RT-MAE 9607

REGRESSION MODELS FOR BIVARIATE COUNTS

by

Linda Lee Ho
and
Julio da Motta Singer

Palavras Chaves: Bivariate counts; multivariate Poisson log-normal distribution; multivariate Poisson disbribution; regression models.
(Key words)

Classificação AMS: J99, H20.
(AMS Classification)

- Fevereiro de 1996 -

REGRESSION MODELS FOR BIVARIATE COUNTS

Linda Lee Ho
Departamento de Engenharia de Produção, Escola Politécnica
Universidade de São Paulo
Caixa Postal 61548
São Paulo, SP 05424-970
Brazil

Julio da Motta Singer
Departamento de Estatística, IME
Universidade de São Paulo
Caixa Postal 66281
São Paulo, SP 05389-970
Brazil

Summary

We consider bivariate Poisson and bivariate Poisson log-normal regression models to analyse counts obtained under a stratified sampling scheme. In either case, Newton-Raphson's method is used to obtain the maximum likelihood estimates of the relevant parameters. The proposed techniques are illustrated with numerical examples for which both underlying probabilistic models are compared.

Key Words: Bivariate counts; multivariate Poisson log-normal distribution; multivariate Poisson distribution; regression models.

1 Introduction

We are concerned with the analysis of bivariate counts observed in sampling units obtained according to a stratified sampling scheme. For example, we may consider the data described in Ho (1995) and reproduced in Tables 1.1-1.2 where two types of defects are counted in 430 samples of 100 g of textile fibers produced by four machines, A_1, A_2, B_1 and B_2, with A_j, $j = 1, 2$ from manufacturer A and the other two from manufacturer B. In this context, we may be interested in comparing the expected defect rates as well as in evaluating both the intensity and the homogeneity of the association between the frequencies of the two types of defects.

For such purposes, we consider two probability models: the bivariate Poisson distribution and the bivariate Poisson log-normal distribution. The former has the advantage of simplicity, but the latter has the attractive property of admitting both positive and negative correlation between its two components, as pointed in Aitchison and Ho (1989). Negative correlation models are of interest in situations where the presence of one type of defect inhibits the presence of the other. The case of a single stratum was presented in Aitchison and Ho (1989): here we extend their results to the case of several strata and more generally to incorporate regression models under either one of the probabilistic models mentioned above.

Table 1.1: Frequency of defects produced by machines from manufacturer A.

Machine	Number of defects of type 1	Number of defects of type 2: 0	1	2	3	4	5	6	7	Total
A_1	0	13	26	22	16	3	2	.	.	82
	1	3	9	12	7	5	2	1	.	39
	2	.	2	1	.	.	1	.	.	4
	3	.	1	.	2	.	.	.	1	4
	4	.	.	1	.	.	.	.	.	1
	Total	16	38	38	25	8	5	1	1	130
A_2	0	15	14	11	10	3	1	.	.	54
	1	2	12	6	7	5	3	.	1	36
	2	.	1	2	1	.	2	.	.	6
	3	.	.	.	.	1	1	.	1	3
	4	.	.	.	1	.	.	.	.	1
	Total	17	27	19	19	9	7	0	2	100

In Section 2 we introduce the regression models and indicate how to obtain maximum likelihood estimators of the relevant parameters via Newton-Raphson's method as well as Wald's test statistics for hypotheses of interest. The proposed techniques are illustrated via a numerical example in Section 3. Finally, in Section 4, we present a brief discussion of the results.

2 Regression models

The regression models under investigation here may be expressed as

$$E(\mathbf{Y}_i) = f(\mathbf{X}_i, \Theta) \tag{2.1}$$

Table 1.2: Frequency of defects produced by machines from manufacturer B

Machine	Number of defects of type 1	Number of defects of type 2: 0	1	2	3	4	5	Total
B_1	0	11	35	21	13	5	1	86
	1	1	5	3	1	1	.	11
	2	.	1	.	2	.	.	3
	Total	12	41	24	16	6	1	100
B_2	0	20	35	17	7	4	1	84
	1	4	7	3	2	.	.	16
	Total	24	42	20	9	4	1	100

where $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{iK})$ is a K-dimensional random vector with Y_{ik} denoting the value of the kth variate for the ith sampling unit, $k = 1, \ldots, K$, $\mathbf{X}_i = (X_{1i}, X_{2i}, \ldots, X_{ti})'$ is a vector with the values of t predictor variables, $i = 1, \ldots, n$, Θ is the r-dimensional vector of unknown parameters and the function f relates the dependent variable to the predictor variables.

First, assume that the $\mathbf{Y}_i's$ follow independent multivariate Poisson distributions, i.e. with probability function (see Kawamura, 1979)

$$P(\mathbf{Y}_i) = \begin{cases} \sum_{\mathbf{U} \in \mathbf{A}} \frac{\prod_{\mathbf{m}} \lambda_{\mathbf{m}i}^{U_{\mathbf{m}}} \exp\{-\sum_{\mathbf{m}} \lambda_{\mathbf{m}i}\}}{\prod_{\mathbf{m}} U_{\mathbf{m}}} & \text{if } Y_{i1} \geq 0, \ldots, Y_{iK} \geq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{2.2}$$

where $\mathbf{m}=(m_1, m_2, \ldots, m_K)'$ is a K-dimensional vector, with each element equal one or zero and at least one non-null element, $\lambda_{\mathbf{m}i}$ is a positive constant, $\mathbf{U} = (U_{100\ldots0}, U_{010\ldots0}, \ldots, U_{110\ldots0}, \ldots, U_{111\ldots1})'$ and $\mathbf{A} = \{\mathbf{U} \in \mathbf{N}^{2^K-1}: \sum_{m_1=1} U_{\mathbf{m}}=Y_{i1}, \sum_{m_2=1} U_{\mathbf{m}}=Y_{i2}, \ldots \sum_{m_K=1} U_{\mathbf{m}}=Y_{iK}\}$. Under this setup, a convenient regression model is

$$E(\mathbf{Y}) = \exp\{\mathbf{X}\boldsymbol{\beta}\} \tag{2.3}$$

where $\mathbf{Y} = (\mathbf{Y}_1', \mathbf{Y}_2', \ldots, \mathbf{Y}_n')'$, $\mathbf{X} = (\mathbf{X}_1', \mathbf{X}_2', \ldots, \mathbf{X}_n')'$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \ldots, \boldsymbol{\beta}_t')'$ is a Kt-dimensional vector of unknown parameters. Here, Θ in (2.1) is given by $(\boldsymbol{\beta}', \boldsymbol{\lambda}')'$

where $\boldsymbol{\lambda}$={ $\lambda_{\mathbf{m}}$, such that at least two components of $\mathbf{m}$ are equal to 1}. An example with $K = 2$ and $t = 4$ will be detailed in the final part of this section.

