Hiroshima Math. J.
**35** (2005), 327–349

# A class of multivariate discrete distributions based on an approximate density in GLMM

Tetsuji Tonda

(Received December 14, 2004)
(Revised January 21, 2005)

Abstract. It is well known that the generalized linear mixed model is useful for analyzing the overdispersion and correlation structure for multivariate discrete data. In this paper, we derive an approximation of the density function for the generalized linear mixed model. This approximation is found to satisfy the properties of probability density function under some conditions. Therefore, this approximation can be regarded as a class of multivariate distributions. Estimation of the parameters in this class can be carried out by the maximum likelihood method. We give the likelihood ratio criteria for testing several covariance structures. Several simulation studies were also conducted for the Poisson log-normal model when the proposed density function is regarded as an approximate likelihood of the generalized linear mixed model.

## 1. Introduction

This paper is concerned with the multivariate discrete distributions for count or binary observations. A review of multivariate discrete distributions is given in Joe (1997). An approach for generalizing the univariate discrete distribution to multivariate distribution has been attempted by formulating a mixture model. For multivariate count data, Steyn (1976) proposed the multivariate Poisson normal model by adding a random effect to the mean parameter of the Poisson distribution, but the resultant distribution ignores the fact that the mean parameter is positive. Aitchison and Ho (1989) modified this fault, and proposed the multivariate Poisson log-normal model. For multivariate binary data, Coull and Agresti (2000) have proposed the multivariate Binomial logit-normal model. The generalized linear mixed model is an extension of the generalized linear model (McCullagh and Nelder, 1989), which is generated by adding a random effect to the linear predictor. This model is useful for analyzing the overdispersion and correlations, and includes the multi-

2000 *Mathematics Subject Classification*. primary 62E20, secondly 62H15.

*Key words and phrases*. Conditional independent; Correlation structure; Density approximation; Generalized linear mixed model; Multivariate discrete distribution; Overdispersion; Random effect.

variate Poisson log-normal model and the multivariate Binomial logit-normal model. Although these multivariate models are flexible, their calculations are problematic since they all involve multidimensional integrals and their probability density functions have no closed forms. Therefore, some numerical methods are needed in practice. It is difficult to apply these models to real data analysis when the dimension of response is large, because such complicated calculations are needed that the data analysis takes too much time. In order to avoid calculating a multiple integral, many authors have proposed approximations of the likelihood functions. Breslow and Clayton (1993) have proposed the penalized quasi-likelihood (PQL) method and the marginal quasi-likelihood (MQL) method. The PQL approach produces biased estimates for the regression effects and the variance component of random effects. Breslow and Lin (1995) provided a bias correction for the single variance component model. The MQL estimating equation approach requires the first and second order marginal moments of the responses, but those are not available under the generalized linear mixed model. Breslow and Clayton (1993) have used an approximation of the mean vector and a "working covariance matrix", as in Zeger, Liang and Albert (1988), to construct the estimating equations for the regression parameters. For estimating the variance component, they used a pseudo-likelihood method proposed by Carroll and Ruppert (1982). Sutradhar and Rao (2001) proposed the MQL estimating equation approach for both regression and variance parameters by deriving the marginal moments up to the fourth order. The purpose of this paper is to construct a class of multivariate discrete distributions with an analytical probability density function for a full likelihood analysis. In this sense, our approach differs from the methods mentioned above.

This paper is structured in the following way. In section 2, we present several models for multivariate discrete data. In section 3, we derive an approximation for the density, and examine its basic properties. Based on the approximate density, we propose a new class of multivariate discrete distributions. Sections 4 and 5 present some methods for estimating and testing the parameters. In section 6, we present some simulation studies.

## 2. Multivariate discrete models

This section presents several multivariate discrete models. There are several approaches which extend the univariate discrete model to the multivariate version. In this section, we focus on an approach based on the mixture model. The multivariate models reviewed in this section are connected with the class of multivariate distributions which appears in section 3.

## 2.1. Models for multivariate count data

### 2.1.1. Univariate models

The Poisson model and the Poisson-Gamma model are well-known models for univariate count data. The Poisson-Gamma model is obtained as a Gamma mixture of Poisson distribution. In fact, assume that $Y$ given $Z = z$ has a Poisson distribution with mean $\lambda z$,

$$\mathrm{P}(Y = y|z) = \frac{(\lambda z)^y}{y!} e^{-\lambda z},$$

and $z$ is a random effect parameter having a gamma distribution with mean 1 and variance $\sigma^2$, whose probability density function is given by

$$g(z;\sigma^2) = \frac{(\sigma^2)^{-1/\sigma^2}}{\Gamma(1/\sigma^2)} z^{1/\sigma^2 - 1} e^{-z/\sigma^2}.$$

Then for $y = 0, 1, 2, \ldots,$

$$\begin{aligned} \mathrm{P}(Y = y) &= \int_0^\infty \mathrm{P}(Y = y|z) g(z;\sigma^2) dz \\ &= \frac{\Gamma(y + 1/\sigma^2)}{y!\Gamma(1/\sigma^2)} \left(\frac{\lambda}{\lambda + 1/\sigma^2}\right)^y \left(\frac{1/\sigma^2}{\lambda + 1/\sigma^2}\right)^{1/\sigma^2}. \end{aligned}$$

The resultant distribution is a negative binomial distribution.

In modeling for count data, it is important that the resultant model has an overdispersion, which means that the variance is larger than the mean (in the Poisson model, its variance is equal to the mean). The Poisson-Gamma model has an overdispersion, since $\mathrm{E}(Y) = \lambda$ and $\mathrm{Var}(Y) = \lambda + \lambda^2\sigma^2$. More generally, assume that $Z$ is a random variable with mean $\mu_Z$ and variance $\sigma_Z^2$. Then, we have

$$\begin{aligned} \mathrm{E}(Y) &= \mathrm{E}\{\mathrm{E}(Y|Z)\} = \mathrm{E}(\lambda Z) = \lambda\mu_Z, \\ \mathrm{Var}(Y) &= \mathrm{E}\{\mathrm{Var}(Y|Z)\} + \mathrm{Var}\{\mathrm{E}(Y|Z)\} \\ &= \mathrm{E}(\lambda Z) + \mathrm{Var}(\lambda Z) = \lambda\mu_Z + \lambda^2\sigma_Z^2. \end{aligned} \tag{2.1}$$

Hence $\mathrm{Var}(Y) \geq \mathrm{E}(Y)$ such that it has equality only when $Z$ has a degenerate distribution. Therefore, general mixtures of the Poisson distribution have an overdispersion relative to the Poisson distribution.

### 2.1.2. Multivariate Poisson model

McKendrick (1916, 1926) and Wicksell (1916) have proposed a natural bivariate Poisson model which is expressed by $(Y_1, Y_2) = (Z_1 + Z_{12}, Z_2 + Z_{12})$, where $Z_1$, $Z_2$, $Z_{12}$ are independent Poisson variables with the mean parameters $\theta_1$, $\theta_2$, $\theta_{12}$, respectively. Campbell (1934) showed that the joint probability density function is given by

$$P(Y_1 = y_1, Y_2 = y_2) = e^{-(\theta_1+\theta_2+\theta_{12})} \frac{\theta_1^{y_1}}{y_1!} \frac{\theta_2^{y_2}}{y_2!} \sum_{i=0}^{\min(y_1, y_2)} i!_{y_1}C_{i\,y_2}C_i \left(\frac{\theta_{12}}{\theta_1\theta_2}\right)^i,$$

and the marginal distributions are Poisson$(\theta_1 + \theta_{12})$ and Poisson$(\theta_2 + \theta_{12})$, respectively. Then, the first two moments are given by

$$E(Y_j) = \mathrm{Var}(Y_j) = \theta_j + \theta_{12}, \qquad \mathrm{Cov}(Y_1, Y_2) = \theta_{12},$$

$$\mathrm{Corr}(Y_1, Y_2) = \frac{\theta_{12}}{\sqrt{\theta_1 + \theta_{12}}\sqrt{\theta_2 + \theta_{12}}} \quad (\geq 0).$$

Unfortunately, this multivariate Poisson model dose not support the negative correlation between two count variables (Holgate, 1964) and the overdispersion on the marginal distribution. Furthermore, note that to construct a $p$-variate version, $2^p - 1$ independent Poisson variables are needed.

### 2.1.3. Multivariate Poisson Log-Normal model

We consider another method of generalizing the univariate Poisson model to the multivariate version, based on a mixture model. Steyn (1976) proposed the multivariate Poisson normal model by assuming the multivariate normal distribution on the Poisson mean parameters. Although this allows for rich correlation structures, this model ignores the fact that the mean parameters are positive. Aitchison and Ho (1989) proposed the multivariate Poisson log-normal model, which takes into account the fact that the mean parameters are positive.

