Abstract

In an infinite sequence of independent Bernoulli trials with success probabilities \( p_k = a/(a + b + k - 1) \) for \( k = 1, 2, 3, \ldots \), let \( N_r \) be the number of \( r \geq 2 \) consecutive successes. Expressions for the first two moments of \( N_r \) are derived. Asymptotics of the probability of no occurrence of \( r \) consecutive successes for large \( r \) are obtained. Using an embedding in a marked Poisson process, it is indicated how the distribution of \( N_r \) can be calculated for small \( r \).

Keywords: Bernoulli trial; embedding in Poisson process; random permutation; records; runs of successes; sums of indicators

2000 Mathematics Subject Classification: Primary 60C05
Secondary 60K99

1. Introduction

In this paper we study the number of runs of successes or consecutive records in an infinite sequence of independent Bernoulli trials where the success or record probability in the \( k \)th trial is \( p_k = a/(a + b + k - 1) \) for \( k = 1, 2, 3, \ldots, a > 0 \), and \( b \geq 0 \). Note that, for \( a = 1 \) and \( b = 0 \), the \( p_s \) are the probabilities of records in a sequence of independent, identically distributed, continuous random variables. For this case, results on the distributions of the number of consecutive records are given in Chern et al. (2000), Chern and Hwang (2005), and Hahlin (1995).

For \( a > 0 \) and \( b = 0 \), it follows from general results of Arratia et al. (1992), see also Arratia et al. (2003), on the limiting distributions of the cycle lengths in certain random permutations that the number of double records, that is, two consecutive successes, is Poisson with mean \( a \). For the general case \( a > 0 \) and \( b \geq 0 \), Móri (2001) proved that the distribution of the number of double records is a beta mixture of Poisson distributions. Some recursion formulae concerning multiple records are also given there. An explicit expression for the binomial moment of the number of double records in the first \( n \) trials is derived in Holst (2008a). Letting \( n \to \infty \) gives the limit distribution of Móri. This is also proved by other methods in Holst (2007), Holst (2008b), and Huffer et al. (2009).

Apart from the works of Chern et al. (2000), Chern and Hwang (2005), and Hahlin (1995) for the case in which \( a = 1 \) and \( b = 0 \), the author is not aware of other detailed studies on multiple records. In Section 2 we calculate for general \( a > 0 \) and \( b \geq 0 \) the first two moments of the number of \( r \) consecutive records and obtain an asymptotic expansion for the probability of no occurrence of \( r \) consecutive records. In Section 3 we illustrate how we can obtain distributions for the number of multiple records for small \( r \), triple and quadruple records in particular. For
2. Notation, moments, and occurrence of runs

In the rest of the paper $I_1, I_2, I_3, \ldots$ are independent Bernoulli random variables with

$$P(I_k = 1) = 1 - P(I_k = 0) = \frac{a}{a + b + k - 1},$$

where $a > 0$ and $b \geq 0$. An infinite random sequence $s$ of 1s and 0s is generated by the $I_s$.

The number of $r \geq 2$ consecutive 1s in $S$ can be written as

$$N_r = \sum_{k=1}^{\infty} I_k I_{k+1} \cdots I_{k+r-1}.$$

By the Borel–Cantelli lemma, it follows that $N_r < \infty$ with probability 1. In the following the notation $s^n = s(s+1) \cdots (s+n-1)$ is used for rising factorials.

**Theorem 2.1.** The first two moments of the number of $r$ consecutive records satisfy

$$E(N_r) = \frac{a^r}{(r-1)(a+b)^{r-1}},$$

$$E(N_r^2) = E(N_r) + 2 \sum_{j=r+1}^{2r-1} E(N_j) + \frac{2a}{r-1} E(N_{2r-1}).$$

**Proof.** Using the fact that the $I_s$ are independent, we obtain a telescoping sum, i.e.

$$E(N_r) = \sum_{k=1}^{\infty} E(I_k \cdots I_{k+r-1})$$

$$= \sum_{k=1}^{\infty} \frac{a^r}{(a+b+k-1)^r}$$

$$= \frac{a^r}{r-1} \sum_{k=1}^{\infty} \frac{1}{(a+b+k)^{r-2}} \left( \frac{1}{a+b+k-1} - \frac{1}{a+b+k+r-2} \right)$$

$$= \frac{a^r}{(r-1)(a+b)^{r-1}}.$$

The independence of the $I_s$, the identity $I_j^2 \equiv I_j$, and the formula for $E(N_r)$ give

