

On Some Probabilistic Distributions Related to the Classical Bernoulli Scheme

S. M. Ananjevskii^{a,*} and V. B. Nevzorov^{a,**}

^a St. Petersburg State University, St. Petersburg, 199034 Russia

* e-mail: ananjevskii@mail.ru

** e-mail: vanev@mail.ru

Received September 13, 2021; revised October 22, 2021; accepted December 2, 2021

Abstract—The classical scheme of independent Bernoulli trials, starting with Jacob Bernoulli’s work, has been one of the most popular topics in probability theory for more than three centuries. It is ideally suited for setting and solving various practical problems. A variety of results have been obtained related to modifications of this scheme. However, new situations and problems emerge that require further advancement in the study of various random variables associated to some extent with independent Bernoulli trials. Solutions to some of these problems are presented here.

Keywords: Bernoulli scheme, binomial distribution, geometric distribution, generating functions

DOI: 10.1134/S1063454122020042

1. We consider a sequence of independent random variables (r.v.) X_1, X_2, \dots that take a value of 1 with some probability $0 < p < 1$ and a value of 0 with the probability $q = 1 - p$.

The event $\{X_n = 1\}$ is often referred to as “success” in the n th trial, while the event $\{X_n = 0\}$ is called the “failure.”

Let μ_n denote the number of successes in the first n trials. The distribution of this number is described by the $B(n, p)$ binomial distribution with the probabilities

$$P_r = P\{\mu_n = r\} = C_n^r p^r q^{n-r}, \quad r = 0, 1, \dots, n-1, n.$$

Mathematical expectation is given by the equalities

$$E\mu_n = np, \quad n = 1, 2, \dots$$

Closely related to the Bernoulli scheme are geometric $\text{Geom}(q)$ distributions with the parameter $0 < q < 1$. If trials are carried out until the first “successful” event occurs, the number of failures is described by the geometric distribution since

$$P\{X_1 = 0, X_2 = 0, \dots, X_n = 0, X_{n+1} = 1\} = pq^n, \quad n = 0, 1, 2, \dots,$$

and the mathematical expectation of the number of failed trials is q/p .

If we are interested in the moment v when both outcomes are found in the X_1, X_2, \dots sequence, we obtain

$$P\{v = n\} = pq^{n-1} + qp^{n-1}, \quad n = 2, 3, \dots$$

We now consider a random variable $W(k)$, which is the number of failures that occurred before the emergence of the k th success. In this case, we deal with a sum that consists of k independent random variables distributed geometrically according to $\text{Geom}(q)$. This sum is described by a negative binomial distribution with the probabilities

$$P\{W(k) = n\} = C_{n+k-1}^k p^k q^n, \quad n = 0, 1, 2, \dots$$

and the mathematical expectation

$$EW(k) = kq/p, \quad k = 1, 2, \dots$$

The generating function for the r.v. $W(k)$ has the form

$$Es^{W(k)} = pq s^k / (1 - qs)^k, \quad k = 1, 2, \dots$$

2. We now consider the following situation. A trial is carried out not to the k th successful event but to the moment when the first series that includes no less than k successive successes occurs in the X_1, X_2, \dots sequence.

The first volume of a two-volume book by Feller ([1], ch. 13, Eq. (7.6)) contains the generating function $F(s)$ for the probabilities

$$f(k, n) = P\{L_n\}$$

that the first series that consists of k successful trials is related to the event

$$L_n = \{X_{n-k} = 0, X_{n-k+1} = 1, \dots, X_{n-1} = 1, X_n = 1\}.$$

It has been shown that

$$F(s) = p^k s^k (1 - ps) / (1 - s + qp^k s^{k+1}), \quad k = 1, 2, \dots \quad (1)$$

If we somewhat change the condition and consider the number of trials prior to the moment when such a series emerges, we obtain that the corresponding generating function has the following form

$$F_1(s) = p^k (1 - ps) / (1 - s + qp^k s^{k+1}), \quad k = 1, 2, \dots \quad (2)$$

Next, we deal with the events

$$\begin{aligned} A_0 &= \{X_1 = 0\}, \\ A_r &= \{X_1 = 1, X_2 = 1, \dots, X_{r-1} = 1, X_r = 1, X_{r+1} = 0\}, \quad r = 1, 2, \dots, k-1, \\ A_k &= \{X_1 = 1, X_2 = 1, \dots, X_{k-1} = 1, X_k = 1\}. \end{aligned}$$

It should be noted that

$$P\{A_r\} = qp^r, \quad r = 0, 1, \dots, k-1 \quad \text{and} \quad P\{A_k\} = p^k.$$

What can be said about the number of failures $W(k)$ if the event A_r event occurs? It is clear that the conditional distribution of this quantity provided that the event A_r , $r = 0, 1, 2, \dots, k-1$ occurred coincides with the distribution of the random variable $W_1(k) = 1 + W(k)$. If the event A_k occurred, no failures have been observed.