For individual $i$ $(i = 1, \dots, N)$ and given $\boldsymbol{\lambda}_i^* = (\lambda_{i1}^*, \dots, \lambda_{ip}^*)'$, we assume that $\boldsymbol{y}_i = (Y_{i1}, \dots, Y_{ip})'$ is a vector of $p$ independent Poisson random variables with the mean parameters $\boldsymbol{\lambda}_i^*$, and the conditional probability density function is written by

$$f_P(\boldsymbol{y}_i | \boldsymbol{\lambda}_i^*) = \prod_{j=1}^{p} \frac{(\lambda_{ij}^*)^{y_{ij}}}{y_{ij}!} e^{-\lambda_{ij}^*}. \tag{2.2}$$

Let $\log \boldsymbol{\lambda}_i^*$ be the componentwise log transformation, i.e.,

$$\log \boldsymbol{\lambda}_i^* = (\log \lambda_{i1}^*, \dots, \log \lambda_{ip}^*)'. \tag{2.3}$$

Then, the multivariate Poisson log-normal model is constructed by incorporating a random effect into the mean parameter, such that

$$\log \boldsymbol{\lambda}_i^* = \log \boldsymbol{\lambda} + \boldsymbol{z}_i, \tag{2.4}$$

where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_p)'$ denotes a $p \times 1$ vector of fixed effects, and $\boldsymbol{z}_i = (Z_{i1}, \ldots, Z_{ip})'$ denotes a $p \times 1$ vector of random effects. Assume that $\boldsymbol{z}_i$'s are independently and identically distributed as the multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\Sigma$. In this model, the probability density function of $\boldsymbol{y}_i$ is expressed by

$$f(\boldsymbol{y}_i; \boldsymbol{\lambda}, \Sigma) = \int_{\mathscr{R}^p} f_{\mathrm{P}}(\boldsymbol{y}_i | \boldsymbol{\lambda}_i^*) g(\boldsymbol{z}_i; \mathbf{0}, \Sigma) d\boldsymbol{z}_i, \tag{2.5}$$

where $g(\boldsymbol{z}_i; \mathbf{0}, \Sigma)$ denotes the multivariate normal density function with mean vector $\mathbf{0}$ and covariance matrix $\Sigma$.

Although there is no closed-form expression of this multiple integral, its moment can be easily obtained through the conditional expectation result and the standard properties of the Poisson and log-normal distributions as in (2.1). Let $\sigma_{ij}$ be the $(i, j)$th element of $\Sigma$, and then the first two moments can be written by

$$\mathrm{E}(Y_j) = \lambda_j e^{1/2\sigma_{jj}} \equiv \alpha_j, \tag{2.6}$$

$$\mathrm{Var}(Y_j) = \alpha_j + \alpha_j^2 (e^{\sigma_{ii}} - 1), \qquad \mathrm{Cov}(Y_i, Y_j) = \alpha_i \alpha_j (e^{\sigma_{ij}} - 1), \tag{2.7}$$

$$\mathrm{Corr}(Y_i, Y_j) = \frac{e^{\sigma_{ij}} - 1}{\sqrt{(e^{\sigma_{ii}} - 1 + \alpha_i^{-1})(e^{\sigma_{jj}} - 1 + \alpha_j^{-1})}}. \tag{2.8}$$

From (2.6), (2.7) and (2.8), we can see that the marginal distribution has an overdispersion relative to the Poisson distribution through the parameter $\sigma_{ii}$, and the sign of correlation between the count $Y_i$ and $Y_j$ depends on that of $\sigma_{ij}$. Since $|\mathrm{Corr}(Y_i, Y_j)| < |\mathrm{Corr}(e^{z_i}, e^{z_j})|$, the range of possible correlation values is not as wide as that of the log-normal distribution, but can be close to this range if $\alpha_i$ and $\alpha_j$ are large.

We consider an alternative parameterization by shifting a mean of the random effect $\boldsymbol{z}_i$ as follows,

$$f(\boldsymbol{y}_i; \boldsymbol{\lambda}, \Sigma) = \int_{\mathscr{R}^p} f_{\mathrm{P}}(\boldsymbol{y}_i | \boldsymbol{\lambda}_i^*) g\left(\boldsymbol{z}_i; -\frac{1}{2}\boldsymbol{\sigma}, \Sigma\right) d\boldsymbol{z}_i, \tag{2.9}$$

where $\boldsymbol{\sigma} = (\sigma_{11}, \ldots, \sigma_{pp})'$. Based on (2.9), the first two moments are rewritten by

$$\mathrm{E}(Y_j) = \lambda_j, \tag{2.10}$$

$$\mathrm{Var}(Y_j) = \lambda_j + \lambda_j^2 (e^{\sigma_{ii}} - 1), \qquad \mathrm{Cov}(Y_i, Y_j) = \lambda_i \lambda_j (e^{\sigma_{ij}} - 1), \tag{2.11}$$

$$\mathrm{Corr}(Y_i, Y_j) = \frac{e^{\sigma_{ij}} - 1}{\sqrt{(e^{\sigma_{ii}} - 1 + \lambda_i^{-1})(e^{\sigma_{jj}} - 1 + \lambda_j^{-1})}}. \tag{2.12}$$

### 2.2. Models for multivariate binary data

#### 2.2.1. Univariate model

Assume that $Y$ given $\Pi = \pi$ has a binomial distribution with $m$ trials and the success parameter $\pi$, i.e.,

$$\mathrm{P}(Y = y|\pi) = {}_mC_y \pi^y (1-\pi)^{m-y},$$

where ${}_mC_y = m(m-1)\dots(m-y)/y!$, and $\pi$ is a random variable which follows a beta distribution with two parameters $a, b > 0$, whose probability density function is given by

$$g(\pi; a, b) = \frac{1}{B(a,b)} \pi^{a-1}(1-\pi)^{b-1},$$

where $0 < \pi < 1$ and $B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$. Then, the probability density function of $Y$ is expressed by

$$\begin{aligned} \mathrm{P}(Y = y) &= \int_0^1 \mathrm{P}(Y = y|\pi) g(\pi; a, b) d\pi \\ &= {}_mC_y \frac{B(y+a, m-y+b)}{B(a,b)}. \end{aligned}$$

#### 2.2.2. Multivariate Binomial Logit-Normal model

For individual $i$ $(i = 1, \dots, N)$ and given $\boldsymbol{\pi}_i^* = (\pi_{i1}^*, \dots, \pi_{ip}^*)'$, we assume that $\boldsymbol{y}_i = (Y_{i1}, \dots, Y_{ip})'$ is distributed as a $p$ independent binomial distribution with an index vector $\boldsymbol{m}_i = (m_{i1}, \dots, m_{ip})'$ and success parameter vector $\boldsymbol{\pi}_i^*$, whose conditional probability density function can be written by

$$f_{\mathrm{B}}(\boldsymbol{y}_i|\boldsymbol{\pi}_i^*; \boldsymbol{m}_i) = \prod_{j=1}^{p} {}_{m_{ij}}C_{y_{ij}} (\pi_{ij}^*)^{y_{ij}} (1-\pi_{ij}^*)^{m_{ij}-y_{ij}},$$

Let $\operatorname{logit} \boldsymbol{\pi}_i^*$ be the componentwise logit function, i.e.,

$$\operatorname{logit} \boldsymbol{\pi}_i^* = \left( \log \frac{\pi_{i1}^*}{1-\pi_{i1}^*}, \dots, \log \frac{\pi_{ip}^*}{1-\pi_{ip}^*} \right)'. \tag{2.13}$$

Then the multivariate Binomial logit-normal model is expressed by incorporating a random effect, such that

$$\operatorname{logit} \boldsymbol{\pi}_i^* = \operatorname{logit} \boldsymbol{\pi} + \boldsymbol{z}_i, \tag{2.14}$$

where $\boldsymbol{\pi} = (\pi_1, \dots, \pi_p)'$ denotes a $p \times 1$ vector of fixed effects and $\boldsymbol{z}_i$ is a $p \times 1$ vector of random effects. Assume that $\boldsymbol{z}_i$'s are independently and identically distributed as the multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\Sigma$. Then the probability density function of $\boldsymbol{y}_i$ is written by

$$f(\boldsymbol{y}_i;\boldsymbol{m}_i,\boldsymbol{\pi},\Sigma)=\int_{\mathscr{R}^p} f_{\mathrm{B}}(\boldsymbol{y}_i|\boldsymbol{\pi}_i^*;\boldsymbol{m}_i)g(\boldsymbol{z}_i;\mathbf{0},\Sigma)d\boldsymbol{z}_i, \tag{2.15}$$

where $g(\boldsymbol{z}_i;\mathbf{0},\Sigma)$ denotes the multivariate normal density function with mean vector $\mathbf{0}$ and covariance matrix $\Sigma$. There are no closed-form expressions for its moments. Coull and Agresti (2000) showed asymptotic forms for small $\sigma_{ij}$ which are expressed by

$$\begin{aligned}
\mathrm{E}(Y_j) &\cong m_j\pi_j,\\
\mathrm{Var}(Y_j) &\cong m_j\pi_j(1-\pi_j)+(m_j^2-m_j)\{\pi_j(1-\pi_j)\}^2\sigma_{jj},\\
\mathrm{Cov}(Y_i,Y_j) &\cong m_im_j\pi_i(1-\pi_i)\pi_j(1-\pi_j)\sigma_{ij}, \qquad i\neq j,
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{Corr}(Y_i,Y_j) \cong{}& m_im_j\pi_i(1-\pi_i)\pi_j(1-\pi_j)\sigma_{ij}\\
&\div\{(m_i^2-m_i)\sigma_{ii}\{\pi_i(1-\pi_i)\}^2+m_i\pi_i(1-\pi_i)\\
&\times(m_j^2-m_j)\sigma_{jj}\{\pi_j(1-\pi_j)\}^2+m_j\pi_j(1-\pi_j)\}^{1/2}.
\end{aligned}$$

Coull and Agresti (2000) have conducted a simulation study of various properties of the multivariate Binomial Logit-Normal distribution, and suggested that these approximations tended to break down when $\sigma_{ii}>0.6$.

### 2.3. Generalized Linear Mixed Model

The GLMM (Generalized Linear Mixed Model) is an extension of the GLM (Generalized Linear Model) with the normal random effects included in the linear predictor (see, for example, Gueorguieva, 2001). The GLMM is constructed by two steps: (i) Conditional independent, (ii) Random effects.