$$E(N_r^2) = E\left( \sum_{k=1}^{\infty} I_k I_{k+1} \cdots I_{k+r-1} \sum_{\ell=1}^{\infty} I_{\ell} I_{\ell+1} \cdots I_{\ell+r-1} \right)$$

$$= E(N_r) + 2 E\left( \sum_{k=1}^{\infty} I_k I_{k+1} \cdots I_{k+r-1} \sum_{\ell=k+1}^{k+r-1} I_{\ell} I_{\ell+1} \cdots I_{\ell+r-1} \right)$$

$$+ 2 E\left( \sum_{k=1}^{\infty} I_k I_{k+1} \cdots I_{k+r-1} \sum_{\ell=k+r}^{\infty} I_{\ell} I_{\ell+1} \cdots I_{\ell+r-1} \right).$$
On consecutive records

\[ E(N_r) + 2 \sum_{j=1}^{r-1} E\left( \sum_{k=1}^{\infty} I_k \cdots I_{k+r-1+j} \right) \]

\[ + 2 \sum_{k=1}^{\infty} E\left( I_k \cdots I_{k+r-1} \sum_{\ell=k+r}^{\infty} I_{\ell} \cdots I_{\ell+r-1} \right) \]

\[ = E(N_r) + 2 \sum_{j=1}^{r-1} E(N_{r+j}) + 2 \sum_{k=1}^{\infty} E(I_k \cdots I_{k+r-1}) E\left( \sum_{\ell=k+r}^{\infty} I_{\ell} \cdots I_{\ell+r-1} \right) \]

\[ = E(N_r) + 2 \sum_{j=1}^{r-1} E(N_{r+j}) + 2 \sum_{k=1}^{\infty} \frac{a'}{r-1} (a+b+k-1)^{a'} (r-1)(a+b+k+r)^{r-1} \]

\[ = E(N_r) + 2 \sum_{j=1}^{r-1} E(N_{r+j}) + 2 \frac{a}{r-1} \sum_{k=1}^{\infty} \frac{a'^{2r-1}}{(a+b+k-1)^{2r-1}} \]

\[ = E(N_r) + 2 \sum_{j=r}^{2r-1} E(N_j) + \frac{2a}{r-1} E(N_{2r-1}) \]

proving the assertion.

In particular, for \( a = 1 \) and \( b = 0 \), we have

\[ E(N_r) = \frac{1}{(r-1)! (r-1)} \]

\[ \text{var}(N_r) = \frac{1}{(r-1)! (r-1)} + \sum_{j=ar}^{2r-2} \frac{2}{j! j} + \frac{1}{(2r-2)! (r-1)^2} - \frac{1}{(r-1)!^2 (r-1)^2} \]

in agreement with Chern et al. (2000). For triple records, we obtain \( E(N_3) = \frac{1}{4} \) and \( \text{var}(N_3) = \frac{95}{288} > \frac{1}{4} \). Thus, the distribution of \( N_3 \) is not Poisson. Recall that the number of double records, \( N_2 \), is Poisson with mean 1.

**Corollary 2.1.** For the mean of \( N_r \), as \( r \to \infty \),

\[ E(N_r) = \frac{\Gamma(a+b) a'}{(r-1)! r^{a+b}} \left( 1 + O\left( \frac{1}{r} \right) \right). \]

**Proof.** Using Stirling’s formula for the gamma function, we obtain, as \( r \to \infty \),

\[ E(N_r) = \frac{a'}{(r-1)(a+b)^{r-1}} \]

\[ = \frac{a'}{(r-1)!} \left( 1 + \frac{a+b}{r-1} \right) \frac{(r-1)!}{(a+b)!} \]

\[ = \frac{a'}{(r-1)!} \left( 1 + \frac{a+b}{r-1} \right) \frac{\Gamma(a+b) \Gamma(r)}{\Gamma(a+b+r)} \]

\[ = \frac{\Gamma(a+b) a'}{(r-1)! r^{a+b}} \left( 1 + O\left( \frac{1}{r} \right) \right). \]
Corollary 2.2. The probability of $r \geq 2$ consecutive records satisfies

$$E(N_r) > P(N_r \geq 1) > E(N_r) \left(1 + \frac{b}{a + b + r} + \frac{(a + b)(b + 1)}{(a + b + r)(a + b + r + 1)}\right) \left(1 + \frac{a + b}{r - 1}\right)^{-1}.$$

Proof. As $P(N_r \geq 2) \geq P(I_1 = \cdots = I_{r+1} = 1) > 0$, the left-hand side inequality in the assertion follows. The right-hand side inequality follows from

$$P(N_r \geq 1) > P(I_1 = \cdots = I_r = 1) + P(I_1 = 0, I_2 = \cdots = I_{r+1} = 1)$$

$$= \frac{a^r}{(a + b)^r} \left(1 + \frac{b}{a + b + r} + \frac{(a + b)(b + 1)}{(a + b + r)(a + b + r + 1)}\right)$$

and the above formula for $E(N_r)$.

