APPLICATIONES MATHEMATICAE
40,4 (2013), pp. 421–430

Czesław Stępniak (Rzeszów)

ON DISTRIBUTION OF WAITING TIME FOR THE FIRST FAILURE FOLLOWED BY A LIMITED LENGTH SUCCESS RUN

Abstract. Many doctors believe that a patient will survive a heart attack unless a succeeding attack occurs in a week. Treating heart attacks as failures in Bernoulli trials we reduce the lifetime after a heart attack to the waiting time for the first failure followed by a success run shorter than a given k. In order to test the "true" critical period of the lifetime we need its distribution. The probability mass function and cumulative distribution function of the waiting time are expressed in explicit and concise form by binomial coefficients.

1. Background. Many doctors believe that a patient will survive a heart attack (in medical terminology: myocardial infarction) unless a succeeding attack occurs in a week. In fact, the critical period (for definition see Jones, 2008, p. 372) ranges, depending on the confidence level, from several days to one month (cf. Mover et al., 1964; Wilkinson et al., 1994; Bacos and Mattingly, 1966; Mooe et al., 1997; Witt at al., 2005; Adabag et al., 2008; Saczynski et al., 2008). Treating heart attacks as failures in Bernoulli trials we reduce the lifetime after a heart attack to the waiting time for the first failure followed by a success run shorter than a given k. In order to test the "true" critical period of the lifetime we need its distribution. Thus the distribution of the waiting time for the first failure followed by a success run shorter than k is important.

Recently, simple and compound patterns of successes and failures in Bernoulli trials were investigated very intensively. The main attention fo-

2010 *Mathematics Subject Classification*: Primary 62E15, 05A10; Secondary 60C05, 62N86.
Key words and phrases: heart attack, lifetime, Bernoulli trial, compound pattern, waiting time, probability mass function, cumulative distribution function.

DOI: 10.4064/am40-4-3

cused on the waiting time for a run of k successes and for two failures separated by a success run of length exactly, at least, or at most k. The first problem originated by Feller (1968) has been considered, among others, by Philippou (1986), Aki et al. (1984), Philippou et al. (1983), Philippou and Makri (1986), Aki (1992), Balakrishnan and Koutras (2002), Fu and Lou (2003). Two failures separated by a success run were investigated by Koutras (1996), Sen and Goyal (2000), Antzoulakos (2001), Sarkar et al. (2004), Dafnis and Philippou (2010) and Dafnis et al. (2012).

Huang and Tsai (1991) introduced a more complex pattern where a failure run not longer than k_1 was followed by a success run not longer than k_2. They obtained the probability generating function (PGF) of the number of occurrences of such patterns. Other combinations of failure and success runs (not shorter, not longer, or exactly equal to given numbers k_1 and k_2) were considered recently by Dafnis et al. (2010). Such patterns were named (k_1, k_2) *events*. Not only PGF and its moments but also recurrent equations and approximate formulae for the probability mass function (PMF) were given for the waiting time for the rth (k_1, k_2) event. Some modified patterns, in the Pólya–Eggenberger model, were investigated by Sen et al. (2006). The above mentioned results were obtained by Markov chain techniques and expressed in terms of PGF or its moments. PMF was presented only in approximate form or by recurrence equations.

PGF is convenient for theoretical consideration and for computing of moments while PMF and the cumulative distribution function (CDF) are practical tools used to construct statistical tests and confidence intervals (see, e.g., Lehmann and Romano 2005). On the other hand, it is a rather long way from PGF to PMF (see Feller 1968, Sec. XIII.4). Explicit forms of PMF's were derived for the waiting time for a run of k successes by Philippou and Muwafi (1982), Uppuluri and Patil (1983), and Muselli (1996), and for two failures separated by a success run by Sen and Goyal (2004).

A failure followed by a success run may be treated as a (k_1, k_2) event with $k_2 = 1$. However neither the PMF nor CDF of waiting time for such a pattern has been presented in the literature so far.

