Inflated-Parameter Family of Generalized Power Series Distributions and Their Application in Analysis of Overdispersed Insurance Data

Nikolai Kolev[1], **Leda Minkova**[2] and **Plamen Neytchev**[3]

[1]Department of Statistics - IME, São Paulo University, Brazil,
[2]Department of Probability, Technical University of Sofia, Bulgaria,
[3]Institute of Meteorology and Hydrology, Bulgarian Academy of Sciences

ABSTRACT

In this paper we suggest an extension of the generalized power series distributions by including an additional parameter ρ. It has a natural interpretation in terms of both "zero-inflated" proportion and correlation coefficient. A new concept of ρ-type lack of memory property is introduced. The proposed new discrete distributions are candidates for modeling of correlated count data which exhibit overdispersion. Two alternative representations of the probability mass functions of the extended distributions are given. A real frequency data are fitted successfully with the extended Poisson and negative binomial models.

Key words and phrases: *Collective model of risk theory; correlation; generalized power series distributions; inflated-parameter distributions; overdispersion.*

AMS subject classification: 60E05, 62P05, 60J10.

Corresponding autor's address: Department of Statistics - IME, São Paulo University, C.P. 66281, 05315-970 São Paulo, SP, Brazil; **E-mail address:** nkolev@ime.usp.br

1 Introduction

During the last decade, a vast activity have been observed in generalizing of the classical discrete distributions. The main idea was to apply the extended versions for modeling different kinds of dependent count or frequency data structure in various fields (Econometrics, Insurance, Finance, Biometrics, etc.), see for example Bowers et *al.* (1997), Collett (1991), Johnson et *al.* (1992), Luceño (1995), Rolski et *al.* (1999), Winkelmann (2000) and references therein.

In the general introduction of the recent monographs Bowers et *al.* (1997), Rolski et *al.* (1999), Winkelmann (2000) is emphasized the need for richer classes of probability distributions when modeling count data. Since the probability distributions for counts are nonstandard in the actuarial literature, special attention is paid here for more flexible distributions, since they can be used as a building blocks for improved count data models with immediate application in insurance describing the accumulated claims.

In the present paper we suggest extensions of the classical univariate geometric, negative binomial, Poisson, Bernoulli, binomial and logarithmic series distributions, by including an additional parameter ρ. It has a natural interpretation in terms of "zero-inflation", and because of this we named the corresponding generalized versions adding "inflated-parameter".

After giving notations and preliminary results in Section 2, probability mass functions (PMF) and probability generating functions (PGF) of the corresponding inflated-parameter distributions are presented in Sections 3 through 8. The relationships between the inflated-parameter distributions, according to the remaining parameters, are the same as between their classical analogue. This simply shows that the new generalized distributions compose a new class (family) of discrete distributions and this topic is discussed in Section 9, where two different representations of the corresponding PMF's of the r.v.'s belonging to the inflated-parameter family of the generalized power series distributions are presented. An overdispersed property of the new class according the family of generalized power series distributions is discussed as well as a new constructive interpretation of the parameter ρ is obtained. In Section 10 we use the inflated-parameter Poisson and negative binomial distributions to approximate real frequency data. At the end, some conclusions are given.

2 Notations and Preliminaries

The random variables (r.v.) considered are assumed to be defined on a fixed probability space $(\Omega, \mathcal{F}, \mathbf{P})$. We will deal with a nonnegative r.v.'s representing a number of claims or a claim amount adopting the assumptions of the collective model of risk theory for a fixed period of time. Let N be a nonnegative integer valued r.v. representing a number of claims, with counting density

$$p_k = P(N = k), \quad k = 0, 1, \dots .$$

For the PGF of the r.v. N we will use notation

$$P_N(t) = \sum_{k=0}^{\infty} p_k t^k, \quad |t| < 1.$$

There are at least three particular cases that are applicable in Insurance as a claim number distributions: the Poisson, the binomial and the negative binomial distributions. We write

- (i) $N \sim Po(\lambda)$ if N has a Poisson distribution with parameter $\lambda > 0$, i.e.

$$p_k = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0, 1, \ldots;$$

- (ii) $N \sim Bi(\pi, n)$ if N has a binomial distribution with parameters $\pi \in (0,1)$ and $n \in \{0, 1, \ldots\}$, i.e.

$$p_k = \binom{n}{k}\pi^k(1-\pi)^{n-k}, \quad k = 0, 1, \ldots, n.$$

  When $n = 1$, we obtain the Bernoulli r.v. with a parameter π;

- (iii) $N \sim NB(\pi, r)$ if N has a negative binomial (NB) distribution with parameters $\pi \in (0,1)$ and $r > 0$, i.e.

$$p_k = \binom{r+k-1}{k}\pi^r(1-\pi)^k, \quad k = 0, 1, \ldots.$$

  In the special case $r = 1$, we obtain the geometric distribution with a parameter π, $Ge_0(\pi)$, on the nonnegative integers.

- (iv) $N \sim LS(\pi)$ if N has a logarithmic series distribution with parameter $\pi \in (0,1)$, i.e.

$$p_k = \frac{\pi^k}{-k\log(1-\pi)}, \quad k = 1, 2, \ldots.$$

  The logarithmic series distribution is used rather rarely. It can be obtained as a limiting distribution of the truncated at zero NB distribution. We include the logarithmic series distribution since it can be used to model the numbers of items of a product purchased by a buyer in a specified period of time, e.g. Chatfield et *al.* (1966);

- (v) $N \sim \delta_m$, if N is concentrated on the integer $m \in \{0, 1, \ldots\}$, i.e. degenerated at $N = m$ with

$$p_m = 1, \quad p_k = 0 \quad \text{for} \quad k \neq m.$$

The equality of mean and variance is characteristic of the Poisson distribution and can be referred to as *equidispersion*. Departures from equidispersion can be either as *overdispersion* (variance is greater than mean) or *underdispersion* (variance is less than the mean). The Binomial distribution is underdispersed and the NB distribution is overdispersed according to the Poisson distribution. The logarithmic series distribution displays overdispersion for $0 < -[\log(1-\pi)]^{-1} < 1$ and underdispersion for $-[\log(1-\pi)]^{-1} > 1$.

2.1 Generalized Power Series Distributions and Panjer's Recursion

Many univariate discrete probability distributions, with a single parameter belong to the class of *generalized power series distributions* (GPSD) or to the classes of their generalizations, see Gupta (1974), Consul (1990).

Definition 2.1. The PGF of the GPSD with a parameter $\theta > 0$, is given by the following relation

$$\varphi(t) = \frac{g(\theta t)}{g(\theta)}, \tag{1}$$

where $g(\theta)$ is a positive, finite and differentiable function. For any member of this family, the PMF of the corresponding r.v. X can be written as

$$P(N = k) = \frac{a(k)\theta^k}{g(\theta)}, \quad k \in S; \quad \theta > 0, \tag{2}$$

where S is any nonempty enumerable set of nonnegative integers, $a(k) \geq 0$ and $g(\theta) = \sum_{k \in S} a(k)\theta^k$.

The binomial, the NB, the logarithmic series and the Poisson distributions belong to this class, see Patil (1962). In the binomial and NB cases, the corresponding additional integer parameters n and r are treated as nuisance parameters.

In the particular cases, the functions $a(k)$, $g(\theta)$ and the parameter θ, are given by the following expressions

$$X \sim Bi(\theta, n): \quad a(k) = \binom{n}{k}, \quad g(\theta) = (1+\theta)^n, \quad \theta = \tfrac{\pi}{1-\pi};$$

$$X \sim Po(\theta): \quad a(k) = \tfrac{1}{k!}, \quad g(\theta) = e^\theta, \quad \theta = \lambda;$$

$$X \sim NB(\theta, r): \quad a(k) = \binom{k+r-1}{k}, \quad g(\theta) = (1-\theta)^{-r}, \quad \theta = 1-\pi;$$

$$X \sim LS(\theta): \quad a(k) = \tfrac{1}{k}, \quad g(\theta) = -ln(1-\theta), \quad \theta = 1-\pi.$$

The concept of the GPSD and their extensions is not popular in actuarial literature. The authors prefer to use as a first step the Panjer-recursion

$$p_k = \left(a + \frac{b}{k}\right) p_{k-1}, \quad k = 1, 2, \ldots \tag{3}$$

for some constants $a < 1$, and b, cf. Panjer (1981). In Sundt and Jewell (1981) is shown that the recursion is satisfied if and only if N has a Poisson, a binomial or a NB distribution or $N \sim \delta_0$, respectively. If we start recursion (3) only at $k = 2$ then logarithmic series distribution also satisfies it as well as the truncated versions of the above distributions. The values of the constants a and b in the different cases can be found, for example in Straub (1988), p. 35. Let us only note, that the class of distributions fulfilling (3) is known in actuarial literature as a $R1(a, b)$ class also.

