Journal of Applied Mathematics & Bioinformatics, vol.5, no.3, 2015, 139-154
ISSN: 1792-6602 (print), 1792-6939 (online)
Scienpress Ltd, 2015

Run and Scan statistics models and their applications in transposition systems and networks

Ioannis S. Triantafyllou[1]

Abstract

An intrinsic generalization of the runs principle comes out if instead of focusing strictly on fixed-length sequences with all their positions occupied by successes, we allow the occurrence of a prespecified (usually small) number of failures. Consequently, our study searches out subsequences of consecutive trials which embraces a prespecified proportion (usually large) of successes. Such a configuration, is traditionally called *scan* or *almost perfect run*. In the present article, we study the waiting time until the first appearance of a scan of type r/k in a sequence of n Bernoulli trials, while several recurrence relations for the calculation of probabilities relative to it are deduced. For illustration purposes, we provide numerical results and applications that shed light on interesting aspects of scan statistics modeling.

Keywords: Almost perfect run; Scan statistic; Consecutive type systems; Transposition systems

---

[1] Department of Statistics and Insurance Science, University of Piraeus 80 Karaoli and Dimitriou Str., 18534 Piraeus, Greece. E-mail: itrantal@unipi.gr

Article Info: *Received* : November 12, 2014. *Revised* : February 15, 2015.
Published online : September 15, 2015.

1 Introduction

A commonly understood nontechnical meaning of the term *run* is an uninterrupted sequence. More specifically, referring to an experiment involving different elements, a *run* of a certain type of element is a consecutive sequence of such elements bordered at each end by other types of outcomes. It is straightforward that additional statistics can be defined based on runs. For example, one may consider the maximal or the minimal run length in a sequence of *n* outcomes.

It is true that runs and relative issues have already attracted much research attention. As early as 1940, Wald and Wolfowitz used runs to establish a two sample test that is intuitively simple and examines whether two independent populations follow a common distribution or not. Since then, a variety of different applications of runs and run-based statistics have been appeared in the literature. For example in the field of Statistical Quality Control, Deming (1972) has prescribed a process monitoring scheme in which the maximal run length above and below the median respectively is considered.

Following a parallel approach, number of runs and run lengths has also been used to provide distribution-free tests for asymmetry of the population that is under investigation, see e.g. Modarres and Gastwirth (1998). Moreover, Balakrishnan and Ng (2001) provided a nonparametric statistical run-based test that compares two independent populations, where only few early failures of elements are needed to be observed.

Runs statistics models have also been utilized in constructing start-up demonstration tests. A start-up demonstration test is a mechanism by which the quality of an equipment (such as batteries or fire alarm systems) is evaluated by means of successful start-ups. Hahn and Gage (1983) introduced a start-up demonstration test in which an equipment (unit) that is under investigation, is accepted if a number of consecutive successful start-ups is accomplished (CS model). Moreover, Gera (2004) ignored the number of failures in testing process and proposed accepting the equipment if a number of consecutive successful start-ups or a total number of successes is met (CSTS model). Balakrishnan and Chan (2000) suggested that an equipment should be accepted if a number of consecutive successful start-ups is observed before a certain number of failures; otherwise the equipment should be rejected (CSTF model). It is noteworthy that additional generalizations, modifications of the aforementioned start-up demonstration tests have been introduced in the literature (see, e.g. Eryilmaz and Chakraborti (2008), Balakrishnan *et al*. (2014) or Yalcin and Eryilmaz (2012)).

In addition, runs and more precisely the length of the longest run play an important role in Statistical Theory of Reliability. Indeed, the well-known consecutive k-out-of-n: F system, which fails if and only if at least k consecutive components fail, may be studied by using run statistics modeling. The failure criterion of the aforementioned structure calls for a run whose length is equal to or greater than k (see, e.g. Triantafyllou & Koutras (2008)), while many generalizations can be approached by applying more complicated run rules (see, e.g. Eryilmaz *et al*. (2011) or Triantafyllou and Koutras (2014)).

