{(4)}\right] \end{aligned} \tag{3.4}$$

and the last three equations are identical, exchanging p_1 and p_2 and the superscripts (2), (3), (4) with (5), (6), (7), respectively. Next, we multiply each equation by s^n and sum to infinity. In this case we cannot compute $U_i(s)$ separately from each equation. Instead, we solve a linear equation system of rank 6 (or in general, rank $4k-2$). The resulting generating function is a ratio of order 7 polynomials. The ARL is obtained by differentiating the generating function. The probability function is obtained as described earlier.

Figure 3 illustrates the run-length distribution when adding rule R_2 ($k=2$) to a control chart (with R_1), for shifts of different magnitude. As with runs rules, the addition of this type of scans rules to a control

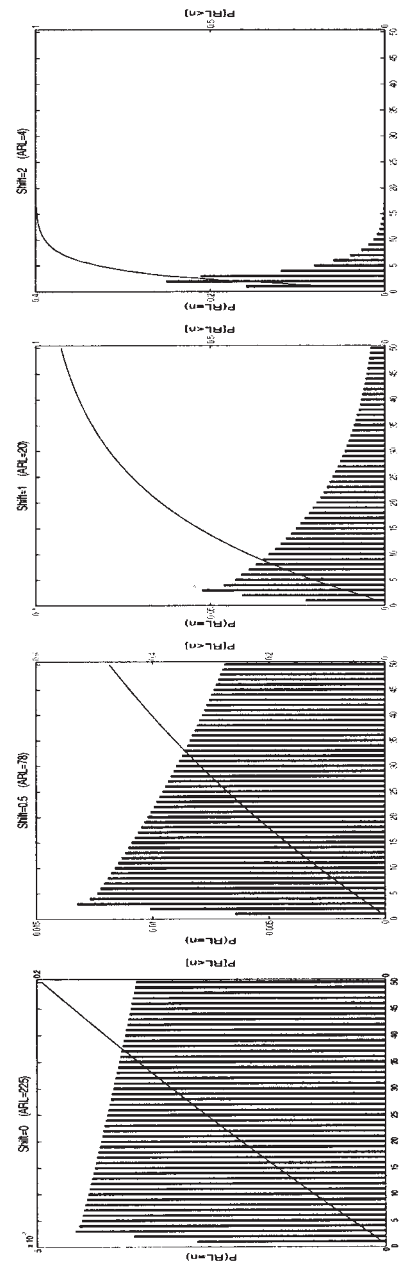

Figure 3. Run-length probability and cumulative distribution, when rule R_2 is added to rule R_1 (the left and right scaling correspond to the pdf and the cdf, respectively).

Table 2. Run-length quartiles and ARL when scan rules R_2 or R_3 are applied with R_1.

	Rule R_2				Rule R_3			
Shift	Q_1	Med	Q_3	ARL	Q_1	Med	Q_3	ARL
0.0	66	157	315	225.44	49	116	229	166.05
0.2	52	124	247	177.56	37	84	166	120.70
0.4	31	73	145	104.46	20	45	88	63.88
0.6	18	41	80	57.92	12	24	46	33.99
0.8	10	23	46	33.12	7	14	26	19.78
1.0	7	14	27	20.01	5	10	17	12.66
1.2	4	9	17	12.81	5	7	11	8.84
1.4	3	6	12	8.69	4	5	8	6.62
1.6	3	5	8	6.21	4	5	6	5.24
1.8	2	4	6	4.66	3	4	5	4.33
2.0	2	3	5	3.65	2	4	5	3.68
2.2	2	3	4	2.96	2	3	4	3.18
2.4	1	2	3	2.48	1	3	4	2.78
2.6	1	2	3	2.13	1	2	4	2.43
2.8	1	2	3	1.87	1	2	3	2.14
3.0	1	1	2	1.68	1	1	2	1.89

chart leads to a highly skewed run-length distribution. As the shift increases, the probabilities of the values k and $k+1$ become more pronounced, and the right tail decreases at a higher rate. In this case, as in the case of runs rules, we suggest using quantiles of the run length distribution rather than relying on the ARL as a single measure of the detection rate. Table 2 gives the run-length median and quartiles beside the corresponding ARL, for shifts of different magnitude, when applying rules R_2 or R_3 with R_1 to a control chart.

