Available online at www.sciencedirect.com
ScienceDirect
Journal of Statistical Planning and Inference 137 (2007) 3990–4002
journal of statistical planning and inference
www.elsevier.com/locate/jspi

# Asymptotic distributions of statistics and parameter estimates for mixed Poisson processes

V. Savani, A. Zhigljavsky*

*Department of Mathematics, Cardiff University, Cardiff, CF24 4AG, UK*

Received 30 March 2006; received in revised form 5 October 2006; accepted 6 April 2007
Available online 27 April 2007

**Abstract**

Mixed Poisson processes have been used as natural models for events occurring in continuous or discrete time. Our main result is the derivation of the joint asymptotic distributions of statistics, including parameter estimators, computed in different time intervals from data generated by mixed Poisson processes. These distributions can be used, for example, to test the hypothesis about the adequacy of the mixed Poisson process against data. We provide some simulation results and test the model on actual market research data.


*MSC:* 62M; 60K10; 62F; 90A40

*Keywords:* Mixed Poisson process; Mixed Poisson distribution; Maximum likelihood; Moment estimators

## 1. Introduction

Mixed Poisson processes have been used as natural models for events occurring in continuous or discrete time in many different fields including accident proneness (Greenwood and Yule, 1920), accidents and sickness (Lundberg, 1964), market research (Ehrenberg, 1988), risk theory (Grandell, 1997) and clinical trials (Cook and Wei, 2003). The main result of this paper is the derivation of the joint asymptotic distributions of statistics, including parameter estimators, computed in different time intervals from data generated by mixed Poisson processes.

The structure of the paper is as follows. In this section we introduce mixed Poisson processes and our main special case, the gamma-Poisson process. We also discuss the market research interpretation of this process which we will use throughout the paper. In Section 2 we give a general expression for the asymptotic covariance matrix of functionals of data from mixed Poisson processes. Using this general setup, in Section 3 we derive the asymptotic distributions between different statistics and estimators computed in different time intervals. These distributions allow hypothesis testing to assess goodness-of-fit of the mixed Poisson process. In Section 4 we apply general results of Section 3 to the case of the gamma-Poisson process. We provide some simulation results and test the model on actual market research data.

* Corresponding author. Tel.: +44 920875076; fax: +44 920874199.
*E-mail addresses:* SavaniV@Cardiff.ac.uk (V. Savani), ZhigljavskyAA@Cardiff.ac.uk (A. Zhigljavsky).

0378-3758/$ - see front matter 
doi:10.1016/j.jspi.2007.04.016

### 1.1. Mixed Poisson processes

Define the multivariate Poisson distribution as

$$\mathbb{P}(\boldsymbol{Z}=\boldsymbol{x}|\Lambda=\lambda)=\prod_{i=0}^{n-1}\frac{[\lambda(t_{i+1}-t_i)]^{x_{i+1}-x_i}}{(x_{i+1}-x_i)!}\,\exp(-\lambda(t_{i+1}-t_i)), \tag{1}$$

where $\lambda>0$ is the intensity, $\boldsymbol{Z}=\{Z(t_1), Z(t_2), \dots, Z(t_n)\}$ is a random vector, the set $\boldsymbol{x}=\{x_0, x_1, x_2, \dots, x_n\}$ is a set of non-negative integers with $0=x_0\leqslant x_1\leqslant\cdots\leqslant x_n$ and $0=t_0\leqslant t_1\leqslant\cdots\leqslant t_n$ represents an increasing sequence of time points. The mixed Poisson process is then defined as a process whose finite-dimensional distributions are

$$\mathbb{P}(\boldsymbol{Z}=\boldsymbol{x})=\int_{0-}^{\infty}\mathbb{P}(\boldsymbol{Z}=\boldsymbol{x}|\Lambda=\lambda)\,\mathrm{d}U_\Lambda(\lambda;\boldsymbol{\theta}). \tag{2}$$

Here $U_\Lambda(\lambda;\boldsymbol{\theta})$ is the distribution function for the random variable $\Lambda$ and $\boldsymbol{\theta}$ is a vector of unknown parameters. The function $U_\Lambda(\lambda;\boldsymbol{\theta})$ is commonly known as the structure distribution of the mixed Poisson process.

The most common distribution function for $\Lambda$ is that of the gamma distribution with probability density function

$$g(\lambda;a,k)=\frac{1}{a^k\Gamma(k)}\lambda^{k-1}\mathrm{e}^{-\lambda/a},\quad a>0,\ \ k>0,\ \ \lambda>0. \tag{3}$$

The mixed Poisson process in this case is often referred to as the gamma-Poisson process. The one-dimensional distribution of the gamma-Poisson process is the negative binomial distribution (NBD) with probabilities

$$p_x=\mathbb{P}(Z(t_1)=x)=\frac{\Gamma(k+x)}{\Gamma(k)x!}\left(\frac{1}{1+at_1}\right)^k\left(\frac{at_1}{1+at_1}\right)^x,\quad x=0,1,2,\dots\ . \tag{4}$$

In literature the parametrization $(m,k)$, where $m=ak$, is often used. In addition to the gamma distribution, Grandell (1997) considered other distributions including beta, shifted-gamma, generalized inverse Gaussian and lognormal distributions as structure distributions for the mixed Poisson process.

### 1.2. Fitting the mixed Poisson process

The fitting of mixed Poisson processes to observed data has mainly focussed on fitting the one-dimensional mixed Poisson distribution when considering data observed over fixed time intervals. Fitting the one-dimensional distribution only gives partial information as to the adequacy of the process being fitted; in particular the dynamical behavior of the mixed Poisson process is not considered by fitting the one-dimensional distribution. The derivation of the joint asymptotic distributions of statistics and estimators allows testing the hypothesis as to whether parameter estimates computed in the two different time intervals could have been generated from the same process. This will allow us to verify the dynamical properties of the underlying model against data.

Note that it is easy to construct methods of testing the adequacy of mixed Poisson processes which are based on testing whether each individual realization follows the standard Poisson process; the distribution of the intensities of the individual Poisson processes can then be checked against a specified structure distribution. However, in practice the individual behavior basically never follows the pure Poisson model (see e.g. Cox and Hinkley, 1978; Ehrenberg, 1988) and therefore the related tests would almost certainly reject the Poisson process assumption. At the same time, it is widely known that the Poisson and mixed Poisson models often work fairly well when the data is aggregated over either time or realizations, or both.

The asymptotic distributions derived in this paper allow us to test the mixed Poisson model hypotheses using the aggregated data, see Section 4.5. We are not aware of any other procedure of testing the dynamics of the mixed Poisson models, except those based on testing individual realizations (but these are not practical). In the practice of market research, when using panel data, we observe multiple realizations of data which can be aggregated.

When only a few events are registered in each individual realization, testing the pure Poisson hypothesis is meaningless as there is not enough data. However, the methodology described in this paper can be perfectly suitable for testing the mixed Poisson model if there are enough realizations in the multiple realization scheme and a suitable aggregation is made.