NEGATIVE BINOMIAL DISTRIBUTION

Zuhriddin A. Nazarov
Technology, Management and Communication Institute in Tashkent.
Department of "Languages, Exact and Social Sciences".
zuhrov13@gmail.com
Munisa O. Ismoilova
Tashkent Perfect University, Center for Digital Learning Technologies.
munisaismoilova9718@gmail.com

Abstract: *This paper provides a concise yet rigorous exposition of the negative binomial distribution, which generalizes the classical binomial model to accommodate overdispersion in count data. Owing to its derivation as a Poisson-Gamma mixture, the distribution possesses remarkable flexibility in modeling heterogeneous stochastic phenomena. It naturally arises in branching processes, risk theory, queueing systems, and waiting time analysis.*

We outline its key probabilistic properties, discuss parameter estimation techniques and examine its asymptotic characteristics. The distribution's theoretical richness and practical relevance make it an indispensable tool in stochastic modeling across various applied domains such as epidemiology, actuarial science, and reliability engineering.

Keywords: *negative binomial distribution, branching processes, stochastic processes, Poisson and Gamma distributions, count data modeling, risk modeling, parameter estimation, queueing theory.*

INTRODUCTION

The negative binomial distribution is a fundamental discrete probability distribution widely used to model count data exhibiting overdispersion, where the variance exceeds the mean. It generalizes the binomial distribution by allowing a random number of trials before a fixed number of successes occur, making it applicable in various stochastic modeling scenarios; see [1]. One of the key areas where the negative binomial distribution plays a crucial role is in the modeling of branching processes, particularly in the study of population dynamics and genetic evolution; see [2]. Additionally, its deep connections with the Poisson and Gamma distributions enable flexible modeling of event occurrences in epidemiology, risk analysis, and queuing systems; see [3]–[4]. The negative binomial distribution can also be derived as a mixture of Poisson distributions with Gamma-distributed rates, which provides a natural framework for handling overdispersed data in practical applications; see [5]. Beyond theoretical significance, the negative binomial distribution has extensive applications in real-world problems, including biological studies, insurance risk assessment, and econometrics; see [6]. Its parameter estimation techniques, including maximum likelihood and Bayesian methods, have been widely explored to improve inference accuracy. This paper aims to discuss the fundamental properties, parameter estimation techniques, and important applications of the negative binomial distribution, emphasizing its practical significance in stochastic modeling.

Definition 1. The binomial distribution is given by

$$\mathrm{B}(n,k,p) := \mathrm{C}_n^k p^k q^{n-k}, \quad 0 \le k \le n, \quad n \in \mathrm{N}, \tag{1}$$

where $0 \le p \le 1$, $q = 1 - p$, $\mathrm{C}_n^k = n!/(k!(n-k)!)$ and N-natural numbers.

The distribution describes the probability of exactly k successes in n trials if the probability of a success in a single trial is p. It was first presented by J.Bernoulli in a work which was posthumously published in [7].

Example 1. If the probability of a sniper hitting the target is 0.7, then how many of the 5 bullets fired are more likely to hit the target?

The number of bullets that can hit the target is $k = 0,1,2,3,4,5.$ Then, we have

$$\mathrm{B}(5,0) := \mathrm{C}_5^0 \cdot (0.7)^0 \cdot (0.3)^{5-0} = (0.3)^5 = 243 \cdot 10^{-5},$$

$$\mathrm{B}(5,1) := \mathrm{C}_5^1 \cdot (0.7)^1 \cdot (0.3)^{5-1} = 5 \cdot 0.7 \cdot (0.3)^4 = 2835 \cdot 10^{-5},$$

$$\mathrm{B}(5,2) := \mathrm{C}_5^2 \cdot (0.7)^2 \cdot (0.3)^{5-2} = 10 \cdot (0.7)^2 \cdot (0.3)^3 = 13230 \cdot 10^{-5},$$

$$\mathrm{B}(5,3) := \mathrm{C}_5^3 \cdot (0.7)^3 \cdot (0.3)^{5-3} = 10 \cdot (0.7)^3 \cdot (0.3)^2 = 30870 \cdot 10^{-5},$$

$$\mathrm{B}(5,4) := \mathrm{C}_5^4 \cdot (0.7)^4 \cdot (0.3)^{5-4} = 5 \cdot (0.7)^4 \cdot 0.3 = 36015 \cdot 10^{-5},$$

$$\mathrm{B}(5,5) := \mathrm{C}_5^5 \cdot (0.7)^5 \cdot (0.3)^{5-5} = (0.7)^5 = 16807 \cdot 10^{-5}.$$

According to the above relations

$$\max_{i=1,2,3,4,5} \{\mathrm{B}(5,i)\} = \mathrm{B}(5,4).$$

Definition 2. The negative binomial distribution is given by

$$\mathrm{B}^-(n,k,p) := \mathrm{C}_{n-1}^{k-1} p^k q^{n-k}, \quad 0 < k \le n, \quad n \in \mathrm{N}, \tag{2}$$

where $\mathrm{C}_{n-1}^{k-1} = (n-1)!/((k-1)!(n-k)!)$.

