



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



# Bayesian analysis of robust Poisson geometric process model using heavy-tailed distributions

Wai-Yin Wan\*, Jennifer So-Kuen Chan

School of Mathematics and Statistics, University of Sydney, NSW 2006, Australia

## ARTICLE INFO

### Article history:

Received 18 November 2009  
Received in revised form 4 June 2010  
Accepted 4 June 2010  
Available online 26 June 2010

### Keywords:

Exponential power distribution  
Geometric process  
Markov chain Monte Carlo algorithm  
Mixture effect  
Outlier diagnosis  
Scale mixture representation

## ABSTRACT

We propose a robust Poisson geometric process model with heavy-tailed distributions to cope with the problem of outliers as it may lead to an overestimation of mean and variance resulting in inaccurate interpretations of the situations. Two heavy-tailed distributions namely Student's  $t$  and exponential power distributions with different tailednesses and kurtoses are used and they are represented in scale mixture of normal and scale mixture of uniform respectively. The proposed model is capable of describing the trend and meanwhile the mixing parameters in the scale mixture representations can detect the outlying observations. Simulations and real data analysis are performed to investigate the properties of the models.

© 2010 Elsevier B.V. All rights reserved.

## 1. Introduction

Most time series measured over continuous range assume a normal error distribution. These traditional time series models are however vulnerable to outliers. Outlier in time series has been realized as an influential factor in model fitting and forecasting. Failure to downweigh the outlying effects may lead to a poor model fit, an overestimation of variance, an inappropriate interpretation and an inaccurate prediction. This issue has received a great deal of attention, therefore several approaches have been developed to reduce the influence of outliers and the distributional deviation in the data analysis. Over the past decades, two main approaches are considered to cope with the overdispersion caused by the outliers. The first approach simply incorporates mixture effects to account for the heterogeneity in the distribution of the data. This can be viewed as a missing data problem assuming that the membership of the data from one of the distributions is unknown and has to be estimated. The mixture model is usually implemented using expectation–maximization (EM) algorithm or Markov chain Monte Carlo (MCMC) sampling algorithm.

The second approach is to adopt a heavy-tailed distribution instead of the commonly used Gaussian distribution as the error distribution of the data. Some popular choices of heavy-tailed distributions include the Student's  $t$ -distribution and a more general class of distributions is the Pearson Type IV distribution (Johnson et al., 1995). Alternatively, the exponential power (EP) distribution which can describe a leptokurtic (positive excess kurtosis) or platykurtic (negative excess kurtosis) shape is another good choice. However, the implementation of these distributions is difficult because the derivation of the marginal posterior distributions of the parameters is intractable using conventional numerical and analytic approximations (Choy and Smith, 1997). To overcome this problem, Box and Tiao (1973) proposed a new exponential power family of normal scale mixtures (SMN) and later Qin et al. (1998) pioneered the scale mixtures of uniform (SMU) which replaced the normal

\* Corresponding author. Tel.: +61 2 92319131.

E-mail addresses: [wwan3887@uni.sydney.edu.au](mailto:wwan3887@uni.sydney.edu.au) (W.-Y. Wan), [jennifer.chan@sydney.edu.au](mailto:jennifer.chan@sydney.edu.au) (J.S.-K. Chan).

distribution in the SMN form by uniform distribution. Theoretically, any distribution that can be expressed by SMN form, also has a SMU representation.

The hierarchical structure of the SMN or SMU representation possesses two prominent advantages: (1) the resulting density contains a mixing parameter which can accommodate the extra-Poisson variation and help to identify the extreme values in outlier diagnosis and (2) the parameter estimation can be simplified by sampling from normal or uniform distribution using Markov Chain Monte Carlo (MCMC) algorithms such as Gibbs sampling. The recent emergence of the software WinBUGS which performs the Bayesian statistical inferences using MCMC algorithms also facilitates the implementation of these representations, thus enhancing their popularity in the context of Bayesian modelling. This hierarchical structure is very practical in insurance applications because it is well known that the normal error distribution falls short of allowing for irregular and extreme claims and hence contaminates the estimation procedure and leads to poor estimation. For instances, Choy and Chan (2003) applied the Student's  $t$ - and EP distributions in SM representations to predict the insurance premiums to be charged on the policyholders in credibility analysis and Chan et al. (2008) predicted and projected the loss reserves data with various heavy-tailed distributions under the generalized- $t$  distribution family expressed in SMU representation.

So far, the techniques of using scale mixture representation apply solely to continuous time series. Yet, discrete count time series is observed in many occasions especially in a medical context. In clinical trials, patients usually have longitudinal measurements and sometimes the appearance of outlying observations in the data set may inflate the mean and variance of the data distribution and have an adverse effect on both the parameter estimation and prediction. Despite overdispersion, trend is often observed in time series and examining the trend patterns provides useful information on the movement of outcomes over time.

To cope with this problem, Thall and Vail (1990) proposed adding cluster-specific and time-specific random effects into the mean link function to give extra-Poisson variation to the data. Thereon, Jowaheer and Sutradhar (2002) incorporated the gamma random effects in the Poisson mixed model and later, Jowaheer et al. (2009) adopted a Poisson mixed model with two independent sources of normal random effects to deal with the problem. Nevertheless, the mixed effects approach which assumes the mean of the Poisson distribution follows gamma or log-normal distribution maybe inadequate to cast light on the outlying effect caused by the extreme observations. To tackle this pitfall, plenty of researches tried various mixing distributions. Some useful ones include generalized inverse Gaussian, generalized gamma, generalized exponential, inverse gamma, etc. Refer to (Gupta and Ong, 2005) for more details. In this paper, we seek a new direction to model overdispersed longitudinally time series of counts due to the presence of outliers. Our proposed model is an extension of the generalized mixture Poisson geometric process (GMPGP) model proposed by Wan and Chan (2009) which is developed from the geometric process (GP) model originated by Lam (1988a,b). The GMPGP model is a two-component mixture model with a mean function to analyze the covariate effect and a ratio function to describe the time effect. Meanwhile, the mixture effects in both mean and ratio functions can capture the population heterogeneity and overdispersion in the data. See Wan (2006) and Wan and Chan (2009) for more details. In this paper, we introduce a robust mixture Poisson geometric process (RMGP) model using some heavy-tailed distributions in scale mixture representation. We assume that the outcome variable has a Poisson distribution with a stochastic mean which forms a latent GP, and the mean of the GP after geometrically discounted by the ratio follows a heavy-tailed distribution such as log- $t$  distribution or log-exponential power (log-EP) distribution represented in SMN and SMU forms respectively. Under the scale mixture representation, the model parameters can be simulated using the MCMC algorithms and the mixing parameters help to identify the extreme values in the outlier diagnosis. To our knowledge, this is a pioneering work in adopting Student's  $t$ - or EP distributions for Poisson time series. Besides, our proposed GP models have a trend component (ratio function) which enables the study of trends of the outcomes over time and can accommodate the clustering effect using a mixture of homogeneous distributions. These make our model more advantageous than many existing Poisson time series models.

To demonstrate the characteristics and application of our models, the paper is presented as follows. First, the development of the robust mixture Poisson geometric process (RMGP) model using Student's  $t$ - and EP distributions from the GP model is described in Section 2. Next, Section 3 introduces the scale mixture representation of the two heavy-tailed distributions and their implementation in the RMGP model. Besides, the hierarchical structure and MCMC algorithms of the models are given followed by the introduction of the model assessment criterion. Furthermore, Section 4 consists of a simulation study of the robust Poisson geometric process (RPGP) models and Section 5 demonstrates an application of the proposed models using the epilepsy data studied by Leppik et al. (1985) with discussion. Lastly, a brief summary is given in Section 6.

## 2. Robust mixture Poisson geometric process model

### 2.1. Geometric process model

Lam (1988a,b) first proposed to model positive continuous data with monotone trend, such as inter-arrival times, by a monotone process called the geometric process (GP) defined as:

**Definition.** Let  $X_1, X_2, \dots$  be a sequence of non-negative random variables. If there exists a positive real number  $a$  such that  $\{Y_t = a^{t-1}X_t, t = 1, 2, \dots\}$  forms a renewal process (RP), then the stochastic process  $\{X_t, t = 1, 2, \dots\}$  is called a geometric process (GP) and the real number  $a$  is called the ratio of the GP.

The GP model asserts that if the ratio  $a$  discounts the  $t$ th outcome  $X_t$  geometrically by  $t - 1$  times, the resulting process  $\{Y_t\}$  becomes stationary and forms a RP, which may follow some parametric distributions  $f(y_t)$  with  $E(Y_t) = \mu$  and  $\text{Var}(Y_t) = \sigma^2$ . Hence, the mean and variance of the GP model are:

$$E(X_t) = \frac{\mu}{a^{t-1}} \quad \text{and} \quad \text{Var}(X_t) = \frac{\sigma^2}{a^{2(t-1)}}$$

respectively. With the ratio parameter, the GP model allows the mean and variance to change over time. In fact, a GP is a monotonic increasing sequence of non-negative random variables if  $a < 1$ , and monotone decreasing if  $a > 1$ . When  $a = 1$ , it becomes a stationary renewal process (RP) which is independently and identically distributed with the same distribution  $f(y_t)$ .

Over the past decade, GP models have been applied to various fields of research (Chan et al., 2006; Lam and Chan, 1998; Lam and Zhang, 2003; Lam et al., 2004) and have been extended to fit different types of data. For example, Chan et al. (2006) introduced the threshold GP model to study the trends of the daily number of infected cases for the severe acute respiratory syndrome (SARS) epidemic outbreak in 2003. Chan and Leung (2010) initiated the binary GP model to study the trends of methadone treatment outcomes. And, Wan and Chan (2009) introduced the generalized mixture Poisson GP (GMPGP) model to analyze the new tumour counts of the bladder cancer patients which is extended from the PGP model initiated by Wan (2006).

### 2.2. Robust PGP model

Let  $W_{it}$  denote the count for subject  $i$  at time  $t$ ,  $i = 1, \dots, m$ ,  $t = 1, \dots, n_i$  and  $n = \sum_{i=1}^m n_i$ . Following the framework of GP model,  $W_{it}$  is assumed to follow a Poisson distribution  $f_P(w_{it}|x_{it})$  with mean  $X_{it}$  which forms a latent GP. Then, we further assume the stochastic process  $\{Y_{it} = a_{it}^{t-1}X_{it}\}$  follows some lifetime distributions  $f(y_{it})$ , such as exponential distribution in Wan (2006) and gamma distribution in Wan and Chan (2009), with mean  $\mu_{it}$ , the resultant model is called Poisson geometric process (PGP) model.

Without loss of generality, we assume that the logarithm of  $Y_{it}$ , i.e.  $Y_{it}^* = \ln Y_{it}$  follows a heavy-tailed distribution  $f(y_{it}^*)$  with a mean  $E(Y_{it}^*) = \mu_{it}^*$ . Then, the marginal pmf for  $W_{it}$  is

$$\begin{aligned} f(w_{it}) &= \int_0^\infty f_P\left(w_{it} \mid \frac{e^{y_{it}^*}}{a_{it}^{t-1}}\right) f(y_{it}^*) dy_{it}^* \\ &= \int_0^\infty \frac{\exp\left(-\frac{e^{y_{it}^*}}{a_{it}^{t-1}}\right) \left(\frac{e^{y_{it}^*}}{a_{it}^{t-1}}\right)^{w_{it}}}{w_{it}!} f(y_{it}^*) dy_{it}^*. \end{aligned} \tag{1}$$

The resultant model is named as robust Poisson GP (RPGP) model. It is essentially a state space model with state variables  $X_{it}$  and has time-evolving mean and ratio functions to accommodate the exogenous effects and non-monotone trends. The mean  $\mu_{it}^*$  and ratio  $a_{it}$  are identity-linked and log-linked respectively to linear functions of covariates defined respectively below:

$$\mu_{it}^* = \beta_{\mu 0} + \beta_{\mu 1} z_{\mu 1it} + \dots + \beta_{\mu q_\mu} z_{\mu q_\mu it} \tag{2}$$

$$\ln a_{it} = \beta_{a 0} + \beta_{a 1} z_{a 1it} + \dots + \beta_{a q_a} z_{a q_a it} \tag{3}$$

where  $z_{jkit}$ ,  $k = 1, \dots, q_j$  are some time-evolving covariates. Hence,  $\{Y_{it}\}$  is no longer a RP but becomes a stochastic process. Similarly, the non-constant ratio function  $a_{it}$  allows a non-monotone trend and different values for  $\beta_{ak}$ ,  $k = 0, \dots, q_a$  can describe a variety of trend patterns. Refer to Wan (2006) for more details.

