Applied Mathematical Sciences, Vol. 6, 2012, no. 137, 6843 - 6856

An EM Algorithm for Multivariate Mixed Poisson Regression Models and its Application

M. E. Ghitany[1], D. Karlis[2], D.K. Al-Mutairi[1] and F. A. Al-Awadhi[1]

[1]Department of Statistics and Operations Research
Faculty of Science, Kuwait University, Kuwait

[2]Department of Statistics
Athens University of Economics, Greece

Abstract

Although the literature on univariate count regression models allowing for overdispersion is huge, there are few multivariate count regression models allowing for correlation and overdiseprsion. The latter models can find applications in several disciplines such as epidemiology, marketing, sports statistics, criminology, just to name a few. In this paper, we propose a general EM algorithm to facilitate maximum likelihood estimation for a class of multivariate mixed Poisson regression models. We give special emphasis to the multivariate negative binomial, Poisson inverse Gaussian and Poisson lognormal regression models. An application to a real dataset is also given to illustrate the use of the proposed EM algorithm to the considered multivariate regression models.

Mathematics Subject Classification: 62J05, 65D15

Keywords: Mixed Poisson distributions, overdispersion, covariates, EM algorithm

1 Introduction

Count data occur in several different disciplines. When only one count variable is considered the literature is vast. There are various models to fit such data and to make inferences on them. For example, traditionally one may use the simple Poisson distribution or extensions like mixed Poisson models which allow for overdispersion in the data (sample variance exceeds the sample mean). When considering jointly two or more count variables, things are more complicated and the literature is much smaller. Modeling correlated count data is also important in certain disciplines, for example in epidemiology when more than one disease is considered or in marketing when purchase frequency of many items is of interest. In the presence of covariates information, for the univariate case the researcher has a plethora of models available, but for the case of multivariate count regression models are less developed.

The kind of models that we will consider in this paper are as follows. Suppose we have m independent Poisson random variables, $X_1, \ldots, X_m$, with respective means $\alpha\theta_1, \ldots, \alpha\theta_m$, where α is a random variable from some mixing distribution. In practice, the mixing distribution introduces overdispersion, but since α is common to all the X_j's it also introduces correlation. We further assume that the parameters θ_j, $j = 1, \ldots, m$, are connected through a log link function with some covariates, namely

$$\log\theta_j = \boldsymbol{\beta}_j^T \mathbf{z}_j, \qquad j = 1, \ldots, m,$$

where $\boldsymbol{\beta}_j$ is a $(p+1)$-dimensional vector of regression coefficients and $\mathbf{z}_j$ is a $(p+1)$-dimensional vector of covariates associated with variate X_j. To ensure model identifiability, it is customary to assume $E(\alpha) = 1$.

The literature on this approach contains the works of Stein and Yuritz (1987), Stein *et al.* (1987) and Kocherlakota (1988) for the case without covariates. Munkin and Trivedi (1999) described multivariate mixed Poisson regression models using a gamma mixing distribution. Gurmu and Elder (2000) used an extended Gamma density as a mixing distribution.

The paper is organized is as follows. In Section 2, we give a detailed description of the proposed multivariate regression models. In Section 3, we consider gamma, inverse-Gaussian and lognormal mixing distributions, each with unit mean, and derive the joint probability mass function (jpmf) of the corresponding multivariate mixed Poisson regression models. Then, we propose a general EM algorithm to facilitate ML estimation for multivariate mixed Poisson regression models in Section 4. Detailed EM algorithms for the considered multivariate mixed Poisson regression models are given in Section 5. In Section 6, we apply the proposed EM algorithm to a real dataset on the demand for health care in Australia using the considered multivariate mixed Poisson regression models. Finally, some concluding remarks are given.

2 The model

The multivariate mixed Poisson regression model considered in this paper is described as follows. Let $X_{ij} \sim Poisson(\alpha_i\theta_{ij})$, $i = 1, \ldots, n, \quad j = 1, \ldots, m$, α_i are independent and identically distributed (i.i.d) random variables from a mixing distribution with cumulative distribution function (c.d.f.) $G(\alpha; \phi)$ where ϕ is a vector of parameters. To allow for regressors, we let $\log\theta_{ij} =$

$\boldsymbol{\beta}_j^T \mathbf{z}_{ij}$, where

$$\boldsymbol{\beta}_j^T = (\beta_{0j}, \beta_{1j}, \ldots, \beta_{pj}), \qquad j = 1, 2, \ldots, m,$$

are $(p+1)$-dimentional vectors of regression coefficients associated with the j-th variable and

$$\mathbf{z}_{ij}^T = (1, z_{1ij}, \ldots, z_{pij}), \qquad i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, m,$$

are $(p+1)$-dimensional vectors of covariates for the i-th observation related to the j-th variable.