So, the GPSD satisfy Panjer's recursion (3). In this paper we will follow the GPSD terminology since, at least historically, it appears earlier.

Recently there are many generalizations of the recursion formula (3) where the densities of counting distributions satisfy certain second and higher order difference equations. Recursions for the evaluation of related compound distributions have been developed in the case of severity distributions which are concentrated on the non-negative integers, see for example Schrörter (1990), Sundt (1992), Dhaene et *al.* (1996).

2.2 Zero-inflated Distributions

Our study is based on the *inflated parameter (zero-modified)* discrete distributions, which are used to model counts that encounter disproportionally large frequencies of zeros, e.g. Johnson at *al.* (1992). Let ξ be an arbitrary nonnegative integer-valued r.v. such that

$$P(\xi = j) = p_j, \quad j = 0, 1, ..., \quad \sum_{j=0}^{\infty} p_j = 1,$$

and let $G_\xi(t) = E(t^\xi)$ be its PGF. An extra proportion of zeros, $\rho \in (0,1)$, is *added* to the proportion of zeros from the distribution of the r.v. ξ, while decreasing the remaining proportions in an appropriate way. The *zero-inflated* modification η of ξ is defined by

$$\begin{aligned} P(\eta = 0) &= \rho + (1-\rho)p_0, \\ P(\eta = j) &= (1-\rho)p_j, \quad j = 1, 2, \ldots . \end{aligned} \tag{4}$$

It has as a PGF

$$G_\eta(t) = \rho + (1-\rho)G_\xi(t). \tag{5}$$

Zero-inflated models address the problem, that the data display a higher fraction of zeros, or non occurrences, than can be possibly explained through any fitted standard count model. The zero-inflated distributions are appropriate alternatives for modeling clustered samples when the population consists of two sub-populations, one containing only zeros, while in the other, counts from a discrete distribution are observed.

If $\rho = 1$, than the corresponding zero-inflated distribution is the degenerated at zero one; if $\rho = 0$, "nothing is changed" in (5), i.e. $G_\eta(t) = G_\xi(t)$.

In general, the inflation parameter ρ may take *negative* values provided that $P(\eta = 0) \geq 0$, i.e., $\rho \geq -\frac{p_0}{1-p_0}$ and therefore $\max\{-1, -\frac{p_0}{1-p_0}\} \leq \rho \leq 0$. This case corresponds to the "opposite" phenomena - "excluding" a proportion of zeros from the basic discrete distribution, if necessary.

In actuarial literature, e.g. Rolski et *al.* (1999) p.35, the zero-inflated distributions are known as a "θ-modification" which can be considered as a reverse truncated operation.

3 Inflated-parameter Geometric Distribution

In this section we suggest a generalization of the usual geometric distribution by including an additional parameter $\rho \in [0,1)$. Several simple interpretations of the proposed distribution are presented in Section 3.1 and a new concept of "ρ-type lack-of memory property" is introduced in Section 3.2.

Let $\{W_1, W_2, \ldots\}$ be an infinite sequence of independent binary variables

$$W_k = \begin{cases} 0 & \text{with probability} \quad 1-\pi, \\ 1 & \text{with probability} \quad \pi, \end{cases}$$

for $k = 1, 2, \ldots$, where the parameter $\pi \in (0, 1)$. In the sequel, we will identify the realization "1" as a "success".

Consider its corresponding *zero-inflated sequence* $\{\overline{W}_1, \overline{W}_2, \ldots\}$, determined for $k = 1, 2, \ldots$ according to (4):

$$\overline{W}_k = \begin{cases} 0 & \text{with probability} \quad (1-\pi)(1-\rho)+\rho, \\ 1 & \text{with probability} \quad (1-\rho)\pi, \end{cases}$$

with $\max\{-1, -\frac{1-\pi}{\pi}\} \leq \rho < 1$. Let the r.v. V be equal to the number of trials that we need to achieve the first observed "success" in the new constructed sequence $\{\overline{W}_1, \overline{W}_2, \ldots\}$ of independent binary variables. The PMF of the r.v. V is given by

$$P(V = k) = [(1-\pi)(1-\rho)+\rho]^{k-1}(1-\rho)\pi, \quad k = 1, 2, \ldots \tag{6}$$

The r.v. V has the usual geometric distribution on the positive integers, $V \sim Ge_1(\pi^*)$, with a parameter

$$\pi^* = (1-\pi)(1-\rho)+\rho = 1-(1-\rho)\pi. \tag{7}$$

Now, let us define the r.v. X by the following relations

$$\begin{aligned} P(X = 0) &= \pi, \\ P(X = k) &= (1-\pi)[(1-\pi)(1-\rho)+\rho]^{k-1}(1-\rho)\pi, \quad k = 1, 2, \ldots \end{aligned} \tag{8}$$

It is easy to verify that the above equations define a proper probability distribution. The corresponding PGF $P_X(s)$ is determined by the following expression

$$P_X(t) = \frac{\pi(1-t\rho)}{1-t[(1-\pi)(1-\rho)+\rho]} = \frac{\pi(1-t\rho)}{1-t(1-\pi+\rho\pi)}. \tag{9}$$

Definition 3.1. We say that the r.v. X defined by (8) (or (9)) has an *inflated-parameter geometric distribution* with parameters $\pi \in (0, 1)$ and $\rho \in (\max\{-1, -\frac{1-\pi}{\pi}\}, 1)$ and will denote this by $X \sim IGe_0(\pi, \rho)$.

Remark 3.1. If $\rho = 0$, the defined inflated-parameter geometric distribution coincides with the usual geometric distribution on the nonnegative integer values, with parameter π, i.e. $Ge_0(\pi) \equiv IGe_0(\pi, 0)$.

Remark 3.2. The mean and the variance of the $IGe_0(\pi, \rho)$ distribution are given by

$$E(X) = \frac{1-\pi}{\pi(1-\rho)} \quad \text{and} \quad Var(X) = \frac{(1-\pi)(1+\pi\rho)}{\pi^2(1-\rho)^2}.$$

3.1 Interpretations

Remark 3.3. Let us consider a time homogeneous Markov chain $\{X_n, n \geq 0\}$ taking values $\{0, 1\}$. We determine the sequence $X_0, X_1, \ldots$ by the distribution of the initial states

$$P(X_0 = 0) = 1 - \pi \quad \text{and} \quad P(X_0 = 1) = \pi$$

and for $n = 0, 1, \ldots$, by the transition probabilities

$$P(X_{n+1} = 0 \mid X_n = 0) = (1 - \pi)(1 - \rho) + \rho, \quad P(X_{n+1} = 1 \mid X_n = 0) = (1 - \rho)\pi$$

and

$$P(X_{n+1} = 0 \mid X_n = 1) = (1 - \pi)(1 - \rho), \quad P(X_{n+1} = 1 \mid X_n = 1) = (1 - \rho)\pi + \rho.$$

For the defined Markov chain the r.v. X determined by (8) has the following *interpretation*: it gives the number of transitions until the first "success" is observed in the sequence $X_0, X_1, \ldots$.

Remark 3.4. Note, that the parameter π^* of the r.v. $V \sim Ge_1(\pi^*)$, given by (7) coincides with the probability $P(\overline{W}_k = 0)$ for the zero-inflated sequence $\{\overline{W}_k, k \geq 1\}$ as well as with the conditional probability $P(X_{n+1} = 0 \mid X_n = 0)$ for the two-state homogeneous Markov chain considered by Remark 3.3.

Let us underline that, in fact, the parameter ρ, represents the proportion of zeros *added* to the usual Bernoulli distribution (when $\rho > 0$), *decreasing* the "successive" value 1 in an appropriate way.