Let $B_1(k) = EW(k)$. We obtain that

$$B_1(k) = (1 - p^k)(1 + EW(k)) = (1 - p^k) + (1 - p^k)B_1(k).$$

Consequently

$$B_1(k) = (1 - p^k) / p^k, \quad k = 1, 2, \dots$$

Let now

$$P(k, s) = Es^{W(k)}$$

denote the generating function of the random variable $W(k)$. Taking into account the possible versions described above, we obtain that the following equalities hold true

$$P(k, s) = p^k + (1 - p^k)Es^{1+W(k)} = p^k + (1 - p^k)sP(k, s)$$

and

$$P(k, s) = p^k / (1 - (1 - p^k)s),$$

i.e., the variable $W(k)$ is described by the geometric $\text{Geom}(1 - p^k)$ distribution with the probabilities

$$P\{W(k) = n\} = p^k (1 - p^k)^n, \quad n = 0, 1, 2, \dots$$

In particular,

$$P\{W(k) = 0\} = p^k$$

and

$$P\{W(k) = 1\} = p^k(1 - p^k).$$

We now proceed to the random variable $R(k)$, the number of successful events before the emergence of the first series of interest for us that contains no less than k successive successes. The equality $R(k) = 0$ holds true if the event A_k occurs. If we deal with the event A_r , $r = 0, 1, 2, \dots, k - 1$, the conditional distribution $R(k)$ coincides with the distribution of the sum $R(k) + r$. For the mathematical expectation $B_2(k) = ER(k)$, we obtain the following relation

$$B_2(k) = \sum_{r=0}^{k-1} (ER(k) + r)p^r q = (1 - p^k)B_2(k) + (1 - kp^{k-1} + (k - 1)p^k)/q^2.$$

We obtain from this formula that

$$B_2(k) = (1 - kp^{k-1} + (k - 1)p^k)/p^{k-1}q.$$

The generating function $R(k, s)$ for the random variable $R(k)$ fulfils the following equality

$$R(k, s) = qR(k, s) + \sum_{r=1}^{k-1} s^r R(k, s)p^r q + p^k,$$

from which we obtain that

$$R(k, s) = p^{k-1}(1 - ps)/(1 - s + qp^{k-1}s^k). \tag{3}$$

If $k = 1$, we arrive at the natural result

$$R(1, s) = 1,$$

which corresponds to the distribution degenerate at zero. If $k = 2$, then

$$R(2, s) = p(1 - ps)/(1 - s + qps^2) = p/(1 - qs),$$

i.e., we obtain a geometric distribution with the probabilities

$$P\{R(2) = n\} = pq^n, \quad n = 0, 1, 2, \dots$$

We now consider a random variable $M(k)$, the number of success series before the momentum when the first series consisting of k or a larger number of successive successful trials emerges.

We now return to the events A_0, A_1, \dots, A_k . If the event A_0 occurs, the conditional distribution of the random variable $M(k)$ coincides with its unconditional distribution. If the event A_k occurs, we obtain that $M(k) = 0$. In all other cases, the conditional distribution of the random variable $M(k)$, provided that the event A_r , $r = 1, 2, \dots, k - 1$ occurs, coincides with the unconditional distribution for the quantity $M(k) + 1$. For the mathematical expectations $a(k) = EM(k)$, we obtain that

$$\begin{aligned} a(k) &= a(k)P(A_0) + (1 + a(k))(P(A_1) + P(A_2) + \dots + P(A_{k-1})) \\ &= qa(k) + (1 + a(k))(qp + qp^2 + qp^{k-1}) = qa(k) + p(1 - p^{k-1}) + a(k)p(1 - p^{k-1}) \\ &= a(k) - p^k a(k) + p(1 - p^{k-1}), \end{aligned}$$

which yields that

$$a(k) = (1 - p^{k-1})/p^{k-1}. \tag{4}$$

If $k = 1$, we naturally obtain $a(1) = 0$. If $k = 2$, then $a(2) = q/p$. If $k = 3$, then $a(3) = (1 - p^2)/p^2$.

It should be noted that the distribution of the number of series of the successes $M(k)$ does not depend on whether the first trial was successful or not.