Let us next consider an experimental trial whose outcomes can be classified into two categories. It is of some interest to investigate whether reasonable criteria providing evidence of clustering of any of the two categories could be established. If this statement can be proved true, then these criteria could be used to detect changes in the underlying process which generates the series of outcomes. Let $X_1, X_2, ..., X_n$ denote a sequence of $n = lm$ $(m \geq 2, l > 1)$ binary outcomes. A common criterion indicates the division of n trials into l disjoint groups of m consecutive trials each and count of the number of successes within each group. The existence of a large number of successes inside any group (say k or more) leads to the conclusion that the underlying process has been moved from its initial state. In case where k is close to m, the corresponding group constitutes an *almost perfect run of successes* or alternatively a scan of type k/m. Scan statistics models find interesting applications in many scientific fields, such as Biology, Statistical

Theory of Reliability or Statistical Quality Control. For example, in order to develop quantitative measures for assessing and interpreting genomic inhomogeneities between different species, molecular biologists compare their DNA sequences and look for long aligned subsequences that match in most of their positions. Apparently, an unusually long match, namely the occurrence of an *almost perfect run*, offers a strong evidence of similarity between subjects that are under investigation.

It is notable that scans modeling is strongly connected to the reliability study of several well-known structures. For example, let us consider the *r*-within-consecutive-*k*-out-of-*n*: *F* system (see, e.g. Triantafyllou and Koutras (2011)), which fails if and only if there exists *k* consecutive components which include among them at least *r* failed ones ($1 \leq r \leq k \leq n$). In terms of scan statistics, the system fails if and only if an almost perfect run of failed components (or equivalently a scan of type *r*/*k*) occurs.

The distribution theory of runs has been mostly developed based on outcomes of Bernoulli trials (independent and identically distributed cases). However, these approaches can be generalized by relaxing the assumption of independence or even the identically distributed part. As already mentioned, let $X_1, X_2, ..., X_n$ denote a sequence of binary trials, each resulting in either a success ($X_i = 1$) or a failure ($X_i = 0$). An immediate generalization of the runs principles arises if instead of looking at fixed-length strings with all their trials resulted in success, one may allow the occurrence of a prespecified maximum number of failures. The term *scan of type k/m* refers to subsequences $X_i, X_{i+1}, ..., X_{i+j-1}$ of length $j \leq m$ such that the number of successes included therein is at least *k*, namely $\sum_{h=i}^{i+j-1} X_r \geq k$. Let $T_k^{(m)}$ denote the waiting time until a scan of type *k*/*m* occurs for the first time. It goes without saying that

$$T_k^{(m)} = \min\left\{ n : \sum_{j=\max(n-m+1,1)}^{n} X_j \geq k \right\} \tag{1.1}$$

and since $X_j = 0$ for all $j \leq 0$, the following ensues (for more details, see Balakrishnan and Koutras (2002))

$$T_k^{(m)} = \min\left\{ n : \sum_{j=n-m+1}^{n} X_j \geq k \right\}. \tag{1.2}$$

Another random variable which is closely related to $T_k^{(m)}$, is the maximum number of successes contained in a moving window of length *m* over a sequence of fixed number of outcomes $X_1, X_2, ..., X_n$, namely

$$S_{n,m} = \max\left\{ \sum_{j=i}^{i+m-1} X_j : 1 \leq i \leq n-m+1 \right\} . \tag{1.3}$$

The variable defined in the last equation, is called *scan statistic*. It is straightforward that since the events $T_k^{(m)} \leq n$ and $S_{n,m} \geq k$ are equivalent, the following holds true

$$P(T_k^{(m)} \leq n) = P(S_{n,m} \geq k) \tag{1.4}$$

For a detailed presentation of scan statistics and their distribution properties, the interested reader is referred to the monograph of Glaz *et al.* (2001).

For illustration purposes, let us consider the next sequence of outcomes

$$S\ S\ S\ F\ F\ F\ F\ S\ S\ F\ S\ S\ F\ S\ S\ S$$

Then

$$T_2^{(3)} = 2,\ T_3^{(4)} = 3,\ T_4^{(5)} = 12,\ T_7^{(9)} = 16$$

$$S_{16,3} = 3,\ S_{16,4} = 3,\ S_{16,5} = 4,\ S_{16,6} = 5,\ S_{16,7} = 5,\ S_{16,8} = 6,\ S_{16,9} = 7 .$$

The waiting time $T_r^{(k)}$ is closely related to the *r*-within-consecutive-*k*-out-of-*n*: *F* structure defined previously. More specifically, if the event $X_i = 1$ represents the failure of the *i*-th component of the aforementioned reliability structure, while $X_i = 0$ expresses its good functioning, then the system's reliability is given as

$$P(T_r^{(k)} > n) = P(S_{n,k} < r) .$$

Based on the above equality, one may exploit well-known results for the *r*-within-consecutive-*k*-out-of-*n*: *F* structure in order to reach conclusions referring to the distribution of the waiting time $T_r^{(k)}$.

