Article

A Statistical Model for Count Data Analysis and Population Size Estimation: Introducing a Mixed Poisson–Lindley Distribution and Its Zero Truncation

Gadir Alomair [1,*], Razik Ridzuan Mohd Tajuddin [2], Hassan S. Bakouch [3,4] and Amal Almohisen [5]

1   Department of Quantitative Methods, School of Business, King Faisal University,
    Al Hofuf 31982, Saudi Arabia
2   Department of Mathematical Sciences, Universiti Kebangsaan Malaysia, Bangi 43600, Malaysia;
    rrmt@ukm.edu.my
3   Department of Mathematics, College of Science, Qassim University, Buraydah 51452, Saudi Arabia;
    h.bakouch@qu.edu.sa or hassan.bakouch@science.tanta.edu.eg
4   Department of Mathematics, Faculty of Science, Tanta University, Tanta 31111, Egypt
5   Department of Statistics and Operations Research, College of Sciences, King Saud University,
    Riyadh 11495, Saudi Arabia; amalmoh@ksu.edu.sa
*   Correspondence: galomair@kfu.edu.sa

Abstract: Count data consists of both observed and unobserved events. The analysis of count data often encounters overdispersion, where traditional Poisson models may not be adequate. In this paper, we introduce a tractable one-parameter mixed Poisson distribution, which combines the Poisson distribution with the improved second-degree Lindley distribution. This distribution, called the Poisson-improved second-degree Lindley distribution, is capable of effectively modeling standard count data with overdispersion. However, if the frequency of the unobserved events is unknown, the proposed distribution cannot be directly used to describe the events. To address this limitation, we propose a modification by truncating the distribution to zero. This results in a tractable zero-truncated distribution that encompasses all types of dispersions. Due to the unknown frequency of unobserved events, the population size as a whole becomes unknown and requires estimation. To estimate the population size, we develop a Horvitz–Thompson-like estimator utilizing truncated distribution. Both the untruncated and truncated distributions exhibit desirable statistical properties. The estimators for both distributions, as well as the population size, are asymptotically unbiased and consistent. The current study demonstrates that both the truncated and untruncated distributions adequately explain the considered medical datasets, which are the number of dicentric chromosomes after being exposed to different doses of radiation and the number of positive Salmonella. Moreover, the proposed population size estimator yields reliable estimates.

1. Introduction

Unobserved events in count data are events that were not recorded. For example, the unobserved events in insurance claims refer to the events where the policyholders do not claim. In some cases, even these events are not identified. For example, the number of times a motorist gets stopped by the police. The number of motorists that did not get stopped cannot be identified, which is known as unobserved events. Modeling count data by examining the observed events or both the observed and unobserved events is typical in statistical modeling. This study would like to explore this area by introducing a mixed

Poisson–Lindley distribution and extending it to its zero-truncated version, as well as population size estimation.

The Lindley distribution was originally introduced in Bayesian analysis as a mixture of an exponential distribution and a gamma distribution [1]. Ghitany et al. [2] conducted a comprehensive examination of its statistical characteristics. Subsequent research has expanded the Lindley distribution into both two-parameter [3–11] and three-parameter variants [12–14], each offering innovative enhancements and broader applications. These distributions have been particularly useful in the fields of survival analysis and reliability assessments.

Sankaran [15] pioneered the use of the Lindley distribution as a mixing distribution with the Poisson distribution, thus creating the Poisson–Lindley (PL) distribution. This model, and the methods to estimate its parameters, were studied extensively in later research [16]. Various mixed PL distributions have been proposed as alternatives to the traditional Poisson and negative binomial models for fitting count data [17–19]. These new distributions share the overdispersion characteristic common to mixed Poisson distributions [20]. Other notable mixed Poisson models include the Poisson Inverse Pareto distribution [21] and the Poisson-transmuted record-type exponential distribution [22], among others.

When count data exclusively contain positive numbers, zero truncation is a common technique used to adjust distributions accordingly. Examples include the zero-truncated Poisson [23], zero-truncated negative binomial [24], and zero-truncated PL [25] distributions. In fields such as criminology, accurately estimating the size of a population when the frequency of non-events is not observable is a significant challenge [26–33]. Rossmo and Routledge [30] highlighted the necessity of understanding the size of a criminal population to inform the creation of effective laws and policies. Such estimations are still relatively rare in medical research but are equally needed.

The current study aims to introduce a new, practical mixed PL distribution by combining an improved second-degree Lindley (ISDL) distribution [7] with a Poisson variable, resulting in what we name the Poisson-improved second-degree Lindley (PISDL) distribution. This constitutes the first aim of this research. Given the ISDL distribution's superior modeling performance [7], we anticipate that the PISDL distribution may outperform the original PL distribution. To accommodate strictly positive data, we also introduce a zero-truncated version of the PISDL (ZTPISDL) distribution, which represents the second aim of this research. Furthermore, we propose an innovative estimator for population size in relation to the ZTPISDL distribution, detailed in Section 5.3, marking the third and most original objective of the paper. Understanding the population size is critical for developing comprehensive policies. This study uses data from epidemiology and cytogenetics to demonstrate the applications of the proposed distributions and estimator.

This paper is structured as follows. Section 1 outlines the study's objectives, building upon previous research on mixed PL distributions, their truncations, and population size estimations. Section 2 presents the PISDL distribution, its statistical properties, and estimation techniques. Section 3 introduces and investigates the zero-truncated PISDL distribution. Section 4 describes the development of a new population size estimator, assuming a PISDL distribution, accompanied by simulation studies in Sections 2–4 to evaluate estimator performance. Section 5 applies the proposed models to medical dataset analyses. Finally, Section 6 concludes with a discussion of the implications, limitations, and future research directions of this study.

2. Poisson-Improved Second-Degree Lindley Distribution

2.1. Probability Mass Function of the PISDL Distribution

Definition 1. *A random variable X is said to follow a Poisson-improved second-degree Lindley (PISDL) distribution with parameter θ if it obeys $X|\theta \sim Poisson(\theta)$ and $\theta \sim ISDL(\lambda)$, where θ, $\lambda > 0$.*

Theorem 1. *Let X be a random variable that follows PISDL with parameter λ; then, the probability mass function (pmf) of X is given as*

$$f(x) = \frac{\lambda^3}{\lambda^2 + 2\lambda + 2} \left[\frac{x(x+5) + 2\lambda(x+2) + \lambda^2 + 5}{(\lambda+1)^{x+3}} \right]; \ x = 0, 1, 2, \ldots, \ \lambda > 0. \quad (1)$$

Proof. Using Definition 1, the pmf of $X|\theta$ is

$$f(x|\theta) = \frac{\theta^x \exp(-\theta)}{x!}; \ x = 0, 1, 2, \ldots, \ \theta > 0.$$

The probability density function of $\theta|\lambda$ is

$$g(\theta) = \frac{\lambda^3}{\lambda^2 + 2\lambda + 2} (1+\theta)^2 \exp(-\lambda\theta); \ \lambda > 0, \theta > 0.$$

The resulting marginal distribution $f(x)$ for the PISDL distribution with parameter λ is obtained as follows:

$$
\begin{aligned}
f(x) &= \int_0^\infty f(x|\theta) g(\theta) d\theta \\
&= \frac{\lambda^3}{(\lambda^2 + 2\lambda + 2)x!} \int_0^\infty \theta^x (1+\theta)^2 \exp[-\theta(\lambda+1)] d\theta \\
&= \frac{\lambda^3}{\lambda^2 + 2\lambda + 2} \left[\frac{x(x+5) + 2\lambda(x+2) + \lambda^2 + 5}{(\lambda+1)^{x+3}} \right].
\end{aligned}
$$

□

Figure 1 shows the pmf plot for the PISDL distribution. In Figure 1, the distribution is skewed to the right, unimodal, and decreasing, which is further supported by the decreasing ratio of probability given as

$$\frac{f(x+1)}{f(x)} = \frac{1}{\lambda+1} \left[1 + \frac{2(x+\lambda+3)}{x(x+5) + 2\lambda(x+2) + \lambda^2 + 5} \right].$$

It is worth noting that the PISDL distribution is actually a three-component mixture distribution that can be written as $f(x) = p_1 f_1(x) + p_2 f_2(x) + p_3 f_3(x)$, where $f_i(x)$ is the pmf of the negative binomial distribution with a parameter number of successes i and proportion $\lambda/(\lambda+1)$. When $i = 1$, $f_i(x)$ is the pmf of the geometric distribution, which is a special case for the negative binomial distribution. The formulae for p_i and $f_i(x)$ for $i = 1, 2, 3$ are given as

$$p_1 = \frac{\lambda^2}{\lambda^2 + 2\lambda + 2}; \ p_2 = \frac{2\lambda}{\lambda^2 + 2\lambda + 2}; \ p_3 = 1 - p_1 - p_2;$$

$$f_1(x) = \frac{\lambda}{(\lambda+1)^{x+1}}; \ f_2(x) = \frac{\lambda^2(x+1)}{(\lambda+1)^{x+2}}; \ f_3(x) = \frac{\lambda^3(x+1)(x+2)}{2(\lambda+1)^{x+3}}.$$

Even though the PISDL distribution is a three-component mixture negative binomial distribution, the existence of three modes cannot be seen in any of the plots in Figure 1 for the selected values of λ. This insinuates that the three modes, which come from the three sub-populations, must be located very close to each other. As mentioned in [34], if the modes of the sub-populations are located very close to each other, then the population will have a single mode. As such, if the existence of the modes of the sub-populations, each with very close mode values, can be certain, then this distribution can be considered as one of the candidates for model fittings.
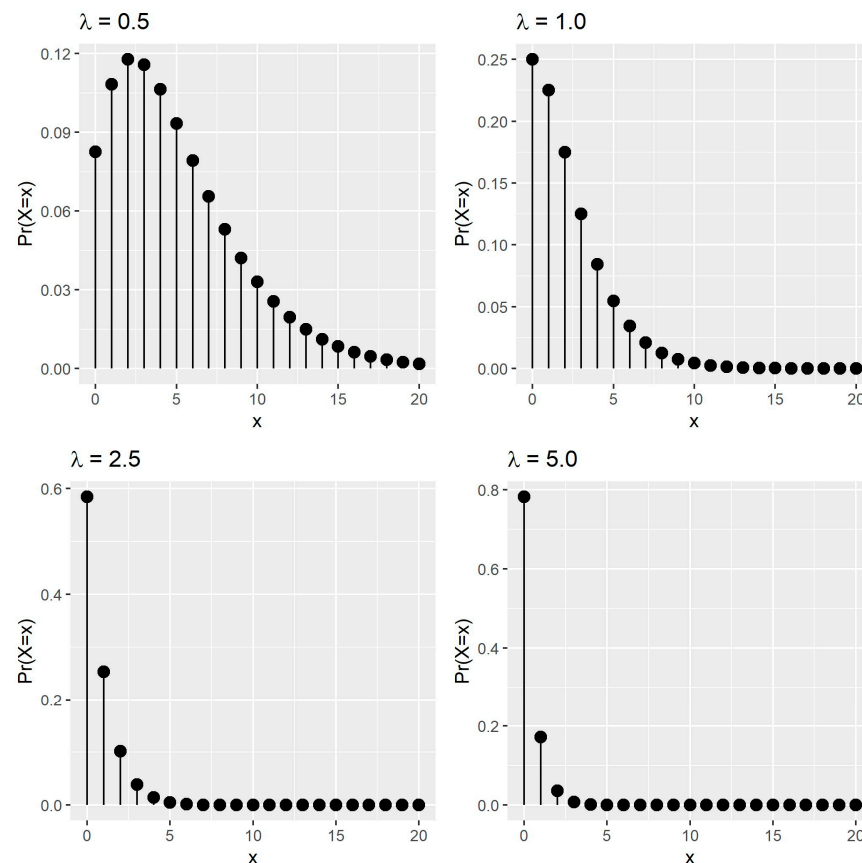
Figure 1. The pmf plot for the PISDL distribution when $\lambda = 0.5,\ 1.0,\ 2.5,\ 5.0$.

