Journal of Applied Statistics

ISSN: 0266-4763 (Print) 1360-0532 (Online) Journal homepage: https://www.tandfonline.com/loi/cjas20

Generalized run tests for statistical process control

Young H. Chun

To cite this article: Young H. Chun (2019) Generalized run tests for statistical process control, Journal of Applied Statistics, 46:12, 2164-2179, DOI: 10.1080/02664763.2019.1581147

To link to this article: https://doi.org/10.1080/02664763.2019.1581147

Published online: 16 Feb 2019.

JOURNAL OF APPLIED STATISTICS
2019, VOL. 46, NO. 12, 2164–2179
https://doi.org/10.1080/02664763.2019.1581147

Taylor & Francis
Taylor & Francis Group

Generalized run tests for statistical process control

Young H. Chun

Department of Information Systems and Decision Sciences, E. J. Ourso College of Business, Louisiana State University, Baton Rouge, LA, USA

ABSTRACT

In a sequence of elements, a *run* is defined as a maximal subsequence of like elements. The number of runs or the length of the longest run has been widely used to test the *randomness* of an ordered sequence. Based on two different sampling methods and two types of test statistics used, run tests can be classified into one of four cases. Numerous researchers have derived the probability distributions in many different ways, treating each case separately. In the paper, we propose a *unified* approach which is based on recurrence arguments of two mutually exclusive sub-sequences. We also consider the sequence of nominal data that has more than two classes. Thus, the traditional run tests for a binary sequence are special cases of our generalized run tests. We finally show that the generalized run tests can be applied to many quality management areas, such as testing changes in process variation, developing non-parametric multivariate control charts, and comparing the shapes and locations of more than two process distributions.

ARTICLE HISTORY
Received 25 February 2018
Accepted 1 February 2019

KEYWORDS
Non-parametric statistics; run test; statistical process control

1. Introduction

In many experiments or observational studies, each element in the sample space can assume only one of two possible outcomes, such as heads or tails, success or failure, or up or down. The order in which the elements of the sample were drawn is frequently available. In any ordered sequence of n elements of two kinds, a 'run' is defined as a succession of similar events preceded and succeeded by different elements. The number of elements in a run is referred to as its length. We can also count the total number of runs in an ordered binary sequence.

Consider, for example, the ordered sequence of 10 binary numbers, 1 1 1 1 0 1 1 0 0 1. The total number of *runs* in the sequence is five (i.e. $\underline{1\ 1\ 1\ 1}\ \underline{0}\ \underline{1\ 1}\ \underline{0\ 0}\ \underline{1}$), among which the number of 1 runs is three and the number of 0 runs is two. In the same sequence, the *length* of the longest 1 run is four (i.e. $\underline{1\ 1\ 1\ 1}$), whereas the *length* of the longest 0 run is two (i.e. $\underline{0\ 0}$). Thus, the *maximum run length* (i.e, the length of the longest 1 or 0 run) is four in the binary sequence.

Various types of *run tests* have been proposed to examine whether or not a two-valued data sequence is randomly generated. The test statistic is often derived from (i) the number

CONTACT Young H. Chun chun@lsu.edu Department of Information Systems and Decision Sciences, E. J. Ourso College of Business, Louisiana State University, Baton Rouge, LA 70803, USA

of runs or (ii) the maximum run length in a given sequence. If the number of runs is too high or too low, for example, we may suspect the statistical independence of each trial or the randomness of the arrangements. Likewise, the maximum run length in a binary sequence can also be used to test the randomness of the ordered sequence.

We can think of two different methods of generating a binary data sequence: (i) binomial sampling and (ii) hypergeometric sampling. First, consider a case of 'binomial sampling' in which a fair coin is flipped 6 times. If each trial is statistically independent, we rarely have a sequence, HHHHHH or TTTTTT. The probability of having only one run or, equivalently, the probability of having the maximum run length of 6 is 3.125% (i.e. $0.5^6 + 0.5^6$). Such a small probability is small enough to reject the null hypothesis, at the 0.05 level of significance, that it is a fair coin. It is also extremely rare to have 6 runs (i.e. HTHTHT or THTHTH) in six trials. The probability that the total number of runs is 6 or, equivalently, the probability of having the maximum run length of only 1 is also 3.125%, which is indicative of non-randomness of the binary sequence.

Second, consider a case of hypergeometric sampling in which 4 girls and 4 boys are sitting in a row of eight seats at a movie theater. If all arrangements are equally probable, it is very unusual that the boys and girls are each sitting together (i.e. BBBBGGGG or GGGGBBBB). The probability of two runs or, equivalently, the probability that the maximum run length is 4 is $2/C(8, 4) = 1/35 = 2.857\%$. This small probability confirms the hunch, at the 0.05 level of confidence, that the separations observed are intentional; the boys and girls are not well acquainted with each other. On the other hand, consider another extreme case in which four boys and four girls are mixed perfectly (i.e. BGBGBGBG or GBGBGBGB). The probability of having 8 runs (i.e. the probability that the maximum run length is 1) is also 2.857%. Such an excessive number of runs indicate intentional mixing; there are 4 pairs of boyfriends and girlfriends.

In the paper, we derive the exact probability distributions of (i) the number of runs and (ii) the maximum run length for a sequence of the categorical variable with *more than two* classes (e.g. win, draw, or lose in a series of chess games). Probability distributions are expressed as *recursive* equations, and can be easily implemented with spreadsheet software such as Microsoft Excel. It can be easily shown that each of the traditional run tests for a binary sequence is a special case of our multinomial run tests in which there are only two classes. We also show that our run tests can be applied to several areas in quality control, such as testing changes in process variation, developing multivariate control charts, and comparing the shapes and locations of more than two populations.

In the next section, we classify various run tests into four distinct cases and define several notations that will be used throughout the paper.

2. Classification of run tests

An ordered sequence of n elements of m different types can be generated by two different sampling methods [24]. First, in *multinomial* sampling, each element is obtained from an *infinite* population with a known probability set $\mathbf{p} = \{p_1, p_2, \ldots, p_m\}$, where $p_1 + p_2 + \ldots + p_m = 1$ and m is the number of distinct classes. As in the Bernoulli scheme, each trial is considered an independent, identically distributed random variable that may take on one of the m possible values, with the outcome j occurring with constant probability p_j. For an ordered sequence of size n, the total number of possible permutations with m distinct classes is m^n.