Apparently, by mixing the Poisson distribution with some heavy-tailed distributions, extra variability will be added to the Poisson distribution which enables the model to accommodate the enlarged variance caused by some extreme observations. In this paper, we consider the Student's  $t$ - and EP distributions because they have different shapes including heavier-than-normal to normal tails in the former as well as the platykurtic shape and leptokurtic shape with a kink in the latter. The various shapes allow the model to be more flexible to capture different kurtoses in the data.

#### 2.2.1. RPGP model with Student's $t$ -distribution

If  $Y_{it}^*$  has a Student's  $t$ -distribution with mean  $\mu_{it}^*$  and variance  $\frac{\nu}{\nu-2}\sigma^2$ , the probability density function of Student's  $t$ -distribution  $f_T(y_{it}^*)$  is given by

$$f_T(y_{it}^*) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\sigma\Gamma\left(\frac{\nu}{2}\right)} \left\{ 1 + \frac{(y_{it}^* - \mu_{it}^*)^2}{\nu\sigma^2} \right\}^{-\frac{(\nu+1)}{2}} \tag{4}$$

where  $\mu_{it}^*$  is given by (2),  $\nu > 2$  and is the degrees of freedom which controls the tails. A small  $\nu$  gives a heavier tail and Student's  $t$ -distribution converges to normal distribution when  $\nu \rightarrow \infty$ . The kurtosis is  $\frac{6}{\nu-4} + 3$  for  $\nu > 4$  which is greater than 3, the kurtosis of the normal distribution.

**Table 1**

Different moments of marginal pmfs for RPGP- $t$  model under a set of floating parameters with fixed values of  $\nu = 10, \sigma = 0.5, \beta_{a0} = 0.5, \beta_{a1} = -0.1, \beta_{\mu 0} = 3, \beta_{\mu 1} = -0.2$ .

Floating parameter	Mean	Variance	Kurtosis	Floating parameter	Mean	Variance	Kurtosis
$\nu = 1$	15.082	152.634	1.0017	$\sigma = 0.1$	12.279	14.110	0.1209
$\nu = 2$	14.381	110.855	1.0036	$\sigma = 0.5$	14.041	81.182	0.5579
$\nu = 5$	14.616	103.899	0.6568	$\sigma = 1.0$	16.143	230.454	0.9817
$\nu = 50$	13.985	69.826	0.3827	$\sigma = 2.0$	15.511	349.030	0.9832
$\beta_{\mu 0} = 1.0$	1.057	1.643	0.0411	$\beta_{a0} = -0.5$	34.452	262.605	0.9605
$\beta_{\mu 0} = 2.0$	2.891	6.827	0.2277	$\beta_{a0} = -0.01$	23.015	175.040	0.8652
$\beta_{\mu 0} = 3.0$	7.832	32.123	0.7399	$\beta_{a0} = 0.5$	14.460	95.392	0.7662
$\beta_{\mu 0} = 4.0$	20.814	153.217	1.1929	$\beta_{a0} = 1.5$	5.344	16.972	0.2819

To study the pmfs of the proposed RPGP- $t$  model, we assume  $q_{\mu} = q_a = 1, z_{\mu 1} = b = 0, 1$  as the covariate effect and  $z_{a1t} = t$  as the time-evolving effect in (2) and (3). Hence the mean function  $\mu_t = \exp(\beta_{\mu 0} + \beta_{\mu 1}b)$  and ratio function  $a_t = \exp(\beta_{a0} + \beta_{a1}t)$ . Fixing  $b = 1$  and  $t = 2$ , we change the values of one of the scale, shape or location parameters each time while keeping the other parameters constant and approximate the pmf in (1) using Monte Carlo integration as described below. Conditional on covariate  $b$  and time  $t$ , the marginal pmf estimator  $\hat{f}_{bt}(w)$ , in general, can be obtained by:

$$\hat{f}_{bt}(w) = \sum_{j=1}^M f_P \left( w \mid \frac{e^{y_{bt}^{*(j)}}}{a_t^{t-1}} \right) = \sum_{j=1}^M \frac{\exp \left( -\frac{e^{y_{bt}^{*(j)}}}{a_t^{t-1}} \right) \left( \frac{e^{y_{bt}^{*(j)}}}{a_t^{t-1}} \right)^w}{w!}, \quad w = 0, 1, \dots, \infty; \quad \text{and}$$

$$y_{bt}^{*(j)} \sim f_T(y_{bt}^{*(j)} \mid \mu_t^*, \sigma, \nu), \quad j = 1, \dots, M \tag{5}$$

where  $M = 10\,000$ , the latent  $y_{bt}^{*(j)}$  are simulated from the Student's  $t$ -distribution in (4) given the parameters  $\mu_{it}^*, \sigma$  and  $\nu$ . Besides the mean and variance, we also study the kurtosis of the pmfs using the method in Gupta and Ong (2005) for a discrete distribution. The relative long-tailedness of the distribution is defined as  $\lim_{w \rightarrow \infty} \hat{f}_{bt}(w + 1) / \hat{f}_{bt}(w)$  where the limit is zero for Poisson distribution. The marginal pmfs are displayed in Fig. 1(a)–(d) with their means, variances and kurtoses summarized in Table 1 below. Note that since different pmfs are drawn on the same graph for comparison, we use curves instead of bars to represent marginal pmfs for better visualization.

Results reveal that in general the location, variability and the tailedness of the marginal pmf depend on parameters  $\beta_{jk}, j = \mu, a; k = 0, 1$ , in which a larger  $\beta_{\mu 0}$  and a smaller  $\beta_{a0}$  lead to a larger mean, variance and kurtosis. Whereas,  $\nu$  and  $\sigma$  control the spread and the tail behaviour of the distribution without altering its mean. A smaller  $\nu$  and a larger  $\sigma$  contribute to a larger variability and a heavier tail and thus the model can accommodate the outlying effect due to extreme values while keeping the mean unchanged. Since the variance of each distribution in Table 1 is substantially larger than the mean, the RPGP- $t$  model is capable of fitting data with overdispersion due to outliers as well as data with equidispersion when  $\sigma$  is small.

2.2.2. RPGP model with EP distribution

If  $Y_{it}^*$  has an exponential power (EP) distribution, also known as generalized error distribution, with mean  $\mu_{it}^*$  and variance  $\sigma^2$ , it has a probability density function

$$f_{EP}(y_{it}^*) = \frac{c_1}{\sigma} \exp \left\{ - \left| \frac{c_0^{1/2} (y_{it}^* - \mu_{it}^*)}{\sigma} \right|^{2/\nu} \right\} \tag{6}$$

where  $c_0 = \Gamma(3\nu/2) / \Gamma(\nu/2), c_1 = c_0^{1/2} / (\nu \Gamma(\nu/2))$  and  $\nu \in (0, 2]$  is a shape parameter which controls the kurtosis. This family subsumes a range of symmetric distributions such as uniform ( $\nu \rightarrow 0$ ) with kurtosis equal to 1.8, normal ( $\nu = 1$ ) with kurtosis equal to 3 and double exponential ( $\nu = 2$ ) with a kurtosis of 6. Its tails can be more platykurtic when  $\nu < 1$  or more leptokurtic when  $\nu > 1$  compared with the normal tail ( $\nu = 1$ ).

By fixing the parameters at  $\beta_{a0} = 0.5, \beta_{a1} = -0.1, \beta_{\mu 0} = 3.0, \beta_{\mu 1} = -0.2, \nu = 1, \sigma = 0.5$  but leaving one floating, the marginal pmfs are again approximated using Monte Carlo integration specified in (5) but replacing  $f_T(y_{bt}^{*(j)} \mid \mu_t^*, \sigma, \nu)$  with  $f_{EP}(y_{bt}^{*(j)} \mid \mu_t^*, \sigma, \nu)$ . The effects of different parameters on the resulting pmfs are illustrated in Table 2 and Fig. 2(a)–(d). Clearly, the tail of the distribution depends on all the parameters and the behaviours of the parameters  $\sigma, \beta_{\mu 0}$  and  $\beta_{a0}$  on the resultant pmf are quite close to those of RPGP- $t$  model. However, comparing Fig. 2(a) with Fig. 1(a), the shape parameter  $\nu$  in the RPGP-EP model allows a wider range of shapes than those of the RPGP- $t$  model as the EP distribution includes both leptokurtic and platykurtic shapes. Hence the RPGP-EP model can accommodate different degrees of tailedness due to moderate to adverse outlying effects.

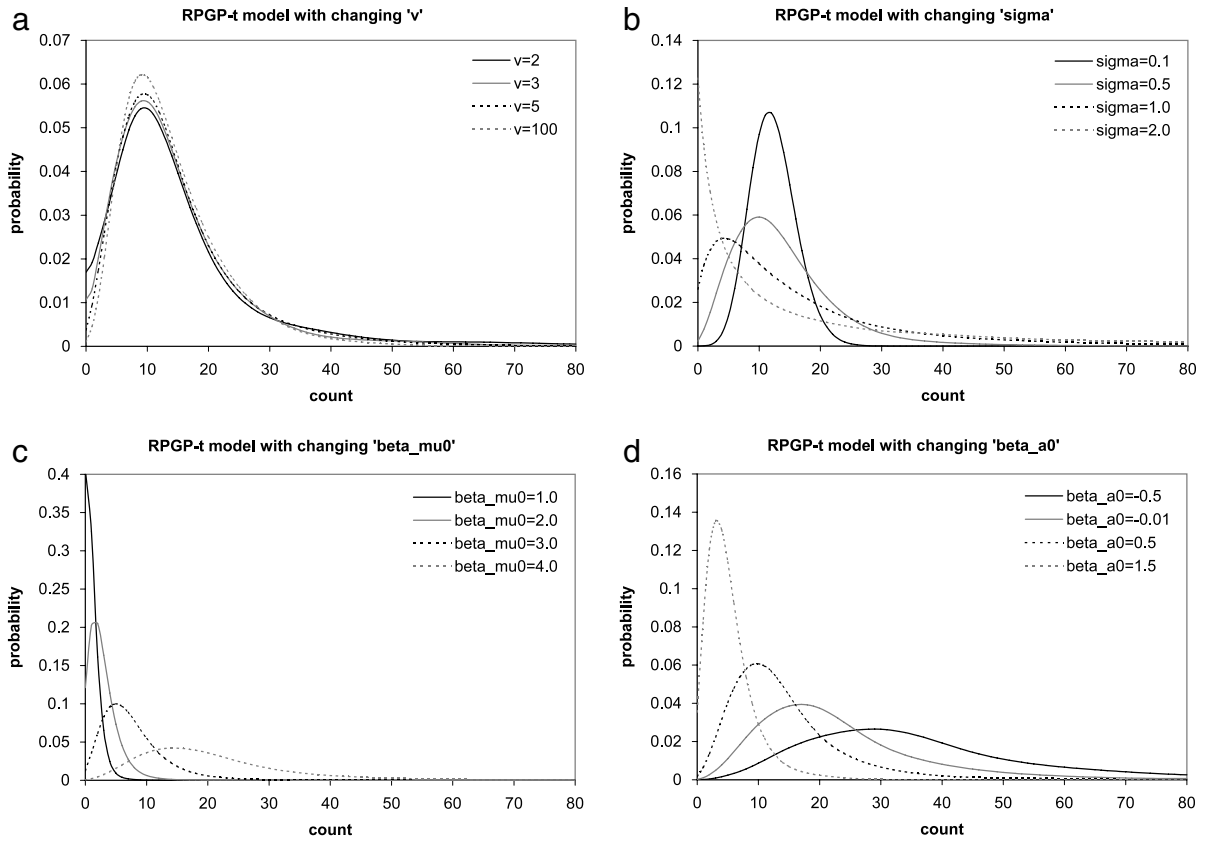


Fig. 1. RPGP-t model with varying  $\nu$ ,  $\sigma$ ,  $\beta_{\mu 0}$  and  $\beta_{a 0}$ .