Remark 3.5. Let us consider the *Correlated binomial distribution*, introduced by Luceño (1995). A r.v. W following this distribution counts the number of "successes" in a sample of n subjects that give equicorrelated binary responses with *correlation coefficient* ρ, probability of success π, under condition that its PMF must depend linearly on ρ. The PGF for $n = 2$ is given by

$$P_W(t) = \rho(1 - \pi + \pi t^2) + (1 - \rho)(1 - \pi + \pi t)^2.$$

From the last expression we obtain the following equations

$$\begin{aligned} P(W = 0) &= (1 - \pi)[\rho + (1 - \rho)(1 - \pi)], \\ P(W = 1) &= 2(1 - \pi)(1 - \rho)\pi, \\ P(W = 2) &= \pi[\rho + (1 - \rho)\pi]. \end{aligned}$$

Now, it is easy to obtain the transition probabilities given by Remark 3.3. Really, if W_1 and W_2 are two equicorrelated binary responses, we have

$$P(W_2 = 1 \mid W_1 = 0) = \frac{P(W_1 = 0, W_2 = 1)}{P(W_1 = 0)} = \frac{\frac{1}{2}P(W_1 + W_2 = 1)}{P(W_1 = 0)} = \frac{\frac{1}{2}P(W = 1)}{P(W_1 = 0)} = (1 - \rho)\pi$$

and

$$P(W_2 = 0 \mid W_1 = 0) = \frac{P(W_1 = 0, W_2 = 0)}{P(W_1 = 0)} = \frac{P(W_1 + W_2 = 0)}{P(W_1 = 0)} = \frac{P(W = 0)}{P(W_1 = 0)},$$

i.e.

$$P(W_2 = 0 \mid W_1 = 0) \quad = \quad (1-\pi)(1-\rho) + \rho.$$

In this case, the r.v. X, given by (8), can be *interpreted* as the number of trials until the first "success" is observed in a sequence $\{W_1, W_2, \ldots\}$ of equicorrelated binary responses. Because of the last interpretation of the r.v. X, one may refer the inflated-geometric distribution $IGe_0(\pi, \rho)$ as *Correlated geometric distribution.*

3.2 ρ-type Lack-of-Memory Property

It is well known, e.g. from Galambos and Kotz (1978), that the equation

$$P(U \geq b + x \mid U \geq b) = P(U \geq x), \quad x \geq 0, \quad b > 0, \tag{10}$$

is true for a r.v. U which is nonnegative and non-degenerate at zero, if and only if it has either the exponential or the geometric distribution. Equation (10) is known as the *lack of memory* property either for the r.v. U or for its distribution function.

Theorem 3.1. *Let $X \sim IGe_0(\pi, \rho)$. Then for any $x \geq 0$ and $b > 0$, the conditional probability $P(X \geq b + x \mid X \geq b)$ has the following equivalent representations:*

(*i*) $[(1-\pi)(1-\rho) + \rho]^x$;

(*ii*) $\frac{(1-\pi)(1-\rho)+\rho}{(1-\pi)} P(X \geq x)$;

(*iii*) $P(X \geq x) + \rho \frac{\pi}{1-\pi} P(X \geq x)$;

(*iv*) $P(X \geq x) + \rho\pi P(X \geq x \mid X > 0)$;

(*v*) $(1-\rho)P(X \geq x) + \rho P(X \geq x \mid X > 0)$;

(*vi*) $P(V \geq x + 1)$,

where the r.v. V is given by (6).

Proof. For any fixed integer $b \geq 1$ from (8) we have

$$P(X \geq b) \quad = \quad \sum_{k=b}^{\infty} (1-\pi)[(1-\pi)(1-\rho) + \rho]^{k-1}(1-\rho)\pi,$$

i.e.

$$P(X \geq b) \quad = \quad (1-\pi)[(1-\pi)(1-\rho) + \rho]^{b-1}.$$

Then for any $x \geq 0$

$$P(X \geq b + x \mid X \geq b) \quad = \quad \frac{P(X \geq b + x)}{P(X \geq b)} \quad = \quad [(1-\pi)(1-\rho) + \rho]^x$$

and the representation *(i)* is obtained.

By simple transformations from *(i)* one can obtain relations *(ii)* - *(v)*. Representation *(vi)* follows from the definition of the r.v. V and *(i)*.

The statement *(v)* from Theorem 3.1 gives us reasons to suggest the following extension of the usual lack-of-memory property.

Definition 3.3. We call that the r.v. U has *ρ-type lack-of-memory property* if

$$P(U \geq b + x \mid U \geq b) = (1-\rho)P(U \geq x) + \rho P(U \geq x \mid U > 0),$$

for any $x \geq 0$ and $b > 0$.

We will not discuss here the characterizations of the ρ-type lack-of-memory property.

4 Inflated-parameter Negative Binomial Distribution

Let r be a positive integer and $X_1, X_2, \dots, X_r$ be independent identically distributed (i.i.d.) r.v.'s having $IGe_0(\pi, \rho)$ distribution, given by (8).

Definition 4.1. We say that the r.v. $Y = X_1 + X_2 + \cdots + X_r$ has an *inflated-parameter negative binomial distribution* with parameters $\pi \in (0,1)$, $\rho \in (\max\{-1, -\frac{1-\pi}{\pi}\}, 1)$ and $r \geq 1$, to be denoted $Y \sim INB(\pi, \rho, r)$.

Since $X_1, X_2, \dots, X_r$ are i.i.d. r.v.'s, each having a PGF given by (9), the PGF of the r.v. $Y \sim INB(\pi, \rho, r)$ has the following form

$$P_Y(t) = \left[\frac{\pi(1-t\rho)}{1-t(1-\pi+\rho\pi)}\right]^r. \tag{11}$$

The PMF of the inflated-parameter negative binomial distribution is given by the next proposition.

Proposition 4.1. *The PMF of the $INB(\pi, \rho, r)$ distributed r.v. Y is given by the following relation*

$$P(Y = y) = \pi^r \sum_{y_1, y_2, \dots} \binom{y_1 + y_2 + \cdots + r - 1}{y_1, y_2, \dots, r-1} [(1-\pi)(1-\rho)]^{y_1+y_2+\cdots} \rho^{y_2+2y_3+\cdots}, \tag{12}$$

where $y = 0, 1, \dots$ and the summation is over all nonnegative integers $y_1, y_2, y_3, \dots$ such that

$$y_1 + 2y_2 + 3y_3 + \cdots = y.$$

Proof. Using some combinatorial equations, from the PGF (11) we consequently obtain

$$
\begin{aligned}
P_Y(t) &= \left[\frac{\pi(1-t\rho)}{1-t(1-\pi+\rho\pi)}\right]^r \\
&= \frac{\pi^r}{\left[1-(1-\pi)(1-\rho)t\frac{1}{1-t\rho}\right]^r} \\
&= \pi^r \sum_{m=0}^{\infty} \binom{-r}{m} \left[-(1-\pi)(1-\rho)t(1+t\rho+t^2\rho^2+\cdots)\right]^m \\
&= \pi^r \sum_{m=0}^{\infty} \binom{m+r-1}{m} [(1-\pi)(1-\rho)t]^m \sum_{m_1,m_2,\ldots} \binom{m}{m_1,m_2,\ldots} (t\rho)^{m_2+2m_3+\cdots},
\end{aligned}
$$

where the last summation is over all nonnegative integers $m_1, m_2, m_3, \ldots$, such that $m_1 + m_2 + m_3 + \cdots = m$. Now, taking into account the equality

$$
\binom{m+r-1}{m}\binom{m}{m_1,m_2,\ldots} = \binom{m+r-1}{m_1,m_2,\ldots,r-1}
$$

we have

$$
P_Y(t) = \pi^r \sum_{m=0}^{\infty} [(1-\pi)(1-\rho)t]^m \sum_{m_1,m_2,\ldots} \binom{m+r-1}{m_1,m_2,\ldots,r-1} (t\rho)^{m_2+2m_3+\cdots}.
$$

Substituting in the last expression $m_i = y_i$, $i \geq 1$ and $m = y - \sum_{j=1}^{\infty}(j-1)y_j$ we finally obtain

$$
P_Y(t) = \sum_{y=0}^{\infty} t^y P(Y=y),
$$

where the probability $P(Y=y)$ is given by (12) for $y \geq 0$.

Remark 4.1. If we put $\rho = 0$ in (12) the PMF of the usual NB distribution is obtained.

Remark 4.2. The probabilities of the first four values of the r.v. $Y \sim INB(\pi, \rho, r)$ are given by the following expressions

$$
\begin{aligned}
P(Y=0) &= \pi^r, \\
P(Y=1) &= \pi^r r(1-\pi)(1-\rho), \\
P(Y=2) &= \pi^r(1-\rho)(1-\pi)\left[\binom{r+1}{2}(1-\rho)(1-\pi)+r\rho\right], \\
P(Y=3) &= \pi^r(1-\rho)(1-\pi)\left[\binom{r+2}{3}(1-\rho)^2(1-\pi)^2+r(r+1)(1-\rho)(1-\pi)\rho+r\rho^2\right],
\end{aligned}
$$

derived from (12).

Remark 4.3. The mean and the variance of the $INB(\pi, \rho, r)$ distribution are given by

$$
E(Y) = \frac{r(1-\pi)}{\pi(1-\rho)} \quad \text{and} \quad Var(Y) = \frac{r(1-\pi)(1+\pi\rho)}{\pi^2(1-\rho)^2}.
$$

5 Inflated-parameter Poisson Distribution

Here we will obtain the PGF and PMF of a new distribution, by finding the limits of the expressions (11) and (12) when

$$r \longrightarrow \infty \quad \text{and} \quad \pi \longrightarrow 1, \quad \text{such that} \quad r(1-\pi) = \lambda = const > 0. \tag{13}$$

The limiting PGF is given by the following proposition.