Second, in *hypergeometric* sampling, each element is taken from a *finite* population with given numbers of elements $\mathbf{d} = \{d_1, d_2, \ldots, d_m\}$, where $n = d_1 + d_2 + \ldots + d_m$ and m is the number of distinct classes (e.g. arrange 2 teachers, 3 parents, and 4 students in a row for a school picture). From the finite population, we randomly select one element at a time without replacement and arrange them in chronological order. In such a case, the total number of distinguishable orderings is $n!/(d_1!d_2! \ldots d_m!)$, and we assume that each of the distinct arrangements is equally likely.

As in the formulation of a dynamic programing problem in operations research, we define the *stage* and the *state* as follows: We are said to be at the ith stage ($i = 1, 2, \ldots, n$) when we obtain the ith outcome in multinomial sampling or draw the ith element in hypergeometric sampling. The state j at stage i is the result of the ith outcome or the type of the ith element. The state at each stage is simply represented as an integer number j, where $j = 1, 2, \ldots, m$. Of course, all the m states are mutually exclusive.

As stated in Lou [21], 'the two most commonly used statistics for testing randomness are the total number of runs and the length of longest run.' To test the randomness of an ordered sequence at a given level of significance, we consider those two types of test statistics in the paper. Let z denote the *total number of runs* and let r be the *maximum run length* in a sequence of size n. The sequence is generated by the multinomial sampling with n and $\mathbf{p} = \{p_1, p_2, \ldots, p_m\}$ or by the hypergeometric sampling with $\mathbf{d} = \{d_1, d_2, \ldots, d_m\}$. Based on the types of the test statistic and the sampling method, a run test can be classified into one of the four different cases as shown in Table 1.

In Table 1, π_z^j in Cases 1 and 3 denotes the probability that the 'number of runs' in the sequence is exactly z when the state at stage n is j (i.e. the last element in the ordered sequence of size n is of type j). Because m sub-sequences are mutually exclusive, the probability π_z^* that the total number of runs is z is simply

$$\pi_z^* = \sum_{j=1}^{m} \pi_z^j. \tag{1}$$

Similarly, ρ_r^j in Cases 2 and 4 represents the probability that the 'maximum run length' in the sequence is *at least* r when the state at stage n is j, where $j = 1, 2, \ldots, m$. Therefore, the probability that the maximum run length is at least r in a sequence of multi-state trials is

$$\rho_r^* = \sum_{j=1}^{m} \rho_r^j. \tag{2}$$

Note that the probability that the maximum run length at stage n is exactly r is simply $\rho_r^* - \rho_{r+1}^*$.

Table 1. Classification of run tests.

	Multinomial sampling with n and $\mathbf{p} = \{p_1, p_2, \ldots, p_m\}$	Hypergeometric sampling with $\mathbf{d} = \{d_1, d_2, \ldots, d_m\}$
Number of runs, z	• Case 1. $\pi_z^j(n, \mathbf{p})$	• Case 3. $\pi_z^j(\mathbf{d})$
Maximum run length, r	• Case 2. $\rho_r^j(n, \mathbf{p})$	• Case 4. $\rho_r^j(\mathbf{d})$

Since the early twentieth century, many researchers have derived the probability distributions for each of the four cases with different methods, treating each case separately. As far as we are aware, no researchers have proposed a unified approach that deals with all the four cases in the same manner. In the paper, we propose a recursive formula for each case, all of which are based on the same simple idea that any ordered sequence can be divided into m mutually exclusive sub-sequences. For various $\mathbf{p}$ and $\mathbf{d}$, we programed the recursive equations in Fortran 90 and cross-checked the frequency distributions with Microsoft Excel.

In the next section, we review some of the relevant literature on the theory and applications of run tests in an ordered sequence.

3. Literature review

3.1. Run tests for a binary sequence

According to Mood [24], 'the distribution theory of runs has had a stormy career,' which dates back to the end of the nineteenth century. However, an actual probability distribution of runs was not derived until 1925 when E. Ising considered the total number of runs in hypergeometric sampling. Wald and Wolfowitz [36] published the same distribution and showed that it is approximately a normal distribution. Wishart and Hirshfeld [38] appear to be the first who considered the total number of runs in binomial sampling and proved that it is also asymptotically normal.

Mood [24] proposed combinatorial methods for the number of runs and showed that the limiting distributions are all normal. The length of the longest run in hypergeometric sampling was investigated in Bateman [3]. Barton and David [2] considered the total number of runs in binomial sampling. Marshall [22] proposed an easy way to compute the mean and variance of the total number of runs in hypergeometric sampling. The same results were obtained by Guenther [15] based on the hypergeometric distribution.

Schilling [31] proposed a classroom experiment in which the results of 200 tosses of a fair coin are simulated. He derived a recursive equation for the probability of having *at most s* successes, while Berresford [4] later found the probability of having *at least s* successes in the same problem. Mogull [23] pointed out one exceptional sequence, in which the nonrandomness cannot be detected by the run test in hypergeometric sampling. Run tests in binomial sampling were suggested by Gelman and Glickman [16] as a classroom experiment that had sparked student involvement in undergraduate courses in statistics and probability. For more detailed history of the theory of runs in a binary sequence, readers are referred to Johnson, Kemp, and Kotz [18, pp.454–455].

3.2. Run tests for a sequence with multiple classes

Although the literature in run tests for a binary sequence is extensive, there has been few research efforts devoted to the problem of specifying the run distributions for a sequence of multi-state trials. Mood [24] is one of the few who derived probability distributions for the number of runs for Case 1 and Case 3 in Table 1. Later, Barton and David [1] showed how the distribution of the number of runs can be built up for a sequence of multi-state trials from that for a binary sequence. They obtained reasonably compact expressions for

Case 3 in Table 1. For a sequence of multi-state trials with n and $\mathbf{p}$, Schwager [32] also proposed recurrence relations and obtained the probability that the 'success' run of a given size occurs. Fu and Koutras [14] dealt with the same problem and derived the probability distributions of various 'success' run statistics.

