Computational Journal of Mathematical and Statistical Sciences
4(xxxx), 1–25
DOI:10.21608/cjmss.2025.357523.1119
https://cjmss.journals.ekb.eg/

Research article

Characterization of Some Probability Distributions Based on Conditional Expectation and Variance

Mahmoud A. Eltehiwy[1,*], **Noura A. Abu El-Magd**[2]

[1] Department of Statistics, Faculty of the Politics and Economics, Beni-Suef University, 62521 Beni-Suef, Egypt; Mahmoud.abdelmone@eps.bsu.edu.eg.

[2] Department of Statistics, Faculty of the Politics and Economics, Beni-Suef University, 62521 Beni-Suef, Egypt; Noura.Shon@eps.bsu.edu.eg.

* **Correspondence:** Mahmoud.abdelmone@eps.bsu.edu.eg

Abstract: The characterizations of probability distributions have been studied by several academics. When a certain distribution is the only one that correlates with a given property, a characterization theorem in probability and statistics is applicable. Additionally, a characterization is a particular distributional or statistical property of a statistic or statistics that describes the corresponding stochastic model in a unique way. Some scholars proposed that real-world data should be characterized under certain criteria before a probability distribution is applied to it. This research investigates the characterization of specific probability distributions using advanced statistical methods. Characterization theorems are fundamental in probability and statistics, as they establish the unique properties that define a given distribution. The study focuses on a novel approach to characterizing truncated negative binomial and logarithmic series distributions through conditional expectation and variance functions. The necessary and sufficient conditions are derived for these characterizations, with a particular emphasis on the failure rate as a key parameter. The findings provide a robust framework for identifying underlying distributions based on their failure rates and conditional properties. An illustrative example is presented to demonstrate the practical application of this methodology, showcasing its potential for uncovering the properties of unobserved distributions.

Keywords: Characterization, conditional expectation and variance, truncated negative binomial distribution, logarithmic series distribution.

Mathematics Subject Classification: 62E10, 39B55, 62E15.

Received: 2 February 2025; Revised: 19 May 2025; Accepted: 27 May 2025; Online: 11 June 2025.

1. Introduction

In statistical inference, probability distribution characterizations are crucial topics. The literature has published numerous articles on characterizing distributions in statistics, including those by Arnold [3], Srivastava [33], Sheetrari [31], Zijlstra [43], and Dallas [9]. Numerous results are covered in the survey book by Kagan et al. [19], the most current remarks by Galambos and Kotz [11], and Nagaraja [23]. The survey of Kotz [20] has a fair summary of the main characterization directions. In particular, several characterizations of the exponential distribution have been extensively discussed because of its simplicity and applicability. Quite a lot of attention has been paid to the characterization of the Poisson distribution by Rao and Rubin [26], which used the Bernstein representation theorem. Srivastava and Srivastava [34] used similar concepts to characterize the binomial model. Shanbhag and Clark [29] established a variation of the Rao-Rubin findings. The 3×3 Bhattacharya covariance matrix's diagonality was used by Shanbhag [28] to characterize the Poisson and binomial distributions from the family of random variables with exponential-type distribution. Gamma distribution and negative binomial distribution (NBD) were characterized by Osaki and Li [25]. Using the mean residual life function (MRF) in combination with the failure rate of the characterized distributions, Ahmed [1] used the connection between conditional expectations and hazard rate functions to characterize the beta, binomial, and Poisson distributions. Kyriakoussis [21] characterized the modified logarithmic series distribution (LSD) using the conditional distribution of two random variables, X and Y, and the regression function $E(\mathrm{Y} \mid Y = y)$. Ramalingam and Jagbir [30] used the uniformly minimum variance unbiased estimator of θ in a sample of size n to characterize LSD. He also used his characterization to build a test statistic to test whether the sample follows an LSD. Su and Huang [36] investigated conditional expectation-based distribution characterizations. Nanda [24] conducted a study on the characterizations of entirely continuous random variables using the failure rate functions and MRF. The characterization of distributions by truncated moments was introduced by Laurent [22]. The distributions were described by Gupta and Gupta [12] using the seconds of residual life.

In life testing situations, if X is the life of a component, then the random variable $(X - m \mid X > m)$ is called the residual life function (RLF), and the quantity $E(X - m \mid X > m)$ is called the MRF. Another quantity which has been of great interest is the variance of RLF, $\sigma_F^2(m) = V(X \mid X > m)$, see, e.g. Dallas [8], Hall and Hellner [14], and Yehia and Ahmed [39]. Gupta [13] studied the monotonicity properties of $\sigma_F^2(m)$ for some non-parametric families of distributions commonly appearing in reliability theory.

New advancements have further enriched the field by introducing novel distributions and characterization methods tailored to complex data scenarios. For instance, Almetwally et al. [2] characterized the Marshall-Olkin alpha power Rayleigh distribution using moment-based properties, showing its utility in engineering applications. Similarly, Hamedani et al. [15] explored the type I quasi-Lambert family through regression-based characterizations, offering insights into flexible modeling. Ibrahim et al. [17] expanded the Nadarajah-Haghighi model with copula-based approaches, incorporating censored and uncensored validation, while Shehata et al. [32] validated the Nikulin-Rao-Robson distribution using goodness-of-fit measures for censored data. In the realm of lifetime distributions, Almetwally et al. [40] proposed an overview of discrete distributions, and Alyami et al. [41] introduced an new life time distribution model with applications in medical and engineering. Furthermore, Yousof et al. [42] developed a discrete generator for count data, characterized through moment and reliability properties, aligning closely with the discrete modeling focus of our study. These contemporary works underscore the ongoing evolution of characterization techniques, motivating our investigation into distributions with specific structural constraints, such as truncation. This study provides the necessary and sufficient conditions for characterizing the zero-truncated negative binomial distribution (ZTNBD) and LSD using conditional expectation and variance in terms of failure rates. Previous works, such as those of Ahmed [1] for binomial distribution, Osaki and Li [25] for NBD or Kyriakoussis [21] for LSD, have utilized different approaches (e.g., regression functions or mean residual life). Our approach leverages the failure rate, a widely interpretable and measurable quantity in reliability and survival analysis, to uniquely identify these distributions, offering a fresh perspective on their statistical properties. The application of conditional expectation and variance to truncated distributions such as ZTNBD is particularly innovative. Although the standard NBD has been extensively studied, the zero-truncated variant, common in data sets that exclude zero counts (e.g., healthcare, economics, and marketing), has received less attention in terms of characterization. In this study, we try to fill this gap by providing a robust theoretical framework tailored to such scenarios. By integrating failure rates into the characterization process, our work unifies concepts from reliability theory with probabilistic modeling. This is distinct from previous characterizations that relied on other statistical properties (e.g., Bhattacharya covariance by Shanbhag [28] or truncated moments by Laurent [22]). This unification enhances the theoretical understanding of these distributions and their behavior under truncation. A necessary and sufficient condition for identifying ZTNBD is given in Section 2. A necessary and sufficient condition required for determining LSD is provided in Section 3. Examples of the uses of these results are provided in Section 4.

2. Characterization of Negative Binomial Truncated at Zero

NBD is widely recognized for its ability to model count data characterized by overdispersion, where the variance exceeds the mean. This distribution finds applications across various disciplines due to its flexibility and robustness. For instance, in accident statistics, it is used to model the frequency of incidents, while in demographic studies, it effectively represents stochastic changes in population sizes (see Hilbe [16]). In psychology, NBD is often applied to model event frequencies, whereas in economics, it serves as a logarithmic distribution for analyzing time series data; see Cameron and Trivedi [4]. In addition, it is frequently employed in market research to analyze consumer behavior, in medical studies for disease modeling, and in military contexts for resource allocation. Its utility also extends to library science, where it models book-lending patterns. NBD's adaptability to a wide range of data scenarios, particularly those involving overdispersion, has made it a preferred alternative to the Poisson distribution, which assumes equal mean and variance; see Ridout et al. [27]. NBD has the form

$$f(z, N, \theta) = \binom{N+z-1}{N-1}(\theta)^N(1-\theta)^z, \quad z = 0, 1, \ldots; N > 0 \tag{2.1}$$

so that

$$f(0) = (\theta)^N.$$

ZTNBD is used to model scenarios where zero occurrences are excluded, making it suitable for datasets that inherently lack zero counts. To calculate probabilities for the truncated distribution, the probability mass function (PMF) of the standard NBD must be adjusted by dividing it by $[1 - f(0)]$, where $f(0)$

represents the probability of zero occurrences. This adjustment ensures that the probabilities sum to one, maintaining the properties of a valid distribution. The resulting PMF is expressed in terms of the number of failures (z) observed before achieving a fixed number of successes (N), and it is applicable in cases such as modeling repeated attempts or trial outcomes where a minimum threshold must be met (see Hilbe [16]). This distribution is particularly valuable in fields like economics, healthcare, and marketing, where truncated data is common, and it offers a flexible framework for analyzing non-zero event frequencies. The PMF is

$$f(z, N, \theta) = \frac{\binom{N+z-1}{N-1}(\theta)^N(1-\theta)^Z}{1-(\theta)^N}, \tag{2.2}$$

where $z = 1, 2, \ldots$, and $f(z)$ represents the probability of observing z failures before achieving the N$^{\text{th}}$ success in a total of $(N + z)$ trials.