In the present paper, we consider α as a continuous random variable with support on $(0, \infty)$. In order to achieve identifiability of the model, we assume that $E(\alpha) = 1$. Note that α is considered as a frailty.

The joint probability mass function (jpmf) of the given model (dropping the observation specific subscript i) is given by

$$P(x_1, \ldots, x_m; \phi) = \int_0^\infty \prod_{j=1}^m \frac{\exp(-\alpha\theta_j)(\alpha\theta_j)^{x_j}}{x_j!} \, g(\alpha; \phi) d\alpha, \tag{1}$$

where $g(\alpha; \phi)$ is the probability density function (pdf) of α.

Some properties associated with the mixed Poisson model given in (1) are as follows.

(i) The marginal distribution of $X_j, \; j = 1, 2, \ldots, m$, is a mixed Poisson distribution with the same mixing distribution $g(\alpha; \cdot)$.

(ii) The variance of X_j is

$$Var(X_j) = \theta_j \; (1 + \theta_j \; \sigma^2),$$

where σ^2 is the variance of α.

(iii) The covariance between X_j and X_k is

$$Cov(X_j, X_k) = \theta_j \; \theta_k \; \sigma^2, \qquad j \neq k.$$

Since $\sigma^2 > 0$, the covariance (correlation) is always positive.

(iv) The generalized variance ratio (GVR) between a multivariate mixed Poisson model, i.e. $X_j \sim Poisson(\alpha\theta_j), j = 1, \ldots, m$, $\alpha \sim G(\cdot; \phi)$ and a simple Poisson model, i.e. $Y_j \sim Poisson(\theta_j), j = 1, \ldots, m$, is given by

$$GVR = \frac{\sum\limits_{j=1}^m Var(X_j) + 2 \sum\limits_{j<k} Cov(X_j, X_k)}{\sum\limits_{j=1}^m Var(Y_j)} = 1 + \sigma^2 \sum_{j=1}^m \theta_j.$$

Note that $GVR > 1$ for a continuous mixing distribution. Hence, the mixing distribution introduces a multivariate overdispersion. Also, the GVR increases as the variance of the mixing distribution increases.

3 Some multivariate mixed Poisson regression models

In this section, we consider some multivariate mixed Poisson regression models based on gamma, inverse-Gaussian and log-normal mixing distributions. These models will be called Multivariate negative binomial, Multivariate Poisson-Inverse Gaussian and Multivariate Poisson-lognormal, respectively.

3.1. Multivariate negative binomial

The negative binomial is traditionally the most widely known and applied mixed Poisson distribution. So a natural choice for the mixing distribution is to consider a gamma (G) distribution with pdf

$$g(\alpha;\gamma) = \frac{\gamma^\gamma}{\Gamma(\gamma)}\alpha^{\gamma-1}\exp(-\gamma\alpha), \qquad \alpha,\gamma > 0,$$

i.e. a gamma density such that $E(\alpha) = 1$ and $\sigma^2 = 1/\gamma$.

The resulting multivariate negative binomial (MNB) distribution has jpmf:

$$P_G(x_1,\ldots,x_m;\gamma) = \frac{\Gamma(\sum\limits_{j=1}^m x_j + \gamma)}{\Gamma(\gamma)\prod\limits_{j=1}^m x_j!}\frac{\gamma^\gamma\prod\limits_{j=1}^m \theta_j^{x_j}}{(\gamma + \sum\limits_{j=1}^m \theta_j)^{\sum\limits_{j=1}^m x_j + \gamma}}. \tag{2}$$

To allow for regressors, we assume that $\log\theta_{ij} = \boldsymbol{\beta}_j^T\mathbf{z}_{ij}$.

3.2. Multivariate Poisson-Inverse Gaussian

The multivariate Poisson inverse Gaussian (MPIG) model is based on an inverse Gaussian (IG) mixing distribution with pdf

$$g(\alpha;\delta) = \frac{\delta}{\sqrt{2\pi}}\exp(\delta^2)\ \alpha^{-3/2}\ \exp\left(-\frac{\delta^2}{2}\left(\frac{1}{\alpha}+\alpha\right)\right), \qquad \alpha,\delta > 0.$$

i.e. an inverse Gaussian distribution such that $E(\alpha) = 1$ and $\sigma^2 = 1/\delta^2$.