Solution methods. The distribution expresses the probability of having to wait exactly n trials until k successes have occurred if the probability of a success in a single trial is p. The above form of the negative binomial distribution is often referred to as the Pascal distribution after the french mathematician, physicist and philosopher B.Pascal. The distribution is sometimes expressed in terms of the number of failures occurring while waiting for k successes, $N = n - k$, in which case we write

$$\mathrm{B}^-(N,k,p) := \bar{\mathrm{C}}_k^N p^k q^N, \quad N \in \mathrm{N}_0 := \{0\} \cup \mathrm{N}, \tag{3}$$

where

$$\bar{\mathrm{C}}_k^N = \mathrm{C}_{N+k-1}^N = \frac{(N+k-1)!}{N! \cdot (k-1)!}.$$

Changing variables, for this last form, to x and k instead of p and k we sometimes use

$$B^{-}(N,\xi,k) := \bar{C}_k^N \cdot \frac{\xi^n \cdot k^k}{(\xi+k)^{N+k}} = \bar{C}_k^N \cdot \left(\frac{\xi}{\xi+k}\right)^N \cdot \left(\frac{k}{\xi+k}\right)^k.$$

The distribution may also be generalized to real values of k, although this may seem obscure from the above probability view-point, writing the binomial coefficient as $(N+k-1)(N+k-2)\cdots(k+1)k/n!$.

Now we will calculate the numerical characteristics of the negative binomial distribution. In the first form given above the expectation value, variance, 3rd and 4th central moments of the distribution are

$$E(n) = \frac{k}{p}, \quad Var(n) = \frac{kq}{p^2}, \quad \mu_3 = \frac{kq(2-p)}{p^3} \quad \text{and} \quad \mu_4 = \frac{kq(p^2-6p+6+3kq)}{p^4}.$$

The coefficients of skewness and kurtosis (the coefficients of excess) are

$$\gamma_1 = \frac{2-p}{\sqrt{kq}} \quad \text{and} \quad \gamma_2 = \frac{p^2-6p+6}{kq}.$$

In the second formulation above $p(n)$, the only difference is that the expectation value becomes

$$E(N) = E(n) - k = \frac{k(1-p)}{p} = \frac{kq}{p}$$

while higher moments remain unchanged as they should since we have only shifted the scale by a fixed amount. In the last form given, using the parameters ξ and k, the expectation value and the variance are

$$E(N) = \xi \quad \text{and} \quad Var(N) = \xi + \frac{\xi^2}{k}.$$

Definition 3. The Poisson distribution is given by

$$P(n,\lambda) := \frac{\lambda^n e^{-\lambda}}{n!}, \quad n \in N_0, \qquad (4)$$

where the parameter λ is a real positive variable.

It is named after the french mathematician S.D.Poisson who was the first to present this distribution in 1837 (implicitly the distribution was known already in the beginning of the 18th century). The Poisson distribution is one of the most important distributions in statistics with many applications. Along with the properties of the distribution we give a few examples here but for a more thorough description we refer to standard text-books.

Example 2. Find the probability that at least 3 of the 1460 students in the faculty were born on the same day.

Let $p = 1/365$ and $n = 1460$, $\lambda = np = 4$. Then we have

$$P(A) = 1 - \sum_{k=0}^{2} \frac{4^k e^{-4}}{k!} \approx 0.61, \qquad (5)$$

where $P(A)$ the probability that at least 3 students were born on the same day.

Definition 4. The Gamma distribution is given by

$$f(x,\alpha,\beta) := \frac{\alpha(\alpha x)^{\beta-1}e^{-\alpha x}}{\Gamma(\beta)}, \qquad (6)$$

where the parameters α and β are positive real variables as is the variable x and $\Gamma(\beta)$ is the gamma function. For all positive integers, $\Gamma(k) = (k-1)!$. Note that the parameter α is simply a scale factor.

For $\beta \le 1$ the distribution is J-shaped and for $\beta > 1$ it is unimodal with its maximum at $x = (\beta - 1)/\alpha$. In the special case where β is a positive integer this distribution is often referred to as the Erlangian distribution. For $\beta = 1$ we obtain the exponential distribution and with $\alpha = 1/2$ and $\beta = n/2$ with n an integer we obtain the chi-squared distribution with n degrees of freedom.