This distribution is reparameterized in the statistical literature using $Q = \frac{1}{\theta}$ and $P = \frac{1-\theta}{\theta}$, such that $Q - P = 1$, enabling more flexible analysis in various applications. Its PMF accounts for the truncation at zero by adjusting the standard negative binomial PMF to ensure valid probability assignments, particularly for scenarios involving non-zero events; see Hilbe [16].

$$f(z) = \left(1 - Q^{-N}\right)^{-1} \binom{N+z-1}{N-1} \left(\frac{P}{Q}\right)^z \left(1 - \frac{P}{Q}\right)^N. \tag{2.3}$$

ZTNBD is characterized by its mean $E(Z) = \frac{NP}{1-Q^{-N}}$ and variance $V(Z) = \frac{NPQ}{1-Q^{-N}}\left[1 - \frac{P}{Q}\right]\left[\left(1 - Q^{-N}\right)^{-1} - 1\right]$, providing a robust framework for modeling overdispersed count data in fields such as healthcare, finance, and ecological studies. The use of conditional expectations further enhances its applicability by allowing precise modeling of truncated data scenarios.

Theorem 2.1. *Let Z be a positive discrete random variable with PMF defined in (2.3). Let $\frac{P}{Q} = q$, and $\left(1 - \frac{P}{Q}\right) = \omega$, then Z is ZTNBD with PMF $f(z; \omega)$ if and only if (iff)*

$$E(Z \mid Z > r) = \frac{Nq}{\omega} + \frac{r+1}{\omega} h(r+1), \quad \text{for all integers } r \geq 0 \tag{2.4}$$

where $h(r+1) = f(r+1)/\bar{F}(r+1)$ is a failure rate at $r+1$ and $\bar{F}(r+1)$ is the survival function at $r+1$.

Under the conditions of Theorem 2.1, we start by proving the following set of lemmas.

Lemma 2.1.

$$f(r+2; \omega) = \frac{q^{r+1}}{N}\binom{N+r+1}{N-1} f(1; \omega), \quad r = 0, 1, 2, \ldots$$

Proof:

$$f(r+2; \omega) = \left(1 - Q^{-N}\right)^{-1} \binom{N+r+2-1}{N-1} q^{r+2}\omega^N, \quad r = 0, 1, 2, \ldots$$

$\omega + q = 1, \qquad 0 \leq \omega, q \leq 1$

$$
\begin{aligned}
&= \left(1 - Q^{-N}\right)^{-1} \frac{(N+r+1)!}{(N-1)!(r+2)!} q^{r+2}\omega^N \\
&= \frac{q(N+r+1)}{r+2}\left(1 - Q^{-N}\right)^{-1} \binom{N+r}{N-1} q^{r+1}\omega^N \\
&= \frac{q(N+r+1)}{r+2} f(r+1;\omega) \\
&\vdots \\
&\vdots \\
&= \frac{q^{r+1}}{N}\binom{N+r+1}{N-1} f(1;\omega)
\end{aligned}
$$

Lemma 2.2. $S_A = \sum_{z=r+1}^{\infty} z f(z;\omega) = \frac{Nq}{\omega}\bar{F}(r+1;\omega) + \frac{r+1}{\omega} f(r+1;\omega)$.
Proof:

$$
\begin{aligned}
S_A &= \sum_{z=r+1}^{\infty} z f(z;\omega) \\
&= \left(1 - Q^{-N}\right)^{-1} \sum_{z=r+1}^{\infty} z \frac{(N+z-1)!}{(N-1)!z!} q^z \omega^N \\
&= \left(1 - Q^{-N}\right)^{-1} \sum_{z=r+1}^{\infty} \frac{(N+z-1)!}{(N-1)!(z-1)!} q^z \omega^N, \text{ using } (z-1) = j \\
&= \left(1 - Q^{-N}\right)^{-1} \sum_{j=r}^{\infty} \frac{(N+j)!}{(N-1)!(j)!} q^{j+1} \omega^N \\
&= Nq\left(1 - Q^{-N}\right)^{-1} \sum_{j=r}^{\infty} \binom{N+j-1}{N-1} q^j \omega^N \\
&\quad + q\left(1 - Q^{-N}\right)^{-1} \sum_{j=r}^{\infty} j \binom{N+j-1}{N-1} q^j \omega^N \\
&= Nq\bar{F}(r;\omega) + q\left[r f(r;\omega) + S_A\right] \\
\therefore S_A - qS_A &= Nq\bar{F}(r;\omega) + qr f(r;\omega) \\
\therefore \omega S_A &= Nq\bar{F}(r;\omega) + qr f(r;\omega), \text{ so} \\
S_A &= \frac{Nq\bar{F}(r;\omega)}{\omega} + \frac{qr f(r;\omega)}{\omega}.
\end{aligned}
$$

Since

$$
\bar{F}(r+1;\omega) = \bar{F}(r;\omega) - f(r;\omega),
$$

so

$$
S_A = \frac{Nq}{\omega}[\bar{F}(r+1;\omega) + f(r;\omega)] + \frac{qr}{\omega} f(r;\omega)
$$

$$S_A = \frac{Nq}{\omega}\bar{F}(r+1;\omega) + \frac{q(N+r)}{\omega} f(r;\omega)$$

By Lemma 2.1

$$S_A = \frac{Nq}{\omega}\bar{F}(r+1;\omega) + \frac{r+1}{\omega} f(r+1;\omega).$$

Proof of Theorem 2.1

A. Necessity. Using Lemma 2.2, we have

$$\begin{aligned} E(Z \mid Z > r) &= S_A/\bar{F}(r+1;\omega) \\ &= \frac{Nq}{\omega} + \frac{r+1}{\omega} h(r+1), \text{ for all integers } r \geq 0 \end{aligned}$$

B. Sufficiency. Eq (2.4) may be written using an unknown function $f(z)$:

$$\sum_{z=r+1}^{\infty} z f(z) = \frac{Nq}{\omega} \sum_{z=r+1}^{\infty} f(z) + \frac{r+1}{\omega} f(r+1).$$

As a result, we may get the corresponding set of equations shown below:

$$(r+1)f(r+1) = \frac{Nq}{\omega} f(r+1) + \frac{r+1}{\omega} f(r+1) - \frac{r+2}{\omega} f(r+2), \quad r = 0, 1, 2, \ldots$$

We've rearranged both sides:

$$f(r+2) = \frac{q(N+r+1)}{r+2} f(r+1).$$

By using Lemma 2.1 to the recurrent formula above, we get

$$f(z;\omega) = \frac{q^{z-1}}{N} \binom{N+z-1}{N-1} f(1;\omega), \quad z = 1, 2, \ldots$$

Using $\sum_{z=1}^{\infty} f(z;\omega) = 1$, we have

$$f(1;\omega) = N\left(1 - Q^{-N}\right)^{-1} q\omega^N.$$

Then

$$f(z;\omega) = \left(1 - Q^{-N}\right)^{-1} \binom{N+z-1}{N-1} \left(\frac{P}{Q}\right)^z \left(1 - \frac{P}{Q}\right)^N, \quad z = 1, 2, \ldots$$

It is the ZTNBD PMF.

We use conditional variance to characterize a ZTNBD in the following:

Theorem 2.2. *A positive discrete random variable Z has a ZTNBD iff*

$$V(Z \mid Z > r) = A(h), \tag{2.5}$$

where $A(h) = \frac{Nq}{\omega^2} - \frac{(r+1)^2}{\omega^2} h^2(r+1) + \frac{r+1}{\omega^2}(1 + r\omega - Nq)h(r+1)$. *For all integers* $r \geq 0$, *where* $h(r+1) = f(r+1)/\bar{F}(r+1)$ *is the failure rate at* $r+1$.

Under the conditions of Theorem 2.2, we start by proving the following set of lemmas.