Combining the above results we obtain the following asymptotic expansion.

Corollary 2.3. The probability of no occurrence of $r$ consecutive records satisfies

$$P(N_r = 0) = 1 - \frac{\Gamma(a + b) a^r}{(r - 1)! r^{a+b}} \left(1 + O\left(\frac{1}{r}\right)\right)$$

as $r \to \infty$.

For $a = 1$ and $b = 0$, we obtain

$$E(N_r) = \frac{1}{(r - 1)! (r - 1)} > P(N_r \geq 1) > \frac{1}{r!} + \frac{1}{(r + 2)!},$$

which implies that

$$P(N_r = 0) = 1 - \frac{1}{(r - 1)! (r - 1)} + O\left(\frac{1}{r! r}\right),$$

in agreement with Corollary 2 of Chern and Hwang (2005), proved in a different way. Using recursions of generating functions, $P(N_r = 0)$ is numerically computed in Table 1 of Chern et al. (2000). With these values in the above inequality for $P(N_r \geq 1)$ we obtain, for $r = 3, 4, 5, 6,$

$$0.250 > 0.192 > 0.175, \quad 0.0556 > 0.0451 > 0.0431,$$

$$0.0104 > 0.00875 > 0.00853, \quad 0.00167 > 0.00144 > 0.00141.$$

The simple approximation $P(N_r \geq 1) \approx 1/r! + 1/(r + 2)!$ is accurate; for example, $P(N_7 \geq 1) = 0.000203$ compared to $1/7! + 1/9! = 0.000201$. This is not surprising: if a long run of records occurs then, with high probability, it should start in the first or third trial.

3. Embedding and multiple records

Apart from special cases, no simple formula for the distribution of $N_r$ seems possible to give. However, using the embedding below, we will indicate how the distribution can be calculated for any $a > 0$ and $b \geq 0$. 
As an illustration, we will study triple records and briefly quadruple records. We can identify

\[ \Pi_1 p \]

and

\[ \Pi_1 \]

The superposition \( \Pi_1 \) is a Poisson process of intensity 1. Suppose that the

\( k \)th point of \( \Pi_1 \) originates from the process \( \Pi_{1p} \). Define a sequence \( \delta \) of 1s and 0s by letting the first 1 occur at position \( L_0 \), the second at position \( L_0 + L_1 \), the third at \( L_0 + L_1 + L_2 \), and so on. It was proved in Holst (2008b) that the sequences \( \delta \) and \( \delta_p \) have the same distribution. We can identify \( \delta \) and \( \delta_p \) with each other.

Note that a point in \( \Pi_1 \) originating from \( \Pi_{1p} \) corresponds to a double record in \( \delta \). As the number of points in \( \Pi_1 \) is Poisson with mean \( \int_0^\infty pe^{-t/a} dt = ap \), it follows that, conditional on the Beta\((a, b)\) random variable \( P = p \), the number of double records, \( N_2 \), is Poisson with mean \( ap \), as was proved in Móri (2001). Also, note that, conditional on \( P = p \), the superposition \( \Pi_{2p} \) of \( \Pi_{2p}, \Pi_{3p}, \ldots \) is a Poisson process of intensity \( 1 - pe^{-t/a} \) and independent of \( \Pi_{1p} \).

First consider the probability of no triple record:

\[
P(N_3 = 0) = P(N_2 \leq 1) + \sum_{j=2}^\infty P(N_2 = j, N_3 = 0).
\]

For the number of \( r \) consecutive records, \( P(N_r = k) = \sum_{j=0}^\infty P(N_2 = j, N_r = k) \), where the probability \( P(N_2 = j, N_r = k) \) can, at least in principle, be found using the embedding above. As an illustration, we will study triple records and briefly quadruple records.

