*Brazilian Journal of Probability and Statistics*
2020, Vol. 34, No. 1, 90–111
https://doi.org/10.1214/18-BJPS414


# On the Nielsen distribution

**Fredy Castellares[a], Artur J. Lemonte[b] and Marcos A. C. Santos[a]**

[a]*Departamento de Estatística, ICEX, Universidade Federal de Minas Gerais*
[b]*Departamento de Estatística, CCET, Universidade Federal do Rio Grande no Norte*

**Abstract.** We introduce a two-parameter discrete distribution that may have a zero vertex and can be useful for modeling overdispersion. The discrete Nielsen distribution generalizes the Fisher logarithmic (i.e., logarithmic series) and Stirling type I distributions in the sense that both can be considered displacements of the Nielsen distribution. We provide a comprehensive account of the structural properties of the new discrete distribution. We also show that the Nielsen distribution is infinitely divisible. We discuss maximum likelihood estimation of the model parameters and provide a simple method to find them numerically. The usefulness of the proposed distribution is illustrated by means of three real data sets to prove its versatility in practical applications.

## 1 Introduction

Count data occur in many practical problems as, for example, the number of occurrences of thunderstorms in a calendar year, the number of accidents, the number of absences, the number of days lost, the number of insurance claims, the number of kinds of species in ecology, and so on. Discrete distributions, which describe count phenomena, have been proposed in the statistical literature in recent years, due perhaps to advances in computational methods which enable us to compute, straightforwardly, the numerical value of special functions such as hypergeometric series. To mention a few, but not limited to, the readers are referred to Roy (2004), Inusah and Kozubowski (2006), Kozubowski and Inusah (2006), Krishna and Pundir (2009), Jazi, Lai and Alamatsaz (2010), Nooghabi, Roknabadi and Borzadaran (2011), Englehardt and Li (2011), Nekoukhou, Alamatsaz and Bidram (2013), Barbiero (2014), among many others.

It is well known that the classical Poisson distribution is applied in many scientific fields involving count data, mainly because of its simplicity. As pointed out by Gómez-Déniz, Sarabia and Calderin-Ojeda (2011), most frequencies of event occurrence can be described initially by a Poisson distribution. However, a major drawback of this distribution is the fact that the variance is restricted to be equal to the mean, a situation that may not be consistent with observation. So, alternative discrete probability distributions, such as the negative binomial distribution, are preferred for modeling the phenomena under study. Additionally, many of these phenomena, such as individual automobile insurance claims, are characterized by two features: (i) overdispersion, that is, the variance is greater than the mean; (ii) zero-inflated (or zero vertex), that is, the presence of a high percentage of zero values in the empirical distribution. In view of this, many attempts have been made in the statistical literature to propose new discrete family of distributions for the distribution of the number of counts.

It is worth emphasizing that the majority of the new discrete distributions proposed recently in the statistical literature are obtained by discretizing a known continuous distribution (see the above references). Instead, we will introduce a new two-parameter discrete distribution on the basis of a series expansion presented in Nielsen (1906). As we will see latter,

*Key words and phrases.* Discrete distribution, Fisher logarithmic distribution, Nielsen expansion, Stirling distributions.

Received December 2017; accepted August 2018.

the new distribution is very simple to deal with, since its probability mass function does not contain any complicated function. Further, it is very flexible and it also presents the twofold characteristic stated above: it can have a zero vertex, and it is overdispersed. Therefore, it may be a natural candidate for fitting phenomena of this nature. The real data examples provided here, and the comparison with the Negative Binomial (NB) distribution (the most important and popular two-parameter discrete distribution for overdispersed data), show that the proposed distribution has an outstanding performance. In addition to the NB distribution, we also consider the Zero-Inflated Poisson (ZIP) and Zero-Inflated NB (ZINB) distributions in the real data applications for the sake of comparison.

The main aim of this paper is to introduce a new two-parameter discrete family of distributions with the hope that the new distribution may have a 'better fit' compared to the NB distribution (and other ones) in certain practical situations. Additionally, we will provide a comprehensive account of the mathematical properties of the proposed new family of distributions. As we will see later, the formulas related with the new distribution are simple and manageable, and with the use of modern computer resources and its numerical capabilities, the proposed distribution may prove to be an useful addition to the arsenal of applied statisticians in discrete data analysis.

In order to introduce the new discrete distribution, we consider the following series expansion provided by Nielsen (1906)

$$\left[-\frac{\log(1-z)}{z}\right]^{\alpha} = 1 + \alpha z \sum_{n=0}^{\infty} \psi_n(n+\alpha) z^n, \qquad \alpha \in \mathbb{R}, |z| < 1, \tag{1}$$

where the coefficients $\psi_n(\cdot)$ are Stirling polynomials. According to Ward (1934), these coefficients can be expressed in the form

$$\begin{aligned}\psi_{n-1}(w) = \frac{(-1)^{n-1}}{(n+1)!}\Bigg[ & H_n^{n-1} - \frac{w+2}{n+2} H_n^{n-2} + \frac{(w+2)(w+3)}{(n+2)(n+3)} H_n^{n-3} - \cdots \\ & + (-1)^{n-1} \frac{(w+2)(w+3)\cdots(w+n)}{(n+2)(n+3)\cdots(2n)} H_n^0 \Bigg],\end{aligned} \tag{2}$$

where $H_n^m$ are positive integers defined recursively by $H_{n+1}^m = (2n+1-m)H_n^m + (n-m+1)H_n^{m-1}$, with $H_0^0 = 1$, $H_{n+1}^0 = 1 \times 3 \times 5 \times \cdots \times (2n+1)$, $H_{n+1}^n = 1$. The first six polynomials are $\psi_0(w) = 1/2$, $\psi_1(w) = (2+3w)/24$, $\psi_2(w) = (w+w^2)/48$, $\psi_3(w) = (-8-10w+15w^2+15w^3)/5760$, $\psi_4(w) = (-6w-7w^2+2w^3+3w^4)/11520$ and $\psi_5(w) = (96+140w-224w^2-315w^3+63w^5)/2{,}903{,}040$.

We would like to point out that according to another definition,[1] the polynomials $S_0(w) = 1$ and $S_n(w) = n!(w+1)\psi_{n-1}(w)$, for $n \geq 1$, are also known as Stirling polynomials. In this paper, we use this terminology to refer to the polynomials $\psi_n(w)$ in accordance with Nielsen (1906) and Ward (1934).

We have the following propositions.

**Proposition 1.** *The expansion* (1) *is absolutely convergent.*

**Proof.** The proof can be found in Flajonet and Sedgewick (2009, p. 385). □

**Proposition 2.** *The expansion* (1) *can be rewritten as*

$$1 = \frac{z^{\alpha}}{[-\log(1-z)]^{\alpha}} \sum_{m=0}^{\infty} \rho_m(\alpha) z^m, \qquad 0 < z < 1, \tag{3}$$

*where* $\rho_0(\alpha) = 1$ *and* $\rho_m(\alpha) = \alpha \psi_{m-1}(\alpha + m - 1)$ *for* $m \geq 1$.

[1] See, for example, http://mathworld.wolfram.com/StirlingPolynomial.html.

**Proof.** We can express

$$\left[-\frac{\log(1-z)}{z}\right]^{\alpha} = 1 + \alpha \sum_{n=1}^{\infty} \psi_{n-1}(n-1+\alpha) z^n$$
$$= \sum_{n=0}^{\infty} \rho_n(\alpha) z^n,$$

where $\rho_0(\alpha) = 1$, and $\rho_n(\alpha) = \alpha\psi_{n-1}(\alpha + n - 1)$ for $n \geq 1$. It then follows that

$$1 = \frac{z^{\alpha}}{[-\log(1-z)]^{\alpha}} \sum_{m=0}^{\infty} \rho_m(\alpha) z^m, \qquad 0 < z < 1.$$

□

**Proposition 3.** *The coefficients* $\rho_m(\alpha)$ *in* (3) *satisfy*

$$\rho_m(\alpha) = \alpha\psi_{m-1}(m + \alpha - 1) > 0,$$

*for* $m \geq 1$ *and* $\alpha > 0$.

**Proof.** In order to proof that $\psi_{m-1}(m + \alpha - 1) > 0$ for $m \geq 1$ and $\alpha > 0$, we shall use the representation of Stirling polynomials given by Graham, Knuth and Patashnik (1994). The Nielsen expansion in (1) can be rewritten as

$$\left[-\frac{\log(1-z)}{z}\right]^{\alpha} = 1 + \alpha \sum_{n=1}^{\infty} \sigma_n(n+\alpha) z^n, \qquad \alpha > 0, |z| < 1,$$

where the polynomials $\sigma_n(\cdot)$ are defined as (Graham, Knuth and Patashnik, 1994)

$$\sigma_n(n+\alpha) = \frac{\sum_{k=0}^{n-1} \left\langle\!\left\langle {n \atop k} \right\rangle\!\right\rangle \binom{n+k+\alpha}{2n}}{\alpha(\alpha+1)\cdots(\alpha+n)}.$$

Here, $\left\langle\!\left\langle {n \atop k} \right\rangle\!\right\rangle$ are the Eulerian numbers of the second kind, which satisfy the recurrence relation

$$\left\langle\!\!\left\langle {n \atop m} \right\rangle\!\!\right\rangle = (2n - m - 1) \left\langle\!\!\left\langle {n-1 \atop m-1} \right\rangle\!\!\right\rangle + (m+1) \left\langle\!\!\left\langle {n-1 \atop m} \right\rangle\!\!\right\rangle, \qquad n, m \in \mathbb{N},$$

with initial condition

$$\left\langle\!\!\left\langle {0 \atop 0} \right\rangle\!\!\right\rangle = 1, \qquad \left\langle\!\!\left\langle {0 \atop m} \right\rangle\!\!\right\rangle = 1, \qquad m \neq 0.$$

The Eulerian numbers of the second kind are not negative (Graham, Knuth and Patashnik, 1994, p. 271). From definition of $\sigma_n(\alpha + n)$, we have that $\sigma_n(\alpha + n) > 0$ for all $\alpha > 0$ and $n \in \mathbb{N}$. Hence, by noting that

$$\psi_{n-1}(n + \alpha - 1) = \sigma_n(n + \alpha) > 0, \qquad n \geq 1, \forall\alpha > 0,$$

the result holds. □

The rest of this paper is organized as follows. In Section 2, we introduce the new discrete distribution. Structural properties related to the new distribution are provided in Section 3. Estimation of model parameters is discussed in Section 4. Section 5 deals with applications to real data sets. The Section 6 ends up the paper with some final comments.