The cumulative distribution function (cdf) for the PISDL distribution is given in Equation (2) and visualized in Figure 2. In Figure 2, it is clear that the PISDL distribution has a valid cdf since $F(x) \to 1$ as $x \to \infty$.

$$F(x) = 1 - \frac{\lambda^4 + 2\lambda^3(x+3) + \lambda^2(x^2 + 7x + 13) + 2\lambda(x+4) + 2}{(\lambda^2 + 2\lambda + 2)(\lambda+1)^{x+3}}. \tag{2}$$

Based on Equation (2), the survival function $S(x)$ can be obtained and given as

$$S(x) = \frac{\lambda^4 + 2\lambda^3(x+3) + \lambda^2(x^2 + 7x + 13) + 2\lambda(x+4) + 2}{(\lambda^2 + 2\lambda + 2)(\lambda+1)^{x+3}}. \tag{3}$$

The plot for the survival function of the PISDL distribution is given in Figure 3.

The hazard rate function $h(x)$ is defined by taking the ratio of the pmf to the survival function, i.e., $h(x) = f(x)/S(x)$, and is given as

$$h(x) = \frac{\lambda^3\left[x(x+5) + 2\lambda(x+2) + \lambda^2 + 5\right]}{\lambda^4 + 2\lambda^3(x+3) + \lambda^2(x^2 + 7x + 13) + 2\lambda(x+4) + 2}. \tag{4}$$

The hazard rate function plot is given in Figure 4. In Figure 4, it can be noted that the hazard rate functions show an increasing pattern with a limiting value of λ, meaning $\lim_{x \to \infty} h(x) = \lambda$.
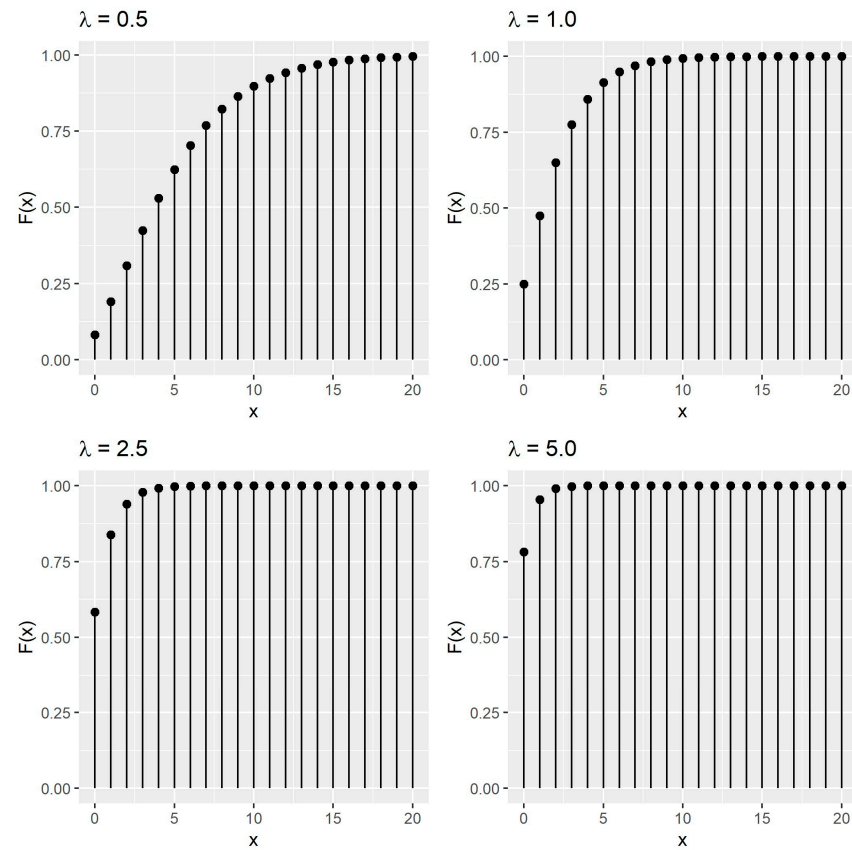
Figure 2. The cdf plot for the PISDL distribution when $\lambda = 0.5$, 1.0, 2.5, 5.0.
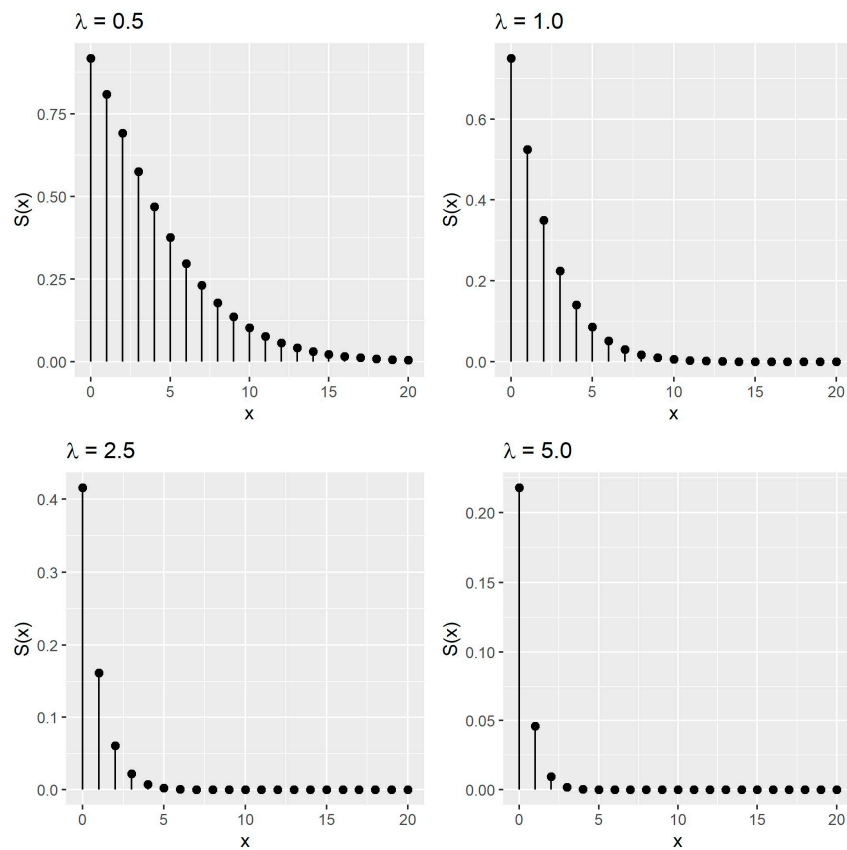
Figure 3. The survival function plot for the PISDL distribution when $\lambda = 0.5$, 1.0, 2.5, 5.0.

Figure 4. The hazard rate plot for the PISDL distribution when $\lambda = 0.5$, 1.0, 2.5, 5.0.
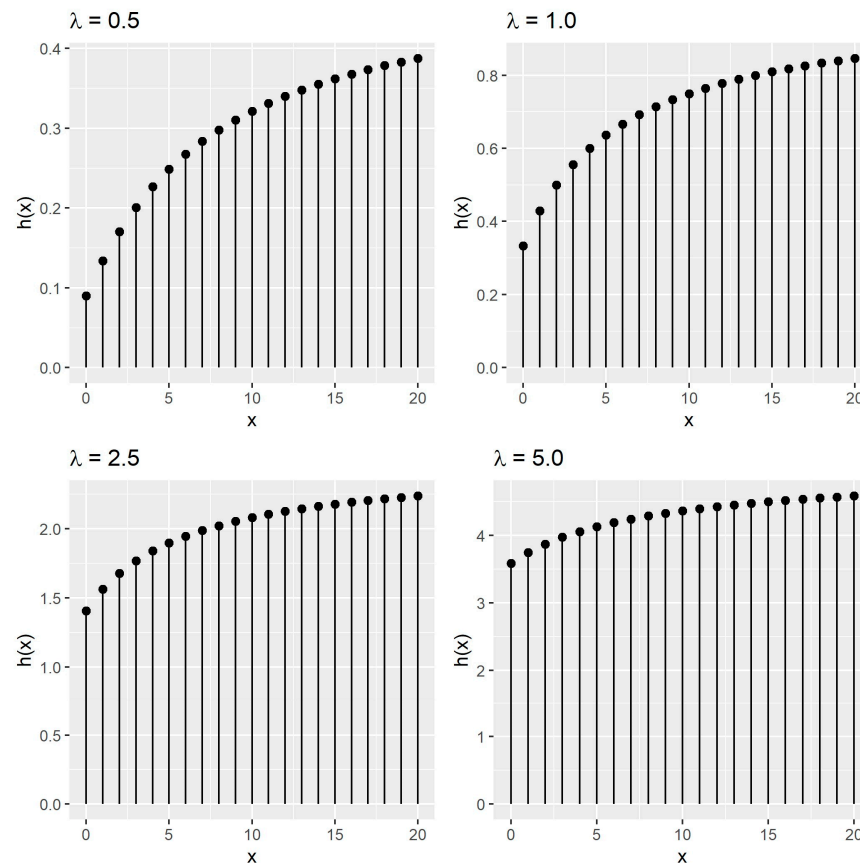
2.2. Some Statistical Properties of the PISDL Distribution

The kth moment in the origin of the PISDL distribution can be written by the following generic expression:

$$E\left(X^k\right) = \sum_{x=0}^{\infty} x^k f(x). \tag{5}$$

In particular, the first two moments in the origin using Equation (5) are obtained, respectively, as

$$E(X) = \frac{\lambda^2 + 4\lambda + 6}{\lambda(\lambda^2 + 2\lambda + 2)}, \ E\left(X^2\right) = \frac{(\lambda + 2)\left(\lambda^2 + 4\lambda + 6\right) + 4(\lambda + 3)}{\lambda^2(\lambda^2 + 2\lambda + 2)}.$$

Hence, the index of dispersion (IOD_X) can be written as

$$\begin{aligned} IOD_X \ &= \frac{Var(X)}{E(X)} = \frac{E\left(X^2\right) - [E(X)]^2}{E(X)} = \frac{E\left(X^2\right)}{E(X)} - E(X) \\ &= 1 + \tfrac{1}{\lambda}\left(\frac{\lambda^4 + 2\left(4\lambda^3 + 12\lambda^2 + 12\lambda + 6\right)}{\lambda^4 + 6\lambda^3 + 16\lambda^2 + 20\lambda + 12}\right). \end{aligned} \tag{6}$$

Since the $IOD_X > 1$, the PISDL distribution is overdispersed for all λ.