In this note we derive, in a combinatorial manner, both the PMF and CDF of the waiting time for the first failure followed by a success run shorter than k. Our formulae are given in explicit and concise form involving binomial coefficients.

2. First steps towards distribution of the waiting time. The waiting time for the first failure followed by a success run shorter than k may be expressed by the following model.

Let $X = (X_1, X_2, \dots)$ be a sequence of independent identically distributed random variables taking values 1 or 0 with probabilities p and

$q = 1 - p$, where 0 is interpreted as failure. Given a positive integer k called the *critical period,* define a statistic $T = t(X)$ as the minimal integer n such that $X_n = 0$ and either $n \le k$, or $X_m = 0$ for some m satisfying $0 < n - m \le k$. The statistic T is said to be the *waiting time for the first failure followed by a success run shorter than k.* We are interested in the distribution

$$p_n = P(T = n), \qquad n = 1, 2, 3, \ldots.$$

Let us start from some special cases. If $n \in \{1, \ldots, k\}$ then the waiting time coincides with the time of the first failure. Thus we get the formula

$$p_n = qp^{n-1}.$$

Now let $n \in \{k+1, \ldots, 2k+1\}$. In this case $T = n$ only if $X_n = X_m = 0$ for some m with $0 < n - m \le k$. We have $n - (k+1)$ such possibilities. This leads to the formula

$$p_n = (n - k - 1)q^2p^{n-2} \tag{1}$$

for all $n \in \{k+1, \ldots, 2k+1\}$.

In particular we get

Corollary 1. $P(T = k+1) = 0$.

The case $n \in \{2k+2, 2k+3, \ldots, 3k+2\}$ is a little more complex.

If $n = 2k+2$ then $T = n$ if and only if the failures appear at times n and m, where $0 < n - m \le k$. Hence $m \in \{n-k, n-k+1, \ldots, n-1\}$, and therefore

$$p_{2k+2} = kq^2p^{2k}.$$

If $n = 2k+3$ then $T = n$ may occur in the presence of either two or three failures. In the case of three failures the integer m has to satisfy the additional condition $m \ge 2(k+1)$, and so the failures appear at times $k+1, 2k+2, 2k+3$. For two failures, m may be an arbitrary element of $\{k+3, \ldots, 2k+2\}$. Thus

$$p_{2k+3} = kq^2p^{2k+1} + q^3p^{2k}.$$

It will be shown in Section 4 that

$$p_n = kq^2p^{n-2} + \binom{n-2k-1}{2}q^3p^{n-3} \tag{2}$$

for all $n \in \{2k+2, 2k+3, \ldots, 3k+2\}$.

Remark 2. For $n = 2k+1$ the formulae (1) and (2) coincide.

In order to derive a general formula for all n, we need some auxiliary results.

3. Auxiliary results. Throughout this paper we use the Newton symbol $\binom{n}{k}$ defined as $\frac{n!}{k!(n-k)!}$ if $0 \le k \le n$ and 0 otherwise. By the well known recurrence formula $\binom{n+1}{k+1} = \binom{n}{k} + \binom{n}{k+1}$ we get directly

$$\sum_{i=0}^{k} \binom{r+i-1}{r-1} = \binom{r+k}{r} \tag{3}$$

for all positive integers r and k (cf. Feller 1968, p. 64).

Going back to our considerations of Section 2 one can observe that the PMF of the waiting time for the first failure followed by a success run shorter than k may be expressed in terms of m-element sequences $x_1, \ldots, x_m$, where $m > k \ge 0$, such that

1° $x_i \in \{0,1\}$ for $i = 1, \ldots, m$,
2° $x_m = 0$,
3° each zero in the sequence is preceded by at least k consecutive ones.

Denote by $S_{k,m}$ the set of all such sequences.