## 2 The new discrete distribution

By considering Propositions 2 and 3, we define the two-parameter discrete distribution named as the *discrete Nielsen* ('dN' for short) distribution. We have the following definition.

**Definition 1.** The probability mass function of $X$ with dN distribution is given by

$$\Pr(X = x) = \frac{p^{\theta+x}\rho_x(\theta)}{[-\log(1-p)]^{\theta}}, \qquad x = 0, 1, 2, \ldots, \tag{4}$$

where $p \in (0, 1)$, $\theta > 0$, $\rho_0(\theta) = 1$,

$$\rho_x(\theta) = \theta\psi_{x-1}(\theta + x - 1), \qquad x = 1, 2, \ldots,$$

and the coefficients $\psi_x(\cdot)$ are the Stirling polynomials given in (2).

If $X$ follows a dN distribution, then the notation used is $X \sim \mathrm{dN}(p, \theta)$. The dN probability mass function in (4) is very simple and does not involve any complicated function. Additionally, there is no functional relationship between the parameters $p \in (0, 1)$ and $\theta > 0$, and they vary freely in the parameter space. We have that

$$\Pr(X = 0) = \frac{p^{\theta}}{[-\log(1-p)]^{\theta}}, \qquad \Pr(X = 1) = \frac{\theta p^{\theta+1}}{2[-\log(1-p)]^{\theta}},$$

and the other probabilities can be easily computed. For fixed $p \in (0, 1)$, it follows that $\Pr(X = 0) \to 1$ as $\theta \to 0^+$, which means that the dN model can have a zero vertex. However, all zeros must be interpreted as observational zeros from the dN distribution, that is, the proposed dN model does not act a zero-inflated model. Figure 1 displays some possible shapes of the dN probability mass function given by expression (4). Note that the mode moves away from zero with increasing $\theta$, for $p$ fixed, indicating that the new discrete distribution is very versatile.

From Nielsen expansion (1) and when $\alpha = 1$, it follows that

$$\left[-\frac{\log(1-z)}{z}\right] = 1 + z\sum_{n=0}^{\infty}\psi_n(n+1)z^n, \qquad 0 < z < 1.$$

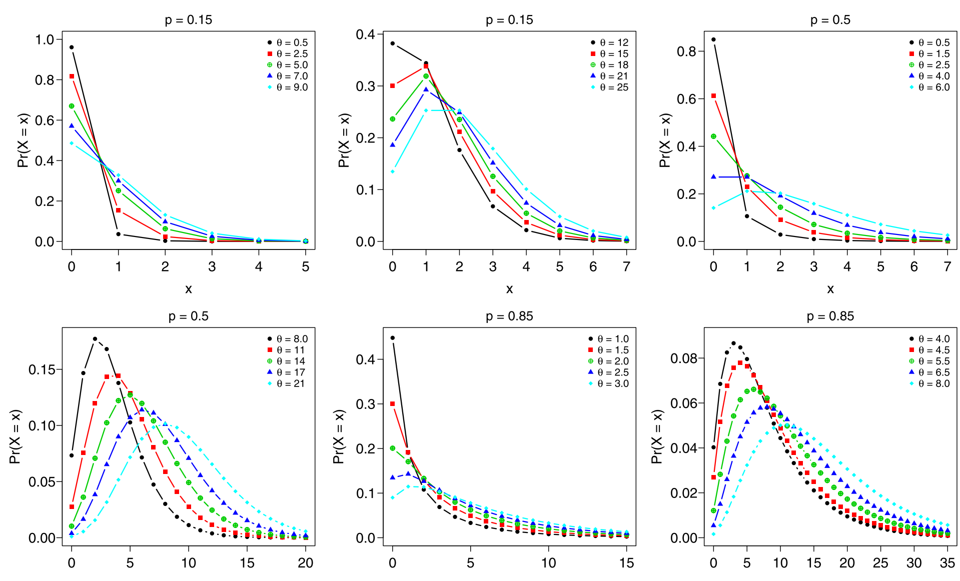


**Figure 1** *The dN probability mass functions for different values of parameters.*

Additionally, we have that

$$-\log(1-z) = z + \frac{z^2}{2} + \frac{z^3}{3} + \cdots, \qquad 0 < z < 1.$$

After some algebra, we obtain

$$\begin{aligned}\left[-\frac{\log(1-z)}{z}\right] &= 1 + \frac{z}{2} + \frac{z^2}{3} + \cdots \\ &= 1 + z\left[\frac{1}{2} + \frac{z}{3} + \cdots\right] \\ &= 1 + z\sum_{n=0}^{\infty} \frac{z^n}{n+2}.\end{aligned}$$

Since $\psi_n(n+1) = 1/(n+2)$, we have that

$$\begin{aligned}1 &= \frac{z}{[-\log(1-z)]} + \sum_{n=0}^{\infty} \frac{z^{n+2}}{(n+2)[-\log(1-z)]} \\ &= \sum_{n=1}^{\infty} \frac{z^n}{n[-\log(1-z)]}.\end{aligned}$$

Hence, for $\theta = 1$, the dN distribution can be reduced to the Fisher logarithmic distribution (Fisher, Corbet and Williams, 1943) given by

$$\Pr(X = x) = \frac{p^x}{x[-\log(1-p)]}, \qquad x = 1, 2, \ldots,$$

where $p \in (0, 1)$; that is, the Fisher logarithmic distribution can be considered a displacement of the dN distribution from 1 (one).

As noted early, the proposed dN distribution has a very simple form. Tractability of the probability mass function may be a great advantage in computing the probabilities as well as structural properties from that equation. As pointed out recently by Jones (2015), "it can certainly be argued that in an age of fast computers, mathematical tractability is not an issue of overwhelming importance. However, straightforward mathematical formulae describing features of distributions remain a springboard to insight, interpretation and clarity of exposition, as well as improving computational speed and convenience. All other things being equal—which they never are!—tractability is still to be preferred to non-tractability." In view of this, the tractability of the two-parameter dN distribution is very welcome and, as a consequence, all properties derived in the next section (Section 3) have very simple forms. Again, as pointed out recently by Jones (2015), "the role of parsimony in statistical modelling, to aid interpretation (that word again!), to facilitate estimation and particularly prediction, affording generalisability of results by avoiding overfitting, is clear. In developing families of distributions, however, the watchword has usually been flexibility, and parsimony is little mentioned." We have that the dN distribution has only two parameters ($p \in (0, 1)$ and $\theta > 0$) which has facilitated the estimation of these parameters by the maximum likelihood method (see Section 4). In short, we have proposed a very simple discrete distribution with only two parameters and which is very flexible.

## 3 Properties

In what follows, we study several structural properties of the two-parameter dN distribution. We have the following propositions.

**Proposition 4.** *Let $X \sim \mathrm{dN}(p,\theta)$. Then, the probability generating function is*

$$G_X(s) = \mathbb{E}(s^X) = \left[\frac{\log(1-ps)}{s\log(1-p)}\right]^\theta, \qquad 0 < s < \frac{1}{p}, \theta \neq 1.$$

*For $\theta = 1$, it follows that*

$$G_X(s) = \mathbb{E}(s^X) = \frac{\log(1-ps)}{\log(1-p)}, \qquad 0 < s < \frac{1}{p}.$$

**Proof.** For $\theta \neq 1$, we have that

$$G_X(s) = \mathbb{E}(s^X) = \sum_{x=0}^{\infty} s^x \frac{p^{\theta+x}\rho_x(\theta)}{[-\log(1-p)]^\theta}.$$

After some algebra, we obtain

$$G_X(s) = \left[\frac{\log(1-ps)}{s\log(1-p)}\right]^\theta \sum_{x=0}^{\infty} \frac{(sp)^{x+\theta}\rho_x(\theta)}{[-\log(1-sp)]^\theta} = \left[\frac{\log(1-ps)}{s\log(1-p)}\right]^\theta,$$

where

$$\sum_{x=0}^{\infty} \frac{(sp)^{x+\theta}\rho_x(\theta)}{[-\log(1-sp)]^\theta} = 1, \qquad 0 < s < \frac{1}{p}.$$

For $\theta = 1$, we have that

$$\begin{aligned} G_X(s) = \mathbb{E}(s^X) &= \sum_{x=1}^{\infty} s^x \frac{p^x}{x[-\log(1-p)]} \\ &= \frac{\log(1-ps)}{\log(1-p)} \sum_{x=1}^{\infty} \frac{(sp)^x}{x[-\log(1-sp)]} \\ &= \frac{\log(1-ps)}{\log(1-p)}, \end{aligned}$$

where

$$\sum_{x=1}^{\infty} \frac{(sp)^x}{x[-\log(1-sp)]} = 1, \qquad 0 < s < \frac{1}{p}.$$

□

**Proposition 5.** *Let $X \sim \mathrm{dN}(p,\theta)$. Then, the moment generating function is*

$$M_X(t) = \mathbb{E}(\mathrm{e}^{tX}) = \mathrm{e}^{-\theta t}\left[\frac{\log(1-p\mathrm{e}^t)}{\log(1-p)}\right]^\theta, \qquad t < -\log(p), \theta \neq 1.$$

*For $\theta = 1$, it follows that*

$$M_X(t) = \mathbb{E}(\mathrm{e}^{tX}) = \frac{\log(1-p\mathrm{e}^t)}{\log(1-p)}, \qquad t < -\log(p).$$

**Proof.** For $\theta \neq 1$, we have that

$$M_X(t) = \mathbb{E}(\mathrm{e}^{tX}) = \sum_{x=0}^{\infty} \mathrm{e}^{tx} \frac{p^{\theta+x}\rho_x(\theta)}{[-\log(1-p)]^\theta}.$$