**(i) Conditional independent:** Let $\boldsymbol{y}_i=(Y_{i1},\dots,Y_{ip})'$ denote a $p\times 1$ observation vector and $\boldsymbol{\theta}_i=(\theta_{i1},\dots,\theta_{ip})'$ denote a $p\times 1$ unknown parameter vector on individual $i$ $(i=1,\dots,N)$. Given $\boldsymbol{\theta}_i$, the conditional probability density function of $\boldsymbol{y}_i$ is expressed by a product of $p$ independent exponential family, i.e.,

$$f_{\exp}(\boldsymbol{y}_i|\boldsymbol{\theta}_i)=\prod_{j=1}^{p}\exp\left\{\frac{y_{ij}\theta_{ij}-b_j(\theta_{ij})}{\phi_j}+c_j(y_{ij},\phi_j)\right\},$$

Table 1. Characterictics of some common univariate distributions in the exponential family.

| | Normal | Poisson | Binomial | Gamma |
|---|---|---|---|---|
| Notation | $\mathrm{N}(\mu, \sigma^2)$ | $\mathrm{Poisson}(\mu)$ | $\mathrm{B}(m, \pi)/m$ | $\mathrm{Gamma}(\mu, v)$ |
| Range of $y$ | $(-\infty, \infty)$ | $0, 1, 2, \dots$ | $0, 1/m, 2/m, \dots, m/m$ | $(0, \infty)$ |
| $\phi$ | $\sigma^2$ | $1$ | $1/m$ | $v^{-1}$ |
| $b(\theta)$ | $\theta^2/2$ | $e^\theta$ | $\log(1+e^\theta)$ | $-\log(-\theta)$ |
| $c(y, \phi)$ | $-\dfrac{y^2 + \phi\log(2\pi\phi)}{2}$ | $-\log y!$ | $\log {}_nC_{ny}$ | $v\log(vy) - \log y - \log \Gamma(v)$ |
| $\mu = \mathrm{E}(Y)$ | $\theta$ | $e^\theta$ | $1/(1+e^{-\theta})$ | $-1/\theta$ |
| Link function: $h(\mu)$ | identity | log | logit | reciprocal |
| Variance function: $v(\mu)$ | $1$ | $\mu$ | $\mu(1-\mu)$ | $\mu^2$ |

* This table is from McCullagh and Nelder (1989).

where $\phi_j$ is a dispersion parameter and $b_j(\cdot)$, $c_j(\cdot)$ are known functions (see, e.g., Table 1). From the properties of the exponential family, the mean vector and covariance matrix are expressed by

$$\mathrm{E}(\boldsymbol{y}_i|\boldsymbol{\theta}_i) = \boldsymbol{\mu}_i = b'(\boldsymbol{\theta}_i), \qquad \mathrm{Cov}(\boldsymbol{y}_i|\boldsymbol{\theta}_i) = \Phi V(\boldsymbol{\mu}_i), \tag{2.16}$$

where $\boldsymbol{\mu}_i = (\mu_{i1}, \dots, \mu_{ip})'$,

$$b'(\boldsymbol{\theta}_i) = \frac{\partial}{\partial \boldsymbol{\theta}_i} b(\boldsymbol{\theta}_i), \qquad b(\boldsymbol{\theta}_i) = (b_1(\theta_{i1}), \dots, b_p(\theta_{ip}))',$$

$V(\boldsymbol{\mu}_i) = \mathrm{diag}(v_1(\mu_{i1}), \dots, v_p(\mu_{ip}))$, $\Phi = \mathrm{diag}(\phi_1, \dots, \phi_p)$ and $v_j(\cdot)$ is a variance function defined by $v_j(\mu_{ij}) = b_j''(\theta_{ij})$ (see, e.g., Table 1).

**(ii) Random effects:** Then the GLMM is expressed by incorporating a random effect $\boldsymbol{z}_i = (Z_{i1}, \dots, Z_{ip})'$ into $\boldsymbol{\theta}_i$, such that

$$\boldsymbol{\theta}_i = h(\boldsymbol{\mu}_i) = \boldsymbol{\eta} + \boldsymbol{z}_i, \qquad \boldsymbol{z}_i \sim \text{i.i.d. } \mathrm{N}_p(\mathbf{0}, \Sigma). \tag{2.17}$$

Here $h(\boldsymbol{\mu}_i) = (h_1(\mu_{i1}), \dots, h_p(\mu_{ip}))'$ and $h_j(\cdot)$ is a link function (see, e.g., Table 1).

Under the model (2.17), the probability density function of $\boldsymbol{y}_i$ is written by

$$f(\boldsymbol{y}_i; \boldsymbol{\eta}, \Sigma) = \int_{\mathscr{R}^p} f_{\exp}(\boldsymbol{y}_i|\boldsymbol{\theta}_i) g(\boldsymbol{z}_i; \mathbf{0}, \Sigma) d\boldsymbol{z}_i, \tag{2.18}$$

where $g(\boldsymbol{z}_i; \mathbf{0}, \Sigma)$ denotes the multivariate normal density function with mean vector $\mathbf{0}$ and covariance matrix $\Sigma$.

## 3. A class of multivariate distribution

In this section, we derive an asymptotic approximation expression for the density function of GLMM, and show that our approximation satisfies the properties of the probability density function. In order to avoid calculating the multidimensional integral (2.18), many authors have proposed approximations of the likelihood function. It is known that the approximate inference based on such quasi-likelihood has biases (see, for example, Breslow and Lin, 1995).

We derive an approximation of the density function (2.18) based on Taylor expansion about the mean of the random effect, $\boldsymbol{z}_i = \mathbf{0}$ (or $\boldsymbol{\theta} = \boldsymbol{\eta}$), up to the 2nd order. In the following, we drop the suffix $i$ for simplicity of notation.

$$\begin{aligned} f(\boldsymbol{y}; \boldsymbol{\eta}, \Sigma) &= \mathrm{E}_{\boldsymbol{z}}\{f_{\exp}(\boldsymbol{y}|\boldsymbol{\eta} + \boldsymbol{z})\} \\ &\cong \mathrm{E}_{\boldsymbol{z}}\left\{ f_{\exp}(\boldsymbol{y}; \boldsymbol{\eta}) + \left[\frac{\partial f_{\exp}}{\partial \boldsymbol{z}}\right]_{\boldsymbol{z}=\mathbf{0}}' \boldsymbol{z} + \frac{1}{2}\boldsymbol{z}' \left[\frac{\partial^2 f_{\exp}}{\partial \boldsymbol{z}\partial \boldsymbol{z}'}\right]_{\boldsymbol{z}=\mathbf{0}}' \boldsymbol{z} \right\} \\ &= f_{\exp}(\boldsymbol{y}; \boldsymbol{\eta}) + \frac{1}{2}\ \mathrm{tr}\left( \left[\frac{\partial^2 f_{\exp}}{\partial \boldsymbol{z}\partial \boldsymbol{z}'}\right]_{\boldsymbol{z}=\mathbf{0}} \Sigma \right) \\ &= f_{\exp}(\boldsymbol{y}; \boldsymbol{\eta})\left\{ 1 + \frac{1}{2}\ \mathrm{tr}(W\Sigma) \right\}, \end{aligned}$$

where $W = (w_{ab})$ is a $p \times p$ matrix whose elements are given by

$$\begin{aligned} w_{aa} &\equiv \frac{\partial^2}{\partial z_a^2} f_{\exp}(\boldsymbol{y}|\boldsymbol{\eta} + \boldsymbol{z})\bigg|_{\boldsymbol{z}=\mathbf{0}} \\ &= \left(\frac{y_a - b_a'(\eta_a)}{\phi_a}\right)^2 - \frac{b_a''(\eta_a)}{\phi_a} = \left(\frac{y_a - \mu_a}{\phi_a}\right)^2 - \frac{v_a(\mu_a)}{\phi_a}, \end{aligned} \tag{3.1}$$

$$\begin{aligned} w_{ab} &\equiv \frac{\partial^2}{\partial z_a \partial z_b} f_{\exp}(\boldsymbol{y}|\boldsymbol{\eta} + \boldsymbol{z})\bigg|_{\boldsymbol{z}=\mathbf{0}} \\ &= \left(\frac{y_a - b_a'(\eta_a)}{\phi_a}\right)\left(\frac{y_b - b_b'(\eta_b)}{\phi_b}\right) = \left(\frac{y_a - \mu_a}{\phi_a}\right)\left(\frac{y_b - \mu_b}{\phi_b}\right). \end{aligned} \tag{3.2}$$

Let

$$\tilde{f}(\boldsymbol{y}; \boldsymbol{\eta}, \Sigma) = f_{\exp}(\boldsymbol{y}; \boldsymbol{\eta})\left\{ 1 + \frac{1}{2}\ \mathrm{tr}(W\Sigma) \right\}, \tag{3.3}$$

where

$$f_{\exp}(\boldsymbol{y};\boldsymbol{\eta}) = \prod_{j=1}^{p} \exp\left\{\frac{y_j\eta_j - b_j(\eta_j)}{\phi_j} + c(y_j,\phi)\right\},$$

and $W$ is given by (3.1) and (3.2). Then we have the following theorem.