Table 2

Different moments of marginal pmfs for RPGP-EP model under a set of floating parameters with fixed values of  $\nu = 1$ ,  $\sigma = 0.5$ ,  $\beta_{a 0} = 0.5$ ,  $\beta_{a 1} = -0.1$ ,  $\beta_{\mu 0} = 3$ ,  $\beta_{\mu 1} = -0.5$ .

Floating parameter	Mean	Variance	Kurtosis	Floating parameter	Mean	Variance	Kurtosis
$\nu = 0.1$	9.379	16.987	0.0925	$\sigma = 0.1$	9.048	9.871	0.0782
$\nu = 0.5$	9.629	20.468	0.1312	$\sigma = 0.5$	10.331	42.865	0.2938
$\nu = 1.0$	10.174	39.574	0.3281	$\sigma = 1.0$	13.891	197.640	0.9747
$\nu = 2.0$	12.592	192.360	0.9835	$\sigma = 2.0$	14.627	327.645	0.9840
$\beta_{\mu 0} = 1.0$	1.382	1.956	0.0388	$\beta_{a 0} = -0.5$	26.592	198.533	0.9605
$\beta_{\mu 0} = 2.0$	3.771	7.695	0.0982	$\beta_{a 0} = -0.01$	16.845	104.392	0.8652
$\beta_{\mu 0} = 3.0$	10.386	46.856	0.3061	$\beta_{a 0} = 0.5$	10.216	46.338	0.7662
$\beta_{\mu 0} = 4.0$	27.046	199.776	0.7261	$\beta_{a 0} = 1.5$	3.788	10.512	0.2819

### 2.2.3. Robust mixture PGP model

Overdispersion may arise due to clustering effects, hence an alternative way to tackle overdispersion is to add mixture effects into the mean and ratio functions. Suppose that there are  $G$  groups of subjects who have different trend patterns and each subject has the probability  $\pi_l$  of coming from group  $l$ ,  $l = 1, \dots, G$ . Conditional on group  $l$ , the marginal pmf for  $W_{it}$  is given by:

$$f_l(w_{it}) = \int_0^\infty \frac{\exp\left(-\frac{e^{y_{it}^*}}{a_{itl}^{t-1}}\right) \left(\frac{e^{y_{it}^*}}{a_{itl}^{t-1}}\right)^{w_{it}}}{w_{it}!} f(y_{it}^*) dy_{it}^* \tag{7}$$

and the group-specific mean  $\mu_{itl}$  and ratio  $a_{itl}$  functions become

$$\mu_{itl}^* = \beta_{\mu 0l} + \beta_{\mu 1l}z_{\mu 1it} + \dots + \beta_{\mu q_{\mu}l}z_{\mu q_{\mu}it}, \quad \text{and} \tag{8}$$

$$\ln a_{itl} = \beta_{a 0l} + \beta_{a 1l}z_{a 1it} + \dots + \beta_{a q_{a}l}z_{a q_{a}it} \tag{9}$$

respectively. The resulting model is named as robust mixture PGP (RMPGP) model in which  $f(y_{it}^*)$  is given by (4) or (6) for the RMPGP-t or RMPGP-EP model. To illustrate the distribution of a 2-group RMPGP-EP model, its pmf  $(\pi_1 f_1(w_{it}) + (1 - \pi_1)$

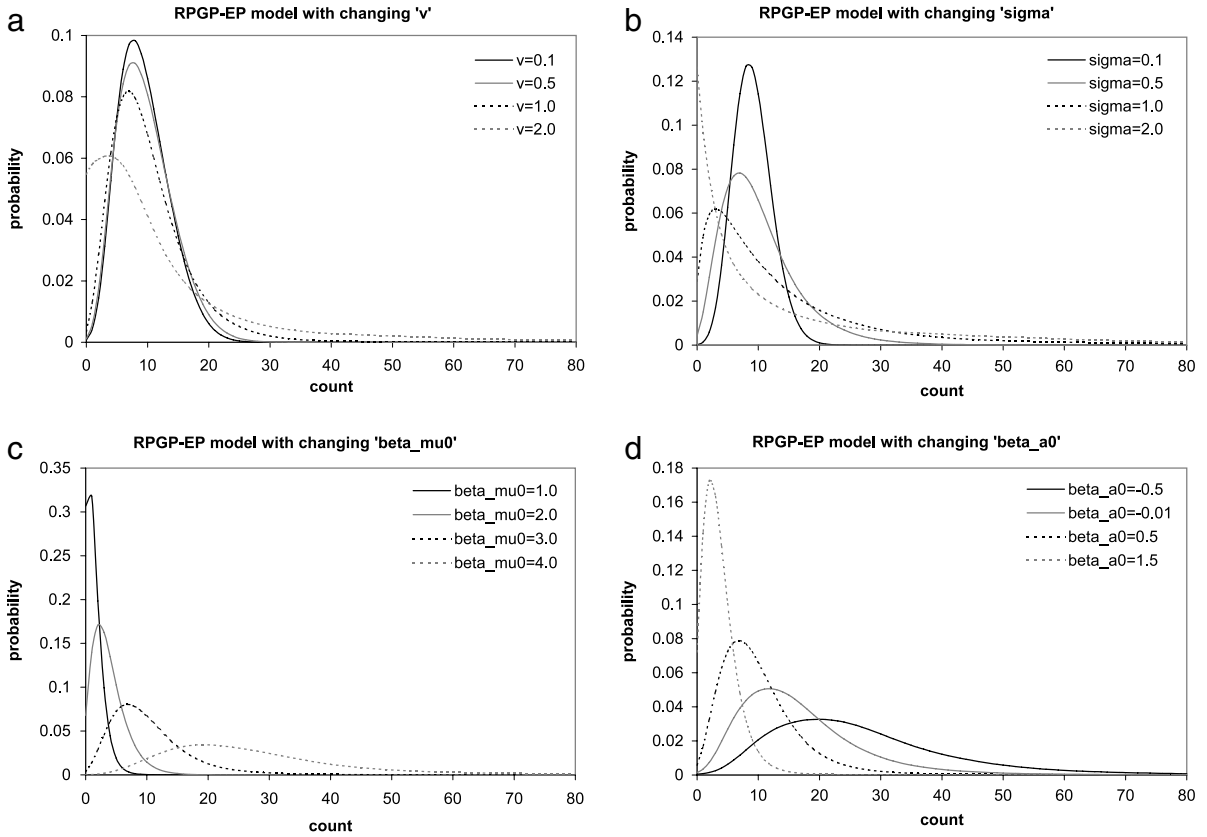


Fig. 2. RPGP-EP model with varying  $\nu$ ,  $\sigma$ ,  $\beta_{\mu 0}$  and  $\beta_{a 0}$ .

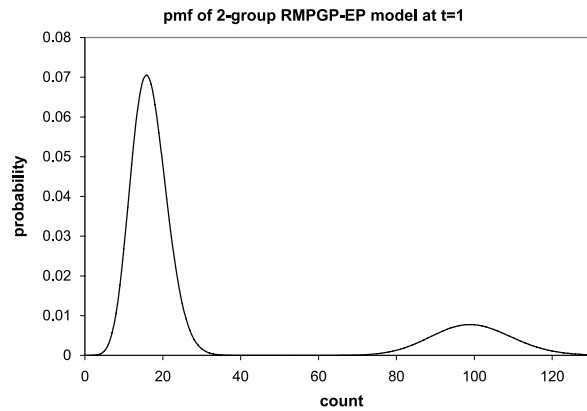


Fig. 3. pmf of 2-group RMPGP-EP model at  $t = 1$ .

$f_2(w_{it})$  where  $f_l(w_{it})$ ,  $l = 1, 2$  is given by (7) is plotted in Fig. 3 by assuming  $G = 2$ ,  $q_\mu = q_a = 1$ ,  $t = 1$ ,  $z_{\mu li} = 1$  and  $z_{a li} = t$ . For  $l = 1$ , we set  $\pi_1 = 0.8$ ,  $\beta_{a01} = -0.1$ ,  $\beta_{a11} = 0.05$ ,  $\beta_{\mu 01} = 3$ ,  $\beta_{\mu 11} = -0.2$ ,  $\nu_1 = 0.2$  and  $\sigma_1 = 0.2$  while for  $l = 2$ , we use  $\beta_{a02} = 0.1$ ,  $\beta_{a12} = -0.01$ ,  $\beta_{\mu 02} = 5$ ,  $\beta_{\mu 12} = -0.4$ ,  $\nu_2 = 1.9$  and  $\sigma_2 = 0.01$ . Fig. 3 clearly displays the two distinct modes in the distribution with a larger mode ( $l = 1$ ) at smaller values of  $W$  and a smaller one ( $l = 2$ ) at larger values of  $W$  representing the outliers. This explains how the incorporation of mixture effects in the RMPGP-EP model can accommodate overdispersion due to clustering effects.

### 3. Bayesian inference

Performing statistical inference using classical methods like maximum likelihood approach is cumbersome when the data distribution has no closed form because the likelihood function involving high-dimensional integration is intractable. To



avoid such numerical difficulties, we use a Bayesian approach via MCMC algorithms to convert the optimization problem to a sampling problem. Since the non-conjugate structure in the posterior distribution for both RMPGP models and the absolute term in the density function of the EP distribution complicate the sampling algorithms, representing the heavy-tailed distributions in a scale mixture form produces a simpler set of full conditional posterior distributions for the parameters and alleviates the computational burden of the Gibbs sampler in the MCMC algorithms.

### 3.1. Scale mixture representation of heavy-tailed distributions

Choy and Smith (1997) has shown that Student’s  $t$ - and EP distributions can be expressed in scale mixture representation to facilitate the simulation in the MCMC algorithms via a Bayesian hierarchical structure. However, the ways they handle outliers are different. Choy and Walker (2003) revealed that the former downweighs the extreme values, whereas the latter merely downweighs or bounds the influence of the outliers. Thus, it will be interesting to study their performances in outlier diagnosis. In the following, the Student’s  $t$ -distribution expressed in SMN form and the EP distribution represented in SMU form will be discussed in details.

#### 3.1.1. Student’s $t$ -distribution in SMN representation

Assume that a continuous random variable  $Y$  has a Student’s  $t$ -distribution  $f_T(y)$  with location  $\mu$ , scale  $\sigma^2$  and degrees of freedom  $\nu$ . The probability density function of  $Y$  is said to have a SMN representation if it can be expressed as

$$f_T(y|\mu, \sigma, \nu) = \int_0^\infty f_N\left(y \mid \mu, \frac{\sigma^2}{\lambda}\right) f_G\left(\lambda \mid \frac{\nu}{2}, \frac{\nu}{2}\right) d\lambda \tag{10}$$

where  $f_N(\cdot|c, d)$  represents a normal distribution with mean  $c$  and variance  $d$ ,  $f_G(\cdot|c, d)$  refers to a gamma distribution with mean  $c/d$  and variance  $c/d^2$ ,  $\nu$  is a shape parameter and  $\lambda$  is a mixing parameter which can be used to identify outlier. An outlier is indicated if  $\lambda$  is substantially small as small value implies that the normal distribution in (10) has an inflated variance and hence helps to downweigh its influence on the variance  $\sigma^2$ .