Lemma 2.3.

$$B = \sum_{z=r+1}^{\infty} z^2 f(z;\omega) = \frac{Nq(Nq+1)}{\omega^2}\bar{F}(r+1) + \frac{r+1}{\omega^2}(Nq - rq + r + 1)f(r+1). \tag{2.6}$$

Proof:

$$\begin{aligned} B = \sum_{z=r+1}^{\infty} z^2 f(z;\omega) &= \sum_{z=r+1}^{\infty} z(z-1) f(z;\omega) + \sum_{z=r+1}^{\infty} z f(z;\omega) \\ &= \sum_{z=r+1}^{\infty} z(z-1) f(z;\omega) + S_A, \end{aligned}$$

where

$$S_A = \sum_{z=r+1}^{\infty} z f(z;\omega)$$

So,

$$\begin{aligned} \sum_{z=r+1}^{\infty} z(z-1) f(z;\omega) &= \left[1 - Q^{-N}\right]^{-1} \sum_{z=r+1}^{\infty} z(z-1)\frac{(N+z-1)!}{(N-1)!z!} q^z \omega^N \\ &= \left[1 - Q^{-N}\right]^{-1} \sum_{z=r+1}^{\infty} \frac{(N+z-1)!}{(N-1)!(z-2)!} q^z \omega^N \end{aligned}$$

using $(z-2) = j$

$$= \left[1 - Q^{-N}\right]^{-1} \sum_{j=r-1}^{\infty} \frac{(N+j+1)!}{(N-1)!j!} q^{j+2} \omega^N$$

$$\begin{aligned} = N^2 q^2 \left[1 - Q^{-N}\right]^{-1} &\sum_{j=r-1}^{\infty} \frac{(N+j-1)!}{(N-1)!j!} q^j \omega^N \\ &+ 2Nq^2 \left[1 - Q^{-N}\right]^{-1} \sum_{j=r-1}^{\infty} j\frac{(N+j-1)!}{(N-1)!j!} q^j \omega^N \\ &+ q^2 \left[1 - Q^{-N}\right]^{-1} \sum_{j=r-1}^{\infty} j^2\frac{(N+j-1)!}{(N-1)!j!} q^j \omega^N \\ &+ q^2 \left[1 - Q^{-N}\right]^{-1} \sum_{j=r-1}^{\infty} j\frac{(N+j-1)!}{(N-1)!j!} q^j \omega^N \\ &+ q^2 N \left[1 - Q^{-N}\right]^{-1} \sum_{j=r-1}^{\infty} \frac{(N+j-1)!}{(N-1)!j!} q^j \omega^N \end{aligned}$$

$$\begin{aligned} &= N^2 q^2 \bar{F}(r-1) + 2Nq^2\left[(r-1)f(r-1) + rf(r) + S_A\right] \\ &+ q^2\left[(r-1)^2 f(r-1) + r^2 f(r) + B\right] + q^2[(r-1)f(r-1) \\ &+ rf(r) + S_A] + Nq^2\bar{F}(r-1). \end{aligned}$$

So,

$$
\begin{aligned}
B = & N^2q^2\bar{F}(r-1) + 2Nq^2(r-1)f(r-1) + 2Nrq^2f(r) + 2Nq^2S_A \\
& + q^2(r-1)^2f(r-1) + r^2q^2f(r) + q^2B + q^2(r-1)f(r-1) + rq^2f(r) \\
& + q^2S_A + Nq^2\bar{F}(r-1) + S_A.
\end{aligned}
$$

Since

$$
\bar{F}(r) = \bar{F}(r-1) - f(r-1),
$$

and

$$
\begin{aligned}
f(r) &= \frac{q(N+r-1)}{r} f(r-1), \\
f(r+1) &= \frac{q(N+r)}{r+1} f(r).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
B = & \frac{1}{\omega\,(1-q^2)}[Nq(Nq+1)(1+q)]\bar{F}(r+1) \\
& + \frac{r+1}{\omega\,(1-q^2)}[(1+q)(r-rq+Nq+1)]f(r+1) \\
= & \frac{1}{\omega(1-q)}[Nq(Nq+1)]\bar{F}(r+1) + \frac{r+1}{\omega(1-q)}[r-rq+Nq+1]f(r+1) \\
= & \frac{1}{\omega^2}\left[Nq(Nq+1)\bar{F}(r+1) + \frac{r+1}{\omega^2}[Nq-rq+r+1]f(r+1).\right.
\end{aligned}
$$

Then

$$
E\left(Z^2 \mid Z > r\right) = B/\bar{F}(r+1) = \frac{Nq(Nq+1)}{\omega^2} + \frac{(r+1)(Nq+r\omega+1)}{\omega^2}h(r+1). \tag{2.7}
$$

***Proof of Theorem* 2.2**

A. Necessity. Using Lemma 2.3, we have the following:

$$
\begin{aligned}
V(Z \mid Z > r) = & E\left(Z^2 \mid Z > r\right) - E^2(Z \mid Z > r) \\
= & \frac{Nq(Nq+1)}{\omega^2} + \frac{(r+1)(Nq-rq+r+1)}{\omega^2}h(r+1) - \frac{N^2q^2}{\omega^2} - \frac{2Nq(r+1)}{\omega^2}h(r+1) \\
& - \frac{(r+1)^2}{\omega^2}h^2(r+1) \\
= & \frac{Nq}{\omega^2} + \frac{(r+1)(1-Nq+r\omega)}{\omega^2}h(r+1) - \frac{(r+1)^2}{\omega^2}h^2(r+1); \quad \text{for all integers } r \geq 0.
\end{aligned}
$$

B. Sufficiency. Eq (2.7) can be written with an unknown function $f(z)$:

$$
\sum_{z=r+1}^{\infty} z^2 f(z) = \frac{Nq(Nq+1)}{\omega^2}\sum_{z=r+1}^{\infty} f(z) + \frac{r+1}{\omega^2}[Nq-rq+r+1]f(r+1).
$$

Therefore, we can get the following equivalent set of equations

$$(r+1)^2 f(r+1) = \frac{Nq(Nq+1)}{\omega^2} f(r+1) + \frac{r+1}{\omega^2}(Nq - rq + r + 1) f(r+1) - \frac{r+2}{\omega^2}$$
$$(Nq - (r+1)q + r + 2) f(r+2);\; r = 0, 1, 2, \ldots$$

Rearranging both sides, we have

$$f(r+2) = \frac{q(N+r+1)}{r+2} f(r+1).$$

Solving the above recurrence formula, we have, by Lemma 2.1

$$f(z;\omega) = \frac{q^{z-1}}{N}\binom{N+z-1}{N-1} f(1;\omega); \quad z = 1, 2, \ldots$$

Using the fact

$$\sum_{z=1}^{\infty} f(z;\omega) = 1, \quad \text{we have}$$

$$f(1,\omega) = N\left[1 - Q^{-N}\right]^{-1} q\omega^N, \quad \text{Viz,}$$

$$f(z;\omega) = \left[1 - Q^{-N}\right]^{-1}\binom{N+z-1}{N-1} q^z \omega^N, \quad z = 1, 2, \ldots$$

which is the ZTNBD.

3. Characterization of Logarithmic Series Distribution

LSD was first proposed by Fisher et al. [10] to study butterfly populations in the Malay Peninsula. Williams [37] used LSD to model various datasets, such as the distribution of parasites, including the count of head lice per host. Furthermore, Williams [36] applied this distribution to analyze the number of publications by entomologists. Another notable application was by Williamson and Bretherton [38], who utilized it to address an inventory control issue in the steel industry.

In a series of publications, including works by Chatfield et al. [7] and Chatfield ([5], [6]), the authors discussed the use of LSD to model the distribution of quantities of a product purchased by a buyer within a specific time frame. They noted that LSD can serve as a useful approximation for NBD when the parameter N is very low (e.g., less than 0.1). The advantage of LSD lies in its reliance on only one parameter θ, as opposed to the two parameters, N and p, required for NBD. In view of the wide application of LSD, a characterization technique for such a distribution is presented.

The following theorems characterize LSD among discrete distributions defined on the set of positive integers.