Theorem 3.1. *If $\Sigma$ is a positive semi-definite matrix and*

$$\sum_{j=1}^{p} \frac{1}{\phi_j} v_j(\mu_j)\sigma_{jj} \le 2, \tag{3.4}$$

*then $\tilde{f}(\boldsymbol{y};\boldsymbol{\eta},\Sigma)$ is a proper probability density function.*

Proof. Let $\mathrm{E}_{\exp}(\cdot)$ denote the expectation under the exponential family density $f_{\exp}(\boldsymbol{y};\boldsymbol{\eta})$, it follows $\mathrm{E}_{\exp}(w_{ij}) = 0$ from (3.1) and (3.2). Then, we have

$$\int_{\mathscr{R}^p} \tilde{f}(\boldsymbol{y};\boldsymbol{\eta},\Sigma)d\boldsymbol{y} = \mathrm{E}_{\exp}\left\{1 + \frac{1}{2}\ \mathrm{tr}(W\Sigma)\right\} = 1 + \sum_{a=1}^{p}\sum_{b=1}^{p} \mathrm{E}(w_{ab})\sigma_{ab} = 1.$$

Moreover, note that let $\boldsymbol{x} = (x_1,\dots,x_p)'$ and $x_j = (y_j - \mu_j)/\phi_j$ $(j = 1,\dots,p)$, then $W = \boldsymbol{x}\boldsymbol{x}' - \mathrm{diag}(v_1(\mu_1)/\phi_1,\dots,v_p(\mu_p)/\phi_p)$. Therefore, for any $\boldsymbol{y} \in \mathscr{R}^p$,

$$\begin{aligned}
\tilde{f}(\boldsymbol{y};\boldsymbol{\eta},\Sigma) &= f_{\exp}(\boldsymbol{y};\boldsymbol{\eta})\left\{1 + \frac{1}{2}\ \mathrm{tr}(W\Sigma)\right\} \\
&= f_{\exp}(\boldsymbol{y};\boldsymbol{\eta})\left\{1 + \frac{1}{2}\left(\boldsymbol{x}'\Sigma\boldsymbol{x} - \sum_{j=1}^{p}\frac{v_j(\mu_j)}{\phi_j}\sigma_{jj}\right)\right\} \\
&\ge f_{\exp}(\boldsymbol{y};\boldsymbol{\eta})\left\{1 - \frac{1}{2}\sum_{j=1}^{p}\frac{v_j(\mu_j)}{\phi_j}\sigma_{jj}\right\}
\end{aligned}$$

This yield

$$\tilde{f}(\boldsymbol{y};\boldsymbol{\eta},\Sigma) \ge 0 \Leftrightarrow \sum_{j=1}^{p}\frac{1}{\phi_j}v_j(\mu_j)\sigma_{jj} \le 2.$$

From these results, $\tilde{f}(\boldsymbol{y};\boldsymbol{\eta},\Sigma)$ is the proper probability density function under the condition (3.4). □

From this theorem, we note that under the condition (3.4), $\tilde{f}(\boldsymbol{y};\boldsymbol{\eta},\Sigma)$ can be regarded as a class of multivariate density functions rather than an approximate density function of (2.18). We denote this class of distributions by $\mathrm{ME}_p(\boldsymbol{\eta},\Sigma)$.

Next, we consider the first two moments of $\mathrm{ME}_p(\boldsymbol{\eta},\Sigma)$. Let $m_j^{(r)} = \mathrm{E}_{\exp}\{(Y_j - \mu_j)^r\}$ for $r = 1,\dots,4$. Then these moments are written by

$$m_j^{(2)} = \phi_j b_j' = \phi_j v_j(\mu_j), \qquad m_j^{(3)} = \phi_j^2 b_j^{(3)}, \qquad m_j^{(4)} = \phi_j^3 b_j^{(4)} + 3\{m_j^{(2)}\}^2, \quad (3.5)$$

where $b_j^{(r)}$ denotes the $r$th order derivative of $b_j(\theta_j)$ with respect to $\theta_j$. Using these expressions, we obtaine the following lemma.

LEMMA 3.1.

$$\mathrm{E}_{\exp}\{\mathrm{tr}(W\Sigma)\} = 0,$$

$$\mathrm{E}_{\exp}\{(Y_j - \mu_j)^r \ \mathrm{tr}(W\Sigma)\} = \frac{m_j^{(r+2)} - m_j^{(r)} m_j^{(2)}}{\phi_j^2} \sigma_{jj},$$

$$\mathrm{E}_{\exp}\{(Y_i - \mu_i)(Y_j - \mu_j) \ \mathrm{tr}(W\Sigma)\} = \frac{2 m_i^{(2)} m_j^{(2)}}{\phi_i \phi_j} \sigma_{ij} \qquad (i \neq j).$$

PROOF.

$$\mathrm{E}_{\exp}\{\mathrm{tr}(W\Sigma)\} = \sum_{a=1}^{p} \sum_{b=1}^{p} \mathrm{E}_{\exp}(w_{ab}) \sigma_{ab},$$

$$\mathrm{E}_{\exp}\{(Y_j - \mu_j)^r \ \mathrm{tr}(W\Sigma)\} = \sum_{a=1}^{p} \sum_{b=1}^{p} \mathrm{E}_{\exp}\{(Y_j - \mu_j)^r w_{ab}\} \sigma_{ab},$$

$$\mathrm{E}_{\exp}\{(Y_i - \mu_i)(Y_j - \mu_j) \ \mathrm{tr}(W\Sigma)\} = \sum_{a=1}^{p} \sum_{b=1}^{p} \mathrm{E}_{\exp}\{(Y_i - \mu_i)(Y_j - \mu_j) w_{ab}\} \sigma_{ab} \qquad (i \neq j).$$

From (3.1) and (3.2),

$$\mathrm{E}_{\exp}(w_{ab}) = 0,$$

$$\begin{aligned} \mathrm{E}_{\exp}\{(Y_j - \mu_j)^r w_{aa}\} &= \frac{1}{\phi_a^2} \mathrm{E}_{\exp}[(Y_j - \mu_j)^r \{(Y_a - \mu_a)^2 - m_a^{(2)}\}] \\ &= \begin{cases} \dfrac{1}{\phi_j^2} (m_j^{(r+2)} - m_j^{(r)} m_j^{(2)}) & (j = a), \\ 0 & (\text{otherwise}), \end{cases} \end{aligned}$$

and for $a \neq b$,

$$\mathrm{E}_{\exp}\{(Y_j - \mu_j)^r w_{ab}\} = \frac{1}{\phi_a \phi_b} \mathrm{E}_{\exp}\{(Y_j - \mu_j)^r (Y_a - \mu_a)(Y_b - \mu_b)\} = 0,$$

$$\mathrm{E}_{\exp}\{(Y_i-\mu_i)(Y_j-\mu_j)w_{ab}\} = \frac{1}{\phi_a\phi_b}\mathrm{E}_{\exp}\{(Y_i-\mu_i)(Y_j-\mu_j)(Y_a-\mu_a)(Y_b-\mu_b)\}$$
$$= \begin{cases} \dfrac{1}{\phi_i\phi_j} m_i^{(2)} m_j^{(2)} & (i=a, j=b \text{ or } i=b, j=a), \\ 0 & (\text{otherwise}). \end{cases}$$

Summarizing these results lead to Lemma 3.1. □

Using Lemma 3.1 and (3.5), we obtaine the following theorem.

THEOREM 3.2. *The mean, variance and covariance of* $\mathrm{ME}_p(\boldsymbol{\eta}, \Sigma)$ *are expressd as follow,*

$$\mathrm{E}(Y_j) = \mu_j + \frac{1}{2} b_j^{(3)} \sigma_{jj},$$
$$\mathrm{Var}(Y_j) = \phi_j v_j(\mu_j) + \left(v_j(\mu_j)^2 + \frac{1}{2}\phi_j b_j^{(4)}\right)\sigma_{jj} - \frac{1}{4}\{b_j^{(3)}\}^2 \sigma_{jj}^2,$$
$$\mathrm{Cov}(Y_i, Y_j) = v_i(\mu_i) v_j(\mu_j) \sigma_{ij} - \frac{1}{4} b_i^{(3)} b_j^{(3)} \sigma_{ii}\sigma_{jj}.$$

If $\sigma_{jj}$'s are small, the variance and covariance of $\mathrm{ME}_p(\boldsymbol{\eta}, \Sigma)$ are simplified as in the following corollary.

COROLLARY 3.1. *For small* $\sigma_{jj}$, *the variance and covariance of* $\mathrm{ME}_p(\boldsymbol{\eta}, \Sigma)$ *are written by*

$$\mathrm{Var}(Y_j) \cong \phi_j v_j(\mu_j) + \left(v_j(\mu_j)^2 + \frac{1}{2}\phi_j b_j^{(4)}\right)\sigma_{jj},$$
$$\mathrm{Cov}(Y_i, Y_j) \cong v_i(\mu_i) v_j(\mu_j) \sigma_{ij}.$$

Using Theorem 3.1 and Table 1, it is easy to generalize the univariate distribution in the exponential family to the multivariate version. The following examples are the Poisson and the Binomial types. These examples are related to the multivariate Poisson log-normal distribution and the multivariate Binomial logit-normal distribution in section 2.

