PLISKA
STUDIA MATHEMATICA
BULGARICA

ПЛИСКА
БЪЛГАРСКИ
МАТЕМАТИЧЕСКИ
СТУДИИ

Pliska Stud. Math. Bulgar. **13** (2000), 151-154

PLISKA
STUDIA MATHEMATICA
BULGARICA

A CHARACTERIZATION OF THE NEGATIVE BINOMIAL DISTRIBUTION

Nikolay Kolev, Leda Minkova

Only a few characterizations have been obtained in literatute for the negative binomial distribution (see Johnson et *al.*, Chap. 5, 1992). In this article a characterization of the negative binomial distribution related to random sums is obtained which is motivated by the geometric distribution characterization given by Khalil et *al.* (1991). An interpretation in terms of an unreliable system is given.

Consider a sequence $X_1, X_2, \ldots$ of non-negative integer-valued independent identically distributed random variables (iid. r.v.'s) defined by

$$p_k = P(X_1 = k) \geq 0, \qquad k \geq 0, \qquad \sum_{k=0}^{\infty} p_k = 1$$

and let $Y_1, Y_2, \ldots$ be another sequence of non-negative integer-valued iid. r.v.'s, independent of the sequence $X_1, X_2, \ldots$, given by

$$q_k = P(Y_1 = k) \geq 0, \qquad k \geq 0, \qquad \sum_{k=0}^{\infty} q_k = 1, \quad \text{with} \quad q_0 < 1.$$

We call $\{Y_n, n \geq 1\}$ the truncating process. Let us define the r.v.'s $N_0 = 0$,

$$N_i = inf\{k > N_{i-1} : X_k < Y_k\}, \qquad i = 1, 2, \ldots, r$$

and $Z_0 = 0$,

$$Z_r = \sum_{i=1}^{r} \left\{ \sum_{j=1+N_{i-1}}^{N_i - 1} Y_j + X_{N_i} \right\}. \tag{1}$$

The r.v. Z_r represents the total truncated sum until the moment when for r-th time, $r \geq 1$, the truncating process $\{Y_n, n \geq 1\}$ has greater jump than the corresponding jump of the process $\{X_n, n \geq 1\}$.

Let us consider the following unreliable system described by Dimitrov et *al*. (1991): during a process operating time a flow of implicit breakdowns with constant intensity arise in a random way, leading to incorrect final results. In such cases, it is profitable to introduce a strategy for making intermediate correctness test control and copies to remember the process states at some chosen moments. If the implicit breakdown is discovered by the test, the process continues from the last successful copied state. The tests and copies control schedule helps to economize the total process duration.

Consider the sequence $\{\alpha_k, k \geq 1\}$ of time intervals between consecutive copies and two independent renewal processes $\{\beta_k, k \geq 1\}$ and $\{\gamma_k, k \geq 1\}$, being testing and copying time durations, correspondingly. Next define

$$X_k = \begin{cases} \alpha_k + \beta_k + \gamma_k, & \text{if no breakdown is discovered by the test,} \\ 0, & \text{otherwise} \end{cases}$$

and

$$Y_k = \begin{cases} 0, & \text{if no breakdown is discovered by the test,} \\ \alpha_k + \beta_k, & \text{otherwise.} \end{cases}$$

for $k \geq 1$. Now, it is clear that Z_r defined by (1) can be interpreted as the total time duration of the unreliable server until the successful finish of the service, if the corresponding time duration without breakdowns is previously known.

Let us denote by $G_U(s) = E[s^U]$ the probability generating function of any integer-valued r.v. U, $|s| \leq 1$. Under the above notations the following theorem is true.

Theorem 1 *The distribution of Z_r is determined by its probability generating function*

$$G_{Z_r}(s) = \left[\frac{G_1(s)}{1 - G_2(s)}\right]^r, \tag{2}$$

where

$$G_1(s) = E\left[s^{X_1}\mathbf{I}(X_1 < Y_1)\right] = \sum_{k=0}^{\infty} p_k s^k \sum_{m=k+1}^{\infty} q_m, \tag{3}$$

$$G_2(s) = E\left[s^{Y_1}\mathbf{I}(X_1 \geq Y_1)\right] = \sum_{k=0}^{\infty} q_k s^k \sum_{m=k}^{\infty} p_m \tag{4}$$

and $\mathbf{I}(\bullet)$ means the indicator function.

