

Universidade Federal de São Carlos
Centro de Ciências Exatas e de Tecnologia
Departamento de Estatística

Inferência em Distribuições Discretas Bivariadas

VERÔNICA AMPARO QUISPE CHIRE

UFSCar - São Carlos/SP
Dezembro/2013

Universidade Federal de São Carlos
Centro de Ciências Exatas e de Tecnologia
Departamento de Estatística

Inferência em Distribuições Discretas Bivariadas

VERÔNICA AMPARO QUISPE CHIRE

ORIENTADOR: PROF. DR. LUIS ERNESTO BUENO SALASAR
COORIENTADORA: PROF. DR. VERA LUCIA D. TOMAZELLA

Dissertação apresentada ao Departamento de Estatística da Universidade Federal de São Carlos - DEs/UFSCar como parte dos requisitos para obtenção do título de mestre em Estatística.

**Ficha catalográfica elaborada pelo DePT da
Biblioteca Comunitária da UFSCar**

Q8iq

Quispe Chire, Verônica Amparo.
Inferência em distribuições discretas bivariadas / Verônica
Amparo Quispe Chire. -- São Carlos : UFSCar, 2013.
60 f.

Dissertação (Mestrado) -- Universidade Federal de São
Carlos, 2013.

1. Inferência (Estatística). 2. Distribuições discretas
bivariadas. 3. Método da redução trivariada. 4. Excesso de
zeros. I. Título.

CDD: 519 (20^a)



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia

Programa de Pós-Graduação em Estatística

Via Washington Luís, Km 235 - C.P.676 - CGC 45358058/0001-40

FONE: (016) 3351-8292 – Email: ppgest@ufscar.br

13565-905 - SÃO CARLOS-SP - BRASIL

FOLHA DE APROVAÇÃO

Aluno(a) : Verônica Amparo Quispe Chire

DISSERTAÇÃO DE MESTRADO DEFENDIDA E APROVADA EM 26/11/2013
PELA COMISSÃO JULGADORA:

Presidente

Prof. Dr. Luis Ernesto Bueno Salasar (DEs-UFSCar/Orientador)

1º Examinador

Prof. Dr. José Galvão Leite (IME-USP)

2º Examinador

Profa. Dra. Vera Lucia Damasceno Tomazella (DEs-UFSCar/Co-Orientadora)

3º Examinador

Prof. Dr. Victor Eliseo Leiva Sanchez (UV)

Agradecimentos

Primeiramente, agradeço a Deus pelo Seu amor incondicional, Sua fidelidade e por ter me guiado e encorajado na busca de mais um sonho.

Aos meus pais Felipe Santiago e Nelly Santos, que me criaram com muito amor e carinho e sempre me apoiaram e acreditaram nas minhas escolhas.

Aos meus queridos irmãos Arturo e Edher pelo companheirismo e pelos muito momentos de alegria, e as minhas queridas irmãs Hanny, Danieli e Oneide, pelo carinho e compreensão e apoio.

Ao professor Dr. Luis Ernesto Bueno Salasar pela orientação, paciência, apoio e pelas ideias deste trabalho.

A professora Dr. Vera Lucia Damasceno Tomazella pela co-orientação e incentivo.

Aos professores Márcio Luis Lanfredi Viola e Maria Aparecida de Paiva Franco, membros da banca do exame de qualificação, pelas correções e sugestões feitas.

Aos professores do Departamento de Estatística da Universidade Federal de São Carlos, em especial os professores Márcio Luis Lanfredi Viola e Adriano Polpo de Campo pelas contribuições neste trabalho.

A secretaria Maria Isabel do Departamento de Estatística da Universidade Federal de São Carlos pela atenção e amizade.

As minhas queridas professoras de inglês, Renata e Marina.

Ao Professor Dr. Luiz Carlos Baida e a professora Dr. Rita de Cássia Pavani Lamas da UNESP por acreditarem neste sonho.

As minhas amigas Patricia, Francielle e Liliane de São Paulo, onde este sonho começou, pelo carinho, apoio e amizade.

Aos amigos do mestrado pela amizade o companheirismo. Especialmente a Katherine, Lorena e Kelly pela força e ajuda que sempre me deram.

Finalmente, agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo auxílio concedido durante este trabalho.