3.4. More Complicated Scans Rules, and Combinations of Runs and Scans Rules

The method described in Sec. 3 can be utilized to handle any type of runs or scans rule, as well as combinations of such rules. The rank of the linear equation system is a function of the history that is taken into account by the rule and of the number and type of rules combined.

From a mathematical point of view, the linear equation system is always well defined. The number of equations, however, may be large. For example, combining rules R_1, R_2, and R_3 results in 142 suffixes, and the addition of rule R_4 to these three rules results in 792 suffixes. When the number of linear equations is very large, it is easier to solve them if the size of the shift and the relevant zones of the chart are specified numerically. The symbolic component of the linear equation then depends on s alone, thereby reducing the complexity of the computation. It is useful to note, however, that the addition of many rules to an ordinary Shewhart chart will cause an increased rate of false alarm, and is therefore generally not recommended.

Since following the method manually can become laborious, we automated it through a computer program (written in Matlab). The program finds all the relevant suffixes, constructs the linear equations, solves them, and gives the run-length distribution. In other words, it automates steps 1–5 of the method, and then gives expressions for the run-length generating and probability functions. A copy of this program can be obtained from the first author.

4. COMPARING OUR METHOD WITH THE MARKOV CHAIN METHOD

The Markov chain approach, which was introduced by Champ and Woodall (1987), was the first method that enabled the exact computation of the run-length distribution for various runs and scans rules. Using the Markov chain method involves mapping all the possible states of the chain, computing the transition matrix (usually a large, sparse matrix), and using it to obtain recursive formulas for the run-length distribution. Their Fortran program (Champ and Woodall, 1990) automates this process, yielding the pdf. In order to compute the exact ARL, the transition matrix (which is usually very large) must be inverted. To overcome this difficulty, Woodall and Reynolds (1983) suggested that in some cases an approximation can be used, based on the geometric limiting form of the run-length distribution.

In comparison, our method involves listing all the suffixes that lead to an alarm. A short computer program enumerates these suffixes for any runs or scans rule or for a combination of rules. The computational effort in our method is in solving a set of linear equations symbolically. This step has also been automated through a computer program. In addition to the exact run-length distribution, our method leads to an expression for the probability generating function. To derive the moments of the distribution (e.g., the ARL) is thus immediate.

Finally, our method can lead to both *recursive* and *non-recursive* formulas for the run-length probabilities. When the entire distribution is required, the recursive formulas are preferable. However, to obtain a single probability (e.g., for estimation) it is more efficient to use the non-recursive method. This is especially important in our case, as the run-length distributions have very long right tails.

5. COMPARISON OF WESTERN ELECTRIC RULES

In order to decide which supplemental runs or scans rules to apply for monitoring a process, one should consider both practical matters (e.g., production speed) as well as the waiting time distribution that is expected for different rules. For example, the "8 points on one side of the centerline" runs rule (R_4) might be inappropriate when detection within less than 8 samples is required. But what are the differences between short-term rules such as R_2 and R_3?

Although tables with quantiles and ARLs for the above Western Electric rules were published (Palm, 1990), they were not used for the purpose of comparing the different rules. To learn more about the differences between the Western Electric rules and about the added value of using runs and scans rules over a simple Shewhart chart we compare the run-length distribution when applying each of the rules R_2–R_6 to a control chart, as well as the simple rule R_1. In the last section, we showed that the run-length distribution is highly skewed. We therefore compare the run-length distributions via their quartiles using box plots. We use a version of a box-plot where the whiskers extend to the 5th and 95th quantiles. Unlike in the display of data, where the classical box-plot is helpful for detecting outliers, the version we use is more suitable for describing theoretical, particularly skewed distributions.