Applying the SMN form (10) for (4) in the RMPGP- $t$  model, the marginal pmf for  $W_{it}$  in (7) becomes:

$$\int_0^\infty \frac{\exp\left(-\frac{e^{y_{it}^*}}{a_{it}^{t-1}}\right) \left(\frac{e^{y_{it}^*}}{a_{it}^{t-1}}\right)^{w_{it}}}{w_{it}!} \int_0^\infty f_N\left(y_{it}^* \mid \mu_{it}^*, \frac{\sigma_1^2}{\lambda_{it}}\right) f_G\left(\lambda_{it} \mid \frac{\nu_l}{2}, \frac{\nu_l}{2}\right) d\lambda_{it} dy_{it}^*.$$

#### 3.1.2. Exponential power distribution in SMU representation

Theoretically, any distribution that can be expressed in SMN form, also has a SMU representation (Qin et al., 1998). To simplify the implementation of the MCMC sampling algorithm, Walker and Gutiérrez-Peña (1999) first proposed to express the EP distribution in SMU representation. In a slightly different form, Chan et al. (2008) write the EP distribution in SMU form as follow:

$$f_{EP}(y|\mu, \sigma) = \int_0^\infty f_U\left(y \mid \mu - \sigma\lambda^{\nu/2}, \mu + \sigma\lambda^{\nu/2}\right) f_G\left(\lambda \mid 1 + \frac{\nu}{2}, \frac{1}{2}\right) d\lambda \tag{11}$$

where  $f_U(\cdot|c, d)$  is a uniform distribution on the interval  $(c, d)$  and again  $\lambda$  is a mixing parameter. Different from the Student’s  $t$ -distribution, the larger the  $\lambda$ , the wider is the range of the uniform distribution to accommodate a possible outlier.

In the RMPGP-EP model, if we replace (6) with (11), the marginal pmf for  $W_{it}$  in (7) becomes:

$$\int_0^\infty \frac{\exp\left(-\frac{e^{y_{it}^*}}{a_{it}^{t-1}}\right) \left(\frac{e^{y_{it}^*}}{a_{it}^{t-1}}\right)^{w_{it}}}{w_{it}!} \int_0^\infty f_U\left(\mu_{it}^* - \sigma_1\lambda_{it}^{\nu_l/2}, \mu_{it}^* + \sigma_1\lambda_{it}^{\nu_l/2}\right) f_G\left(\lambda_{it} \mid 1 + \frac{\nu_l}{2}, \frac{1}{2}\right) d\lambda_{it} dy_{it}^*.$$

### 3.2. MCMC algorithms

To implement the MCMC algorithms, WinBUGS, an interactive Windows version of the BUGS program for Bayesian analysis of complex statistical models using MCMC techniques is used. For the RMPGP- $t$  and RMPGP-EP models, the hierarchical structure under the Bayesian framework is outlined in the following:

$$w_{it} \sim I_{i1}f_P\left(\frac{e^{y_{it1}^*}}{a_{it1}^{t-1}}\right) + \dots + I_{iG}f_P\left(\frac{e^{y_{itG}^*}}{a_{itG}^{t-1}}\right)$$

$$y_{it}^* \sim \begin{cases} f_N\left(\mu_{it}^*, \frac{\sigma_1^2}{\lambda_{it}}\right) \text{ and } \lambda_{it} \sim f_G\left(\frac{\nu_l}{2}, \frac{\nu_l}{2}\right) & \text{for RMPGP-}t \text{ model;} \\ f_U\left(\mu_{it}^* - \sigma_1\lambda_{it}^{\nu_l/2}, \mu_{it}^* + \sigma_1\lambda_{it}^{\nu_l/2}\right) \text{ and } \lambda_{it} \sim f_G\left(1 + \frac{\nu_l}{2}, \frac{1}{2}\right) & \text{for RMPGP-EP model.} \end{cases}$$



where  $\mu_{itl}^*$  and  $a_{itl}$  are given by (8) and (9) and  $I_{il}$  is the group membership indicator for subject  $i$  such that  $I_{il} = 1$  if he/she comes from group  $l$  and zero otherwise. In order to construct the posterior density, some prior distributions are assigned to the model parameters as follows:

$$\beta_{jkl} \sim f_N(0, \tau_{jkl}^2), \quad j = \mu, a; \quad k = 0, 1, \dots, q_j; \quad l = 1, \dots, G \tag{12}$$

$$\sigma_l^2 \sim f_G(c_l, d_l) \tag{13}$$

$$v_l \sim \begin{cases} f_U(0.01, h_l) & \text{for RMPGP-}t \text{ model;} \\ f_U(0, 2) & \text{for RMPGP-EP model} \end{cases} \tag{14}$$

$$(I_{i1}, \dots, I_{iG})^T \sim \text{multinomial}(1, \pi_1, \dots, \pi_G) \tag{15}$$

$$(\pi_1, \dots, \pi_G)^T \sim f_{\text{Dir}}(\alpha_1, \dots, \alpha_G) \tag{16}$$

where  $c_l, d_l, h_l$  are some positive constants,  $f_G(c, d)$  denotes the inverse gamma density given by

$$f_G(x) = \frac{d^c}{\Gamma(c)} x^{-(c+1)} e^{-d/x}$$

and  $f_{\text{Dir}}(\alpha)$  represents a Dirichlet distribution, a conjugate to multinomial distribution, with parameters  $\alpha = (\alpha_1, \dots, \alpha_G)$ . In case of a 2-group ( $G = 2$ ) mixture model, (15) can be simplified to  $I_{i1} \sim \text{Bernoulli}(\pi_1)$ ,  $I_{i2} = 1 - I_{i1}$  and (16) becomes a uniform prior  $f_U(0, 1)$  for  $\pi_1$  with  $\pi_2 = 1 - \pi_1$ . With the posterior means  $\hat{I}_{il}$  of the group membership indicators  $I_{il}$ , patient  $i$  is classified to group  $l'$  if  $\hat{I}_{il'} = \max_l \hat{I}_{il}$ .

According to Bayes' theorem, the posterior density is proportional to the joint densities of complete data likelihood and prior probability distributions. For the RMPGP- $t$  and RMPGP-EP models, the complete data likelihood functions  $L_T(\theta)$  and  $L_{EP}(\theta)$  for the observed data  $w_{it}$  and missing data  $\{y_{itl}^*, \lambda_{itl}, I_{il}\}$  are:

$$L_T(\theta) = \prod_{i=1}^m \prod_{l=1}^G \left\{ \pi_l \prod_{t=1}^{n_i} f_P(w_{it} | \beta_{al}, y_{itl}^*) f_N(y_{itl}^* | \beta_{\mu l}, \sigma_l, \lambda_{itl}) f_G(\lambda_{itl} | v_l) \right\}^{I_{il}} \quad \text{and}$$

$$L_{EP}(\theta) = \prod_{i=1}^m \prod_{l=1}^G \left\{ \pi_l \prod_{t=1}^{n_i} f_P(w_{it} | \beta_{al}, y_{itl}^*) f_U(y_{itl}^* | \beta_{\mu l}, \sigma_l, v_l, \lambda_{itl}) f_G(\lambda_{itl} | v_l) \right\}^{I_{il}}. \tag{17}$$

The vector of model parameters is  $\theta = (\theta_1^T, \dots, \theta_m^T, \pi_1, \dots, \pi_{G-1})^T$  where  $\theta_l = (\beta_{\mu l}, \beta_{al}, \sigma_l, v_l)^T$ ;  $i = 1, \dots, m, t = 1, \dots, n_i, l = 1, \dots, G, \beta_{\mu l} = (\beta_{\mu 0l}, \dots, \beta_{\mu q_{\mu}l})^T$  and  $\beta_{al} = (\beta_{a0l}, \dots, \beta_{a q_{a}l})^T$ .

Treating  $\{y_{itl}^*, \lambda_{itl}, I_{il}\}$  as missing observations, the joint posterior density of the RMPGP-EP model is illustrated as follows:

$$f(\mathbf{w}, \mathbf{I}, \mathbf{y}^*, \boldsymbol{\lambda} | \boldsymbol{\beta}, \boldsymbol{\sigma}, \mathbf{v}, \boldsymbol{\pi}) \propto L_{EP}(\theta) \left( \prod_{j=\mu, a} \prod_{k=0}^{q_j} \prod_{l=1}^G f_N(\beta_{jkl} | 0, \tau_{jkl}^2) \right) \left( \prod_{l=1}^G f_G(\sigma_l^2 | a_l, b_l) \right) \left( \prod_{l=1}^G f_U(v_l | 0, 2) \right) f_{\text{Dir}}(\boldsymbol{\pi} | \boldsymbol{\alpha})$$

where  $\mathbf{w} = (w_{11}, w_{12}, \dots, w_{mn_m})^T, \mathbf{I} = (I_{11}, I_{12}, \dots, I_{mG})^T, \mathbf{y}^* = (y_{111}^*, y_{121}^*, \dots, y_{mn_m G}^*)^T, \boldsymbol{\lambda} = (\lambda_{111}, \lambda_{121}, \dots, \lambda_{mn_m G})^T, \boldsymbol{\beta} = (\beta_{a01}, \dots, \beta_{a q_a G}, \beta_{\mu 01}, \dots, \beta_{\mu q_{\mu} G})^T, \boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_G)^T, \mathbf{v} = (v_1, \dots, v_G)^T$  and  $\boldsymbol{\pi} = (\pi_1, \dots, \pi_G)^T$ . The complete data likelihood  $L_{EP}(\theta)$  of the RMPGP-EP model is given by (17) and the priors are given by (12)–(16). In Gibbs sampling, the unknown parameters are simulated iteratively from their univariate conditional posterior distributions which are proportional to the joint posterior density of complete data likelihood and prior densities.

The univariate full conditional posterior densities for each of the unknown model parameters  $\theta = (\boldsymbol{\beta}, \boldsymbol{\sigma}, \mathbf{v}, \boldsymbol{\pi}, \boldsymbol{\lambda})$  and the latent group indicator  $I_{il}$  are given by:

$$f(\beta_{\mu kl} | \mathbf{w}, \mathbf{I}, \mathbf{y}^*, \boldsymbol{\lambda}, \boldsymbol{\beta}^-, \boldsymbol{\sigma}, \mathbf{v}, \boldsymbol{\pi}) \propto \exp\left(-\frac{\beta_{\mu kl}^2}{2\tau_{\mu kl}^2}\right) \quad \text{restricted to}$$

$$\beta_{\mu kl} \in \left[ z_{\mu kit}^{-1} \max_{i,t} \left( y_{itl}^* - \sigma_l \lambda_{itl}^{v_l/2} - \sum_{\kappa \neq k} \beta_{\mu \kappa l} z_{\mu \kappa it} \right), z_{\mu kit}^{-1} \min_{i,t} \left( y_{itl}^* + \sigma_l \lambda_{itl}^{v_l/2} - \sum_{\kappa \neq k} \beta_{\mu \kappa l} z_{\mu \kappa it} \right) \right],$$

if  $z_{\mu kit} \neq 0$

$$f(\beta_{akl} | \mathbf{w}, \mathbf{I}, \mathbf{y}^*, \boldsymbol{\lambda}, \boldsymbol{\beta}^-, \boldsymbol{\sigma}, \mathbf{v}, \boldsymbol{\pi}) \propto \exp\left[-\sum_{i=1}^m I_{il} \sum_{t=1}^{n_i} \left\{ \frac{e^{y_{itl}^*}}{a_{itl}^{t-1}} + \beta_{akl} z_{akit} (t-1) w_{it} \right\} - \frac{\beta_{akl}^2}{2\tau_{akl}^2}\right]$$