The mode of the PISDL distribution can be obtained by maximizing the log pmf of the PISDL distribution or equivalently by solving the quadratic equation $Ax_m{}^2 + Bx_m + C = 0$,

where $x_m \geq 0$ is the mode of the distribution, where $A = \ln(\lambda + 1)$, $B = (2\lambda + 5)\ln(\lambda + 1) - 2$, and $C = (5 + \lambda(\lambda + 4))\ln(\lambda + 1) - 2\lambda - 5$. As a result, the solution for the equation is

$$x_m = \frac{2 - (5 + 2\lambda)\ln(1 + \lambda) + \sqrt{4 + (5 + 4\lambda)[\ln(1 + \lambda)]^2}}{2\ln(1 + \lambda)}.$$

It can be shown from Figure 1 that $x_m > 0$, for $0 < \lambda < 1$ and when $\lambda \geq 1$, then $x_m = 0$. The moment, the probability and the cumulant generating functions are given, respectively, is

$$M_X(t) = \frac{\lambda^3}{\lambda^2 + 2\lambda + 2}\left(s + 2s^2 + 2s^3\right); \ s = s(t) = [\lambda + 1 - \exp(t)]^{-1},$$

$$G_X(t) = \frac{\lambda^3}{\lambda^2 + 2\lambda + 2}\left(r + 2r^2 + 2r^3\right); \ r = s(\ln t) = (\lambda + 1 - t)^{-1},$$

$$C_X(t) = \ln M_X(t) = 3\ln\lambda - \ln\left(\lambda^2 + 2\lambda + 2\right) + \ln\left(s + 2s^2 + 2s^3\right).$$

2.3. Parameter Estimation of the PISDL Distribution

The parameter of the PISDL distribution needs to be estimated before modeling real-world datasets. Here, parameter estimation is based on the two commonly used estimation methods, which are the methods of moments and maximum likelihood.

2.3.1. Method of Moments Estimator

The method of moments estimator can be obtained by equating the sample mean to the population mean. Therefore, the moment estimator of λ, hereby denoted as $\widetilde{\lambda}$, can be obtained by solving the following equation:

$$\overline{x} = \frac{\widetilde{\lambda}^2 + 4\widetilde{\lambda} + 6}{\widetilde{\lambda}\left(\widetilde{\lambda}^2 + 2\widetilde{\lambda} + 2\right)}, \tag{7}$$

or equivalently solving the cubic equation $A\widetilde{\lambda}^3 + B\widetilde{\lambda}^2 + C\widetilde{\lambda} + D = 0$, where $A = \overline{x}$, $B = 2\overline{x} - 1$, $C = 2\overline{x} - 4$, and $D = -6$.

2.3.2. Maximum Likelihood Estimator

Note that the log-likelihood function can be written as $l = \ln L = \sum_{i=1}^{n} \ln f(x_i) = \sum_{x=0}^{\infty} n_x \ln[f(x)]$, where n_x is the frequency of x-valued data. Hence, l is given as

$$l = 3n\ln\lambda - n\ln(\lambda^2 + 2\lambda + 2) - n(\overline{x} + 3)\ln(\lambda + 1)$$
$$+ \sum_{x=0}^{\infty} n_x \ln\left[x(x + 5) + 2\lambda(x + 2) + \lambda^2 + 5\right].$$

By differentiating l with respect to λ, we obtain

$$\frac{dl}{d\lambda} = \frac{3n}{\lambda} - \frac{2n(\lambda + 1)}{\lambda^2 + 2\lambda + 2} - \frac{n(\overline{x} + 3)}{\lambda + 1} + \sum_{x=0}^{\infty} n_x \frac{2(\lambda + x + 2)}{x(x + 5) + 2\lambda(x + 2) + \lambda^2 + 5}, \tag{8}$$

in which equating it to zero and solving it yields the maximum likelihood estimator (MLE) for λ, which is abbreviated as $\hat{\lambda}$. Equivalently, one can directly maximize the l to obtain a similar result. To estimate the variance of $\hat{\lambda}$, Fisher's information about λ and $I(\lambda)$ needs to be obtained, and it is given as

$$I(\lambda) = -\frac{4\lambda^6 + 7\lambda^5 - 7\lambda^4 - 40\lambda^3 - 52\lambda^2 - 36\lambda - 12}{\lambda^2(\lambda+1)^2(\lambda^2+2\lambda+2)^2}$$

$$+\frac{4\lambda^3}{(\lambda+1)^3(\lambda^2+2\lambda+2)}\sum_{x=0}^{\infty}\frac{(\lambda+x+2)^2}{[x(x+5)+2\lambda(x+2)+\lambda^2+5](\lambda+1)^x}. \tag{9}$$

The summation term cannot be written in a closed form, but one may use several Lerch Transcendent [35] functions for it. However, it is only practical to leave the summation term as is. So, the variance of $\hat{\lambda}$ can be written as $I^{-1}(\lambda)/n$ and, subsequently, the $(1-\alpha)100\%$ confidence interval can be written as $\hat{\lambda} \mp z_\alpha I^{-1/2}(\lambda)/\sqrt{n}$. If $\alpha = 0.025$, and then the 95% confidence interval will be obtained. The summation term reduces considerably to a simple form when $\hat{\lambda}$ is substituted, thus resulting in a constant. This will eventually give an estimated variance of $\hat{\lambda}$. For example, for $n = 100$ and when $\hat{\lambda} = 1.8793$, the summation equates to 1.4284, which then results in the estimated variance of 0.0145.

2.4. Simulation Study

A simulation study is conducted to assess the performance of the two earlier estimators in estimating the parameter of the PISDL distribution. The algorithm for the simulation study is as follows:

Step 1: Generate $N = 1000, 2000, \ldots, 10,000$ random data that follows the PISDL distribution with $\lambda = 0.5$.

Step 2: Obtain the estimated λ using MLE and moment estimator.

Step 3: Repeat Steps 1–2 for a total of 2000 iterations and obtain the estimates.

Step 4: Calculate the mean absolute deviation, *MAD*, and the mean squared error values, *MSEs*, given, respectively, as $MAD = \sum_{i=1}^{2000}|\check{\lambda} - \lambda|/2000$ and $MSE = \sum_{i=1}^{2000}(\check{\lambda} - \lambda)^2/2000$, where $\check{\lambda}$ can be the MLE or moment estimator for λ.

Step 5: Repeat Steps 3–4 for $\lambda = 2.0, 5.0$.

Generally, if the *MAD* and *MSE* approach zero as *N* increases, then the estimate is asymptotically unbiased and consistent. For the simulation study, R software version 3.0.2 is used, and the estimated parameter using MLE is obtained using the *optim* command to optimize the log-likelihood value. The results of the simulation study are presented in Figure 5. In Figure 5, for any value of λ, as *N* increases, the *MAD* and the *MSE* of the MLE and the moment estimator decrease, suggesting that both estimates are asymptotically unbiased and consistent.
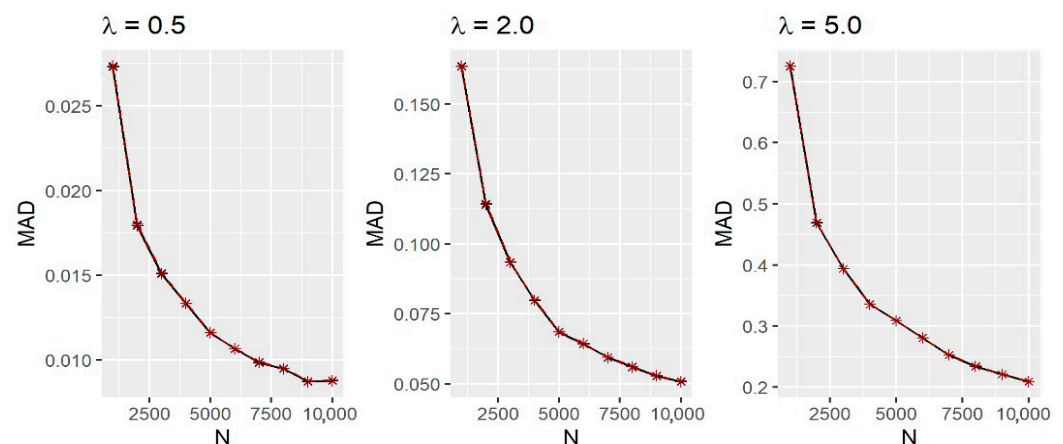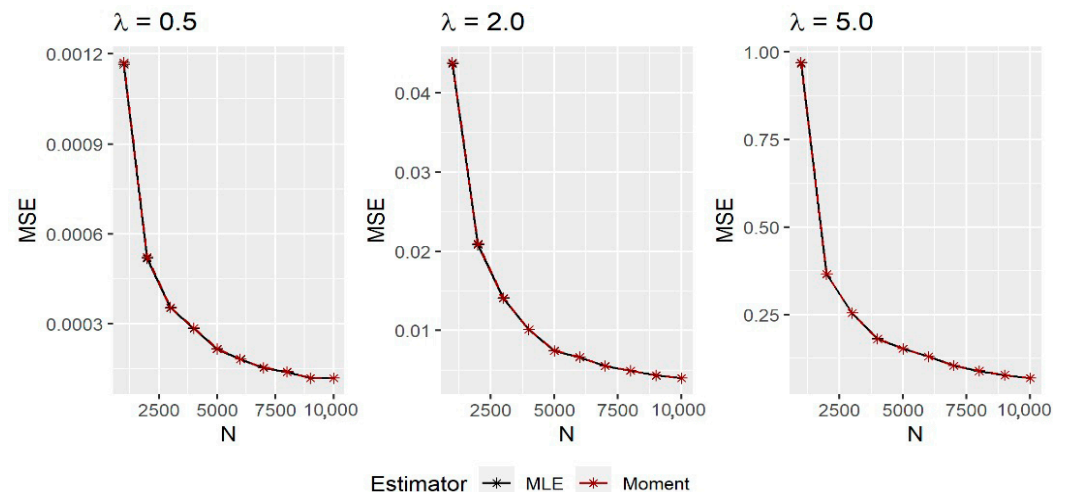

Figure 5. *Cont.*

Figure 5. The *MAD* and the *MSE* of the MLE and moment estimator for the *PISDL* distribution when $\lambda = 0.5,\ 2.0,\ 5.0$ and $N = 1000,\ 2000,\dots,\ 10,000$.