LEMMA 3. *For any nonnegative integers k, m and r such that $m > k$ the number of sequences in $S_{k,m}$ including exactly $r+1$ zeros is given by*

$$N_{k,m,r} = \binom{m-(r+1)k-1}{r}.$$

Proof. We observe that $N_{k,m,r}$ may be expressed as the number of allocations of $m-(r+1)$ indistinguishable balls into $r+1$ distinguishable cells so that each cell is occupied by at least k balls. First we put k balls into each of $r+1$ cells. The remaining $b = m-(r+1)(k+1)$ ones are allocated in an arbitrary way to $c = r+1$ cells. We have

$$\binom{b+c-1}{c-1} = \binom{m-(r+1)k-1}{r}$$

such ways (cf. Feller, 1968, p. 38). This completes the proof. ∎

From Lemma 3, by our convention concerning the symbol $\binom{n}{k}$ we get the following corollary.

COROLLARY 4. *For given m and k the number $N_{k,m,r}$ is positive if and only if $r \le \left[\frac{m}{k+1}\right] - 1$, where $[a]$ means the integer part of a.*

In the next sections we also need the following elementary result.

LEMMA 5. *Let A_m and B_{mr}, for $m \in I$ and $r = 0, \ldots, n_i$, be random events such that $A_m \cap A_{m'} = \emptyset$ and $B_{mr} \cap B_{mr'} = \emptyset$ for all $m \ne m'$ and $r \ne r'$. Assume also that A_m is independent of B_{mr} for all m, r. Then*

$$P\Big(\bigcup_{m\in I}\Big[A_m \cap \bigcup_{r=0}^{n_i} B_{mr}\Big]\Big) = \sum_m \sum_r P(A_m)P(B_{mr}).$$

Proof. This follows directly from the Formula of Incompatible and Exhaustive Cases (e.g. Brémaud, 1988, p. 19). ■

4. Probability mass function of the waiting time. We continue considering the waiting time for the first failure followed by a success run shorter than k, computed in Section 2 for $n \le 2k+1$.

If $n > 2k+1$, then the number m appearing in the definition of the waiting time runs over the set $I = \{n-k, n-k+1, \ldots, n-1\}$. Given n define the random events

$$A_m = \{\omega : X_m(\omega) = 0, X_{m+1}(\omega) = 1, \ldots, X_{n-1}(\omega) = 1,\ X_n(\omega) = 0\},$$

for $m \in I$. We note that these events are disjoint and

$$P(A_m) = p^{n-m-1}q^2. \tag{4}$$

Now for every m let us consider the random events

$$B_{mr} = \Big\{\omega : (X_1(\omega), \ldots, X_{m-1}(\omega), 0) \in S_{k,m} \text{ and } \sum_{i=1}^{m-1} X_i(\omega) = m - r - 1\Big\},$$

where $S_{k,m}$ appeared in Lemma 3, for $r = 0, 1, \ldots, \big[\frac{m}{k+1}\big] - 1$ (cf. Corollary 4), These events are also disjoint and, according to that lemma,

$$P(B_{mr}) = \binom{n-(r+1)k-1}{r} p^{m-r-1}q^r. \tag{5}$$

Since the random variables $X_1, \ldots, X_n$ are independent, each random event B_{mr} is independent of A_m. Hence the assumption of Lemma 5 is met. The main result in this paper is the following theorem.

THEOREM 6. *The probability mass function of the waiting time for the first failure followed by a success run shorter than k is given by the formula*

$$p_n = P(T = n) = \begin{cases} p^n t & \text{if } n = 1, \ldots, k, \\ (n-k-1)p^n t^2 & \text{if } n = k+1, \ldots, 2k+1, \\ p^n t^2 \displaystyle\sum_{r=0}^{[\frac{n-k-2}{k+1}]} a_r t^r & \text{if } n > 2k+14, \end{cases} \tag{6}$$

where $t = q/p$ and

$$a_r = a_{r;k,n} = \binom{n-(r+1)k-1}{r+1} - \binom{n-(r+2)k-1}{r+1}, \qquad r = 0, \ldots, \left[\frac{n-k-2}{k+1}\right]. \tag{7}$$