The resulting MPIG distribution has jpmf:

$$P_{IG}(x_1,\ldots,x_m;\delta) = \frac{2\delta\exp(\delta^2)}{\sqrt{2\pi}}K_{\sum\limits_{j=1}^m x_j - 1/2}(\delta\Delta)\left(\frac{\delta}{\Delta}\right)^{\sum\limits_{j=1}^m x_j - 1/2}\prod_{j=1}^m \frac{\theta_j^{x^j}}{x_j!}, \tag{3}$$

where $\Delta = \sqrt{\delta^2 + 2\sum\limits_{j=1}^m \theta_j}$ and $K_r(x)$ denotes the modified Bessel function of the third kind of order r. To allow for regressors we assume $\log\theta_{ij} = \boldsymbol{\beta}_j^T\mathbf{z}_{ij}$.

Properties of the distribution given in (3) can be found in Stein and Yuritz (1987) and Stein et al. (1987). Our proposed MPIG model generalizes the one in Dean et al. (1989).

Note that both the gamma and IG mixing distributions are special cases of a larger family of distributions called the Generalized inverse Gaussian (GIG) family of distributions, see Jorgensen (1982), with density function

$$g(\alpha;\lambda,\gamma,\delta)=\left(\frac{\gamma}{\delta}\right)^{\lambda}\frac{\alpha^{\lambda-1}}{2K_{\lambda}(\delta\gamma)}\exp\left(-\frac{1}{2}\left(\frac{\delta^2}{\alpha}+\gamma^2\alpha\right)\right),\qquad \alpha>0,$$

where the parameter space is given by

$$\{\lambda<0,\gamma=0,\delta>0\}\cup\{\lambda>0,\gamma>0,\delta=0\}\cup\{-\infty<\lambda<\infty,\gamma>0,\delta>0\}.$$

This distribution will be denoted by $GIG(\lambda,\gamma,\delta)$.

The gamma distribution arises when $\lambda>0,\gamma>0,\delta=0$ and the IG distribution arises when $\lambda=-1/2,\gamma>0,\delta>0$.

3.3. Multivariate Poisson-lognormal

Consider another plausible and commonly used mixing distribution, namely a log-normal (LN) distribution with density

$$g(\alpha;\nu)=\frac{1}{\sqrt{2\pi}\ \nu\ \alpha}\ \exp\left(-\frac{(\log(\alpha)+\nu^2/2)^2}{2\nu^2}\right),\qquad \alpha,\nu>0,$$

such that $E(\alpha)=1$ and $\sigma^2=\exp(\nu^2)-1$.

The resulting multivariate Poisson-lognormal (MPLN) has jpmf:

$$P_{LN}(x_1,\ldots,x_m;\nu)=\int_0^{\infty}\prod_{j=1}^{m}\frac{\exp(-\alpha\theta_j)(\alpha\theta_j)^{x_j}}{x_j!}\ \frac{\exp\left(-\frac{(\log(\alpha)+\nu^2/2)^2}{2\nu^2}\right)}{\sqrt{2\pi}\ \nu\ \alpha}d\alpha.\quad (4)$$

Unfortunately, the last integral cannot be simplified and hence numerical integration is needed. To allow for regressors we assume $\log\theta_{ij}=\boldsymbol{\beta}_j^T\mathbf{z}_{ij}$. The MPLN distribution given in (4) is different than the one used in Aitchison and Ho (1989).

4 General EM algorithm for ML estimation

One of the main obstacles for using multivariate mixed Poisson regression models is that the log-likelihood is complicated and hence its maximization needs a special effort. For example, Munkin and Trivedi (1999) used a simulated maximum likelihood method. In this section, we develop an EM type algorithm which is easy to use for finding the MLEs of the model's parameters, since we do not use derivative based optimization.

The EM algorithm is a powerful algorithm for ML estimation for data containing missing values or being considered as containing missing values. This formulation is particularly suitable for distributions arising as mixtures since the mixing operation can be considered as producing missing data. The unobserved quantities are simply the realizations α_i of the unobserved mixing parameter for the i-th observation. Hence at the E-step one needs to calculate the conditional expectation of some functions of α_i's and then to maximize the likelihood of the complete model which reduces to maximizing the likelihood of the mixing density. For more details, see Karlis (2001, 2005). Formally, the EM algorithm can be described as follows:

- E-Step Using the current estimates, say $\phi^{(r)}$ taken from the r-th iteration, calculate the pseudo-values $t_{si} = E(h_s(\alpha_i) \mid \boldsymbol{x}_i, \phi^{(r)})$, for $i = 1, \ldots, n, s = 1, \ldots, k$, where $h_s(.)$ are certain functions.
- M-Step Use the pseudovalues t_{si} from the E-step to maximize the likelihood of the mixing distribution and obtain the updated estimates $\phi^{(r+1)}$.
- Iterate between the E-step and the M-step until some convergence criterion is satisfied, for example the relative change in log-likelihood between two successive iterations is smaller than 10^{-8}.