Now suppose that the $\mathbf{Y}_i's$ follow independent multivariate Poisson log-normal distributions, i.e. with probability function (see Aitchinson and Ho, 1989)

$$P(\mathbf{Y}_i \mid \mathbf{X}_i, \boldsymbol{\beta}, \boldsymbol{\Sigma}_i) = \int_{R_+^K} \prod_{j=1}^{K} f(Y_{ij} \mid \lambda_j) g(\boldsymbol{\lambda} \mid \mathbf{X}_i, \boldsymbol{\beta}, \boldsymbol{\Sigma}_i) d\boldsymbol{\lambda} \tag{2.4}$$

where $f(Y_{ij} \mid \lambda_j)$ is the probability function of a Poisson distribution with parameter λ_j, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_K)'$ and $g(\boldsymbol{\lambda} \mid \mathbf{X}_i, \boldsymbol{\beta}, \boldsymbol{\Sigma}_i)$ is the density function of a multivariate log-normal distribution with parameters $\mathbf{X}_i\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}_i$. In this context, a convenient regression model is

$$E(\mathbf{Y}) = \mathbf{exp}\{\mathbf{X}\boldsymbol{\beta} + \frac{1}{2}\mathbf{vecdiag}(\boldsymbol{\Sigma})\} \tag{2.5}$$

with $\mathbf{vecdiag}(\boldsymbol{\Sigma}) = (\text{vecdiag}(\boldsymbol{\Sigma}_1)', \text{vecdiag}(\boldsymbol{\Sigma}_2)', \ldots, \text{vecdiag}(\boldsymbol{\Sigma}_n)')'$, where $\text{vecdiag}(\boldsymbol{\Sigma}_i)$ is a K-dimensional vector with elements equal to the diagonal elements of $\boldsymbol{\Sigma}_i$. Here, $\boldsymbol{\Theta}$ in (2.1) is compounded by the elements of $\boldsymbol{\beta}$ and the non-redundant elements of $\boldsymbol{\Sigma}_i$, $i = 1, \ldots, n$.

In order to calculate (2.4), we may consider the transformation $\boldsymbol{\lambda} \to \mathbf{v}$ given by

$$\boldsymbol{\lambda} = \mathbf{exp}\{\mathbf{X}_i\boldsymbol{\beta} + \mathbf{T}_i\mathbf{v}\sqrt{2}\} \tag{2.6}$$

where $\mathbf{T}_i$ is a triangular matrix satisfying $\boldsymbol{\Sigma}_i = \mathbf{T}_i\mathbf{T}_i'$.

Using (2.5), the probability function (2.4) may be expressed as

$$P(\mathbf{Y}_i \mid \mathbf{X}_i, \boldsymbol{\beta}, \mathbf{T}_i) = \pi^{-K/2} \int_{R^K} H(\mathbf{Y}_i, \mathbf{X}_i, \boldsymbol{\beta}, \mathbf{T}_i, \mathbf{v}) \exp(-\mathbf{v}'\mathbf{v}) d\mathbf{v} \tag{2.7}$$

where

$$\begin{aligned} H(\mathbf{Y}_i, \mathbf{X}_i, \boldsymbol{\beta}, \mathbf{T}_i, \mathbf{v}) = \exp\{\mathbf{Y}_i'(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{T}_i\mathbf{v}\sqrt{2}) &- \mathbf{1}_K' \,\mathbf{exp}(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{T}_i\mathbf{v}\sqrt{2}) \\ &- \mathbf{1}_K'\mathbf{ln}[\boldsymbol{\Gamma}(\mathbf{Y}_i + \mathbf{1}_K)]\}, \end{aligned} \tag{2.8}$$

$\boldsymbol{\Gamma}$ denotes the gamma function and $\mathbf{1}_K$ is a K-dimensional vector with all elements equal to one. In this way, the probability in (2.6) can be calculated by a multivariate Hermitian integration. Details are presented in the Appendix.

The maximum likelihood estimator $\widehat{\boldsymbol{\Theta}}$ of the parameter vector under either (2.3) or (2.8) may be obtained via Newton-Raphson's method, which consists of iterations of

$$\widehat{\boldsymbol{\Theta}}^{(w)} = \widehat{\boldsymbol{\Theta}}^{(w-1)} - \left(\frac{\partial^2}{\partial\boldsymbol{\Theta}\partial\boldsymbol{\Theta}'} \log L_n(\boldsymbol{\Theta}) \mid_{\boldsymbol{\Theta}=\widehat{\boldsymbol{\Theta}}^{(w-1)}}\right)^{-1} \left(\frac{\partial}{\partial\boldsymbol{\Theta}} \log L_n(\boldsymbol{\Theta}) \mid_{\boldsymbol{\Theta}=\widehat{\boldsymbol{\Theta}}^{(w-1)}}\right) \tag{2.9}$$

w= 1, 2, ..., where $\boldsymbol{\Theta}^{(0)}$ is an initial known value and $L_n(\boldsymbol{\Theta})=\prod_{i=1}^n P(\mathbf{Y}_i)$ is the likelihood function for a sample of size n. The iterative process terminates when $\|\widehat{\boldsymbol{\Theta}}^{(w)} - \widehat{\boldsymbol{\Theta}}^{(w-1)}\| < \epsilon$, for some fixed $\epsilon > 0$.

The first and second order derivates of $\log L_n(\Theta)$ with respect to Θ are respectively

$$\frac{\partial}{\partial \Theta} \log L_n(\Theta) = \sum_{i=1}^{n} \frac{1}{P(\mathbf{Y}_i)} \frac{\partial}{\partial \Theta} P(\mathbf{Y}_i) \tag{2.10}$$

and

$$\begin{aligned} \frac{\partial^2}{\partial \Theta \partial \Theta'} \log L_n(\Theta) &= \sum_{i=1}^{n} \frac{1}{P(\mathbf{Y}_i)} \frac{\partial^2}{\partial \Theta \partial \Theta'} P(\mathbf{Y}_i) \\ &\quad - \frac{1}{P(\mathbf{Y}_i)^2} \left(\frac{\partial}{\partial \Theta} P(\mathbf{Y}_i) \right) \left(\frac{\partial}{\partial \Theta'} P(\mathbf{Y}_i) \right). \end{aligned} \tag{2.11}$$

For bivariate counts ($K = 2$) and in a particular situation with t strata (with n_j denoting the sample size in the jth stratum, $j = 1, \ldots, t$), a typical specification matrix is given by

$$\mathbf{X} = \begin{pmatrix} \mathbf{1}_{n_1} \otimes \mathbf{A}_1 \\ \mathbf{1}_{n_2} \otimes \mathbf{A}_2 \\ \cdots \\ \mathbf{1}_{n_t} \otimes \mathbf{A}_t \end{pmatrix} \tag{2.12}$$

with $\mathbf{A}_j$ denoting a $(2 \times 2t)$ matrix such that

$$\mathbf{A}_1 = \begin{bmatrix} \mathbf{I}_2 & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} \end{bmatrix},$$

$$\mathbf{A}_2 = \begin{bmatrix} \mathbf{0} & \mathbf{I}_2 & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} \end{bmatrix},$$

$$\ldots\ldots$$

$$\mathbf{A}_t = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{I}_2 \end{bmatrix}.$$