EXAMPLE 3.1. POISSON TYPE:
*Under the Poisson distribution, Table* 1 *shows* $\mu_j = \exp(\eta_j)$, $\phi_j = 1$ *and* $v_j(\mu_j) = \mu_j = \exp(\eta_j)$. *Let* $\lambda_j = e^{\eta_j}$ *and* $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_p)'$, *then the probability density function of Poisson type is expressed by*

$$\tilde{f}(\boldsymbol{y}; \boldsymbol{\lambda}, \Sigma) = f_{\mathrm{P}}(\boldsymbol{y}; \boldsymbol{\lambda})\left\{1 + \frac{1}{2}\,\mathrm{tr}(W\Sigma)\right\}, \qquad f_{\mathrm{P}}(\boldsymbol{y}; \boldsymbol{\lambda}) = \prod_{j=1}^{p} \frac{\lambda_j^{y_j}}{y_j!} e^{-\lambda_j}, \quad (3.6)$$

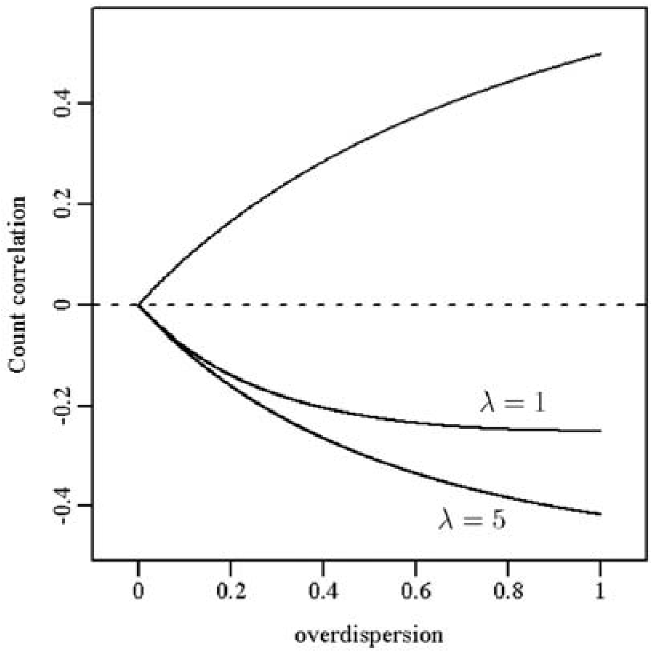


**Fig. 1.** The regions of count correlation and overdispersion attainable by the symmetric bivariate poisson log-normal model.

*where* $W = (\boldsymbol{y} - \boldsymbol{\lambda})(\boldsymbol{y} - \boldsymbol{\lambda})' - \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$, *and the condition* (3.4) *can be reduced to*

$$\sum_{j=1}^{p} \lambda_j \sigma_{jj} \leq 2.$$

*Moreover, for small* $\sigma_{jj}$, *the first two moments are expressed by*

$$\mathrm{E}(Y_j) = \lambda_j\left(1 + \frac{1}{2}\sigma_{jj}\right), \tag{3.7}$$

$$\mathrm{Var}(Y_j) \cong \lambda_j + \lambda_j\left(1 + \frac{1}{2}\lambda_j\right)\sigma_{jj}, \qquad \mathrm{Cov}(Y_i, Y_j) \cong \lambda_i\lambda_j\sigma_{ij}. \tag{3.8}$$

(3.7) *and* (3.8) *also correspond to the approximations of the exact moments of the multivariate Poisson log-normal distribution described in section* 2. *Figure* 2 *to* 5 *are the surface of* (3.6) *for several cases of* $(\lambda_1, \lambda_2, \sigma_{11}, \sigma_{22}, \sigma_{12})$.

Example 3.2. Binomial type:

*Under the Binomial distribution, Table* 1 *shows* $\mu_j = \{1 + \exp(-\eta_j)\}^{-1}$, $\phi_j = 1/n_j$ *and* $v_j(\mu_j) = \mu_j(1 - \mu_j)$. *Let* $\pi_j = \mu_j$ *and* $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_p)'$, *then the probability density function of the Binomial type is expressed by*

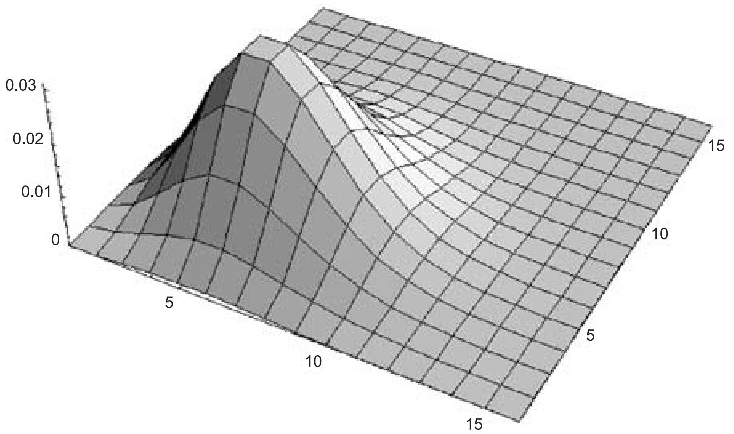


**Fig. 2.** $(\lambda_1, \lambda_2, \sigma_{11}, \sigma_{22}, \sigma_{12}) = (5, 5, 0, 0, 0)$.

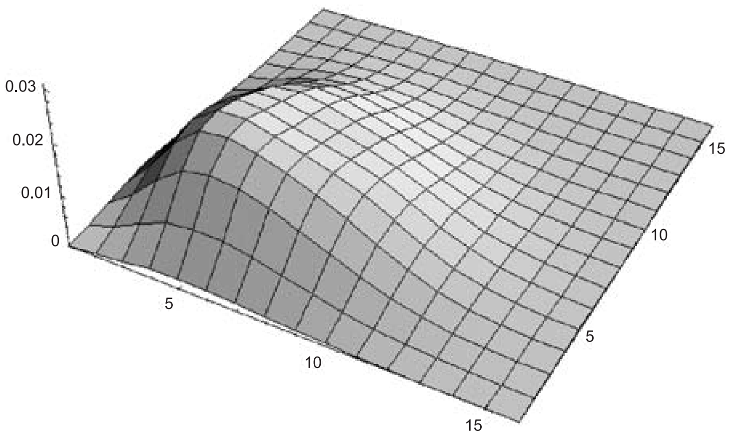


**Fig. 3.** $(\lambda_1, \lambda_2, \sigma_{11}, \sigma_{22}, \sigma_{12}) = (5, 5, 0.1, 0.1, 0)$

$$\tilde{f}(\boldsymbol{y}; \boldsymbol{\pi}, \boldsymbol{m}, \Sigma) = f_{\mathrm{B}}(\boldsymbol{y}; \boldsymbol{\pi}, \boldsymbol{m})\left\{1 + \frac{1}{2}\ \mathrm{tr}(W\Sigma)\right\},$$

*where* $W = (\boldsymbol{y} - \boldsymbol{m} \cdot \boldsymbol{\pi})(\boldsymbol{y} - \boldsymbol{m} \cdot \boldsymbol{\pi})' - \mathrm{diag}(m_1\pi_1(1 - \pi_1), \dots, m_p\pi_p(1 - \pi_p))$, $\boldsymbol{m} \cdot \boldsymbol{\pi} = (m_1\pi_1, \dots, m_p\pi_p)'$, *and the condition* (3.4) *can be reduced to*

$$\sum_{j=1}^{p} m_j\pi_j(1 - \pi_j)\sigma_{jj} \le 2.$$

*Moreover, for small* $\sigma_{jj}$, *the first two moments are expressed by*

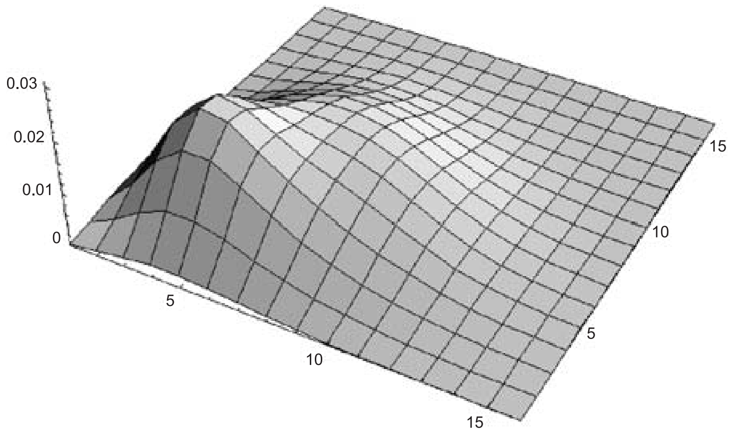


**Fig. 4.** $(\lambda_1, \lambda_2, \sigma_{11}, \sigma_{22}, \sigma_{12}) = (5, 5, 0.1, 0.1, 0.1)$

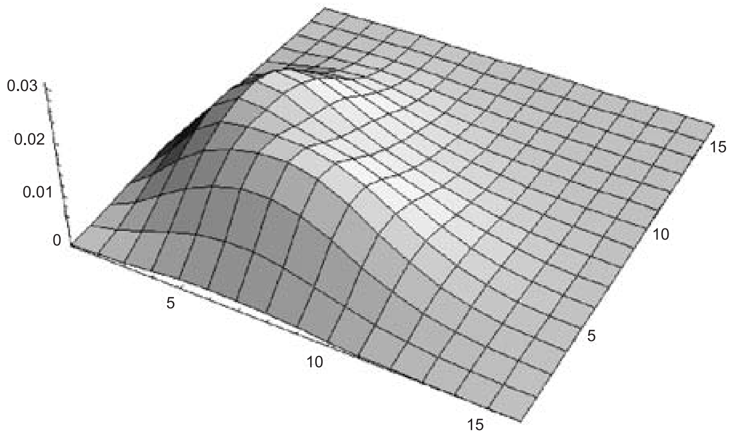


**Fig. 5.** $(\lambda_1, \lambda_2, \sigma_{11}, \sigma_{22}, \sigma_{12}) = (5, 5, 0.1, 0.1, -0.1)$

$$\mathrm{E}(Y_j) = \pi_j + \pi_j(1-\pi_j)\left(\frac{1}{2} - \pi_j\right)\sigma_{jj}$$

$$\mathrm{Var}(Y_j) \cong \frac{\pi_j(1-\pi_j)}{n_j}\left[1 + \left\{\frac{1}{2} + (n_j - 3)\pi_j(1-\pi_j)\right\}\sigma_{jj}\right]$$

$$\mathrm{Cov}(Y_i, Y_j) \cong \pi_i(1-\pi_i)\pi_j(1-\pi_j)\sigma_{ij}.$$

Next, we consider the improvement of approximation when (3.3) is used as an approximation formula. If the response variables are independent, the density function can be expressed by a product of the marginal distributions. If $\sigma_{jj}$ is small, it can be expanded by

$$\begin{aligned}\tilde{f}(\boldsymbol{y};\boldsymbol{\eta},\Sigma) &= f_{\exp}(\boldsymbol{y};\boldsymbol{\eta})\prod_{j=1}^{p}\left\{1+\frac{1}{2}w_{jj}\sigma_{jj}\right\}\\ &\cong f_{\exp}(\boldsymbol{y};\boldsymbol{\eta})\left\{1+\frac{1}{2}\sum_{j=1}^{p}w_{jj}\sigma_{jj}+\frac{1}{4}\sum_{i<j}^{p}w_{ii}\sigma_{ii}w_{jj}\sigma_{jj}\right\}.\end{aligned}$$