As shown in Table 1, two of the most popular statistics for run tests are (i) the number of runs and (ii) the maximum run length [24]. However, some researchers are interested in a pre-specified pattern of outcomes that may occur at some point in the series of trials. Schwager [32] derived the probability of the occurrence of a subsequence in the series of n trials in multinomial sampling with $\mathbf{p}$. He also discussed a wide variety of the applications of run tests, including DNA sequencing, psychological achievement testing, human and animal behavior, radar astronomical observations, and non-parametric hypothesis testing. Fu [13] proposed a Markov chain imbedding technique and found the distribution of waiting time until a pre-specified subsequence occurs in a sequence of multi-state trials.

3.3. Recursive equation for run tests

As anticipated, the problem of deriving the probability distributions of run statistics is a combinatorial one, and the whole development of run tests depends on some identities in combinatory analysis. Thus, the explicit formula for run distributions are rather complex, and are 'not at first sight very illuminating' [21]. For computational efficiency, the probability distributions of the number of runs are often approximated as normal distributions.

Recently, several authors have proposed a simple *recursion* formula that can be used to generate the exact probability distributions for run tests. For example, Schilling [31] obtained a recursive equation for the longest run in a binary sequence that is generated by binomial sampling. He showed that a spreadsheet software can easily be used to generate the exact probability distribution for any moderate size n. For a similar problem, Riehl [29] derived the distribution of the total number of runs in a fixed number n of throws of a biased coin with p. With his relatively simple recursive formula, it is easy to develop a spreadsheet or to write a computer program that computes the exact distribution for any given n and p.

The distribution theory of run has a long and rich history, but we are not aware of any research efforts devoted to developing a *unified* approach for all the four cases in Table 1 for a sequence of nominal data with more than two classes. In the paper, we develop *recursive* equations for an ordered sequence of more than two classes. The recursive formula can be shown to be more computationally efficient.

3.4. Applications of run tests

Run tests for a binary sequence have been widely used to detect hidden effects in many practical situations. In agriculture, for example, long rows of unhealthy plants in a cornfield may be suggestive of a contagious disease [12, pp. 42–43]. The meteorologist observes successions of similar weather to discover a holdover effect that would be useful in weather forecasting [4].

In sports, unusually long runs of success might indicate that the player or the team has 'hot hands' [34]. The 'random walk' model in the financial market can be tested by counting the number of runs in a sequence of daily closing prices [11]. In political science, the sequence of Democratic and Republican presidents in the White House may reveal some hidden aspects [17].

The sequence of a consumer's shopping choice between different stores has been studied with multinomial run tests [35]. In health care industry, the monthly revenue generated by an ambulatory urgent care clinic is either above or below the target level. Such a binary sequence can be used to predict the 'runs of a weak business' [26].

Run tests have also been widely applied to statistical quality control. In process control, for example, a bottling machine will slightly over-fill or under-fill most bottles of soda, and unusually long runs of one or the other indicate a mal-adjustment of the filling machine [39]. More detailed reviews of applications of run tests to quality control are given at the beginning of Section 6.

4. Multinomial sampling

4.1. Case 1: number of runs

In multinomial sampling with n and $\mathbf{p}$, let $\pi_z^j(n, \mathbf{p})$ be the probability that the total number of runs is z when the state is j at the last stage n. Such an event can be achieved in two different ways. First, the number of runs is already z when the state is j at the previous stage $n-1$. Adding another element of the *same* type j does not change the number of runs z. Second, the number of runs is $z-1$ when the state is anything but j at stage $n-1$. In such a case, adding an element of different type j increases the number of runs from $z-1$ to z.

Thus, we have the following recurrence relations:

$$\pi_z^j(n, \mathbf{p}) = p_j \pi_z^j(n-1, \mathbf{p}) + p_j \sum_{\forall k, k \neq j} \pi_{z-1}^k(n-1, \mathbf{p}), \quad \text{for } z \geq 2 \text{ and } j = 1, 2, \dots, m. \tag{3}$$

The recursive equations can be easily implemented with Microsoft Excel with $m+1$ worksheets. For spreadsheet implementation, it is more convenient to use the following recurrence relations:

$$\pi_z^j(n, \mathbf{p}) = p_j[\pi_z^j(n-1, \mathbf{p}) - \pi_{z-1}^j(n-1, \mathbf{p}) + \pi_{z-1}^*(n-1, \mathbf{p})], \quad \text{for } j = 1, 2, \dots, m. \tag{4}$$

The probability that all the elements in the sequence of size n are of type j or, in other words, the number of runs is 1 is ${p_j}^n$. Thus, the boundary condition for the recursive equation is

$$\pi_z^j(n, \mathbf{p}) = \begin{cases} {p_j}^n & \text{if } z = 1 \\ 0 & \text{if } z \geq n. \end{cases} \tag{5}$$

Using Microsoft Excel spreadsheet, we obtained the frequency distribution of the number of runs z in the multinomial sequence of size $n = 6$ with various probability sets $\mathbf{p}$. When the probability set is $\mathbf{p} = \{.1, .2, .3, .4\}$, for example, the probability that the total number of runs is $z = 4$ is 29.494% as shown in Table 2. The average number of runs in the sequence of size $n = 6$ is 4.5000, and the variance is 1.1300.

Table 2. Frequency distributions of the number of runs z when $n = 6$ in multinomial sampling.

Probability set **p**	Number of runs z 1	2	3	4	5	6	Total	$E[z]$	$Var[z]$
{1/3, 2/3}	65	124	248	184	92	16	729	3.2222	1.4321
{1/6, 2/6, 3/6}	397	1691	5002	7598	6409	2231	23,328	4.0556	1.3117
{.1, .2, .3, .4}	489	3279	13,314	29,494	35,509	17,915	100,000	4.5000	1.1300

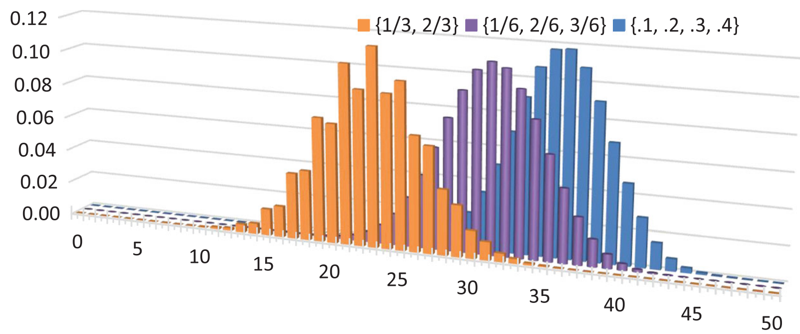

Figure 1. Total number of runs z for various probability sets **p** when $n = 50$.

