Statistical Methodology 29 (2016) 32–50

Contents lists available at ScienceDirect

Statistical Methodology

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

A two-parameter general inflated Poisson distribution: Properties and applications

Athanasios C. Rakitzis [a,*], Philippe Castagliola [a], Petros E. Maravelakis [b]

[a] *LUNAM Université, Université de Nantes & IRCCyN UMR CNRS 6597, Nantes, France*

[b] *Department of Business Administration, University of Piraeus, 80 Karaoli & Dimitriou Street, 18534, Piraeus, Greece*

HIGHLIGHTS

- A new two-parameter general inflated Poisson distribution is proposed and studied.
- The proposed model can be effectively used for fitting underdispersed and overdispersed count data.
- A general mechanism for inflating the zero and non-zero values of any discrete distribution is offered.
- Real-data examples illustrate the practical usefulness of the proposed model.

ARTICLE INFO

Article history:
Received 7 March 2015
Received in revised form 16 October 2015
Accepted 26 October 2015
Available online 3 November 2015

MSC:
62F99
62P99
62E15

Keywords:
Count data
EM algorithm
Maximum likelihood estimation
Overdispersion
Poisson distribution
Underdispersion

ABSTRACT

In this work, we propose and study a two-parameter modification of the ordinary Poisson distribution that is suitable for the modeling of non-typical count data. This model can be viewed as an extension of the zero-inflated Poisson distribution. We derive the proposed model as a special case of a general one and focus our study on it. The theoretical properties for each model are given, while estimation methods for the two-parameter model are discussed as well. Three practical examples illustrate its usefulness. The results show that the proposed model is very flexible in the modeling of various types of count data.

* Corresponding author.
E-mail addresses: arakitz@aegean.gr (A.C. Rakitzis), philippe.castagliola@univ-nantes.fr (P. Castagliola), maravel@unipi.gr (P.E. Maravelakis).

http://dx.doi.org/10.1016/j.stamet.2015.10.002
1572-3127/

1. Introduction

The Poisson distribution is a widely used probability model in the description of count related data that arise very frequently across many disciplines. When a large number of zero counts is present in the data, there is evidence of overdispersion and the ordinary Poisson distribution is not an appropriate model, because it often underestimates the observed dispersion.

The use of the zero-inflated Poisson (ZIP) distribution has been considered by several authors as an appropriate model for count data with an excessive number of zeros. Xie and Goh [31] and Xie et al. [32] used the ZIP distribution in the context of statistical quality control, in order to describe a zero-defect process, subject to random shocks. According to this model, the random shocks will cause some defects (non-conforming items) and the number of defects follows a Poisson distribution with parameter λ. Note also that the shocks occur independently with parameter $1 - \phi$, where $\phi \in [0, 1]$ is the zero-inflated parameter. Other applications of the ZIP model can be found in manufacturing (Lambert [17], Bae et al. [1]), epidemiology (Bohning [4]), biology (Yip [33]), healthcare (Brown et al. [5]), insurance (Yip [34]) and accident data (Lord et al. [18], Ullah et al. [29]). Bohning [3] provided several practical examples where the ZIP distribution is the appropriate model for the available data. Also, he discussed the numerical aspects of parameter estimation for the ZIP distribution, using the *Computer Assisted Analysis of Mixtures* (C.A.MAN) package.

A natural extension of the ZIP model is to assume that random shocks influence not only the zero values but also all the first $r + 1$ values $\{0, 1, \dots, r\}$ of a Poisson distribution with parameter λ. The idea is that instead of a single inflated parameter $\phi_0 \in [0, 1]$, which only impacts on the smallest value $x = 0$, there are $r + 1$ inflated parameters $\phi_0, \phi_1, \dots, \phi_r$, which impact on the first $r + 1$ smallest values $x = 0, \dots, r$. When data are not inflated only at "zero", they are usually referred to as "non-zero" or as "general inflated" models. It seems that Yoneda [35] was the first who studied a modification of the Poisson distribution with arbitrary probabilities to the first $r + 1$ values and the remaining probabilities proportional to $e^{-x}\lambda^x/x!$. Padney [22] proposed the Poisson distribution inflated at point "8", as an appropriate model for the number of flowers of the plant *Primula Veris*. Murat and Szynal [21] considered modified Power Series distributions inflated at point r, $r > 0$, and extended the results of Gupta et al. [13] who studied the distributional properties of zero-inflated Power Series distributions. Melkersson and Rooth [20] proposed the *zero-two inflated* Poisson (ZTIP) distribution (i.e., their model is only inflated at "0" and "2") as an appropriate model for describing the Swedish women's fertility. Recently, Begum et al. [2] proposed and studied an inflated Poisson distribution (called *Generalized Inflated Poisson*), considering also special cases of it in order to fit the data of Melkersson and Rooth [20].

The main motivation of our work is to develop an appropriate discrete probability model that generalizes the ZIP model while it can be used for modeling unusual count data, i.e., data with various level of dispersion or non-unimodal data. Moreover, it would be desirable to have the least possible number of model parameters, in order to keep things simple. According to Weiß [30] (see also Sellers et al. [26]), when the equidispersion property of the Poisson distribution is violated, there is a vast amount of research articles that proposes several alternative models in order to handle the overdispersion of the data but there are only few, when underdispersion is present. Moreover, we will focus on two-parameter models, in order to have only one additional parameter that controls the degree of inflation in all the first $r + 1$ values of the Poisson distribution.

The paper is laid out as follows: In Section 2 we start with a general model that consists of $r + 2$ parameters, for inflating the first $r+1$ values of the Poisson distribution, and then we derive as a special case, the proposed two-parameter model. The properties for both models are given. In Section 3 we discuss the ability of the proposed model in describing overdispersed and underdispersed data while estimation methods are discussed in Section 4. Three practical applications, concerning the fitting of the considered models, are provided in Section 5 while conclusions and future research are discussed in Section 6.

2. The proposed model

In the current section, we present two discrete probability models that can be considered as generalizations of the well known zero-inflated Poisson distribution (see, for example, Johnson et al.