Theorem 3.1. *Let Z be a non-negative discrete random variable with PMF $f(z;\theta)$ and mean $\alpha\theta(1-\theta)^{-1}$; then Z has a LSD with*

$$f(z;\theta) = \frac{\alpha\theta^z}{z}, \quad z = 1, 2, \ldots; \quad 0 < \theta < 1, \text{ and } \alpha = -1/\ln(1-\theta); \text{ iff}$$

$$E(Z \mid Z > r) = \frac{r+1}{1-\theta} h(r+1), \qquad \textit{for all integer } r \geq 0. \tag{3.1}$$

Proof:
We prove the theorem in two parts:
Necessity: If Z follows the LSD, then the conditional expectation $E(Z \mid Z > r)$ satisfies the given form in (3.1).
Sufficiency: If the conditional expectation $E(Z \mid Z > r)$ satisfies the given form in (3.1), then Z must follow the LSD.
1. Necessity
Assume Z follows the LSD, we need to show that the conditional expectation $\mathbb{E}[Z \mid Z > r]$ satisfies

$$\mathbb{E}[Z \mid Z > r] = \frac{r+1}{1-\theta} \cdot \frac{f(r+1)}{\overline{F}(r+1)},$$

$$\mathbb{E}[Z \mid Z > r] = \frac{\sum_{z=r+1}^{\infty} z f(z;\theta)}{\sum_{z=r+1}^{\infty} f(z;\theta)}.$$

Substitute the PMF $f(z;\theta) = \frac{\alpha\theta^z}{z}$. Then

$$\mathbb{E}[Z \mid Z > r] = \frac{\sum_{z=r+1}^{\infty} z \cdot \frac{\alpha\theta^z}{z}}{\sum_{z=r+1}^{\infty} \frac{\alpha\theta^z}{z}}.$$

Simplify the numerator and denominator:

$$\mathbb{E}[Z \mid Z > r] = \frac{\sum_{z=r+1}^{\infty} \theta^z}{\sum_{z=r+1}^{\infty} \frac{\theta^z}{z}}.$$

The numerator, $\sum_{z=r+1}^{\infty} \theta^z$, is an infinite geometric series starting at $z = r+1$. The first term is θ^{r+1}, and the common ratio is θ. For a geometric series,

$$\sum_{k=0}^{\infty} ar^k = \frac{a}{1-r}, \qquad \text{where } |r| < 1,$$

we adjust the starting index accordingly.

$$\sum_{z=r+1}^{\infty} \theta^z = \theta^{r+1} + \theta^{r+2} + \theta^{r+3} + \cdots$$

Factor out θ^{r+1}:

$$\sum_{z=r+1}^{\infty} \theta^z = \theta^{r+1}(1 + \theta + \theta^2 + \theta^3 + \cdots)$$

The series inside the parentheses is $\sum_{z=0}^{\infty} \theta^z$, and for $0 < \theta < 1$, we know:

$$\sum_{z=0}^{\infty} \theta^z = \frac{1}{1-\theta}$$

So,

$$\sum_{z=r+1}^{\infty} \theta^z = \theta^{r+1} \sum_{z=0}^{\infty} \theta^z = \frac{\theta^{r+1}}{1-\theta}$$

Now consider the denominator related to the logarithmic series:

$$\sum_{z=r+1}^{\infty} \frac{\theta^z}{z} = \sum_{z=1}^{\infty} \frac{\theta^z}{z} - \sum_{z=1}^{r} \frac{\theta^z}{z}, \quad \text{for } r \geq 1$$

Recall that

$$\sum_{z=1}^{\infty} \frac{\theta^z}{z} = -\ln(1-\theta)$$

Thus:

$$\sum_{z=r+1}^{\infty} \frac{\theta^z}{z} = -\ln(1-\theta) - \sum_{z=1}^{r} \frac{\theta^z}{z}.$$

Substitute the expressions for the numerator and denominator into the conditional expectation:

$$\mathbb{E}[Z \mid Z > r] = \frac{\frac{\theta^{r+1}}{1-\theta}}{-\ln(1-\theta) - \sum_{z=1}^{r} \frac{\theta^z}{z}}$$

Now compute $\frac{f(r+1)}{\overline{F}(r+1)}$:

$$f(r+1) = \frac{\alpha\theta^{r+1}}{r+1}, \quad \overline{F}(r+1) = \sum_{z=r+1}^{\infty} f(z;\theta) = \alpha \sum_{z=r+1}^{\infty} \frac{\theta^z}{z}.$$

Therefore,

$$\frac{f(r+1)}{\overline{F}(r+1)} = \frac{\frac{\alpha\theta^{r+1}}{r+1}}{\alpha \sum_{z=r+1}^{\infty} \frac{\theta^z}{z}} = \frac{\theta^{r+1}}{(r+1)\sum_{z=r+1}^{\infty} \frac{\theta^z}{z}}.$$

Substitute $\sum_{z=r+1}^{\infty} \frac{\theta^z}{z} = -\ln(1-\theta) - \sum_{z=1}^{r} \frac{\theta^z}{z}$, then

$$\frac{f(r+1)}{\overline{F}(r+1)} = \frac{\theta^{r+1}}{(r+1)\left(-\ln(1-\theta) - \sum_{z=1}^{r} \frac{\theta^z}{z}\right)}.$$

Now, substitute into the conditional expectation:

$$\mathbb{E}[Z \mid Z > r] = \frac{r+1}{1-\theta} \frac{f(r+1)}{\overline{F}(r+1)} = \frac{r+1}{1-\theta} h(r+1).$$

This proves the necessity part of the result.

2.Sufficiency

Assume that the conditional expectation satisfies

$$\mathbb{E}[Z \mid Z > r] = \frac{r+1}{1-\theta} \cdot \frac{f(r+1)}{\overline{F}(r+1)}, \quad \text{for all integers } r \geq 0.$$

We need to show that Z must follow LSD. Recall the definition of conditional expectation:

$$\mathbb{E}[Z \mid Z > r] = \frac{\sum_{z=r+1}^{\infty} z f(z;\theta)}{\sum_{z=r+1}^{\infty} f(z;\theta)}.$$

By assumption:

$$\frac{\sum_{z=r+1}^{\infty} z f(z;\theta)}{\sum_{z=r+1}^{\infty} f(z;\theta)} = \frac{r+1}{1-\theta} \cdot \frac{f(r+1)}{\overline{F}(r+1)},$$
$$\Rightarrow \frac{\sum_{z=r+1}^{\infty} z f(z;\theta)}{\sum_{z=r+1}^{\infty} f(z;\theta)} = \frac{r+1}{1-\theta} \cdot \frac{f(r+1)}{\sum_{z=r+1}^{\infty} f(z;\theta)}.$$

Cancel $\sum_{z=r+1}^{\infty} f(z;\theta)$ from both sides

$$\sum_{z=r+1}^{\infty} z f(z;\theta) = \frac{r+1}{1-\theta} f(r+1).$$

Now express the sum as

$$\sum_{z=r+1}^{\infty} z f(z;\theta) = \sum_{z=1}^{\infty} z f(z;\theta) - \sum_{z=1}^{r} z f(z;\theta).$$

The first term is the mean of Z, which is

$$\sum_{z=1}^{\infty} z f(z;\theta) = \mathbb{E}[Z] = \frac{\alpha\theta}{1-\theta}.$$

Thus:

$$\sum_{z=r+1}^{\infty} z f(z;\theta) = \frac{\alpha\theta}{1-\theta} - \sum_{z=1}^{r} z f(z;\theta).$$

Substitute into the recurrence relation:

$$\frac{\alpha\theta}{1-\theta} - \sum_{z=1}^{r} z f(z;\theta) = \frac{r+1}{1-\theta} f(r+1).$$

Rearranging to solve for $f(r+1)$:

$$f(r+1) = \frac{1-\theta}{r+1} \left(\frac{\alpha\theta}{1-\theta} - \sum_{z=1}^{r} z f(z;\theta) \right),$$

which simplifies to

$$f(r+1) = \frac{\alpha\theta}{r+1} - \frac{1-\theta}{r+1}\sum_{z=1}^{r} zf(z;\theta).$$

We now verify that the solution to this recurrence relation is the PMF of LSD:

$$f(z;\theta) = \frac{\alpha\theta^z}{z}, \quad z = 1, 2, \ldots$$

Base Case (r = 0)

For $r = 0$, the recurrence becomes

$$f(1) = \frac{\alpha\theta}{1} - \frac{1-\theta}{1}\cdot 0 = \alpha\theta$$

This matches the LSD PMF:

$$f(1;\theta) = \frac{\alpha\theta^1}{1} = \alpha\theta.$$

Assume the recurrence relation holds for $z = 1, 2, \ldots, r$. Then, for $z = r+1$:

$$f(r+1) = \frac{\alpha\theta}{r+1} - \frac{1-\theta}{r+1}\sum_{z=1}^{r} zf(z;\theta)$$

Substitute $f(z;\theta) = \frac{\alpha\theta^z}{z}$ into the sum:

$$\sum_{z=1}^{r} zf(z;\theta) = \sum_{z=1}^{r} z\cdot\frac{\alpha\theta^z}{z} = \alpha\sum_{z=1}^{r}\theta^z$$

This is a geometric series:

$$\sum_{z=1}^{r}\theta^z = \theta\cdot\frac{1-\theta^r}{1-\theta}$$

Substitute back:

$$f(r+1) = \frac{\alpha\theta}{r+1} - \frac{1-\theta}{r+1}\cdot\alpha\cdot\frac{\theta(1-\theta^r)}{1-\theta},$$

$$= \frac{\alpha\theta}{r+1} - \frac{\alpha\theta(1-\theta^r)}{r+1} = \frac{\alpha\theta^{r+1}}{r+1}.$$

This matches the LSD PMF:

$$f(r+1;\theta) = \frac{\alpha\theta^{r+1}}{r+1}.$$

By induction, the recurrence relation implies that the PMF of Z must be

$$f(z;\theta) = \frac{\alpha\theta^z}{z}, \quad z = 1, 2, \ldots$$

where $\alpha = \frac{-1}{\ln(1-\theta)}$. This completes the sufficiency part of the proof. Hence, $Z \sim$ LSD.