After some algebra, we obtain

$$M_X(t) = \left[\frac{\log(1-p\mathrm{e}^t)}{\mathrm{e}^t\log(1-p)}\right]^\theta \sum_{x=0}^{\infty} \frac{(p\mathrm{e}^t)^{x+\theta}\rho_x(\theta)}{[-\log(1-p\mathrm{e}^t)]^\theta}$$
$$= \mathrm{e}^{-\theta t}\left[\frac{\log(1-p\mathrm{e}^t)}{\log(1-p)}\right]^\theta,$$

where

$$\sum_{x=0}^{\infty} \frac{(p\mathrm{e}^t)^{x+\theta}\rho_x(\theta)}{[-\log(1-p\mathrm{e}^t)]^\theta} = 1, \qquad t < -\log(p).$$

For $\theta = 1$, we have

$$M_X(t) = \mathbb{E}(\mathrm{e}^{tX}) = \sum_{x=1}^{\infty} \mathrm{e}^{tx}\frac{p^x}{x[-\log(1-p)]}$$
$$= \frac{\log(1-p\mathrm{e}^t)}{\log(1-p)}\sum_{x=1}^{\infty}\frac{(p\mathrm{e}^t)^x}{x[-\log(1-p\mathrm{e}^t)]}$$
$$= \frac{\log(1-p\mathrm{e}^t)}{\log(1-p)},$$

where

$$\sum_{x=1}^{\infty}\frac{(p\mathrm{e}^t)^x}{x[-\log(1-p\mathrm{e}^t)]} = 1, \qquad t < -\log(p).$$ □

**Proposition 6.** *Let* $X \sim \mathrm{dN}(p,\theta)$. *Then, the cumulant generating function is*

$$K_X(t) = \log\big[M_X(t)\big] = -\theta t + \theta\big\{\log\big[-\log(1-p\mathrm{e}^t)\big] - \log\big[-\log(1-p)\big]\big\},$$

*where* $t < -\log(p)$ *and* $\theta \neq 1$. *For* $\theta = 1$, *it follows that*

$$K_X(t) = \log\big[M_X(t)\big] = \log\big[-\log(1-p\mathrm{e}^t)\big] - \log\big[-\log(1-p)\big],$$

*where* $t < -\log(p)$.

**Proof.** The result follows directly from Proposition 5. □

**Proposition 7.** *Let* $X \sim \mathrm{dN}(p,\theta)$. *Then, the characteristic function is given by*

$$\phi_X(t) = \mathbb{E}(\mathrm{e}^{\mathrm{i}tX}) = \mathrm{e}^{-\mathrm{i}\theta t}\left[\frac{\log(1-p\mathrm{e}^{\mathrm{i}t})}{\log(1-p)}\right]^\theta, \qquad t\in\mathbb{R}, \theta\neq 1.$$

*For* $\theta = 1$, *it follows that*

$$\phi_X(t) = \mathbb{E}(e^{\mathrm{i}tX}) = \frac{\log(1-p\mathrm{e}^{\mathrm{i}t})}{\log(1-p)}, \qquad t\in\mathbb{R}.$$

*where* $\mathrm{i} = \sqrt{-1}$ *is the imaginary number.*

**Proof.** The proof is similar to that of Proposition 5 just considering the logarithm function for complex variables. □

It follows from Proposition 5 that the ordinary moments of $X \sim \mathrm{dN}(p, \theta)$, for $\theta \neq 1$, are given by

$$\mu_r' = \mathbb{E}(X^r) = \left\{ \frac{d^r}{dt^r} \mathrm{e}^{-\theta t} \left[ \frac{\log(1 - p\mathrm{e}^t)}{\log(1 - p)} \right]^\theta \right\}_{t=0}.$$

For example, the mean (i.e., $\mu_1'$) and variance are

$$\mathbb{E}(X) = \theta \left( \frac{p}{(1-p)[-\log(1-p)]} - 1 \right),$$

$$\mathbb{VAR}(X) = \theta \frac{p[-\log(1-p) - p]}{[(1-p)\log(1-p)]^2}.$$

We have that the mean and variance increase as the values of $p \in (0, 1)$ and $\theta > 0$ increase. Table 1 lists the values of the mean and variance with varying values of $p$ and $\theta$. The skewness and kurtosis of $X$ can be calculated from the ordinary moments using well-known relationships. Figure 2 shows how these measures vary with respect to $p$ and $\theta$. Note that the skewness and kurtosis of the dN distribution can be quite pronounced, and the values of both measures decrease as the values of $p \in (0, 1)$ and $\theta > 0$ increase.

**Table 1** *Mean (above) and variance (below) of the dN distribution*

| $\theta \setminus p$ | 0.1 | 0.3 | 0.5 | 0.7 | 0.9 |
|---|---|---|---|---|---|
| 0.2 | 0.011 | 0.040 | 0.089 | 0.188 | 0.582 |
| | 0.012 | 0.055 | 0.161 | 0.541 | 4.762 |
| 0.5 | 0.027 | 0.101 | 0.221 | 0.469 | 1.454 |
| | 0.030 | 0.136 | 0.402 | 1.352 | 11.904 |
| 0.8 | 0.044 | 0.161 | 0.354 | 0.750 | 2.327 |
| | 0.048 | 0.218 | 0.643 | 2.163 | 19.047 |
| 1.2 | 0.065 | 0.242 | 0.531 | 1.126 | 3.490 |
| | 0.072 | 0.327 | 0.965 | 3.245 | 28.571 |
| 1.6 | 0.087 | 0.323 | 0.708 | 1.501 | 4.654 |
| | 0.095 | 0.436 | 1.286 | 4.327 | 38.094 |
| 2.5 | 0.136 | 0.504 | 1.107 | 2.345 | 7.272 |
| | 0.149 | 0.682 | 2.010 | 6.760 | 59.522 |
| 4.0 | 0.218 | 0.806 | 1.771 | 3.752 | 11.635 |
| | 0.238 | 1.091 | 3.216 | 10.817 | 95.236 |

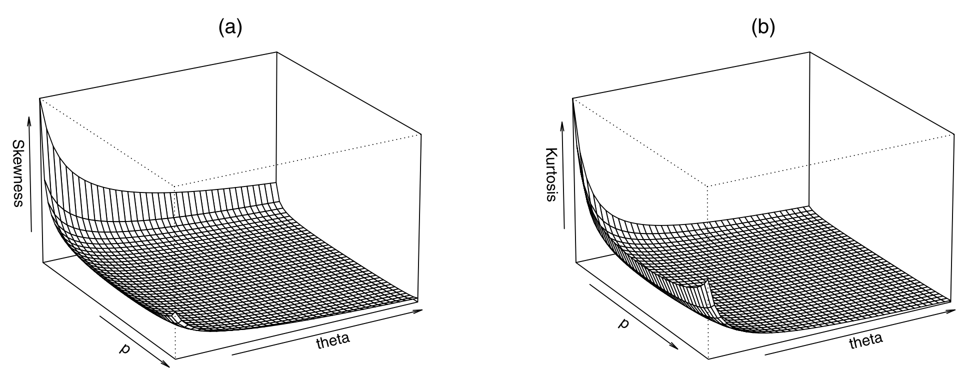


**Figure 2** *Skewness* (a) *and kurtosis* (b) *of the dN distribution as functions of $p$ and $\theta$.*

The index of dispersion (a normalized measure of the dispersion of a probability distribution) of the dN distribution, defined as $I_d = \mathbb{VAR}(X)/\mathbb{E}(X)$, takes the form

$$I_d = \frac{p[-\log(1-p)-p]}{[(1-p)\log(1-p)]^2}\left(\frac{p}{(1-p)[-\log(1-p)]}-1\right)^{-1},$$

which is independent of the parameter $\theta$. It follows that $I_d > 1$ for $p \in (0,1)$, and $\lim_{p\to 0^+} I_d = 1$ and $\lim_{p\to 1^-} I_d = \infty$, which implies that the dN distribution is suitable for modeling count data with overdispersion, like the NB distribution.

**Proposition 8.** *Let $X_1 \sim \mathrm{dN}(p,\theta_1)$ and $X_2 \sim \mathrm{dN}(p,\theta_2)$ be two independent random variables. Define $Z = X_1 + X_2$. The probability mass function (convolution) of $Z$ is given by*

$$\Pr(Z=z) = \frac{p^{\theta_1+\theta_2+z}\rho_z(\theta_1+\theta_2)}{[-\log(1-p)]^{\theta_1+\theta_2}}, \qquad z = 0,1,2,\ldots,$$

*where $p \in (0,1)$, $\theta_1 > 0$ (with $\theta_1 \neq 1$), $\theta_2 > 0$ (with $\theta_2 \neq 1$), $\rho_0(\theta_1+\theta_2) = 1$, and*

$$\rho_z(\theta_1+\theta_2) = (\theta_1+\theta_2)\psi_{z-1}(\theta_1+\theta_2+z-1), \qquad z = 1,2,\ldots.$$

*For $\theta_1 = \theta_2 = 1$, we have that*

$$\Pr(Z=z) = \frac{2!p^z}{z![-\log(1-p)]^2}\left\lfloor {z \atop 2} \right\rfloor, \qquad z = 2,3,\ldots,$$

*where $\left\lfloor {n \atop k} \right\rfloor$ (for $n \in \mathbb{N}$ and $k \in \mathbb{N}$) are Stirling numbers of the first kind, and they can be calculated by the recurrence relation*

$$\left\lfloor {n+1 \atop k} \right\rfloor = n\left\lfloor {n \atop k} \right\rfloor + \left\lfloor {n \atop k-1} \right\rfloor,$$

*with the initial conditions*

$$\left\lfloor {0 \atop 0} \right\rfloor = 1, \qquad \left\lfloor {n \atop 0} \right\rfloor = \left\lfloor {0 \atop n} \right\rfloor = 0.$$

**Proof.** By using $\Pr(Z=z) = \sum_{x=0}^{z}\Pr(X_1=x)\Pr(X_2=z-x)$ for $\theta_1 \neq 1$ and $\theta_2 \neq 1$, we obtain

$$\Pr(Z=z) = \sum_{x=0}^{z}\left[\frac{p}{[-\log(1-p)]}\right]^{\theta_1} p^x\rho_x(\theta_1)\left[\frac{p}{[-\log(1-p)]}\right]^{\theta_2} p^{z-x}\rho_{z-x}(\theta_2).$$

After some algebra, we have

$$\Pr(Z=z) = \left[\frac{p}{[-\log(1-p)]}\right]^{\theta_1+\theta_2} p^z \sum_{x=0}^{z}\rho_x(\theta_1)\rho_{z-x}(\theta_2).$$