Suppose that the number of runs z is less than or equal to 2 in the sequence of $n = 6$ and $p = \{.1, .2, .3, .4\}$. Then, the p-value is $(489 + 3279)/100{,}000 = 0.03768$ from Table 2 and we reject the null hypothesis of randomness at the 0.05 level of significance.

Our computation results further revealed that the expected value and the variance of the total number of runs z are linear functions of the number of trials n. When the probability set is $\mathbf{p} = \{.1, .2, .3, .4\}$, for example, the expected value is $E[z] = 0.3 + 0.7n$, and the variance is shown to be $Var[z] = -0.25 + 0.23n$ for $n > 1$.

Figure 1 displays the probability distribution of the total number of runs z in the multinomial sequence of size $n = 50$. When the probability set is $\mathbf{p} = \{.1, .2, .3, .4\}$, for example, the average number of runs is shown to be $E[z] = 35.30$ and the variance is $Var[z] = 11.25$. The bell-shaped curves in Figure 1 indicate that the probability mass function of z can be approximated by a normal distribution with the mean $E[z]$ and the variance $Var[z]$. The normal approximation is more appealing if the probability set **p** is uniformly distributed and/or n is relatively large.

4.2. *Case 2: maximum run length*

In multinomial sampling with n and $\mathbf{p}$, let $\rho_r^j(n, \mathbf{p})$ be the probability that the maximum run length is at least r when the state is j at the last stage n. There are two cases in which the maximum run length is at least r when the state is j at stage n. First, the maximum run length is at least r at the previous stage $n - 1$, and the state at the current stage n is simply j. Second, the maximum run length at stage $n - r$ is less than r, and the state at stage $n - r$ is anything but j. In such a case, the maximum run length becomes r if the states at the last r stages are all j.

Table 3. Frequency distributions of the maximum run length r in the multinomial sequence of size $n = 6$.

Probability	Maximum run length r								
set **p**	1	2	3	4	5	6	Total	$E[r]$	$Var[r]$
{1/3, 2/3}	16	216	236	128	68	65	729	3.2894	1.6515
{1/6, 2/6, 3/6}	2231	11,088	6373	2377	862	397	23,328	2.5603	1.0674
{.1, .2, .3, .4}	17,915	53,216	20,849	5909	1622	489	100,000	2.2157	0.8017

Thus, the recurrence relation is

$$\rho_r^j(n,\mathbf{p}) = p_j\rho_r^*(n-1,\mathbf{p}) + \left((1-p_j) - \sum_{\forall k,k\neq j} \rho_r^k(n-r,\mathbf{p})\right) {p_j}^r, \quad \text{for } j = 1,2,\ldots,m. \tag{6}$$

To implement a spreadsheet program such as Microsoft Excel, it is more convenient to use the following expression:

$$\rho_r^j(n,\mathbf{p}) = p_j\rho_r^*(n-1,\mathbf{p}) + [(1-p_j) - \rho_r^*(n-r,\mathbf{p}) + \rho_r^j(n-r,\mathbf{p})]{p_j}^r. \tag{7}$$

The boundary condition for the recursive equation is

$$\rho_r^j(n,\mathbf{p}) = \begin{cases} {p_j}^n & \text{if } r = n \\ 0 & \text{if } r > n. \end{cases} \tag{8}$$

Table 3 displays the frequency distributions of the maximum run length r in the sequence of size $n = 6$ that is generated by multinomial sampling with a probability set $\mathbf{p}$. For example, when the probability set is $\mathbf{p} = \{.1, .2, .3, .4\}$, the probability that the maximum run length is $r = 4$ is 5.909%.

As shown in Figure 2, the frequency distribution of the maximum run length is skewed to the right, showing a long right tail of large values. When the probability set is $\mathbf{p} = \{1/3, 2/3\}$ and the number of elements is $n = 50$, the expected value and the variance of the maximum run length r are 7.9471 and 7.7178, respectively. When the probability set is $\mathbf{p} = \{.1, .2, .3, .4\}$, the expected value and the variance of the maximum run length decrease to 4.2045 and 1.5567, respectively.

5. Hypergeometric sampling

5.1. Case 3: number of runs

In hypergeometric sampling with $\mathbf{d} = \{d_1, d_2, \ldots, d_m\}$, let $\pi_z^j(\mathbf{d})$ be the probability that the total number of runs is z when the state is j at the last stage $n = d_1 + d_2 + \ldots + d_m$. For notational convenience, let $\pi_z^j(d_k - 1)$ denote the probability that the total number of runs is z when the latest state is j and only the number of elements of type k is reduced by 1; i.e. $\mathbf{d} = \{d_1, d_2, \ldots, d_{k-1}, \ldots, d_m\}$.

When the state at the last stage n is j, the total number of runs will be z in two different cases. First, the state is j at the previous stage $n-1$ and the number of runs is already z. Adding another element of the same type j at the current stage n does not change the

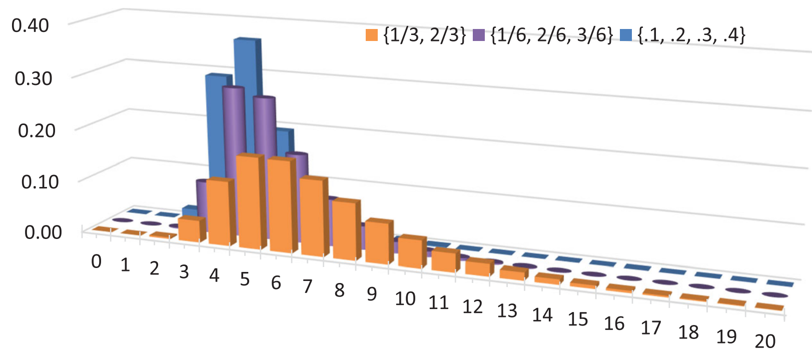

Figure 2. Maximum run length r for various probability sets $\mathbf{p}$ when $n = 50$.