3. Zero-Truncated Poisson-Improved Second-Degree Lindley Distribution

Truncation is a widely used trait in real-world situations in a variety of domains, including industry [23–25], medicine [23,24], and many more. The progression of a disease that is not an increasing function but will stabilize after a certain period is an example of truncation. Therefore, a flexible truncated count data distribution is introduced by truncating the PISDL distribution at zero, yielding a zero-truncated Poisson-improved second-degree Lindley distribution (ZTPISDL). It was observed that the PISDL and the PL distributions are equally competent based on the two datasets considered in Section 5. Therefore, it is expected that the ZTPISDL distribution to be as competent as the zero-truncated PL (ZTPL) distribution. The development and the statistical properties of the ZTPISDL distribution are discussed in the following sections.

3.1. Probability Mass Function of the ZTPISDL Distribution

Definition 2. *A random variable Y is said to follow a ZTPISDL distribution if it obeys $X|\lambda \sim PISDL(\lambda)$ and $p(y) = f(y)/[1 - f(0)]$, where $p(y)$ is the pmf of Y.*

Theorem 2. *Let Y be a random variable that follows the ZTPISDL with parameter λ, and then the pmf of Y is given as*

$$p(y) = \frac{\lambda^3}{\lambda^4 + 6\lambda^3 + 13\lambda^2 + 8\lambda + 2}\left[\frac{y(y+5) + 2\lambda(y+2) + \lambda^2 + 5}{(\lambda+1)^y}\right],\ y = 1, 2, 3, \dots \quad (10)$$

Proof. Using Definition 2, the pmf of Y is obtained. $\square$

Figure 6 shows the pmf plot for the ZTPISDL distribution that follows similar shapes as the pmf plot for the PISDL distribution, which is skewed to the right, unimodal, and decreasing. Using the log of $p(y)$ defined in (10), the mode y_{mod} of the distribution can be obtained by solving the following formula:

$$y_{\text{mod}} = \frac{-D + \sqrt{D^2 - 4AE}}{2A},$$

where $A = \ln(\lambda + 1)$, $D = A(2\lambda + 5) - 2$, and $E = A\left(\lambda^2 + 4\lambda + 5\right) - 2\lambda - 5$.
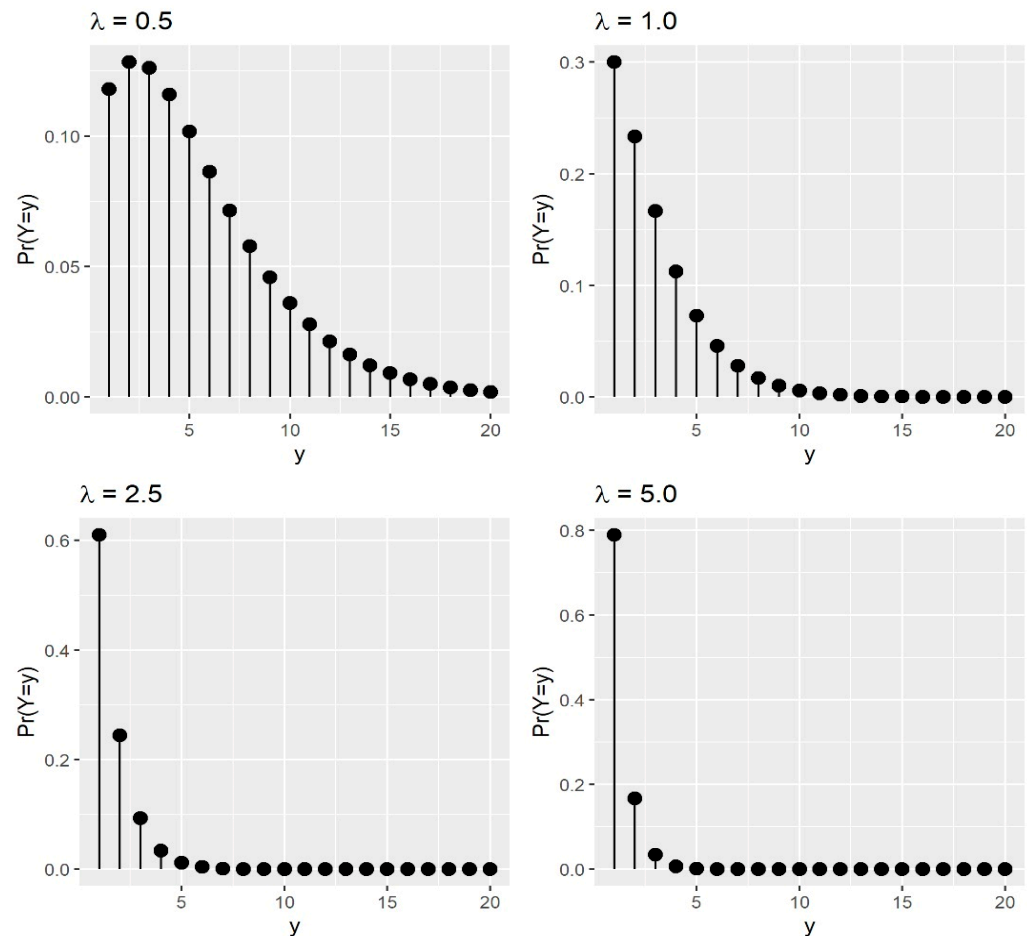
Figure 6. The pmf plot for the *ZTPISDL* distribution $\lambda = 0.5$, 1.0, 2.5, 5.0.

3.2. Some Statistical Properties of the ZTPISDL Distribution

If $Y \sim ZTPISDL$ for $y = 1, 2, 3, \ldots$, then the kth moment for Y can be easily obtained because it satisfies $E(Y^k) = E(X^k)/[1 - f(0)]$. The first two moments in the origin are obtained and can be, respectively, written as

$$E(Y) = \frac{(\lambda + 1)^3 (\lambda^2 + 4\lambda + 6)}{\lambda(\lambda^4 + 6\lambda^3 + 13\lambda^2 + 8\lambda + 2)},$$

$$E(Y^2) = \frac{(1 + \lambda)^3 ((\lambda + 2)(\lambda^2 + 4\lambda + 6) + 4(3 + \lambda))}{\lambda^2 (2 + \lambda(1 + \lambda)(8 + \lambda(5 + \lambda)))}.$$

Using a similar approach as the IOD_X for X, the index of dispersion for Y (IOD_Y) can be written as

$$IOD_Y = \frac{Var(Y)}{E(Y)} = \frac{1}{\lambda} - \frac{2(2 + \lambda)}{6 + \lambda(4 + \lambda)} + \frac{4 + 2\lambda(8 + \lambda(5 + \lambda))}{2 + \lambda(1 + \lambda)(8 + \lambda(5 + \lambda))}.$$

Based on the IOD_Y, the ZTPISDL distribution is underdispersed (overdispersed) when $\lambda > (<)1.51494$. The ZTPISDL distribution is only equidispersed when $\lambda = 1.51494$. The recurrence probability for Y is similar for X, except the x is substituted with y because

$$\frac{p(y + 1)}{p(y)} = \frac{f(y + 1)/[1 - f(0)]}{f(y)/[1 - f(0)]} = \frac{f(y + 1)}{f(y)}.$$

The generating functions for both ZTPISDL and PISDL distributions can be related since their pmfs are related as well. Therefore, their relationships are given, respectively, be-

low so that the moment generating, the cumulant generating, and the probability generating functions for Y can be worked out as follows:

$$M_Y(t) = \frac{M_X(t) - f(0)}{1 - f(0)},$$

$$C_Y(t) = \ln[M_Y(t)] = \ln[M_X(t) - f(0)] - \ln[1 - f(0)],$$

$$G_Y(t) = M_Y(\ln t) = \frac{M_X(\ln t) - f(0)}{1 - f(0)} = \frac{G_X(t) - f(0)}{1 - f(0)}.$$

3.3. Parameter Estimation of the ZTPISDL Distribution

The parameter of the ZTPISDL distribution can be estimated using the moment and MLE techniques. The moment estimator of λ, hereby denoted as $\tilde{\lambda}$, can be obtained by solving the following equation:

$$\overline{y} = \frac{\left(\tilde{\lambda} + 1\right)^3 \left(\tilde{\lambda}^2 + 4\tilde{\lambda} + 6\right)}{\tilde{\lambda}\left(\tilde{\lambda}^4 + 6\tilde{\lambda}^3 + 13\tilde{\lambda}^2 + 8\tilde{\lambda} + 2\right)}, \tag{11}$$

or equivalently solving the quintic equation $A\tilde{\lambda}^5 + B\tilde{\lambda}^4 + C\tilde{\lambda}^3 + D\tilde{\lambda}^2 + E\tilde{\lambda} + F = 0$, such that $A = 1 - \overline{y}$, $B = 7 - 6\overline{y}$, $C = 21 - 13\overline{y}$, $D = 31 - 8\overline{y}$, $E = 22 - 2\overline{y}$, and $F = 6$. For MLE, the log-likelihood function l is given as

$$l = 3n \ln \lambda - n \ln\left(\lambda^4 + 6\lambda^3 + 13\lambda^2 + 8\lambda + 2\right) - n\overline{y}\ln(\lambda + 1)$$

$$+ \sum_{y=1}^{\infty} n_y \ln\left[y(y+5) + 2\lambda(y+2) + \lambda^2 + 5\right].$$

By differentiating l with respect to λ, we obtain

$$\frac{dl}{d\lambda} = \frac{3n}{\lambda} - \frac{n\left(4\lambda^3 + 18\lambda^2 + 26\lambda + 8\right)}{\lambda^4 + 6\lambda^3 + 13\lambda^2 + 8\lambda + 2} - \frac{n\overline{y}}{\lambda + 1} + \sum_{y=1}^{\infty} n_y \frac{2(\lambda + y + 2)}{y(y+5) + 2\lambda(y+2) + \lambda^2 + 5}, \tag{12}$$

in which equating it to zero and solving it implies the MLE for λ, which is denoted as $\hat{\lambda}$. Equivalently, one can directly maximize the l directly to obtain a similar result.

