Insurance: Mathematics and Economics 48 (2011) 406–412

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

journal homepage: www.elsevier.com/locate/ime

A new discrete distribution with actuarial applications

Emilio Gómez-Déniz [a,*], José María Sarabia [b], Enrique Calderín-Ojeda [a]

[a] Department of Quantitative Methods, University of Las Palmas de Gran Canaria, Gran Canaria, Spain

[b] Department of Economics, University of Cantabria, Santander, Spain

ARTICLE INFO

Article history:
Received September 2010
Received in revised form
January 2011
Accepted 16 January 2011

MSC:
60E05
62E99
60J80

Keywords:
Claim
Compound
Geometric distribution
Overdispersion
q-series
Recursion
Unimodality

ABSTRACT

A new discrete distribution depending on two parameters, $\alpha < 1, \alpha \neq 0$ and $0 < \theta < 1$, is introduced in this paper. The new distribution is unimodal with a zero vertex and overdispersion (mean larger than the variance) and underdispersion (mean lower than the variance) are encountered depending on the values of its parameters. Besides, an equation for the probability density function of the compound version, when the claim severities are discrete is derived. The particular case obtained when α tends to zero is reduced to the geometric distribution. Thus, the geometric distribution can be considered as a limiting case of the new distribution. After reviewing some of its properties, we investigated the problem of parameter estimation. Expected frequencies were calculated for numerous examples, including short and long tailed count data, providing a very satisfactory fit.

1. Introduction

The development of parametric families of discrete distributions, which describe count phenomena, and the study of their properties have been a persistent theme of 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. Count data occur in many practical problems, for example the number of events such as insurance claims, the number of kinds of species in ecology, etc.

Most frequencies of event occurrence can be described initially by a Poisson distribution. Nevertheless, 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. In view of this, alternative probability distributions, such as the negative binomial and generalized Poisson, among others, are preferred for modeling the phenomena under study.

Many of these phenomena, such as individual automobile insurance claims, are characterized by two features: (1) Overdispersion, i.e., the variance is greater than the mean; (2) Zero-inflated, i.e. the presence of a high percentage of zero values in the empirical distribution. In view of this, many attempts have been made in statistical literature, and particularly in the actuarial field, to find a probabilistic model for the distribution of the number of counts.

In this paper, a new discrete distribution is introduced and obtained from the cumulative distribution function (cdf) defined as follows. For a random variable N taking non-negative integer $\{0, 1, \ldots\}$, the expression

$$P_n = \Pr(N \leq n) = 1 - \frac{\log(1-\alpha\theta^{n+1})}{\log(1-\alpha)}, \tag{1}$$

is a genuine cdf, depending on two parameters, $\alpha < 1, \ \alpha \neq 0$ and $0 < \theta < 1$, since it is simple to see that it is non-negative and strictly increasing and goes to one when n goes to infinity.

This new distribution can be considered as an alternative one to the negative binomial, Poisson-inverse Gaussian, hyper-Poisson and generalized Poisson distributions, among others. As we will see later, the new distribution is unimodal with a zero vertex.

By taking in (1) limit when the parameter α tends to zero and applying L'Hospital's rule it is easy to derive that (1) is reduced to $P_n = 1 - \theta^{n+1}$, i.e. the cdf of the geometric distribution with parameter $0 < \theta < 1$ and probability mass function (henceforth, pmf) $\Pr(N = n) = (1-\theta)\theta^n$. Therefore, since the geometric distribution is obtained as a limiting case, it constitutes another

* Corresponding author. Tel.: +34 928451803; fax: +34 928451829.
E-mail address: egomez@dmc.ulpgc.es (E. Gómez-Déniz).

0167-6687/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.insmatheco.2011.01.007

distribution related to the geometric distribution; see, for example Jain and Consul (1971), Philippou et al. (1983), Tripathi et al. (1987), Kemp (2008), Makčutek (2008) and Gómez-Déniz (2010).

The new discrete distribution proposed here presents the two-fold characteristic stated above: it has a zero vertex and can be overdispersed; therefore, it is a candidate for fitting phenomena of this nature. Both the examples of real data provided here and the comparison with other distributions in the literature show that the distribution has an outstanding performance.

One of the advantages of the new distribution is its simplicity (its pmf contains no special function) and flexibility.

To the best of our knowledge, the discrete distribution presented here has not been previously addressed in statistical and actuarial literature.

The structure of the paper is as follows. In Section 2, we present the new discrete distribution and some of its properties. Also, some results obtained for the particular case $0 < \alpha < 1$ are presented, including an equation for the pmf of the compound version, when the claim severities are discrete. Section 3 addresses the estimation of the two parameters of the new distribution. In Section 4 the expected frequencies are calculated for numerous examples, and the distribution was found to provide a very satisfactory fit. In the final section, some conclusions are drawn.

2. The new distribution and its properties

2.1. The general case

In this section we define the new probability mass function and its properties. We begin by taking into account that for a random variable N taking non-negative integers $\{0, 1, \ldots\}$ with cdf as in (1), then the pmf is given by

$$p_n = \Pr(N = n) = \frac{\log(1-\alpha\theta^n) - \log(1-\alpha\theta^{n+1})}{\log(1-\alpha)}. \tag{2}$$

From (1) we have that the survival function of a random variable N with pmf (2) is given by

$$\bar{P}_n = 1 - P_{n-1} = \frac{\log(1-\alpha\theta^n)}{\log(1-\alpha)}. \tag{3}$$

It is well known that the unimodality property is a significant feature in many statistical distributions. The following result shows that the new discrete distribution is unimodal with a zero vertex.