$$f(y_{itl}^* | \mathbf{w}, \mathbf{I}, \mathbf{y}^{*-}, \boldsymbol{\lambda}, \boldsymbol{\beta}, \boldsymbol{\sigma}, \mathbf{v}, \boldsymbol{\pi}) \propto \exp\left(-\frac{e^{y_{itl}^*}}{a_{itl}^{t-1}} + y_{itl}^* w_{it}\right) \quad \text{restricted to } y_{itl}^* \in (\mu_{itl}^* - \sigma_l \lambda_{itl}^{v_l/2}, \mu_{itl}^* + \sigma_l \lambda_{itl}^{v_l/2})$$

$$\begin{aligned}
 f(\sigma_l^2 | \mathbf{w}, \mathbf{I}, \mathbf{y}^*, \boldsymbol{\lambda}, \boldsymbol{\beta}, \boldsymbol{\sigma}^-, \mathbf{v}, \boldsymbol{\pi}) &\propto \sigma_l^{-\left(\sum_{i=1}^m n_i l_{il} + 2a_l + 2\right)} \exp\left(-\frac{b_l}{\sigma_l^2}\right) \text{ restricted to } \sigma_l^2 > \max_{i,t} \left\{ \frac{(y_{itl}^* - \mu_{itl}^*)^2}{\lambda_{itl}^{v_l}} \right\} \\
 f(v_l | \mathbf{w}, \mathbf{I}, \mathbf{y}^*, \boldsymbol{\lambda}, \boldsymbol{\beta}, \boldsymbol{\sigma}, \mathbf{v}^-, \boldsymbol{\pi}) &\propto \left[ \Gamma\left(\frac{v_l}{2} + 1\right) 2^{\frac{v_l}{2}} \right]^{-\sum_{i=1}^m n_i l_{il}} \text{ restricted to } v_l > \max_{i,t} \frac{\ln\left(\frac{y_{itl}^* - \mu_{itl}^*}{\sigma_l^2}\right)^2}{\ln \lambda_{itl}} \\
 f(\mathbf{I}_i | \mathbf{w}, \mathbf{y}^*, \mathbf{I}^-, \boldsymbol{\lambda}, \boldsymbol{\beta}, \boldsymbol{\sigma}, \mathbf{v}, \boldsymbol{\pi}) &\propto \prod_{l=1}^G \frac{\pi_l \prod_{t=1}^{n_i} \frac{\exp\left(-\frac{e^{y_{itl}^*} - \lambda_{itl}}{a_{itl}^{t-1}}\right) \left(\frac{e^{y_{itl}^*}}{a_{itl}^{t-1}}\right)^{w_{it}}}{\sigma_l \Gamma\left(\frac{v_l}{2} + 1\right) 2^{\frac{v_l}{2}}}}{\sum_{l'=1}^G \left\{ \pi_{l'} \prod_{t=1}^{n_i} \frac{\exp\left(-\frac{e^{y_{itl'}^*} - \lambda_{itl'}}{a_{itl'}^{t-1}}\right) \left(\frac{e^{y_{itl'}^*}}{a_{itl'}^{t-1}}\right)^{w_{it}}}{\sigma_{l'} \Gamma\left(\frac{v_{l'}}{2} + 1\right) 2^{\frac{v_{l'}}{2}}}\right\}} = \prod_{i=1}^G \pi_{il}^{I_{il}} \\
 &= \text{Multinomial}(1, \pi_{i1}', \dots, \pi_{iG}') \\
 f(\lambda_{itl} | \mathbf{w}, \mathbf{I}, \mathbf{y}^*, \boldsymbol{\lambda}^-, \boldsymbol{\beta}, \boldsymbol{\sigma}, \mathbf{v}, \boldsymbol{\pi}) &\propto \exp\left(-\frac{\lambda_{itl}}{2}\right) \text{ restricted to } \lambda_{itl} > \left(\frac{y_{itl}^* - \mu_{itl}^*}{\sigma_l}\right)^{2/v_l} \\
 f(\pi_l | \mathbf{w}, \mathbf{I}, \mathbf{y}^*, \boldsymbol{\lambda}, \boldsymbol{\beta}, \boldsymbol{\sigma}, \mathbf{v}, \boldsymbol{\pi}^-) &\propto \pi_l^{\sum_{i=1}^m l_{il} + a_l - 1}
 \end{aligned}$$

where  $\boldsymbol{\beta}^-, \mathbf{y}^{*-}, \boldsymbol{\lambda}^-, \boldsymbol{\sigma}^-, \mathbf{v}^-, \boldsymbol{\pi}^-$  and  $\mathbf{I}^-$  and are vectors of  $\boldsymbol{\beta}, \mathbf{y}^*, \boldsymbol{\lambda}, \boldsymbol{\sigma}, \mathbf{v}, \boldsymbol{\pi}$  and  $\mathbf{I}$  excluding  $\beta_{jkl}, y_{itl}^*, \lambda_{itl}, \sigma_l, v_l, \pi_l$  and  $I_i$  respectively. The MCMC algorithms are implemented using WinBUGS where 55 000 iterations are executed for each model and the first 5000 iterations are discarded as the burn-in period. Thereafter parameters are sub-sampled from every 50th iteration to reduce the auto-correlation in the sample. This results in  $M = 1000$  simulated posterior samples of every parameter and parameter estimates are given by their sample means or medians. History plots and auto-correlation function (ACF) plots of each parameter are examined to ensure convergence and independence amongst the parameters.

### 3.3. Model selection criterion

In our analysis, we adopt the deviance information criterion (DIC), originated by Spiegelhalter et al. (2002), as the model selection criterion. Deviance information criterion (DIC) is the sum of posterior mean deviance  $\bar{D}(\boldsymbol{\theta})$  measuring the model fit and an effective dimension  $p_D$  which accounts for the model complexity. For the RMPGP model, the DIC is defined as

$$\begin{aligned}
 DIC &= \overline{D(\boldsymbol{\theta})} + p_D \\
 &= -\frac{4}{M} \sum_{j=1}^M \sum_{i=1}^m \sum_{l=1}^G I_{il}^{(j)} \left[ \ln \pi_l^{(j)} + \sum_{t=1}^{n_i} \ln \left\{ f_P(w_{it} | y_{itl}^{*(j)}, \boldsymbol{\beta}_{al}^{(j)}) f_D(y_{itl}^{*(j)} | \boldsymbol{\beta}_{\mu l}^{(j)}, \sigma_l^{(j)}, v_l^{(j)}) \right\} \right] \\
 &\quad + 2 \sum_{i=1}^m \sum_{l=1}^G \bar{I}_{il} \left[ \ln \bar{\pi}_l + \sum_{t=1}^{n_i} \ln \left\{ f_P(w_{it} | \bar{y}_{itl}^*, \bar{\boldsymbol{\beta}}_{al}) f_D(\bar{y}_{itl}^* | \bar{\boldsymbol{\beta}}_{\mu l}, \bar{\sigma}_l, \bar{v}_l) \right\} \right] \tag{18}
 \end{aligned}$$

where  $D = T$  or  $EP$ ,  $f_D(\cdot)$  are densities given by (4) and (6) respectively, and  $\theta^{(j)}$  and  $\bar{\theta}$  represent the  $j$ th posterior sample and posterior mean of parameter  $\theta$ ,

$$\bar{I}_{il}^{(j)} = \frac{\pi_l^{(j)} \prod_{t=1}^{n_i} \left\{ f_P(w_{it} | y_{itl}^{*(j)}, \boldsymbol{\beta}_{al}^{(j)}) f_D(y_{itl}^{*(j)} | \boldsymbol{\beta}_{\mu l}^{(j)}, \sigma_l^{(j)}, v_l^{(j)}) \right\}}{\sum_{l'=1}^G \pi_{l'}^{(j)} \prod_{t=1}^{n_i} \left\{ f_P(w_{it} | y_{itl'}^{*(j)}, \boldsymbol{\beta}_{al'}^{(j)}) f_D(y_{itl'}^{*(j)} | \boldsymbol{\beta}_{\mu l'}^{(j)}, \sigma_{l'}^{(j)}, v_{l'}^{(j)}) \right\}}$$

and  $\bar{I}_{il}$  is defined in a similar way by replacing  $y_{itl}^{*(j)}$  and  $\boldsymbol{\theta}^{(j)}$  with  $\bar{y}_{itl}^*$  and  $\bar{\boldsymbol{\theta}}$ . The rule of thumb is that the smaller the DIC, the better is the model.

### 4. Simulation

To investigate the properties of the RPGP- $t$  and RPGP-EP models, we conduct a simulation study in which  $r = 100$  data sets are simulated from each of the RPGP models based on a set of true parameters. We set  $q_\mu = q_a = 1$  and each data set contains  $m = 80$  time series of length  $n_i = 8$  from  $m_0 = 40$  subjects in the control group ( $b = 0$ ) and another  $m_1 = 40$  subjects in the treatment group ( $b = 1$ ) with  $z_{\mu 1it} = b$  in the mean function (2). The degrees of freedom  $\nu$  is set to include

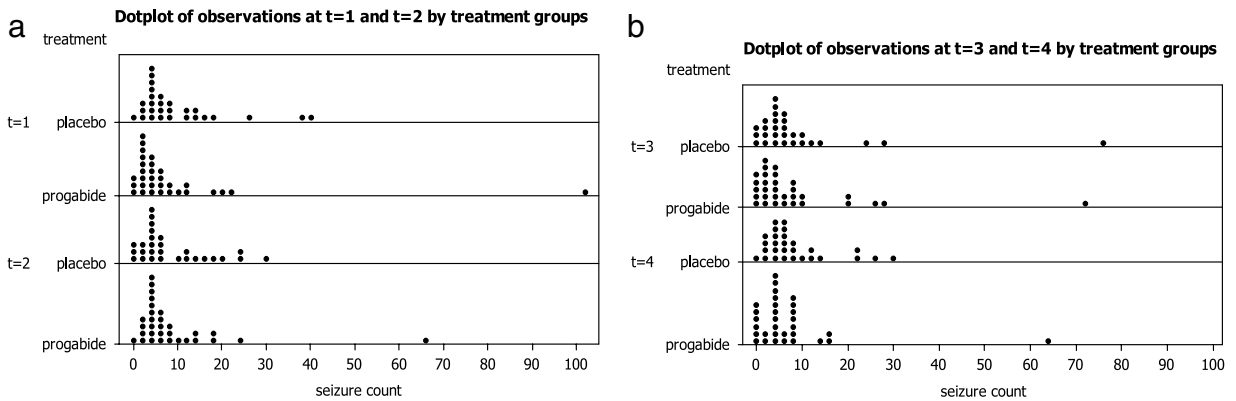


Fig. 4. Dotplots of seizure counts at different two-week intervals.

heavy ( $\nu = 2.5$ ) and light ( $\nu = 50$ ) tails for Student’s  $t$ -distribution and platykurtic ( $\nu = 0.1$ ) and leptokurtic ( $\nu = 1.8$ ) shapes for EP distribution. The parameters  $\beta_{ak}$  in the ratio function (3) with  $z_{a1it} = t$  are also set to include different trend patterns by varying the sign and magnitude of the true values.