Resumo

A análise de dados bivariados pode ser encontrada nas mais diversas áreas do conhecimento, quando os dados de interesse são obtidos de forma pareada e apresentam correlação entre as contagens.

Neste trabalho são apresentados os modelos Poisson bivariado de Holgate, Poisson generalizado bivariado e Poisson bivariado inflacionado de zeros, os quais são úteis na modelagem de dados de contagem bivariados correlacionados. Aplicações ilustrativas serão apresentadas para estes modelos e a comparação entre eles será realizada pelos critérios de seleção de modelos AIC e BIC, assim como pelo teste da razão de verossimilhança assintótico. Particularmente, propomos uma abordagem Bayesiana para os modelos Poisson bivariado de Holgate e Poisson Inflacionado de zeros, baseada no algoritmo Gibbs sampling com dados ampliados.

Palavras-chave: distribuições discretas bivariadas, método da redução trivariada, excesso de zeros, algoritmo EM, dados ampliados.

Abstract

The analysis of bivariate data can be found in several areas of knowledge, when the data of interest are obtained in a paired way and present correlation between counts.

In this work the Holgate bivariate Poisson, bivariate generalized Poisson and bivariate zero-inflated Poisson models are presented, which are useful to the modeling of bivariate count data correlated. Illustrative applications are presented for these models and the comparison between them is made by using criteria of model selection AIC and BIC, as well as the asymptotic likelihood ratio test. Particularly, we propose a Bayesian approach to the Holgate bivariate Poisson and bivariate zero-inflated Poisson models, based in the Gibbs sampling algorithm with data augmentation.

Keywords: bivariate discrete distributions, trivariate reduction method, excess zeros, EM algorithm, data augmentation.

Sumário

Lista de Abreviaturas	v
1 Introdução	1
1.1 Definições preliminares	2
1.1.1 Variáveis aleatórias discretas univariadas	2
1.1.2 Variáveis aleatórias discretas bivariadas	3
1.2 Métodos de estimação	5
1.2.1 Método dos momentos	5
1.2.2 Método de máxima verossimilhança	6
1.2.3 Algoritmo EM	6
1.3 Critérios de seleção	7
1.4 Objetivo do trabalho	7
1.5 Organização do trabalho	8
2 Distribuição Poisson Bivariada de Holgate	9
2.1 Distribuição Poisson univariado	9
2.2 Distribuição Poisson bivariada de Holgate	9
2.3 Estimação pontual	14
2.3.1 Método dos momentos	15
2.3.2 Método de máxima verossimilhança	16
2.4 Teste de hipótese	19
2.4.1 Teste de independência	19
2.5 Aplicação	20
3 Distribuição Poisson Generalizada Bivariada	22
3.1 Distribuição Poisson generalizada	22
3.2 Distribuição Poisson generalizada bivariada	23
3.3 Estimação pontual	24
3.3.1 Método dos momentos	25
3.3.2 Método de máxima verossimilhança	27
3.4 Teste de hipótese	28

3.4.1	Teste de independência	28
3.4.2	Teste para a distribuição PB	29
3.5	Aplicação	29
4	Distribuição Poisson Bivariada Inflacionada de Zeros	32
4.1	Distribuição Poisson inflacionada de zeros	32
4.2	Distribuição Poisson bivariada inflacionada de zeros	33
4.3	Estimação pontual	35
4.3.1	Método de máxima verossimilhança	35
4.3.2	Função de verossimilhança baseada nos dados aumentados	36
4.3.3	Estimação via algoritmo EM	39
4.4	Teste de hipótese	40
4.4.1	Teste de hipótese para o parâmetro de inflação	41
4.4.2	Teste para a distribuição ZIPD	41
4.4.3	Teste para a distribuição PD	42
4.5	Aplicação	42
5	Abordagem Bayesiana dos Modelos PB e BZIP	45
5.1	Introdução	45
5.2	Distribuição a posteriori do modelo PB a partir de dados ampliados	46
5.3	Distribuição a posteriori do modelo BZIP a partir de dados ampliados	49
5.4	Aplicação	51
6	Conclusões e Propostas Futuras	56