The convolution of Stirling polynomials has the form (Graham, Knuth and Patashnik, 1994, Cap. 6)

$$\sum_{k=0}^{n} r\sigma_k(k+r)s\sigma_{n-k}\big(s+[n-k]\big) = (r+s)\sigma_n\big(n+[r+s]\big).$$

From this expression, we have

$$\sum_{x=0}^{z}\theta_1\sigma_x(x+\theta_1)\theta_2\sigma_{z-x}\big(\theta_2+[z-x]\big) = (\theta_1+\theta_2)\sigma_z\big(z+[\theta_1+\theta_2]\big) = \rho_z(\theta_1+\theta_2),$$

by using $\rho_0(\alpha) =: \alpha\sigma_0(\alpha) = 1$, and $\rho_x(\alpha) = \alpha\psi_{x-1}(\alpha + x - 1) = \alpha\sigma_x(x+\alpha)$ with $\alpha > 0$ and $x = 1, 2, \ldots$. Hence, it follows that

$$\Pr(Z = z) = \left[\frac{p}{[-\log(1-p)]}\right]^{\theta_1+\theta_2} \rho_z(\theta_1+\theta_2) p^z, \qquad z = 0, 1, 2, \ldots.$$

For $\theta_1 = \theta_2 = 1$, we have that

$$\Pr(Z = z) = \sum_{x=0}^{z} \left[\frac{1}{[-\log(1-p)]}\right] \frac{p^x}{x} \left[\frac{1}{[-\log(1-p)]}\right] \frac{p^{z-x}}{(z-x)}.$$

After some algebra, we obtain

$$\begin{aligned}\Pr(Z = z) &= \left[\frac{1}{[-\log(1-p)]}\right]^2 \frac{p^z}{z} \sum_{x=1}^{z-1} \frac{z}{x(z-x)} \\ &= \left[\frac{1}{[-\log(1-p)]}\right]^2 p^z \frac{2}{z} \sum_{x=1}^{z-1} \frac{1}{x}.\end{aligned}$$

The $n$th partial sum of the divergent harmonic series, given by

$$M_n = \sum_{k=1}^{n} \frac{1}{k},$$

is called the $n$th harmonic number. We have that

$$\left\lfloor \begin{matrix} n \\ 2 \end{matrix} \right\rfloor = (n-1)! M_{n-1},$$

and hence

$$\frac{2}{z} \sum_{x=1}^{z-1} \frac{1}{x} = \frac{2!}{z!}(z-1)! M_{z-1} = \frac{2!}{z!} \left\lfloor \begin{matrix} n \\ 2 \end{matrix} \right\rfloor.$$

From the above expression, we obtain

$$\Pr(Z = z) = \left[\frac{1}{[-\log(1-p)]}\right]^2 \frac{2!}{z!} \left\lfloor \begin{matrix} z \\ 2 \end{matrix} \right\rfloor p^z, \qquad z = 2, 3, 4 \ldots,$$

which concludes the proof. □

The generalization of Proposition 8 is provided in the following proposition.

**Proposition 9.** *Let $X_1, X_2, \ldots, X_n$ be $n$ independent random variables, where $X_k \sim \mathrm{dN}(p, \theta_k)$ for $k = 1, 2, \ldots, n$. Define $Z = X_1 + \cdots + X_n$. If $\theta_k \neq 1$ for $k = 1, 2, \ldots, n$, then*

$$Z \sim \mathrm{dN}(p, \theta_1 + \cdots + \theta_n).$$

*If $\theta_k = 1$ for $k = 1, 2, \ldots, n$, then*

$$\Pr(Z = z) = \frac{p^z}{[-\log(1-p)]^n} \frac{n!}{z!} \left\lfloor \begin{matrix} z \\ n \end{matrix} \right\rfloor, \qquad z = n, n+1, n+2, \ldots.$$

**Proof.** The case $\theta_k = 1$ for $k = 1, 2, \ldots, n$ can be found in Patil (1963). If $\theta_k \neq 1$ for $k = 1, \ldots, n$, we consider the inversion theorem (Feller, 1971, p. 511). The characteristic function of $Z$ is given by

$$\phi_Z(t) = \mathbb{E}(\mathrm{e}^{\mathrm{i}tZ}) = \prod_{k=1}^{n} \phi_{X_k}(t).$$

Additionally, we have that

$$\phi_{X_k}(t) = \mathbb{E}(\mathrm{e}^{\mathrm{i}tX_k}) = \mathrm{e}^{-\mathrm{i}t\theta_k}\left[\frac{\log(1-p\mathrm{e}^{\mathrm{i}t})}{\log(1-p)}\right]^{\theta_k}, \qquad t \in \mathbb{R}, \theta_k \neq 1,$$

and hence we obtain

$$\phi_Z(t) = \left[\frac{p}{\log(1-p)}\right]^{\theta_1+\cdots+\theta_n}\left[\frac{\log(1-p\mathrm{e}^{\mathrm{i}t})}{p\mathrm{e}^{\mathrm{i}t}}\right]^{\theta_1+\cdots+\theta_n}.$$

Let $\theta = \theta_1 + \cdots + \theta_n$. We have

$$\phi_Z(t) = \left[\frac{p}{\log(1-p)}\right]^{\theta}\sum_{m=0}^{\infty}\rho_m(\theta)p^m\mathrm{e}^{\mathrm{i}tm}, \qquad 0 < p < 1.$$

By using the inversion theorem, there exists a probability function of the form

$$\Pr(Z = z) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\phi_Z(t)\mathrm{e}^{-\mathrm{i}zt}\,dt,$$

which is given by

$$\Pr(Z = z) = \left[\frac{p}{\log(1-p)}\right]^{\theta}\sum_{m=0}^{\infty}\rho_m(\theta)p^m\frac{1}{2\pi}\int_{-\pi}^{\pi}\mathrm{e}^{\mathrm{i}(m-z)t}\,dt,$$

where

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\mathrm{e}^{\mathrm{i}(m-z)t}\,dt = \begin{cases}1, & m = z,\\ 0, & m \neq z.\end{cases}$$

Hence, it follows that

$$\Pr(Z = z) = \left[\frac{p}{\log(1-p)}\right]^{\theta}\rho_z(\theta)p^z, \qquad z = 0, 1, \ldots,$$

which concludes the proof. □

**Proposition 10.** *Let $X \sim \mathrm{dN}(p, \theta)$, where $p \in (0, 1)$ and $\theta = n \in \mathbb{N}$. Then, the probability mass function of $X$ takes the form*

$$\Pr(X = x) = \frac{p^{n+x}}{[-\log(1-p)]^n}\frac{n!}{(n+x)!}\left[\begin{matrix}x+n\\ n\end{matrix}\right], \qquad x = 0, 1, 2, \ldots. \tag{5}$$

**Proof.** From Graham, Knuth and Patashnik (1994, Cap. 7), we have that

$$\left[\frac{-\log(1-p)}{p}\right]^n = 1 + \sum_{x=1}^{\infty}\frac{n!}{(n+x)!}\left[\begin{matrix}x+n\\ n\end{matrix}\right]p^x.$$

Additionally, for $x = 0, 1, 2, \ldots$ and $n = 1, 2, \ldots$, it follows that

$$\psi_x(x+n) = \frac{(n-1)!}{(n+1+x)!}\left[\begin{matrix}n+1+x\\ n\end{matrix}\right],$$

which completes the proof. □

Let $Z$ be a random variable with probability mass function given by

$$\Pr(Z = z) = \frac{p^z}{[-\log(1-p)]^n}\frac{n!}{z!}\left[\begin{matrix}z\\ n\end{matrix}\right], \qquad z = n, n+1, n+2, \ldots.$$

We will name the above distribution as the discrete Stirling type I (‘dS’ for short) distribution with parameters $p \in (0, 1)$ and $n \in \mathbb{N}$, say $Z \sim \mathrm{dS}(p, n)$. The dS distribution appeared for the first time in Patil (1963) as a convolution of $n$ random variables with Fisher logarithmic distribution; see also Patil and Wani (1965). We have the following proposition.

**Proposition 11.** *Let $X \sim \mathrm{dN}(p, n)$ and $Z \sim \mathrm{dS}(p, n)$, where $n \in \mathbb{N}$ and $0 < p < 1$. Then, the dS distribution is a displacement of the dN distribution from $n \in \mathbb{N}$ and, additionally, is fulfilled $Z = X + n$.*

**Proof.** The proof follows from equation (5) and by defining $Z = X + n$. □

Finally, we have the following proposition.

**Proposition 12.** *Let $X \sim \mathrm{dN}(p, \theta)$, where $p \in (0, 1)$ and $\theta > 0$ with $\theta \neq 1$. Then, the dN distribution is infinitely divisible.*

**Proof.** Let $X_{k,n} \sim \mathrm{dN}(p, \theta/n)$ be independent and identically distributed random variables for all $k$, and $n$ fixed. Define $X = X_{1,n} + X_{2,n} + \cdots + X_{n,n}$. Then, we have that

$$\phi_X(t) = \prod_{k=1}^{n} \phi_{X_{k,n}}(t) = [\phi_{X_{1,n}}(t)]^n.$$

It then follows that

$$\phi_X(t) = \left\{ \mathrm{e}^{-\mathrm{i}t\theta/n} \left[ \frac{\log(1 - p\mathrm{e}^{\mathrm{i}t})}{\log(1 - p)} \right]^{\theta/n} \right\}^n = \mathrm{e}^{-\mathrm{i}t\theta} \left[ \frac{\log(1 - p\mathrm{e}^{\mathrm{i}t})}{\log(1 - p)} \right]^{\theta},$$

which concludes the proof. □

The infinitely divisible distribution plays an important role in many areas of statistics, for example, in stochastic processes and in actuarial statistics. When a distribution $G$ is infinitely divisible, then for any integer $j \geq 2$, there exists a distribution $G_j$ such that $G$ is the $j$-fold convolution of $G_j$, namely, $G = G_j^{*j}$. Additionally, since the new two-parameter dN distribution is infinitely divisible, an upper bound for its variance can be obtained when $\theta \neq 1$, which is given by

$$\mathbb{VAR}(X) \geq \frac{\Pr(X = 1)}{\Pr(X = 0)} = \frac{p\theta}{2};$$

see, for example, Johnson and Kotz (1982, p. 75).