In this case $\mathbf{m} \in \{(1,0), (0,1), (1,1)\}$ and $\lambda_{\mathbf{m}i} = \gamma_{\mathbf{m}j}$ for all the sampling units belonging to the jth stratum, $j = 1, \ldots, t$. Thus, in (2.1), $\Theta = \{(\gamma_{10j}, \gamma_{01j}, \gamma_{11j})', j = 1, \ldots, t\}$. Now considering the model

$$\begin{aligned} \gamma_{10j} + \gamma_{11j} &= \lambda_{1j}^* = \exp(\beta_{1j}) \\ \gamma_{01j} + \gamma_{11j} &= \lambda_{2j}^* = \exp(\beta_{2j}) \\ \gamma_{11j} &= \theta_j^2 \end{aligned} \tag{2.13}$$

$j = 1, \ldots, t$ and writing $\boldsymbol{\beta} = (\beta_{11}, \beta_{21}, \ldots, \beta_{1t}, \beta_{2t})'$, the expression (2.2) for all sampling units belonging to the jth stratum reduces to

$$P(Y_{i1}, Y_{i2}) = |\det\left(\frac{\partial \boldsymbol{\Lambda}_j}{\partial \boldsymbol{\Theta}_j}\right)| \sum_{u=0}^{\min(Y_{i1}, Y_{i2})} \exp\left[-\mathbf{1}_2'\exp(\mathbf{X}_i\boldsymbol{\beta}) + \theta_j^2\right] \times$$
$$\times \frac{[\mathbf{e}_1'\exp(\mathbf{X}_i\boldsymbol{\beta}) - \theta_j^2]^{(Y_{i1}-u)}}{(Y_{i1}-u)!} \frac{[\mathbf{e}_2'\exp(\mathbf{X}_i\boldsymbol{\beta}) - \theta_j^2]^{(Y_{i2}-u)}}{(Y_{i2}-u)!} \frac{\theta_j^{2u}}{u!} \qquad (2.14)$$

where $\boldsymbol{\Lambda}_j = (\lambda_{1j}^*, \lambda_{2j}^*, \gamma_{11j})'$, $\mathbf{e}_1=(1,0)'$, $\mathbf{e}_2=(0,1)'$ and $\boldsymbol{\Theta}_j = [\beta_{1j}, \beta_{2j}, \theta_j]'$. Then it follows that

$$\frac{\partial}{\partial \boldsymbol{\Theta}_j} P(\mathbf{Y}_i) = \begin{bmatrix} \frac{\partial}{\partial \boldsymbol{\beta}} P(\mathbf{Y}_i) \\ \\ \frac{\partial}{\partial \theta_j} P(\mathbf{Y}_i) \end{bmatrix}, \qquad (2.15)$$

where

$$\text{i)} \quad \frac{\partial}{\partial \boldsymbol{\beta}} P(\mathbf{Y}_i) = \mathbf{X}_i'\text{diag}\left(\frac{\partial}{\partial \boldsymbol{\lambda}_j^*} P(\mathbf{Y}_i)\right) \exp(\mathbf{X}_i\boldsymbol{\beta}) \qquad (2.16)$$

with $\boldsymbol{\lambda}_j^* = (\lambda_{1j}^*, \lambda_{2j}^*)'$,

$$\frac{\partial}{\partial \lambda_{1j}^*} P(\mathbf{Y}_i) = -P(\mathbf{Y}_i) + P(\mathbf{Y}_i - \mathbf{e}_1),$$

and

$$\frac{\partial}{\partial \lambda_{2j}^*} P(\mathbf{Y}_i) = -P(\mathbf{Y}_i) + P(\mathbf{Y}_i - \mathbf{e}_2).$$

$$\text{ii)} \quad \frac{\partial}{\partial \theta_j} P(\mathbf{Y}_i) = [P(\mathbf{Y}_i) - P(\mathbf{Y}_i - \mathbf{e}_1) - P(\mathbf{Y}_i - \mathbf{e}_2) + P(\mathbf{Y}_i - \mathbf{1}_2)]\, 2\theta_j \qquad (2.17)$$

since

$$\frac{\partial}{\partial \theta_j} P(\mathbf{Y}_i) = \frac{\partial}{\partial \gamma_{11j}} P(\mathbf{Y}_i) \frac{\partial \gamma_{11j}}{\partial \theta_j}$$

where

$$\frac{\partial}{\partial \gamma_{11j}} P(\mathbf{Y}_i) = [P(\mathbf{Y}_i) - P(\mathbf{Y}_i - \mathbf{e}_1) - P(\mathbf{Y}_i - \mathbf{e}_2) + P(\mathbf{Y}_i - \mathbf{1}_2)]. \qquad (2.18)$$

Also,

$$\frac{\partial^2}{\partial \Theta_j' \partial \Theta_j} P(\mathbf{Y}_i) = \begin{bmatrix} \frac{\partial^2}{\partial \beta' \partial \beta} P(\mathbf{Y}_i) & \frac{\partial^2}{\partial \beta' \partial \theta_j} P(\mathbf{Y}_i) \\ \frac{\partial^2}{\partial \theta_j \partial \beta} P(\mathbf{Y}_i) & \frac{\partial^2}{\partial \theta_j^2} P(\mathbf{Y}_i) \end{bmatrix} \tag{2.19}$$

where

$$\text{i)} \qquad \frac{\partial^2}{\partial \beta' \partial \beta} P(\mathbf{Y}_i) \quad = \quad \mathbf{X}_i' \mathbf{L} \mathbf{X}_i \tag{2.20}$$

with

$$\mathbf{L} = \operatorname{diag}\left[\frac{\partial}{\partial \boldsymbol{\lambda}_j^*} P(\mathbf{Y}_i)\right] \operatorname{diag}\left[\exp(\mathbf{X}_i\beta)\right] + \frac{\partial^2}{\partial \boldsymbol{\lambda}_j^{*\prime} \partial \boldsymbol{\lambda}_j^*} P(\mathbf{Y}_i) \left[\exp(\mathbf{X}_i\beta)\right] \left[\exp(\mathbf{X}_i\beta)\right]',$$

$$\text{ii)} \qquad \frac{\partial^2}{\partial \beta' \partial \theta_j} P(\mathbf{Y}_i) = 2\theta_j \mathbf{S} \tag{2.21}$$

with

$$\mathbf{S} = \left[\frac{\partial}{\partial \beta'} P(\mathbf{Y}_i) - \frac{\partial}{\partial \beta'} P(\mathbf{Y}_i - \mathbf{e}_1) - \frac{\partial}{\partial \beta'} P(\mathbf{Y}_i - \mathbf{e}_2) + \frac{\partial}{\partial \beta'} P(\mathbf{Y}_i - \mathbf{1}_2)\right]$$

where

$$\frac{\partial}{\partial \beta} P(\mathbf{Y}_i - \mathbf{e}_1), \frac{\partial}{\partial \beta} P(\mathbf{Y}_i - \mathbf{e}_2) \text{ and } \frac{\partial}{\partial \beta} P(\mathbf{Y}_i - \mathbf{1}_2)$$