### 3.1. Triple records

First consider the probability of no triple record:

\[ P(N_3 = 0) = P(N_2 \leq 1) + \sum_{j=2}^\infty P(N_2 = j, N_3 = 0). \]
**Proposition 3.1.** For two or three double records and no triple record,

\[ P(N_2 = 2, N_3 = 0) = P(N_2 = 2) - \sum_{\ell=0}^{\infty} P(N_2 = 2 + \ell) \frac{(\ell + 1)!}{(a + 1)^{\ell+1}} \]

and

\[ P(N_2 = 3, N_3 = 0) \]

\[ = P(N_2 = 3) - \sum_{\ell=0}^{\infty} P(N_2 = 3 + \ell) \left( \frac{(\ell + 1)!}{(a + 1)^{\ell+1}} + \frac{(\ell + 2)!}{(a + 2)^{\ell+1}} - \frac{(\ell + 2)! (\ell + 1)}{(a + 1)^{\ell+1}} \right). \]

**Proof.** Conditional on \( P = p \), we can argue as follows. Given two fixed numbers \( 0 < u < v \), the number of points of \( \Pi_{\geq 2p} \) in the interval \((u, v]\) is Poisson with mean

\[ \int_u^v (1 - pe^{-t/u}) \, dt = v - u - ap(e^{-u/a} - e^{-v/a}). \]

The event \('N_2 = j, N_3 = 0'\) means that the Poisson process \( \Pi_p \) has \( j \) points and that between all adjacent points there is at least one point from \( \Pi_{\geq 2p} \). Given that the \( j \) points of \( \Pi_p \) occur at \( t_1 < t_2 < \cdots < t_j \), the probability that \( N_3 = 0 \) is

\[ \prod_{k=1}^{j-1} (1 - \exp(-t_{k+1} - t_k - ape^{-t_{k+1}/a} + ape^{-t_k/a})). \]

As the points of the Poisson process \( \Pi_p \) occur with intensity \( pe^{-t/u} \), the probability of having no points of \( \Pi_p \) outside a fixed discrete finite set is \( e^{-ap} \). Hence,

\[ P(N_2 = j, N_3 = 0 \mid P = p) \]

\[ = \int \cdots \int I(0 < t_1 < \cdots < t_j) e^{-ap} p e^{-t_1/a} \cdots e^{-t_j/a} \]

\[ \times \prod_{k=1}^{j-1} (1 - \exp(-t_{k+1} + t_k - ape^{-t_{k+1}/a} + ape^{-t_k/a})) \, dt_1 \cdots dt_j \]

\[ = e^{-ap} \int_0^v \cdots \int_0^v \prod_{k=1}^{j-1} \left( 1 - \left( \frac{x_k}{x_{k+1}} \right)^{\frac{a}{x_{k+1}}} \right) \, dx_1 \cdots dx_j. \]

For \( j = 2 \), we find, with \( u = x_1/x_2 \), the series expansion and the beta function:

\[ P(N_2 = 2, N_3 = 0 \mid P = p) \]

\[ = e^{-ap} \int_0^{ap} x_2 \int_0^1 (1 - u^a e^{u_2(1-u)}) \, du \, dx_2 \]

\[ = e^{-ap} \int_0^{ap} x_2 \left( 1 - \sum_{\ell=0}^{\infty} \frac{x_2^\ell}{\ell!} \int_0^1 u^\ell (1-u)^\ell \, du \right) \, dx_2 \]

\[ = e^{-ap} \int_0^{ap} x_2 \left( 1 - \sum_{\ell=0}^{\infty} \frac{x_2^\ell}{(a + 1)^{\ell+1}} \right) \, dx_2 \]

\[ = P(N_2 = 2 \mid P = p) - \sum_{\ell=0}^{\infty} \frac{P(N_2 = 2 + \ell \mid P = p)(\ell + 1)!}{(a + 1)^{\ell+1}}. \]
which gives the unconditional probability in the assertion. With similar calculations we obtain
the formula for \( P(N_2 = 3, N_3 = 0) \).

Analogously, it is possible to calculate \( P(N_2 = j, N_3 = 0) \) for \( j = 4, 5, \ldots \).
At least two double records are needed to obtain a triple record. Therefore,

\[
P(N_3 = 1) = \sum_{j=2}^{\infty} P(N_2 = j, N_3 = 1),
\]

which can be calculated in a similar way as \( P(N_3 = 0) \). Note that

\[
P(N_2 = 2, N_3 = 1) = P(N_2 = 2) - P(N_2 = 2, N_3 = 0).
\]

The integral

\[
P(N_2 = j, N_3 = 1 \mid P = p) = e^{-ap} \int_0^{ap} \int_0^{x_1} \cdots \int_0^{x_j} \prod_{k=1}^{j-1} \left( \frac{x_k}{x_{k+1}} \right)^a e^{x_{k+1}-x_k} \left( \frac{x_k}{x_{k+1}} \right)^a e^{x_{j+1}-x_{j}} \, dx_1 \cdots dx_j
\]

has to be evaluated. Similar expressions can be given for \( P(N_3 = 2), P(N_3 = 3), \ldots \).