The rest of the paper is organized as follows. In Section 2, we prove general results for the waiting time $T_r^{(k)}$. More specifically, we prove recurrence relations for the calculation of probabilities, that are relevant to the variable $T_r^{(k)}$, with respect to the design parameters r, n, k. Finally, in Section 3 we display several numerical results based on the propositions established earlier, while an application in transposition systems is illustrated in detail.

2 General results

Let $X_1, X_2, ..., X_n$ be a sequence of Bernoulli trials each resulting in either a *success* ($X_i = 1$) or a *failure* ($X_i = 0$). In this section, we study the waiting time for the first occurrence of a scan of type *r*/*k*, namely the time until *k* consecutive trials appear such that the number of successes contained therein is at least *r* ($r \le k$). As already mentioned, the random variable $T_r^{(k)}$ given in (1.1), describes the above scan waiting time. In case of independent and identically distributed trials, the distribution properties of $T_r^{(k)}$ have been well-studied in the literature (see, e.g. Balakrishnan and Koutras (2002) or Bersimis *et al.* (2014)). In the sequel, we study the waiting time $T_r^{(k)}$ under the exchangeability assumption, meaning that the trials have identical distributions, but they are not necessarily independent, that is they may affect one another.

If $X_{i:n}$ ($1 \le i \le n$) denotes the $i-$th order statistic of the sample $X_1, X_2, ..., X_n$, it goes without saying that the following quantity

$$p_i(r,k,n) = P(T_r^{(k)} = X_{i:n}) \tag{2.1}$$

expresses the probability that a scan of type r/k occurs for the first time upon the event $X_{i:n}$. In words, $p_i(r,k,n)$ demonstrates the prospect that under the accomplishment of n Bernoulli trials, a subsequence of k uninterrupted trials such that the number of successes actualized therein is equal or greater than r, occurs.

Let us next denote by $c_i(r,k,n)$ the number of sets with exactly i successes (among the n trials) that result in the appearance of a scan of type r/k for the first time. We next prove recurrence relations for the calculation of probability $p_i(r,k,n)$ for the special case $r=2$ with respect to design parameters k, n and i.

Proposition 2.1. Let $X_1, X_2, ..., X_n$ be a sequence of exchangeable Bernoulli trials. The probability $p_i(2,k,n)$ that the appearance of a scan of type $2/k$ for the first time is carried out upon the occurrence of $X_{i:n}$, satisfies the following recurrence relation with respect to i

$$\begin{aligned}(n)_{i+1}[p_{i+1}(2,k,n)-p_i(2,k,n)] = 2(n-i)(n-(k-1)(i-1))_i - (n-i(k-1))_{i+1} \\ -(n-i)(n-i+1)(n-(k-1)(i-2))_{i-1}\end{aligned} \tag{2.2}$$

for all values $i=1,2,...,n$ and $k \le n$, while

$$(n)_i = \frac{n!}{(n-i)!}.$$

Proof. The probability $p_i(r,k,n)$ is related to the quantities $c_i(r,k,n)$ defined earlier, through the system of equations

$$p_i(r,k,n) = \binom{n}{n-i+1}^{-1} c_{n-i+1}(r,k,n) - \binom{n}{n-i}^{-1} c_{n-i}(r,k,n) \qquad i=1,2,...,n \,. \tag{2.3}$$