Table 4. Frequency distributions of the number of runs z in hypergeometric sampling with various $\mathbf{d}$.

Element	Number of runs z											
set, $\mathbf{d}$	1	2	3	4	5	6	7	8	9	10	Total	$E[z]$	$Var[z]$
{1, 2}	0	2	1	0	0	0	0	0	0	0	3	2.3333	0.2222
{1, 2, 3}	0	0	3	9	13	5	0	0	0	0	30	4.6667	0.7556
{1, 2, 3, 4}	0	0	0	4	36	170	455	695	561	179	2100	8.0000	1.3333

number of runs z. Second, the number of runs is $z-1$ when the state is anything but j at the previous stage $n-1$. In such a case, adding an element of type j increases the number of runs from $z-1$ to z at the current stage n.

Thus, we have the following recurrence relation:

$$\pi_z^j(\mathbf{d}) = \frac{d_j}{n}\pi_z^j(d_j - 1) + \frac{d_j}{n}\sum_{\forall k, k \neq j} \pi_{z-1}^k(d_k - 1), \quad \text{for } j = 1, 2, \ldots, m. \tag{9}$$

For a spreadsheet program, we may simply use the following recursive equation:

$$\pi_z^j(\mathbf{d}) = \frac{d_j}{n}[\pi_z^j(d_j - 1) - \pi_{z-1}^j(d_j - 1) + \pi_{z-1}^*(d_j - 1)], \quad \text{for } j = 1, 2, \ldots, m. \tag{10}$$

The number of runs is 1 if all the elements in the sequence of size n are of the same type j. Thus, the boundary condition is

$$\pi_r^j(\mathbf{d}) = \begin{cases} 1 & \text{if } d_j = n \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

Using Microsoft Excel, we found the frequency distributions of the number of runs z for various cases of hypergeometric sampling and reported the results in Table 4. Consider an ordered sequence of 1 blue ball, 2 red balls, 3 green balls, and 4 yellow balls. We know that there are 10! / (1! 2! 3! 4!) = 12,600 possible arrangements. When the element set is $\mathbf{d} = \{1, 2, 3, 4\}$, the possible number of runs is between 4 and 10. As shown in Table 4, the average number of runs is 8.000 and the variance is 1.333. Note that the order of d_j in the element set $\mathbf{d}$ does not make a difference in the frequency distribution.

5.2. Case 4: maximum run length

As in the multinomial sampling, let $\rho_r^j(\mathbf{d})$ be the probability that the maximum run length is at least r when the state is j at the last stage $n = d_1 + d_2 + \ \ldots \ + d_m$. As in Case 2, there are two ways in which the maximum run length is at least r when the state is j at stage n.

First, the maximum run length is at least r at the previous stage $n-1$, and the state at the current stage n is simply j. Second, the maximum run length at stage $n-r$ is less than r, and the state at stage $n-r$ is anything but j. However, the states at the next r stages are all j consecutively.

Thus, the recurrence relation for the maximum run length r is

$$\rho_r^j(\mathbf{d}) = \frac{d_j}{n}\rho_r^*(d_j-1) + \left[\frac{n-d_j}{n-r} - \sum_{\forall k,\, k\neq j} \rho_r^k(d_k - r)\right]\binom{n}{r}^{-1}\binom{d_j}{r}. \tag{12}$$

The following recurrence relation is easier to implement with spread sheet software:

$$\rho_r^j(\mathbf{d}) = \frac{d_j}{n}\rho_r^*(d_j-1) + \left[\frac{n-d_j}{n-r} - \rho_r^*(d_j - r) + \rho_r^j(d_j - r)\right]\binom{n}{r}^{-1}\binom{d_j}{r},$$
$$\text{for } j = 1, 2, \ldots, m. \tag{13}$$

The boundary condition can be shown to be

$$\rho_r^j(\mathbf{d}) = \begin{cases} 0 & \text{if } d_j = n < r \\ 1 & \text{if } d_j = n \geq r. \end{cases} \tag{14}$$

For various cases of hypergeometric sampling, we found the frequency distributions of the maximum run length r with spreadsheet software. The results are reported in Table 5. When the element set is $\mathbf{d} = \{1, 2, 3, 4\}$, for example, the possible value of the maximum run length is between 1 and 4. As shown in Table 5, the average run length is 2.2243 and its variance is 0.4111.

We proposed a unified approach to four different types of run tests in Table 1 that are all based on *recursion* relations. The unified approach, which is computationally much more efficient, can be easily implemented for practitioners with spreadsheet software such as Microsoft Excel. Our run tests for a sequence with multiple classes are also more general than the traditional run tests for a binary sequence.

Table 5. Frequency distributions of the maximum run length r in hypergeometric sampling with various **d**.

Element	Maximum run length r							
set, **d**	1	2	3	4	5	Total	E[r]	Var[r]
{1, 2}	1	2	0	0	0	3	1.6667	0.2222
{1, 2, 3}	5	19	6	0	0	30	2.0333	0.3656
{1, 2, 3, 4}	179	1341	510	70	0	2100	2.2243	0.4111

6. Run tests for statistical quality control

Run tests that can detect any non-randomness in an ordered sequence have been successfully applied to many practical situations such as statistical quality control. In the section, we first review various applications of run tests to quality control, and then focus on three potential areas of applications of our generalized run tests – such as testing changes in process variation, developing non-parametric multivariate control charts, and comparing the shape and location of more than two production processes.

6.1. Applications to quality control

As pointed out by Chakraborti [7], most process control charts are based on the assumption of a specific form of a probability distribution, such as the normal distribution. In many practical applications, however, we often do not enough data to justify this assumption, and *non-parametric* control charts have been used particularly for short production runs [40]. In such a job-shop production environment, we don't have enough measurements to estimate the process distribution and its parameter values. With not enough sample observations, traditional control charts based on three-sigma control limits are not appropriate.