### 3.2. Triple records for \( a = 1 \) and \( b = 0 \)

For record probabilities in a sequence of independent, identically distributed, continuous
random variables, we have \( a = 1 \) and \( b = 0 \). Using the above results, we find, after some
computations, that

\[
P(N_2 = j) = \frac{e^{-1}}{j!}, \quad P(N_2 \leq 1) = 0.735 \, 758 \, 88,
\]

\[
P(N_2 = 2, N_3 = 0) = 0.066 \, 990 \, 05, \quad P(N_2 = 3, N_3 = 0) = 0.005 \, 047 \, 15.
\]

\[
P(N_2 = 4, N_3 = 0) = 0.000 \, 207 \, 57, \quad P(N_2 = 5, N_3 = 0) = 0.000 \, 005 \, 32.
\]

Summing these probabilities gives

\[
P(N_2 \leq 5, N_3 = 0) = 0.808 \, 008 \, 98.
\]

By subtle asymptotic analyses of a recurrence and a differential equation, a complicated formula
for the probability generating function of \( N_3 \) is given in Chern et al. (2000); see also Chern and
Hwang (2005). From this, they obtained

\[
P(N_3 = 0) = 0.808 \, 009 \, 125 \, 346 \, 368 \ldots
\]

For general \( a > 0 \) and \( b \geq 0 \), the methods used in Chern et al. are not applicable.

### 3.3. Quadruple records

To obtain a quadruple record, that is, four consecutive records, at least three double records
are needed. Hence,

\[
P(N_4 = 0) = P(N_2 \leq 2) + \sum_{j=3}^{\infty} P(N_2 = j, N_4 = 0).
\]
Proposition 3.2. For three double records and no quadruple record,
\[
P(N_2 = 3, N_4 = 0) = P(N_2 = 3) - \sum_{\ell=0}^{\infty} P(N_2 = 3 + \ell) \frac{(\ell + 2)! (\ell + 1)}{(a + 1)^{\ell+2}}.
\]

Proof. As in the proof of Proposition 3.1, we see that
\[
P(N_2 = 3, N_4 = 1 | P = p)
\]
\[
= e^{-ap} \int \int \int \mathbf{1}(0 < x_1 < x_2 < x_3 < ap) \left( \frac{x_1}{x_3} \right)^a e^{x_3-x_1} dx_1 dx_2 dx_3
\]
\[
= e^{-ap} \int_0^{ap} \int_0^{x_3} \left( \frac{x_1}{x_3} \right)^a e^{x_3-x_1} (x_3 - x_1) dx_1 dx_3
\]
\[
= \sum_{\ell=0}^{\infty} \frac{(ap)^{\ell+3} e^{-ap}}{\ell!} \frac{1}{\ell + 3} \int_0^{1} u^\ell (1 - u)^{\ell+1} du
\]
\[
= \sum_{\ell=0}^{\infty} \frac{(ap)^{\ell+3}}{(\ell + 3)!} \frac{e^{-ap} (\ell + 2)! (\ell + 1)}{(a + 1)^{\ell+2}},
\]
from which the assertion follows.

Similar formulae can be derived for \( P(N_2 = j, N_4 = 0) \) for \( j = 4, 5, \ldots \).

For \( a = 1 \) and \( b = 0 \), we find that \( P(N_2 \leq 2) = 0.91970, P(N_2 = 3, N_4 = 0) = 0.03097 \), and, after computations similar to those for triple records, \( P(N_2 = 4, N_4 = 0) = 0.00391 \). Thus, \( P(N_2 \leq 4, N_4 = 0) = 0.95458 \). Chern et al. (2000) obtained the value \( P(N_4 = 0) = 0.95490 \). Recall that
\[
1 - E(N_4) = 1 - \frac{1}{3! 3} = 0.94444 < P(N_4 = 0) = 0.95490 < 1 - \frac{1}{4!} - \frac{1}{6!} = 0.95694;
\]
see the end of Section 2.

References