## 4 Parameter estimation

In the following, we address the problem of estimating the dN parameters. We consider the maximum likelihood (ML) method to estimate the unknown parameters $p \in (0, 1)$ and $\theta > 0$. Let $x_1, \ldots, x_n$ be a sample of size $n$ obtained from $X \sim \mathrm{dN}(p, \theta)$. The log-likelihood function for the model parameters can be expressed as

$$\ell(p, \theta) = n\theta \log\left[ \frac{p}{-\log(1 - p)} \right] + \log(p) \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \log[\rho_{x_i}(\theta)], \tag{6}$$

where

$$\rho_{x_i}(\theta) = \begin{cases} 1, & x_i = 0, \\ \theta \psi_{x_i - 1}(\theta + x_i - 1), & x_i = 1, 2, \ldots. \end{cases}$$

The ML estimates $\widehat{p}$ and $\widehat{\theta}$ of $p$ and $\theta$, respectively, can be obtained by maximizing the log-likelihood function $\ell(p,\theta)$ with respect to $p$ and $\theta$. However, we can show from the likelihood equations that, for given $p$, the ML estimate of $\theta$ becomes

$$\widehat{\theta}(p) = -\frac{\bar{x}}{p}\left[\frac{1}{p} + \frac{(1-p)^{-1}}{\log(1-p)}\right]^{-1}, \tag{7}$$

where $\bar{x} = n^{-1}\sum_{i=1}^{n} x_i$. By replacing $\theta$ by $\widehat{\theta}(p)$ in the log-likelihood function in (6), we obtain the profile log-likelihood function for $p$ as

$$\ell^*(p) = n\widehat{\theta}(p)\log\left[\frac{p}{-\log(1-p)}\right] + \log(p)\sum_{i=1}^{n} x_i + \sum_{i=1}^{n}\log\left[\rho_{x_i}(\widehat{\theta}(p))\right].$$

Let $\Lambda_n(p)$ be the geometric mean given by

$$\Lambda_n(p) = \left(\prod_{i=1}^{n}\frac{p^{\widehat{\theta}(p)+x_i}\rho_{x_i}(\widehat{\theta}(p))}{[-\log(1-p)]^{\widehat{\theta}(p)}}\right)^{1/n}.$$

Hence, the profile log-likelihood function for $p$ reduces simply to

$$\ell^*(p) = n\log\left[\Lambda_n(p)\right]. \tag{8}$$

The profile log-likelihood function $\ell^*(p)$ in equation (8) plotted against $p$ for a trial series of values determines numerically the value of the ML estimate of $p$ which maximizes (8). We only need to find the value $\widehat{p}$ such that

$$\widehat{p} = \arg\max_{p}\{\Lambda_n(p)\}, \qquad p \in (0, 1).$$

Once the ML estimate $\widehat{p}$ is obtained from the plot, it can be substituted into equation (7) to produce the unrestricted ML estimate $\widehat{\theta} = \widehat{\theta}(\widehat{p})$. It should be mentioned that the above procedure is very simple to deal with and therefore it can be easily considered in any statistical computing program.

Since the new parametric dN model corresponds to a regular ML problem, we have that the standard asymptotics apply; that is, the ML estimators of the model parameters are asymptotically normal, asymptotically unbiased and have asymptotic variance-covariance matrix given by the inverse of the expected Fisher information matrix. Let $\boldsymbol{K}(p,\theta)$ be the unit (per observation) expected Fisher information matrix for the parameter vector $(p,\theta)$. So, when $n$ is large and under some mild regularity conditions, we have that

$$\sqrt{n}\begin{pmatrix}\widehat{p}-p\\ \widehat{\theta}-\theta\end{pmatrix} \overset{a}{\sim} \mathcal{N}_2\left(\begin{pmatrix}0\\0\end{pmatrix}, \boldsymbol{K}(p,\theta)^{-1}\right),$$

where "$\overset{a}{\sim}$" means approximately distributed, and $\boldsymbol{K}(p,\theta)^{-1}$ is the inverse of $\boldsymbol{K}(p,\theta)$. Unfortunately, there is no closed-form expression for the matrix $\boldsymbol{K}(p,\theta)$. However, the asymptotic behavior remains valid if the expected information matrix $\boldsymbol{K}(p,\theta)$ is approximated by the average matrix evaluated at $(\widehat{p},\widehat{\theta})$, say $n^{-1}\boldsymbol{J}_n(\widehat{p},\widehat{\theta})$, where $\boldsymbol{J}_n(p,\theta)$ is the observed Fisher information matrix. So, it is useful to obtain an expression for $\boldsymbol{J}_n(p,\theta)$, which can be used to obtain asymptotic standard errors for the ML estimates. We have that

$$\boldsymbol{J}_n(p,\theta) = \begin{bmatrix} J_{pp} & J_{p\theta}\\ J_{p\theta} & J_{\theta\theta}\end{bmatrix},$$

whose elements are provided in Appendix A. The above asymptotic normal distribution can be used to construct approximate confidence intervals for the parameters; that is, we have the asymptotic confidence intervals $\widehat{p} \pm \Phi^{-1}(1-\alpha/2)\,\mathrm{se}(\widehat{p})$ and $\widehat{\theta} \pm \Phi^{-1}(1-\alpha/2)\,\mathrm{se}(\widehat{\theta})$

**Table 2** *Descriptive statistics*

| | Automobile claim | Accident proneness | Chromatid aberrations |
|---|---|---|---|
| $n$ | 4000 | 165 | 400 |
| Mean | 0.0865 | 1.3450 | 0.5475 |
| Variance | 0.1225 | 4.3958 | 1.1256 |
| Skewness | 5.3180 | 2.9674 | 3.1222 |
| Kurtosis | 41.007 | 15.243 | 15.683 |
| CV | 4.0470 | 1.5636 | 1.9378 |
| ID | 1.4164 | 3.2672 | 2.0507 |

CV: Coefficient of variation ($= s/\bar{x}$); ID: Index of dispersion ($= s^2/\bar{x}$).

for $p$ and $\theta$, respectively, both with asymptotic coverage of $100(1-\alpha)\%$. Here, se$(\cdot)$ is the square root of the diagonal element of $\boldsymbol{J}_n(\widehat{p}, \widehat{\theta})^{-1}$ corresponding to each parameter (i.e., the asymptotic standard error), and $\Phi^{-1}(\cdot)$ denotes the standard normal quantile function.

Next, we conduct some Monte Carlo simulation experiments to evaluate the performance of the ML estimators $\widehat{p}$ and $\widehat{\theta}$ in estimating $p$ and $\theta$, respectively. In order to generate random values from $X \sim \mathrm{dN}(p, \theta)$, the usual method for discrete distributions can be used (see, for example, Ross, 2013, Ch. 4); that is, generate $u \sim \mathcal{U}(0, 1)$ and set $X = j$ if $\sum_{k=0}^{j} P_k < u < \sum_{k=0}^{j+1} P_k$, where $j = 0, 1, 2, \ldots$, and $P_k = \Pr(X = k)$ is the probability mass function given in (4). The simulation was performed using the R program (R Core Team, 2016), and the number of Monte Carlo replications was 10,000. The evaluation of point estimation was performed based on the following quantities for each sample size: the mean, the bias and the root mean squared error (RMSE), which are computed from 10,000 Monte Carlo replications. We also consider the coverage probability (CP) of the 90% and 95% intervals of the dN model parameters. We set the sample size at $n = 150$, 250 and 400, and consider $p = 0.4$ and 0.7, and $\theta = 1.5$, 2.5, 3.5 and 5.0. The simulation results are provided in the Appendix B. These results reveal interesting information. The ML estimators $\widehat{p}$ and $\widehat{\theta}$ have negative and positive biases, respectively, in all cases considered; that is, it seems that the parameters $p$ and $\theta$ are underestimated and overestimated, respectively. However, the ML estimates are stable and, in general, are close to the true values of the parameters for the sample sizes considered. Additionally, as the sample size increases, the bias and RMSE decrease, as expected. Regarding interval estimation, it is clear that the asymptotic CIs for the dN model parameters have very good empirical coverages, presenting CP near the respective nominal levels in all cases.

## 5 Empirical illustrations

We illustrate the usefulness of the two-parameter dN distribution by considering three real data sets. All computations were done using the R program, which is a free software environment for statistical computing and graphics. The first data set corresponds to the number of automobile insurance claims per policy over a fixed period of time (Gossiaux and Lemaire, 1981); the second data set represents the number of accidents of workers in a particular division of a large steel corporation in an observational period of six months (Sichel, 1951); and the third data set represents the number of chromatid aberrations in 24 hours (Catcheside, Lea and Thoday, 1946a, 1946b). Table 2 gives a descriptive summary for the data sets. From this table, we have that the sample index of dispersion is greater than 1, which indicates that the dN distribution may be suitable to fit these data sets.

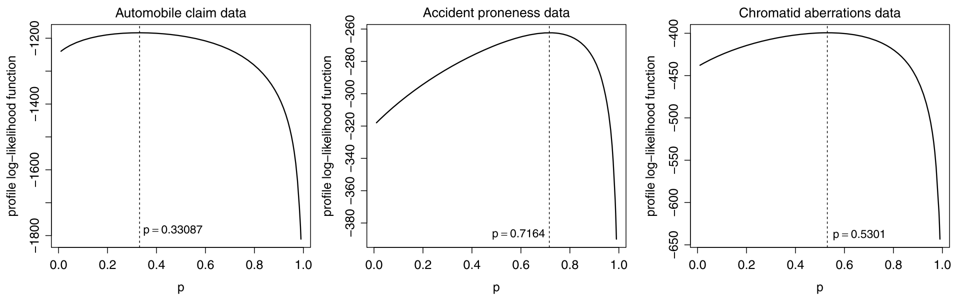


**Figure 3** *The profile log-likelihood curve for $p$.*

**Table 3** *Parameter estimates; dN model*

| | Automobile claim | | |
|---|---|---|---|
| Parameter | ML estimate | SE | 90% CI |
| $p$ | 0.3309 | 0.0416 | (0.2627; 0.3991) |
| $\theta$ | 0.3749 | 0.0618 | (0.2735; 0.4763) |
| | Accident proneness | | |
| Parameter | ML estimate | SE | 90% CI |
| $p$ | 0.7164 | 0.0532 | (0.6292; 0.8035) |
| $\theta$ | 1.3394 | 0.2580 | (0.9166; 1.7627) |
| | Chromatid aberrations | | |
| Parameter | ML estimate | SE | 90% CI |
| $p$ | 0.5301 | 0.0601 | (0.4315; 0.6286) |
| $\theta$ | 1.1089 | 0.2179 | (0.7517; 1.4665) |

The parameter $p$ of the dN model was estimated using the profile log-likelihood function in (8). Figure 3 displays the profile log-likelihood curves plotted against the parameter $p \in (0, 1)$, where their respective maximum occur near $p = 0.33087$ for the automobile claim data, near $p = 0.71640$ for the accident proneness data, and near $p = 0.53010$ for the chromatid aberrations data. Table 3 lists the ML estimates, asymptotic standard errors (SE), and the 90% confindence intervals (CI) for the model parameters.