Lista de Abreviaturas

\mathbb{R}	Conjunto dos números reais
\mathbb{N}	Conjunto dos números naturais
f.p.	Função de probabilidades
f.g.p.	Função geradora de probabilidades
PB	Poisson bivariada de Holgate
PD	Poisson dupla
PGB	Poisson generalizada bivariada
BZIP	Poisson bivariada inflacionada de zeros
ZIPD	Poisson dupla inflacionada de zeros
EMV	Estimador de máxima verossimilhança
EMM	Estimador de momentos
EP	Erro padrão
MCMC	Métodos de Monte Carlo via cadeias de Markov

Capítulo 1

Introdução

A análise de dados pareados pode ser encontrada nas mais diversas áreas. A distribuição Poisson bivariada tem se destacado nos casos em que esses dados pareados são de contagem. Na área da saúde, Kocherlakota & Kocherlakota (2001) investigaram o número de ataques epiléticos em pacientes antes e depois da aplicação de uma determinada droga. Karlis & Ntzoufras (2005) avaliaram o número de consultas ao médico e o número total de medicamentos prescritos aos pacientes. No esporte, Karlis & Ntzoufras (2003) verificaram o número de pontos ganhos marcados entre equipes adversárias.

A distribuição de Poisson bivariada foi primeiramente discutida por Campbell (1934) e Aitken (1936), mas somente após o trabalho de Holgate (1964) a distribuição foi amplamente considerada na literatura. A distribuição de Poisson bivariada também foi estudada por Kawamura (1973) quem obteve a distribuição a partir do limite da distribuição Binomial bivariada. Em outros trabalhos, Kawamura (1984,1985) desenvolveu métodos inferenciais do modelo. Karlis & Ntzoufras (2003) constataram que o alto custo computacional para a estimação dos parâmetros da distribuição impediu a utilização do modelo Poisson bivariado de Holgate nas décadas de 70 e 80. A partir da década de 90, com o desenvolvimento de recursos computacionais, surgiram os primeiros modelos de regressão Poisson bivariado de Holgate apresentados por Jung & Winkelmann (1993), Kocherlakota & Kocherlakota (2001), Ho & Singer (2001) e Karlis & Ntzoufras (2005), entre outros.

Nas aplicações citadas acima, nota-se que os dados de contagem bivariados apresentam sobredispersão e excesso de zeros. Sobredispersão é um fenômeno comum que ocorre na modelagem de dados discretos e cuja ocorrência é caracterizada quando a variância é maior que a média. Visando a modelagem de dados com essa característica, Famoye & Consul (1995) propuseram a distribuição de Poisson generalizada bivariada, a qual é derivada a partir da distribuição de Poisson generalizada definida por Consul & Jain (1973) usando o método da redução trivariada. A distribuição de Poisson generalizada bivariada tem como caso particular o modelo de Poisson bivariado de Holgate. Vernic (1997) comparou estes modelos utilizando dados reais e obteve um melhor ajuste do modelo Poisson generalizado bivariado. O excesso de zeros em dados de contagem bivariados é uma ocorrência de certa forma, comum em diversas aplicações, que pode ser causada por uma combinação dos chamados zeros estruturais e zeros amostrais. Estas classificações foram mencionadas em Ridout *et al.* (1998). No primeiro caso, os zeros estruturais ocorrem

de forma inevitável. No último, os zeros amostrais ocorrem ao acaso. Com o objetivo de melhorar o ajuste dos modelos aos dados com excesso de zeros, modelos bivariados inflacionados de zeros foram discutidos por diversos autores. Dentre eles, Li *et al.* (1999) propuseram vários tipos de modelos de Poisson multivariado inflacionado de zeros e suas propriedades, entre eles o modelo Poisson bivariado inflacionado de zeros; Walhin (2001) discutiu a construção de três modelos de Poisson bivariado inflacionado de zeros e realizou uma comparação entre eles; Krishna & Tukaram (2011) definiram a distribuição series de potência bivariada inflacionada de zeros e suas propriedades. A distribuição de Poisson bivariada inflacionada de zeros, abordada em Li *et al.* (1999) foi proposta para ajustar dados com maior ocorrência de zeros que o esperado no modelo Poisson bivariado de Holgate.

Na abordagem Bayesiana, as distribuições discretas bivariadas, segundo Arab *et al.* (2011), envolvem dificuldades computacionais que têm impedido pesquisadores de usar essa abordagem para estas distribuições. Os recentes avanços na modelagem Bayesiana e, especificamente, a melhoria de métodos computacionais, como os MCMC, forneceram mecanismos para facilitar a análise Bayesiana das distribuições discretas bivariadas, como as de Poisson bivariada de Holgate e Poisson bivariada inflacionada de zeros.