For a linear function $h_s(\alpha_i)$, the conditional posterior expectations can be easily and accurately obtained as it will be shown in the next section. For more complicated functions, if the exact solution is not available, one may proceed either by appropriate approximations based on Taylor approximations or by numerical approximations including numerical integration and/or simulation based approximations. All the above solutions seem to work well in practice.

The M-step is somewhat obvious and depends on the assumed mixing distribution. For some distributions a special iterative scheme may be appropriate, and perhaps another EM algorithm.

5 EM algorithm for some multivariate mixed Poisson regression models

We start this section with a preliminary lemma for calculating positive or negative moments of the posterior distribution of $\alpha \mid \boldsymbol{x}$. These moments will be very useful for implementing the E-step of the EM algorithm described in Section 4.

Let $\mathbf{1}_j$ denote the vector with all entries 0 apart from the j-th entry which is 1, i.e. $\mathbf{1}_1 = (1, 0, 0, \ldots, 0)$, $\mathbf{1}_2 = (0, 1, 0, \ldots, 0)$ etc.

Lemma 1. Suppose that the vector $\boldsymbol{X} = (X_1, \ldots, X_m)$ follows a multivariate mixed Poisson model with jpmf $P(\boldsymbol{x}; \phi) = P(x_1, \ldots, x_m; \phi)$ given in (1). Then, for $r = 1, 2, \ldots$ and $j = 1, 2, \ldots m$,

$$E(\alpha^r \mid \boldsymbol{x}) = \frac{(x_j+1)\ldots(x_j+r)}{\theta_j^r} \frac{P(\boldsymbol{x} + r\mathbf{1}_j; \phi)}{P(\boldsymbol{x}; \phi)},$$

and

$$E(\alpha^{-r} \mid \boldsymbol{x}) = \frac{\theta_j^r}{x_j(x_j-1)\ldots(x_j-r+1)} \frac{P(\boldsymbol{x} - r\mathbf{1}_j; \phi)}{P(\boldsymbol{x}; \phi)}, \quad x_j \neq 0, 1, \ldots, r-1.$$

Proof. Without loss of generality, we prove the lemma for the case $j = 1$. The posterior density $g(\alpha \mid \boldsymbol{x})$ is given by

$$g(\alpha \mid \boldsymbol{x}) = \frac{P(\boldsymbol{x}|\alpha)\ g(\alpha; \phi)}{P(\boldsymbol{x}; \phi)}.$$

Thus, for $r = 1, 2, \ldots,$ we have

$$\begin{aligned} E(\alpha^r \mid \boldsymbol{x}) &= \frac{\int_0^\infty \alpha^r \prod_{j=1}^m \frac{\exp(-\alpha\theta_j)\ (\alpha\theta_j)^{x_j}}{x_j!}\ g(\alpha; \phi) d\alpha}{P(\boldsymbol{x}; \phi)} \\ &= \frac{\int_0^\infty \frac{(x_1+1)\ldots(x_1+r) \exp(-\alpha\theta_1)(\alpha\theta_1)^{x_1+r}}{\theta_1^r\ (x_1+r)!} \prod_{j=2}^m \frac{\exp(-\alpha\theta_j)(\alpha\theta_j)^{x_j}}{x_j!} g(\alpha; \phi) d\alpha}{P(\boldsymbol{x}; \phi)} \\ &= \frac{(x_1+1)\ldots(x_1+r)}{\theta_1^r} \frac{P(x_1+r, x_2, \ldots, x_m; \phi)}{P(\boldsymbol{x}; \phi)}, \end{aligned}$$

The proof for $E(\alpha^{-r} \mid \boldsymbol{x})$ is similar. ∎

Note that when $x_j = 0, 1, \ldots, r-1$, the negative moments $E(\alpha^{-r} \mid \boldsymbol{x})$ must be calculated individually for each mixing distribution.