Table 4 lists the observed and expected frequencies, log-likelihood function values evaluated at the ML estimates, Pearson goodness-of-fit chi-squared statistics ($\chi^2$) and the corresponding $p$-values. From the values of this table we have that the dN distribution seems to give a satisfactory fit on the basis of the $\chi^2$ statistics and the corresponding $p$-values. It worth emphasizing that some classes were combined in the calculation of the Pearson statistic. Groupings were done in order that the expected frequencies are large so that the chi-squared approximation to the Pearson statistic is tenable.

From Table 2, we have that the sample index of dispersion is greater than 1 (one) for the data sets, which indicates some evidence of overdispersion. Undoubtedly, the most useful and important two-parameter distribution for modeling count data with overdispersion is the NB distribution. Hence, the natural question is how the NB distribution fits these data. The NB probability mass function, specified in terms of its mean, $\mu$ say, is given by

$$\Pr(Y = y) = \left(\frac{\phi}{\phi + \mu}\right)^{\phi} \left(\frac{\mu}{\phi + \mu}\right)^{y} \frac{\Gamma(y + \phi)}{\Gamma(\phi)\Gamma(y + 1)}, \qquad y = 0, 1, 2, \ldots,$$

**Table 4** *Fit of the data sets*; *dN model*

| Automobile claim | | | Accident proneness | | | Chromatid aberrations | | |
|---|---|---|---|---|---|---|---|---|
| Count | Observed | Expected | Count | Observed | Expected | Count | Observed | Expected |
| 0 | 3719 | 3719.17 | 0 | 77 | 77.44 | 0 | 268 | 270.14 |
| 1 | 232 | 230.67 | 1 | 36 | 37.15 | 1 | 87 | 79.40 |
| 2 | 38 | 38.95 | 2 | 24 | 20.00 | 2 | 26 | 29.21 |
| 3 | 7 | 8.43 | 3 | 13 | 11.51 | 3 | 9 | 11.88 |
| 4 | 3 | 2.05 | 4 | 4 | 6.91 | 4 | 4 | 5.11 |
| 5 | 1 | 0.53 | 5 | 3 | 4.27 | 5 | 2 | 2.28 |
| | | | 6 | 2 | 2.69 | 6 | 1 | 1.05 |
| | | | 7 | 1 | 1.72 | 7 | 3 | 0.49 |
| | | | 8 | 2 | 1.12 | | | |
| | | | 9 | 2 | 0.73 | | | |
| | | | 10 | 0 | 0.48 | | | |
| | | | 15 | 1 | 0.07 | | | |
| Maximum log-likelihood | | −1183.432 | | | −262.30 | | | −399.41 |
| $\chi^2$ | | 1.055 | | | 2.257 | | | 1.924 |
| Degrees of freedom | | 2 | | | 3 | | | 2 |
| $p$-value | | 0.590 | | | 0.521 | | | 0.382 |

**Table 5** *Fit of the data sets*; *NB model*

| Automobile claim | | | Accident proneness | | | Chromatid aberrations | | |
|---|---|---|---|---|---|---|---|---|
| Count | Observed | Expected | Count | Observed | Expected | Count | Observed | Expected |
| 0 | 3719 | 3719.22 | 0 | 77 | 77.94 | 0 | 268 | 270.18 |
| 1 | 232 | 229.90 | 1 | 36 | 35.84 | 1 | 87 | 78.55 |
| 2 | 38 | 39.91 | 2 | 24 | 20.04 | 2 | 26 | 29.84 |
| 3 | 7 | 8.42 | 3 | 13 | 11.86 | 3 | 9 | 12.22 |
| 4 | 3 | 1.93 | 4 | 4 | 7.22 | 4 | 4 | 5.19 |
| 5 | 1 | 0.46 | 5 | 3 | 4.47 | 5 | 2 | 2.25 |
| | | | 6 | 2 | 2.79 | 6 | 1 | 0.99 |
| | | | 7 | 1 | 1.76 | 7 | 3 | 0.44 |
| | | | 8 | 2 | 1.11 | | | |
| | | | 9 | 2 | 0.71 | | | |
| | | | 10 | 0 | 0.45 | | | |
| | | | 15 | 1 | 0.05 | | | |
| Maximum log-likelihood | | −1183.55 | | | −262.60 | | | −399.86 |
| $\widehat{\mu}$ | | 0.0865 | | | 1.3455 | | | 0.5475 |
| | | (0.0260) | | | (0.3869) | | | (0.1539) |
| $\widehat{\phi}$ | | 0.2166 | | | 0.6986 | | | 0.6200 |
| | | (0.0364) | | | (0.1430) | | | (0.1270) |
| $\chi^2$ | | 1.4168 | | | 2.350 | | | 2.4159 |
| Degrees of freedom | | 2 | | | 3 | | | 2 |
| $p$-value | | 0.492 | | | 0.503 | | | 0.299 |

where $\mu > 0$ and $\phi > 0$. It can be shown that the variance can be written as $\mu + \mu^2/\phi$ and hence the parameter $\phi$ is referred to as the "dispersion parameter". Table 5 presents the observed and expected frequencies, log-likelihood function values evaluated at the ML estimates, ML estimates, asymptotic SEs (between parentheses), $\chi^2$ statistics and the corresponding $p$-values. From this table we have that the NB distribution provides a good fit for these data sets on the basis of the $\chi^2$ statistics and the corresponding $p$-values. However, by comparing the Tables 4 and 5, we may conclude that the new two-parameter dN distribution is slightly better than the NB distribution for modeling these data sets; that is, the dN distribution provides a slight improvement over to the BN distribution to fit these data sets.

It is also interesting to consider some zero-inflated models to fit the data sets. In short, these models are designed to deal with situations where there is an "excessive" number of individuals with a count of 0 (zero). On this regard, we shall consider the ZIP and ZINB distributions. The ZIP probability mass function is given by

$$\Pr(Y=y)=\begin{cases}\omega+(1-\omega)\mathrm{e}^{-\lambda}, & y=0,\\ (1-\omega)\dfrac{\mathrm{e}^{-\lambda}\lambda^{y}}{y!}, & y=1,2,\ldots,\end{cases}$$

where $\lambda > 0$, whereas the ZINB probability mass function takes the form

$$\Pr(Y=y)=\begin{cases}\omega+(1-\omega)\left(\dfrac{\phi}{\phi+\mu}\right)^{\phi}, & y=0,\\ (1-\omega)\left(\dfrac{\phi}{\phi+\mu}\right)^{\phi}\left(\dfrac{\mu}{\phi+\mu}\right)^{y}\dfrac{\Gamma(y+\phi)}{\Gamma(\phi)\Gamma(y+1)}, & y=1,2,\ldots.\end{cases}$$

Here, $\omega \in (0, 1)$ is the probability of extra zeros. The ZIP and ZINB distributions tend to the Poisson and NB distributions, respectively, as $\omega \to 0$. Tables 6 and 7 present, for the ZIP and ZINB models, respectively, the observed and expected frequencies, log-likelihood function values evaluated at the ML estimates, ML estimates, asymptotic SEs (between parentheses), $\chi^2$ statistics and the corresponding $p$-values. On the basis of the $\chi^2$ statistics and the corresponding $p$-values, we have that the ZIP model is not suitable to fit the data sets (see Table 6). Table 7 indicates that the ZINB model is adequate to fit the data sets (see the $\chi^2$ statistics and the corresponding $p$-values), however, note that the ML estimates of $\omega$ are near zero and hence the NB should be preferable since it has less parameters to be estimated

**Table 6** *Fit of the data sets; ZIP model*

| Automobile claim | | | Accident proneness | | | Chromatid aberrations | | |
|---|---|---|---|---|---|---|---|---|
| Count | Observed | Expected | Count | Observed | Expected | Count | Observed | Expected |
| 0 | 3719 | 3719.03 | 0 | 77 | 77.01 | 0 | 268 | 268.02 |
| 1 | 232 | 224.67 | 1 | 36 | 23.18 | 1 | 87 | 71.79 |
| 2 | 38 | 48.50 | 2 | 24 | 26.19 | 2 | 26 | 40.03 |
| 3 | 7 | 6.98 | 3 | 13 | 19.72 | 3 | 9 | 14.88 |
| 4 | 3 | 0.75 | 4 | 4 | 11.14 | 4 | 4 | 4.15 |
| 5 | 1 | 0.07 | 5 | 3 | 5.03 | 5 | 2 | 0.93 |
| | | | 6 | 2 | 1.89 | 6 | 1 | 0.17 |
| | | | 7 | 1 | 0.61 | 7 | 3 | 0.03 |
| | | | 8 | 2 | 0.17 | | | |
| | | | 9 | 2 | 0.04 | | | |
| | | | 10 | 0 | 0.01 | | | |
| | | | 15 | 1 | 0.00 | | | |
| Maximum log-likelihood | −1187.78 | | | −285.57 | | | −413.15 | |
| $\widehat{\lambda}$ | | 0.4318 | | | 2.2591 | | | 1.1153 |
| | | (0.0518) | | | (0.1767) | | | (0.1116) |
| $\widehat{\omega}$ | | 0.7997 | | | 0.4045 | | | 0.5091 |
| | | (0.0224) | | | (0.0451) | | | (0.0440) |
| $\chi^2$ | | 14.845 | | | 15.492 | | | 4.709 |
| Degrees of freedom | | 2 | | | 3 | | | 2 |
| $p$-value | | <0.001 | | | <0.001 | | | <0.001 |