Proof. For $n \le 2k+1$ formula (6) was derived in Section 2.
If $n > 2k+1$ then

$$p_n = P\Big(\bigcup_{m\in I}\Big[A_m \cap \bigcup_{r=0}^{[\frac{m}{k+1}]-1} B_{mr}\Big]\Big).$$

Lemmas 3 and 5, via formulae (4) and (5), yield

$$p_n = \sum_{m=n-k}^{n-1}\sum_{r=0}^{[\frac{m}{k+1}]-1}\binom{m-(r+1)k-1}{r}p^{n-r-2}q^{r+2}.$$

Since $m \le n-1$ and $\binom{m-(r+1)k-1}{r} = 0$ for $r > \left[\frac{m}{k+1}\right]-1$ (cf. Corollary 4) one can write

$$p_n = \sum_{m=n-k}^{n-1}\sum_{r=0}^{[\frac{n-1}{k+1}]-1}\binom{m-(r+1)k-1}{r}p^{n-r-2}q^{r+2}.$$

By changing the order of summation we get

$$p_n = p^n t^2 \sum_{r=0}^{[\frac{n-k-2}{k+1}]} a_r t^r,$$

where

$$\begin{aligned} a_r &= \sum_{m=n-k}^{n-1}\binom{m-(r+1)k-1}{r} \\ &= \binom{n-(r+1)k-2}{r} + \dots + \binom{n-(r+2)k-1}{r} \quad \text{for } r = 0,\dots,\left[\frac{n-k-2}{k+1}\right]. \end{aligned}$$

Finally, by the identity (3), the last formula reduces to (7). ∎

Some special cases of (6) and (7) are presented in the following remarks.

Remark 7. For all $k \ge 1$ and $n > 2k+1$, $a_{0;k,n} = k$.

Remark 8. For all $k \ge 1$ and $n > 2k+1$ and for $r = \left[\frac{n-k-2}{k+1}\right]$,

$$a_{r;k,n} = \begin{cases} \binom{n-(r+1)k-1}{r+1} - 1 & \text{if } n = i(k+1) \text{ for some integer } i, \\ \binom{n-(r+1)k-1}{r+1} & \text{otherwise.} \end{cases}$$

Remark 9. For $n \in \{2k+2, 2k+3, \dots, 3k+2\}$ formulae (6) and (2) coincide.

Proof. The case $r = 0$ may be verified directly. For $r = \left[\frac{n-k-2}{k+1}\right]$ we only need to note that

$$\binom{n - \left(\left[\frac{n-k-2}{k+1}\right] + 2\right)k - 1}{\left[\frac{n-k-2}{k+1}\right] + 1} > 0$$

if and only if $n - \left[\frac{n+k}{k+1}\right](k+1) \geq 0$. On the other hand, the last inequality is equivalent to $n = i(k+1)$ for some i. Moreover, in this case,

$$\binom{n - \left(\left[\frac{n-k-2}{k+1}\right] + 2\right)k - 1}{\left[\frac{n-k-2}{k+1}\right] + 1} = \binom{i-1}{i-1} = 1,$$

implying the desired result. ■

5. Cumulative distribution function of the waiting time. In this section we derive an explicit form of the CDF corresponding to the PMF given by (6). To this end, we will consider the same three cases.

If $n \in \{1, \ldots, k\}$ then

$$F(n) = q \sum_{i=1}^{n} p^{i-1} = 1 - p^n.$$

In particular, $F(k) = 1 - p^k$.

Now let $n \in \{k+1, \ldots, 2k+1\}$. In this case

$$\begin{aligned} F(n) &= 1 - p^k + \left(\frac{q}{p}\right)^2 \sum_{i=k+1}^{n} (i-k-1)p^i \\ &= 1 - p^k + \frac{q^2}{p} \sum_{i=k+1}^{n} ip^{i-1} - (k+1)\left(\frac{q}{p}\right)^2 \sum_{i=k+1}^{n} p^i. \end{aligned}$$

We note that $\sum_{i=k+1}^{n} ip^{i-1}$ may be written in the form

$$\frac{d}{dp} \sum_{i=k+1}^{n} p^i = \frac{[(k+1)p^k - (n+1)p^n]q + (p^{k+1} - p^{n+1})}{q^2}.$$