In order to derive the EM algorithm it is interesting to note that the complete data log-likelihood factors in to two parts. Consider as complete data $(X_{i1}, X_{i2}, \ldots, X_{im}, \alpha_i)$. i.e. the observed data together with the unobserved value of the mixing distribution. Then the complete data log-likelihood takes the form

$$\ell_C(\Theta) = \sum_{i=1}^n \sum_{j=1}^m [-\alpha_i\theta_{ij} + x_{ij}\log(\alpha_i\theta_{ij}) - \log(x_{ij}!)] + \sum_{i=1}^n \log(g(\alpha_i; \phi))$$

where Θ stands for all the parameters of the model. Clearly, the complete data log-likelihood can be split in two parts. The parameters ϕ of the mixing distribution appear only in the second part and thus the expectations needed to be calculated at the E-step are simply those related to the mixing distribution only. For estimating parameters ϕ it suffices to maximize the log-likelihood of the mixing distribution replacing the unobserved quantities of α_i with their expectations.

We now describe in detail the EM algorithm for the multivariate mixed Poisson regression models considered in section 3.

5.1. Multivariate negative binomial regression model

For the gamma mixing distribution, the posterior distribution of $\alpha \mid \boldsymbol{x}$ is a gamma distribution with parameters $\gamma + \sum_{j=1}^{m} x_j$ and $\gamma + \sum_{j=1}^{m} \theta_j$. For the E-step, we use Lemma 1 or the posterior mean to compute $E(\alpha \mid \boldsymbol{x}_i)$ and we use the posterior distribution to compute $E(\log \alpha_i \mid \boldsymbol{x}_i)$. The EM-algorithm, in this case, is as follows:

- *E-Step:* Calculate, for all $i = 1, 2, \dots, n$,

$$s_i = E(\alpha_i \mid \boldsymbol{x}_i) = \frac{\gamma + \sum_{j=1}^{m} x_{ij}}{\gamma + \sum_{j=1}^{m} \theta_{ij}},$$

$$t_i = E(\log(\alpha_i) \mid \boldsymbol{x}_i) = \Psi(\gamma + \sum_{j=1}^{m} x_{ij}) - \log(\gamma + \sum_{j=1}^{m} \theta_{ij}),$$

  where $\Psi(.)$ is the digamma function and $\theta_{ij} = \exp(\boldsymbol{\beta}_j^T \mathbf{z}_{ij})$, $i = 1, 2, \dots, n$, $j = 1, 2, \dots, m$, are obtained using the current values of $\boldsymbol{\beta}_j$.

- *M-step:*
  - Update the regression parameters $\boldsymbol{\beta}_j$, $j = 1, 2, \dots, p$, using the pseudo-values s_i as offset values and by fitting a simple Poisson regression model and
  - Update γ by

$$\gamma_{new} = \gamma_{old} - \frac{\Psi(\gamma_{old}) - \log(\gamma_{old}) - \bar{t} + \bar{s} - 1}{\Psi'(\gamma_{old}) - 1/\gamma_{old}}$$

    where $\Psi'(.)$ denotes the trigamma function, $\bar{s}$ and $\bar{t}$ are the sample means of $s_1, s_2, \dots s_n$ and $t_1, t_2, \dots t_n$, respectively. This is the one step ahead Newton iteration.

5.2. Multivariate Poisson-inverse Gaussian regression model

For the IG mixing distribution, the posterior distribution of $\alpha \mid \boldsymbol{x}$ is a GIG distribution, namely $\alpha \mid \boldsymbol{x} \sim GIG(\sum_{j=1}^{m} x_j - 1/2, \sqrt{2\sum_{j=1}^{m}\theta_j + \delta^2}, \delta)$. For the E-step, we use Lemma 1 to compute $E(\alpha \mid \boldsymbol{x}_i)$ and $E(\alpha^{-1} \mid \boldsymbol{x}_i)$.

The EM-algorithm, in this case, is as follows:

For the M-step of the algorithm, we need to derive the expectations of the form $E(\alpha \mid \boldsymbol{x}_i)$ and $E(\alpha^{-1} \mid \boldsymbol{x}_i)$ which can be easily derived since we know that the density is GIG. Also, simple expectations can be based on Lemma 1. Hence, the EM algorithm is as follows (using the results from Lemma 1).