In order to improve the accuracies of the approximation near $\rho_{ij}=0$, we suggest adding the correct term as follows:

$$\tilde{f}(\boldsymbol{y};\boldsymbol{\eta},\Sigma) = f_{\exp}(\boldsymbol{y};\boldsymbol{\eta})\left\{1+\frac{1}{2}\ \mathrm{tr}(W\Sigma)+\frac{1}{4}\sum_{i<j}^{p}w_{ii}w_{jj}\sigma_{ii}\sigma_{jj}r(\rho_{ij})\right\}. \tag{3.9}$$

where $r(\rho_{ij})$ is a smooth weight function which satisfies $r(0)=1$ and $r(1)=r(-1)=0$, for example $r(\rho)=0.5\{1+\cos(\rho\pi)\}$. The last term in (3.9) corresponds to the higher order term in the independent case. Note that (3.9) may not be the density probability function under the condition (3.4).

## 4. Parameter estimation

The maximum likelihood estimation of the parameters $\boldsymbol{\theta}=(\boldsymbol{\eta},\Sigma)$ is complicated computationally for the models described in section 2 because of the multidimensional integrals in (2.5), (2.15) and (2.18). In previous works, some numerical optimizations were applied. Aitchison and Ho (1989) and Coull and Agresti (2000) used the multidimensional Gauss-Hermite quadrature for (2.5) and (2.15), respectively. Chib and Winkelmann (2001) have applied the Markov Chain Monte Carlo method for (2.5). Gueorguieva (2001) has proposed the EM algorithm for (2.18). However, when the dimension of response is large, these numerical optimization based methods are difficult to apply real data analysis, because of enormous amounts of calculation. Therefore, as another approach, we use the maximum likelihood estimator based on the proposed p.d.f. (3.3) regarded as a full likelihood,

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}\in\Theta}\ \tilde{\ell}(\boldsymbol{\theta}), \qquad \tilde{\ell}(\boldsymbol{\theta}) = \sum_{i=1}^{N}\log\tilde{f}(\boldsymbol{y}_i;\boldsymbol{\eta},\Sigma).$$

Note that the condition (3.4) restricts the parameter space of $\boldsymbol{\theta}$ as follows,

$$\Theta = \left\{\boldsymbol{\theta}\,|\sum_{j=1}^{p}\frac{1}{\phi_j}v(\mu_j)\sigma_{jj}\le 2\right\}.$$

The maximum likelihood estimator due to (3.3) can be obtained through the numerical optimization based on the SPIDER algorithm proposed by Ohtaki and Izumi (1999).

## 5. Tests of covariance structures

Based on the type of covariance parameter $\Sigma$, there are a number of interesting hypotheses. For example, in the bivariate case, $\boldsymbol{y} = (Y_1, Y_2)'$, we are interested in the following hypotheses,

$$\text{(i)} \quad H_0 : \sigma_{12} = 0, \qquad \text{(ii)} \quad H_0 : \sigma_{22} = \sigma_{12} = 0, \qquad \text{(iii)} \quad H_0 : \Sigma = O.$$

The hypothesis (i) leads to an independent univariate generalized linear mixed model, hypothesis (iii) presents an independent generalized linear model, and hypothesis (ii) is an intermediate hypothesis of (i) and (iii) which presents that $Y_1$ and $Y_2$ are independent, are a generalized linear mixed model and a generalized linear model, respectively. These hypotheses can be examined by the likelihood ratio test based on (3.3),

$$T = -2\{\tilde{\ell}(\boldsymbol{\theta}_0) - \tilde{\ell}(\boldsymbol{\theta})\},$$

where $\boldsymbol{\theta}_0$ is the maximum likelihood estimator under the null hypothesis. Since the parameter values under the null hypothesis (ii) or (iii) is on the boundary of the parameter space, the asymptotic distribution of the likelihood ratio test statistic is not a chi-squared distribution. Self and Liang (1987) derived the asymptotic distribution of the likelihood ratio statistics under such a situation. The hypotheses (ii) and (iii) correspond to *Case* 6 and *Case* 9 in Self and Liang (1987). From their results, the asymptotic distributions under the null hypotheses (i), (ii) and (iii) are given by

$$\text{(i)} \quad \mathrm{P}(\chi_1^2 \le x),$$

$$\text{(ii)} \quad \frac{1}{2}\{\mathrm{P}(\chi_1^2 \le x) + \mathrm{P}(\chi_2^2 \le x)\},$$

$$\text{(iii)} \quad \frac{1}{4}\{\mathrm{P}(\chi_1^2 \le x) + 2\mathrm{P}(\chi_2^2 \le x) + \mathrm{P}(\chi_3^2 \le x)\}.$$

respectively, where $\chi_k^2$ denotes the central chi-squared variable with $k$ degrees of freedom.

## 6. Simulation study

The numerical experiments are studied under two situations (simulations I and II). In simulation I, it is assumed that the true distribution is the GLMM, and in simulation II, it is assumed that the dataset are generated by the proposed class of distribution (3.3). In each experiment, the number of repetitions is 1,000.

**Simulation I:** The purpose of this simulation study is to see the tendencies of the maximum likelihood estimator and the likelihood ratio test criterion

Table 2. MLEs under the case I.

| $\sigma^2$ | $\rho$ | $\hat{\lambda}_1$ | | $\hat{\lambda}_2$ | | $\hat{\sigma}_1^2$ | | $\hat{\sigma}_2^2$ | | $\hat{\rho}$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.1 | 0.0 | 5.044 | (0.137) | 5.033 | (0.138) | 0.083 | (0.016) | 0.083 | (0.016) | 0.022 | (0.156) |
| | 0.1 | 5.026 | (0.141) | 5.038 | (0.143) | 0.083 | (0.018) | 0.084 | (0.018) | 0.101 | (0.162) |
| | 0.2 | 5.039 | (0.146) | 5.042 | (0.146) | 0.083 | (0.018) | 0.081 | (0.017) | 0.193 | (0.161) |
| | 0.3 | 5.057 | (0.145) | 5.051 | (0.147) | 0.083 | (0.017) | 0.081 | (0.018) | 0.312 | (0.182) |
| | 0.4 | 5.055 | (0.149) | 5.054 | (0.143) | 0.080 | (0.017) | 0.080 | (0.017) | 0.421 | (0.180) |
| | 0.5 | 5.063 | (0.154) | 5.060 | (0.149) | 0.079 | (0.016) | 0.079 | (0.018) | 0.512 | (0.179) |
| 0.05 | 0.0 | 5.017 | (0.134) | 5.008 | (0.122) | 0.049 | (0.016) | 0.049 | (0.016) | 0.020 | (0.273) |
| | 0.1 | 4.999 | (0.118) | 5.001 | (0.125) | 0.048 | (0.016) | 0.048 | (0.016) | 0.130 | (0.295) |
| | 0.2 | 5.015 | (0.128) | 5.013 | (0.128) | 0.049 | (0.016) | 0.048 | (0.016) | 0.220 | (0.276) |
| | 0.3 | 5.003 | (0.124) | 5.012 | (0.128) | 0.047 | (0.017) | 0.048 | (0.016) | 0.282 | (0.275) |
| | 0.4 | 5.019 | (0.125) | 5.011 | (0.119) | 0.047 | (0.015) | 0.048 | (0.015) | 0.396 | (0.271) |
| | 0.5 | 5.010 | (0.125) | 5.007 | (0.125) | 0.048 | (0.016) | 0.048 | (0.016) | 0.491 | (0.256) |

* The number of repetitions is 1,000 and the sample size is 400.

when the proposed density function is regarded as an approximate density of GLMM. It is assumed that the true model is the symmetric bivariate Poisson log-normal model, such as

$$\boldsymbol{\lambda} = \lambda \mathbf{1}_2, \qquad \Sigma = \sigma^2\{\rho \mathbf{1}_2 \mathbf{1}_2' + (1-\rho) I_2\}.$$

Tables 2, 3 and 4 give the maximum likelihood estimates of $\lambda_j$, $\sigma_{jj}$ $(j = 1, 2)$ and $\rho$ based on (3.3) under the following two cases

$$\text{Case I.} \qquad \lambda = 5, \qquad \sigma^2 = 0.1, 0.05, \qquad \rho = 0.0, \dots, 0.5$$

$$\text{Case II.} \qquad \lambda = 1, \qquad \sigma^2 = 0.5, \qquad \rho = 0.0, \dots, 0.5.$$

In both of cases, the sample size is 400. The parenthesized values are standard deviations which is defined by $\{\sum_{r=1}^{R}(\hat{\theta}_r - \bar{\theta})^2/R\}^{1/2}$ where $\bar{\theta} = \sum_{r=1}^{R}\hat{\theta}_r/R$, $R$ denotes the number of repetitions and $\theta_r$ denotes the maximum likelihood estimate of $\lambda_1$, $\lambda_2$, $\sigma_1^2$, $\sigma_2^2$, $\rho$ based on $r$th dataset.