Proposition 5.1. *Under the limiting conditions (13) the following relation is true*

$$\lim_{r\to\infty}\lim_{\pi\to 1} P_Y(t) = exp\left[\frac{\lambda(t-1)}{1-t\rho}\right], \tag{14}$$

where $P_Y(t)$ the PGF given by (11).

Proof. Taking logarithm on both sides of (11) we have

$$lnP_Y(t) = r\left\{ln[1-(1-\pi+\rho\pi t)] - ln[1-t(1-\pi+\rho\pi)]\right\}.$$

Using the Taylor expansion of the logarithmic function $ln(1-x)$, after some simple transformations we obtain that

$$lnP_Y(t) = r(1-\pi)(t-1)\left\{1+\frac{1}{2}[2\rho\pi t+(1-\pi)(t+1)]\right.$$
$$\left.+\frac{1}{3}[3(\rho\pi t)^2+3\rho\pi t(t+1)(1-\pi)+(t^2+t+1)(1-\pi)^2]+\cdots\right\}.$$

Now using the limiting conditions (13) we finally have

$$\lim_{r\to\infty}\lim_{\pi\to 1} lnP_Y(t) = \lambda(t-1)[1+\rho t+(\rho t)^2+\cdots] = \frac{\lambda(t-1)}{1-\rho t}.$$

Taking anti-logarithm in the last relation we obtain (14).

Remark 5.1. If we put $\rho = 0$ in the limiting PMF given by right side of the relation (14) we obtain the PGF of the usual Poisson distribution with parameter $\lambda > 0$.

Therefore, we have a reason to define the corresponding r.v. by the following definition.

Definition 5.1. We say that the r.v. Z has an *inflated-parameter Poisson distribution* with parameters $\lambda > 0$ and $\rho \in [0,1)$, and will denote this by $Z \sim IPo(\lambda,\rho)$, if its PGF $P_Z(t)$ is represented by the following equation

$$P_Z(t) = exp\left[\frac{\lambda(t-1)}{1-t\rho}\right]. \tag{15}$$

By analogy with the NB case, we will obtain the PMF of the inflated-parameter Poisson distribution by the following proposition.

Proposition 5.2. *The PMF of the $IPo(\lambda,\rho)$ distributed r.v. Z is given by the following relation*

$$P(Z=z) \;=\; \sum_{z_1,\,z_2,\ldots} \frac{e^{-\lambda}}{z_1!z_2!\ldots}[\lambda(1-\rho)]^{z_1+z_2+\cdots}\rho^{z_2+2z_3+\cdots}, \tag{16}$$

where $z=0,1,\ldots$ and the summation is over all nonnegative integers $z_1,z_2,z_3,\ldots$ such that

$$z_1+2z_2+3z_3+\cdots=z.$$

Proof. From the PGF (15) we have

$$\begin{aligned} P_Z(t) &= exp\left[\lambda(t-1)(1+\rho t+\rho^2t^2+\cdots)\right] \\ &= exp\left\{\lambda[-1+(1-\rho)t+\rho(1-\rho)t^2+\rho^2(1-\rho)t^3+\cdots]\right\}. \end{aligned}$$

Using the Taylor expansion of the exponential function $exp(x)$, we obtain

$$P_Z(t) \;=\; \sum_{n=0}^{\infty}\frac{\lambda^n[-1+(1-\rho)t+\rho(1-\rho)t^2+\rho^2(1-\rho)t^3+\cdots]^n}{n!},$$

i.e.

$$P_Z(t)=\sum_{n=0}^{\infty}\frac{\lambda^n}{n!}\sum_{n_0,\,n_1,\ldots}\binom{n}{n_0,n_1,\ldots}(-1)^{n_0}(1-\rho)^{n_1+n_2+\cdots}\rho^{n_2+2n_3+\cdots}t^{n_1+2n_2+3n_3+\cdots},$$

where the last summation is over all nonnegative integers $n_0,n_1,n_2,\ldots$, such that $n_0+n_1+n_2+\cdots=n$. Substituting in the last expression $n_i=z_i,\ i\geq 0$ and $n=z-\sum_{j=0}^{\infty}(j-1)z_j$ we obtain

$$P_Z(t) \;=\; \sum_{z=0}^{\infty}t^z\sum_{z_1,\,z_2,\ldots}\frac{[\lambda(1-\rho)]^{z_1+z_2+\cdots}\rho^{z_2+2z_3+\cdots}}{z_1!z_2!\ldots}\sum_{z_0=0}^{\infty}\frac{(-\lambda)^{z_0}}{z_0!},$$

where $z=0,1,\ldots$ and the last summation is over all nonnegative integers $z_1,z_2,z_3,\ldots$ such that $z_1+2z_2+3z_3+\cdots=z$. Therefore,

$$P_Z(t) \;=\; \sum_{z=0}^{\infty}t^zP(Z=z),$$

where the probability $P(Z=z)$ is given by (16) for $z\geq 0$.

Remark 5.2. From (16) we obtain the probabilities of the first four values of the r.v. $Z\sim IPo(\lambda,\rho)$, given by the following expressions

$$\begin{aligned} P(Z=0) &= e^{-\lambda}, \\ P(Z=1) &= e^{-\lambda}\lambda(1-\rho), \\ P(Z=2) &= e^{-\lambda}\lambda(1-\rho)\left[\tfrac{\lambda(1-\rho)}{2!}+\rho\right], \\ P(Z=3) &= e^{-\lambda}\lambda(1-\rho)\left[\tfrac{\lambda^2(1-\rho)^2}{3!}+\lambda(1-\rho)\rho+\rho^2\right]. \end{aligned}$$

Remark 5.3. We will show how to obtain the PMF (16) of the inflated-parameter Poisson r.v. from the PMF (12) of the inflated-parameter NB distribution by using the limiting conditions (13).

It is easy to show that the following equality is true

$$\binom{y_1 + y_2 + \cdots + r - 1}{y_1, y_2, \ldots, r - 1} = \prod_{i=1}^{\infty} \frac{(r - 1 + y_i)(r - 1 + y_i - 1) \cdots [r - 1 + y_i - (y_i - 1)]}{y_i!} = \prod_{i=1}^{\infty} A_i.$$

Then (12) can be represented as

$$P(Y = y) = \pi^r \sum_{y_1, y_2, \ldots} (1 - \rho)^{y_1 + y_2 + \cdots} \rho^{y_2 + 2y_3 + \cdots} A_1 (1 - \pi)^{y_1} \prod_{i=2}^{\infty} A_i (1 - \pi)^{y_i},$$

Under the limiting conditions (13) the following two relations

$$\lim_{r \to \infty} \lim_{\pi \to 1} \pi^r = \lim_{r \to \infty} \left(1 - \frac{\lambda}{r}\right)^r = e^{-\lambda} \quad \text{and} \quad \lim_{r \to \infty} (1 - \pi)^{y_i} = \frac{\lambda^{y_i}}{r^{y_i}}, \quad i = 1, 2, \ldots$$

are valid. Then

$$\lim_{r \to \infty} \lim_{\pi \to 1} A_1 \pi^r (1 - \pi)^{y_1} = \frac{e^{-\lambda} \lambda^{y_1}}{y_1!} \quad \text{and} \quad \lim_{r \to \infty} \lim_{\pi \to 1} A_i = \frac{\lambda^{y_i}}{y_i!}, \quad i = 2, 3, \ldots$$

Therefore,

$$\lim_{r \to \infty} \lim_{\pi \to 1} P(Y = y) = P(Z = z),$$

where $P(Z = z)$ is given by (16) for $z = 0, 1, \ldots$.

Remark 5.4. The mean and the variance of the $IPo(\pi, \lambda)$ distribution are given by

$$E(Z) = \frac{\lambda}{(1 - \rho)} \quad \text{and} \quad Var(Z) = \frac{\lambda(1 + \rho)}{(1 - \rho)^2}.$$

Remark 5.5. Let us note that the $IPo(\lambda, \rho)$ coincides with the Pólya-Aeppli distribution, introduced by Evans (1953).

6 Inflated-parameter Bernoulli Distribution

Until now, we defined the inflated-parameter geometric distribution, the inflated-parameter NB distribution (by summing r i.i.d. $IGe_0(\pi, \rho)$ r.v.'s) and the inflated-parameter Poisson distribution (from the $INB(\pi, \rho, r)$ distributed r.v. by using the limiting conditions (13)).

Our aim is to obtain the same inflated-parameter Poisson distribution (given by (15) or (16)), but starting from an appropriate defined *inflated-parameter Bernoulli distribution* (as in the classical theory).

Let us define the r.v. Q as follows:

$$\begin{aligned} P(Q = 0) &= 1 - \pi, \\ P(Q = k) &= \pi \rho^{k-1} (1 - \rho), \quad k = 1, 2, \ldots \end{aligned} \tag{17}$$

The above equations define a proper probability distribution *on the set of the nonnegative integers.*

Definition 6.1. We call the r.v. Q defined by (17) an *inflated-parameter Bernoulli distributed* with parameters $\pi \in (0,1)$ and $\rho \in [0,1)$, and we will denote this by $Q \sim IBe(\pi,\rho)$.