- *E-Step:* Calculate, for all $i = 1, 2, \ldots, n$,

$$s_i = E(\alpha_i \mid \boldsymbol{x}_i) = \frac{x_{i1}+1}{\theta_{i1}} \frac{P_{IG}(x_{i1}+1, x_{i2}, \ldots, x_{im})}{P_{IG}(\boldsymbol{x}_i; \delta)},$$

$$t_i = E(\alpha_i^{-1} \mid \boldsymbol{x}_i) = \begin{cases} \frac{1+\delta\sqrt{\delta^2+2\sum_{j=1}^{m}\theta_{ij}}}{\delta^2}, & \text{if } \boldsymbol{x}_i = \mathbf{0}, \\ \frac{\theta_{ij}}{x_{ij}} \frac{P_{IG}(\boldsymbol{x}_i - \mathbf{1}_j; \delta)}{P_{IG}(\boldsymbol{x}_i; \delta)}, & \text{if } x_{ij} > 0, \end{cases}$$

  where $\theta_{ij} = \exp(\boldsymbol{\beta}_j^T \mathbf{z}_{ij})$, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$, are obtained using the current values of $\boldsymbol{\beta}_j$ and $P_{IG}(\boldsymbol{x}_i; \delta)$ is given in (3).

- *M-step:*
  - Update the regression parameters $\boldsymbol{\beta}_j$, $j = 1, 2, \ldots, m$, using the pseudo-values s_i as the offset values and by fitting a simple Poisson regression model and
  - Update δ by

$$\delta_{new} = \frac{1}{\sqrt{\bar{s} + \bar{t} - 2}}.$$

5.3. Multivariate Poisson-Lognormal regression model

For the multivariate Poisson log-normal model defined in (4) we can again derive an EM algorithm. However, in this case, since the expectations are not easy to obtain we may switch to a Monte Carlo EM approach.

The complete data log-likelihood takes the form

$$\begin{aligned} \ell_C(\Theta) = & \sum_{i=1}^{n}\sum_{j=1}^{m}\{-\alpha_i\theta_{ij} + x_{ij}\log(\alpha_i\theta_{ij}) - \log(x_{ij}!)\} + \\ & \sum_{i=1}^{n}\left\{-\frac{1}{2}\log(\pi) - \log(\nu) - \log(\alpha_i) - \frac{(\log(\alpha_i) + \nu^2/2)^2}{2\nu^2}\right\}. \end{aligned}$$

Thus, the expectations needed for the M-step are $E(\alpha_i)$ and $E((\log \alpha_i)^2)$.

Hence the algorithm can be written as

- *E-Step:* Calculate, for all $i = 1, 2, \ldots, n$,

  $$s_i = E(\alpha_i \mid \boldsymbol{x}_i) = \frac{(x_{i1}+1)}{\theta_{i1}} \frac{P_{LN}(x_{i1}+1, \ldots, x_{im}; \nu)}{P_{LN}(x_{i1}, \ldots, x_{im}; \nu)},$$

  $$t_i = E((\log \alpha_i)^2 \mid \boldsymbol{x}_i) = \frac{\int_0^\infty (\log \alpha_i)^2 \prod_{j=1}^m \frac{\exp(-\alpha_i\theta_{ij})(\alpha_i\theta_{ij})^{x_{ij}}}{x_{ij}!} \frac{\exp\left(-\frac{(\log(x_{ij})+\nu^2/2)^2}{2\nu^2}\right)}{\sqrt{2\pi}\nu\alpha_i} d\alpha_i}{P_{LN}(x_{i1}, \ldots, x_{im}; \nu)},$$

  Unfortunately, this expectation cannot be simplified. On the other hand, it can be evaluated numerically. Alternatively, a Monte Carlo approach is also possible using a rejection algorithm.

- *M-step:*
  - Update the regression parameters $\boldsymbol{\beta}_j$, $j = 1, 2, \ldots, m$, using the pseudo-values s_i as the offset values and by fitting a simple Poisson regression model and
  - Update ν by

    $$\hat{\nu}_{new} = 2(\sqrt{1+\bar{t}} - 1).$$

5.4. Computational issues

In this section, we point out some computational issues related to the implementation of the EM algorithm for the three considered multivariate mixed Poisson regression models.

(i) Good starting values for the parameters can be obtained by fitting simple univariate Poisson or negative binomial regressions. These will give starting values for the regression coefficients. We also emphasize that for the MNB regression model, since the M-step is in fact a Newton-Raphson iteration, one may obtain inadmissible values if the starting values are bad. Hence for this case the choice of initial values needs special attention. For example, as already mentioned, α relates to correlation and over-dispersion of the data. Hence, the starting value for the parameter of the mixing distribution may be chosen by equating the overdispersion of the model to the average of the observed overdispersion.