Table 2 and 3 are results of MLEs based on (3.3), and Table 4 is based on (3.9). Table 5 shows the results of the likelihood ratio tests for three covariance structures. The column labeled "Null" means the actual test sizes. The settings of the true parameters are $\lambda_j = 5$, $d = 0.1, 0.3, 0.5$ and $\rho = 0.0, \dots,$ 0.5, where $d$ means the index of overdispersion relative to the Poisson distribution defined by $d = \lambda(e^{\sigma^2} - 1)$ described in Aitchison and Ho (1989). In each cell, the top, middle and bottom lines mean that the sample sizes are 50, 100 and 200.

From Tables 2, 3 and 4, it can be seen that that there are some biases for the estimator of the variance component and instability for the estimator of the

Table 3. MLEs under the case II.

| $\sigma^2$ | $\rho$ | $\hat{\lambda}_1$ | | $\hat{\lambda}_2$ | | $\hat{\sigma}_1^2$ | | $\hat{\sigma}_2^2$ | | $\hat{\rho}$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.5 | 0.0 | 1.046 | (0.070) | 1.048 | (0.075) | 0.391 | (0.093) | 0.400 | (0.094) | 0.079 | (0.193) |
| | 0.1 | 1.041 | (0.069) | 1.046 | (0.073) | 0.395 | (0.097) | 0.387 | (0.094) | 0.187 | (0.210) |
| | 0.2 | 1.051 | (0.074) | 1.050 | (0.079) | 0.389 | (0.096) | 0.383 | (0.096) | 0.279 | (0.211) |
| | 0.3 | 1.052 | (0.079) | 1.050 | (0.073) | 0.372 | (0.088) | 0.377 | (0.096) | 0.406 | (0.198) |
| | 0.4 | 1.049 | (0.074) | 1.051 | (0.077) | 0.355 | (0.096) | 0.363 | (0.096) | 0.536 | (0.220) |
| | 0.5 | 1.056 | (0.072) | 1.054 | (0.074) | 0.344 | (0.091) | 0.343 | (0.088) | 0.653 | (0.220) |

* The number of repetitions is 1,000 and the sample size is 400.

Table 4. MLEs with correct term under the case II.

| $\sigma^2$ | $\rho$ | $\hat{\lambda}_1$ | | $\hat{\lambda}_2$ | | $\hat{\sigma}_1^2$ | | $\hat{\sigma}_2^2$ | | $\hat{\rho}$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.5 | 0.0 | 1.046 | (0.069) | 1.047 | (0.077) | 0.426 | (0.109) | 0.415 | (0.097) | 0.030 | (0.199) |
| | 0.1 | 1.048 | (0.073) | 1.048 | (0.071) | 0.424 | (0.096) | 0.423 | (0.107) | 0.125 | (0.198) |
| | 0.2 | 1.049 | (0.069) | 1.048 | (0.077) | 0.416 | (0.097) | 0.414 | (0.100) | 0.202 | (0.209) |
| | 0.3 | 1.057 | (0.076) | 1.051 | (0.074) | 0.412 | (0.098) | 0.415 | (0.103) | 0.290 | (0.193) |
| | 0.4 | 1.053 | (0.075) | 1.059 | (0.073) | 0.403 | (0.101) | 0.406 | (0.102) | 0.397 | (0.217) |
| | 0.5 | 1.060 | (0.077) | 1.056 | (0.074) | 0.409 | (0.110) | 0.400 | (0.101) | 0.474 | (0.241) |

* The number of repetitions is 1,000 and the sample size is 400.

correlation coefficient, even if the sample size is large ($N = 400$). These results show the limits of the Poisson log-normal model (Aitchison and Ho, 1989), which is illustrated as follows.

Under the symmetric bivariate Poisson log-normal model, the first two moments are written by the simple expressions $\mathrm{E}(Y_j) = \lambda$, $\mathrm{Var}(Y_j) = \lambda + \lambda^2(e^{\sigma^2} - 1)$, $\mathrm{Cov}(Y_1, Y_2) = \lambda^2(e^{\sigma_{12}} - 1)$. Then, the overdispersion on the marginal distribution is expressed by $d = \lambda(e^{\sigma^2} - 1)$. Using this expression and (2.8), the correlation coefficient of $Y_1$ and $Y_2$ is rewritten by $\mathrm{Corr}(Y_1, Y_2) = \lambda(e^{\sigma_{12}} - 1)/(d + 1)$. Note that the limiting values of $\sigma_{12}$ are $-\sigma^2$ and $\sigma^2$. The correlation coefficient $\mathrm{Corr}(Y_1, Y_2)$ in this model has the following upper and lower bounds,

$$-\frac{\lambda d}{(d + \lambda)(d + 1)} \leq \mathrm{Corr}(Y_1, Y_2) \leq \frac{d}{d + 1}. \tag{6.1}$$

Figure 1 illustrates the interval of the correlation coefficient supported by the multivariate Poisson Log-Normal model, which is the plot of $d$ and (6.1) in the case of $\lambda = 1, 5$. From Figure 1, it follows that the range of the correlation coefficient supported by this model is not wide, when the mean parameters $\lambda_j$'s are not large. From Table 5, we can see that the actual test sizes are close

Table 5. The actual test sizes and powers of the likelihood ratio tests.

| $H_0$ | $d$ | $N$ | Null | $\rho$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | 0.0 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 |
| (i) | 10% | 50 | 24.0 | — | 22.1 | 24.1 | 21.8 | 23.5 | 23.6 |
| | | 100 | 15.5 | — | 15.4 | 16.1 | 15.3 | 16.6 | 17.6 |
| | | 200 | 10.7 | — | 12.2 | 11.0 | 10.0 | 11.7 | 15.9 |
| | 30% | 50 | 13.1 | — | 13.9 | 14.4 | 15.8 | 17.6 | 19.5 |
| | | 100 | 6.2 | — | 7.6 | 9.6 | 12.8 | 17.7 | 20.3 |
| | | 200 | 5.8 | — | 6.4 | 9.1 | 14.2 | 25.3 | 35.7 |
| | 50% | 50 | 8.2 | — | 8.2 | 11.4 | 13.8 | 17.9 | 25.0 |
| | | 100 | 7.8 | — | 7.9 | 10.7 | 17.9 | 23.4 | 33.1 |
| | | 200 | 3.3 | — | 8.5 | 14.8 | 23.5 | 39.9 | 55.6 |
| (ii) | 10% | 50 | 6.0 | — | 11.3 | 9.8 | 12.0 | 9.8 | 11.0 |
| | | 100 | 4.7 | — | 11.4 | 13.1 | 11.4 | 14.1 | 13.9 |
| | | 200 | 5.2 | — | 18.4 | 16.5 | 18.2 | 20.3 | 19.0 |
| | 30% | 50 | 7.2 | — | 24.2 | 24.8 | 24.5 | 27.8 | 29.6 |
| | | 100 | 5.0 | — | 45.0 | 41.2 | 43.3 | 48.7 | 50.4 |
| | | 200 | 4.3 | — | 69.1 | 69.3 | 68.7 | 76.7 | 78.8 |
| | 50% | 50 | 5.9 | — | 46.6 | 45.3 | 44.8 | 51.4 | 52.4 |
| | | 100 | 6.3 | — | 69.0 | 75.2 | 74.4 | 78.2 | 80.2 |
| | | 200 | 5.4 | — | 93.4 | 94.1 | 94.8 | 97.1 | 97.7 |
| (iii) | 0% | 50 | 5.5 | — | — | — | — | — | — |
| | | 100 | 5.3 | — | — | — | — | — | — |
| | | 200 | 4.9 | — | — | — | — | — | — |
| | 10% | 50 | — | 12.2 | 11.6 | 12.3 | 10.3 | 15.3 | 13.0 |
| | | 100 | — | 15.9 | 15.4 | 17.2 | 18.9 | 18.1 | 19.4 |
| | | 200 | — | 26.8 | 27.0 | 25.2 | 28.1 | 28.4 | 31.2 |
| | 30% | 50 | — | 40.4 | 40.4 | 40.8 | 41.4 | 41.4 | 44.4 |
| | | 100 | — | 64.1 | 63.7 | 66.9 | 69.5 | 69.6 | 69.8 |
| | | 200 | — | 90.6 | 91.4 | 93.5 | 92.6 | 94.0 | 95.3 |
| | 50% | 50 | — | 70.7 | 67.7 | 71.5 | 70.5 | 73.3 | 78.0 |
| | | 100 | — | 93.9 | 94.8 | 94.9 | 94.8 | 95.5 | 94.8 |
| | | 200 | — | 99.8 | 99.7 | 99.9 | 99.9 | 99.9 | 100.0 |

* The number of repetitions is 1,000.

to the corresponding nominal sizes in almost all cases except for the case of hypothesis (i) and small $d$, and that the power of test for the correlation is poor. It can be seen that these poor results are obtained due to the same problem that caused the instability of the estimation of $\rho$.