**Table 7** *Fit of the data sets; ZINB model*

| Automobile claim | | | Accident proneness | | | Chromatid aberrations | | |
|---|---|---|---|---|---|---|---|---|
| Count | Observed | Expected | Count | Observed | Expected | Count | Observed | Expected |
| 0 | 3719 | 3719.22 | 0 | 77 | 77.94 | 0 | 268 | 270.18 |
| 1 | 232 | 229.90 | 1 | 36 | 35.84 | 1 | 87 | 78.55 |
| 2 | 38 | 39.91 | 2 | 24 | 20.04 | 2 | 26 | 29.84 |
| 3 | 7 | 8.42 | 3 | 13 | 11.86 | 3 | 9 | 12.22 |
| 4 | 3 | 1.93 | 4 | 4 | 7.22 | 4 | 4 | 5.19 |
| 5 | 1 | 0.46 | 5 | 3 | 4.47 | 5 | 2 | 2.25 |
| | | | 6 | 2 | 2.79 | 6 | 1 | 0.99 |
| | | | 7 | 1 | 1.76 | 7 | 3 | 0.44 |
| | | | 8 | 2 | 1.11 | | | |
| | | | 9 | 2 | 0.71 | | | |
| | | | 10 | 0 | 0.45 | | | |
| | | | 15 | 1 | 0.05 | | | |
| Maximum log-likelihood | | −1183.55 | | | −262.60 | | | −399.86 |
| $\widehat{\mu}$ | | 0.0865 | | | 1.3455 | | | 0.5475 |
| | | (0.1046) | | | (0.3649) | | | (0.1701) |
| $\widehat{\phi}$ | | 0.2166 | | | 0.6986 | | | 0.6200 |
| | | (0.3258) | | | (0.3732) | | | (0.3383) |
| $\widehat{\omega}$ | | 0.0008 | | | 0.0002 | | | 0.00008 |
| | | (1.2072) | | | (0.2487) | | | (0.2989) |
| $\chi^2$ | | 1.417 | | | 2.350 | | | 2.4159 |
| Degrees of freedom | | 1 | | | 2 | | | 1 |
| $p$-value | | 0.234 | | | 0.309 | | | 0.120 |

(i.e., simpler model). In summary, the above analysis indicates that having a lot of zeros does not necessarily mean that you need a zero-inflated model; see, for example, Allison (2012, Ch. 9).

Finally, it worth emphasizing that we have proposed a two-parameter discrete distribution which seems to give a satisfactory fit (at least) in the three cases considered, on the basis of the $\chi^2$ statistics and the corresponding $p$-values. So, the dN may be a good alternative to the popular NB distribution (as well as some zero-inflated models) in practice. Therefore, we believe the two-parameter dN distribution may be an excellent means of fitting an empirical distribution that presents too many zeros and/or overdispersion.

## 6 Concluding remarks

In this paper, we have introduced a new discrete distribution, so-called the *discrete Nielsen* (dN) distribution. The new class of discrete distributions was obtained from a series expansion provided by Nielsen (1906). The proposed dN distribution is indexed by two parameters and it has a very simple form for its probability mass function. Additionally, it can have a zero vertex, and it is overdispersed. We have provided a comprehensive account of the structural properties of the new discrete distribution, including explicit expressions for the probability generating function, moment generating function, characteristic function, etc. The estimation of the unknown parameters of the dN distribution was approached by the method of maximum likelihood, and a very simple way of computing them numerically was provided. We derive an expression for the observed Fisher information matrix that can be used to compute asymptotic standard errors for the maximum likelihood estimates. From Monte Carlo simulation experiments we verify that the method of maximum likelihood is very effective in estimating the dN model parameters. Applications of the new discrete distribution to real data sets were given to demonstrate that it can be used quite effectively for modeling count data which present too many zeros and/or overdispersion. In conclusion, the dN distribution may provide a rather flexible mechanism for fitting a wide spectrum of discrete real world data sets which may have a lot of zeros and/or overdispersion. We hope that the new discrete distribution may serve as an alternative distribution (among many others) to the negative binomial distribution for modeling count data in several areas.

Finally, note that we can also consider regression structures for the dN model parameters. Let $Y_1, \ldots, Y_n$ be $n$ independent random variables, where each $Y_i$ $(i = 1, \ldots, n)$ follows the dN distribution with parameters $p_i$ and $\theta_i$; that is, $Y_i \sim \mathrm{dN}(p_i, \theta_i)$. Suppose the following functional relations:

$$g_1(p_i) = \eta_{1i} = \boldsymbol{x}_i^\top \boldsymbol{\beta},$$
$$g_2(\theta_i) = \eta_{2i} = \boldsymbol{s}_i^\top \boldsymbol{\tau},$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_r)^\top$ and $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_s)^\top$ are vectors of unknown regression coefficients which are assumed to be functionally independent, $\boldsymbol{\beta} \in \mathbb{R}^r$ and $\boldsymbol{\tau} \in \mathbb{R}^s$ with $r + s < n$, $\eta_{1i}$ and $\eta_{2i}$ are the linear predictors, and $\boldsymbol{x}_i^\top = (x_{i1}, \ldots, x_{in})$ and $\boldsymbol{s}_i^\top = (s_{i1}, \ldots, s_{in})$ are observations on $r$ and $s$ known covariates (or independent variables or regressors). Additionally, the functions $g_1$ and $g_2$ here play a similar role to the link functions of generalized linear models, in the sense that they specifically define how the parameters are linked to linear combinations of the covariates. It is important for the link functions to be injective and the covariates to be linearly independent, so that, with these two conditions, the regression parameters are identifiable. So, the functions $g_1 : (0, 1) \to \mathbb{R}$ and $g_2 : (0, \infty) \to \mathbb{R}$ are assumed to be strictly monotonic and twice differentiable. There are several possible choices for the link functions $g_1(\cdot)$ and $g_2(\cdot)$. For instance: logit $g_1(p) = \log(p/(1 - p))$; probit $g_1(p) = \Phi^{-1}(p)$; and logarithmic $g_2(\theta) = \log(\theta)$. Note that, from the relations given in the previous section between the parameters and moments of the dN distribution, the covariates of the above regression model affect not only the mean but also the variance of the distribution of the observations. An in-depth investigation of such regression model is beyond the scope of the present paper, but certainly is an interesting topic for future work.

## Appendix A: Elements of $\boldsymbol{J}_n(p, \theta)$

The observed information matrix $\boldsymbol{J}_n(p, \theta)$ is given by

$$\boldsymbol{J}_n(p, \theta) = \begin{bmatrix} J_{pp} & J_{p\theta} \\ J_{p\theta} & J_{\theta\theta} \end{bmatrix},$$

whose elements are

$$J_{pp} = -\frac{n\theta[1 + \log(1 - p)]}{[(1 - p)\log(1 - p)]^2} + \frac{n(\theta + \bar{x})}{p^2},$$
$$J_{p\theta} = -\frac{n[(1 - p)\log(1 - p) + p]}{p(1 - p)\log(1 - p)},$$
$$J_{\theta\theta} = \frac{n}{\theta^2} - \sum_{i=1}^{n}\sum_{k=0}^{x_i} \frac{1}{(\theta + k)^2} - \sum_{i=1}^{n} \Delta_i,$$

where

$$\Delta_i = \begin{cases} 0, & x_i = 0, \\ \dfrac{A_{i(1)}}{A_{i(3)}} - \left(\dfrac{A_{i(2)}}{A_{i(3)}}\right)^2, & x_i = 1, 2, \ldots, \end{cases}$$

with

$$A_{i(1)} = \sum_{k=0}^{x_i-1} \frac{\left\langle\!\left\langle {x_i \atop k} \right\rangle\!\right\rangle}{(2x_i)!} \frac{\Gamma(\theta+k+x_i+1)}{\Gamma(\theta+k-x_i+1)} \times \left\{\left[\Psi(\theta+k+x_i+1) - \Psi(\theta+k-x_i+1)\right]^2 + \Psi'(\theta+k+x_i+1) - \Psi'(\theta+k-x_i+1)\right\},$$

$$A_{i(2)} = \sum_{k=0}^{x_i-1} \frac{\left\langle\!\left\langle {x_i \atop k} \right\rangle\!\right\rangle}{(2x_i)!} \frac{\Gamma(\theta+k+x_i+1)}{\Gamma(\theta+k-x_i+1)} \times \left\{\Psi(\theta+k+x_i+1) - \Psi(\theta+k-x_i+1)\right\},$$

$$A_{i(3)} = \sum_{k=0}^{x_i-1} \frac{\left\langle\!\left\langle {x_i \atop k} \right\rangle\!\right\rangle}{(2x_i)!} \frac{\Gamma(\theta+k+x_i+1)}{\Gamma(\theta+k-x_i+1)},$$

where $\Gamma(\cdot)$, $\Psi(\cdot)$ and $\Psi'(\cdot)$ are the gamma, digamma and trigamma functions, respectively.

## Appendix B: Simulation results

The Monte Carlo simulation results are provided in Tables 8 and 9.