Consequently, we get

$$F(n) = 1 - (n-k)p^{n-1} + (n-k-1)p^n.$$

In particular,

$$F(2k+1) = 1 - (k+1)p^{2k} + kp^{2k+1}. \tag{8}$$

Now it remains to consider the case $n > 2k+1$. Then, by (6)–(8),

$$F(n) = 1 - (k+1)p^{2k} + kp^{2k+1} + \sum_{i=2k+2}^{n} \sum_{r=2}^{\left[\frac{i+k}{k+1}\right]} a_{i,r} p^{i-r} q^r,$$

where

$$a_{i,r} = \binom{i-(r-1)k-1}{r-1} - \binom{i-rk-1}{r-1}. \tag{9}$$

The above results are collected in the following theorem.

THEOREM 10. *The cumulative distribution function of the waiting time for the first failure followed by a success run shorter than k is given by the formula*

$$F(n) = \begin{cases} 1-p^n & \text{if } n = 1,\dots,k, \\ 1-(n-k)p^{n-1}+(n-k-1)p^n & \text{if } n = k+1,\dots,2k+1, \\ 1-(k+1)p^{2k}+kp^{2k+1}+\displaystyle\sum_{i=2k+2}^{n}\sum_{r=2}^{[\frac{i+k}{k+1}]} a_{i,r}p^{i-r}q^r & \text{if } n > 2k+1, \end{cases} \tag{10}$$

where $a_{i,r}$ is given by (9)*.*

We note that the last line in (10) may be expressed in a more attractive form:

$$\begin{aligned} & 1-(k+1)p^{2k}+kp^{2k+1}+\sum_{i=2k+2}^{n}\sum_{r=2}^{[\frac{n+k}{k+1}]} a_{i,r}p^{i-r}q^r \\ &\qquad = 1-(k+1)p^{2k}+kp^{2k+1}+\sum_{r=2}^{[\frac{n+k}{k+1}]}\sum_{i=2k+2}^{n} a_{i,r}p^{i-r}q^r. \end{aligned} \tag{11}$$

To this end we only need to use Corollary 4 implying

$$\binom{i-(r-1)k-1}{r-1} = \binom{i-rk-1}{r-1} = 0 \quad \text{for } r > \left[\frac{i}{k+1}\right]$$

and the fact that $i \le n$. However, the representation (11) is less convenient for computation because it involves some combinatorial expressions instead of zeros.

Acknowledgements. This work was partially supported by the Centre for Innovation and Transfer of Natural Sciences and Engineering Knowledge.

References

A. S. Adabag, T. M. Therneau, B. J. Gersh, S. A. Weston and V. L. Roger (2008), *Sudden death after myocardial infarction*, J. Amer. Med. Assoc. 300, 2022–2029.

S. Aki (1992), *Waiting time problems for a sequence of discrete random variables*, Ann. Inst. Statist. Math. 44, 363–378.

S. Aki, H. Kuboki and K. Hirano (1984), *On discrete distributions of order k*, Ann. Inst. Statist. Math. 36, 431–440.

D. L. Antzoulakos (2001), *Waiting times for patterns in a sequence of multistate trials*, J. Appl. Probab. 38, 508–518.

J. M. Bacos and T. W. Mattingly (1966), *The critical period of heart attack*, J. Rehabil. 32, 29–31.

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

P. Brémaud (1988), *An Introduction to Probabilistic Modelling*, Springer, New York.

S. D. Dafnis, D. L. Antzoulakos and A. N. Philippou (2010), *Distributions related to (k_1, k_2) events*, J. Statist. Plann. Inference 140, 1691–1700.

S. D. Dafnis and A. N. Philippou (2010), *On waiting time distributions for the occurrence of patterns*, in: Proceedings of the World Congress on Engineering 2010, Vol. III, London, 1697–1704.