Remark 6.1. If $\rho = 0$, the inflated-parameter Bernoulli distribution $IBe(\pi,0)$ coincides with the usual Bernoulli distributed r.v. with parameter π, taking values 0 and 1.

From (17) we calculate the corresponding PGF $P_Q(t)$ given by the following expression

$$P_Q(t) = 1 - \frac{\pi(1-t)}{1-t\rho}. \tag{18}$$

7 Inflated-parameter Binomial Distribution

Now, if we sum n i.i.d. $IBe(\pi,\rho)$ r.v's we will obtain the r.v. B with the following PGF

$$P_B(t) = \left[1 - \frac{\pi(1-t)}{1-t\rho}\right]^n. \tag{19}$$

Definition 7.1. We call the r.v. B defined by the PGF (19) *inflated-parameter binomial distributed* with parameters $\pi \in (0,1)$, $\rho \in [0,1)$ and n, and will denote this by $B \sim IBi(\pi,\rho,n)$.

The next proposition represents the PMF of the $IBi(\pi,\rho,n)$ distributed r.v.

Proposition 7.1. *The PMF of the $IBi(\pi,\rho,n)$ distributed r.v. B is given by the following relation*

$$P(B=b) = (1-\pi)^n \sum_{b_1,b_2,\dots} \binom{n}{n-b_1-b_2-\cdots, b_1, b_2, \dots} \rho^b \left[\frac{\pi(1-\rho)}{(1-\pi)\rho}\right]^{b_1+b_2+\cdots}, \tag{20}$$

where $b = 0,1,\dots$ and the summation is over all nonnegative integers $b_1, b_2, b_3\dots$ such that

$$b_1 + 2b_2 + 3b_3 \cdots = b.$$

Proof. From the PGF (19) we consequently obtain

$$
\begin{aligned}
P_B(t) &= \left[1-\pi+\frac{\pi(1-\rho)t}{1-\rho t}\right]^n \\
&= (1-\pi)^n \sum_{k=0}^{n} \binom{n}{k} \left[\frac{\pi(1-\rho)\rho t}{\rho(1-\pi)(1-\rho t)}\right]^k \\
&= (1-\pi)^n \sum_{k=0}^{n} \binom{n}{k} \left[\frac{\pi(1-\rho)}{\rho(1-\pi)}(\rho t+\rho^2 t^2+\cdots)\right]^k \\
&= (1-\pi)^n \sum_{k=0}^{n} \binom{n}{k} \sum_{k_1,k_2,\ldots} \binom{k}{k_1,k_2,\ldots} \left[\frac{\pi(1-\rho)}{\rho(1-\pi)}\right]^{k_1+k_2+\cdots} (t\rho)^{k_1+2k_2+\cdots},
\end{aligned}
$$

where the last summation is over all nonnegative integers $k_1, k_2, \ldots$, such that $k_1+k_2+\cdots = k$. Substituting in the last expression $k_i = b_i,\ i \geq 1$ and $k = b - \sum_{j=1}^{\infty}(j-1)b_j$ we finally obtain

$$
P_B(t) = \sum_{b=0}^{\infty} t^b P(B=b),
$$

where the probability $P(B=b)$ is given by (20) for $b \geq 0$.

Remark 7.1. Representation (20) has the following equivalent form:

$$
P(B=b) = \sum_{b_1,b_2,\ldots} \binom{n}{n_0,b_1,b_2,\ldots} (1-\pi)^{n_0}[\pi(1-\rho)]^{b_1+b_2+\cdots} \rho^{b_2+2b_3+\cdots}, \tag{21}
$$

with $b = 0,1,\ldots$ and summation over all nonnegative integers $b_1, b_2, b_3, \ldots$, such that $b_1 + 2b_2 + 3b_3 + \cdots = b$, under condition that $n_0 + b_1 + b_2 + \cdots = n$.

Remark 7.2. Let us note that our $IBi(\pi,\rho,n)$ distributed r.v. can take *infinite number* of non-negative values, since by construction the corresponding inflated-parameter Bernoulli distributed r.v. can take all non-negative values.

Remark 7.3. The first four values of the probabilities of a r.v. $B \sim IBi(\pi,\rho,n)$ are given by the following expressions

$$
\begin{aligned}
&P(B=0) = (1-\pi)^n, \\
&P(B=1) = (1-\pi)^{n-1}\pi(1-\rho)n, \\
&P(B=2) = (1-\pi)^{n-2}\pi(1-\rho)\left[\binom{n}{2}\pi(1-\rho)+n(1-\pi)\rho\right], \\
&P(B=3) = (1-\pi)^{n-3}\pi(1-\rho)\left[\binom{n}{3}\pi^2(1-\rho)^2+n(n-1)(1-\pi)\pi(1-\rho)\rho+n(1-\pi)^2\rho^2\right].
\end{aligned}
$$

Now, by putting in (19)

$$
n \longrightarrow \infty \quad \text{and} \quad \pi \longrightarrow 0, \quad \text{such that} \quad n\pi = \lambda = const > 0, \tag{22}
$$

we would like to obtain in the limit the PGF (15) of the inflated-parameter Poisson distribution. Indeed,

$$\lim_{n\to\infty}\lim_{\pi\to 0} P_B(t) = \lim_{n\to\infty}\left[1 - \frac{\lambda(1-t)}{n(1-\rho t)}\right]^n = exp\left[\frac{\lambda(t-1)}{1-t\rho}\right].$$

Remark 7.4. The PMF (16) can be obtained from the PMF (21) by using the limiting conditions (22) also, but we omit these calculations (compare with Remark 5.3).

Remark 7.5. The mean and the variance of the $IBi(\pi,\rho,n)$ distribution are given by

$$E(B) = \frac{n\pi}{1-\rho} \quad \text{and} \quad Var(B) = \frac{n\pi(1-\pi+\rho)}{(1-\rho)^2}.$$

8 Inflated-parameter Logarithmic Series Distribution

Let $Y \sim INB(\pi,\rho,r)$ and its PGF $P_Y(t)$ is given by (11). From (12) we find that $P(Y = 0) = \pi^r$. Then $P(Y > 0) = 1 - \pi^r$. Now, let us consider the *truncated at zero* $INB(\pi,\rho,r)$ distributed r.v. Y_1. Its PMF is given by the following relation

$$P(Y_1 = y_1) = \frac{P(Y = y_1)}{1-\pi^r}, \quad y_1 = 1, 2, \ldots$$

The corresponding PGF $P_{Y_1}(t)$ has the form

$$P_{Y_1}(t) = \frac{1}{1-\pi^r}\left[P_Y(t) - \pi^r\right] = \frac{\pi^r}{1-\pi^r}\left\{\frac{(1-\rho t)^r - [1-t(1-\pi+\rho\pi)]^r}{[1-t(1-\pi+\rho\pi)]^r}\right\}.$$

Assuming $r \longrightarrow 0$ in the last expression, after using the L'Hôpital's rule, we obtain the relation

$$\lim_{r\to 0} P_{Y_1}(t) = ln\left[\frac{1-\rho t}{1-t(1-\pi+\rho\pi)}\right](-ln\pi)^{-1}.$$

If we denote by L the r.v. having the limiting PGF, then the following equality is fulfilled

$$P_L(t) = ln\left[1 + \frac{(1-\pi)(1-\rho)t}{1-t(1-\pi+\rho\pi)}\right](-ln\pi)^{-1}.$$

After simple transformations, the PGF $P_L(t)$ can be given by the following equivalent representation

$$P_L(t) = ln\left[1 - \frac{(1-\pi)(1-\rho)t}{1-\rho t}\right]^{-1}(ln\pi)^{-1}. \tag{23}$$

Remark 8.1. If we put $\rho = 0$ in the last expression, we derive the PGF of the usual logarithmic series distribution.

So, we are ready to give the following definition.

Definition 8.1. We say that the r.v. L has an *inflated-parameter logarithmic series distribution* with parameters $\pi \in (0,1)$ and $\rho \in [0,1)$, and we will denote this by $L \sim ILS(\pi,\rho)$, if its PGF $P_L(t)$ is given by (23).

The following proposition represents the PMF of the defined inflated-parameter logarithmic series distribution.