As an alternative, we may simply observe if the measurements are above or below the target level, and test the randomness of the binary sequence [25] and [30]. For example, the room temperature measured at every hour could be above (+) or below (-) the target level. Based on the binary sequence (+ + - - + ... + - +), we can count (i) the total number of runs and (ii) the maximum run length, and test if the process is 'in control' or 'out of control'. Likewise, 'up' and 'down' can be used to form a binary sequence and to detect the presence of any assignable causes of variability in the quality of manufactured product [41].

We can also observe the sequence of defective and non-defective items from a manufacturing process, and test its randomness. In service quality management, we may consider the binary sequence of 'late' and 'on time' flight arrivals, 'satisfactory' and 'unsatisfactory' customer feedbacks, and the like, and then test for unlikely conditions indicative of any special causes. As shown in Bonnini, Corain, and Salmaso [6], non-parametric methods can be also used to support the development of new, successful industrial products on the basis of their experimental performances.

Many researchers, including Qui and Li [27] and [28], have shown that non-parametric control charts perform much better than the traditional parametric control charts when observed data are not normally distributed. For an extensive review of the literature on non-parametric or distribution-free control charts for univariate variable data, readers are referred to Chakraborti, Van Der Laan, and Bakir [8] and Chakraborti, Qiu, and Mukherjee [9].

6.2. Process variation

Monitoring multivariate production processes is one of the important and challenging problems in statistical process control [10]. For example, Li, Zou, Wang, and Huwang [20] developed a new multivariate nonparametric method for monitoring shape parameters

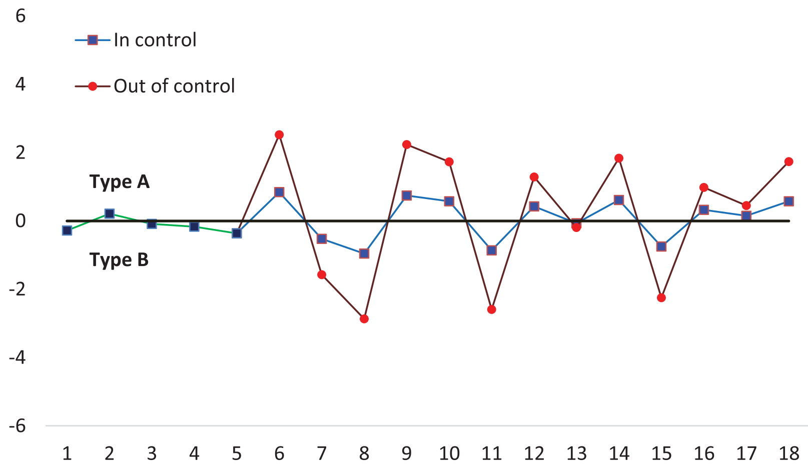

Figure 3. Run test for a categorical variable with two classes {A, B}.

of a production process. Their method is based on a powerful multivariate spatial-sign test and exponentially weighted moving average (EWMA) control scheme. Recently, Li, Xie, and Zhou [19] proposed a similar method that can be used to monitor the location and scale parameters of a univariate distribution. Zhang, Li, and Li [42] developed a new distribution-free control chart by integrating the nonparametric two-sample Cramér-von Mises test and the exponentially weighted moving average control scheme.

The generalized run tests we proposed in the paper can be easily used to monitor any changes in process variation. Consider, for example, the two production processes in Figures 3 and 4. The original process, which is 'in control', is generated with normal random numbers with $\mu = 0$ and $\sigma = 1$. After time $t = 5$, we purposefully increased the process standard deviation to $\sigma = 3$, with the same process mean $\mu = 0$, to represent an 'out of control' status. According to Figure 3 with only two classes {A, B}, the sequences of runs in the two production processes are exactly the same; i.e. {BABBBA ... BAAA}. With traditional run tests developed for the binary sequence, we fail to detect the change in process variation after time $t = 5$.

In Figure 4, on the other hand, the quality characteristic is discretized into three classes {A, B, C}. As shown in Figure 4, the 'out of control' process has a different sequence of runs than the 'in control' process has. A fewer number of Type B observations after time $y = 5$ is a clear indication that the process variation has been increased.

6.3. *Multivariate control chart*

Statistical process control charts have been widely used in industry for monitoring the stability of multivariate processes. For example, Zhou, Zi, Geng, and Li [43] recently developed a new non-parametric procedure that integrates a two-sample multivariate run test and a change-point method.

Another application of our generalized run tests in quality control is non-parametric multivariate control charts for short production run [5]. Consider a bivariate production process as shown in Figure 5. To represent the heights (X_i) and the weights (Y_i) of 100 items,

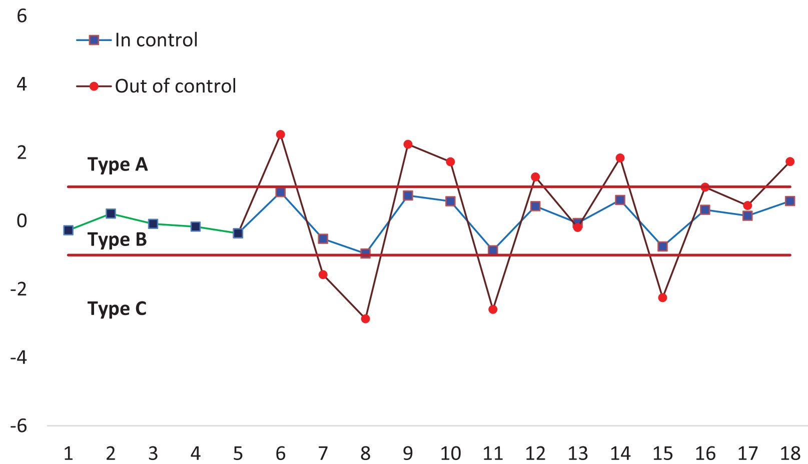

Figure 4. Run test for a categorical variable with three classes {A, B, C}.