We can easily observe that the number of sets $c_i(r,k,n)$ for the special case $r=2$ can be expressed as

$$c_i(2,k,n) = \binom{2n - i(k+1) - k(n-1) + 1}{n-i}. \tag{2.4}$$

(note that an equivalent expression has been already appeared in Naus (1968)). We then substitute the last formula in (2.3) and the following ensues

$$p_{i+1}(2,k,n) - p_i(2,k,n) =$$

$$= \frac{1}{n!}\left\{\frac{2(n-i)!(n-(k-1)(i-1)}{(n-k(i-1)-1)!} - \frac{(n-i-1)!(n-ik+i)!}{(n-ki-1)!} - \frac{(n-i+1)!(n-ki+2k+i-2)!}{(n-ki+2k-1)!}\right\}.$$

Finally, we result in equation (2.2) after some algebraic manipulations. □

Proposition 2.2. Let $X_1, X_2, ..., X_n$ be a sequence of exchangeable Bernoulli trials. The probability $p_i(2,k,n)$ that the appearance of a scan of type 2/*k* for the first time is carried out upon the occurrence of $X_{i:n}$, satisfies the following recurrence relation with respect to *n*

$$\begin{aligned} n!(n-1)_i[p_i(2,k,n) - p_i(2,k,n-1)] \\ = (n-i(k-1)+k-2)_i + (n-i)(n-i+1)(n-(k-1)(i-2))_{i-1} \\ -(n-i)[(n-k(i-2)+i-3)_{i-1} + (n-(i-1)(k-1))_i] \end{aligned} \tag{2.5}$$

for all values $i = 1,2,...,n$ and $k \le n$, while

$$(n)_i = \frac{n!}{(n-i)!}.$$

Proof. Since the probability $p_i(r,k,n)$ can be expressed via equation (2.3), we recall relation (2.4) and the following equality is derived for the special case $r = 2$

$$p_i(2,k,n) - p_i(2,k,n-1) =$$

$$= \frac{1}{n!(n-1)!}\left\{\frac{(n-i-1)!(n-k(i-1)+i-2)!}{(n-k(i-1)-2)!} + \frac{(n-i+1)!(n-k(i-2)+i-2)!}{(n-k(i-2)-1)!}\right.$$

$$\left. - \frac{(n-i)!(n-k(i-2)+i-3)!}{(n-k(i-2)-2)!} - \frac{(n-i)!(n-k(i-1)+i-1)!}{(n-k(i-1)-1)!}\right\}.$$

The desired result is derived after straightforward algebraic maneuvering. □

Proposition 2.3. Let $X_1, X_2, ..., X_n$ be a sequence of exchangeable Bernoulli trials. The probability $p_i(2,k,n)$ that the appearance of a scan of type 2/*k* for the first time is carried out upon the occurrence of $X_{i:n}$, satisfies the following recurrence relation with respect to *n* and *i*

$$\begin{aligned}(n)_i(n+1)_i[p_{i-1}(2,k,n+1)-p_i(2,k,n)] \\ = (n+1)_i(n-(i-1)(k-1))_i-(n-i+1)(n+1)_i(n-(k-1)(i-2))_{i-1} \\ -(n-i+2)[(n-k(i-2)+i-1)_{i-1}-(n-i+3)(n-k(i-3)+i-2)_{i-2}]\end{aligned} \tag{2.6}$$

for all values $i=1,2,...,n$ and $k \le n$, while

$$(n)_i = \frac{n!}{(n-i)!}.$$

Proof. Recalling equations (2.3) and (2.4), we may write the following recurrence with respect to design parameters *n* and *i*

$$\begin{aligned}p_{i-1}(2,k,n+1)-p_i(2,k,n) \\ = \frac{1}{n!}\left\{\frac{(n-i)!(n-k(i-1)+i-1)!}{(n-k(i-1)-1)!}-\frac{(n-i+1)!(n-k(i-2)+i-2)!}{(n-k(i-2)-1)!}\right\} \\ +\frac{1}{(n+1)!}\left\{\frac{(n-i+3)!(n-ik+3k+i-2)!}{(n-k(i-3))!}-\frac{(n-i+2)!(n-ik+2k+i-1)!}{(n-k(i-2))!}\right\}.\end{aligned}$$

