



Multivariate generalized Poisson geometric process model with scale mixtures of normal distributions



Jennifer So Kuen Chan*, Wai Yin Wan

School of Mathematics and Statistics, University of Sydney, NSW 2006, Australia
Bureau of Crime Statistics and Research, NSW 2006, Australia

ARTICLE INFO

Article history:

Received 27 April 2013

Available online 15 February 2014

AMS subject classifications:

62P25

62M10

62J12

62Hxx

62F15

Keywords:

Generalized Poisson distribution

Geometric process

Markov chain Monte Carlo algorithm

Multivariate log- t distribution

Scale mixtures of normals

ABSTRACT

This paper proposes a new model named as the multivariate generalized Poisson log- t geometric process (MGPLTGP) model to study multivariate time-series of counts with overdispersion or underdispersion, non-monotone trends within each time-series and positive or negative correlation between pairs of time-series. This model assumes that the multivariate counts follow independent generalized Poisson distributions with an additional parameter to adjust for different degrees of dispersion including overdispersion and underdispersion. Their means after discounting the trend effect geometrically by ratio functions form latent stochastic processes and follow a multivariate log- t distribution with a flexible correlation structure to capture both positive correlation and negative correlation. By expressing the multivariate Student's t -distribution in scale mixtures of normals, the model can be implemented through Markov chain Monte Carlo algorithms via the user-friendly WinBUGS software. The applicability of the MGPLTGP model is illustrated through an analysis of the possession and/or use of two illicit drugs, amphetamines and narcotics in New South Wales, Australia.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

The modelling of multivariate time-series of counts has wide applications in different areas. However researches for multivariate count models are relatively limited due to the computational difficulties in implementation. This is true also in other context, for example, hypothesis testing for dispersion when the methods proposed for continuous data (for a recent proposal, see [23]). Multivariate normal (MN) distribution is commonly used as an alternative choice to model discrete data [11]. Unfortunately, it becomes inappropriate when the count data is skewed, resulting from small means and/or zero-inflation.

In order to study multivariate time-series of counts with different properties in dispersion, trend and correlation, this paper proposes a new model namely the multivariate generalized Poisson log- t geometric process (MGPLTGP) model. This model is shown to have several advantages over some existing models in the literature. Amongst these models, models with bivariate Poisson distribution proposed by Kocherlakota and Kocherlakota [16] and multivariate Poisson (MP) distribution by Johnson et al. [10] expressed each component of the MP distribution as a sum of two independent univariate Poisson random variables in which one variable is common in all the sums. In this way, the model has a closed form pdf as the marginal distribution is essentially the simple Poisson distribution with mean equals the variance and the covariance between all pairs of variables is the mean of the common Poisson variable. However, the equal and positive correlation between all pairs of Poisson variables is very restrictive and the model is only applicable to equidispersed data.

* Corresponding author at: School of Mathematics and Statistics, University of Sydney, NSW 2006, Australia.

E-mail address: jchan@maths.usyd.edu.au (J.S.K. Chan).

Thereafter, Karlis and Meligkotsidou [12] extended the MP distribution to allow different covariance for each pair of variables. Nevertheless, the restriction on positive correlation and equidispersion still remain unsolved. To deal with negative correlation and overdispersion, a number of researches have considered using a mixed model approach. These MP mixed models can be classified into two types. The first type of model contains a MP distribution with mean follows a univariate mixing distribution [15]. However this model, though allows overdispersion, can only apply to positively correlated multivariate counts as the covariance function is always positive. The second type of models adopt a multivariate mixing distribution with possible negative correlation on the mean vectors of the MP distributions. However this type of MP mixed model, though are suitable for modelling overdispersed count data still cannot cope with underdispersed data. Moreover, the resulting distribution are so complicated that in practice most models consider only a special case in which components of the MP distribution are assumed to be independent [14].

To simplify the model, this paper adopts the MP mixed model of the second type with independent generalized Poisson (GP) distribution [7] for each time-series. The multivariate mixing distribution captures different covariance structures using different covariance matrices. Moreover, as non-stationarity is often prominent in time-series data, this paper further extends the geometric process model pioneered by Lam [17,18] for studying trend dynamic in the multivariate GP mixed model. The geometric process model was first applied to model inter-arrival times with monotone trend in reliability problems. Later on, Wan and Chan [28] proposed the Poisson geometric process (PGP) model which is essentially a Poisson-gamma mixed model to model longitudinal time-series of counts with a trend movement. The model is further extended to allow mixture effect and overdispersion due to zero-inflation. In addition, Wan and Chan [29] introduced the robust PGP model which is a Poisson mixed model with heavy-tailed mixing distribution such as Student's *t* or exponential power distributions. The thick tails of the distributions enhance extra Poisson variability to handle serious overdispersion due to extreme observations. To model underdispersion, Wan and Chan [30] adopt the GP distribution to handle count data with under or overdispersion. This generalized Poisson geometric process (GPGP) model is found to be the most comprehensive PGP models.