3.4. Simulation Study

A simulation study is conducted to assess the performance of the obtained estimators for the parameter of the ZTPISDL distribution. The algorithm for the simulation study is similar to the one in Section 2.4, except that the data are generated using the ZTPISDL distribution. Similarly, R software is used, and the estimated parameter using MLE is obtained using the *optim* command to optimize the log-likelihood value. The findings of the simulation study are shown in Figure 7. Figure 7 shows that for any value of λ, as N increases, the *MAD* and the *MSE* of both estimates fall, indicating that they are both asymptotically unbiased and consistent.

When dealing with a truncated distribution, the population size is usually unknown and needs to be estimated. Assuming that a population follows the ZTPISDL distribution, a population size estimator is developed and studied. The discussion on the population size is provided in the next section.
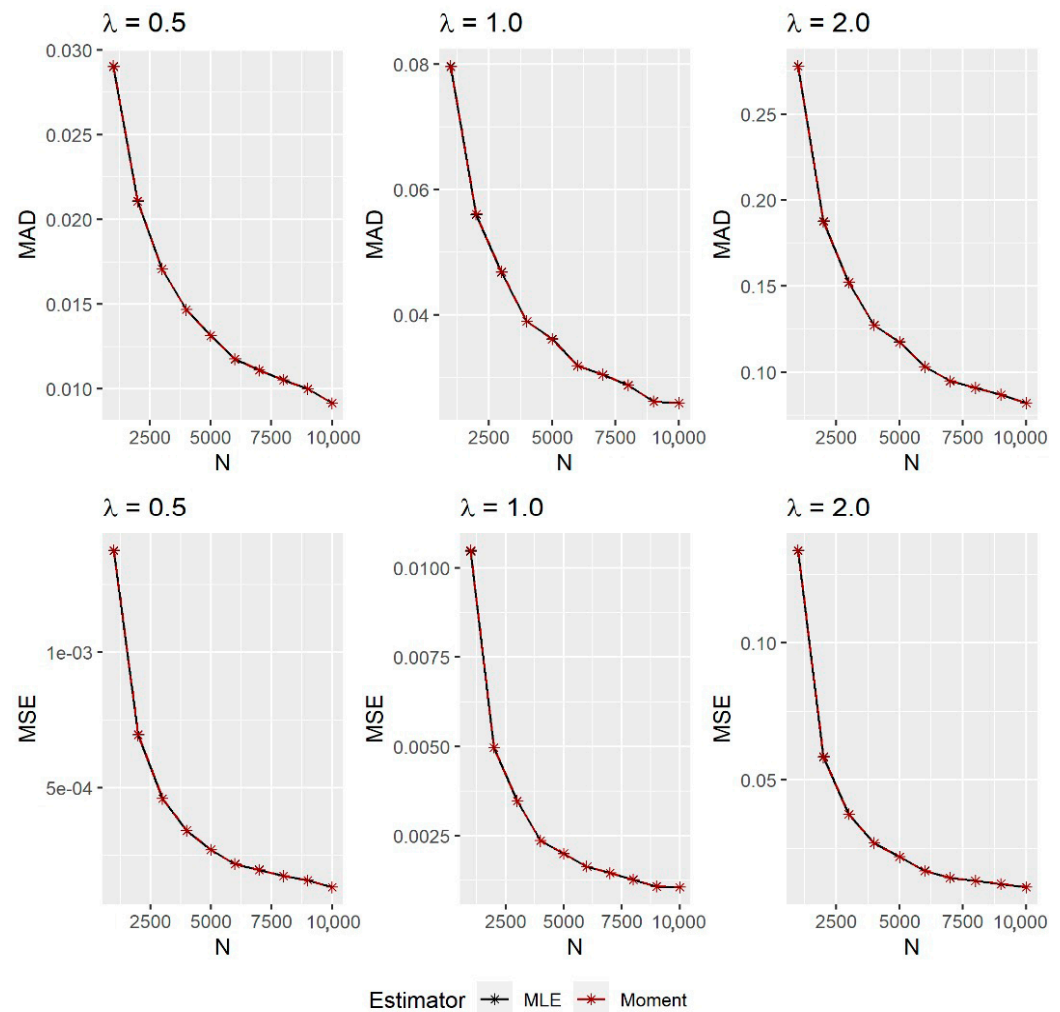
Figure 7. The *MAD* and the *MSE* of the MLE and moment estimator for the *ZTPISDL* distribution when $\lambda = 0.5$, 1.0, 2.0 and $N = 1000, 2000, \ldots, 10,000$.

4. Population Size Estimation

4.1. Horvitz–Thompson Estimator under ZTPISDL Distribution (HT-ZTPISDL)

A popular estimator for the population size is the Horvitz–Thompson estimator [36], which includes information from both truncated and untruncated distributions. A Horvitz–Thompson estimator has the following form:

$$\hat{N} = \frac{n}{1 - \Pr(X = 0|\hat{\omega})},$$

where $\hat{\omega}$ is the estimator for ω for a distribution and n is the sample size. Basically, the estimated parameter of the truncated distribution is substituted into the probability mass function of the untruncated distribution for the unobserved events. The estimator for the population size $\hat{N}$, which follows a ZTPISDL distribution in the form of the Horvitz–Thompson estimator (HT-ZTPISDL), is given as

$$\hat{N} = \frac{n(\hat{\lambda} + 1)^3(\hat{\lambda}^2 + 2\hat{\lambda} + 2)}{\hat{\lambda}^4 + 6\hat{\lambda}^3 + 13\hat{\lambda}^2 + 8\hat{\lambda} + 2}, \tag{13}$$

where $\hat{\lambda}$ is the MLE for λ in the ZTPISDL distribution.

4.2. Variance and Confidence Interval for HT-ZTPISDL

Böhning [37] has provided a simple yet informative method for obtaining a variance for any population size estimator using the conditional expectation technique. The variance of HT-ZTPISDL can be written as

$$
Var(\hat{N}) = Var_{\hat{\lambda},n}\left[\frac{n}{g(\hat{\lambda})}\right] = Var_n\left\{E_{\hat{\lambda}|n}\left[\frac{n}{g(\hat{\lambda})}\right]\right\} + E_n\left\{Var_{\hat{\lambda}|n}\left[\frac{n}{g(\hat{\lambda})}\right]\right\},
\qquad (14)
$$

where $g(\lambda) = 1 - f(0)$. Observe that the variation in estimating the population size comes from two sources of variation. The first term in Equation (14) explains the binomial variation involved in sampling n units of data with population size N and probability $g(\lambda)$ [37]. The second term in Equation (14) explains the variation that occurs when estimating parameter λ using n observed data [37]. Using the delta method for the first term of Equation (14), we obtain

$$
E_{\hat{\lambda}|n}\left[\frac{n}{g(\hat{\lambda})}\right] = \frac{n}{g(\lambda)}.
$$

As $n \sim binomial(N, g(\lambda))$, we obtain

$$
Var_n\left\{E_{\hat{\lambda}|n}\left[\frac{n}{g(\hat{\lambda})}\right]\right\} = Var_n\left[\frac{n}{g(\lambda)}\right] = \frac{Ng(\lambda)[1-g(\lambda)]}{g(\lambda)^2}.
$$

The equation above is further estimated by substituting $Ng(\lambda)$ with n and λ with $\hat{\lambda}$, yielding

$$
\widehat{Var}_n\left\{E_{\hat{\lambda}|n}\left[\frac{n}{g(\hat{\lambda})}\right]\right\} \approx \frac{n\left[1-g(\hat{\lambda})\right]}{g(\hat{\lambda})^2}.
$$

For the ZTPISDL distribution, we obtain

$$
g(\lambda) = \frac{\lambda^4 + 6\lambda^3 + 13\lambda^2 + 8\lambda + 2}{(\lambda+1)^3(\lambda^2 + 2\lambda + 2)}.
$$

Therefore,

$$
\widehat{Var}_n\left\{E_{\hat{\lambda}|n}\left[\frac{n}{g(\hat{\lambda})}\right]\right\} \approx \frac{n\hat{\lambda}^3(\hat{\lambda}+1)^3(\hat{\lambda}^2 + 2\hat{\lambda} + 2)(\hat{\lambda}^2 + 4\hat{\lambda} + 5)}{\left(\hat{\lambda}^4 + 6\hat{\lambda}^3 + 13\hat{\lambda}^2 + 8\hat{\lambda} + 2\right)^2}.
\qquad (15)
$$

Now, consider the second term of Equation (14), and assume that

$$
E_n\left\{Var_{\hat{\lambda}|n}\left[\frac{n}{g(\hat{\lambda})}\right]\right\} \approx Var_{\hat{\lambda}|n}\left[\frac{n}{g(\hat{\lambda})}\right].
$$

Using the delta method, we obtain

$$
Var_{\hat{\lambda}|n}\left[\frac{n}{g(\hat{\lambda})}\right] = n^2 Var_{\hat{\lambda}|n}\left[\frac{1}{g(\hat{\lambda})}\right] \approx n^2\left[\frac{g'(\lambda)}{g(\lambda)^2}\right]^2 Var(\hat{\lambda}),
$$

where $Var(\hat{\lambda}) = I^{-1}(\lambda)/n$ and $I(\lambda)$ is given in Equation (9). Therefore,

$$
Var_{\hat{\lambda}|n}\left[\frac{n}{g(\hat{\lambda})}\right] = \frac{n\lambda^4(\lambda+1)^4\left(\lambda^4 + 10\lambda^3 + 37\lambda^2 + 52\lambda + 30\right)^2}{\left(\lambda^4 + 6\lambda^3 + 13\lambda^2 + 8\lambda + 2\right)^4} I^{-1}(\lambda).
$$