Figure 4 describes the run-length distribution when the process is in-control. The six boxes, from top to bottom, correspond to supplemental rules R_4, R_3, R_2, R_5, R_6 and the simple rule R_1. The boxes are ordered according to location (and dispersion). When the process is in-control, long run-lengths or equivalently, rare false alarms,

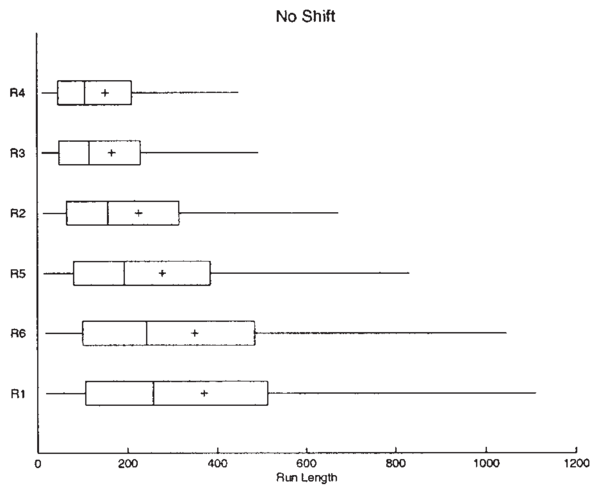

Figure 4. Run-length box plots for an in-control process (whiskers extend to the 5th and 95th quantiles).

are desirable. In other words, we would prefer a rule with a high-valued location and small dispersion. Applying this criterion, no rule is optimal relative to the others. Note that this is true for the simple rule R_1 as well. This result gives a new insight into the addition of the Western Electric rules. It is well-known that the addition of a runs or scans rule decreases the in-control ARL, and based on this fact a simple Shewhart chart is preferred to one with supplemental rules for in-control processes. However, the box plots in Fig. 4 reveal a more complete picture where although the ARL and the median run-length of R_1 are the largest, so is its spread! Therefore, using a simple chart as opposed to one with supplemental rules, we would expect less false alarms *on average* but the timings of false alarms would be more variable.

Figure 5 describes the run-length distribution when there is a shift in the mean, of sizes 0.5, 1, and 2 standard deviations, respectively. As the shift increases, the ordering by location changes for some rules in comparison to the in-control situation (e.g., rules R_3 and R_4). In some cases the nature of the distribution changes. One instance of this is the change of skewness as a function of the shift size. The ARL (denoted by "+" in the box) which is always larger than the median for small shifts, is smaller than the median for large shifts in some cases (e.g., for R_6).