If we now consider the number $N(k)$ of the series of failures before the emergence of a group of k successive successes, two possible situations can be distinguished. If the first trial was successful, $N(k) = M(k)$. Otherwise we obtain that $N(k) = M(k) + 1$. For the mathematical expectations $b(k) = EN(k)$, we arrive at the equality

$$b(k) = pB_2(k) + q(1 + B_2(k)) = q + B_2(k) = (q^2 p^{k-1} + 1 - kp^{k-1} + (k - 1)p^k)/p^{k-1}q. \tag{5}$$

If we also consider the total number $S(k)$ of the series of successes and the series of failures before the emergence of k successive successful trials, the distribution of this variable coincides with a mixture with weights p and q of the random variables $2M(k)$ and $2M(k) + 1$.

We now determine the form of the generating functions $M(k, s)$, $N(k, s)$, and $S(k, s)$ for the random variables $M(k)$, $N(k)$, and $S(k)$. It should be noted that the following formulas hold true:

$$N(k, s) = pM(k, s) + qsM(k, s) = (p + qs)M(k, s) \quad (6)$$

and

$$S(k, s) = pM(k, s^2) + qsM(k, s^2) = (p + qs)M(k, s^2). \quad (7)$$

Returning to the events A_0, A_1, \dots, A_k , we can derive the following equality for the function $M(k, s)$:

$$M(k, s) = qM(k, s) + (pq + p^2q + \dots + p^{k-1}q)sM(k, s) + p^k = qM(k, s) + (p - p^k)sM(k, s) + p^k,$$

from which it follows that

$$M(k, s) = p^{k-1}/(1 - s + sp^{k-1}). \quad (8)$$

Equation (8) yields in particular that

$$P\{M(k) = n\} = p^{k-1}(1 - p^{k-1})^n, \quad n = 0, 1, 2, \dots \quad (9)$$

Taking into account equalities (6) and (7), we also obtain that

$$N(k, s) = (p + qs)p^{k-1}/(1 - s + sp^{k-1}) \quad (10)$$

and

$$S(k, s) = (p + qs)p^{k-1}/(1 - s^2 + p^{k-1}s^2). \quad (11)$$

The following result (problem 24) can be found in ([1], ch. 13).

Let $q_{n,l} = P\{Q(n) = l\}$ be the probability to have in n Bernoulli trials exactly l series of success of the length r . Then the generating function $Q_n(s) = Es^{Q(n)}$ is the coefficient of s^n in the formula

$$(1 - p^r s^r)/[1 - s + qp^r s^{r+1} - (1 - ps)p^r s^r q].$$

We now present some statements that enable us to find the corresponding distributions for the number $V(k, r)$ of the series consisting of exactly $r < k$ successes in the scheme presented above. In this situation, the event A_r is associated with the random variable $V(k, r) + 1$. If the event A_k occurs, there are no such series. In all other cases, we deal with the conditional distribution that coincides with the distribution of the random variable $V(k, r)$. For the generating function

$$V(k, r, s) = Es^{V(k,r)},$$

the equality holds true

$$V(k, r, s) = p^k + p^r qsV(k, s) + (1 - p^r q - p^k)V(k, r, s),$$

which yields that

$$V(k, r, s) = p^{k-r}/(p^{k-r} + q - sq). \quad (12)$$

For $r = 1$, we obtain, in particular, that

$$V(k, 1, s) = p^{k-1}/(p^{k-1} + q - sq). \quad (13)$$

Using Eq. (13), we obtain, in particular, that

$$P\{V(k, 1) = n\} = q^n p^{k-1}/(q + p^{k-1})^{n+1}, \quad n = 0, 1, 2, \dots \quad (14)$$

We also note that the formulas for the generating functions

$$V(k, k-1, s) = p/(1 - sq), \quad k = 2, 3, \dots$$

do not depend on k . In these cases, we obtain that

$$P\{V(k, k - 1) = n\} = pq^n, \quad n = 0, 1, 2, \dots,$$

i.e., we deal with a geometric distribution.

3. We now continue considering a sequence of independent random variables X_1, X_2, \dots , which take a value of 1 with some probability $0 < p < 1$ and a value of 0 with the probability $q = 1 - p$ until the emergence of the first series of k successes. We have presented above the distributions for the number of successes and failures, the number of the series of successes, and the number of the series of failures in the sequence before the emergence of k successive successes. We have also considered the distribution for the number of the series of success of a certain length $r = 1, 2, \dots, k - 1$ that precede the emergence of the first series of k successes.

We now return to the consideration of the series of failures observed before the emergence of k successively recorded successes. It should be noted that suchlike series already can consist of any number of successive failures. Equation (10) provides a formula for the generating function of the random variable $N(k)$, the number of failure series. We now can consider the series of such failures of a fixed length $r = 1, 2, \dots$. Let $W(r, k)$ denote their number.