Afterwards both models are fit to each data set using the Bayesian approach implemented by WinBUGS and R2WinBUGS and the parameter estimate  $\hat{\theta}$  is given by the average of the  $M = 1000$  posterior medians  $\hat{\theta}_j$ . To examine the bias and precision of the MCMC sampling algorithms, the standard deviation ( $SD$ ) of  $\hat{\theta}_j$  over  $r = 100$  simulated data sets for each parameter  $\theta$  is reported. To assess the accuracy of parameter estimates when the data set is simulated and fitted to the same model, we calculate the mean squared error ( $MSE$ ) for each parameter  $\theta$  which is given by:

$$MSE = \frac{1}{r} \sum_{j=1}^r (\hat{\theta}_j - \theta)^2.$$

For model selection, the average  $DIC$  and average squared error ( $ASE$ ) proposed by Wegman (1972) is used to assess the quality of the density estimator on the true pmf. Denote the true pmf of the RMPGP model by  $f_{bt}(w)$  at time  $t$  with covariate  $b$ , the  $ASE$  is used to compare the performance of the two models on the same simulated data set and is defined as

$$ASE = \sum_{b=0,1} \psi_b \sum_{t=1}^{n_i} \sum_{j=1}^r \sum_{w=0}^{\infty} \left\{ \hat{f}_{jbt}(w) - f_{bt}(w) \right\}^2$$

where  $\hat{f}_{jbt}(w)$  is the pmf estimator of (1) using Monte Carlo integration described in (5) for counts  $w$  in the  $j$ th simulated set at time  $t$  with treatment effect  $b$  and  $\psi_b = m_b/m$  is the weight associated with the control ( $b = 0$ ) or treatment group ( $b = 1$ ). Clearly, the smaller the  $ASE$ , the closer is the estimated pmf to the true one and thus the better is the model performance.

Table 3 summarizes the results of the four set of simulation experiments with the first two data sets simulated from the RPGP- $t$  model and the next two simulated from the RPGP-EP model. In general, the MCMC algorithms give unbiased and precise results as both  $MSE$  and  $SD$  of most parameters are reasonably small. Moreover, the values of the shape parameter  $\nu$  of the two models match with each other in terms of the tailedness. For example, the small  $\hat{\nu} = 5.286$  of the Student’s  $t$ -distribution agrees with  $\hat{\nu} = 1.8234$  which is close to 2 in the EP distribution. However, it is noticed that  $\nu$  of the RPGP- $t$  model has relatively lower precision and higher bias reflecting the higher level of difficulty in estimating the tailedness of the heavy-tailed Student’s  $t$ -distribution.

In model comparison, although the RPGP- $t$  model has a slightly smaller  $ASE$  (0.04504 versus 0.04817 averaged over the four simulated sets), the RPGP-EP model outperforms the RPGP- $t$  model in  $DIC$  (1968.4 versus 1979.48). All in all, the simulation experiment shows that the performance of the MCMC algorithms for the two models is satisfactory and the estimated pmfs  $\hat{f}_{jbt}(w)$  approximate the true pmfs  $f_{bt}(w)$  reasonably well. While EP distribution can be platykurtic or leptokurtic with a kink and Student’s  $t$ -distribution can give a very heavy tail when the outlying effect is tremendous, the two RPGP models are suitable under different circumstances.

### 5. Real data analysis

We illustrate the usefulness of our proposed models through the epilepsy data which can be found in Thall and Vail (1990). The data were collected from a clinical trial of 59 epileptics by Leppik et al. (1985). In the randomized controlled trial,

**Table 3**

Parameter estimates, SD, MSE, DIC and ASE under 4 simulated data sets based on 4 sets of true parameters from different models.

Data	Model	Parameter	True	Estimate	SD	MSE	DIC	ASE		
1	RPGP- <i>t</i>	$\beta_{a0}$	-2.0	-1.9320	0.0379	0.0060	<b>2315.55</b>	<b>0.0371</b>		
		$\beta_{a1}$	0.2	0.1960	0.0056	0.0000				
		$\beta_{\mu 0}$	-1.5	-1.6016	0.0307	0.0113				
		$\beta_{\mu 1}$	-0.5	-0.3232	0.0381	0.0327				
		$\nu$	2.5	5.2860	1.3266	9.5040				
		$\sigma$	0.1	0.4244	0.0377	0.1066				
		$\beta_{a0}$		-1.8844	0.0382				2357.72	0.0592
	RPGP-EP	$\beta_{a1}$		0.1892	0.0052					
		$\beta_{\mu 0}$		-1.5034	0.0254					
		$\beta_{\mu 1}$		-0.3433	0.0416					
		$\nu$		1.8234	0.1532					
		$\sigma$		0.1052	0.0132					
		2	RPGP- <i>t</i>	$\beta_{a0}$	-1.0	-0.8651	0.0546	0.0211	1834.54	<b>0.0367</b>
				$\beta_{a1}$	0.2	0.1820	0.0110	0.0004		
$\beta_{\mu 0}$	1.0			0.9794	0.0236	0.0010				
$\beta_{\mu 1}$	-0.5			-0.4528	0.0329	0.0033				
$\nu$	50.0			50.9375	10.6352	112.86				
$\sigma$	0.02			0.0203	0.0004	0.0000				
$\beta_{a0}$				-0.8523	0.0540		<b>1816.94</b>	0.0377		
RPGP-EP	$\beta_{a1}$			0.1803	0.0108					
	$\beta_{\mu 0}$			0.9907	0.0178					
	$\beta_{\mu 1}$			-0.4541	0.0389					
	$\nu$			1.0469	0.2979					
	$\sigma$			0.1014	0.0128					
	3		RPGP- <i>t</i>	$\beta_{a0}$		-0.4852	0.0795		1883.96	<b>0.0543</b>
				$\beta_{a1}$		0.0987	0.0110			
$\beta_{\mu 0}$				0.7384	0.0920					
$\beta_{\mu 1}$				-0.2350	0.0454					
$\nu$				73.6282	18.1843					
$\sigma$				0.3152	0.0654					
$\beta_{a0}$		-0.5		-0.4289	0.0729	0.0103	<b>1860.39</b>	0.0550		
RPGP-EP		$\beta_{a1}$	0.1	0.0926	0.0105	0.0002				
		$\beta_{\mu 0}$	1.0	0.8287	0.0769	0.0352				
		$\beta_{\mu 1}$	-0.5	-0.2812	0.0656	0.0521				
		$\nu$	0.1	0.2005	0.1178	0.0238				
		$\sigma$	0.5	0.6272	0.0701	0.0210				
		4	RPGP- <i>t</i>	$\beta_{a0}$		0.9938	0.0694		1883.86	0.0521
				$\beta_{a1}$		-0.0996	0.0088			
$\beta_{\mu 0}$				1.7407	0.0784					
$\beta_{\mu 1}$				-0.2256	0.0549					
$\nu$				2.2267	1.1224					
$\sigma$				0.1682	0.0028					
$\beta_{a0}$	1.0			1.1048	0.0663	0.0153	<b>1838.56</b>	<b>0.0407</b>		
RPGP-EP	$\beta_{a1}$		-0.1	-0.1110	0.0087	0.0002				
	$\beta_{\mu 0}$		2.0	1.9388	0.0502	0.0062				
	$\beta_{\mu 1}$		-0.5	-0.3768	0.0743	0.0207				
	$\nu$		1.8	1.8430	0.1174	0.0155				
	$\sigma$		0.2	0.1945	0.0225	0.0005				

$m = 59$  patients suffering from simple or complex partial seizures were assigned to either an antiepileptic drug progabide ( $z_{\mu i1} = 1$ ) or a placebo ( $z_{\mu i1} = 0$ ) with no intrinsic therapeutic value. The seizure counts were recorded at a two-week interval for an eight-week period ( $n_i = 4$ ) with no dropout or missing cases. As shown in Table 4, the seizure counts exhibit a prominent extra-Poisson variation with large variance to mean ratios at all time  $t$  due to some outlying observations as displayed in Fig. 4(a) and (b). To assess the overdispersion in the data, we fit a simple Poisson regression model using a mean link function  $\eta_{it} = \exp(\beta_0 + \beta_1 z_{\mu i1} + \beta_2 t)$  and a PGP model using a mean function  $\mu_{it} = \exp(\beta_{\mu 0} + \beta_{\mu 1} z_{\mu i1})$  and a ratio function  $a_{it} = \exp(\beta_{a0} + \beta_{a1} t)$ . The mean and variance under both models (indicated by ‘\*’ in Table 4) are equivalent and are given by  $\eta_{it}$  and  $\frac{\mu_{it}}{a_{it} - 1}$  respectively. Obviously, neither of the two simple models, as restricted by their equidispersed property, can capture the overdispersion. Besides, the higher mean seizure counts of the placebo group indicate that ‘treatment group’ is a feasible covariate. Moreover, the gradually decreasing seizure counts over time for both placebo and progabide groups suggests that time  $t$  maybe a possible time-evolving effect. In addition, population heterogeneity in terms of trend pattern

**Table 4**  
Mean and variance for the observed epilepsy data and under two simple fitted and the best models.

	Model	Time				Overall
		$t = 1$	$t = 2$	$t = 3$	$t = 4$	
Placebo mean	Observed	9.3571	8.2857	8.7857	8.0000	8.6071
	Poisson reg.*	9.3658	8.8385	8.3410	7.8714	8.6020
	PGP*	9.2073	8.9707	8.4645	7.7348	
	RMPGP-EP	10.0465	10.4001	9.9565	8.7905	
Variance	Observed	102.757	66.6561	215.286	57.9259	107.934
	RMPGP-EP	187.187	214.021	198.254	141.916	
Progabide mean	Observed	8.5806	8.4194	8.1290	6.7419	7.9677
	Poisson reg.*	8.6700	8.1819	7.7214	7.2867	7.9557
	PGP*	8.5212	8.3023	7.8337	7.1584	
	RMPGP-EP	6.8235	7.2885	6.9278	5.9980	
Variance	Observed	332.718	140.652	193.049	126.6645	193.966
	RMPGP-EP	92.114	124.374	107.435	77.371	
Overall mean	Observed	8.9492	8.3559	8.4407	7.3390	8.2712
	Poisson reg.*	8.9940	8.4920	8.0179	7.5704	8.2648
	PGP*	8.8729	8.6469	8.1573	7.4495	
	RMPGP-EP	8.3531	8.7652	8.3652	7.3233	
Variance	Observed	220.084	103.785	200.182	92.8831	152.607
	RMPGP-EP	137.234	166.918	150.536	108.003	

and count level are also detected intuitively. In consideration of these and the clinical interest of examining trend patterns, we adopt the RMPGP models to analyze the epilepsy data.

Referring to the MCMC algorithms detailed in Section 3.2, our prior specifications are mostly non-informative except for  $\nu_l$  in the RMPGP- $t$  model. In both RMPGP models, we assign  $\tau_{jkl}^2 = 1000$  in (12),  $c_l = d_l = 0.001$  in (13) and  $\alpha_l = 1/G$  in (16) for a  $G$ -group ( $G \geq 2$ ) model. For  $\nu_l$  in the RMPGP- $t$  model, we take  $h_l = 20$  since there is a high degree of overdispersion in the seizure counts. After implementing the MCMC algorithms in WinBUGS, the posterior sample means are adopted as parameter estimates since the posterior densities of most model parameters are highly symmetric and the posterior sample mean are close to the posterior sample median. Table 5 summarizes the parameter estimates, standard errors (SE), 95% credibility interval (CI) and model selection criterion of the fitted models.

We first fitted a simple RPGP model with treatment group ( $z_{\mu lit} = 0, 1$ ) as the covariate in the mean function  $\mu_{lit}$  in (8) and two-week interval ( $z_{ait} = t = 1, 2, 3, 4$ ) as the time-evolving effect in the ratio function  $a_{itl}$  in (9). The negative  $\beta_{\mu 1}$  in both RPGP models indicate the treatment effect that patients receiving progabide are associated with lower seizure counts. However, within the treatment group, it is explicit that some of these patients have abnormally high seizure counts. Simply fitting a simple RPGP model may fail to allow for the clustering effect among patients receiving the same treatment. We therefore fitted a 2-group RMPGP model and also attempted to fit a 3-group RMPGP model using both Student's  $t$ - and EP distributions but the results indicated that one of the groups in the 3-group RMPGP model degenerated and hence the models were discarded.