Neste contexto Bayesiano, vários trabalhos foram propostos com respeito ao modelo Poisson bivariado de Holgate. Por exemplo, Tsionas (1999) abordou uma análise Bayesiana da distribuição de Poisson multivariada, baseada no algoritmo Gibbs sampling com dados aumentados. Stern & Zacks (2002) discutiram um teste de independência entre duas variáveis aleatórias de Poisson sobre a distribuição de Poisson bivariada de Holgate, utilizando o teste Bayesiano Evidence Test. Karlis & Tsiamyrtzis (2008) discutiram a análise Bayesiana do modelo Poisson bivariado de Holgate e ofereceram uma estimação da distribuição a posteriori do modelo. Recentemente, Majumdar & Gries (2010) discutiram a abordagem Bayesiana do modelo de regressão Poisson bivariado inflacionado de zeros utilizando os métodos MCMC.

1.1 Definições preliminares

1.1.1 Variáveis aleatórias discretas univariadas

Definição 1 *Uma variável aleatória X é discreta quando ela assume valores no conjunto $A = \{x_i, i = 1, 2, \dots\} \subseteq \mathbb{R}$ finito ou infinito enumerável. A função de probabilidades de X , $p_X(x_i) = P(X = x_i)$, para $i = 1, 2, \dots$, satisfaz*

$$i) \quad p_X(x_i) \geq 0, \quad \text{para todo } x_i \in A;$$

$$ii) \quad \sum_{i=1}^{\infty} p_X(x_i) = 1.$$

Definição 2 *Seja X uma variável aleatória discreta que assume valores inteiros e não negativos. A f.g.p. de X é definida em um intervalo $I \subseteq \mathbb{R}$ como*

$$g_X(t) = E[t^X] = \sum_{i=1}^{\infty} p_X(x_i)t^{x_i}, \quad t \in I,$$

desde que a série convirja absolutamente em I .

Observe que o domínio de definição I da f.g.p. $g_X(t)$ contém o intervalo $[-1, 1]$. De fato, para $|t| \leq 1$ temos que

$$0 \leq \sum_{i=1}^{\infty} |p_X(x_i)t^{x_i}| \leq \sum_{i=1}^{\infty} p_X(x_i)|t^{x_i}| \leq \sum_{i=1}^{\infty} p_X(x_i) = 1.$$

Teorema 1.1.1 *Existe uma correspondência biunívoca entre f.p. e f.g.p. para variáveis aleatórias discretas que assumem valores em \mathbb{N} .*

Demonstração: A k -ésima derivada de $g_X(t)$, denotada por $g_X^{(k)}(t)$, em relação a t e $k \geq 1$, é obtida pela sua diferenciação termo a termo e é dada por

$$g_X^{(k)}(t) = \sum_{n=k}^{\infty} n(n-1)(n-2)\dots(n-k+1)t^{n-k}P(X=n),$$

para $t \in I$, o que implica

$$g_X^{(k)}(0) = k!P(X=k).$$

Logo,

$$P(X=k) = \frac{g_X^{(k)}(0)}{k!}, \quad k=0,1,\dots$$

na qual definimos $g_X^{(0)}(t) = g_X(t)$. Logo, podemos obter a f.p. de X a partir da f.g.p. de X . ■

1.1.2 Variáveis aleatórias discretas bivariadas

Definição 3 *Um vetor aleatório (X, Y) é discreto quando assume ele valores em um conjunto $S = \{(x_i, y_j), i, j = 1, 2, \dots\} \subseteq \mathbb{R}^2$ finito ou infinito enumerável. A f.p. de (X, Y) , dada por $p_{X,Y}(x_i, y_j) = P(X = x_i, Y = y_j)$, para $i, j = 1, 2, \dots$, satisfaz*

$$i) \quad p_{X,Y}(x_i, y_j) \geq 0 \quad \text{para todo } (x_i, y_j) \in S;$$

$$ii) \quad \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} p_{X,Y}(x_i, y_j) = 1.$$

Definição 4 *Duas variáveis aleatórias X e Y discretas são independentes se, e somente se, $p_{X,Y}(x_i, y_j) = p_X(x_i)p_Y(y_j)$ para todo $(x_i, y_j) \in S = \{(x_i, y_j), i, j = 1, 2, \dots\} \subseteq \mathbb{R}^2$.*