Proposition 8.1. *The PMF of the $ILS(\lambda,\rho)$ distributed r.v. L is given by the following relation*

$$P(L=l) = \sum_{l_1,l_2,\ldots} \frac{(-1+l_1+l_2+\cdots)!}{(-ln\pi)l_1!l_2!\ldots}[(1-\pi)(1-\rho)]^{l_1+l_2+\cdots}\rho^{l_2+2l_3+\cdots}, \tag{24}$$

where $l = 1,2,\ldots$ and the summation is over all nonnegative integers $l_1,l_2,l_3,\ldots$ such that

$$l_1 + 2l_2 + 3l_3 + \cdots = l.$$

Proof. From (23) we have

$$\begin{aligned} P_L(t) &= ln\left[1-(1-\pi)(1-\rho)t(1+\rho t+\rho^2t^2+\cdots)\right](ln\pi)^{-1} \\ &= -(ln\pi)^{-1}\sum_{i=0}^{\infty}\frac{[t(1-\pi)(1-\rho)(1+\rho t+\rho^2t^2+\cdots)]^{i+1}}{i+1}, \end{aligned}$$

after using the Taylor expansion of the logarithmic function $ln(1-x)$. Therefore,

$$P_L(t) = -(ln\pi)^{-1}\sum_{i=0}^{\infty}\frac{1}{i+1}\sum_{n_1,n_2,\ldots}\binom{i+1}{n_1,n_2,\ldots}[(1-\rho)(1-\pi)]^{n_1+n_2+\cdots}\rho^{n_2+2n_3+\cdots}t^{n_1+2n_2+3n_3+\cdots}.$$

The last summation is over all nonnegative integers $n_1,n_2,\ldots$, such that $n_1+n_2+\cdots = i+1$. Substituting $n_i = l_i$, $i \geq 1$ and $i+1 = l - \sum_{j=1}^{\infty}(j-1)l_j$ we obtain

$$P_L(t) = -(ln\pi)^{-1}\sum_{l=1}^{\infty}t^l\sum_{l_1,l_2,\ldots}\frac{(-1+l_1+l_2+\cdots)!}{l_1!l_2!\ldots}[(1-\rho)(1-\pi)]^{l_1+l_2+\cdots}\rho^{l_2+2l_3+\cdots},$$

i.e. we derived the PMF (24).

Proposition 8.2. *Let the r.v. $Y \sim INB(\pi,\rho,r)$. Then the following convergence is fulfilled*

$$\lim_{r\to 0} P(Y=y \mid Y\geq 1) = P(L=y),$$

where the r.v. $L \sim ILS(\pi,\rho)$.

Proof. Since $P(Y\geq 1) = 1-\pi^r$, we have

$$P(Y=y \mid Y\geq 1) = \frac{P(Y=y, Y\geq 1)}{P(Y\geq 1)} = \frac{P(Y=y,Y\geq 1)}{1-\pi^r}.$$

From the last relation and (12) we have

$$P(Y=y|Y\geq 1) = \frac{\pi^r}{1-\pi^r}\sum_{y_1,y_2,\ldots}\binom{y_1+y_2+\cdots+r-1}{y_1,y_2,\ldots,r-1}[(1-\pi)(1-\rho)]^{y_1+y_2+\cdots}\rho^{y_2+2y_3+\cdots}$$

$$= \frac{r\pi^r}{1-\pi^r}\sum_{y_1,y_2,\ldots}\frac{(y_1+y_2+\cdots+r-1)\cdots(r+2)(r+1)}{y_1!y_2!\ldots}[(1-\pi)(1-\rho)]^{y_1+y_2+\cdots}\rho^{y_2+2y_3+\cdots}.$$

Notice that according to the L'Hôpital's rule $\lim_{r\to 0}\frac{r\pi^r}{1-\pi^r} = -(ln\pi)^{-1}$. Then it can be seen that $\lim_{r\to 0} P(Y=y|Y\geq 1)$ converges to the PMF $P(L=y), y=1,2,\ldots$ of the inflated-parameter logarithmic series distribution as given by (24).

Remark 8.2. The classical NB and logarithmic series distributions are related with the same limiting results stated by the last two propositions, see Qu at *al.* (1990).

Remark 8.3. The probabilities of the first three values of the r.v. $L\sim ILS(\pi,\rho)$, are given by the following expressions

$$\begin{aligned} P(L=1) &= -(ln\pi)^{-1}(1-\rho)(1-\pi),\\ P(L=2) &= -(ln\pi)^{-1}(1-\rho)(1-\pi)\left[\tfrac{(1-\rho)(1-\pi)}{2}+\rho\right],\\ P(L=3) &= -(ln\pi)^{-1}(1-\rho)(1-\pi)\left[\tfrac{(1-\rho)^2(1-\pi)^2}{3}+(1-\rho)(1-\pi)\rho+\rho^2\right].\end{aligned}$$

Remark 8.4. The mean and the variance of of the r.v. $L\sim ILS(\pi,\rho)$ are given by the following expressions

$$E(L) = \frac{-(1-\pi)}{\pi(1-\rho)ln\pi} \quad \text{and} \quad Var(L) = \frac{-(1-\pi)[ln\pi(1+\pi\rho)+1-\pi]}{\pi^2(1-\rho)^2(ln\pi)^2}.$$

9 A Family of Inflated-parameter GPSD

Having in hands our inflated-parameter discrete distributions studied in the previous sections, it is natural to propose an "inflated-parameter" generalization of the GPSD. In this section we define the family of inflated-parameter GPSD. We give common representation of the PMF's and PGF's in the corresponding subsections. An overdispersed property of the new family is discussed and a new constructive interpretation of the additional parameter ρ is given. We will assume hereafter that $\rho\in[0,1)$.

9.1 Common Representation of the PMF's

One can observe that the PMF's of the inflated-parameter binomial, Poisson, negative binomial and logarithmic series distributions given by (21), (16), (12) and (24) correspondingly, have similar representations according to the additional parameter ρ. Therefore, we can expect a common expression of the corresponding PMF's, as states the following proposition.

Proposition 9.1. *The PMF's given by (21), (16), (12) and (24), correspondingly, have the following common representation*

$$P(N = k) = \frac{1}{g(\theta)} \sum_{k_1, k_2, \ldots} a(k)[\theta(1-\rho)]^{k_1+k_2+\cdots} \rho^{k_2+2k_3+\cdots}, \tag{25}$$

with $k = 0, 1, 2, \ldots$, $\rho \in [0,1), \theta > 0$, *and the summation is on the set of all nonnegative integers* $k_1, k_2, \ldots$, *such that* $k_1 + 2k_2 + \cdots = k$. *If the r.v.* $N \sim ILS(\theta, \rho)$, *its realizations begin from* 1 *and the summation in (25) is over the nonnegative integers, such that* $k_1 + 2k_2 + \ldots = k + 1$.

In the particular cases, the functions $a(k)$, $g(\theta)$ *and the parameter* θ, *are given by the following expressions*

$N \sim IBi(\theta, \rho, n)$: $a(k) = \binom{n}{n-k_1-k_2-\cdots, k_1, k_2, \ldots}$, $g(\theta) = (1+\theta)^n$, $\theta = \frac{\pi}{1-\pi}$;

$N \sim IPo(\theta, \rho)$: $a(k) = \frac{1}{k_1! k_2! \ldots}$, $g(\theta) = e^{\theta}$, $\theta = \lambda$,

$N \sim INB(\theta, \rho, r)$: $a(k) = \binom{k_1+k_2+\cdots+r-1}{k_1, k_2, \ldots, r-1}$, $g(\theta) = (1-\theta)^{-r}$, $\theta = 1 - \pi$,

$N \sim ILS(\theta, \rho)$: $a(k) = \frac{(-1+k_1+k_2+\cdots)!}{k_1! k_2! \ldots}$, $g(\theta) = -ln(1-\theta)$, $\theta = 1 - \pi$.

Proof. Using simple transformations one can obtain the above relations from (21), (16), (12) and (24), respectively.

Definition 9.1. The r.v. N belongs to the family of *Inflated-parameter GPSD* with parameters $\theta > 0$ and $\rho \in [0,1)$ if its PMF can be represented by (25).

Remark 9.1. Let us note that the defined family is different than the corresponding family studied by Gupta et *al.* (1995).

The following statement gives alternative expressions for the corresponding PMF's.

Proposition 9.2. *The PMF's of the* $IBi(\theta, \rho, n)$, $IPo(\theta, \rho)$, $INB(\theta, \rho, r)$ *and* $ILS(\theta, \rho)$ *distributed r.v.'s can be given by the following equivalent expressions*

$$N \sim IBi(\pi, \rho, n): \quad P(N = k) = \sum_{i=1}^{min(k,n)} \binom{n}{i} \binom{k-1}{i-1} [\pi(1-\rho)]^i (1-\pi)^{n-i} \rho^{k-i};$$

$$N \sim IPo(\lambda, \rho): \quad P(N = k) = e^{-\lambda} \sum_{i=1}^{k} \binom{k-1}{i-1} \frac{[\lambda(1-\rho)]^i \rho^{k-i}}{i!};$$

$$N \sim INB(\pi, \rho, r): \quad P(N = k) = \pi^r \sum_{i=1}^{k} \binom{k-1}{i-1} \binom{r+i-1}{i} [(1-\pi)(1-\rho)]^i \rho^{k-i};$$

$$N \sim ILS(\pi, \rho): \quad P(N = m) = (-ln\pi)^{-1} \sum_{i=1}^{m} \binom{m-1}{i-1} \frac{[(1-\pi)(1-\rho)]^i \rho^{m-i}}{i},$$

where $k = 0, 1, \ldots, m = 1, 2, \ldots$ *and it is assumed that* $\sum_{i=1}^{0} = 1$.