(ii) The expectations given in Lemma 1 are computationally suitable since the direct evaluation of the moments sometimes may be more time consuming. Note also that the denominator of each expectation in Lemma 1 is the probability needed for the log-likelihood evaluation and hence in most cases it is easily available.

(iii) The conditional expectation of $\alpha|\boldsymbol{x}$ itself can be of interest in some applications and is interpreted as a frailty. It represents the "importance" of each observation in the data. An advantage of the algorithm is that this quantity is easily available after completion of the EM algorithm.
(iv) The standard errors can be obtained using the standard approach of Louis (1982) for the standard errors for the EM algorithm. Alternatively one can obtain them numerically by the Hessian of the log-likelihood.

6 An Application

We analyze the *demand for health care in Australia* dataset, first used in Cameron and Trivedi (1998). There are 5190 observations. We have used three dependent variables, namely the total number of prescribed medications used in the past two days (PRESCRIB), the total number of nonprescribed medications used in the past two days (NONPRESC) and the number of consultations with non-doctor health professionals in the past two weeks (NONDOC). As covariates, we used the general health questionnaire score using Goldberg's method. The considered covariates are:
(i) hscore: (high score indicates bad health),
(ii) chcond (chronic condition: 1 if chronic condition(s) but not limited in activity, 0 other),
(iii) sex (1 if female, 0 if male),
(iv) age (in years divided by 100),
(v) income (annual income in Australian dollars divided by 1000: measured as mid-point of coded ranges).

We have fitted the MNB, MPIG and MPLN regression models using the EM algorithm described in section 5. Standard errors are obtained as usual from the EM output. As starting values for all models, we used the regression coefficients from separate Poisson regressions and for the overdispersion parameter α, the average of the observed overdispersions. We have checked with other initial values and we always converged to the same solution. For each model, we needed less than 100 iterations. The stopping criterion was to stop iterating when the relative change in the log-likelihood was smaller than 10^{-8}. All computing was made using the statistical computing environment language `R`. The MPLN regression model needed more time since we have to evaluate numerically the integrals. However, even this took less than 3 minutes in an Intel Duo 2Ghz processor. For the other two models, we needed less than 2 minutes computing time.

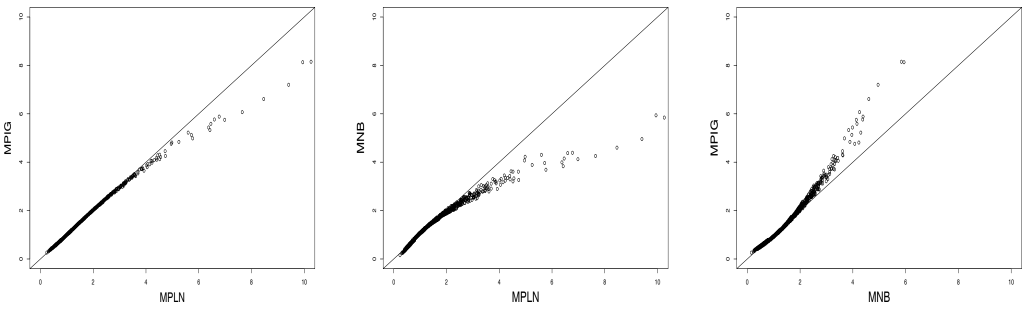
Table 1 shows the estimates and their standard errors of the parameters for MNB, MPIG and MPLN regression models, respectively. Table 1 also shows the log-likelihood of these models. Comparing the log-likelihoods, the MPLN regression model is the best choice and the MNB regression model is the worst.

		MNB		MPIG		MPLN	
		estimate	*s.e.*	*estimate*	*s.e.*	*estimate*	*s.e.*
PRESCRI	(Intercept)	-2.245	0.080	-2.275	0.081	-2.290	0.080
	sex	0.597	0.044	0.618	0.045	0.621	0.044
	age	2.840	0.107	2.847	0.109	2.859	0.107
	income	-0.058*	0.061	-0.057*	0.062	-0.054*	0.061
	hscore	0.121	0.008	0.122	0.008	0.122	0.008
	chcond	0.389	0.040	0.424	0.041	0.429	0.040
NONPRESCRI	(Intercept)	-1.328	0.088	-1.347	0.089	-1.361	0.089
	sex	0.252	0.054	0.267	0.055	0.271	0.055
	age	-0.586	0.139	-0.584	0.141	-0.575	0.141
	income	0.265	0.072	0.266	0.073	0.269	0.073
	hscore	0.083	0.011	0.083	0.011	0.083	0.011
	chcond	0.268	0.054	0.295	0.054	0.302	0.054
NONDOCO	(Intercept)	-2.984	0.130	-3.012	0.130	-3.026	0.131
	sex	0.442	0.073	0.462	0.073	0.466	0.073
	age	2.169	0.176	2.170	0.177	2.183	0.178
	income	-0.183*	0.104	-0.180*	0.105	-0.176*	0.105
	hscore	0.169	0.011	0.170	0.011	0.170	0.012
	chcond	-0.043*	0.067	-0.010*	0.068	-0.006*	0.068
		$\hat{\gamma}$=1.780	0.090	$\hat{\delta}$=1.227	0.037	$\hat{\nu}$=0.727	0.089
	loglik	-13011.32		-12981.97		-12977.22	