Table 6. MLEs for the Poisson type.

| $\lambda$ | $\sigma^2$ | $\sigma_{12}$ | $N$ | $\lambda_1$ | | $\lambda_2$ | | $\sigma_{11}$ | | $\sigma_{22}$ | | $\sigma_{12}$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | 0.05 | −0.03 | 50 | 4.981 | (0.375) | 4.989 | (0.384) | 0.053 | (0.049) | 0.052 | (0.050) | −0.033 | (0.040) |
| | | | 100 | 5.003 | (0.271) | 4.988 | (0.270) | 0.051 | (0.036) | 0.049 | (0.035) | −0.032 | (0.024) |
| | | | 200 | 5.015 | (0.179) | 4.999 | (0.180) | 0.049 | (0.025) | 0.049 | (0.024) | −0.030 | (0.017) |
| | | 0.00 | 50 | 4.971 | (0.370) | 4.982 | (0.370) | 0.051 | (0.048) | 0.054 | (0.050) | 0.000 | (0.040) |
| | | | 100 | 4.999 | (0.268) | 5.007 | (0.265) | 0.050 | (0.035) | 0.049 | (0.034) | 0.001 | (0.025) |
| | | | 200 | 4.996 | (0.185) | 5.005 | (0.184) | 0.049 | (0.024) | 0.048 | (0.024) | −0.001 | (0.017) |
| | | 0.03 | 50 | 5.001 | (0.375) | 5.022 | (0.378) | 0.050 | (0.048) | 0.052 | (0.049) | 0.035 | (0.042) |
| | | | 100 | 5.006 | (0.257) | 5.012 | (0.258) | 0.049 | (0.035) | 0.049 | (0.033) | 0.030 | (0.026) |
| | | | 200 | 5.004 | (0.189) | 5.000 | (0.182) | 0.049 | (0.025) | 0.049 | (0.024) | 0.031 | (0.017) |
| | 0.10 | −0.06 | 50 | 4.993 | (0.403) | 5.010 | (0.417) | 0.100 | (0.062) | 0.098 | (0.062) | −0.064 | (0.047) |
| | | | 100 | 4.979 | (0.278) | 5.003 | (0.276) | 0.102 | (0.041) | 0.101 | (0.039) | −0.062 | (0.026) |
| | | | 200 | 5.011 | (0.195) | 4.997 | (0.186) | 0.099 | (0.026) | 0.100 | (0.028) | −0.060 | (0.017) |
| | | 0.00 | 50 | 5.003 | (0.427) | 5.014 | (0.416) | 0.102 | (0.061) | 0.102 | (0.063) | −0.001 | (0.042) |
| | | | 100 | 5.014 | (0.287) | 5.008 | (0.292) | 0.099 | (0.041) | 0.101 | (0.040) | 0.000 | (0.027) |
| | | | 200 | 5.006 | (0.202) | 4.998 | (0.201) | 0.099 | (0.028) | 0.101 | (0.028) | 0.000 | (0.019) |
| | | 0.06 | 50 | 5.012 | (0.398) | 5.022 | (0.384) | 0.097 | (0.056) | 0.098 | (0.056) | 0.064 | (0.042) |
| | | | 100 | 5.001 | (0.282) | 5.008 | (0.274) | 0.101 | (0.039) | 0.100 | (0.041) | 0.061 | (0.027) |
| | | | 200 | 5.010 | (0.185) | 4.998 | (0.192) | 0.100 | (0.026) | 0.101 | (0.026) | 0.061 | (0.018) |
| | 0.15 | −0.09 | 50 | 4.983 | (0.387) | 5.000 | (0.385) | 0.154 | (0.062) | 0.155 | (0.063) | −0.092 | (0.041) |
| | | | 100 | 5.007 | (0.255) | 5.010 | (0.258) | 0.150 | (0.039) | 0.151 | (0.037) | −0.092 | (0.025) |
| | | | 200 | 5.002 | (0.179) | 4.993 | (0.181) | 0.152 | (0.026) | 0.150 | (0.026) | −0.091 | (0.018) |
| | | 0.00 | 50 | 5.004 | (0.411) | 5.012 | (0.407) | 0.155 | (0.064) | 0.150 | (0.062) | 0.001 | (0.041) |
| | | | 100 | 5.000 | (0.270) | 4.999 | (0.273) | 0.153 | (0.042) | 0.150 | (0.040) | 0.001 | (0.029) |
| | | | 200 | 4.990 | (0.194) | 5.000 | (0.195) | 0.152 | (0.028) | 0.151 | (0.027) | 0.001 | (0.020) |
| | | 0.09 | 50 | 4.975 | (0.387) | 5.025 | (0.412) | 0.157 | (0.065) | 0.158 | (0.065) | 0.095 | (0.046) |
| | | | 100 | 5.002 | (0.261) | 4.993 | (0.273) | 0.153 | (0.039) | 0.149 | (0.039) | 0.092 | (0.026) |
| | | | 200 | 4.999 | (0.188) | 4.997 | (0.178) | 0.150 | (0.026) | 0.151 | (0.027) | 0.091 | (0.019) |

* The number of repetitions is 1,000.

**Simulation II:** In this simulation, we generate the dataset from the proposed class of distribution (3.3). The conditional distribution approach (Johnson, 1987, Chapter 3) is used to generate a $p$-variate random vector from $\mathrm{ME}(\boldsymbol{\eta}, \Sigma)$. The purpose is to see the numerical behavior of MLEs. It is assumed that the true model is the Poisson type as in Example 3.1, and the settings of the parameters are $\lambda_j = \lambda = 5$, $\sigma_{jj} = \sigma^2 = 0.05, 0.10, 0.15$ $(j = 1, 2)$ and $\sigma_{12} = -0.6\sigma^2, 0, 0.6\sigma^2$. Table 6 gives the MLEs of $(\lambda_1, \lambda_2, \sigma_{11}, \sigma_{22}, \sigma_{12})$. From Table 6, it can be seen that the performances of the MLEs are good in almost all cases, except for the estimation of $\sigma_{12}$ in the case in which the sample size is small.

From these simulation studies, we recommend using (3.3) not as the approximation, but as the density function. The advantage of our approach is that our method needs no assumptions on the conditional distribution. Therefore, it is possible to apply a similar approach to other conditional distributions not included in the exponential family. For example, assuming that the conditional distribution is the Weibull distribution, a new multivariate distribution for analyzing multivariate survival data can be constructed.

## Acknowledgment

I wish to express my deepest gratitude to Prof. Y. Fujikoshi of Hiroshima University for his valuable advices and encouragements. Also I would like to thank Prof. M. Ohtaki of Hiroshima University for his helpful comments and suggestions. Further, I am grateful to Dr. H. Wakaki and Dr. K. Satoh of Hiroshima University and Dr. H. Yanagihara of Tsukuba University for their supports and encouragements.

## References

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

[2] Breslow, N. E. and Clayton, D. G. Approximate inference in generalized linear mixed models. J. Amer. Statist. Assoc, **88** (1993), 9–25.

[3] Breslow, N. E. and Lin, X. Bias correction in generalized linear mixed models with a single component of dispersion. Biometrika, **82** (1995), 81–92.

[4] Campbell, J. T. The Poisson correlation function. Proc. Edin. Math. Soc. Ser. 2, **4** (1934), 18–26.

[5] Carroll, R. J. and Ruppert, D. Robust estimation in heterscedastic linear models. Ann. Statist., **10** (1982), 429–441.

[6] Chib, S. and Winkelmann, R. Markov chain Monte Carlo analysis of correlated count data. J. Bus. Econ. Stat., **19** (2001), 428–435.

[7] Coull, B. A. and Agresti, A. Random effects modeling of multiple binomial responses using the multivariate binomial logit-normal distribution. Biometrics, **56** (2000), 73–80.

[8] Crow, E. L. and Shimizu, K. Lognormal distributions: Statistics: Textbooks and Monographs, 88. Marcel Dekker, Inc., New York 1988.

[9] Gueorguieva, R. A multivariate generalized linear mixed model for joint modelling of clustered outcomes in the exponential family. Stat. Model., **1** (2001), 177–193.

[10] Holgate, P. Estimation for the bivariate Poisson distribution. Biometrika, **51** (1964), 241–245.

[11] Joe, H. Multivariate models and dependence concepts: Monographs on Statistics and Applied Probability, 73. Chapman & Hall, London 1997.

[12] Johnson, M. E. Multivariate statisticsl simulation. John Wiley & Sons, New-York 1987.

[13] McCullagh, P and Nelder, J. A. Generalized linear models. Chapman and Hall, London 1989.

[14] McKendrick, A. G. Applications of the kinetic theory of gases to vital phenomena. Indian J. Med. Res., **3** (1916), 667–687.

[15] McKendrick, A. G. Applications of mathematics to medical problems. Proc. Edin. Math. Soc., **44** (1926), 98–130.

[16] Ohtaki, M. and Izumi, S. Globally convergent algorithm without derivatives for maximizing a multivariate function. Proceedings of Development of statistical theories and their application for complex nonlinear data. Hiroshima, Japan: Hiroshima University 1999.

[17] Self, S. G. and Liang, K.-Y. Asymptotic properties of maximum likelihood estimator and likelihood ratio tests under nonstandard conditions. J. Amer. Statist. Assoc, **82** (1987), 605–610.

[18] Steyn, K. S. On the multivariate poisson normal distribution. J. Amer. Statist. Assoc, **71** (1976), 233–236.

[19] Sutradhar, B. C. and Rao, R. P. On marginal quasi-likelihood inference in generalized linear mixed models. J. Multivariate Anal., **76** (2001), 1–34.

[20] Zeger, S. L., Liang, K.-Y. and Albert, P. S. Models for longitudinal data: A generalized estimating equation approach. Biometrics, **44** (1988), 1049–1060.

*Tetsuji Tonda*
*Department of Environmetrics and Biometrics*
*Research Institute for Radiation Biology and Medicine*
*Hiroshima University*
1-2-3 *Kasumi, Minami-Ku, Hiroshima* 734-8553, *Japan*
*e-mail: ttetsuji@hiroshima-u.ac.jp*