Not surprisingly, both RMPGP models give parallel results as they share some common model properties except the shape of the distribution of  $y_{itl}^*$ . In the RMPGP- $t$  model, two distinct groups of patients were identified with the first group of patients having generally higher seizure counts and is named as the high-level group ( $l = 1$ ). Within this group, 54% of the patients are receiving progabide and they have lower seizure counts in general ( $\beta_{\mu 11} < 0$ ) than those receiving placebo. Whereas in the low-level ( $l = 2$ ) group, 49% of the patients belong to the progabide group and again they generally have less epileptic seizures during the studying period ( $\beta_{\mu 12} < 0$ ). Besides, the ratio function  $a_{it1}$  in (9) reveals that there is a slightly decreasing trend in the seizure counts in the high-level group while  $a_{it2}$  indicates that no obvious trend is detected in the low-level group. In addition, comparing with the low-level group, the relatively smaller  $\nu_1$  shows that the high-level group has a higher degree of overdispersion in the seizure counts due to the existence of some abnormally large observations as revealed in Fig. 4(a) and (b).

As expected, the RMPGP-EP model gives consistent results in terms of trend pattern and treatment effect. Moreover, the group membership of the patients has a close affinity with that of RMPGP- $t$  model and two diverse groups, high-level and low-level groups, are recognized as well. But for the high-level group, despite the comparable mean level, the ratio function shows that the seizure count increases in the second 2-week interval before it drops in the next two 2-week intervals. Coherently, the estimate of  $\nu_1 = 1.49$  agrees with the small  $\nu_1 = 7.21$  in the RMPGP- $t$  model indicating that the distribution of the seizure counts in the high-level group has a heavier tail to account for the higher degree of overdispersion. On the other hand, the smaller  $\nu_2 = 0.24$  indicates the data distribution is more uniform in the low-level group.

For model selection, the smaller DIC given by (18) as shown in Table 5 for the RMPGP-EP model manifests its better model fit on the epileptic seizure counts after accounting for the model complexity. A plausible explanation is that the EP distribution has a more flexible tail behaviour and thus provides a better fit to the data. To further investigate this, their observed and fitted pmfs for the low-level group are illustrated in Fig. 5(a)–(d) at different time points, in which the observed

**Table 5**  
Parameter estimates with SE and DIC for the epilepsy data.

Parameter	RMPGP- <i>t</i> model			RMPGP-EP model			
	Estimate	SE	95% CI	Estimate	SE	95% CI	
<i>G</i> = 1	$\beta_{a0}$	-0.6435	0.1464	(-0.9193, -0.3188)	0.0073	0.2780	(-0.5049, 0.5725)
	$\beta_{a1}$	0.1218	0.0456	(0.0281, 0.2094)	0.0089	0.0675	(-0.1284, 0.1359)
	$\beta_{\mu 0}$	1.2970	0.1132	(1.0730, 1.4930)	1.7460	0.1582	(1.4750, 2.0890)
	$\beta_{\mu 1}$	-0.2177	0.1393	(-0.4838, 0.0574)	-0.2637	0.1699	(-0.5564, 0.1181)
	$\nu$	73.210	41.920	(9.1650, 145.00)	1.1920	0.1883	(0.8772, 1.5170)
	$\sigma$	0.9826	0.0669	(0.8607, 1.1290)	0.9036	0.1616	(0.6618, 1.2000)
	DIC	1923.08			1945.40		
<i>G</i> = 2	$\beta_{a01}$	-0.0241	0.4366	(-0.8885, 0.8952)	-0.2410	0.3122	(-0.7459, 0.5274)
	$\beta_{a11}$	0.0192	0.1062	(-0.1984, 0.2282)	0.0739	0.0738	(-0.1149, 0.1963)
High-level ( <i>l</i> = 1)	$\beta_{\mu 01}$	2.7070	0.1993	(2.2910, 3.0860)	2.6700	0.1569	(2.4100, 2.9850)
	$\beta_{\mu 11}$	-0.3544	0.1922	(-0.7197, 0.0426)	-0.3727	0.1342	(-0.5992, -0.0921)
	$\nu_1$	7.2110	4.7430	(1.9570, 18.630)	1.4860	0.2559	(0.9858, 1.9210)
	$\sigma_1$	0.5965	0.1029	(0.3955, 0.7890)	0.4841	0.1304	(0.2969, 0.7882)
	$\pi_1$	0.4055	0.0738	(0.2603, 0.5523)	0.4069	0.0688	(0.2730, 0.5421)
	$\beta_{a02}$	0.2058	0.2378	(-0.1916, 0.6775)	0.2169	0.2271	(-0.2113, 0.6907)
	$\beta_{a12}$	-0.0464	0.0589	(-0.1611, 0.0579)	-0.0474	0.0581	(-0.1634, 0.0613)
Low-level ( <i>l</i> = 2)	$\beta_{\mu 02}$	1.2750	0.1252	(1.0220, 1.5230)	1.2750	0.0982	(1.0880, 1.4600)
	$\beta_{\mu 12}$	-0.3587	0.1656	(-0.6734, -0.0460)	-0.4193	0.1477	(-0.7189, -0.1340)
	$\nu_2$	10.930	5.2690	(1.6580, 19.560)	0.2424	0.1813	(0.0118, 0.6970)
	$\sigma_2$	0.3286	0.1148	(0.1028, 0.5491)	0.7327	0.1058	(0.5191, 0.9433)
	DIC	1700.90			1660.66		

pmf  $f_{il}(w)$  for group  $l$  at time  $t$  is generally given by

$$f_{il}(w) = \sum_{b=0,1} \psi_b \left\{ \frac{\sum_{i=1}^m \hat{I}_{il} I(W_{it} = w) I(z_{\mu 1i} = b)}{\sum_{w'=0}^{\infty} \sum_{i=1}^m \hat{I}_{il} I(W_{it} = w') I(z_{\mu 1i} = b)} \right\}, \quad w = 0, 1, 2, \dots \tag{19}$$

where  $I(W_{it} = w)$  is an indicator which returns 1 when  $W_{it} = w$  for patient  $i$  at time  $t$  in group  $l$  and 0 otherwise,  $I(z_{\mu 1i} = b)$  indicates the treatment group  $b$  of patient  $i$ ,  $\hat{I}_{il}$  is the posterior mean of the group membership indicator and  $\psi_b$  is the weight associated with the placebo ( $b = 0$ ) or progabide group ( $b = 1$ ). On the other hand, the fitted pmf  $\hat{f}_{il}(w)$  is simply obtained by the numerical approximation described in (5) based on the parameter estimates  $\theta_l$  and is weighted by  $\psi_b$ . Based on Fig. 5(a)–(d), both models imitate the observed pmfs pretty well. However the estimated trend in the ratio function cannot accommodate the upsurge in observation  $w = 4$  at  $t = 2$  resulting in a discrepancy between the observed and fitted pmfs in Fig. 5(b). It is not surprised to find that the two RMPGP models give similar pmfs as both have the capability of modelling highly overdispersed data. Nevertheless, despite of the affinity, the slightly heavier tail in the distribution of the RMPGP-EP model possibly gives rise to the better DIC.

In the best model, the RMPGP-EP model, since extra variation is added to the mean of the Poisson distribution, the variances of the estimated pmf  $\hat{f}_{il}(w)$  of each treatment group and the overall variances which comprises the variance of expectation and the expectation of variance conditional on the mixture group being known, show a dramatic improvement and are reported in Table 4.

Advantageously, implementing the model using Bayesian approach enables us to study the latent stochastic process  $y_{itl}^*$  and the mixing parameters  $\lambda_{itl}$  which can be output in the coda in WinBUGS. Under the RMPGP-EP model, to examine the density of the unobserved  $y_{itl}^*$  which are simulated from an EP distribution with location parameter  $\mu_{itl}^*$ , scale parameter  $\sigma_1^2$  and shape parameter  $\nu_l$ , we compare the densities of the posterior samples of  $y_{itl}^*$  with the normal distribution which has mean and standard deviation of the posterior samples and the EP distribution with same mean and standard deviation and a shape parameter  $\hat{\nu}_l$ . Four selected  $y_{itl}$  from each cluster group  $l$  and treatment group  $z_{\mu 1i}$  are illustrated in the following Fig. 6. Obviously,  $y_{it1}$ 's have a leptokurtic shape whereas  $y_{it2}$ 's appear to be more uniform than the normal distribution. These agree with the results in Table 5 that the shape parameter of the high-level group is larger than that of the low-level group ( $\nu_1 > \nu_2$ ) and explain how the EP distribution can downweigh the outlying effect.

Last but not least, the outlier diagnosis is performed using the mixing parameters in the better model, the RMPGP-EP model. For each cluster group  $l$ , the mixing parameters  $\lambda_{itl}$  are plotted with the standardized observations  $w'_{itl}$  of patients who are classified to group  $l$  in Fig. 7. For fair comparison, the seizure count is standardized as  $w'_{itl} = |w_{it} - \hat{w}_{itl}| / \hat{\sigma}_{w_{itl}}$  where  $\hat{w}_{itl}$  and  $\hat{\sigma}_{w_{itl}}$  are respectively the mean and standard deviation of the estimated pmf  $f_{il}(w)$  in (19) at time  $t$  under group  $l$  of the RMPGP-EP model. An unusually large  $\lambda_{itl}$  indicates that the  $w_{it}$  is possibly an outlier under group  $l$ . Both mixing parameters  $\lambda_{itl}$  and standardized seizure counts  $w'_{itl}$  are sorted by groups ( $l = 1, 2$ ) for better visualization and are graphed in Fig. 7.

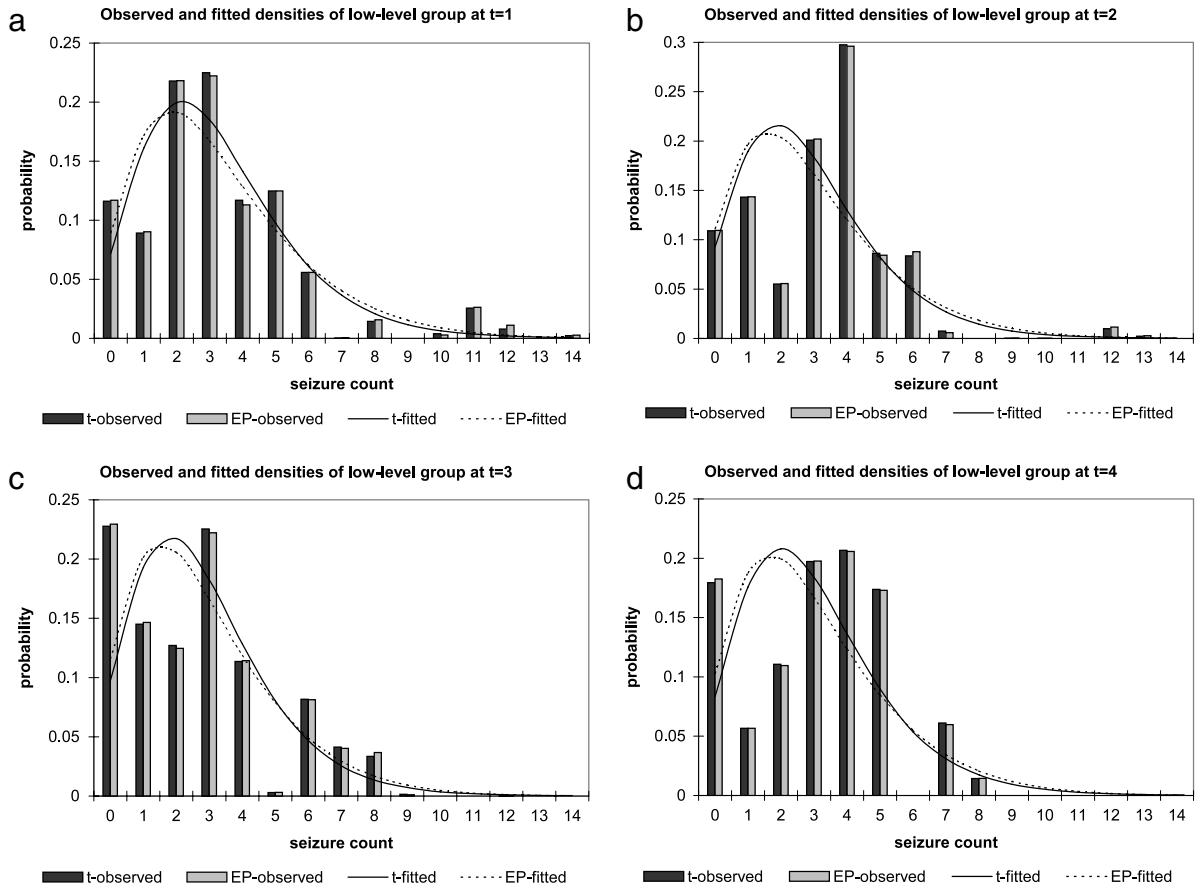


Fig. 5. Distributions of low-level group at  $t = 1, 2, 3$  and 4.