Table 1: Results from fitting the three regression models to the demand for health care dataset. An * indicates that the test for the regression coefficient is not significant.

The significance of the covariates was assessed by a Wald type statistic.

Recall that the posterior expectation of α for each observation is available as a byproduct of the EM algorithm. This quantity can be seen as a frailty implying the tendency of the patient to need medical treatment. We have plotted these quantities for any two regression models in Figure 1. Agreement between the two models is implied by the diagonal line. Figure 1 shows agreement between the MPIG and MPLN regression models and disagreement between these models and the MNB regression model.



(a) (b) (c)

Figure 1: Plot of the $E(\alpha_i \mid x_{i1}, x_{i2}, x_{i3})$ for all the observations. The diagonal line implies agreement between the two models.

7 Concluding Remarks

In this paper, we have described an EM algorithm for multivariate mixed Poisson regression models, *viz.* MNB, MPIG and MPLN regression models. The algorithm is used to estimate easily the regression coefficients and the parameter of the mixing distribution. The proposed EM algorithm is applied to a dataset on the demand for health care in Australia. It was noticed that the MPLN regression model provides a better fit than the other two regression models.

It is important that the proposed algorithm facilitates the fitting of various multivariate mixed Poisson regression models, thus providing alternative models to the multivariate negative binomial one. In this paper we used MPIG and MPLN regression models which are new in the literature.

Also note that the algorithm is applicable for certain other choices of mixing distribution, while it is easily programmable and avoids overflow problems which may occur via other numerical maximization schemes. Furthermore, the computation of the joint probability mass function, which is typically an integral, is avoided and this reduces computational problems.

References

[1] J. Aitchison and C.H. Ho, The multivariate oisson-log normal distribution, *Biometrika*, **76** (1989), 643–653.

[2] C.A. Cameron and P.K. Trivedi, *Regression Analysis of Count Data*, Cambridge University Press, New York, 1989.

[3] C. Dean and J.F. Lawless, A mixed Poisson-inverse Gaussian regression model, *Canadian Journal of Statistics*, **17** (1989), 171-182.

[4] S. Gurmu and J. Elder, Generalized bivariate Count data regression models, *Economics Letters*, **68** (2000), 31-36.

[5] B. Jorgensen, *Statistical Properties of the Generalized Inverse Gaussian Distribution,* Lecture Notes in Statistics, Volume 9, Springer-Verlag, New York, 1982.

[6] D. Karlis, A general EM approach for maximum likelihood estimation in mixed Poisson regression models, *Statistical Modelling*, **1** (2001), 305-318.

[7] D. Karlis, An EM algorithm for mixed Poisson distributions, *ASTIN Bulletin*, **35** (2005), 3-24.

[8] S. Kocherlakota, On the compounded bivariate Poisson distribution: A unified treatment, *Annals of the Institute of Statistical Mathematics*, **40** (1988), 61-76.

[9] T.A. Louis, Finding the observed information matrix when using the EM algorithm, *Journal of the Royal Statistical Society B*, **44** (1982), 226-233.

[10] M.K. Munkin and P.K. Trivedi, Simulated maximum likelihood estimation of multivariate mixed-Poisson regression models, with application, *Econometrics Journal*, **2** (1999), 29-48.

[11] G. Stein and J.M. Yuritz, Bivariate compound Poisson distributions, *Communications in Statistics-Theory and Methods*, **16** (1987), 3591-3607.

[12] G.Z. Stein, W. Zucchini and J.M. Juritz, Parameter estimation for the Sichel distribution and its multivariate extension, *Journal of the American Statistical Association*, **82** (1987), 938-944.

Received: March, 2012