Definition 5. The logarithmic distribution is given by

$$\Lambda(r,p) := -\frac{(1-p)^r}{r\ln p}, \quad r \in \mathbb{N}, \qquad (7)$$

where the parameter $0 < p < 1$ is a real variable.

Results and discussions. Generating functions serve as an important tool in the analysis of probability distributions. The probability generating function of the negative binomial distribution is given by

$$\Phi(z) = \left(\frac{pz}{1-qz}\right)^k$$

in the first case ($p(n)$) and

$$\Phi(z) = \left(\frac{k}{k+(1-z)\xi}\right)^k$$

in the second case ($p(N)$) for the two different parameterizations.

The negative binomial distribution exhibits several profound and practically important connections with other classical probability distributions, which not only enhance its theoretical value but also widen its scope of application in various fields such as reliability theory, queueing systems, risk modeling, and biological processes. In particular, these connections reveal underlying structural similarities and allow for insightful interpretations in terms of mixture models, limiting behaviors, and compound distributions. For example, the negative binomial distribution can be represented as a Poisson-Gamma mixture, which offers a natural explanation for its overdispersion property relative to the Poisson distribution. Furthermore, under certain parameterizations, the negative binomial distribution converges to the Poisson distribution, establishing a limiting relationship that bridges discrete distributions under specific conditions. In the subsequent subsections, we systematically explore and discuss some of these fundamental connections, highlighting both their mathematical formulations and their implications for statistical modeling and inference.

Theorem 1. Let $N > 0$, $k > 0$ and $\lambda > 0$. Then

$$\lim_{k\to\infty} B^-(N,\lambda,k) = P(N,\lambda).$$

Proof of Theorem 1. In what follows, we consider the negative binomial distribution in the form

$$B^-(N,\lambda,k) := \bar{C}_k^N \cdot \frac{\lambda^n \cdot k^k}{(\lambda+k)^{N+k}} = \bar{C}_k^N \cdot \left(\frac{\lambda}{\lambda+k}\right)^N \cdot \left(\frac{k}{\lambda+k}\right)^k = \quad (8)$$
$$= C_{N+k-1}^N \cdot \left(\frac{\lambda/k}{1+\lambda/k}\right)^N \cdot \left(\frac{1}{1+\lambda/k}\right)^k.$$

Then we have

$$C_{N+k-1}^N = \frac{(N+k-1)!}{N!\cdot(k-1)!} \to \frac{k^N}{N!}, \quad (9)$$

$$\left(\frac{1}{1+\lambda/k}\right)^k \to e^{-\lambda} \quad \text{and} \quad \left(\frac{\lambda/k}{1+\lambda/k}\right)^N \cdot k^N \to \lambda^N \quad as\ k\to\infty. \quad (10)$$

From the combination of relations (9) and (10) above, we obtain the proof of the theorem.

Theorem 2. Let $z = N/\lambda$ and $k << \lambda \to \infty$. Then

$$\lim_{n\to\infty} B^-(N,\lambda,k) = \frac{f(z,k,k)}{\lambda}.$$

Proof of Theorem 2. By substituting $z = N/\lambda$ in equation (8), we have

$$B^-(z,\lambda,k) = B^-(N,\lambda,k)\frac{dN}{dz} = \lambda C_{z\lambda+k-1}^{z\lambda} \cdot \left(\frac{\lambda/k}{1+\lambda/k}\right)^{z\lambda} \cdot \left(\frac{1}{1+\lambda/k}\right)^k =$$
$$= \frac{(z\lambda+k-1)(z\lambda+k-2)\cdot\ldots\cdot(z\lambda+1)}{\Gamma(k)}\lambda k^k \cdot \left(\frac{1}{1+k/\lambda}\right)^{z\lambda} \cdot \left(\frac{1}{k+\lambda}\right)^k \to$$
$$\to \lambda k^k \cdot \frac{(z\lambda)^{k-1}}{\Gamma(k)} \cdot \left(\frac{1}{1+k/\lambda}\right)^{z\lambda} \cdot \left(\frac{1}{k+\lambda}\right)^k =$$
$$= \frac{z^{k-1}k^k}{\Gamma(k)} \cdot \left(\frac{1}{1+k/\lambda}\right)^{z\lambda} \cdot \left(\frac{\lambda}{k+\lambda}\right)^k \to \frac{z^{k-1}k^k e^{-kz}}{\Gamma(k)},$$

(11)

where we have used that $[\lambda/(k+\lambda)]^k \to 1$ and

$$\left(\frac{1}{1+k/\lambda}\right)^{z\lambda} \to e^{-kz} \quad \text{as}\ n\to\infty. \quad (12)$$

Thus, it has been demonstrated that as $n \to \infty$ and under the condition $\xi \gg k$, the limiting distribution of the variable z converges to the Gamma distribution. The proof is completed.