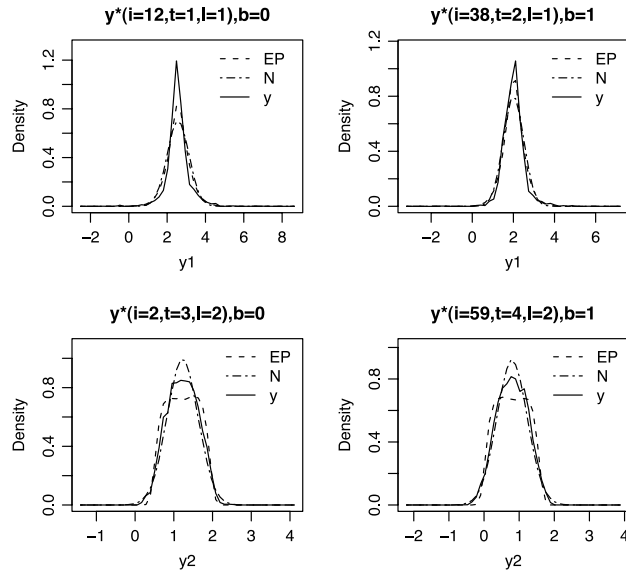


Fig. 6. Comparison of densities of  $y_{it}$ 's with normal and EP distributions.

Clearly, all the top 10 (5%) outlying counts with the first 10 largest  $\lambda_{itl}$  locate in the high-level group ( $l = 1$ ) due to the presence of some extreme observations. They are highlighted in Fig. 7 with the corresponding rank and observation in parentheses. Not surprisingly, the correlation between  $\lambda_{it1}$  and  $w'_{it1}$  is high ( $r = 0.9731$ ) which signifies the appropriateness



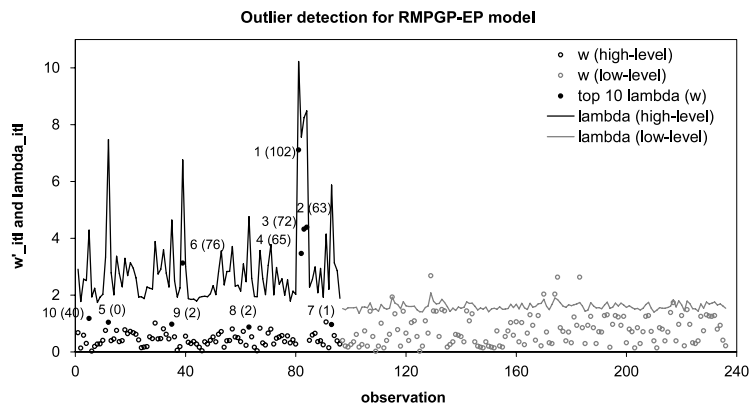


Fig. 7. Outlier diagnosis using  $\lambda_{itl}$  in the RMPGP-EP model.

of using the mixing parameters in outlier diagnosis. While the outlying effect is not substantial in the low-level group ( $l = 2$ ),  $\lambda_{it2}$  appears to be relatively smaller.

The top 4 outlying values come from the same patient with ID 49 who has abnormally high seizure counts at each 2-week interval. Besides, 2 large observations are identified from patients with ID 8, and 25. In spite of large observation, four small seizure counts are also classified as outlier from patients with ID 10, 24, 39 and 56. Knowing which patients are associated with abnormal seizure counts, specialists can pay more attention to their abnormalities and alternative treatments maybe considered. At the same time, the RMPGP-EP model has downweighed the outlying effect and thus the general trend pattern is not distorted by the extreme observations.

## 6. Conclusion

In this paper, we propose using a Poisson geometric process (PGP) model to analyze repeated measurements of counts over time. The model is essentially a state space model with a ratio function which describes the direction and strength of the trend and a mean function which reveals the initial level and studies the covariate effect. However, the PGP model fails to allow for extra-Poisson variation when extreme observations appear in the data. Ignoring the outlying effects may also lead to overestimated mean and variance resulting in invalid interpretation and prediction. As remedies, two methods are suggested to account for overdispersion which include adopting a heavy-tailed distribution and incorporating mixture effect. The latter can handle clustering effect in the data and overdispersion arisen from that may also be captured.

As a new direction to account for population heterogeneity, we apply the heavy-tailed distributions in the modelling of time series of counts and pioneer the robust Poisson geometric process (RPGP) model. This model allows the mean  $X_{itl}$  of the Poisson distribution to follow a GP and the logarithm of the underlying stochastic process  $\{Y_{it} = a_{it}^{l-1} X_{it}\}$  follows a heavy-tailed distribution. The resultant model is called the RPGP model. By varying a set of model parameters, the properties of the RPGP models are reported in Tables 1 and 2 and their pmfs are revealed in Figs. 1 and 2. Although the marginal pmfs do not have a closed-form, Monte Carlo integration in (5) can be used to approximate the pmf, and hence the mean as well as the variance. Tables 1 and 2 show that the model can accommodate both equidispersed and overdispersed data with varying degrees of kurtosis.

The RPGP models and its extension, the RMPGP models to allow clustering effects are implemented using a Bayesian approach. Expressing the heavy-tailed distributions in a scale mixture form facilitates the model implementation using MCMC algorithms and the mixing parameters enables us to perform outlier diagnosis as shown in the real data analysis. Here, the Student's  $t$ -distribution in scale mixture of normal (SMN) and exponential power (EP) distribution in scale mixture of uniform (SMU) are adopted. The resultant models can be efficiently implemented via the user-friendly software, WinBUGS. Moreover, the posterior densities for the MCMC algorithm are derived for the RMPGP-EP model.

The simulation study shows that the performances of the two RMPGP models are comparable and satisfactory. In case of data with very long tail, the RMPGP- $t$  model seems to fit better since Student's  $t$ -distribution allows a much heavier than normal distribution. On the other hand, the RMPGP-EP model gives a better fit in the analysis of epilepsy data with diverse degree of overdispersion across mixture groups as EP distribution has a more flexible tail which can be either leptokurtic or platykurtic.

One pitfall of our proposed model is that taking log-transformation of the latent  $Y_{itl}$  inevitably causes those data associated with close-to-zero means being identified as outliers. Hence when zero is dominant in the data, the proposed RMPGP models can be extended to include a zero-altered component (Wan and Chan, 2009).

## Appendix

See Tables 1–5 and Figs. 1–7.

## References

- Box, G.E.P., Tiao, G.C., 1973. *Bayesian Inference in Statistical Analysis*. Addison-Wesley, Massachusetts.
- Chan, J.S.K., Choy, S.T.B., Makov, U.E., 2008. Robust Bayesian analysis of loss reserve data using the generalized-t distribution. *ASTIN Bulletin* 38, 207–230.
- Chan, J.S.K., Leung, D.Y.P., 2010. Binary geometric process model for the modelling of longitudinal binary data with trend. *Computational Statistics* doi:10.1007/s00180-010-0190-8.
- Chan, J.S.K., Yu, P.L.H., Lam, Y., Ho, A.P.K., 2006. Modelling SARS data using threshold geometric process. *Statistics in Medicine* 25, 1826–1839.
- Choy, S.T.B., Chan, C.M., 2003. Scale mixtures distributions in insurance applications. *ASTIN Bulletin* 33, 93–104.
- Choy, S.T.B., Smith, A.M.F., 1997. Hierarchical models with scale mixtures of normal distributions. *TEST* 6, 205–221.
- Choy, S.T.B., Walker, S.G., 2003. The extended exponential power distribution and Bayesian robustness. *Statistics & Probability Letters* 65, 227–232.
- Gupta, R.C., Ong, S.H., 2005. Analysis of long-tailed count data by Poisson mixtures. *Communications in statistics—Theory and Methods* 34, 557–573.
- Johnson, N.L., Kotz, S., Balakrishnan, N., 1995. *Continuous Univariate Distributions*, vol. 2, second ed.. Wiley, New York.
- Jowaheer, V., Sutradhar, B.C., 2002. Analysing longitudinal count data with overdispersion. *Biometrika* 89, 389–399.
- Jowaheer, V., Sutradhar, B.C., Sneddon, E., 2009. On familial Poisson mixed models with multi-dimensional random effects. *Journal of Statistical Computation and Simulation* 79, 1043–1062.
- Lam, Y., 1988a. Geometric process and replacement problem. *Acta Mathematicae Applicatae Sinica* 4, 366–377.
- Lam, Y., 1988b. A note on the optimal replacement problem. *Advances in Applied Probability* 20, 479–482.
- Lam, Y., Chan, J.S.K., 1998. Statistical inference for geometric processes with lognormal distribution. *Computational Statistics and Data Analysis* 27, 99–112.
- Lam, Y., Zhang, Y.L., 2003. A geometric-process maintenance model for a deteriorating system under a random environment. *IEEE Transactions on Reliability* 52, 83–89.
- Lam, Y., Zhu, L.X., Chan, J.S.K., Liu, Q., 2004. Analysis of data from a series of events by a geometric process model. *Acta Mathematicae Applicatae Sinica. English Series* 20, 263–282.
- Leppik, I.E., Dreifuss, F.E., Bowman-Cloyd, T., Santilli, N., Jacobs, M., Crosby, C., Cloyd, J., Stockman, J., Graves, N., Sutula, T., Welty, T., Vickery, J., Brundage, R., Gumnit, R., Guitierrez, A., 1985. A double-blind crossover evaluation of progabide in partial seizures. *Neurology* 35, 285.
- Qin, Z., Damien, P., Walker, S., 1988. Uniform scale mixture models with applications to Bayesian inference. Working Paper #98 005. University of Michigan Business School, Michigan.
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A., 2002. Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B* 64, 583–616.
- Thall, P.F., Vail, S.C., 1990. Some covariance models for longitudinal count data with overdispersion. *Biometrics* 46, 657–671.
- Walker, S.G., Gutiérrez-Peña, E., 1999. Robustifying Bayesian procedures. In: Bernardo, J.M., Dawid, A.P., Smith, A.F.M. (Eds.), *Bayesian Statistics*, vol. 6. Oxford, New York, pp. 685–710.
- Wan, W.Y., 2006. Analysis of Poisson count data using geometric process model. M. Phil. Thesis. University of Hong Kong, Hong Kong.
- Wan, W.Y., Chan, J.S.K., 2009. A new approach for handling longitudinal count data with zero-inflation and overdispersion: Poisson geometric process model. *Biometrical Journal* 51, 556–570.
- Wegman, E.J., 1972. Nonparametric probability density estimation: a comparison of density estimation methods. *Journal of Statistical Computation and Simulation* 1, 225–245.