S. D. Dafnis, A. N. Philippou and D. L. Antzoulakos (2012), *Distribution of patterns of two successes separated by a string of $k-2$ failures*, Statist. Papers 53, 323–344.

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

J. C. Fu and W. Y. W. Lou (2003), *Distribution Theory of Runs and Patterns and its Applications: A Finite Markov Chain Imbedding Approach*, World Sci., River Edge, NJ.

W. Huang and C. Tsai (1991), *On a modified binomial distribution of order k*, Statist. Probab. Lett. 11, 125–131.

B. D. Jones (2008), *Comprehensive Medical Terminology*, 3rd ed., Delmar Publ., Albany, NY.

M. Koutras (1996), *On a waiting time distribution in a sequence of Bernoulli trials*, Ann. Inst. Statist. Math. 48, 789–806.

E. L. Lehmann and J. P. Romano (2005), *Testing Statistical Hypotheses*, Springer, New York.

T. Mooe, P. Ericsson and B. Stegmayr (1997), *Ischemic stroke after acute myocardial infarction*, Stroke 28, 762–767.

M. M. Mover, D. I. Miller and M. M. Nachlas (1964), *Clinical features relevant to possible resuscitation in death after acute myocardial infraction*, Am. Heart J. 67, 437–444.

M. Muselli (1996), *Simple expressions for success run distributions in Bernoulli trials*, Statist. Probab. Lett. 31, 101–128.

A. N. Philippou (1986), *Distributions and Fibonacci polynomials of order k, longest runs, and reliability of consecutive k-out-of-n:F systems*, in: A. N. Philippou et al. (eds.), Fibonacci Numbers and their Applications, Reidel, Dordrecht, 203–227.

A. N. Philippou, C. Georgiou and G. N. Philippou (1983), *A generalized geometric distribution and some its properties*, Statist. Probab. Lett. 1, 171–175.

A. N. Philippou and F. S. Makri (1986), *Successes, runs and longest runs*, Statist. Probab. Lett. 4, 101–105.

A. N. Philippou and A. A. Muwafi (1982), *Waiting for the kth consecutive success and the Fibonacci sequence of order k*, Fibonacci Quart. 20, 28–32.

J. S. Saczynski, F. A. Spencer, J. M. Gore, J. H. Gurwitz, J. Yarzebski, D. Lessard and R. J. Goldberg (2008), *Twenty-year trends in the incidence of stroke complicating acute myocardial infarction*, Arch. Intern. Med. 168, 2104–2110.

A. Sarkar, K. Sen and Anuradha (2004), *Waiting time distributions of runs in higher order Markov chains*, Ann. Inst. Statist. Math. 56, 317–349.

K. Sen, M. Agarwal and G. N. Bhattacharya (2006), *Pólya–Eggenberger F-S models of order (k_1, k_2)*, Studia Sci. Math. Hungar. 43, 1–31.

K. Sen and B. Goyal (2000), *Distributions of the numbers of success runs of length* r *and failure runs of length* r', J. Indian Statist. Assoc. 38, 1–22.

K. Sen and B. Goyal (2004), *Distributions of patterns of two failures separated by success runs of length* k, J. Korean Statist. Soc. 33, 35–58.

V. R. Uppuluri and S. A. Patil (1983), *Waiting times and generalized Fibonacci sequences*, Fibonacci Quart. 21, 242–249.

P. Wilkinson, K. Laji, K. Ranjadayalan, L. Parsons, and A. D. Timmis (1994), *Acute myocardial infarction in women: survival analysis in first six months*, British Med. J. 309, 566–569.

B. J. Witt, R. D. Brown Jr., S. J. Jacobsen, S. A. Weston, B. P. Yawn and V. L. Roger (2005), *A community-based study of stoke incidence after myocardial infarction*, Ann. Intern. Med. 143, 785–792.

Czesław Stępniak
Faculty of Mathematics and Natural Sciences
University of Rzeszów
Pigonia 1
35-310 Rzeszów, Poland
E-mail: stepniak@umcs.lublin.pl

Received on 31.1.2013;
revised version on 17.9.2013 (2170)