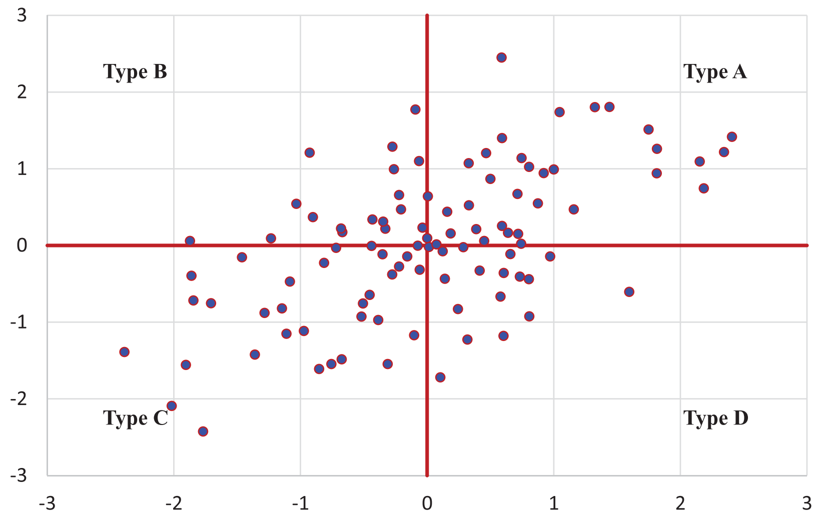

Figure 5. Run test for a categorical variable with four classes {A, B, C, D}.

we generated bivariate normal random numbers with the correlation coefficient $\rho = +0.6$. If we have enough sample observations, we can estimate the variance-covariance matrix of the bivariate process and use the Hotelling's T^2 control chart to monitor the mean vector of the process.

If the sample observations are not enough as in short production run, we may consider a non-parametric control chart. Based on the target values of the height and the weight, the two-dimensional space can be divided into four quadrants {A, B, C, D}. From the ordered sequence of A, B, C, and D, we can count the number of runs and the maximum run length. Based on those test statistics, we can test the randomness of the multivariate sequence and detect any shifts in the process means and variations.

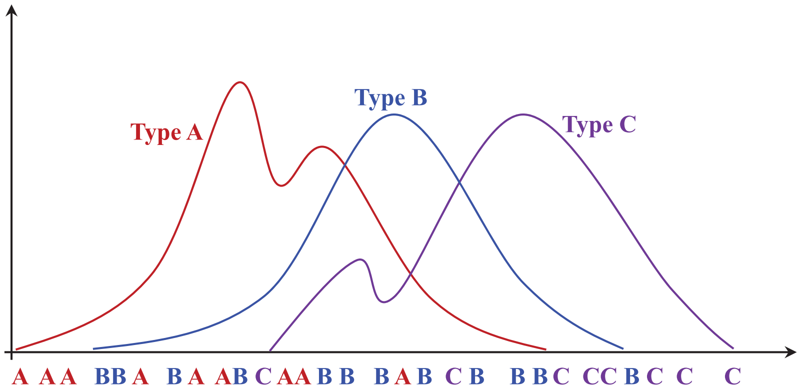

Figure 6. Run test for a categorical variable with four classes {A, B, C, D}.

6.4. Comparison of more than two populations

For small samples, Wald and Wolfowitz [36] used a run test to test if two samples have come from the same continuous distribution. They used the number of runs in the binary sequence as a test statistic. We can simply extend their idea to the case of more than two populations. The procedure is as follow.

First, take independent random samples of sizes $n_1, n_2, \ldots$, and n_m from m populations (e.g. m machines). Then, combine all the sample observations and arrange the combined data in ascending order. For the ordered sequence of m classes, we can identify (i) the number of runs and (ii) the maximum run length, and use them as test statistics.

As shown in Figure 6, for example, if the total number of runs is too small or the maximum run length is too large, we have sufficient evidence that three process distributions are not identical (i.e. the locations and the shapes of the population distributions are different.)

7. Concluding remarks

In this paper, we proposed several types of run tests for the sequence of outcomes or elements of *more than two* kinds. In a multinomial process with three types of elements, for example, each of the three different elements {A, B, C} is presented with probabilities, p_1, p_2, and p_3. In a hypergeometric process with four types of elements of size d_1, d_2, d_3, and d_4, we assume that all the elements {A, B, C, D} are mixed together and arranged in random order. In both cases, we can use the number of runs and the maximum run length to test the randomness of the ordered sequence of outcomes and elements of more than two kinds.

The theory of runs has grown markedly in popularity and in a variety of applications. Various run tests have been used to check for randomness in a sample distribution, to examine a distribution of regression residuals for non-randomness, to test for both trends and cyclical patterns with temporal data, and to test the association between two variables. Although a run test is known as a ‘quick and dirty’ method, it is popular due to its simplicity of use and its versatility in a wide variety of applications.

One of the popular application areas is statistical process control [37] and sampling inspection [33]. Most job-shops are characterized by short production runs, and many of these shops produce parts on production runs of less than 50 units. In such a job-shop manufacturing environment, the routine use of traditional control charts and hypothesis testing methods appear to be somewhat of a challenge, as not enough units are produced in any one batch to estimate the population distribution and its parameter values.

In such a job-shop environment with short production runs, practitioners can use our non-parametric run tests developed for nominal data with more than two classes. Furthermore, our unified approach that is based on the recursive equations can be easily implemented with any spreadsheet software such as Microsoft Excel.

Acknowledgements

I'd like to express my gratitude for the financial support from *Cherie H. Flores Endowed Chair of MBA Studies* at Louisiana State University. I'd also like to express my sincere thanks to an anonymous referee and the associate editor for their thorough and knowledgeable reviews that helped me improve the original manuscript significantly.

Disclosure statement

No potential conflict of interest was reported by the author.