In the GPGP model, each time series W_{it} , $t = 1, \dots, n$ follows an independent GPD with the mean being a latent GP, X_{it} , $t = 1, \dots, n$ and the corresponding latent detrended stochastic process is given by $Y_{it} = X_{it}/a^{t-1}$ for some ratio $a > 0$. We assign a log multivariate-*t* (MT) distribution as the mixing distribution to the latent variables (Y_{1t}, \dots, Y_{mt}) such that its mean and covariance matrix can accommodate covariate effects and different correlation structures respectively. MT distribution is preferred to MN distribution adopted in [1,21] as MT distribution provides more flexible tails for handling outlying observations. The resultant model is essentially a multivariate version of the model combining the methodologies of robust PGP and GPGP models [29,30] and is called multivariate generalized Poisson log-*t* geometric process (MGPLTGP) model.

For model implementation, the expectation-maximization (EM) algorithm in the likelihood approach [11,13] becomes computational intensive due to the complexity of the joint probability function as the number of dimension increases. To avoid the evaluation of the complex joint probability function, Karlis and Xekalaki [14] adopted a Bayesian approach by constructing a simple Gibbs sampler to simulate the parameters from their full conditional posterior distributions. The MLT mixing distribution is expressed in scale mixtures of MLNs to facilitate the sampling from multivariate normal distribution using Markov chain Monte Carlo (MCMC) algorithms. Moreover the mixing parameters in the scale mixtures representation help to identify extreme observations in the outlier diagnosis. This method is adopted to estimate the parameters of the MGPLTGP model applied to study the trends and correlation in a bivariate time-series of arrests on use or possession of two illicit drugs in Sydney from January 1995 to December 2008. MGPLTGP model is shown to outperform the MP and MP mixed models.

The rest of the paper is organized as follows. Section 2 briefly reviews the well-established MP models and MP mixed models on which our proposed model is built. Section 3 introduces the development of the PGP, robust PGP and GPGP models from the basic geometric process model. Section 4 investigates the proposed MGPLTGP model using scale mixtures of MLNs. In Section 5, we discuss the implementation of MGPLTGP model using MCMC algorithms followed by the introduction of the model assessment criterion. Then a real data is analysed using the MGPLTGP model and compared to MP and mixed MP models in Section 6. The last Section contains some concluding remarks with plausible future extensions.

2. Review of multivariate Poisson models

2.1. Multivariate Poisson model

Suppose that V_i , $i = 0, \dots, m$ are independent Poisson random variables with mean η_i . Then the random variables $W_i = V_i + V_0$, $i = 1, \dots, m$ follow jointly a multivariate Poisson (MP) distribution where m represents the dimension of the distribution [22]. Denoting $\mathbf{W} = (W_1, \dots, W_m)$, the joint probability mass function (pmf) is given by:

$$f_{MP}(\mathbf{w}|\boldsymbol{\eta}) = \exp\left(-\sum_{i=0}^m \eta_i\right) \prod_{i=1}^m \frac{\eta_i^{w_i}}{w_i!} \sum_{i=0}^{\xi} \left\{ \prod_{j=1}^m \binom{w_j}{i} (i!)^{m-1} \left(\frac{\eta_0}{\prod_{k=1}^m \eta_k}\right)^i \right\}$$

where the observed vector $\mathbf{w} = \{w_1, w_2, \dots, w_m\}$ and $\xi = \min(w_1, w_2, \dots, w_m)$. Marginally, each W_i follows a Poisson distribution and its mean, variance and covariance are given by:

$$E(W_i) = \text{Var}(W_i) = \eta_i + \eta_0 \quad \text{and} \quad \text{Cov}(W_i, W_j) = \eta_0 \quad \text{for all } i \neq j.$$

Download English Version:

<https://daneshyari.com/en/article/1145572>

Download Persian Version:

<https://daneshyari.com/article/1145572>

[Daneshyari.com](https://daneshyari.com)