Proposition 1. *The pmf given in (2) is unimodal with a modal value at $n = 0$.*

Proof. Letting (2) define p_n also for non-integer values of n, then we obtain that for $n \geq 0$

$$\begin{aligned} p'_n \log(1-\alpha) &= \frac{\mathrm{d}}{\mathrm{d}n}(\log(1-\alpha\theta^n) - \log(1-\alpha\theta^{n+1})) \\ &= -\frac{\alpha\theta^n \log\theta}{1-\alpha\theta^n} + \frac{\alpha\theta^{n+1}\log\theta}{1-\alpha\theta^{n+1}} \\ &= -\alpha\theta^n \log\theta \left(\frac{1}{1-\alpha\theta^n} - \frac{\theta}{1-\alpha\theta^{n+1}}\right) \\ &= -\alpha\theta^n \log\theta \frac{1-\theta}{(1-\alpha\theta^{n+1})(1-\alpha\theta^n)}, \end{aligned}$$

that is,

$$p'_n = \frac{-\alpha}{\log(1-\alpha)} \frac{(1-\theta)\theta^n \log\theta}{(1-\alpha\theta^{n+1})(1-\alpha\theta^n)} < 0.$$

Hence, p_n is decreasing. □

For further results concerning the unimodality of discrete distributions, see also Keilson and Gerber (1971), Medgyessy (1972) and Abouammoh (1987).

The mean value can be written as

$$\mu(\alpha, \theta) = E(N) = \frac{1}{\log(1-\alpha)} \sum_{n=1}^{\infty} \log(1-\alpha\theta^n). \tag{4}$$

Observe that (4) can be rewritten as

$$E(N) = \frac{\log((\alpha\theta; \theta)_\infty)}{\log(1-\alpha)},$$

where $(a; q)_\infty$, the q-Pochhammer symbol (see Askey, 1980, p. 347 and Andrews et al., 1999, p. 488), is defined as

$$(a; q)_\infty = \prod_{k=0}^{\infty}(1-aq^k), \quad 0 < q < 1, \tag{5}$$

$(a; q)_\infty$ is also called q-Pochhammer symbol, since it is a q analog of the usual Pochhammer symbol. The q-Pochhammer symbol is quickly evaluated and readily available in standard software such as Mathematica 7.0.

The second moment around the origin is

$$E(N^2) = \frac{1}{\log(1-\alpha)} \sum_{n=1}^{\infty}(2n-1)\log(1-\alpha\theta^n). \tag{6}$$

Therefore, the variance can be written as

$$\mathrm{Var}(N) = \frac{1}{\log(1-\alpha)} \sum_{n=1}^{\infty}(2n-1)\log(1-\alpha\theta^n) - (\mu(\alpha,\theta))^2.$$

Now, we have that

$$\frac{\partial\mu(\alpha,\theta)}{\partial\theta} = \frac{-\alpha}{\log(1-\alpha)} \sum_{n=0}^{\infty} \frac{(n+1)\theta^n}{1-\alpha\theta^{n+1}} > 0. \tag{7}$$

On the other hand,

$$\begin{aligned} &\frac{\partial\mu(\alpha,\theta)}{\partial\alpha} \\ &= \frac{1}{(1-\alpha)\log(1-\alpha)} \left(\mu(\alpha,\theta) - (1-\alpha)\sum_{n=0}^{\infty} \frac{\theta^{n+1}}{1-\alpha\theta^{n+1}}\right) \\ &= \frac{1}{(1-\alpha)\log^2(1-\alpha)} \\ &\times \sum_{n=0}^{\infty} \frac{(1-\alpha\theta^{n+1})\log(1-\alpha\theta^{n+1}) - (1-\alpha)\theta^{n+1}\log(1-\alpha)}{1-\alpha\theta^{n+1}}. \end{aligned} \tag{8}$$

Let us now consider the following function $g(\alpha) = (1-\alpha\theta^{n+1})\log(1-\alpha\theta^{n+1}) - (1-\alpha)\theta^{n+1}\log(1-\alpha)$ for fixed $0 < \theta < 1$. It is easy to see that $\frac{\mathrm{d}g(\alpha)}{d\alpha} = \theta^{n+1}(\log(1-\alpha) - \log(1-\alpha\theta^{n+1}))$, this expression is positive for $\alpha < 0$ and negative for $0 < \alpha < 1$. Therefore, $g(\alpha)$ is a decreasing function for $0 < \alpha < 1$ and an increasing function for $\alpha < 0$. Now, having into account that $g(0^+) = g(0^-) = 0$, $g(1^-) = (1-\theta^{n+1})\log(1-\theta^{n+1}) < 0$, $g(-\infty) = -\infty$ as it is easily verified, we conclude that (8) is always negative.

In conclusion, we have that $\frac{\partial\mu(\alpha,\theta)}{\partial\alpha} < 0$ for $\alpha < 1$, $\alpha \neq 0$. Therefore, the mean increases with θ and decreases with α. Furthermore, the mean value tends to 0 when α tends to 1 to the left and to $\theta/(1-\theta)$, the mean of the geometric distribution with parameter $0 < \theta < 1$, when α tends to zero.

The quantile n_γ of (2) can be obtained from (1) and it is given by

$$n_\gamma = \left[\frac{1}{\log\theta}\log\left(\frac{1-(1-\alpha)^{1-\gamma}}{\alpha}\right) - 1\right],$$

where $[\cdot]$ denotes the integer part. In particular, the median is

$$n_{0.5} = \left[\frac{1}{\log\theta}\log\left(\frac{1-\sqrt{1-\alpha}}{\alpha}\right) - 1\right].$$