Proof. We will demonstrate how to obtain the PMF for the r.v. $N \sim IPo(\lambda, \rho)$. The remaining expressions can be deduced in a similar way.

The starting point here is to use the following relation

$$\frac{1}{(1-y)^j} = \sum_{l=0}^{\infty} \binom{l+j-1}{l} y^l, \quad 0 < y < 1, \tag{26}$$

valid for any $j = 1, 2, \ldots$.

For the PGF (15) we have

$$P_N(t) = exp\left\{\lambda[-1 + (1-\rho)t + \rho(1-\rho)t^2 + \rho^2(1-\rho)t^3 + \cdots]\right\}.$$

Using the Taylor expansion of the exponential function $exp(x)$, after some algebra we obtain

$$P_N(t) = e^{-\lambda}\left\{1 + \sum_{j=1}^{\infty} \frac{[\lambda(1-\rho)t]^j}{j!} \frac{1}{(1-\rho t)^j}\right\},$$

Applying (26) in the last expression we have

$$\sum_{k=0}^{\infty} P(N=k)t^k = e^{-\lambda}\left\{1 + \sum_{j=1}^{\infty} \frac{[\lambda(1-\rho)t]^j}{j!} \sum_{l=0}^{\infty} \binom{l+j-1}{l} \rho^l t^l\right\}.$$

The PMF of the r.v. $N \sim IPo(\lambda, \rho)$ given by the proposition is obtained by equating the coefficients of t^k on both sides of the last equality for fixed $k = 0, 1, 2, \ldots$.

9.2 Common Representation of the PGF's

Let $P_N(t)$ and $P_X(t)$ be PGF's of the non-negative integer valued r.v.'s N and X, respectively. Let N denote the number of claims and let X_i denote the amount of the i-th claim, $i = 1, 2, \ldots$. Let $\{X_1, X_2, \ldots\}$ be a sequence of i.i.d. r.v.'s, with common PGF $P_X(t)$. Assume that the X_i's are also independent on N and consider the random sum

$$S = X_1 + X_2 + \cdots + X_N,$$

with the convention that $S = 0$ if $N = 0$. Then S equals aggregate claims, and the corresponding PGF $P_S(t) = P_N(P_X(t))$, see for example Bower et *al.* (1997). This situation describes the portfolio of insurance policies during a given length of time.

Now, if N belongs to the family of GPSD with parameter θ defined by (1) and X is an arbitrary discrete distribution, then the resulting random sum S has a PGF given by the following expression

$$P_S(t) = \frac{g(\theta P_X(t))}{g(\theta)}, \tag{27}$$

where the possible choices of the function $g(\theta)$ are given by Proposition 9.1, see Hirano et *al.* (1984).

According to Proposition 9.1 the inflated-parameter binomial, Poisson, NB and logarithmic series distributions have a common representation for their PMF's. Therefore, one can expect that they have the corresponding common representation of their PGF's. This is precised by the following statement, which gives the PGF of the inflated-parameter GPSD defined by Definition 9.1.

Proposition 9.3. *The PGF of the inflated-parameter GPSD is given by (27), where the functions $g(\theta)$ are given by Proposition 9.1 and*

$$P_X(t) = \frac{t(1-\rho)}{1-t\rho}. \tag{28}$$

Proof: Using simple transformations from the corresponding functions $g(\theta)$, given by Proposition 9.1, relations (27) and (28), one can get easy the PGF's (19), (15), (11) and (23), respectively.

Remark 9.3. In fact, Proposition 9.3 gives a constructive representation of the distributions belonging to the family of inflated-parameter GPSD. Indeed, (28) is the PGF of the geometric distribution with parameter $1-\rho$ and taking positive integer values.

In terms of the collective risk model, this means that the aggregated claim S has inflated-parameter GPSD when the individual claims have geometric distribution with parameter $1-\rho$, i.e. $X_i \sim Ge_1(1-\rho)$, and number of claims N for a given length of time is a r.v. belonging to the usual family of GPSD with parameter θ.

Remark 9.4. Proposition 9.3 gives, in fact, a new interpretation of the additional parameter ρ (being a parameter of geometric distribution), different than "zero-inflated" proportion and correlation coefficient, as discussed earlier.

Remark 9.5. From (27) and (1) it is easy to establish, that the variance-mean ratio of the inflated-parameter GPSD is greater than the corresponding variance-mean ratio of the original GPSD, i.e. our new family is *overdispersed* according the family of GPSD, if the additional parameter $\rho \in (0,1)$.

10 Applications

In this section we will approximate the frequency data given in the column headed "Observed" of the Table 10.1 by using $IPo(\lambda, \rho)$ and $INB(\pi, \rho, r)$ distributions. The statistics are taken from Daykin et *al.* (1994) p. 52, and relate to claims under UK comprehensive motor policies. The 421240 policies were classified according to the number of claims in the year 1968.

Let us denote by $\bar{X}_n$ and σ_n^2 the sampling mean and variance. Then the average number of claims per policy is $\bar{X}_n = 0.13174$ and $\sigma_n^2 = 0.13852$.

In the column headed "Poisson" of the Table 10.1 are given the corresponding expected values by using the usual Poisson distribution with a parameter $\lambda = \bar{X}_n = 0.13174$. The column of the Table 10.2 headed "NB" sets out the resulting NB approximation with parameters $\pi = 0.951$ and $r = 2.558$. The last rows show the corresponding values of the Pearon's χ^2.

The value of the chi-square in the Poisson case is too high, so the insufficiency of the Poisson law for the data is evident. The reason is that the sampling variance is greater than the sampling mean, whereas they should be almost equal if the Poisson law were valid. The value of χ^2 in the NB case is 9.18 which gives probability 0.05 for 5 degrees of freedom, so that the representation is acceptable.

We will not discuss here the Maximum Likelihood (ML) estimates of the parameters and their properties, but they can be calculated numerically. Here we give the corresponding results for comparison only. The ML estimates are obtained by a direct minimization approach of the log-likelihood following Mickey and Britt (1974). The minimization procedure is based on derivative-free algorithm for nonlinear least squares proposed by Ralston and Jennrich (1978).

10.1 $IPo(\lambda, \rho)$-case

The mean and variance of the $IPo(\lambda, \rho)$ distribution are given by Remark 5.4. Solving the corresponding system we obtain the following moment estimates for the parameters λ and ρ:

$$\hat{\rho} = \frac{\sigma_n^2 - \bar{X}_n}{\sigma_n^2 + \bar{X}_n} \quad \text{and} \quad \hat{\lambda} = \frac{2\bar{X}_n^{\,2}}{\sigma_n^2 + \bar{X}_n}.$$

Remark 10.1. The same estimates can be found, for example, in Johnson et *al.* (1992) pp. 381-382, where they were reported for the Pólya-Aeppli distribution, see Remark 5.5.

In our case we obtain the values

$$\hat{\rho} = 0.0251 \quad \text{and} \quad \hat{\lambda} = 0.12843.$$

The corresponding expected values are given in the column headed "IPo" of the Table 10.1.

Table 10.1. Poisson case

k	Observed	Poisson	IPo	IPo-ML
0	370412	369246.88	370469.93	370435.30
1	46545	48643.57	46385.30	46447.48
2	3935	3204.09	4068.21	4045.88
3	317	140.70	296.20	291.57
4	28	4.63	19.14	18.61
5	3	0.12	1.13	1.09
≥ 6	0	0.01	0.07	0.06
	Chi-square	667.52	13.60	13.61

The comparison of the expected values given in the columns headed "Poisson" and "IPo" shows that $IPo(\lambda, \rho)$ distribution fits the observed frequencies much better than the usual

Poisson distribution, which has a shorter tail than the data. The value of χ^2 is 13.60 which gives probability 0.04 for 5 degrees of freedom, so the approximation by $IPo(\lambda, \rho)$ distribution is acceptable (observe that the use of the NB distribution is preferable if our criterion is the value of the χ^2 statistics).

Remark 10.2. We calculated a positive value for the moment estimate of the parameter ρ. This can be interpreted in the following way: the observed number of zeros is more than it can be predicted by the usual Poisson distribution (as it can be seen from the Table 10.1).

We obtain the following ML estimates

$$\hat{\rho}_{ML} = 0.02441 \quad \text{and} \quad \hat{\lambda}_{ML} = 0.12852$$

with $\chi^2 = 13.61$. One can see that the ML estimates of the parameters are close to the values of the corresponding moment estimates. The corresponding expected values are given in the last column of the Table 10.1.