can be obtained by applying (2.16) conveniently,

$$\text{iii)} \qquad \frac{\partial^2}{\partial \theta_j^2} P(\mathbf{Y}_i) = 4\theta_j^2 \mathbf{Q} + 2\left[P(\mathbf{Y}_i) - P(\mathbf{Y}_i - \mathbf{e}_1) - P(\mathbf{Y}_i - \mathbf{e}_2) + P(\mathbf{Y}_i - \mathbf{1}_2)\right] \tag{2.22}$$

with

$$\mathbf{Q} = \left[\frac{\partial}{\partial \gamma_{11j}} P(\mathbf{Y}_i) - \frac{\partial}{\partial \gamma_{11j}} P(\mathbf{Y}_i - \mathbf{e}_1) - \frac{\partial}{\partial \gamma_{11j}} P(\mathbf{Y}_i - \mathbf{e}_2) + \frac{\partial}{\partial \gamma_{11j}} P(\mathbf{Y}_i - \mathbf{1}_2)\right]$$

where

$$\frac{\partial}{\partial \gamma_{11j}} P(\mathbf{Y}_i - \mathbf{e}_1), \frac{\partial}{\partial \gamma_{11j}} P(\mathbf{Y}_i - \mathbf{e}_2) \text{ and } \frac{\partial}{\partial \gamma_{11j}} P(\mathbf{Y}_i - \mathbf{1}_2)$$

can be obtained by applying (2.18) conveniently.

Now, considering the Poisson log-normal setup defined by (2.6), and assuming further that $\mathbf{T}_1 = \mathbf{T}_2 = \ldots = \mathbf{T}_n = \mathbf{T}$ (which implies $\mathbf{\Sigma}_1=\mathbf{\Sigma}_2=\ldots=\mathbf{\Sigma}_n=\mathbf{\Sigma}$), we may write

$$\frac{\partial}{\partial\Theta}P(\mathbf{Y}_i) = \begin{bmatrix} \frac{\partial}{\partial\beta}P(\mathbf{Y}_i) \\ \\ \frac{\partial}{\partial\mathbf{T}^*}P(\mathbf{Y}_i) \end{bmatrix}, \tag{2.23}$$

with $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \ldots, \boldsymbol{\beta}_t')'$ and $\mathbf{T}^* = (\tau_{11}, \tau_{21}, \tau_{22})'$ where

$$\text{i)} \qquad \frac{\partial}{\partial\beta}P(\mathbf{Y}_i) = \mathbf{X}_i'\left[\mathbf{Y}_i P(\mathbf{Y}_i) - (\mathbf{Y}_i + \mathbf{1}_2)\#\mathbf{A}\right] \tag{2.24}$$

with

$$\mathbf{A} = \begin{bmatrix} P(\mathbf{Y}_i + \mathbf{e}_1) \\ \\ P(\mathbf{Y}_i + \mathbf{e}_2) \end{bmatrix},$$

and # denoting the Hadamard product (see Searle, 1982).

$$\text{ii)} \qquad \frac{\partial}{\partial\mathbf{T}^*}P(\mathbf{Y}_i) = \begin{bmatrix} \mathbf{e}_1'\mathbf{T}'\mathbf{M}\mathbf{e}_1 & \mathbf{e}_1'\mathbf{T}'\mathbf{M}'\mathbf{e}_2 & \mathbf{e}_2'\mathbf{T}'\mathbf{M}\mathbf{e}_2 \end{bmatrix}' \tag{2.25}$$

with

$$\mathbf{M} = \mathbf{Y}_i\mathbf{Y}_i'P(\mathbf{Y}_i) - \mathbf{Y}_i(\mathbf{Y}_i' + \mathbf{1}_2')\#\mathbf{A}' - \mathbf{A}\#(\mathbf{Y}_i + \mathbf{1}_2)\mathbf{Y}_i' - \text{diag}\left[(\mathbf{Y}_i + \mathbf{1}_2)\#\mathbf{A}\right] + \mathbf{B} \tag{2.26}$$

and

$$\mathbf{B} = \begin{bmatrix} (Y_{i1}+1)(Y_{i1}+2)P(\mathbf{Y}_i + 2\mathbf{e}_1) & (Y_{i1}+1)(Y_{i2}+1)P(\mathbf{Y}_i + \mathbf{1}_2) \\ \cdot & \\ (Y_{i1}+1)(Y_{i2}+1)P(\mathbf{Y}_i + \mathbf{1}_2) & (Y_{i2}+1)(Y_{i2}+2)P(\mathbf{Y}_i + 2\mathbf{e}_2) \end{bmatrix}.$$

Also,

$$\frac{\partial^2}{\partial\Theta'\partial\Theta}P(\mathbf{Y}_i) = \begin{bmatrix} \frac{\partial^2}{\partial\beta'\partial\beta}P(\mathbf{Y}_i) & \frac{\partial^2}{\partial\beta'\partial\mathbf{T}^*}P(\mathbf{Y}_i) \\ \\ \frac{\partial^2}{\partial\mathbf{T}^{*\prime}\partial\beta}P(\mathbf{Y}_i) & \frac{\partial^2}{\partial\mathbf{T}^{*\prime}\partial\mathbf{T}^*}P(\mathbf{Y}_i) \end{bmatrix} \tag{2.27}$$

where

$$\begin{aligned} \text{i)} \qquad \frac{\partial^2}{\partial\beta'\partial\beta}P(\mathbf{Y}) &= \mathbf{X}_i'\{\mathbf{Y}_i\mathbf{Y}_i'P(\mathbf{Y}_i) - \mathbf{Y}_i(\mathbf{Y}_i' + \mathbf{1}_2')\#\mathbf{A}' - \mathbf{A}\#(\mathbf{Y}_i + \mathbf{1}_2)\mathbf{Y}_i' \\ &\quad -\text{diag}[(\mathbf{Y}_i + \mathbf{1}_2)\#\mathbf{A}] + \mathbf{B}\}\mathbf{X}_i. \end{aligned} \tag{2.28}$$

$$\text{ii)} \quad \frac{\partial^2}{\partial\beta'\partial\mathbf{T}^*}P(\mathbf{Y}_i) = \begin{bmatrix} \tau_{11}\mathbf{C} + \tau_{21}\mathbf{D} \\ \tau_{11}\mathbf{D} + \tau_{21}\mathbf{E} \\ \tau_{22}\mathbf{E} \end{bmatrix} \tag{2.29}$$

where

$$\begin{aligned} \mathbf{C} &= Y_{i1}^2\frac{\partial}{\partial\beta'}P(\mathbf{Y}_i) - (2Y_{i1}+1)(Y_{i1}+1)\frac{\partial}{\partial\beta'}P(\mathbf{Y}_i+\mathbf{e}_1) \\ &\quad +(Y_{i1}+1)(Y_{i1}+2)\frac{\partial}{\partial\beta'}P(\mathbf{Y}_i+2\mathbf{e}_1), \end{aligned}$$

$$\begin{aligned} \mathbf{D} &= Y_{i1}Y_{i2}\frac{\partial}{\partial\beta'}P(\mathbf{Y}_i) - Y_{i1}(Y_{i2}+1)\frac{\partial}{\partial\beta'}P(\mathbf{Y}_i+\mathbf{e}_2) - Y_{i2}(Y_{i1}+1)\frac{\partial}{\partial\beta'}P(\mathbf{Y}_i+\mathbf{e}_1) \\ &\quad +(Y_{i1}+1)(Y_{i2}+1)\frac{\partial}{\partial\beta'}P(\mathbf{Y}_i+\mathbf{1}_2). \end{aligned}$$

and

$$\begin{aligned} \mathbf{E} &= Y_{i2}^2\frac{\partial}{\partial\beta'}P(\mathbf{Y}_i) - (2Y_{i2}+1)(Y_{i2}+1)\frac{\partial}{\partial\beta'}P(\mathbf{Y}_i+\mathbf{e}_2) \\ &\quad +(Y_{i2}+1)(Y_{i2}+2)\frac{\partial}{\partial\beta'}P(\mathbf{Y}_i+2\mathbf{e}_2). \end{aligned}$$