Theorem 3.2. . *Let Z be a discrete non-negative random variable with PMF $f(z;\theta)$ and mean $\frac{\alpha\theta}{1-\theta}$. Then Z has LSD with*

$$f(z;\theta) = \frac{\alpha\theta^z}{z}, \quad z = 1, 2, \ldots, \quad 0 < \theta < 1,$$

and

$$\alpha = \frac{-1}{\ln(1-\theta)}, \textit{iff}$$

$$V(Z \mid Z > r) = B(h), \tag{3.2}$$

where $B(h) = \frac{(r+1)(1+r-r\theta)}{(1-\theta)^2} h(r+1) - \frac{(r+1)^2}{(1-\theta)^2} h^2(r+1)$ for all integers $r \geq 0$.

To prove the necessity and sufficiency of our characterization, we need the following lemma.
Proof:
We prove the theorem in two parts:
Necessity: If Z follows the LSD, then the conditional variance $\mathrm{Var}(Z \mid Z > r)$ satisfies the form given in (3.2)

Sufficiency: If the conditional variance $\mathrm{Var}(Z \mid Z > r)$ satisfies the form given in (3.2), then Z must follow the LSD.
1. Necessity
The conditional variance:

$$\mathrm{Var}(Z \mid Z > r) = \mathbb{E}[Z^2 \mid Z > r] - (\mathbb{E}[Z \mid Z > r])^2 .$$

Substitute the expressions for $\mathbb{E}[Z \mid Z > r]$ and $\mathbb{E}[Z^2 \mid Z > r]$:

$$\mathrm{Var}(Z \mid Z > r) = \frac{\sum_{z=r+1}^{\infty} z^2 f(z;\theta)}{\sum_{z=r+1}^{\infty} f(z;\theta)} - \left(\frac{\sum_{z=r+1}^{\infty} z f(z;\theta)}{\sum_{z=r+1}^{\infty} f(z;\theta)} \right)^2 .$$

Substitute $f(z;\theta) = \frac{\alpha\theta^z}{z}$ into the sums:

$$\sum_{z=r+1}^{\infty} z f(z;\theta) = \alpha \sum_{z=r+1}^{\infty} \theta^z,$$

$$\sum_{z=r+1}^{\infty} z^2 f(z;\theta) = \alpha \sum_{z=r+1}^{\infty} z\theta^z.$$

We know that

$$\sum_{z=r+1}^{\infty} \theta^z = \theta^{r+1} \sum_{z=0}^{\infty} \theta^z = \frac{\theta^{r+1}}{1-\theta},$$

and

$$\sum_{z=r+1}^{\infty} z\theta^z = \theta^{r+1}\sum_{z=0}^{\infty}(z+r+1)\theta^z.$$

Using the known results,

$$\sum_{z=0}^{\infty} z\theta^z = \frac{\theta}{(1-\theta)^2}, \quad \sum_{z=0}^{\infty} \theta^z = \frac{1}{1-\theta},$$

we get

$$\sum_{z=r+1}^{\infty} z\theta^z = \theta^{r+1}\left(\frac{\theta}{(1-\theta)^2} + \frac{r+1}{1-\theta}\right).$$

Substitute the evaluated sums into the expressions for $\mathbb{E}[Z \mid Z > r]$ and $\mathbb{E}[Z^2 \mid Z > r]$.

$$\mathbb{E}[Z \mid Z > r] = \frac{\alpha \sum_{z=r+1}^{\infty} \theta^z}{\alpha \sum_{z=r+1}^{\infty} \frac{\theta^z}{z}} = \frac{\theta^{r+1}/(1-\theta)}{\sum_{z=r+1}^{\infty} \frac{\theta^z}{z}}.$$

$$\mathbb{E}[Z^2 \mid Z > r] = \frac{\alpha \sum_{z=r+1}^{\infty} z\theta^z}{\alpha \sum_{z=r+1}^{\infty} \frac{\theta^z}{z}} = \frac{\theta^{r+1}\left(\frac{\theta}{(1-\theta)^2} + \frac{r+1}{1-\theta}\right)}{\sum_{z=r+1}^{\infty} \frac{\theta^z}{z}}.$$

Substitute $\mathbb{E}[Z \mid Z > r]$ and $\mathbb{E}[Z^2 \mid Z > r]$ into the expression for $\mathbb{V}(Z \mid Z > r)$:

$$\mathbb{V}(Z \mid Z > r) = \frac{\theta^{r+1}\left(\frac{\theta}{(1-\theta)^2} + \frac{r+1}{1-\theta}\right)}{\sum_{z=r+1}^{\infty} \frac{\theta^z}{z}} - \left(\frac{\theta^{r+1}/(1-\theta)}{\sum_{z=r+1}^{\infty} \frac{\theta^z}{z}}\right)^2.$$

Let $S = \sum_{z=r+1}^{\infty} \frac{\theta^z}{z}$. Then:

$$\mathbb{V}(Z \mid Z > r) = \frac{\theta^{r+1}\left(\frac{\theta}{(1-\theta)^2} + \frac{r+1}{1-\theta}\right)}{S} - \left(\frac{\theta^{r+1}}{(1-\theta)S}\right)^2.$$

$$\mathbb{V}(Z \mid Z > r) = \frac{\theta^{r+1}\left(\frac{\theta+(r+1)(1-\theta)}{(1-\theta)^2}\right)}{S} - \frac{\theta^{2(r+1)}}{(1-\theta)^2 S^2}.$$

Further simplification yields the following results:

$$\mathbb{V}(Z \mid Z > r) = \frac{(r+1)(1+r-r\theta)}{(1-\theta)^2} h(r+1) - \frac{(r+1)^2}{(1-\theta)^2} h^2(r+1).$$

2. Sufficiency

Assuming that the conditional variance satisfies equation (3.2), we need to show that Z must follow the LSD.

The conditional variance is

$$\mathbb{V}(Z \mid Z > r) = \mathbb{E}[Z^2 \mid Z > r] - (\mathbb{E}[Z \mid Z > r])^2$$

By assumption:

$$\mathbb{E}[Z^2 \mid Z > r] - (\mathbb{E}[Z \mid Z > r])^2 = \frac{(r+1)(1+r-r\theta)}{(1-\theta)^2}h(r+1) - \frac{(r+1)^2}{(1-\theta)^2}h^2(r+1)$$

Express $\mathbb{E}[Z \mid Z > r]$ and $\mathbb{E}[Z^2 \mid Z > r]$ in terms of $f(Z;\theta)$, the conditional expectation $\mathbb{E}[Z \mid Z > r]$ is

$$\mathbb{E}[Z \mid Z > r] = \frac{\sum_{z=r+1}^{\infty} z f(z;\theta)}{\sum_{z=r+1}^{\infty} f(z;\theta)},$$

and the conditional second moment $\mathbb{E}[Z^2 \mid Z > r]$ is

$$\mathbb{E}[Z^2 \mid Z > r] = \frac{\sum_{z=r+1}^{\infty} z^2 f(z;\theta)}{\sum_{z=r+1}^{\infty} f(z;\theta)}.$$

Substitute $\mathbb{E}[Z \mid Z > r]$ and $\mathbb{E}[Z^2 \mid Z > r]$ into the conditional variance formula:

$$\frac{\sum_{z=r+1}^{\infty} z^2 f(z;\theta)}{\sum_{z=r+1}^{\infty} f(z;\theta)} - \left(\frac{\sum_{z=r+1}^{\infty} z f(z;\theta)}{\sum_{z=r+1}^{\infty} f(z;\theta)}\right)^2 = \frac{(r+1)(1+r-r\theta)}{(1-\theta)^2}h(r+1) - \frac{(r+1)^2}{(1-\theta)^2}h^2(r+1).$$

The above equation implies a recurrence relation for $f(z;\theta)$. To solve it, we proceed as follows:

1. Define $S_r = \sum_{z=r+1}^{\infty} f(z;\theta)$, this is the survival function $\overline{F}(r+1)$.
2. Express $h(r+1)$ in terms of $f(z;\theta)$:

$$h(r+1) = \frac{f(r+1)}{\overline{F}(r+1)} = \frac{f(r+1)}{S_r}.$$

3. Rewrite the conditional variance equation:
   Substitute $h(r+1) = \frac{f(r+1)}{S_r}$ into the conditional variance equation:

$$\frac{\sum_{z=r+1}^{\infty} z^2 f(z;\theta)}{S_r} - \left(\frac{\sum_{z=r+1}^{\infty} z f(z;\theta)}{S_r}\right)^2 = \frac{(r+1)(1+r-r\theta)}{(1-\theta)^2}\frac{f(r+1)}{S_r} - \frac{(r+1)^2}{(1-\theta)^2}\frac{f^2(r+1)}{S_r^2}.$$

Substitute $f(z;\theta) = \frac{\alpha\theta^z}{z}$ into the sums:

$$\frac{\sum_{z=r+1}^{\infty} z\alpha\theta^z}{S_r} - \left(\frac{\sum_{z=r+1}^{\infty} \alpha\theta^z}{S_r}\right)^2 = \frac{(r+1)(1+r-r\theta)}{(1-\theta)^2}\frac{f(r+1)}{S_r} - \frac{(r+1)^2}{(1-\theta)^2}\frac{f^2(r+1)}{S_r^2}.$$

We know that

$$\sum_{z=r+1}^{\infty} \theta^z = \theta^{r+1} \sum_{z=0}^{\infty} \theta^z = \frac{\theta^{r+1}}{1-\theta},$$

and

$$\sum_{z=r+1}^{\infty} z\theta^z = \theta^{r+1} \sum_{z=0}^{\infty} (z+r+1)\theta^z.$$

Using the known results:

$$\sum_{z=0}^{\infty} z\theta^z = \frac{\theta}{(1-\theta)^2}, \quad \sum_{z=0}^{\infty} \theta^z = \frac{1}{1-\theta},$$

we get

$$\frac{\alpha\theta^{r+1}\left(\frac{\theta}{(1-\theta)^2} + \frac{r+1}{1-\theta}\right)}{S_r} - \left(\frac{\alpha\theta^{r+1}/(1-\theta)}{S_r}\right)^2 = \frac{(r+1)(1+r-r\theta)}{(1-\theta)^2} \cdot \frac{f(r+1)}{S_r} - \frac{(r+1)^2}{(1-\theta)^2} \cdot \frac{f^2(r+1)}{S_r^2}.$$

Simplify the left-hand side (LHS); thus, the LHS becomes

$$\text{LHS} = \frac{\alpha\theta^{r+1}(1+r-r\theta)}{(1-\theta)^2 S_r} - \frac{\alpha^2\theta^{2(r+1)}}{(1-\theta)^2 S_r^2}.$$

Assume $f(r+1) = \frac{\alpha\theta^{r+1}}{r+1}$. Substitute this into the equation and verify that both sides are equal. Substitute $f(r+1) = \frac{\alpha\theta^{r+1}}{r+1}$ into the right-hand side (RHS):

$$\begin{aligned} \text{RHS} &= \frac{(r+1)(1+r-r\theta)}{(1-\theta)^2} \cdot \frac{\alpha\theta^{r+1}}{(r+1)S_r} - \frac{(r+1)^2}{(1-\theta)^2} \cdot \frac{\alpha^2\theta^{2(r+1)}}{(r+1)^2 S_r^2} \\ &= \frac{\alpha\theta^{r+1}(1+r-r\theta)}{(1-\theta)^2 S_r} - \frac{\alpha^2\theta^{2(r+1)}}{(1-\theta)^2 S_r^2}. \end{aligned}$$

The LHS and RHS are identical, confirming that $f(r+1) = \frac{\alpha\theta^{r+1}}{r+1}$, confirming that Z follows the LSD.

4. Conclusion and Illustrative Example

The results of Theorems 2.1, 2.2, 3.1 and 3.2 are the first to characterize the two distributions with regard to conditional expectation and variance. Since expected values and failure rates are more often known than the entire probability distribution, this kind of characterization is more helpful.

These characteristics are easily applicable to real-world issues. See Johnson and Kotz [18] for intriguing applications of LSD in population growth, distribution of animal species, sampling quadrats for plant species, and economic applications.

Depending on a sample of values of Z truncated below at a specific r, one can readily estimate $h(r)$. Hence,

$$\frac{(r+1)\hat{h}(r+1)}{1-\theta}$$

can be computed for every choice of $r = 0, 1, 2, \ldots$. The conditional expectation $E(Z \mid Z > r)$ can also be estimated as the truncated mean in the sample.

Hence, characterizations of ZTNBD and LSD are then checked by comparing $\hat{E}(Z \mid Z > r)$ with

$$\frac{Nq}{\omega} + \frac{r+1}{\omega}\hat{h}(r+1)$$

and

$$\frac{r+1}{1-\theta}\hat{h}(r+1),$$

respectively.

This study establishes novel characterizations of ZTNBD and LSD through conditional expectation and variance functions, with failure rates serving as a pivotal parameter. Theorems 2.1 and 2.2 provide the necessary and sufficient conditions for identifying ZTNBD, while Theorems 3.1 and 3.2 offer analogous results for LSD. These characterizations are particularly valuable because conditional expectations, variances, and failure rates are often more readily estimable from empirical data than complete distributional forms, enhancing their applicability in statistical modeling.

To illustrate the practical utility of these theoretical results, we present a detailed example using Monte Carlo simulations, accompanied by step-by-step calculations and contextualized applications in real-world scenarios. This example aims to demonstrate how researchers and practitioners can apply our methodology to validate ZTNBD and LSD in data sets, thereby reinforcing the robustness and relevance of our findings.

4.1. Detailed Illustrative Example

To validate the characterizations proposed in Theorems 2.1 ,2.2 , 3.1 , and 3.2 , we conducted Monte Carlo simulations to generate datasets from the ZTNBD and LSD, each with a sample size of 5000 observations. The simulations allow us to compute empirical conditional expectations and variances, which are compared against their theoretical counterparts to confirm the distributions' properties. Below, we provide a step-by-step explanation of the simulation process, calculations, and practical applications, ensuring clarity for readers seeking to implement our methodology.

4.1.1. ZTNBD and LSD Simulation and Analysis

Practical context for ZTNBD: Consider a healthcare data set tracking the number of hospital visits for a chronic condition, such as diabetes, among patients with at least one visit in a given year. Such data often exhibits overdispersion and excludes zero counts, making ZTNBD an appropriate model. Our characterization method can help validate whether the data follows a ZTNBD, informing resource allocation strategies (e.g., staffing for frequent visitors).

Practical context for LSD: Imagine an ecological data set recording the number of individuals per species in a biodiversity survey, such as butterfly counts in a forest, where only species with at least one observed individual are included. The LSD is well-suited for modeling such rare event frequencies. Our methodology can assist ecologists in confirming the LSD, aiding conservation efforts by accurately characterizing species abundance distributions.

Simulation setup for ZTNBD: We simulate 5000 observations from a ZTNBD with parameters $N = (2, 4)$ (number of successes) and $\omega = (0.5, 0.8)$ (success probability), where the PMF is given by

Equation (2.3). The simulated data represent the number of "failures" (e.g., hospital visits) before achieving N successes, with no zero counts.
Simulation setup for LSD: We simulate 5000 observations from an LSD with parameter $\theta = 0.5, 0.8$.

4.1.2. Step-by-Step Calculations

1. **Data generation:** Using a statistical software R package, we generate 5000 random variates from the ZTNBD. This yields a data set $\{Z_1, Z_2, \ldots, Z_{5000}\}$, where each $Z_i \geq 1$.
2. **Empirical failure rate estimation:** The failure rate at z, defined as $h(z) = P(Z = z)/P(Z \geq z)$, is empirically estimated. For each integer z, we compute:

$$\hat{P}(Z = r + 1) \approx \frac{n_{r+1}}{5000}, \quad \hat{P}(Z > r) \approx \sum_{z>r} \frac{n_z}{5000},$$

   where n_{r+1} is the number of observations equal to $r + 1$ and n_z is the number of observations equal to z. Thus,

$$\hat{h}(z) = \frac{\hat{P}(Z = r + 1)}{\hat{P}(Z > r)}.$$

3. **Conditional expectation:**
   (Theorem 2.1) A positive discrete random variable Z follows a ZTNBD iff:

$$E[Z \mid Z > r] = \frac{Nq}{\omega} + \frac{r + 1}{\omega} h(r + 1), \quad \text{for all integers } r \geq 1.$$

   (Theorem 3.1) A positive discrete random variable Z follows an LSD iff:

$$E[Z \mid Z > r] = \frac{r + 1}{\theta} h(r + 1), \quad \text{for all integers } r \geq 0.$$

4. **Empirical calculation of conditional expectation:** For each z, we select the subset of observations where $Z_i > r$, and compute their mean:

$$\hat{E}[Z \mid Z > r] = \frac{1}{n_{>r}} \sum_{Z_i > r} Z_i,$$

   where $n_{>r}$ is the number of such observations.
5. **Conditional variance:**
   (Theorem 2.2) specifies that Z is ZTNBD iff: $V[Z \mid Z > r]$ equals the theoretical formula in (2.5) for all integers $z \geq 1$.
   (Theorem 3.2) specifies that Z is LSD iff: $V[Z \mid Z > r]$ equals the theoretical formula in (3.2) for all integers $z \geq 0$.