Definição 5 *Seja (X, Y) um vetor aleatório discreto com f.p. $p_{X,Y}(x_i, y_j) = P(X = x_i, Y = y_j)$, $(x_i, y_j) \in \mathbb{N}^2$, para $i, j = 1, 2, \dots$. A f.g.p. bivariada de (X, Y) é definida em um retângulo $I \times J \subseteq \mathbb{R}^2$ como*

$$g_{X,Y}(t_1, t_2) = E[t_1^X t_2^Y] = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} p_{X,Y}(x_i, y_j) t_1^{x_i} t_2^{y_j}, \quad (t_1, t_2) \in I \times J,$$

desde que a série convirja absolutamente em $I \times J$.

Observe que o domínio de definição $I \times J$ da f.g.p $g_{X,Y}(t_1, t_2)$ contém o intervalo $[-1, 1] \times [-1, 1]$. Note que para $|t_1| \leq 1$ e $|t_2| \leq 1$ temos que

$$0 \leq \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} |p_{X,Y}(x_i, y_j) t_1^{x_i} t_2^{y_j}| \leq \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} p_{X,Y}(x_i, y_j) = 1.$$

Teorema 1.1.2 *Existe uma correspondência biunívoca entre f.p. e a f.g.p. para variáveis aleatórias discretas bivariadas que assumem valores em \mathbb{N}^2 .*

Demonstração: A derivada parcial de $g_{X,Y}(t_1, t_2)$, r vezes em relação a t_1 e s vezes em relação a t_2 , $r \geq 1, s \geq 1$, é obtida pela sua diferenciação termo a termo e é dada por

$$\frac{\partial^{r+s} g_{X,Y}(t_1, t_2)}{\partial t_1^r \partial t_2^s} = \sum_{x=r}^{\infty} \sum_{y=s}^{\infty} x(x-1) \dots (x-r+1) y(y-1) \dots (y-s+1) t_1^{x-r} t_2^{y-s} p(x, y),$$

para $(t_1, t_2) \in I \times J$.

Disto, temos que

$$\frac{\partial^{r+s} g_{X,Y}(0, 0)}{\partial t_1^r \partial t_2^s} = r!s!p(r, s),$$

o que implica que f.p de (X, Y) é dada por

$$\begin{aligned} p(r, s) &= \frac{1}{s!r!} \left. \frac{\partial^{r+s} g_{X,Y}(t_1, t_2)}{\partial t_1^r \partial t_2^s} \right|_{t_1=0, t_2=0}, \quad r, s = 1, 2, \dots, \\ p(0, s) &= \frac{1}{s!} \left. \frac{\partial^s g_{X,Y}(0, t_2)}{\partial t_2^s} \right|_{t_2=0}, \quad s = 1, 2, \dots, \\ p(r, 0) &= \frac{1}{r!} \left. \frac{\partial^r g_{X,Y}(t_1, 0)}{\partial t_1^r} \right|_{t_1=0}, \quad r = 1, 2, \dots, \\ p(0, 0) &= g_{X,Y}(0, 0). \end{aligned}$$

■

Teorema 1.1.3 *Seja $g_{X,Y}(t_1, t_2)$ a f.g.p. conjunta das variáveis aleatórias X e Y definida em $(t_1, t_2) \in I \times J$. Então, X e Y são independentes se, e somente se, $g_{X,Y}(t_1, t_2) = g_X(t_1)g_Y(t_2)$ para todo $t_1 \in I, t_2 \in J$, em que $g_X(t_1) = g_{X,Y}(t_1, 1)$ e $g_Y(t_2) = g_{X,Y}(1, t_2)$ são as f.g.p. de X e Y , respectivamente.*

Demonstração: Se X e Y são independentes, então

$$g_{X,Y}(t_1, t_2) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} t_1^{x_i} t_2^{y_j} p(x_i, y_j) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} t_1^{x_i} t_2^{y_j} p_X(x_i) p_Y(y_j) = \sum_{i=1}^{\infty} t_1^{x_i} p_X(x_i) \sum_{j=1}^{\infty} t_2^{y_j} p_Y(y_j) = g_X(t_1) g_Y(t_2).$$