10.2 $INB(\pi, \rho, r)$-case

To estimate the parameters of the $INB(\pi, \rho, r)$ distribution we need additionally the third moment together with sampling mean and variance.

The mean and the variance of the r.v. $X \sim INB(\pi, \rho, r)$ are given by Remark 4.3. From the PGF (11), after some algebra we obtain the following relation for the third moment

$$E(X^3) = \frac{r(1-\pi)}{\pi}\left[1 + 4\rho + \rho^2 + \frac{3(r+1)(\rho+1)(1-\pi)}{\pi} + \frac{(r+1)(r+2)(1-\pi)^2}{\pi^2}\right].$$

The solution of the corresponding system gives the following procedure for calculation the moment estimates of the parameters.

Step 1. The moment estimate of the parameter ρ is a solution of the following quadratic equation

$$a\rho^2 + b\rho + c = 0,$$

where

$$a = \left(\bar{X}_n + \frac{\sigma_n^2}{\bar{X}_n}\right)\left(\bar{X}_n + \frac{2\sigma_n^2}{\bar{X}_n}\right) - \frac{m_3 - \sigma_n^2}{\bar{X}_n},$$

$$b = 2 - 2\left(\bar{X}_n + \frac{\sigma_n^2}{\bar{X}_n}\right)\left(\bar{X}_n + \frac{2\sigma_n^2}{\bar{X}_n}\right) + 2\frac{m_3}{\bar{X}_n},$$

and

$$c = \left(\bar{X}_n + \frac{\sigma_n^2}{\bar{X}_n}\right)\left(\bar{X}_n + \frac{2\sigma_n^2}{\bar{X}_n}\right) - \frac{m_3 + \sigma_n^2}{\bar{X}_n}$$

with m_3 being the third sample moment;

Step 2. The moment estimate for the parameter π is given by

$$\hat{\pi} = \left[\frac{\sigma_n^2}{\bar{X}_n}(1 - \hat{\rho}) - \hat{\rho}\right]^{-1},$$

where $\hat{\rho}$ is the result from Step 1;

Step 3. Finally, the moment estimate of the parameter r can be calculated by the following formula

$$\hat{r} = \frac{\bar{X}_n \hat{\pi}(1-\hat{\rho})}{1-\hat{\pi}},$$

where $\hat{\rho}$ and $\hat{\pi}$ are the calculated values from Step 1 and Step 2, correspondingly.

For the considered data we obtain the following moment estimates

$$\hat{\rho} = -.03869, \quad \hat{\pi} = 0.88428 \quad \text{and} \quad \hat{r} = 1.04564.$$

In the column headed "INB" of the Table 10.2 are given the corresponding estimated frequencies, when using the computed moment estimates of the parameters. One can see that the $INB(\pi, \rho, r)$ distribution fits the observed frequencies perfectly.

Table 10.2. NB case

k	Observed	NB	INB	INB-ML
0	370412	370459.94	370409.99	370412.37
1	46545	46413.30	46553.37	46545.63
2	3935	4043.97	3922.17	3928.82
3	317	300.92	325.21	324.23
4	28	20.48	26.85	26.58
5	3	1.32	2.21	2.17
≥ 6	0	0.09	0.20	0.19
	Chi-square	9.18	0.78	0.76

Remark 10.3. We calculated a negative value for the moment estimate of the parameter ρ. This is possible (see Section 2.2) and can be interpreted in the following way: the observed number of zeros is less than predicted by the classical NB distribution. Let us note, that in the Poisson case we observed just the opposite situation (compare with Remark 10.2).

We obtain the following ML estimates for the parameters

$$\hat{\rho}_{ML} = -.03670, \quad \hat{\pi}_{ML} = 0.88748 \quad \text{and} \quad \hat{r}_{ML} = 1.07727,$$

with $\chi^2 = 0.76$. The corresponding estimated frequencies are given in the last column of the Table 10.2.

Remark 10.4. The computer program code in FORTRAN for computation of the corresponding PMF's is available from the third author upon request.

11 Conclusions

In this paper we introduce extensions of some classical univariate discrete distributions. This can be considered as a new method for adding a parameter to a family of distributions. The natural interpretation of the additional parameter $\rho \in [0, 1)$ being "zero-inflated" parameter (see Section 2.2), correlation coefficient (see Remark 3.5) and parameter of a mixing geometric distribution (see Remark 9.3), gives possibility to use the proposed class of inflated-parameter GPSD for modeling dependent count or frequency data structures, which naturally appear in Insurance, Finance and Economics. The corresponding variance-mean ratios show that the inflated-parameter distributions are overdispersed according to their univariate analogue. The results in Section 9 show that it is possible to define different classes of extended GPSD, taking a mixing discrete distribution, different than the geometric one. The simulation results from the last two subsections show that the inclusion of the additional parameter ρ improves significantly the corresponding approximations of our frequency data when using $IPo(\lambda, \rho)$ and $INB(\pi, \rho, r)$ distributions.

This paper is only a starting point, giving a theoretical basis for the distributions that belong to the new class of inflated-parameter GPSD. We are sure, the additional parameter ρ will lead to a second-order difference equation, which will help to estimate effectively the tails of distributions. Some future investigations related with this topic as well as with some statistical score-tests are currently in progress.

12 Acknowledgment

This work is sponsored by the CKER of the Society of Actuaries.

REFERENCES

Bowers, N.L,, Gerber, H.U., Hickman, J.C., Jones, D.A. and Nesbitt, C.J. (1997). *Actuarial Mathematics*, 2nd ed. Schaumburg, Ill.: Society of Actuaries.

Chatfield, C., Ehrenberg, A.S.C., and Goodhart, G.J. (1966). Progress on a simplified model of stationary purchasing behavior (with discussion), *J. of the Royal Statistical Society* **A 129**, 317-367.

Consul, P.C. (1990). New class of location-parameter discrete probability distributions and their characterizations, *Communications in Statistics: Theory and Methods* **19 (12)**, 4653-4666.

Collett, D. (1991). *Modelling Binary Data*. Chapman and Hall: London.

Daykin, C.D., Pentikäinen, T. and Pesonen, M. (1994). *Practical Risk Theory for Actuaries*, Chapman & Hall.

Dhaene, J., Willmot, G. and Sundt, B. (1996). Recursions for distributions functions and stop-loss transforms, Proc. *XXVII ASTIN Colloquium*, Copenhagen, 497-515.

Evans, D.A. (1953). Experimental evidence concerning contagious distributions in ecology, *Biometrika* **4**, 186-211.

Galambos, J. and Kotz, S. (1978). *Characterization of Probability Distributions.* Lecture Notes in Math. 675. Springer-Verlag, New York.

Gupta, R.C. (1974). Modified power series distributions and some of its applications, *Sankhyā: Series B* **35**, 288-298.

Gupta, L.P., Gupta, R.C. and Thripati, R.C. (1995). Inflated modified power series distributions with applications, *Communications in Statistics: Theory and Methods* **24 (9)**, 2355-2374.

Hirano, K., Kuboki, H., Aki, S. and Kyribayashi, K. (1984). *Figures of Probability Density Functions - Discrete Univariate Case*, Computer Science Monographs, Tokyo.

Johnson, N.L., Kotz, S. and Kemp, A.W. (1992). *Univariate Discrete Distributions.* Wiley Series in Probability and Mathematical Statistics. 2nd edition.

Luceño, A. (1995). A family of partially correlated Poisson models for over-dispersion, *Computational Statistics & Data Analysis* **20**, 511-520.

Panjer, H.H. (1981). Recursive evaluation of a family of compound distributions. *ASTIN Bulletin* **12**, 22-26.

Patil, G.P. (1962). Certain properties of the generalized power distributions, *Ann. Inst. Stat. Math.* **14**, 179-182.

Qu, Y., Beck, G.J. and Williams, G.W. (1990). Polya-Eggenberhger distribution: parameter estimation and hypothesis tests, *Biom. J.* **32**, 229-242.

Rolski, T., Schmidli, H., Schmidt, V. and Teugels, J. (1999). *Stochastic Processes for Insurance and Finance*, John Wiley & Sons, Chichester.

Schrörter, K.J. (1990). On a family of counting distributions and recursions for related compound distributions, *Scand. Actuarial J.*, 161-175.

Straub, E. (1988). *Non-life Insurance Mathematics*, Springer-Verlag.

Sundt, B. (1992). On some extensions of Panjer's class of counting distributions, *ASTIN Bulletin* **22**, 61-80.

Sundt, B. and Jewell, W.S. (1981). Further results on recursive evaluation of compound distributions, *ASTIN Bulletin* **12**, 27-39.

Winkelmann, R. (2000). *Econometric Analysis of Count Data*, 3rd Edition. Springer-Verlag: Berlin.