By substituting λ with $\hat{\lambda}$, the second term of Equation (14) can be estimated as

$$\hat{E}_n\left\{Var_{\hat{\lambda}|n}\left[\frac{n}{g(\hat{\lambda})}\right]\right\} = \frac{n\hat{\lambda}^4(\hat{\lambda}+1)^4(\hat{\lambda}^4 + 10\hat{\lambda}^3 + 37\hat{\lambda}^2 + 52\hat{\lambda} + 30)^2}{(\hat{\lambda}^4 + 6\hat{\lambda}^3 + 13\hat{\lambda}^2 + 8\hat{\lambda} + 2)^4}I^{-1}(\hat{\lambda}). \quad (16)$$

By combining Equations (15) and (16), the variance of the HT-ZTPISDL can be estimated as

$$\begin{aligned}
\widehat{Var}(\hat{N}) &= \frac{n\hat{\lambda}^3(\hat{\lambda}+1)^3(\hat{\lambda}^2 + 2\hat{\lambda} + 2)(\hat{\lambda}^2 + 4\hat{\lambda} + 5)}{(\hat{\lambda}^4 + 6\hat{\lambda}^3 + 13\hat{\lambda}^2 + 8\hat{\lambda} + 2)^2} \\
&\quad + \frac{n\hat{\lambda}^4(\hat{\lambda}+1)^4(\hat{\lambda}^4 + 10\hat{\lambda}^3 + 37\hat{\lambda}^2 + 52\hat{\lambda} + 30)^2}{(\hat{\lambda}^4 + 6\hat{\lambda}^3 + 13\hat{\lambda}^2 + 8\hat{\lambda} + 2)^4}I^{-1}(\hat{\lambda}).
\end{aligned} \quad (17)$$

Therefore, the 95% confidence interval for the estimator can be written as $\hat{N} \mp z_{0.025}\sqrt{\widehat{Var}(\hat{N})}$, where $z_{0.025} = 1.96$.

4.3. Simulation Study

A simulation study is conducted to assess the performance of the HT-ZTPISDL estimator in estimating the population size when the data are generated from the ZTPISDL distribution. The algorithm for the simulation study is as follows:

Step 1: Generate $N = 1000, 2000, \ldots, 10,000$ random data, which follow the ZTPISDL distribution with $\lambda = 0.5$.

Step 2: Obtain $\hat{\lambda}$ using the MLE and use $\hat{\lambda}$ to obtain $\hat{N}$.

Step 3: Repeat Steps 1–2 for a total of 2000 iterations and obtain the estimates.

Step 4: Calculate the relative absolute error, RAB values, and the relative standard deviation, RSd values, given, respectively, as $RAB = \left|\overline{N} - N\right|/N$ and $RSd = \frac{1}{N}\sqrt{\sum_{i=1}^{2000}\left(\hat{N} - \overline{N}\right)^2/2000}$, where $\overline{N} = \sum_{i=1}^{2000}\hat{N}/2000$.

Step 5: Repeat Steps 3–4 for $\lambda = 2.0, 5.0$.

R software is used to obtain $\hat{\lambda}$ using the *optim* command, and $\hat{N}$ is obtained by plugging in $\hat{\lambda}$ in Equation (10). The results of the simulation study are presented in Figure 8. In this figure, for any value of λ, as N increases, the MAD and the MSE of $\hat{N}$ decreases, suggesting that $\hat{N}$ is asymptotically unbiased and consistent.
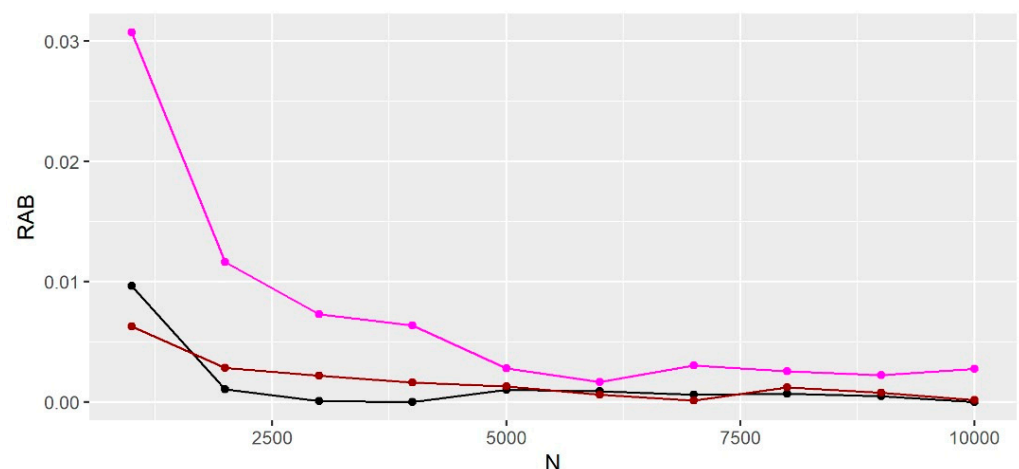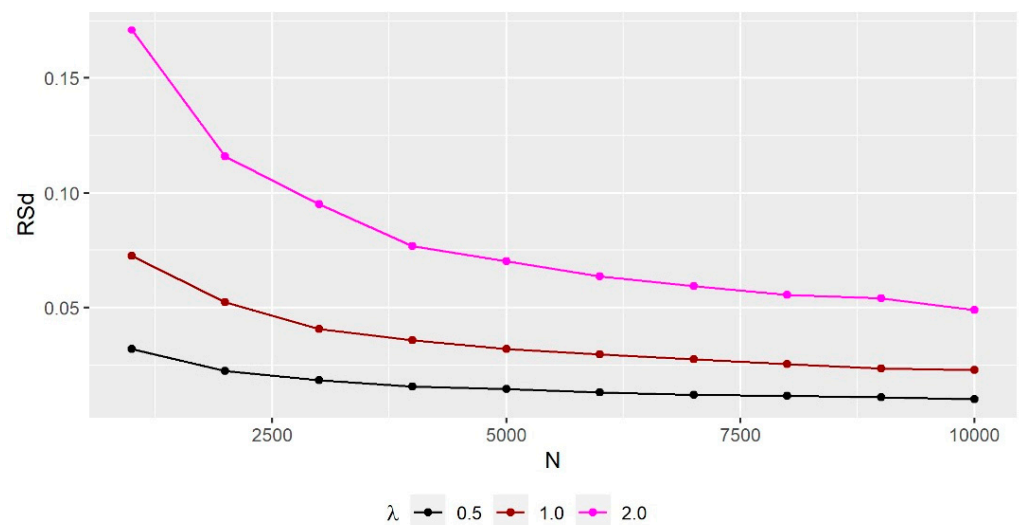

Figure 8. *Cont.*

Figure 8. The *RAB* and the *RSd* of the estimated population size when $\lambda = 0.5,\ 1.0,\ 2.0$.

5. Medical Data Applications

The applications in the medical datasets are segregated into three subsections with respect to the PISDL distribution, the ZTPISDL distribution, and the estimation of the population size. For a comparison of the model fittings, Akaike's information criterion, *AIC* [38], and Bayesian information criterion, *BIC* [39], are used.

5.1. Model Fittings Using the PISDL Distribution

Two datasets on the number of dicentric chromosomes after being exposed to different doses of radiation (0.405 and 0.600) that were studied by Puig and Barquinero [40] are considered in fitting using the PISDL distribution. The two datasets are overdispersed with dispersion values of 1.2704 and 1.2178, respectively. Since the PISDL distribution is closely related to the Poisson and the PL distributions, the model fittings from the PISDL distribution are compared with those from the Poisson and the PL distributions. The results of the model fittings for the two datasets using the three distributions are summarized in Tables 1 and 2.

Table 1. Model fittings of the number of dicentric chromosomes after being exposed to a 0.405 radiation dose using Poisson, PL, and PISDL distributions.

		Distributions			
x	n_x	**Poisson**	*PL*	*PISDL* (MLE)	*PISDL* (Moment)
0	437	426.55	433.87	433.73	433.64
1	66	84.50	72.04	72.30	72.36
2	15	8.37	11.81	11.75	11.77
3	1	0.55	1.91	1.87	1.88
4	1	0.03	0.37	0.35	0.37
Total	520	520.00	520.00	520.00	520.00
Parameter					
$\hat{\lambda}$		0.1981	-	6.5464	6.5392
$\hat{\theta}$		-	5.7953	-	-
Max log-likelihood		−285.14	−279.40	−279.45	
AIC		572.27	560.80	560.90	-
BIC		576.52	565.05	565.15	-
χ^2		11.55	1.13	1.23	1.22
df		1	1	1	1
p-value		0.0007	0.2878	0.2674	0.2694

Table 2. Model fittings of the number of dicentric chromosomes after being exposed to a 0.600 radiation dose using Poisson, PL, and PISDL distributions.

		Distributions			
x	n_x	Poisson	PL	*PISDL* (MLE)	*PISDL* (Moment)
0	473	456.69	475.79	474.91	475.07
1	119	147.65	117.81	118.96	118.88
2	34	23.87	28.53	28.55	28.50
3	3	2.57	6.79	6.64	6.62
4	2	0.22	2.08	1.94	1.93
Total	631	631.00	631.00	631.00	631.00
Parameter					
$\hat{\lambda}$		0.3233	-	4.4001	4.4046
$\hat{\theta}$		-	3.7420	-	-
Max log-likelihood		−469.65	−464.09	−463.96	
AIC		941.30	930.17	929.91	-
BIC		945.75	934.61	934.36	-
χ^2		11.85	2.77	2.54	2.54
df		1	2	2	2
p-value		0.0006	0.2503	0.2808	0.2808

In both tables, the Poisson distribution does not fit the data based on the *p*-value. On the other hand, the model fittings based on both PL and PISDL distributions gave similar *AIC* and *BIC* values, as well as non-significant *p*-values, indicating that both distributions can be used for describing the number of dicentric chromosomes after being exposed to radiation of different doses. However, the first dataset was fitted better by the PL distribution, whereas the second dataset was fitted better by the PISDL distribution. Therefore, it is reasonable to suggest that both PL and PISDL distributions are equally competent and can be selected as the best distributions in explaining the number of dicentric chromosomes after being exposed to two different doses of radiation.

The comparison between the empirical plots of the data and the fitted values based on Equation (1) in Theorem 1 for the considered datasets, (i) the number of dicentric chromosomes after being exposed to a 0.405 radiation dose, and (ii) the number of dicentric chromosomes after being exposed to a 0.600 radiation dose are presented in Figure 9.

5.2. Model Fittings Using ZTPISDL Distribution

A dataset on the number of positive samples of *Salmonella* data, which was initially given in a survey study by Snow et al. [41] and later summarized by Arnold et al. [42], is considered in model fitting using the zero-truncated Poisson (ZTP), the zero-truncated PL (ZTPL), and the ZTPISDL distributions. The data refer to the number of farms with at least one positive sample of *Salmonella*. The dataset is overdispersed with a dispersion value of 1.4381. The results of the model fittings are given in Table 3.