The proof is complete after straightforward algebraic operations. □

Proposition 2.4. Let $X_1, X_2, ..., X_n$ be a sequence of exchangeable Bernoulli trials. The probability $p_i(2,k,n)$ that the appearance of a scan of type 2/*k* for the first time is carried out upon the occurrence of $X_{i:n}$, satisfies the following recurrence relation with respect to *k* and *i*

$$\begin{aligned}(n)_{i+1}[p_{i+1}(2,k+1,n)-p_i(2,k,n)] \\ = (n-i)(n-(i-1)(k-1))_i+(n-k(i-1))_i-(n-ik)_{i+1} \\ -(n-i)(n-i+1)(n-(k-1)(i-2))_{i-1}\end{aligned} \tag{2.7}$$

for all values $i=1,2,...,n$ and $k \le n$, while

$$(n)_i = \frac{n!}{(n-i)!}.$$

Proof. Recalling equations (2.3) and (2.4), we may write the following recurrence with respect to design parameters k and i

$$
\begin{aligned}
& p_{i+1}(2,k+1,n) - p_i(2,k,n) \\
&\quad = \frac{1}{n!}\left\{(n-i)!\left[\frac{(n-ik+k+i-1)!}{(n-ik+k-1)!} + \frac{(n-k(i-1))!}{(n-i(k+1)+k)!}\right]\right. \\
&\qquad \left. - \frac{(n-i+1)!(n-(k-1)(i-2))!}{(n-k(i-2)-1)!} - \frac{(n-i-1)!(n-ik)!}{(n-i(k+1)-1)!}\right\}.
\end{aligned}
$$

The proof is complete after straightforward algebraic operations. □

Proposition 2.5. Let $X_1, X_2, ..., X_n$ be a sequence of exchangeable Bernoulli trials. The probability $p_i(2,k,n)$ that the appearance of a scan of type 2/k for the first time is carried out upon the occurrence of $X_{i:n}$, satisfies the following recurrence relation with respect to k and n

$$
\begin{aligned}
& n(n-1)_i[p_i(2,k+1,n-1) - p_i(2,k,n)] \\
&\quad = (n-i)[n(n-k(i-2)-1)_{i-1} + (n-(k-1)(i-1))_i \\
&\qquad - (n-i+1)(n-(i-2)(k-1))_{i-1}] - n(n-k(i-1)-1)_i
\end{aligned}
\tag{2.8}
$$

for all values $i = 1,2,...,n$ and $k \le n$, while

$$(n)_i = \frac{n!}{(n-i)!}.$$

Proof. Recalling equations (2.3) and (2.4), we may write the following recurrence with respect to design parameters k and n

$$
\begin{aligned}
& p_i(2,k+1,n-1) - p_i(2,k,n) \\
& = \frac{1}{n!}\left\{\frac{(n-i)!(n-k(i-1)+i-1)!}{(n-k(i-1)-1)!} - \frac{(n-i+1)!(n-k(i-2)+i-2)!}{(n-k(i-2)-1)!}\right\} \\
&\qquad + \frac{1}{(n+1)!}\left\{\frac{n-i)!(n-ik+2k-1)!}{(n-i(k+1)+2k)!} - \frac{(n-i-1)!(n-ik+k-1)!}{(n-i(k+1)+k-1)!}\right\}.
\end{aligned}
$$

The proof is complete after straightforward algebraic operations. □

Proposition 2.6. Let $X_1, X_2, ..., X_n$ be a sequence of exchangeable Bernoulli trials. The probability $p_i(2,k,n)$ that the appearance of a scan of type 2/*k* for the first time is carried out upon the occurrence of $X_{i:n}$, satisfies the following recurrence relation with respect to *i*, *k* and *n*

$$\begin{aligned}(n)_{i+2}[p_{i+1}(2,k+1,n-1)-p_i(2,k,n)] \\ =(n-i)(n-i-1)[(n-(i-1)(k-1))_i-(n-i+1)(n-(i-2)(k-1))_{i-1}] \\ -n(n-ik-1)_{i+1}+n(n-i+1)(n-k(i-1)-1)_i\end{aligned} \quad (2.9)$$

for all values $i = 1,2,...,n$ and $k \le n$, while

$$(n)_i = \frac{n!}{(n-i)!}.$$

Proof. Recalling equations (2.3) and (2.4), we may write the following recurrence with respect to design parameters *k* and *n*

$$\begin{aligned}&p_{i+1}(2,k+1,n-1)-p_i(2,k,n) \\ &=\frac{1}{n!}\left\{\frac{(n-i)!(n-k(i-1)+i-1)!}{(n-k(i-1)-1)!}-\frac{(n-i+1)!(n-k(i-2)+i-2)!}{(n-k(i-2)-1)!}\right\} \\ &\quad+\frac{1}{(n+1)!}\left\{\frac{(n-i-1)!(n-ik+k-1)!}{(n-i(k+1)+k-1)!}-\frac{(n-ik-1)!(n-i-2)!}{(n-i(k+1)-2)!}\right\}.\end{aligned}$$