Proof. Consider the decomposition of $G_{Z_r}(s)$ after the first jump of the processes $\{X_n, n \geq 1\}$ and $\{Y_n, n \geq 1\}$. Then

$$G_{Z_r}(s) = E\left[s^{Z_r}\mathbf{I}(X_1 < Y_1)\right] + E\left[s^{Z_r}\mathbf{I}(X_1 \geq Y_1)\right].$$

At first, let us suppose that $X_1 < Y_1$. In this case $Z_r = X_1 + T_1$, where T_1 means the total truncated sum after the first jump of both processes. The truncating process

$\{Y_n, n \geq 1\}$ is independent of the process $\{X_n, n \geq 1\}$ by assumption. Then, the r.v. T_1 is independent of X_1 and T_1 has the same distribution as Z_{r-1}. Therefore

$$E\left[s^{Z_r}\mathbf{I}(X_1 < Y_1)\right] = E\left[s^{X_1}\mathbf{I}(X_1 < Y_1)\right] E\left[s^{Z_{r-1}}\right] = G_1(s)G_{Z_{r-1}}(s).$$

We obtain the relation (3) by the following equations

$$G_1(s) = E\left[s^{X_1}\mathbf{I}(X_1 < Y_1)\right] = \sum_{k=0}^{\infty} P(X_1 = k)s^k P(X_1 < Y_1) = \sum_{k=0}^{\infty} p_k s^k \sum_{m=k+1}^{\infty} q_m.$$

Similarly, if $X_1 \geq Y_1$, we have $Z_r = Y_1 + T_2$, where T_2 is the total truncated sum after the first jump. Since after the first jump X_1 dominates Y_1, the r.v. T_2 is independent of Y_1 and has the same distribution as Z_r. Then

$$E\left[s^{Z_r}\mathbf{I}(X_1 \geq Y_1)\right] = E\left[s^{Y_1}\mathbf{I}(X_1 \geq Y_1)\right] E\left[s^{Z_r}\right] = G_2(s)G_{Z_r}(s).$$

In this case

$$G_2(s) = E\left[s^{Y_1}\mathbf{I}(X_1 \geq Y_1)\right] = \sum_{k=0}^{\infty} P(Y_1 = k)s^k P(X_1 \geq Y_1) = \sum_{k=0}^{\infty} q_k s^k \sum_{m=k}^{\infty} p_m,$$

as was stated by (4).

Combining both cases we have

$$G_{Z_r}(s) = G_1(s)G_{Z_{r-1}}(s) + G_2(s)G_{Z_r}(s).$$

The last equation is fulfilled for any integer $r \geq 1$, and using it iteratively we obtain

$$G_{Z_r}(s) = \left[\frac{G_1(s)}{1 - G_2(s)}\right]^r G_{Z_0}(s).$$

By convention $G_{Z_0}(s) = 1$, since $Z_0 = 0$ and therefore (2) is derived. □

Corollary. *Let $X_1, X_2, \ldots$ be geometrically distributed with parameter $p \in (0,1)$. For any truncating process $\{Y_n, n \geq 1\}$ with $q_0 < 1$, the distribution of Z_r is negative binomial with parameters p and r.*

Proof. In this case $p_k = (1-p)p^k$, $k = 0, 1, \ldots$ and from (3) and (4) we have

$$G_1(s) = \frac{1-p}{1-ps} G_{Y_1}(ps) \qquad \text{and} \qquad G_2(s) = 1 - G_{Y_1}(ps).$$

Substituting the last two expressions in (2) we obtain

$$G_{Z_r}(s) = \left(\frac{1-p}{1-ps}\right)^r,$$

which is the probability generating function of the negative binomial distribution. □

Remark. We acknowledge that the last two proofs are highly influenced by the corresponding proofs of Theorem 1 and Corollary 3 in Khalil et *al.* (1991), correspondingly. The statement of the Theorem 1 can be obtained also from Theorem 1 in Khalil et *al.* (1991). In fact, if we write $Z_r = \sum_{i=1}^{r} U_i$, with

$$U_i = \sum_{j=1+N_{i-1}}^{N_i - 1} Y_j + X_{N_i},$$

then the r.v.'s $U_1, \ldots, U_r$ are independent and with the same distribution.

The following characterization theorem of the negative binomial distribution in terms of random sums is obtained as a direct consequence of the above results.

Theorem 2 *Let us consider the geometric truncating process* $\{Y_n, n \geq 1\}$ *with parameter* $q \in (0,1)$. *Then the r.v.* Z_r *given by (1), is negative binomial distributed with parameters* p *and* r *iff* $X_1, X_2, \ldots$ *are geometrically distributed with parameter* p.

Let us note, that the necessary part of the Theorem 2 is true even if the truncating process $\{Y_n, n \geq 1\}$ is not geometric, as it was shown by the Corollary.

Bibliography

[1] B. Dimitrov, Z. Khalil, N. Kolev, P. Petrov On the optimal total processing time using checkpoints. *IEEE Transactions on Software Engineering*, **SE-17**, (1991), 436-442.

[2] N. L. Johnson, S. Kotz, A. W. Kemp. Univariate discrete distributions. Wiley Series in Probability and Mathematical Statistics. Second edition, 1992.

[3] Z. Khalil, B. Dimitrov, J. P. Dion. A characterization of the geometric distribution related to random sums. *Communication and Statistics - Stochastic Models* (1991) **7**, 321-326.

Nikolay Kolev
Dept.of Probability and Statistics
Institute of Mathematics
Bulgarian Academy of Sciences
Acad. G.Bontchev str., bl. 8
1113 Sofia, Bulgaria
e-mail: nkolev@math.bas.bg

Leda Minkova
Department of Stochastics and Optimization
Technical University of Sofia