Por outro lado, se $g_{X,Y}(t_1, t_2) = g_X(t_1)g_Y(t_2)$, então

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} t_1^{x_i} t_2^{y_j} p_{X,Y}(x_i, y_j) = \sum_{i=1}^{\infty} t_1^{x_i} p_X(x_i) \sum_{j=1}^{\infty} t_2^{y_j} p_Y(y_j) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} t_1^{x_i} t_2^{y_j} p_X(x_i) p_Y(y_j).$$

Da equação acima temos $p_{X,Y}(x_i, y_j) = p_X(x_i)p_Y(y_j)$ para todo $(x_i, y_j) \in \mathbb{N}^2$. Portanto, X e Y são variáveis aleatórias independentes. ■

1.2 Métodos de estimação

Nesta seção, vamos considerar que a f.p. $p(x, y)$ é indexada de um vetor de parâmetros $\theta \in \Theta \subseteq \mathbb{R}^q$, para $q \geq 1$, a qual será denotada por $p(x, y|\theta)$.

Diversos métodos de estimação tem sido sugeridos na literatura para estimar $\theta = (\theta_1, \theta_2, \dots, \theta_q)$, como, por exemplo, os métodos dos momentos e de máxima verossimilhança, que discutiremos a seguir.

1.2.1 Método dos momentos

O método dos momentos é bastante simples e intuitivo. Este método consiste em igualar os momentos populacionais a seus respectivos momentos amostrais.

Seja $\mu'_{r,s} = E[X^r Y^s] < \infty$, o (r, s) -ésimo momento das variáveis aleatórias X e Y e $\mu_{r,s} = E[(X - \mu_X)^r (Y - \mu_Y)^s] < \infty$, o (r, s) -ésimo momento central das variáveis aleatórias X e Y , em que $\mu_X = E[X] = \mu'_{1,0}$, $\mu_Y = E[Y] = \mu'_{0,1}$, para $r, s = 0, 1, \dots$

Considere uma amostra aleatória bivariada $(X_1, Y_1), \dots, (X_n, Y_n)$ de tamanho n de uma f.p. $p(x, y|\theta)$, e sejam $n_{ij} = \sum_{\kappa=1}^n I_{\{(i,j)\}}(X_{\kappa}, Y_{\kappa})$ a frequência de cada par (i, j) e $D = \{(i, j) : n_{ij} > 0\}$ o conjunto dos pares observados, para $i, j = 0, 1, \dots$. Desta forma, o tamanho da amostra é dado por

$$n = \sum_{(i,j) \in D} n_{ij}.$$

O (r, s) -ésimo momento amostral será denotado por

$$m'_{r,s} = \frac{1}{n} \sum_{(i,j) \in D} i^r j^s n_{ij}, \quad r, s = 0, 1, \dots,$$

e o (r, s) -ésimo momento amostral central será representado por

$$m_{r,s} = \frac{1}{n} \sum_{(i,j) \in D} (i - \bar{X})^r (j - \bar{Y})^s n_{ij},$$

sendo

$$\bar{X} = \frac{1}{n} \sum_{(i,j) \in D} i n_{ij}, \quad \bar{Y} = \frac{1}{n} \sum_{(i,j) \in D} j n_{ij}.$$

Nos casos em que é possível se obter $\boldsymbol{\theta}$ em função dos momentos amostrais a partir da resolução de um sistema de equações em que se iguala momentos populacionais aos respectivos momentos amostrais, denominamos esta função como estimador de momentos.

1.2.2 Método de máxima verossimilhança

Seja $(x_1, y_1), \dots, (x_n, y_n)$ os valores observados de uma amostra aleatória bivariada $(X_1, Y_1), \dots, (X_n, Y_n)$ de um vetor (X, Y) que possui a f.p. $p(x, y|\boldsymbol{\theta})$, para $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^q$, e $q \geq 1$. A sua função de verossimilhança para $\boldsymbol{\theta}$ é dada por

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n p(x_i, y_i|\boldsymbol{\theta}),$$

e seu logaritmo é $l(\boldsymbol{\theta}) = \sum_{i=1}^n \log \{p(x_i, y_i|\boldsymbol{\theta})\}$. A estimativa de máxima verossimilhança $\hat{\boldsymbol{\theta}}$ é o valor de $\boldsymbol{\theta}$ que maximiza a função de verossimilhança $L(\boldsymbol{\theta})$ ou, equivalentemente, que maximiza o seu logaritmo $l(\boldsymbol{\theta})$.