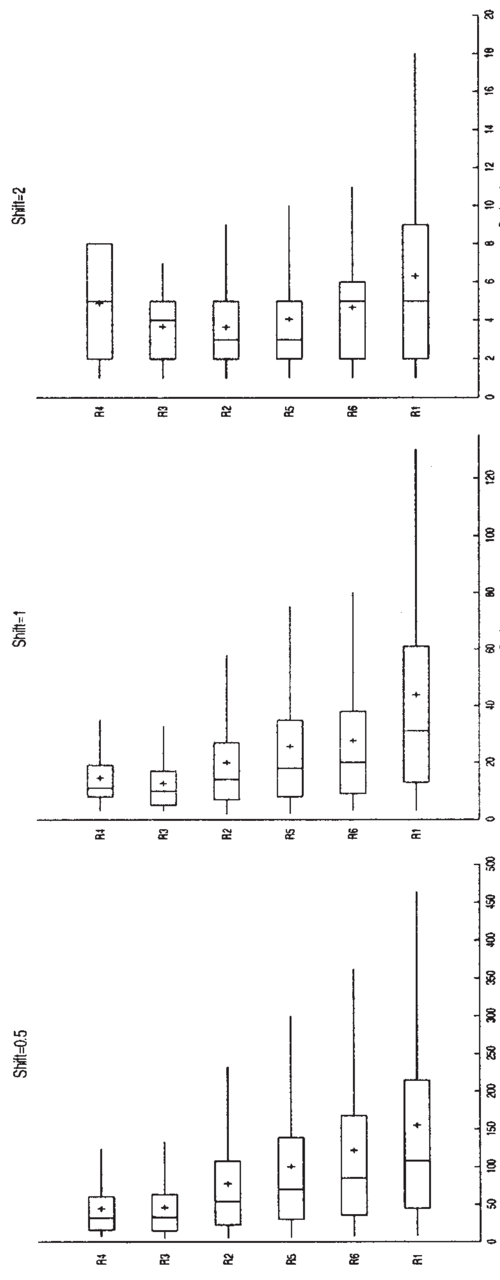

Figure 5. Run-length Box Plots for a Shift in the Process Mean (whiskers extend to the 5th and 95th quantiles).

Another example is: while most rules lead to a right-tailed distribution as the shift increases, rule R_4 tends more and more to concentrate the alerts on the 8th sample (this can also be seen in Fig. 2).

The two figures also suggest that the popular notion that a simple Shewhart chart is favorable to a chart with supplemental rules for detecting medium to large shifts, is not always true.

When comparing rules that have a similar structure, a counter-intuitive phenomenon is revealed: Some pairs of rules, such as R_2 and R_5 ("two within three consecutive points fall in $(-3,-2)$ or $(2,3)$" and "two consecutive points fall in $(-3,-2)$ or $(2,3)$") exhibit a similar behavior, both when the process is in-control and when it shifts. These rules are defined quite similarly, and therefore this relationship might be expected. However, following this reasoning, we would expect the pair R_3 and R_6 ("four within five consecutive points fall in $(-3,-1)$ or $(1,3)$ " and "five consecutive points fall in $(-3,-1)$ or $(1,3)$") to have a similar run-length distribution. The plots show us that this is not the case.

Finally, note that the scale of the X-axis in Figs. 4 and 5, which gives the time until signaling in units of samples, greatly reduces in magnitude from the in-control case to the out-of-control cases. This difference in scale may have implications on the preference between rules. It could be the case where the difference between the signaling behaviors of some rules (e.g., rules R_1 and R_4 for a shift of size 2) is practically insignificant. In other words, the time scale related to the process is substantial for determining preferences between different rules.

6. A WEB APPLICATION

To make our theoretical derivations useful for practitioners, we created a web application (http://iew3.technion.ac.il/sqconline/runplot.html) that receives input from the user regarding the runs or scans rule they would like to learn about. The input screen is user friendly and simple, and there is a linked glossary for further explanations. The computations are then done in the background, and the output is a plot of the run-length distribution. This gives a visual and immediate picture of the waiting time until signaling an alarm when using that rule, and the attached probabilities. In addition, the plot includes the cumulative distribution, so that the median (and possibly other percentiles) can be immediately approximated.

Using a web application not only enables practitioners to learn about existing rules (such as the Western Electric rules), but it has the flexibility to evaluate new customized rules that may be more appropriate for the specific process that is being monitored. For example, if a process has a weekly component and only positive shifts are of interest, the user may be interested in a rule that signals if, say, a run of seven consecutive points fall above the centerline. Using our web application, one can try out several existing and new rules, and decide between them based on their theoretical properties, rather than rely on intuition.

7. SUMMARY

The addition of various stopping rules, which are based on runs and scans, was suggested by several authors in order to increase the sensitivity of the classical Shewhart control chart to small shifts in the process. These supplementary rules are mostly rules of the thumb, rather than theoretically based, and only few authors have investigated them theoretically.

The most popular measure for studying and comparing the performance of such charts, under shifts of different magnitude, is the average run length (given either by an exact numerical form or by an approximation).