**Table 8** *Simulation results*; $p = 0.4$

| | $\theta = 1.5$ | | $\theta = 2.5$ | | $\theta = 3.5$ | | $\theta = 5.0$ | |
|---|---|---|---|---|---|---|---|---|
| | $\widehat{p}$ | $\widehat{\theta}$ | $\widehat{p}$ | $\widehat{\theta}$ | $\widehat{p}$ | $\widehat{\theta}$ | $\widehat{p}$ | $\widehat{\theta}$ |
| | | | | $n = 150$ | | | | |
| Mean | 0.3726 | 2.2656 | 0.3864 | 3.1270 | 0.3769 | 4.4530 | 0.3820 | 6.0282 |
| Bias | −0.0274 | 0.7656 | −0.0136 | 0.6270 | −0.0231 | 0.9530 | −0.0180 | 1.0282 |
| RMSE | 0.1281 | 1.7923 | 0.1065 | 2.3963 | 0.1047 | 3.2097 | 0.0959 | 3.3565 |
| CP(90%) | 87.2 | 90.8 | 89.3 | 90.0 | 89.8 | 91.2 | 91.3 | 92.3 |
| CP(95%) | 94.0 | 94.3 | 95.0 | 93.2 | 94.7 | 93.5 | 96.3 | 94.5 |
| | | | | $n = 250$ | | | | |
| Mean | 0.3823 | 1.7762 | 0.3906 | 2.7558 | 0.3893 | 3.8884 | 0.3918 | 5.4957 |
| Bias | −0.0177 | 0.2762 | −0.0094 | 0.2558 | −0.0107 | 0.3884 | −0.0082 | 0.4957 |
| RMSE | 0.0920 | 0.8986 | 0.0802 | 1.0315 | 0.0768 | 1.4762 | 0.0744 | 2.0900 |
| CP(90%) | 90.0 | 93.8 | 89.3 | 91.3 | 90.5 | 91.8 | 89.5 | 90.5 |
| CP(95%) | 94.8 | 95.3 | 94.0 | 93.2 | 94.5 | 94.0 | 94.8 | 93.3 |
| | | | | $n = 400$ | | | | |
| Mean | 0.3873 | 1.6516 | 0.3946 | 2.6642 | 0.3967 | 3.6996 | 0.3900 | 5.3646 |
| Bias | −0.0127 | 0.1516 | −0.0054 | 0.1642 | −0.0033 | 0.1996 | −0.0100 | 0.3646 |
| RMSE | 0.0696 | 0.5230 | 0.0642 | 0.8254 | 0.0627 | 1.0114 | 0.0600 | 1.4032 |
| CP(90%) | 91.2 | 92.5 | 91.2 | 93.3 | 89.3 | 89.8 | 89.2 | 91.8 |
| CP(95%) | 96.3 | 95.5 | 95.5 | 95.8 | 94.8 | 93.8 | 94.3 | 94.7 |

**Table 9** *Simulation results*; $p = 0.7$

| | $\theta = 1.5$ | | $\theta = 2.5$ | | $\theta = 3.5$ | | $\theta = 5.0$ | |
|---|---|---|---|---|---|---|---|---|
| | $\widehat{p}$ | $\widehat{\theta}$ | $\widehat{p}$ | $\widehat{\theta}$ | $\widehat{p}$ | $\widehat{\theta}$ | $\widehat{p}$ | $\widehat{\theta}$ |
| | | | | $n = 250$ | | | | |
| Mean | 0.6902 | 1.5761 | 0.6930 | 2.5869 | 0.6939 | 3.6129 | 0.6898 | 5.2688 |
| Bias | −0.0098 | 0.0761 | −0.0070 | 0.0869 | −0.0061 | 0.1129 | −0.0102 | 0.2688 |
| RMSE | 0.0610 | 0.3766 | 0.0508 | 0.5141 | 0.0465 | 0.6425 | 0.0493 | 1.0487 |
| CP(90%) | 89.8 | 92.3 | 90.0 | 90.8 | 91.8 | 93.0 | 89.2 | 91.2 |
| CP(95%) | 95.7 | 95.7 | 94.8 | 94.3 | 96.3 | 96.0 | 94.2 | 95.2 |
| | | | | $n = 250$ | | | | |
| Mean | 0.6934 | 1.5423 | 0.6944 | 2.5658 | 0.6962 | 3.5689 | 0.694 | 5.1330 |
| Bias | −0.0066 | 0.0423 | −0.0056 | 0.0658 | −0.0038 | 0.0689 | −0.0060 | 0.1330 |
| RMSE | 0.0458 | 0.2585 | 0.0400 | 0.3875 | 0.0361 | 0.4990 | 0.0364 | 0.7258 |
| CP(90%) | 89.7 | 90.3 | 89.7 | 90.5 | 91.2 | 91.7 | 87.5 | 88.2 |
| CP(95%) | 95.7 | 95.2 | 95.2 | 95.8 | 95.8 | 95.0 | 93.2 | 94.2 |
| | | | | $n = 400$ | | | | |
| Mean | 0.6968 | 1.5166 | 0.6978 | 2.5423 | 0.6960 | 3.5606 | 0.6974 | 5.0616 |
| Bias | −0.0032 | 0.0166 | −0.0022 | 0.0423 | −0.0040 | 0.0606 | −0.0026 | 0.0616 |
| RMSE | 0.0349 | 0.1913 | 0.0333 | 0.3142 | 0.0289 | 0.3976 | 0.0269 | 0.5275 |
| CP(90%) | 91.5 | 90.7 | 88.2 | 88.0 | 91.3 | 90.2 | 90.0 | 91.5 |
| CP(95%) | 96.7 | 95.0 | 93.5 | 93.3 | 94.7 | 95.3 | 95.2 | 96.2 |

## Acknowledgments

We are very grateful to two anonymous referees for the valuable comments and suggestions which have improved the first version of the paper. Fredy Castellares gratefully acknowledges the financial support from FAPEMIG (Belo Horizonte/MG, Brazil). Artur Lemonte gratefully acknowledges the financial support of the Brazilian agency CNPq (grant 301808/2016–3).

## References

Allison, P. D. (2012). *Logistic Regression Using SAS: Theory and Application*, 2nd ed. Cary, North Carolina: SAS Institute Inc.

Barbiero, A. (2014). An alternative discrete skew Laplace distribution. *Statistical Methodology* **16**, 47–67. MR3110887 https://doi.org/10.1016/j.stamet.2013.07.002

Catcheside, D. G., Lea, D. E. and Thoday, J. M. (1946a). Types of chromosome structural change induced by the irradiation of Tradescantia microspores. *Journal of Genetics* **47**, 113–136.

Catcheside, D. G., Lea, D. E. and Thoday, J. M. (1946b). The production of chromosome structural changes in Tradescantia microspores in relation to dosage, intensity and temperature. *Journal of Genetics* **47**, 137–149.

Englehardt, J. D. and Li, R. C. (2011). The discrete Weibull distribution: An alternative for correlated counts with confirmation for microbial counts in water. *Risk Analysis* **31**, 370–381.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications. Volume II*. New York: John Wiley & Sons. MR0270403

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**, 42–58.

Flajonet, P. and Sedgewick, R. (2009). *Analytic Combinatorics*. New York: Cambridge University Press. MR2483235 https://doi.org/10.1017/CBO9780511801655

Gómez-Déniz, E., Sarabia, J. M. and Calderin-Ojeda, E. (2011). A new discrete distribution with actuarial applications. *Insurance Mathematics & Economics* **48**, 406–412. MR2820054 https://doi.org/10.1016/j.insmatheco.2011.01.007

Gossiaux, A. and Lemaire, J. (1981). Methodes d'ajustement de distributions de sinistres. *Bulletin of the Association of Swiss Actuaries* **81**, 87–95.

Graham, R., Knuth, D. and Patashnik, O. (1994). *Concrete Mathematics: A Foundation for Computer Science*, 2nd ed. New York: Addison-Wesley. MR1397498

Inusah, S. and Kozubowski, T. J. (2006). A discrete analogue of the Laplace distribution. *Journal of Statistical Planning and Inference* **136**, 1090–1102. MR2181990 https://doi.org/10.1016/j.jspi.2004.08.014

Jazi, M. A., Lai, C. D. and Alamatsaz, M. H. (2010). A discrete inverse Weibull distribution and estimation of its parameters. *Statistical Methodology* **7**, 121–132. MR2591715 https://doi.org/10.1016/j.stamet.2009.11.001

Johnson, N. L. and Kotz, S. (1982). Developments in discrete distributions, 1969–1980. *International Statistical Review* **50**, 71–101. MR0668611 https://doi.org/10.2307/1402460

Jones, M. C. (2015). On families of distributions with shape parameters. *International Statistical Review* **83**, 175–192. MR3377071 https://doi.org/10.1111/insr.12055

Kozubowski, T. J. and Inusah, S. (2006). A skew Laplace distribution on integers. *Annals of the Institute of Statistical Mathematics* **58**, 555–571. MR2327893 https://doi.org/10.1007/s10463-005-0029-1

Krishna, H. and Pundir, P. S. (2009). Discrete Burr and discrete Pareto distributions. *Statistical Methodology* **6**, 177–188. MR2649616 https://doi.org/10.1016/j.stamet.2008.07.001

Nekoukhou, V., Alamatsaz, M. H. and Bidram, H. (2013). Discrete generalized exponential distribution of a second type. *Statistics* **47**, 876–887. MR3175721 https://doi.org/10.1080/02331888.2011.633707

Nielsen, N. (1906). *Handbuch der Theorie der Gammafunction*. Leipzig: Teubner.

Nooghabi, M. S., Roknabadi, A. H. R. and Borzadaran, G. M. (2011). Discrete modified Weibull distribution. *Metron* **LXIX**, 207–222. MR3041266 https://doi.org/10.1007/BF03263557

Patil, G. P. (1963). Minimum variance unbiased estimation and certain problems of additive number theory. *The Annals of Mathematical Statistics* **34**, 1050–1056. MR0155388 https://doi.org/10.1214/aoms/1177704029

Patil, G. P. and Wani, J. K. (1965). On certain structural properties of the logarithmic series distribution and the first type Stirling distribution. *Sankhya Series A* **27**, 271–280. MR0202233

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ross, S. M. (2013). *Simulation*, 5th ed. London: Academic Press.

Roy, D. (2004). Discrete Rayleigh distribution. *IEEE Transactions on Reliability* **53**, 255–260.

Sichel, H. S. (1951). The estimation of the parameters of a negative binomial distribution with special reference to psychological data. *Psychometrika* **16**, 107–127. MR0042100 https://doi.org/10.1007/BF02313431

Ward, M. (1934). The representation of Stirling's numbers and Stirling's polynomials as sums of factorial. *American Journal of Mathematics* **56**, 87–95. MR1507004 https://doi.org/10.2307/2370916

F. Castellares
M. A. C. Santos
Departamento de Estatística, ICEX
Universidade Federal de Minas Gerais
Belo Horizonte/MG
Brazil
E-mail: fwcc29@gmail.com
msantos@est.ufmg.br

A. J. Lemonte
Departamento de Estatística, CCET
Universidade Federal do Rio Grande no Norte
Natal/RN
Brazil
E-mail: arturlemonte@gmail.com