In Table 3, the ZTP distribution does not provide a good fit to the data based on the *p*-value. On the contrary, the ZTPISDL and the ZTPL distributions provide a good fit to the data based on the *AIC* and *BIC* values, as well as the non-significant *p*-value. However, the dataset was fitted better by the ZTPISDL distribution. Therefore, this suggests that the ZTPISDL is the best distribution in describing the number of positive samples of *Salmonella* data. The comparison between the empirical plots of the data and the fitted value based on Equation (10) in Theorem 2 for the above dataset is presented in Figure 10.
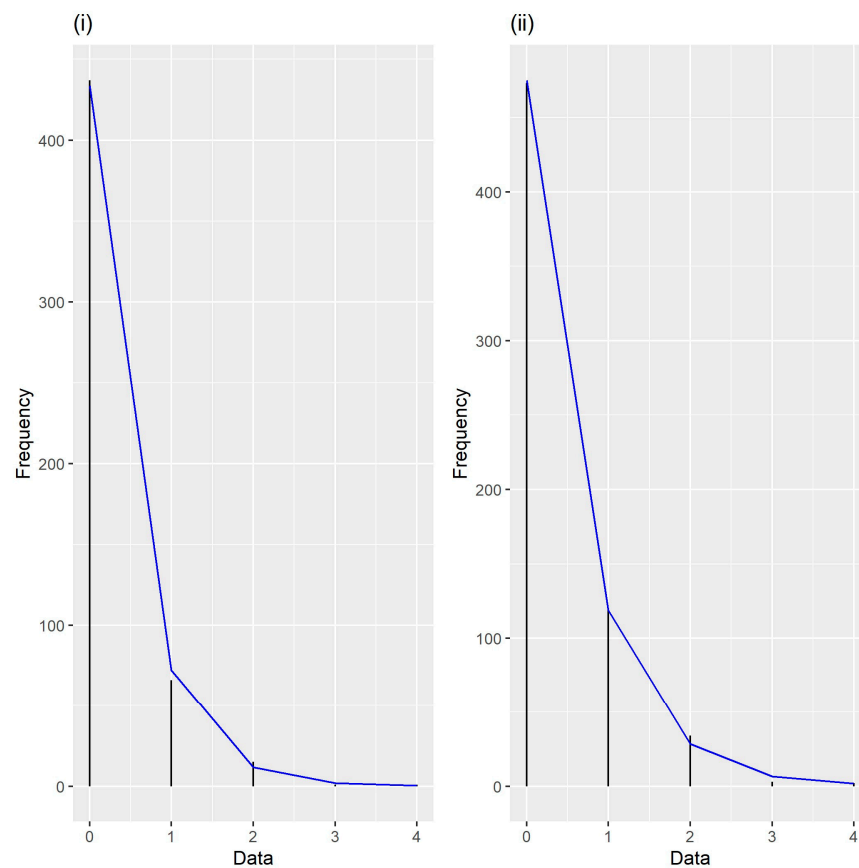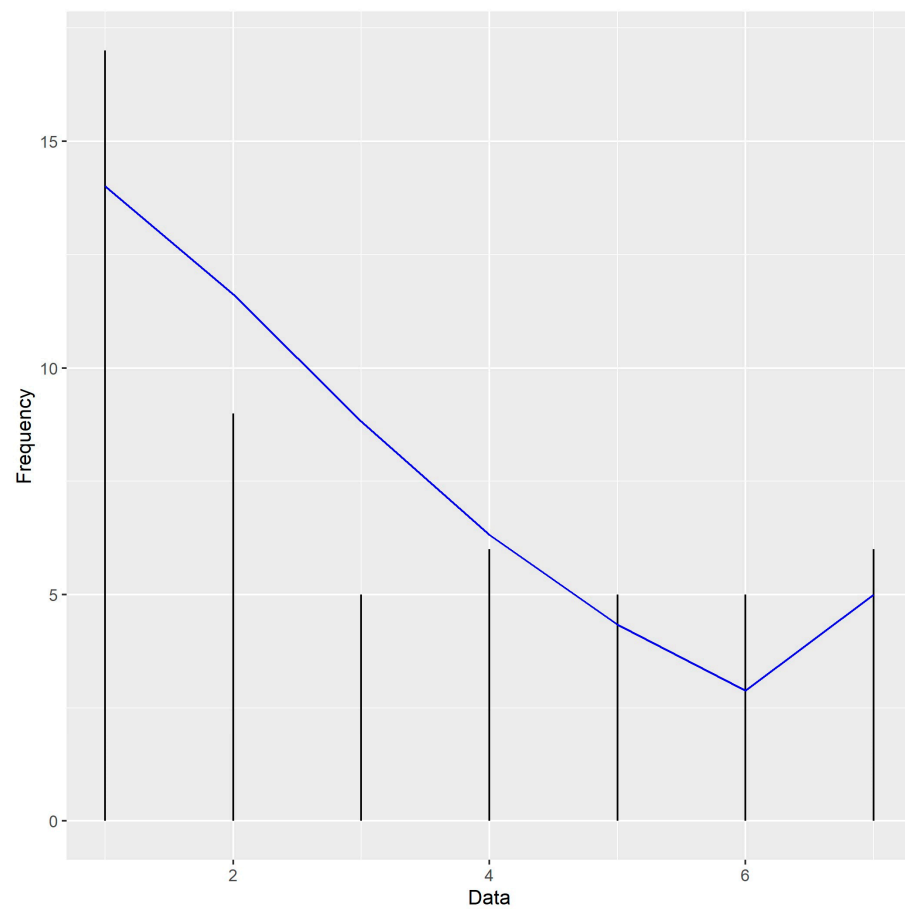
Figure 9. Plots of the empirical (vertical black line) and the fitted (blue line) for (**i**) the number of dicentric chromosomes after being exposed to a 0.405 radiation dose and (**ii**) the number of dicentric chromosomes after being exposed to a 0.600 radiation dose.

Table 3. Model fittings of the number of positive samples of *Salmonella* using the ZTP, ZTPL, and ZTPISDL distributions.

y	n_y	Distributions			
		ZTP	*ZTPL*	*ZTPISDL* **(MLE)**	*ZTPISDL* **(Moment)**
1	17	7.88	15.06	14.01	14.00
2	9	12.12	11.55	11.62	11.62
3	5	12.44	8.45	8.82	8.82
4	6	9.57	5.99	6.32	6.32
5	5	5.89	4.15	4.34	4.34
6	5	3.02	2.83	2.89	2.89
7	6	2.08	4.97	5.00	5.01
Total	53	53.00	53.00	53.00	53.00
Parameter					
$\hat{\lambda}$		3.0778	-	0.8932	0.8928
$\hat{\theta}$		-	0.6660	-	-
Max log-likelihood		−110.64	−105.28	−105.10	
AIC		223.27	212.55	212.19	-
BIC		225.24	214.53	214.16	-
χ^2		24.10	3.61	4.16	4.16
df		4	3	4	4
p-value		<0.0001	0.3068	0.3848	0.3848

Figure 10. A plot of the empirical (vertical black line) and the fitted values (blue line) for the number of positive samples of *Salmonella*.

5.3. Estimating Population Size

The dataset studied in Section 5.2 on the number of positive samples of *Salmonella* data is considered to estimate the population size of the sample. The Horvitz–Thompson estimator based on the proposed ZTPISDL distribution is compared with those based on the ZTP and the ZTPL distributions. The Horvitz–Thompson estimators based on the ZTP [43,44] and the ZTPL [32] distributions are, respectively, given as

$$\hat{N}_{ZTP} = \frac{n}{1 - \exp(-\hat{\lambda}_{ZTP})},$$

and

$$\hat{N}_{ZTPL} = \frac{n(\hat{\theta} + 1)^3}{\hat{\theta}^2 + 3\hat{\theta} + 1},$$

where $\hat{\lambda}_{ZTP}$ refers to the MLE of λ for the ZTP distribution and $\hat{\theta}$ refers to the MLE of θ for the ZTPL distribution. The estimated population sizes and their corresponding standard deviations, as well as the lower and upper limits for the 95% confidence interval, are presented in Table 4. Based on Table 4, the estimated population size based on the ZTPISDL distribution using the MLE is 66.64 with a 95% confidence interval between 57.04 and 76.24. Since the ZTPISDL distribution based on the MLE provides the best fit for the data (refer to Table 3), the resulting estimated population size is acceptable.

Table 4. The estimated population size, the standard deviation, and the lower and upper limits for the 95% confidence interval of the population size estimator based on the ZTP, ZTPL, and ZTPISDL distributions.

Distributions	Estimated $\hat{N}$	SD	95% Lower Limit	95% Upper Limit
ZTP	55.56	1.761	52.11	59.01
ZTPL	71.21	4.045	63.28	79.14
ZTPISDL (MLE)	66.64	4.896	57.04	76.24
ZTPISDL (Moment)	63.10	4.961	53.38	72.82

6. Conclusions, Limitations, and Future Research

A tractable one-parameter Poisson-improved second-degree Lindley (PISDL) distribution has been proposed to address the need for modeling count data exhibiting overdispersion. This distribution is composed of three negative binomial distributions, each with fixed mixing proportions and parameters, allowing for the fitting of datasets originating from three sub-populations whose modes are proximal. However, if the three modes are far from each other and clearly visible from the plots of the datasets, the PISDL distribution may not be a good candidate for model fittings. The hazard rate function for the PISDL distribution showed an increasing shape. Parameters of the PISDL distribution have been estimated using maximum likelihood estimation (MLE) and moment methods, and both were found to be asymptotically unbiased and consistent. It has been observed from model fittings that the PISDL distribution performs on par with the PL distribution and surpasses the standard Poisson distribution in describing the number of dicentric chromosomes post-exposure to various radiation doses.

Given that data may not always present the frequency of unobserved events and exhibit dispersion, methodologies like zero truncation or the size-biased approach are employed. Zero truncation is a commonly favored method for handling datasets lacking frequencies of non-observed events. Hence, a zero-truncated version of the proposed distribution, named the zero-truncated PISDL (ZTPISDL) distribution, has been proposed to accommodate data exhibiting either over- or underdispersion. Parameters of the ZTPISDL distribution estimated by the MLE and moment methods have also been shown to be asymptotically unbiased and consistent. When applied to datasets, the MLE technique for estimating parameters of the ZTPISDL distribution has provided the best fit in comparison with zero-truncated Poisson and zero-truncated PL distributions.