References

[1] D.E. Barton, F.N. David, *Multiple runs*, Biometrika 44 (1957), pp. 168–178.
[2] D.E. Barton, F.N. David, *Non-randomness in a sequence of two alternatives: II. Runs test*, Biometrika 45 (1958), pp. 253–256.
[3] G. Bateman, *On the power function of the longest run as a test for randomness in a sequence of alternatives*, Biometrika 35 (1948), pp. 97–112.
[4] G.C. Berresford, *Runs in coin tossing: randomness revealed*, Coll. Math. J. 33 (2002), pp. 391–394.
[5] S. Bersimis, S. Psarakis, J. Panaretos, *Multivariate statistical process control charts: an overview*, Qual. Reliab. Eng. Int. 23 (2007), pp. 517–543.
[6] S. Bonnini, L. Corain, L. Salmaso, *A new statistical procedure to support industrial research into new product development*, Qual. Reliab. Eng. Int. 22 (2006), pp. 555–566.
[7] S. Chakraborti, *Nonparametric (distribution-free) quality control charts*, Encycl. Stat. Sci. (2011), pp. 1–27. doi:10.1002/9781118445112.stat02699.
[8] S. Chakraborti, P. van der Laan, S.T. Bakir, *Nonparametric control charts: an overview and some results*, J. Qual. Technol. 33 (2001), pp. 304–315.
[9] S. Chakraborti, P. Qiu, A. Mukherjee, *Editorial to the special issue: nonparametric statistical process control charts*, Qual. Reliab. Eng. Int. 31 (2015), pp. 1–2.
[10] N. Chen, X. Zi, C. Zou, *A distribution-free multivariate control chart*, Technometrics 58 (2016), pp. 448–459.
[11] P.F. Christoffersen, F.X. Diebold, *How relevant is volatility forecasting for financial risk management?* Rev. Econ. Stat. 82 (2000), pp. 12–22.
[12] W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 1, John Wiley and Sons, New York, 1968.
[13] J. Fu, *Distribution theory of runs and patterns associated with a sequence of multi-state trials*, Satat. Sin. 6 (1996), pp. 957–974.
[14] J.C. Fu, M.V. Koutras, *Distribution theory of runs: a Markov chain approach*, J. Am. Stat. Assoc. 89 (1994), pp. 1050–1058.
[15] D.J. Gans, *Persistence of party success in American presidential elections*, J. Interdiscip. Hist. 16 (1985), pp. 221–237.

[16] A. Gelman, M.E. Glickman, *Some class-participation demonstrations for introductory probability and statistics*, J. Educ. Behav. Stat. 25 (2000), pp. 84–100.
[17] W.C. Guenther, *Some remarks on the runs test and the use of the hypergeometric distribution*, Am. Stat. 32 (1978), pp. 71–73.
[18] N.L. Johnson, A.W. Kemp, S. Kotz. *Univariate Discrete Distributions*, 3rd ed., John Wiley & Sons, New York, 2005.
[19] Z. Li, M. Xie, M. Zhou, *Rank-based EWMA procedure for sequentially detecting changes of process location and variability*, Qual. Technol. Quant. Manage. 15 (2018), pp. 354–373.
[20] Z. Li, C. Zou, Z. Wang, L. Huwang, *A multivariate sign chart for monitoring process shape parameters*, J. Qual. Technol. 45 (2013), pp. 149–165.
[21] W.Y.Q. Lou, *On runs and longest run tests: a method of finite Markov chain imbedding*, J. Am. Stat. Assoc. 91 (1996), pp. 1595–1601.
[22] C.W. Marshall, *A simple derivation of the mean and variance of the number of runs in an ordered sample*, Am. Stat. 24 (1970), pp. 27–28.
[23] R.G. Mogull, *The one-sample runs test: a category of exception*, J. Educ. Behav. Stat. 19 (1994), pp. 296–303.
[24] A.M. Mood, *The distribution theory of runs*, Ann. Math. Stat. 11 (1940), pp. 367–392.
[25] F. Mosteller, *Note on an application of runs to quality control charts*, Ann. Math. Stat. 12 (1941), pp. 228–232.
[26] L.A. Moyé, A.S. Kapadia, I.M. Cech, *A prediction model for outpatient visits and cash flow: developments in prospective run theory*, Statistician 39 (1990), pp. 399–414.
[27] P. Qiu, Z. Li, *Distribution-free monitoring of univariate processes*, Stat. Probab. Lett. 81 (2011), pp. 1833–1840.
[28] P. Qiu, Z. Li, *On nonparametric statistical process control of univariate processes*, Technometrics 53 (2011), pp. 390–405.
[29] G. Riehl, *More computer generated thinking*, Teach. Stat. 16 (1994), pp. 9–11.
[30] B. Saperstein, *On the occurrence of n successes within N Bernoulli trials*, Technometrics 15 (1973), pp. 809–818.
[31] M.F. Schilling, *The longest run of heads*, Coll. Math. J. 21 (1990), pp. 196–207.
[32] S.J. Schwager, *Run probabilities in sequences of Markov-dependent trials*, J. Am. Stat. Assoc. 78 (1983), pp. 168–175.
[33] G. Shmueli, A. Cohen, *Run-related probability functions applied to sampling inspection*, Technometrics 42 (2000), pp. 188–202.
[34] A. Tversky, T. Gilovich, *The cold facts about the "hot hands" in basketball*, Chance 2 (1989), pp. 16–21.
[35] W.R. Vanhonacker, *Testing the exact order of an individual's choice process in an information-theoretic framework*, J. Marketing Res. 22 (1985), pp. 377–387.
[36] A. Wald, J. Wolfowitz, *On a test whether two samples are from the same population*, Ann. Math. Stat. 11 (1940), pp. 147–162.
[37] P. Walker, *Non-parametric methods in quality control*, Statistician 22 (1973), pp. 59–67.
[38] J. Wishart, H.O. Hirshfeld, *A theorem concerning the distribution of joins between line segments*, J. London Math. Soc. 11 (1936), p. 227.
[39] J. Wolfowitz, *On the theory of runs with some applications to quality control*, Ann. Math. Stat. 14 (1943), pp. 280–288.
[40] C.M. Wright, D.E. Booth, M.Y. Hu, *Joint estimation: SPC method for short-run autocorrelated data*, J. Qual. Technol. 33 (2001), pp. 365–378.
[41] D. Xiang, F. Tsung, X. Pu, *Statistical process control for latent quality characteristics using the up-and-down test*, Technometrics 59 (2017), pp. 496–507.
[42] J. Zhang, E. Li, Z. Li, *A Cramér-von Mises test-based distribution-free control chart for joint monitoring of location and scale*, Comput. Ind. Eng. 110 (2017), pp. 484–497.
[43] M. Zhou, X. Zi, W. Geng, Z. Li, *A distribution-free multivariate change-point model for statistical process control*, Commun. Stat. Simul. Comput. 44 (2015), pp. 1975–1987.