Here

$$\frac{\partial}{\partial\beta'}P(\mathbf{Y}_i+\mathbf{e}_1),\ \frac{\partial}{\partial\beta'}P(\mathbf{Y}_i+2\mathbf{e}_1),\ \frac{\partial}{\partial\beta'}P(\mathbf{Y}_i+\mathbf{e}_2),\ \frac{\partial}{\partial\beta'}P(\mathbf{Y}_i+2\mathbf{e}_2) \text{ and } \frac{\partial}{\partial\beta'}P(\mathbf{Y}_i+\mathbf{1}_2)$$

can be obtained by applying (2.24) conveniently.

$$\text{iii)} \quad \frac{\partial^2}{\partial\mathbf{T}^{*\prime}\partial\mathbf{T}^*}P(\mathbf{Y}_i) = \begin{bmatrix} \tau_{11}\mathbf{F} + \tau_{21}\mathbf{G} \\ \tau_{11}\mathbf{G} + \tau_{21}\mathbf{H} \\ \tau_{22}\mathbf{H} \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1'\mathbf{M} & 0 \\ \mathbf{e}_2'\mathbf{M} & 0 \\ \mathbf{0}_2' & \mathbf{M}_{2.2} \end{bmatrix} \tag{2.30}$$

where $\mathbf{M}_{i,j}$ denotes the (i,j)th element of the matrix $\mathbf{M}$ in (2.26), $\mathbf{0}_2$ is a bidimensional vector with all elements equal to zero.

$$\begin{aligned} \mathbf{F} &= Y_{i1}^2\frac{\partial}{\partial\mathbf{T}^{*\prime}}P(\mathbf{Y}_i) - (2Y_{i1}+1)(Y_{i1}+1)\frac{\partial}{\partial\mathbf{T}^{*\prime}}P(\mathbf{Y}_i+\mathbf{e}_1) \\ &\quad +(Y_{i1}+1)(Y_{i1}+2)\frac{\partial}{\partial\mathbf{T}^{*\prime}}P(\mathbf{Y}_i+2\mathbf{e}_1). \end{aligned}$$

$$
\begin{aligned}
\mathbf{G} &= Y_{i1}Y_{i2}\frac{\partial}{\partial \mathbf{T}^{*\prime}}P(\mathbf{Y}_i) - Y_{i1}(Y_{i2}+1)\frac{\partial}{\partial \mathbf{T}^{*\prime}}P(\mathbf{Y}_i+\mathbf{e}_2) \\
&\quad - Y_{i2}(Y_{i1}+1)\frac{\partial}{\partial \mathbf{T}^{*\prime}}P(\mathbf{Y}_i+\mathbf{e}_1) \\
&\quad + (Y_{i1}+1)(Y_{i2}+1)\frac{\partial}{\partial \mathbf{T}^{*\prime}}P(\mathbf{Y}_i+\mathbf{1}_2).
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbf{H} &= Y_{i2}^2\frac{\partial}{\partial \mathbf{T}^{*\prime}}P(\mathbf{Y}_i) - (2Y_{i2}+1)(Y_{i2}+1)\frac{\partial}{\partial \mathbf{T}^{*\prime}}P(\mathbf{Y}_i+\mathbf{e}_2) \\
&\quad + (Y_{i2}+1)(Y_{i2}+2)\frac{\partial}{\partial \mathbf{T}^{*\prime}}P(\mathbf{Y}_i+2\mathbf{e}_2).
\end{aligned}
$$

Again,

$$
\frac{\partial}{\partial \mathbf{T}^{*\prime}}P(\mathbf{Y}_i+\mathbf{e}_1),\ \frac{\partial}{\partial \mathbf{T}^{*\prime}}P(\mathbf{Y}_i+2\mathbf{e}_1),\ \frac{\partial}{\partial \mathbf{T}^{*\prime}}P(\mathbf{Y}_i+\mathbf{e}_2),\ \frac{\partial}{\partial \mathbf{T}^{*\prime}}P(\mathbf{Y}_i+2\mathbf{e}_2) \text{ and } \frac{\partial}{\partial \mathbf{T}^{*\prime}}P(\mathbf{Y}_i+\mathbf{1}_2)
$$

may be obtained by applying (2.25) conveniently.

Under some regularity conditions, asymptotically, $\sqrt{n}(\widehat{\Theta}-\Theta)$ follows a Normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{I}_{\Theta}^{-1}$, where $\mathbf{I}_{\Theta}$ is Fisher's information matrix, i.e.

$$
\mathbf{I}_{\Theta} = E\left[\frac{\partial}{\partial\Theta}\log P(\mathbf{Y},\Theta)\right]\left[\frac{\partial}{\partial\Theta}\log P(\mathbf{Y},\Theta)\right]'. \tag{2.31}
$$

Some hypotheses of interest can be expressed as

$$
H_0 : \mathbf{R}\Theta = \mathbf{0} \tag{2.32}
$$

where $\mathbf{R}$ is a $(q\times r)$ matrix of constants. They can be tested by Wald's test statistic

$$
\mathbf{W} = \widehat{\Theta}'\mathbf{R}'\left[\mathbf{R}\widehat{Var}(\widehat{\Theta})\mathbf{R}'\right]^{-1}\mathbf{R}\widehat{\Theta} \tag{2.33}
$$

where $\widehat{Var}(\widehat{\Theta})$ is a consistent estimator of $Var(\widehat{\Theta})$ given by

$$
\widehat{Var}(\widehat{\Theta}) = \left\{\sum_{i=1}^{n}\left[\frac{\partial}{\partial\Theta}\log P(\mathbf{Y}_i,\widehat{\Theta})\right]\left[\frac{\partial}{\partial\Theta}\log P(\mathbf{Y}_i,=\widehat{\Theta})\right]'\right\}^{-1};
$$

under (2.32) $\mathbf{W}$ follows asymptotically a $\chi^2_{(q)}$ distribution where q is the number of rows of $\mathbf{R}$.

3 Numerical example

In this section, we apply the methods discussed above to the example mentioned in the introduction. We are interested in comparing the expected defect rates between machines within manufacturers as well as between manufacturers. The p-values corresponding to Pearson's goodness-of-fit test of three probability models are summarized in Table 3.1.