We consider two possible cases.

In the situation when $X_1 = 0$ and $X_1 = 1$, the number of series of failures of length r is denoted as $W_1(r)$ and $W_2(r)$, respectively. Corresponding generating functions $W(r, k, s)$, $W_1(r, s)$, and $W_2(r, s)$ for the random variables $W(r, k)$, $W_1(r)$, and $W_2(r)$ are also needed. We note that

$$W(r, k) = qW_1(r) + pW_2(r). \tag{15}$$

Let $X_1 = 0$. Then

$$P\{W(r, k) = m | X_1 = 0\} = P\{W_1(r) = m\}. \tag{16}$$

If $X_1 = 1$, then

$$P\{W(r, k) = m | X_1 = 1\} = P\{W_2(r) = m\}. \tag{17}$$

We obtain that

$$P\{W(r, k) = m\} = qP\{W_1(r) = m\} + pP\{W_2(r) = m\}. \tag{18}$$

We consider the following set of initial events:

$$A_j = \{X_1 = 0, X_2 = 0, \dots, X_{j-1} = 0, X_j = 1\}, \quad j = 1, 2, \dots, r - 1, r, r + 1, \dots$$

If the event A_{r+1} occurs, which is associated with the conditional probability

$$P\{A_{r+1} | X_1 = 0\} = q^{r-1}p,$$

the distribution of the random variable $W_1(r)$ coincides with the distribution of the random variable $1 + W_2(r)$. In other cases, the distribution $W_1(r)$ coincides with the distribution $W_2(r)$.

Let $X_1 = 1$ now. We consider the events

$$B_j = \{X_1 = 1, X_2 = 1, \dots, X_{j-1} = 1, X_j = 0\}, \quad j = 2, 3, \dots, k,$$

$$B_{k+1} = \{X_1 = 1, X_2 = 1, \dots, X_k = 1\}.$$

In the situations when any of the events B_2, B_3, \dots, B_k occurs, the conditional distribution $W_2(r)$ coincides with the distribution of the random variable $W_1(r)$. In the case B_{k+1} , we obtain that $W_2(r) = 0$.

For the first of these two situations, we obtain an equality for the generating functions of the form

$$W_1(r, s) = pq^{r-1}sW_2(r, s) + (1 - pq^{r-1})W_2(r, s). \tag{19}$$

In the second case, we arrive at the equality

$$W_2(r, s) = (1 - p^{k-1})W_1(r, s) + p^{k-1}. \tag{20}$$

The two last formulas that relate the two generating functions yield the equalities

$$W_1(r, s) = \frac{p^{k-1}(1 + pq^{r-1}s - pq^{r-1})}{p^{k-1} + (pq^{r-1} - p^k q^{r-1})(1-s)}, \quad (21)$$

$$W_2(r, s) = \frac{p^{k-1}}{p^{k-1} + pq^{r-1}(1-s) - p^k q^{r-1}(1-s)} \quad (22)$$

and

$$W(r, k, s) = qW_1(r, s) + pW_2(r, s) = \frac{qp^{k-1} + p^k q^r s - p^k q^r + p^k}{p^{k-1} + (pq^{r-1} - p^k q^{r-1})(1-s)}. \quad (23)$$

In the particular case $k = 2$ and $r = 1$, we obtain the generating function in the form

$$W(1, 2, s) = \frac{1 - pq + pqs}{1 + q - qs}.$$

Other problems related to the scheme of independent Bernoulli trials can be found in publications [2, 3].

ACKNOWLEDGMENTS

We are grateful to peer reviewers for finding minor errors in the manuscript, the removal of which improved the quality of the publication.

ADDITIONAL INFORMATION

Cite this work: S. M. Ananjevskii and V. B. Nevzorov, "On Some Probabilistic Distributions Related to the Classical Bernoulli Scheme," Vest. St. Petersburg. Gos. Univ., 2022, vol. 9 (67), no. 2, pp. 202–209 (in Russian). <https://doi.org/10.21638/spbu01.2022.202>

REFERENCES

1. V. Feller, *An Introduction to Probability Theory and Its Applications* (Wiley, New York, 1968; Mir, Moscow, 1984), Vol. 1.
2. N. Balakrishnan and V. B. Nevzorov, *A Primer on Statistical Distributions* (Wiley, Hoboken, N.J., 2003).
3. O. A. Ivanov, "Bernoulli scheme and generalizations of Fibonacci numbers," Vestn. S.-Peterb. Univ., Ser. 1: Mat., Mekh., Astron. **1(59)**, 529–532 (2014).

Translated by M. Shmatikov