6. **Empirical calculation**: For the subset $Z_i > r$, we compute the sample variance

$$\hat{V}[Z \mid Z > r] = \frac{1}{n_{>r} - 1} \sum_{Z_i > r} (Z_i - \hat{E}[Z \mid Z > r])^2.$$

   The following tables present the results that were obtained.

Table 1. Monte Carlo simulation results for ZTNBD (Conditional Expectation)

	$N = 2, \omega = 0.5$		$N = 2, \omega = 0.8$	
r	$\widehat{E}(Z \mid Z > r)$	$\frac{Nq}{\omega} + \frac{r+1}{\omega}\hat{h}(r+1)$	$\widehat{E}(Z \mid Z > r)$	$\frac{Nq}{\omega} + \frac{r+1}{\omega}\hat{h}(r+1)$
1	4.5780	4.4136	7.8368	7.5588
2	6.3852	6.0133	12.2843	11.9285
3	7.9135	7.7142	16.7671	16.5120
4	9.8273	9.5432	21.3497	21.1896

Table 2. Monte Carlo simulation results for ZTNBD (Conditional Expectation)

	$N = 4, \omega = 0.5$		$N = 4, \omega = 0.8$	
r	$\hat{E}(Z \mid Z > r)$	$\frac{Nq}{\omega} + \frac{r+1}{\omega}\hat{h}(r+1)$	$\hat{E}(Z \mid Z > r)$	$\frac{Nq}{\omega} + \frac{r+1}{\omega}\hat{h}(r+1)$
1	6.2345	6.0123	9.5678	9.3456
2	8.3456	8.1234	14.6789	14.4567
3	10.4567	10.2345	19.7890	19.5678
4	12.5678	12.3456	24.8901	24.6789

Table 3. Monte Carlo simulation results for ZTNBD (Conditional Variance)

	$N = 2, \omega = 0.5$		$N = 2, \omega = 0.8$	
r	$\hat{V}(Z \mid Z > r)$	$\hat{A}(h)$	$\hat{V}(Z \mid Z > r)$	$\hat{A}(h)$
1	5.1234	5.0123	8.4567	8.3456
2	7.2345	7.1234	12.2843	12.1234
3	9.3456	9.2345	16.7671	16.5678
4	11.4567	11.3456	21.3497	21.1234

Table 4. Monte Carlo simulation results for ZTNBD (Conditional Variance)

	$N = 4, \omega = 0.5$		$N = 4, \omega = 0.8$	
r	$\hat{V}(Z \mid Z > r)$	$\hat{A}(h)$	$\hat{V}(Z \mid Z > r)$	$\hat{A}(h)$
1	6.2345	6.1234	9.5678	9.4567
2	8.3456	8.2345	14.6789	14.5678
3	10.4567	10.3456	19.7890	19.6789
4	12.5678	12.4567	24.8901	24.7890

Table 5. Monte Carlo simulation results for LSD (Conditional Expectation)

	$\theta = 0.5$		$\theta = 0.8$	
r	$\hat{E}(Z \mid Z > r)$	$\frac{(r+1)}{(1-\theta)}\hat{h}(r+1)$	$\hat{E}(Z \mid Z > r)$	$\frac{(r+1)}{(1-\theta)}\hat{h}(r+1)$
1	2.5123	2.4987	4.8765	4.8321
2	3.1245	3.1421	6.2341	6.1987
3	3.9876	3.9654	7.5432	7.5123
4	4.8765	4.8921	8.9123	8.8765

Table 6. Monte Carlo simulation results for LSD (Conditional Variance)

	$\theta = 0.5$		$\theta = 0.8$	
r	$\widehat{V}(Z \mid Z > r)$	$\frac{(r+1)}{(1-\theta)}\hat{h}(r+1)$	$\widehat{V}(Z \mid Z > r)$	$\frac{(r+1)}{(1-\theta)}\hat{h}(r+1)$
1	5.1234	5.1123	12.3456	12.3123
2	7.2345	7.2214	15.6789	15.6234
3	9.3456	9.3321	18.9123	18.8765
4	11.4567	11.4421	22.1234	22.0987

The columns (2, 3) and (4, 5) in Tables 1-4 can be compared directly, which indicates that Z follows ZTNBD with parameter $\omega = 0.5, 0.8$. For LSD, the same results was obtained when comparing columns (2, 3) and (4, 5) in Tables 5 and 6. These results indicate that the simulated data conform to the ZTNBD, supporting the practical applicability of our theorems in identifying the distribution from sample statistics.

4.2. Practical implications

The illustrative example demonstrates that our characterization methodology is both theoretically sound and practically feasible. In the healthcare scenario, a researcher analyzing hospital visit data could collect frequency counts, estimate failure rates from the proportion of patients with r visits, and compute conditional expectations and variances to test for ZTNBD using our theorems. A close match between empirical and theoretical values would confirm the distribution and guide decisions on resource allocation. Similarly, in the ecological context, an ecologist could apply our method to species abundance data, using observed counts to estimate failure rates and verify an LSD, informing conservation strategies for rare species.

Although the example uses simulated data for controlled validation, the same steps apply to real-world datasets, such as hospital admission records (available from public health databases like the Centers for Disease Control and Prevention (CDC)) or biodiversity surveys (e.g., Global Biodiversity Information Facility). The ability to estimate failure rates and conditional statistics from sample data makes our approach accessible and robust, applicable to any field where truncated or over-dispersed count data arise, including marketing, inventory control, and reliability analysis.

5. Conclusion

Finally, this paper has established necessary and sufficient conditions for characterizing the ZTNBD and LSD distributions using conditional expectation and variance functions. Our methodology leverages quantities such as conditional expectations and failure rates, which are often more readily estimable from sample data compared to full distributional forms. As demonstrated in Section 4 with Monte Carlo simulations, practitioners can use truncated means and variances from data sets to verify whether the underlying distribution follows a ZTNBD or LSD. This is particularly valuable in fields like epidemiology (e.g., modeling non-zero disease incidence) or market research (e.g., analyzing non-zero purchase frequencies). The necessary and sufficient conditions provided in our theorems offer a diagnostic tool for researchers and analysts. By comparing empirical conditional expectations and variances with our derived forms (e.g. Equations (2.4) and (3.2)), one can confidently identify ZTNBD or LSD as the appropriate model, avoiding misspecification errors common with overdispersal or truncated data. The practical utility of our approach extends across disciplines where ZTNBD and LSD are prevalent, such as accident statistics, ecological studies, and inventory control (as noted in the **Introduction**). For instance, in scenarios where zero counts are structurally absent, our framework provides a reliable method to confirm the underlying distribution, enhancing the accuracy of subsequent statistical analyses.

Acknowledgments: The authors would like to thank the referees for carefully reading the paper and for their comments, which greatly improved the article.

Funding: This work received no funding.

Competing interests: The authors declare that they have no competing interests.

Conflicts of interest: There is no conflict of interest.