Any of the three distributions may be considered as a possible probabilistic model to the machine A_1 data. Otherwise, two of them may be considered (discard only the independent Poisson distributions) to the machine A_2 data set. Similar analyses can be carried out for the data sets corresponding to machines B_1 and B_2 suggesting that the three proposed distributions may not be ruled out as underlying probability models.

For illustrative purposes, we consider both the bivariate Poisson log-normal and the bivariate Poisson models for a regression analysis. The parameter estimates were calculated via Newton Raphson's method detailed in Section 2 and implemented using SAS (1988) IML procedure. The same procedure was used to compute Wald's statistics of the form (2.33). Goodness of fit for the different regression models are compared by Akaike's Information Criteria (AIC). The reader is referred to Kendall, Stuart and Ord (1983) for details.

In this context, the indices from $i = 1$ to $i = 130$ correspond the observations for the machine A_1; those from $i = 131$ to $i = 230$ correspond to A_2; those from $i = 231$ to $i = 330$ correspond to B_1 and those from $i = 331$ to $i = 430$ correspond to B_2.

For these data, assuming first underlying bivariate Poisson distributions, a regression model of the form (2.3) may have specification matrix $\mathbf{X}$ with

$$\mathbf{X}_1 = \ldots = \mathbf{X}_{130} = \mathbf{A}_1, \quad \mathbf{X}_{131} = \ldots = \mathbf{X}_{230} = \mathbf{A}_2,$$

$$\mathbf{X}_{231} = \ldots = \mathbf{X}_{330} = \mathbf{B}_1, \quad \mathbf{X}_{331} = \ldots = \mathbf{X}_{430} = \mathbf{B}_2,$$

where

Table 3.1: p-values for the goodness-of-fit tests.

Machine	Bivariate Poisson	Bivariate Poisson Log-Normal	Independent Poisson
A_1	0.8904	0.1998	0.2821
A_2	0.1050	0.1654	0.0328
B_1	0.2890	0.0170	0.6092
B_2	0.7231	0.4153	0.8274

$$\mathbf{A}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \mathbf{A}_2 = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$\mathbf{B}_1 = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}, \mathbf{B}_2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix};$$

here $\boldsymbol{\beta} = (\beta_{1A_1}, \beta_{2A_1}, \ldots, \beta_{1B_2}, \beta_{2B_2})'$ implying different expected defect rates for each machine. Also, using the reparametrization $\theta_i = \sqrt{\lambda_{11i}}$, $i = 1, \ldots, 430$, we have

$$\begin{aligned} \theta_1 = \ldots = \theta_{130} = \theta_{A_1}, \quad \theta_{131} = \ldots = \theta_{230} = \theta_{A_2}, \\ \theta_{231} = \ldots = \theta_{330} = \theta_{B_1}, \quad \theta_{331} = \ldots = \theta_{430} = \theta_{B_2}, \end{aligned} \tag{3.1}$$

which implies a different correlation parameter for each of the four machines. The resulting parameter estimates with their estimated standard errors are presented in Table 3.2.

Table 3.2: Parameter estimates and estimated standard errors under model (3.1).

	Machine			
Parameter	A_1	A_2	B_1	B_2
β_1	−0.7244	−0.4943	−1.7720	−1.8326
	(0.1260)	(0.1280)	(0.2425)	(0.2500)
β_2	0.6737	0.7275	0.5068	0.2624
	(0.0626)	(0.0695)	(0.0774)	(0.0877)
θ	0.4565	0.6618	0.2081	0.0000
	(0.0856)	(0.0721)	(0.1437)	(0.2404)

Hypotheses concerning the equality of both the expected defect rates and of the association between the frequencies of the two types of defects among the four machines can be expressed respectively as

$$H_0 : \boldsymbol{\beta}_{A_1} = \ldots = \boldsymbol{\beta}_{B_2} \tag{3.2}$$

and

$$H_0 : \theta_{A1} = \theta_{A2} = \theta_{B1} = \theta_{B2}. \tag{3.3}$$

The corresponding Wald p-values are respectively $p < 0.0001$ and $p = 0.0033$, indicating strong evidence against them. Similar hypotheses for the machines within each manufacturer may be expressed as

$$H_0 : \begin{cases} \beta_{A1} = \beta_{A2} \\ \beta_{B1} = \beta_{B2} \end{cases} \tag{3.4}$$

and

$$H_0 : \begin{cases} \theta_{A1} = \theta_{A2} \\ \theta_{B1} = \theta_{B2}. \end{cases} \tag{3.5}$$

The corresponding p-values are respectively $p = 0.1959$ and $p = 0.1421$, suggesting that the proposed model can be reduced by taking

$$\mathbf{X}_1 = \ldots = \mathbf{X}_{230} = \mathbf{A},$$
$$\mathbf{X}_{231} = \ldots = \mathbf{X}_{430} = \mathbf{B}$$

where

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \text{ and } \mathbf{B} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

and

$$\begin{aligned} \theta_1 = \ldots = \theta_{230} = \theta_A, \\ \theta_{231} = \ldots = \theta_{430} = \theta_B. \end{aligned} \tag{3.6}$$

For this last model, both the expected defect rates and the association between the frequencies of the two types of defect are homogenous for the different machines from each manufacturer.

The parameter estimates for the reduced model with their standard errors are presented in Table 3.3. The expected defect rates given by $\exp(\beta_j)$, $j = A, B$ and the correlation coefficients between the frequencies of the two types of defects given by

$$\frac{\theta_j^2}{\sqrt{\exp(\mathbf{1}_2'\beta_j)}}, \; j = A, B,$$

under model (3.6) are reproduced in Table 3.4.

The intensity of the association between the frequencies of the two types of defects may also be investigated. The hypotheses $H_0 : \theta_j = 0$, $j = A, B$, correspond to null association. The p-values for the machines from manufacturers A and B are respectively $p < 0.0001$ and $p = 0.2557$ suggesting a positive association between the frequencies of the two types of defects for the machines from the manufacturer A and a null association for the machines from the manufacturer B.

Table 3.3: Parameter estimates under model (3.6).

Manufacturer	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\theta}$
A	−0.6178	0.6975	0.5528
	(0.0893)	(0.0468)	(0.0549)
B	−1.8018	0.39204	0.1128
	(0.1765)	(0.0580)	(0.1738)

Table 3.4: Estimates of the expected defect rates and correlation coefficients (standard errors within parentheses) under model (3.6).

Manufacturer	Type 1 defect rate	Type 2 defect rate	Correlation coeficients
A	0.5391	2.0087	0.2937
	(0.0484)	(0.0934)	(0.0519)
B	0.1650	1.4800	0.0257
	(0.0287)	(0.0860)	(0.0228)

The data set was reanalyzed under a bivariate Poisson log-normal regression model. The model considers a specification matrix similar to that of the model (3.1) and

$$\begin{aligned} &\boldsymbol{\Sigma}_1 = \ldots = \boldsymbol{\Sigma}_{130} = \boldsymbol{\Sigma}_{A_1}, \quad \boldsymbol{\Sigma}_{131} = \ldots = \boldsymbol{\Sigma}_{230} = \boldsymbol{\Sigma}_{A_2}, \\ &\boldsymbol{\Sigma}_{231} = \ldots = \boldsymbol{\Sigma}_{330} = \boldsymbol{\Sigma}_{B_1}, \quad \boldsymbol{\Sigma}_{331} = \ldots = \boldsymbol{\Sigma}_{430} = \boldsymbol{\Sigma}_{B_2}. \end{aligned} \tag{3.7}$$

The corresponding paramater estimates with estimated standard errors (within parentheses) are indicated in Table 3.5.