In this article, we present a method which leads to exact symbolic formulas for both the run-length generating and probability functions (as well as the ARL). The method is based on carrying out five steps. We demonstrated how these steps are carried out for simple runs and scans rules. However, the process has been automated via a computer program, which can be used for more complicated rules and for combination of rules. Our website gives the run-length distribution for some of the most popular runs and scans rules.

Using our method, we show that the run-length distribution is highly skewed when applying these types of rules. We therefore recommend the use of quantiles and box plots over the single ARL measure for assessing and comparing the different rules. Some authors suggested to compare competing rules by comparing the out-of-control ARLs for rules which share the same in-control ARL (Lowry et al., 1995). However, when comparing complete run-length distributions for different rules, a more complex picture is revealed. Rules that share the same in-control ARL can have substantially different distributions. In particular, it seems that the variability increases with the ARL (or median). The story becomes even more complex when we look at the entire out-of-control distributions as well (rather than singly at the out-of-control ARL). The bottom line is that no rule is optimal in the sense of sensitivity. In general, rules that are more sensitive on average, tend to have a larger variability in signaling times. We find that some supplemental rules are in fact advantageous over a simple Shewhart chart when considering both location and dispersion of the in-control and out-of-control run-length distributions.

Finally, our Web application can aid in describing the run-length distribution in a more informative way, thereby leading to an improvement in the user's interpretation and expectations of the rate of alarms caused by charts with various runs and scans rules. Moreover, using our method, one can select an existing or a new rule, according to a required performance of the chart.

REFERENCES

Aki, S. (1992). Waiting time problems for a sequence of discrete random variables. *Annals of the Institute of Statistical Mathematics* 44:363–378.

Balakrishnan, N., Koutras, M. V. (2002). *Runs and Scans with Applications*. John Wiley.

Champ, C. W., Woodall, W. H. (1987). Exact results for Shewhart control charts with supplementary runs rules. *Technometrics* 29(4):393–399.

Champ, C. W., Woodall, W. H. (1990). A program to evaluate the run length distribution of Shewhart control charts with supplementary runs rules. *Journal of Quality Technology* 22:68–73.

Champ, C. W., Woodall, W. H. (1997). Signal probabilities of runs rules supplementing a Shewhart control chart. *Communications in Statistics—Simulation and Computation* 26:1347–1360.

Chao, M. T., Fu, J. C., Koutras, M. V. (1995). Survey of reliability studies of consecutive-k-out-of-n:F and related systems. *IEEE Transactions on Reliability* 44(1):120–127.

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*. Chapter XIII. Vol. 1. 3rd ed. New York: John Wiley.

Glaz, J., Naus J., Wallenstein, S. (2001). *Scan Statistics*. New York: Springer-Verlag,

Johnson, N. L., Kotz, S., Kemp, A. W. (1992). *Univariate Discrete Distributions*. 2nd ed. John Wiley.

Lowry, C. A., Champ, C. W. Woodall, W. H. (1995). The performance of control charts for monitoring process variation. *Communications in Statistics—Simulation and Computation* 24:409–437.

Palm, A. (1990). Tables of run-length percentiles for determining the sensitivity of Shewhart control charts for averages with supplementary runs rules. *Journal of Quality Technology* 22:289–298.

Shmueli, G. (2003). Computing consecutive-type reliabilities non-recursively. *IEEE Transaction on Reliability* 52.

Shmueli, G., Cohen, A. (2000). Run-related probability functions applied to sampling inspection. *Technometrics* 42(2):188–202.

Stanley, R. P. (1990). *Enumerative Combinatorics*. Chapter 4. Vol. I. Wadsworth and Brooks/Cole Advanced Books Software.

Walker, E., Philpot, J. W., Clement, J. (1991). False signal rates for the Shewhart control chart with supplementary runs tests. *Journal of Quality Technology* 23(3):247–252.

Western Electric Company. (1956). *Statistical Quality Control Handbook*.

Woodall, W. H., Reynolds, M. R. Jr. (1983). A discrete markov chain representation of the sequential probability ratio test. *Communications in Statistics—Sequential Analysis* 2:27–44.