When the population size is unknown due to the absence of frequencies for non-observed events in positive count data, estimation has been conducted using the Horvitz–Thompson estimator based on the ZTPISDL distribution. Since the ZTPISDL distribution has provided the best fit for the dataset considered in this study, the acceptance of the resulting estimated population size for the number of positive Salmonella samples is justified. It is suggested that this population size estimate may serve as a lower bound to the actual population size, especially when the ZTPISDL distribution is extended to linear models with the inclusion of relevant covariates. Moreover, the derivation of the variance and confidence interval for the population size estimator is intended to assist policymakers in revising rules and guidelines pertaining to the population, which will, in turn, benefit the population at large.

Despite its flexibility, the PISDL distribution's mixing proportion is solely dependent on λ, which consequently limits its flexibility. By introducing a new parameter (denoted as α) to influence the mixing proportion alongside λ, greater flexibility is achieved, resulting in a more adaptable PISDL distribution. This would, in turn, lead to a more versatile ZTPISDL distribution and improved estimates for population size.

Further research is anticipated to explore additional modifications and applications of the PISDL distribution. These include, but are not limited to, actuarial measures, such as value-at-risk and tail value-at-risk, reliability measures, like hazard rate and entropy, various forms of inflated models, including zero-inflated, k-inflated, and zero-one-inflated

distributions, and weighted models, like size-biased and area-biased distributions. These enhancements are expected to broaden the utility of the PISDL distribution, making it a competitive model in the realm of the statistical literature.

References

1. Lindley, D.V. Fiducial distributions and Bayes' theorem. *J. R. Stat. Soc. Ser. B Methodol.* **1958**, *20*, 102–107. [CrossRef]
2. Ghitany, M.E.; Atieh, B.; Nadarajah, S. Lindley distribution and its application. *Math. Comput. Simul.* **2008**, *78*, 493–506. [CrossRef]
3. Altun, E.; Cordeiro, G.M. The unit-improved second-degree Lindley distribution: Inference and regression modeling. *Comput. Stat.* **2020**, *35*, 259–279. [CrossRef]
4. Asgharzadeh, A.; Bakouch, H.S.; Nadarajah, S.; Sharafi, F. A new weighted Lindley distribution with application. *Braz. J. Probab. Stat.* **2016**, *30*, 1–27. [CrossRef]
5. Ghitany, M.E.; Alqallaf, F.; Al-Mutairi, D.K.; Husain, H.A. A two-parameter weighted Lindley distribution and its applications to survival data. *Math. Comput. Simul.* **2011**, *81*, 1190–1201. [CrossRef]
6. Ghitany, M.E.; Al-Mutairi, D.K.; Balakrishnan, N.; Al-Enezi, L.J. Power Lindley distribution and associated inference. *Comput. Stat. Data Anal.* **2013**, *64*, 20–33. [CrossRef]
7. Karuppusamy, S.; Balakrishnan, V.; Sadasivan, K. Improved second-degree Lindley distribution and its applications. *IOSR J. Math.* **2018**, *13*, 50–56.
8. Nadarajah, S.; Bakouch, H.S.; Tahmasbi, R. A generalized Lindley distribution. *Sankhya B* **2011**, *73*, 331–359. [CrossRef]
9. Shanker, R.; Mishra, A. A quasi Lindley distribution. *Afr. J. Math. Comput. Sci. Res.* **2013**, *6*, 64–71.
10. Shanker, R.; Mishra, A. A two-parameter Lindley distribution. *Stat. Transit. New Ser.* **2013**, *14*, 45–56. [CrossRef]
11. Zakerzadeh, H.; Dolati, A. Generalized Lindley distribution. *J. Math. Ext.* **2009**, *3*, 1–7.
12. Bakouch, H.S.; Al-Zahrani, B.M.; Al-Shomrani, A.A.; Marchi, V.A.; Louzada, F. An extended Lindley distribution. *J. Korean Stat. Soc.* **2012**, *41*, 75–85. [CrossRef]
13. Shanker, R.; Shukla, K.K.; Shanker, R.; Tekie, A.L. A three-parameter Lindley distribution. *Am. J. Math. Stat.* **2017**, *7*, 15–26.
14. Coşkun, K.U.Ş.; Korkmaz, M.Ç.; Kinaci, İ.; Karakaya, K.; Akdoğan, Y. Modified-Lindley distribution and its applications to the real data. *Commun. Fac. Sci. Univ. Ank. Ser. A1 Math. Stat.* **2022**, *71*, 252–272.
15. Sankaran, M. Note: The discrete Poisson-Lindley distribution. *Biometrics* **1970**, 145–149. [CrossRef]
16. Ghitany, M.E.; Al-Mutairi, D.K. Estimation methods for the discrete Poisson–Lindley distribution. *J. Stat. Comput. Simul.* **2009**, *79*, 1–9. [CrossRef]
17. Bhati, D.; Sastry, D.V.S.; Qadri, P.M. A new generalized Poisson-Lindley distribution: Applications and properties. *Austrian J. Stat.* **2015**, *44*, 35–51. [CrossRef]
18. Das, K.K.; Ahmed, I.; Bhattacharjee, S. A new three-parameter Poisson-Lindley distribution for modeling over dispersed count data. *Int. J. Appl. Eng. Res.* **2018**, *13*, 16468–16477.
19. Shanker, R.; Mishra, A. A two-parameter Poisson-Lindley distribution. *Int. J. Stat. Syst.* **2014**, *9*, 79–85.
20. Karlis, D.; Xekalaki, E. Mixed Poisson distributions. *Int. Stat. Rev.* **2005**, 35–58. [CrossRef]
21. Wasinrat, S.; Choopradit, B. The Poisson inverse Pareto distribution and its applications. *Thail. Stat.* **2023**, *21*, 110–124.
22. Erbayram, T.; Akdoğan, A. A new discrete model generated from mixed Poisson transmuted record type exponential distribution. *Ric. Math.* **2023**, 1–23. [CrossRef]
23. David, F.N.; Johnson, N.L. The truncated Poisson. *Biometrics* **1952**, *8*, 275–285. [CrossRef]
24. Sampford, M.R. The truncated negative binomial distribution. *Biometrika* **1955**, *42*, 58–69. [CrossRef]
25. Ghitany, M.E.; Al-Mutairi, D.K.; Nadarajah, S. Zero-truncated Poisson–Lindley distribution and its application. *Math. Comput. Simul.* **2008**, *79*, 279–287. [CrossRef]
26. Böhning, D.; Suppawattanabodee, B.; Kusolvisitkul, W.; Viwatwongkasem, C. Estimating the number of drug users in Bangkok 2001: A capture–recapture approach using repeated entries in one list. *Eur. J. Epidemiol.* **2004**, *19*, 1075–1083. [CrossRef] [PubMed]

27. Bouchard, M. A capture–recapture model to estimate the size of criminal populations and the risks of detection in a marijuana cultivation industry. *J. Quant. Criminol.* **2007**, *23*, 221–241. [CrossRef]
28. Bouchard, M.; Morselli, C.; Macdonald, M.; Gallupe, O.; Zhang, S.; Farabee, D. Estimating risks of arrest and criminal populations: Regression adjustments to capture–recapture models. *Crime Delinq.* **2019**, *65*, 1767–1797. [CrossRef]
29. Cai, T.; Xia, Y. Estimating size of drug users in Macau: An open population capture-recapture model with data augmentation using public registration data. *Asian J. Criminol.* **2018**, *13*, 193–206. [CrossRef]
30. Rossmo, D.K.; Routledge, R. Estimating the size of criminal populations. *J. Quant. Criminol.* **1990**, *6*, 293–314. [CrossRef]
31. Tajuddin, R.R.M.; Ismail, N.; Ibrahim, K. Estimating population size of criminals: A new Horvitz–Thompson estimator under one-inflated positive Poisson–Lindley model. *Crime Delinq.* **2021**, *68*, 1004–1034. [CrossRef]
32. Van Der Heijden, P.G.; Cruyff, M.; Van Houwelingen, H.C. Estimating the size of a criminal population from police records using the truncated Poisson regression model. *Stat. Neerl.* **2003**, *57*, 289–304. [CrossRef]
33. Van der Heijden, P.G.; Cruyff, M.; Böhning, D. Capture recapture to estimate criminal populations. In *Encyclopedia of Criminology and Criminal Justice*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 267–276.
34. Tajuddin, R.R.M.; Ismail, N.; Ibrahim, K. Several two-component mixture distributions for count data. *Commun. Stat. Simul. Comput.* **2022**, *51*, 3760–3771. [CrossRef]
35. Lerch, M. Note sur la function. *Acta Math.* **1887**, *11*, 19–24. [CrossRef]
36. Horvitz, D.G.; Thompson, D.J. A generalization of sampling without replacement from a finite universe. *J. Am. Stat. Assoc.* **1952**, *47*, 663–685. [CrossRef]
37. Böhning, D. A simple variance formula for population size estimators by conditioning. *Stat. Methodol.* **2008**, *5*, 410–423. [CrossRef]
38. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **1974**, *19*, 716–723. [CrossRef]
39. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **1978**, *6*, 461–464. [CrossRef]
40. Puig, P.; Barquinero, J.F. An application of compound Poisson modelling to biological dosimetry, Proceedings of the Royal Society A: Mathematical. *Phys. Eng. Sci.* **2011**, *467*, 897–910.
41. Snow, L.C.; Davies, R.H.; Christiansen, K.H.; Carrique-Mas, J.J.; Wales, A.D.; O'Connor, J.L.; Cook, A.J.; Evans, S.J. Survey of the prevalence of Salmonella species on commercial laying farms in the United Kingdom. *Vet. Rec.* **2007**, *161*, 471–476. [CrossRef]
42. Arnold, M.E.; Papadopoulou, C.; Davies, R.H.; Carrique-Mas, J.J.; Evans, S.J.; Hoinville, L.J. Estimation of Salmonella prevalence in UK egg-laying holdings. *Prev. Vet. Med.* **2010**, *94*, 306–309. [CrossRef] [PubMed]
43. Godwin, R.T.; Böhning, D. Estimation of the population size by using the one-inflated positive Poisson model. *J. R. Stat. Soc. Ser. C Appl. Stat.* **2017**, *66*, 425–448. [CrossRef]
44. Tajuddin, R.R.M.; Ismail, N.; Ibrahim, K. On variance estimation for the population size estimator under one-inflated positive Poisson distribution. *Malays. J. Fundam. Appl. Sci.* **2022**, *18*, 237–244. [CrossRef]