We now consider a reduced model which incorporates the hypothesis

$$H_0 : \begin{cases} \boldsymbol{\Sigma}_{A_1} = \boldsymbol{\Sigma}_{A_2} = \boldsymbol{\Sigma}_A \\ \boldsymbol{\Sigma}_{B_1} = \boldsymbol{\Sigma}_{B_2} = \boldsymbol{\Sigma}_B. \end{cases} \tag{3.8}$$

The Wald statistics p-value for a test of (3.8) is $p = 0.7417$. A further reduction implied by

$$H_0 : \boldsymbol{\Sigma}_A = \boldsymbol{\Sigma}_B \tag{3.9}$$

Table 3.5: Estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ with their estimated standard errors under Model (3.7).

	Machines			
Estimates	A_1	A_2	B_1	B_2
β_1	−0.9630 (0.1969)	−0.7260 (0.1762)	−2.2137 (0.5721)	−1.8326 (0.2500)
β_2	0.6474 (0.0694)	0.6302 (0.0925)	0.5055 (0.0781)	0.2624 (0.0877)
σ_{11}	0.4744 (0.3057)	0.4954 (0.2093)	0.8892 (1.1208)	0.0000 ($< 10^{-4}$)
σ_{12}	0.1577 (0.0894)	0.3214 (0.0907)	0.0481 (0.1315)	0.0000 ($< 10^{-4}$)
σ_{22}	0.0524 (0.0455)	0.2085 (0.0831)	0.0026 (0.0136)	0.0000 ($< 10^{-4}$)

is not supported ($p = 0.0547$). The hypotheses of equal expected defect rates and equal association between the frequencies of the two types of defects can be expressed respectively as

$$H_0 : \boldsymbol{\beta}_{A1} + \frac{1}{2}\text{vecdiag}(\boldsymbol{\Sigma}_A) = \ldots = \boldsymbol{\beta}_{B2} + \frac{1}{2}\text{vecdiag}(\boldsymbol{\Sigma}_B) \tag{3.10}$$

and

$$H_0 : \begin{cases} \sigma_{12A} = \sigma_{12B} \\ \mathbf{1}_2' \left[\boldsymbol{\beta}_{A1} + \frac{1}{2}\text{vecdiag}(\boldsymbol{\Sigma}_A)\right] = \ldots = \mathbf{1}_2' \left[\boldsymbol{\beta}_{B2} + \frac{1}{2}\text{vecdiag}(\boldsymbol{\Sigma}_B)\right], \end{cases} \tag{3.11}$$

where σ_{12j} is the off-diagonal element of $\boldsymbol{\Sigma}_j, j = A, B$. Since the corresponding p-values are < 0.0001, it is unreasonable to confirm either hypotheses. We may also consider

$$H_0 : \boldsymbol{\beta}_{j1} = \boldsymbol{\beta}_{j2} \tag{3.12}$$

$j = A, B$, which corresponds to equality of expected defect rates for the two machines within each manufacturer. Here $p = 0.2287$, suggesting that such reduction is reasonable. For the model under investigation, the homogeneous association hypothesis is

$$H_0 : \mathbf{1}_2'\boldsymbol{\beta}_{j1} = \mathbf{1}_2'\boldsymbol{\beta}_{j2} \tag{3.13}$$

$j = A, B$. However under model (3.8), the hypothesis (3.13) implies (3.14). The test of hypothesis (3.13) also suggests an homogeneous association between the frequencies of the two types of defect in the different machines from each manufacturer. Thus, the proposed model can be reduced by considering the same specification matrix as in model (3.6) while maintaining

$$\begin{aligned} \Sigma_1 = \ldots = \Sigma_{230} = \Sigma_A, \\ \Sigma_{231} = \ldots = \Sigma_{430} = \Sigma_B. \end{aligned} \tag{3.14}$$

The corresponding parameter estimates and their estimated standard errors are presented in Table 3.6. For this last model both the expected defect rates and the association between the frequencies of the two types of defect are homogenous for the different machines from each manufacturer.

Table 3.6: Parameter estimates with their estimated standard errors under model (3.15).

Manufacturer	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_{11}$	$\hat{\sigma}_{21}$	$\hat{\sigma}_{22}$
A	−0.8454	0.6417	0.4572	0.2265	0.1122
	(0.1314)	(0.0560)	(0.1912)	(0.0715)	(0.1260)
B	−1.8579	0.3919	0.1122	0.0053	0.0002
	(0.3382)	(0.0582)	(0.5782)	(0.0553)	(0.0048)

Tests of the hypotheses $H_0 : \sigma_{12j} = 0$, $j = A, B$, produce respectively $p = 0.0008$ and $p = 0.4622$ suggesting a positive association between the frequencies of the two types of defect for the machines from the manufacturer A and a null association for the machines from the manufacturer B confirming our previous results. In Table 3.7, we present the expected defect rates

$$\alpha_j = \exp\left(\beta_j + \frac{1}{2}\text{vecdiag}(\Sigma_j)\right),$$

and the correlation coefficients

$$\frac{\exp(\sigma_{12j}) - 1}{\{[\exp(\sigma_{11j}) - 1 + (\mathbf{e}_1'\alpha_j)^{-1}][\exp(\sigma_{22j}) - 1 + (\mathbf{e}_2'\alpha_j)^{-1}]\}^{1/2}},$$

$j = A, B$, with their estimated standard errors under model (3.15).

Both types of models considered here suggest that the expected defect rates are equal for different machines from each manufacturer and also that the corresponding association between the frequencies of the two types of defects are homogenous. In both cases, a

Table 3.7: Expected defect rates and correlation coefficients with their estimated standard errors estimates under model (3.15).

Manufacturer	Type 1 defect rate	Type 2 defect rate	Correlation coeficients
A	0.5397	2.0093	0.2076
	(0.0562)	(0.1172)	(0.0422)
B	0.1650	1.4800	0.0026
	(0.0290)	(0.0938)	(0.0284)

positive association between the frequencies of the two types of defect is suggested for the machines from the manufacturer A.

The goodness of fit from the different regression models are compared by Akaike's Information Criterion (**AIC**) [see Kendall, Stuart and Ord, 1983]. In Table 3.8, the **AIC** for the different regression models are displayed. According to this criterion the best model is the bivariate Poisson regression model (3.6).

Table 3.8: Akaike's Information Criterion for the regression models.