The proof is complete after straightforward algebraic operations. □

3 Numerical results and applications

In this section, we display several numerical results for the quantity $p_i(r,k,n)$, namely the probability that, among *n* Bernoulli trials $X_1, X_2, ..., X_n$, the first appearance of a scan of type *r*/*k* is observed upon the occurrence of the order statistic $X_{i:n}$, $i = 1,2,...,n$. More specifically, we recall the equations $(2.4)-(2.9)$, proved earlier in the manuscript, in order to calculate probability $p_i(2,k,n)$ for several values of the design parameters k,n. Table 1 depicts the numerical outcomes of the applied recurrences.

Table 1: The probability $p_i(2,k,n)$ for several values of the design parameters k,n

		i									
k	***n***	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
3	**3**	0	1	0							
	4	0	5/6	1/6	0						
	5	0	7/10	3/10	0	0					
	6	0	3/5	2/5	0	0	0				
	7	0	11/21	47/105	1/35	0	0	0			
	8	0	13/28	13/28	1/14	0	0	0	0		
	9	0	5/12	13/28	5/42	0	0	0	0	0	
	10	0	17/45	41/90	17/105	1/210	0	0	0	0	0
4	**4**	0	1	0	0						
	5	0	9/10	1/10	0	0					
	6	0	4/5	1/5	0	0	0				
	7	0	5/7	2/7	0	0	0	0			
	8	0	9/14	5/14	0	0	0	0	0		
	9	0	7/12	17/42	1/84	0	0	0	0	0	
	10	0	8/15	13/30	1/30	0	0	0	0	0	0
5	**5**	0	1	0	0	0					
	6	0	14/15	1/15	0	0	0				
	7	0	6/7	1/7	0	0	0	0			
	8	0	11/14	3/14	0	0	0	0	0		
	9	0	13/18	5/18	0	0	0	0	0	0	
	10	0	2/3	1/3	0	0	0	0	0	0	0
6	**6**	0	1	0	0	0	0				
	7	0	20/21	1/21	0	0	0	0			
	8	0	25/28	3/28	0	0	0	0	0		
	9	0	5/6	1/6	0	0	0	0	0	0	
	10	0	7/9	2/9	0	0	0	0	0	0	0

For illustration purposes, let us consider an application of the theoretical results, proved in the previous section, in the field of Cryptography. More precisely, let us assume that a transposition system is applied in order to encipher a message. In transposition systems, plaintext values are rearranged without otherwise changing them or replacing them with other values. All the plaintext characters that were present before the encipherment are still present after that. The only thing that changes is the order of the initial form of the text. It goes without saying that the rearrangement of the sequence of alphabet letters that are included in the initial message, aims at making it difficult for a reader to understand the meaning of the original text. After transposing the message, we focus on comparing its initial form with the final one. The smaller the number of matches between the two forms of the text, the better the result of the cipher is. According to the transposition system that is under our investigation, an encoding is said to be unacceptable whenever among k consecutive comparisons of the corresponding positions between the initial and the final form of the message, appear at least 2 matches. The failure criterion of the transposition system could be expressed in terms of scan statistics as follows: the encipherment fails, namely the encoding is not reliable enough, whenever a scan of type $2/k$ occurs for the first time. In words, the time (measured in number of Bernoulli trials) until the transposition system fails (for the first time) to produce an encipherment that satisfies the prespecified demands, coincides to the waiting time $T_2^{(k)}$ till the first occurrence of a scan of the prespefied type. Therefore, we may apply the general results proved previously, in order to study in detail the operation of such a transposition system. For example, let us assume that the message that should be cipher by the transposition system is the phrase *ENEMY CLOSE*, while the encoded message is not acceptable if there exist k consecutive positions such that at least 2 matches appear therein. If the required design parameter k is equal to 4, then a plausible rearrangement that leads to failure of the encipherment, is given as follows