Theorem 3. Let $L(N,p)$ - the logarithmic distribution. Then

$$\lim_{k \to 0} B^{-}(N,\xi,k) = L(N,p).$$

Proof of Theorem 3. The probabilities for $N \in \mathbb{N}_0$ are given by

$$\{p(0),\ p(1),\ p(2),\dots\} = p^k \cdot \{1,\ kq,\ k(k+1)q^2/2!,\dots\}$$

if we omit the zero class and renormalize we get

$$\frac{kp^k}{1-p^k}\cdot\{0,\ q,\ (k+1)q^2/2!,\dots\}$$

and if we let $k \to 0$ we finally obtain

$$-\frac{1}{\ln p}\cdot\{0,\ q,\ q^2/2,\dots\},$$

where we have used that

$$\lim_{k \to 0}\frac{p^k}{1-p^k} = -\frac{1}{\ln p}$$

which is easily realized expanding $p^{-k} = \exp\{-k\ln p\}$ into a power series. This we recognize as the logarithmic distribution

$$L(N,p) := -\frac{1}{\ln p}\cdot\frac{(1-p)^N}{N},$$

thus we have shown that omitting the zero class and letting $k \to 0$ the negative binomial distribution becomes the logarithmic distribution. It is a limiting form of the negative binomial distribution when the zero class has been omitted and the parameter $k \to 0$. The theorem is proved.

Appendix. In our subsequent work, we intend to explore the relationship between the negative binomial distribution and the theory of branching processes in greater depth. Specifically, we will examine a process wherein branching transitions from a Poisson distribution to a logarithmic distribution. In such cases, one of the most elegant and effective methods for deriving the resulting distribution is through the application of probability generating functions. The probability generating functions corresponding to the Poisson distribution with mean parameter μ and the logarithmic distribution with parameter p are presented in expression

$$G_P(z) := \exp\{\mu(z-1)\} \quad \text{and} \quad G_L(z) := \frac{\ln(1-zq)}{\ln(1-q)} = \alpha\ln(1-zq),$$

where $\mu > 0$, $0 \le q \le 1$ and $\alpha = 1/\ln p$. For a branching process in n steps

$$G(z) = G_1(G_2(..G_{n-1}(G_n(z))...)),$$

where $G_k(z)$ is the probability generating function in the k th step. In the above case this gives

$$G(z) = G_P(G_L(z)) = \exp\{\mu(\alpha \ln(1 - zq) - 1)\} = \frac{p^k}{(1 - zq)^k},$$

where we have put $k = -\alpha\mu$. This expression can be readily recognized as the probability generating function corresponding to a negative binomial distribution with parameters k and p. Consequently, we have demonstrated that a Poisson distribution with mean μ, when subjected to branching governed by a logarithmic distribution with parameter p, results in a negative binomial distribution characterized by parameters $k = -\alpha\mu = -\mu/\ln p$ and p or $\xi = kq/p$. Conversely, any negative binomial distribution with parameters k and p (or ξ) can be interpreted as arising from a compound process involving a Poisson distribution with parameter $\mu = -k\ln p = k\ln(1 + \xi/k)$ and a logarithmic distribution with parameters p and ξ/μ; see [8]–[10].

Conclusion. The negative binomial distribution is a crucial tool in probability theory and statistical modeling, especially for count data with overdispersion. Its connections with the Poisson and Gamma distributions provide flexibility in various applications, including branching processes, epidemiology, and risk analysis. The distribution’s parameter estimation techniques enhance its usability in real-world problems. By exploring its fundamental properties and applications, this study highlights the significance of the negative binomial distribution in stochastic modeling and encourages further research on its advanced applications.

REFERENCES:

1. Johnson N.L., Kotz S., Kemp A.W. (2005). Univariate Discrete Distributions. Wiley.
2. Athreya K.B., Ney P.E. (2004). Branching Processes. Springer.
3. Cameron, A.C., Trivedi, P.K. (2013). Regression Analysis of Count Data. Cambridge University Press.
4. Kingman J.F.C. (1993). Poisson Processes. Oxford University Press.
5. Hilbe J.M. (2011). Negative Binomial Regression. Cambridge University Press.
6. McCullagh P., Nelder J.A. (1989). Generalized Linear Models. Chapman & Hall.
7. Ars Conjectandi by Jacques Bernoulli, published posthumously in 1713.
8. Borovkov A.A.: Probability Theory. URSS, Moscow, 2009.
9. Себастьянов Б.А.: Курс теории вероятностей и математическая статистика. Москва, Наука, 1982.
10. Феллер В.: Введение в теорию вероятностей и ее приложения. Москва, Мир, 1984.