Regression model	Model	Number of parameters	$\log L_n(\Theta)$	AIC
bivariate Poisson Log-Normal	(3.7)	20	−1003.19	2046.38
	(3.8)	14	−1004.93	2037.87
	(3.15)	10	−1007.78	2035.55
bivariate Poisson	(3.1)	12	−1020.91	2065.83
	(3.6)	06	−1005.33	2022.66

4 Discussion

In the example analysed above, there is no strong evidence that a bivariate Poisson log-normal distribution is more adequate than a bivariate Poisson distribution. For illustrative purposes, we consider simulated counts under a bivariate Poisson log-normal distribution with a covariance parameter equal to -0.0952. The data set is summarized in Table 4.1 and the underlying parameter estimates under different distributions are displayed in Table 4.2 along with their estimated standard errors.

Table 4.1: Observed frequencies of a simulated sample.

	Y_1						
Y_2	0	1	2	3	4	5	Total
0	3	9	3	2	1	1	19
1	7	7	2	1	1	.	18
2	2	3	.	.	1	.	6
3	2	1	1	.	.	.	4
4	2	.	1	.	.	.	3
Total	16	20	7	3	3	1	50

The simulated sample has a negative correlation coefficient (-0.22) and the logarithms of the likelihood are -141.62 and -144.31 respectively, for a bivariate Poisson log-normal underlying distribution and a bivariate Poisson underlying distribution. The AIC's are respectively 293.25 and 294.63. The estimates of the rates under different probabilistic models are very similar although they have lower standard errors under a bivariate Poisson distribution. On the other hand, the estimates of the covariance and variance under a bivariate Poisson log-normal distribution have smaller discrepancies when they are compared to the corresponding sample values. Under a bivariate Poisson distribution, the estimate of the covariance is null. In this case, it seems that a bivariate Poisson log-normal distribution may be more appropriate.

Appendix

Multivariate Hermitian integration of (2.6).

Salzer, Zucker and Capuano (1952) show that integrals of the form

$$\int_{-\infty}^{+\infty} f(v)\exp\{-v^2\}dv$$

may be approximated by

Table 4.2: Parameter estimates and estimated standard errors for the simulated sample.

Parameters	Observed values	Estimates under a bivariate Poisson Log-Normal distribution	Estimates under a bivariate Poisson distribution
$E(Y_1)=1.0000$	1.2000	1.2003 (0.1776)	1.2000 (0.1549)
$E(Y_2)=1.0000$	1.0800	1.0815 (0.2353)	1.0800 (0.1470)
$Var(Y_1)=1.2214$	1.5510	1.5746 (0.4621)	1.2000 (0.1549)
$Cov(Y_1,Y_2)=-0.0952$	−0.3243	−0.2799 (0.5341)	0.0000 ($< 10^{-4}$)
$Var(Y_2)=1.2214$	1.3812	1.4217 (0.3043)	1.0800 (0.1470)

$$\int_{-\infty}^{+\infty} f(v)\exp\{-v^2\}dv = \sum_{i=1}^{n} \alpha_i^{(n)} f(v_i^{(n)}) + R_n$$

where α_i are Christoffel numbers, v_i is the ith zero of the Hermite polynomial and

$$R_n = \frac{\pi^{1/2} f(\epsilon)^{(2n)}}{2^n(2n)(2n-1)\dots(n+2)(n+1)}$$

for some ϵ, $-\infty < \epsilon < \infty$. Thus, multiple integrals as the expression (2.6) may be approximated by

$$\int_{R^K} H(\mathbf{Y},\mathbf{X},\boldsymbol{\beta},\mathbf{T},\mathbf{v})\exp\{(-\mathbf{v}'\mathbf{v})\}dv$$
$$= \sum_{iK=1}^{n} \dots \sum_{i1=1}^{n} \alpha_{iK} \dots \alpha_{i1} H(\mathbf{Y},\mathbf{X},\boldsymbol{\beta},\mathbf{T}, v_{i1}^{(1)}, \dots, v_{iK}^{(K)})$$

where α_{ij} are Christoffel numbers and $v_{ij}^{(j)}$ is the ith zero of the Hermite polynomial, $j = 1, \dots, K$.

Acknowledgements

The authors are grateful to the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for financial support.

References

[1] Aitchison, J. and Ho, C.H. (1989). The multivariate Poisson log-normal distribution. *Biometrika*, **76**,643-53.

[2] Ho, L.L. (1995). *Análise de contagens multivariadas* Doctoral dissertation. Escola Politécnica. Universidade de São Paulo, Brazil.

[3] Holgate, P. (1964). Estimation for the bivariate Poisson distribution. *Biometrika*, **51**, 241-5.

[4] Kawamura, K. (1979). The structure of multivariate Poisson distribution. *Kodai Mathematical Journal*, **2**, 337-45.

[5] Kendall, M.G., Stuart, A. and Ord, J.K. (1983). *The advanced theory of statistics: design and analysis, and time-series*, v.3 4. ed. London: Charles Griffin.

[6] Kocherlakota, S. (1988). On the compounded bivariate distribution: a unified treatment. *Annals of Institute of Mathematical Statistics*, **40**, 61-76.

[7] Salzer, H.E.; Zucker, R.; Capuano, R. (1952). Table of the zeroes and weight factors of the first twenty Hermite polynomials. *Journal of Research National Bureau Standards*, **48**, 111-6.

[8] SAS Institute Inc. (1988). *SAS/IML User's Guide*, Release 6.03 Edition. Cary, NC: SAS Institute Inc.

[9] Searle, S.R. (1982). *Matrix algebra useful for statistics.* New York: John Wiley.

ÚLTIMOS RELATÓRIOS TÉCNICOS PUBLICADOS

9601 - MENTZ, R.P.; MORETTIN, P.A. and TOLOI, C.M.C. Bias correction for estimators of the residual variance in the ARMA (1,1) Model. São Paulo, IME-USP, 1996. 21p. (RT-MAE-9601)

9602 - SINGER, J.M.; PERES, C.A.; HARLE, C.E. Performance of wald's test for the hardy-weiberg equilibrium with fixed sample sizes. São Paulo, IME-USP, 1996.15p. (RT-MAE-9602).

9603 - SINGER, J.M. and E. SUYAMA, E. Dispersion structure, Hierarchical models, Random effects models, Repeated measures. São Paulo, IME-USP, 1996. 21p. (RT-MAE-9603).

9604 - LIMA, A.C.P. and SEN, P.K. A Matrix-Valued Counting Process with First-Order Interactive Intensities. São Paulo, IME-USP, 1996, 25p. (RT-MAE-9604)

9605 - BOTTER, D.A. AND SINGER, J.M. Experimentos com Intercâmbio de Dois Tratamentos e Dois Períodos: Estratégias para Análise e Aspectos Computacionais, São Paulo, IME-USP, 1996. 18p. (RT-MAE-9605)

9606- MORETTIN, P.A. From Fourier to Wavelet Analysis of Time Series. São Paulo, IME-USP, 1996. 11p. (RT-MAE-9606)

The complete list of Relatórios do Departamento de Estatística, IME-USP, will be sent upon request.

Departamento de Estatística
IME-USP
Caixa Postal 66.281
05389-970 - São Paulo, Brasil