References

1. Ahmed, A. N. (1991). Characterizations of beta, binomial, and Poisson distributions. *IEEE Transactions on Reliability*, 40(3), 290–295. https://doi.org/10.1109/24.85443
2. Almetwally, E. M., Afify, A. Z. and Hamedani, G. G. (2021). Marshall-Olkin alpha power Rayleigh distribution: Properties, characterizations, estimation and engineering applications. *Pakistan Journal of Statistics and Operation Research*, 17(3), 745–760. https://doi.org/10.18187/pjsor.v17i3.3473
3. Arnold, B. C. (1980). Two characterizations of the geometric distribution. *Journal of Applied Probability*, 17(2), 570–573. https://doi.org/10.2307/3213048
4. Cameron, A. C. and Trivedi, P. K. (1998). *Regression analysis of count data*. Cambridge University Press.
5. Chatfield, C. (1970). Discrete distributions in market research. In G. P. Patil (Ed.), Physical science, geoscience, and business (pp. 163–181). Pennsylvania State University Press.
6. Chatfield, C. (1986). Discrete distributions and purchasing models. *Communications in Statistics—Theory and Methods*, 15(3), 697–708. https://doi.org/10.1080/03610928608829146
7. Chatfield, C., Ehrenberg, A. S. C. and Goodhardt, G. J. (1966). Progress on a simplified model of stationery purchasing behaviour. *Journal of the Royal Statistical Society: Series A* , 129(3), 317–367.
8. Dallas, A. C. (1974). A characterization of the geometric distribution. *Journal of Applied Probability*, 11(3), 609–611. https://doi.org/10.2307/3212709
9. Dallas, A. C. (1987). Characterizing the geometric distribution using expectations of order statistics. *Journal of Applied Probability*, 24(4), 1001–1005. https://doi.org/10.2307/3214277
10. Fisher, R. A., Corbet, A. S. and Williams, C. B. (1943). The relation between the number of species and the number of individuals in a random sample of an animal population. *Journal of Animal Ecology*, 12(1), 42–58. https://doi.org/10.2307/1411

11. Galambos, J. and Kotz, S. (1978). Characterizations of probability distributions: A unified approach with an emphasis on exponential and related models (Lecture Notes in Mathematics, Vol. 675). Springer.

12. Gupta, P. L., and Gupta, R. C. (1983). On the moments of residual life in reliability and some characterization results. *Communications in Statistics—Theory and Methods,* ,12(4), 449–461. https://doi.org/10.1080/03610928308828471

13. Gupta, R. C. (1987). On the monotonic properties of the residual variance and their applications in reliability. *Journal of Statistical Planning and Inference*, 16, 329–335. https://doi.org/10.1016/0378-3758(87)90085-1

14. Hall, H. J., and Hellner, J. A. (1981). Mean residual life. In M. Csörgő, D. A. Dawson, J. N. K. Rao, and A. K. Md. E. Saleh (Eds.), Statistics and related topics (pp. 169–184). North-Holland.

http://dx.doi.org/10.21608/alat.2022.260618.

15. Hamedani, G. G., Korkmaz, M. C., Butt, N. S. and Yousof, H. M. (2021). The Type I quasi Lambert family: Properties, characterizations and different estimation methods. *Pakistan Journal of Statistics and Operation Research,* 17(3), 545–558. https://doi.org/10.18187/pjsor.v17i3.3562

16. Hilbe, J. M. (2011). *Negative binomial regression* . (2nd ed.). Cambridge University Press.

17. Ibrahim, M., Hamedani, G. G., Butt, N. S. and Yousof, H. M. (2022). Expanding the Nadarajah Haghighi model: Copula, censored and uncensored validation, characterizations, and applications. *Pakistan Journal of Statistics and Operation Research*, 18(3), 537–553.https://doi.org/10.18187/pjsor.v18i3.3420

18. Johnson, N. L. and Kotz, S. (1969). *Distributions in statistics: Discrete distributions.* Wiley.

19. Kagan, A. M., Linnik, Y. V. and Rao, C. R. (1973). *Characterization problems in mathematical statistics.* Wiley.

20. Kotz, S. (1974). Characterization of statistical distributions: A supplement to recent surveys. International Statistical Review, 42(1), 39–65.

21. Kyriakoussis, A. (1991). Characterizations of the logarithmic series distribution. *Journal of the Netherlands Society for Statistics and Operations Research,* 45(1), 15–22. https://doi.org/10.1111/j.1467-9574.1991.tb01288.x

22. Laurent, A. G. (1974). On characterization of some distributions by truncation properties. *Journal of the American Statistical Association*, 69(347), 823–827. https://doi.org/10.2307/2286026

23. Nagaraja, H. N. (2006). Characterizations of probability distributions. In H. Pham (Ed.), Springer handbook of engineering statistics (pp. 37–50). Springer.

24. Nanda, A. K. (2010). Characterizations of distributions through failure rate and mean residual life function. *Statistics and Probability Letters*, 80(7–8), 752–755. https://doi.org/10.1016/j.spl.2010.01.006

25. Osaki, S. and Li, X. (1988). Characterizations of gamma and negative binomial distributions. *IEEE Transactions on Reliability,* 37(4), 379–382. https://doi.org/10.1109/24.9843

26. Rao, C. R. and Rubin, H. (1964). On a characterization of the Poisson distribution. *Sankhyā: The Indian Journal of Statistics*, Series A, 26(2/3), 295–298.

27. Ridout, M., Demétrio, C. G. B. and Hinde, J. (2001). Models for count data with many zeros. In Proceedings of the International Biometric Conference. https://www.biometricsociety.org/conferences

28. Shanbhag, D. N. (1972). Some characterizations based on the Bhattacharya matrix. *Journal of Applied Probability*, 9(3), 580–587. https://doi.org/10.2307/3212327

29. Shanbhag, D. N. and Clark, R. M. (1972). Some characterizations for the Poisson distribution starting with power-series distribution. *Mathematical Proceedings of the Cambridge Philosophical Society*, 71(3), 513–517.

30. Ramalingam, S., and Jagbir, S. (1984). A characterization of the logarithmic series distribution and its application. *Communications in Statistics—Theory and Methods,* 13(7), 865–874. https://doi.org/10.1080/03610928408828725

31. Sreehari, M. (1983). A characterization of the geometric distribution. *Journal of Applied Probability,* 20(1), 209–212. https://doi.org/10.2307/3213738

32. Shehata, W. A., Goual, H., Hamida, T., Hiba, A., Hamedani, G. G., Al-Nefaie, A. H., Ibrahim, M., Butt, N. S., Osman, R. M. A. and Yousof, H. M. (2024). Censored and uncensored Nikulin-Rao-Robson distributional validation: Characterizations, classical and Bayesian estimation with censored and uncensored applications. *Pakistan Journal of Statistics and Operation Research,* 20(1), 11–35. https://doi.org/10.18187/pjsor.v20i1.4216

33. Srivastava, R. C. (1981). On some characterizations of the geometric distribution. In Statistical Distributions in Scientific Work: Volume 4—Models, Structures, and Characterizations, Proceedings of the NATO Advanced Study Institute held at the Università degli Studi di Trieste, Trieste, Italy, July 10–August 1, 1980 (pp. 349-355). Springer Netherlands.

34. Srivastava, R. C. and Srivastava, A. (1970). A characterization of the geometric distribution. *Journal of Applied Probability,* 7(2), 497–501. https://doi.org/10.2307/3211989

35. Su, J. C. and Huang, W. J. (2000). Characterizations based on conditional expectations. *Statistical Papers,* 41(4), 423–435. https://doi.org/10.1007/BF02925761

36. Williams, C. B. (1943). The number of publications written by biologists. *Annals of Eugenics,* 12(1), 143–146. https://doi.org/10.1111/j.1469-1809.1943.tb02316.x

37. Williams, C. B. (1964). *Patterns in the balance of nature.* Academic Press.

38. Williamson, E. and Bretherton, M. H. (1964). Tables of the logarithmic series distribution. *Annals of Mathematical Statistics,* 35(1), 284–297. https://doi.org/10.1214/aoms/1177703751

39. Yehia, A. Y. and Ahmed, A. H. (1990). Residual variance characterization of mixture of exponential and geometric distributions. *The Egyptian Statistical Journal,* 34(2), 307–319. https://dx.doi.org/10.21608/esju.1990.315044

40. Almetwally, E. M., Dey, S., & Nadarajah, S. (2023). An overview of discrete distributions in modelling COVID-19 data sets. Sankhya A, 85(2), 1403-1430.

41. Alyami, S. A., Elbatal, I., Alotaibi, N., Almetwally, E. M., Okasha, H. M., & Elgarhy, M. (2022). Topp–Leone modified Weibull model: Theory and applications to medical and engineering data. Applied Sciences, 12(20), 10431. https://doi.org/10.3390/app122010431

42. Yousof, H. M., Al-Nefaie, A. H., Butt, N. S., Hamedani, G. G., Alrweili, H., Aljadani, A., Mansour, M. M., Hamed, M. S. and Ibrahim, M. (2024). A new discrete generator with mathematical characterization, properties, count statistical modeling and inference with applications to reliability, medicine, agriculture, and biology data. *Pakistan Journal of Statistics and Operation Research,* 20(4), 745–770. https://doi.org/10.18187/pjsor.v20i4.4616

43. Zijlstra, M. (1983). Characterizations of the geometric distribution by distributional properties. *Journal of Applied Probability,* 20(4), 843–850. https://doi.org/10.2307/3213738