ENEMYCLOSE (initial message)

YMNEECOESL (encoded message)

S S S S S F S S F S (sequence of 10 comparisons)

It goes without saying that in the above scheme, there exist $k = 4$ consecutive comparisons (between the initial and the final form of the message) wherein 2 matches (F) appear. Based on Table 3.1, one may immediately observe that under the runs rule $r = 2, k = 4, n = 10$, the probability that the failure of the encoding procedure comes out upon the occurrence of the $i-th$ match is equal to 8/15, 13/30, 1/30 for $i = 2,3,4$ respectively, while for all other $i-$values the corresponding probability equals to zero.

References

[1] N. Balakrishnan and P.S. Chan, Start-up demonstration tests with rejection of units upon observing d failures, *Annals of the Institute of Statistical Mathematics*, **52**, (2000), 184-196.

[2] N. Balakrishnan, M.V. Koutras and F.S. Milienos, Some binary start-up demonstration tests and associated inferential methods, *Annals of the Institute of Statistical Mathematics*, **66**, (2014),759-787.

[3] N. Balakrishnan and M.V. Koutras, *Runs and Scans with Applications*, John Wiley & Sons, N. Y., 2002.

[4] N. Balakrishnan and H.K.T. Ng, A General Maximal Precedence Test, In *System and Bayesian Reliability*, eds. Y. Hayakawa, T. Irony and M. Xie, 105-122, Singapore, World Scientific Publishing Co Pte. Ltd, 2001.

[5] S. Bersimis, M.V. Koutras and G.K. Papadopoulos, Waiting time for an almost perfect run and applications in statistical process control, *Methodology and Computing in Applied Probability*, **16**, (2014), 207-222.

[6] W.E. Deming, Making things right, In *Statistics*: *A Guide to the Unknown*, Eds. J. Tanur *et al.*, Holden-Day, San Francisco, California, 1972.

[7] S. Eryilmaz and S. Chakraborti, On start-up demonstration tests under exchangeability, *IEEE Transactions on Reliability*, **57**, (2008), 627-632.

[8] S. Eryilmaz, M.V. Koutras and I.S. Triantafyllou, Signature based analysis of *m*-consecutive *k*-out-of-*n*: *F* systems with exchangeable components, *Naval Research Logistics*, **58**, (2011), 344-354.

[9] A.E. Gera, Combined *k*-out-of-*n*: *G* and consecutive k_c-out-of-*n*: *G* systems, *IEEE Transactions on Reliability*, **53**, (2004), 523-531.

[10] J. Glaz, J. Naus and S. Wallenstein, *Scan Statistics*. In *Springer Series in Statistics*, Springer, New York, 2001.

[11] G.J. Hahn and J.B. Gage, Evaluation of a start-up demonstration test, *Journal of Quality Technology*, **15**, (1983), 103-105.

[12] R. Modarres and J.L. Gastwirth, Hybrid test for the hypothesis of symmetry, *Journal of Applied Statistics*, **25**, (1998), 777-783.

[13] J. Naus, An extension of the birthday problem, *The American Statistician*, **22**, (1968), 27–29.

[14] I.S. Triantafyllou and M.V. Koutras, On the signature of coherent systems and applications, *Probability in the Engineering and Informational Science*, **22**, (2008), 19–35.

[15] I.S. Triantafyllou and M.V. Koutras, Signature and *IFR* preservation of 2-within-consecutive *k*-out-of-*n*: *F* systems, *IEEE Transactions on Reliability*, **60**, (2011), 315-322.

[16] I.S. Triantafyllou and M.V. Koutras, Reliability properties of (n, f, k) systems, *IEEE Transactions on Reliability*, **63**, (2014),357-366.

[17] A. Wald and J. Wolfowitz, On a test whether two samples are from the same population, *Annals of Mathematical Statistics*, **11**, (1940), 147-162.

[18] F. Yalcin and S. Eryilmaz, Start-up demonstration test based on total successes and total failures with dependent start-ups, *IEEE Transactions on Reliability*, **61**, (2012), 227-230.