Statistics and Probability Letters 92 (2014) 89–94

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

On recovering a mixed Poisson distribution from its left-truncated version

M. Pérez-Casany *, J. Ginebra, J. Valero

Technical University of Catalonia and DAMA-UPC, Spain

ARTICLE INFO

Article history:
Received 13 February 2014
Received in revised form 12 May 2014
Accepted 12 May 2014
Available online 19 May 2014

MSC:
60E05
60E10
60E07

Keywords:
Count variable
Poisson mixture
Zero-truncated distribution
Probability generating function

ABSTRACT

An algorithm is proposed that, starting from the probability generating function of a left-truncation at k of a mixed Poisson distribution, recovers the first $k+1$ probabilities of the untruncated distribution, without the need of eliciting what the mixing distribution is. The result establishes that irrespective of the value where the distribution is truncated, there still remains enough information in the tail so that the initial mixing distribution can be recovered.

1. Introduction

Working with non-negative discrete data, one often faces situations where some of the first positive k values are unobservable. Moreover, the data may also exhibit a variance that is significatively larger than its mean. When both phenomena take place at the same time, one usually fits the data by means of the left-truncation version of a mixed Poisson model. We denote this type of probability models by kT-MP.

It is known that if the model assumed is a kT-MP and the mixing distribution is known, simple calculus allows one to obtain the probability generating function (pgf) of the mixed Poisson (MP) distribution and, from that, the pgf of the kT-MP. This paper goes in the other sense, showing how to recover the first $k+1$ probabilities of the corresponding untruncated MP model starting from the pgf of the kT-MP distribution and the probability at zero of the mixing distribution, without requiring one to know the complete mixing distribution. Most often, this probability at zero will be known to be equal to zero. This is the case, for instance, when a continuous mixing distribution is considered.

The result presented is useful, for example, in the analysis of words or species frequency count data through 0T-MP models, because the probability at zero of the untruncated MP model determines the number of unobserved words or species and hence the size of the vocabulary or population. It is also important in the analysis of protein or social graphs data where only nodes with more than a given positive number of edges are observed, but the interest is in the estimation of the total number of nodes in the graph.

* Corresponding author. Tel.: +34 93 4137689; fax: +34 93 413 77 01.
E-mail address: marta.perez@upc.edu (M. Pérez-Casany).

http://dx.doi.org/10.1016/j.spl.2014.05.008

2. Preliminaries

Given a non-negative integer (count) random variable (r.v.) X such that $P(X = i) = p_i$, its probability generating function (pgf) is defined as: $h(s) = E[s^X] = \sum_{i=0}^{+\infty} p_i s^i$. It is known that two r.v's have the same pgf if, and only if, they have the same probability distribution.

Let $f(s)$ be a function that has derivatives of all orders in a given real interval (a, b). Denoting by $f^{(i)}(s)$ the ith derivative of f at s, it is said that f is *absolutely monotone* (*strictly absolutely monotone*) in (a, b) if, and only if,

$$f^{(i)}(s) \geq (>) 0, \quad \forall s \in (a, b) \text{ and } i = 0, 1, 2, \dots.$$

A real function $h(s)$ is a pgf if, and only if, it is absolutely monotone in $(0, 1)$ and $h(1) = 1$.

A *zero-modification* of X, is a r.v. X^*, such that $P(X^* = 0) = \epsilon + (1 - \epsilon)\, p_0$ and $P(X^* = i) = (1 - \epsilon)\, p_i$ when $i \geq 1$, for a given ϵ such that $\epsilon + (1 - \epsilon)\, p_0 \geq 0$. Positive (negative) values of ϵ yield r.v's with a probability at zero larger (smaller) than p_0 (Johnson et al., 2005).

Let us assume that the Poisson parameter, λ, of a Poisson distributed r.v. follows a non-negative distribution with distribution function U. The resulting r.v. is known as MP r.v. and has probability mass function equal to

$$P(X = i) = \int_0^{+\infty} \frac{\lambda^i}{i!} e^{-\lambda}\, dU(\lambda), \quad \text{for } i = 0, 1, 2, \dots.$$

In Puri and Goldie (1979) it is proved that a non-degenerate count r.v. is MP if, and only if, its pgf is strictly absolutely monotone in $(-\infty, 1)$. The following characterization of MP appears in Valero et al. (2010) and is a reformulation of this result.

Proposition 1. *A function $h_X(s)$ is the pgf of a r.v. X with a MP distribution different from the degenerated distribution at zero if, and only if,* (a) $h_X(1) = 1$, (b) *it is analytical in $(-\infty, 1)$, and* (c) $h_X^{(i)}(s) > 0$ *for all s in $(-\infty, 1)$ and $i = 0, 1, \dots$.*

The following lemma is a consequence of Proposition 1.

Lemma 1. *If X is an MP distributed r.v. with pgf $h_X(s)$, and one denotes by ϵ the probability at zero of the mixing distribution, then $h_X(s)$ is a non-negative increasing function with*

$$\lim_{s \to -\infty} h_X(s) = \epsilon.$$

Proof. This lemma is a consequence of the fact that if $U(0) = \epsilon$, and if U_0 is the distribution function of the zero-truncation of the mixing distribution, then:

$$h_X(s) = \epsilon + (1 - \epsilon) \int_0^{\infty} e^{\lambda\,(s-1)} dU_0(\lambda), \tag{1}$$

which tends to ϵ when s tends to $-\infty$. From (1) one has that in the case where $\epsilon > 0$, X may be posed as a zero-modification of an MP distribution with a pgf that has limit zero at $-\infty$. $\square$

Given a count random variable X and given a positive integer number k, the *left-truncation* of X at k, denoted by X^{kT}, is defined to be the r.v. with

$$P(X^{kT} = i) = \begin{cases} 0 & \text{if } 0 \leq i < k + 1, \\ \dfrac{P(X = i)}{1 - \sum\limits_{j=0}^{k} P(X = j)} & \text{if } i \geq k + 1. \end{cases}$$

It is easy to check that if one denotes by $h_X(s)$ the pgf of X, the pgf of X^{kT} is:

$$h_{X^{kT}}(s) = \frac{h_X(s) - \sum\limits_{j=0}^{k} P(X = j)\, s^j}{1 - \sum\limits_{j=0}^{k} P(X = j)} = \frac{\sum\limits_{j=1}^{+\infty} P(X = k + j) s^{k+j}}{\sum\limits_{j=1}^{+\infty} P(X = k + j)}. \tag{2}$$

Next theorem, in Valero et al. (in press), characterizes the left-truncated MP distributions that have the first $k + 1$ moments finite by means of their pgf.

Theorem 1. *A function $h(s)$ is the pgf of a r.v. with an MP left-truncated at k distribution with the first $k + 1$ moments finite if, and only if, it verifies that:* (a) $h^{(i)}(0) = 0$, $\forall i \in \{0, 1, 2, \dots, k\}$, $h(1) = 1$ *and* $h^{(i)}(s) < +\infty$, $\forall i \in \{1, 2, \dots, k + 1\}$; (b) *it is analytical in $(-\infty, 1)$;* (c) *all the coefficients of the series expansion of $h(s)$ about any point s_0 in $(-\infty, 1)$ are strictly positive except maybe the first $k + 1$, or what is the same, $h^{(k+1)}(s)$ is strictly absolutely monotone;* (d) *there exists a polynomial function $q_k(s) = L_0 + L_1 s + \cdots L_k s^k$ with $L_i > 0$, $\forall i$ such that*

$$\lim_{s \to -\infty} (h(s) + q_k(s)) = 0.$$