As estimativas de máxima verossimilhança dos parâmetros $\theta_1, \dots, \theta_q$ são obtidas resolvendo-se o sistema de equações,

$$\begin{aligned} \frac{\partial l(\boldsymbol{\theta})}{\partial \theta_1} &= 0, \\ &\vdots \\ \frac{\partial l(\boldsymbol{\theta})}{\partial \theta_q} &= 0. \end{aligned}$$

Para garantir que a solução das equações do sistema seja ponto de máximo, a matriz Hessiana da função de verossimilhança de $l(\boldsymbol{\theta})$

$$H = \left[\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right] \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$$

deve ser definida negativa, isto é, $\mathbf{z}' H \mathbf{z} < 0$, onde $\mathbf{z} \neq 0$ é um vetor coluna pertencente a \mathbb{R}^q , sendo cada elemento da matriz H dado por

$$h_{i,j} = \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}, \quad i, j = 1, 2, \dots, q.$$

É importante ressaltar que nem sempre é possível obter os estimadores de máxima verossimilhança de forma analítica, e portanto se faz necessário o uso de alguns métodos numéricos para obter as estimativas.

1.2.3 Algoritmo EM

O algoritmo EM, proposto por Dempster *et al.* (1977) é um amplamente utilizado no cálculo das estimativas de máxima verossimilhança de forma iterativa.

Existem duas principais aplicações do algoritmo EM: a primeira quando os dados são incompletos, devido a problemas ou limitação do processo de observação (dados faltantes) e a segunda, quando a maximização da função de verossimilhança é analiticamente complexa, mas a f.p. pode ser simplificada, admitindo a existência de valores adicionais (dados latentes). A segunda aplicação é mais comum e mais simples computacionalmente. Isto porque o algoritmo EM reduz a complexidade da maximização da função de verossimilhança utilizando os dados completos. Mais detalhes sobre o algoritmo EM pode ser visto em McLachlan & Krishnan (2007).

Seja \mathbf{x} o conjunto de dados observados e \mathbf{w} o conjunto dos dados latentes (não observados). Assim, temos que $\mathbf{x}_c = (\mathbf{x}, \mathbf{w})$ são os dados completos e sua f.p. é $p(\mathbf{x}_c|\boldsymbol{\theta})$, para $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^q$, $q \geq 1$. Denota-se por $l_c(\boldsymbol{\theta}|\mathbf{x}_c)$ o logaritmo da função de verossimilhança dos dados completos. Cada iteração do algoritmo EM consiste de dois passos, um passo E (esperança) e um passo M (maximização), definidos como:

- Passo E: calcule $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^k) = E[l_c(\boldsymbol{\theta}|\mathbf{X}_c)|\mathbf{x}, \boldsymbol{\theta}^k]$, onde a esperança é tomada com respeito a distribuição condicional $p(\mathbf{w}|\mathbf{x}, \boldsymbol{\theta}^k)$, para $k = 1, 2, \dots$
- Passo M: encontre $\boldsymbol{\theta}^{k+1}$ que maximiza $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^k)$.

Estes passos devem ser repetidos até se atingir uma convergência, podendo ser adotado como critério de parada, por exemplo, $\|\boldsymbol{\theta}^{k+1} - \boldsymbol{\theta}^k\|_m < \epsilon$, onde $\|\cdot\|_m$ é a norma do máximo e ϵ é um valor determinado maior que zero.

1.3 Critérios de seleção

Após a obtenção das estimativas dos parâmetros dos modelos propostos pelos métodos de estimação é importante a utilização de um critério que nos possibilite escolher um dos modelos. Os critérios mais usados são AIC [Akaike (1974)] e o BIC [Schwarz (1978)], definidos por $AIC = -2 \log L(\boldsymbol{\theta}) + 2d$ e $BIC = -2 \log L(\boldsymbol{\theta}) + d \log N$, respectivamente, com $L(\boldsymbol{\theta})$ sendo a função de máxima verossimilhança, d a dimensão do vetor de parâmetros e N o tamanho da amostra. O modelo escolhido é aquele que apresenta o menor valor de AIC ou BIC.

1.4 Objetivo do trabalho

O objetivo deste trabalho é modelar dados de contagem bivariados correlacionados sob uma perspectiva clássica e bayesiana. Inicialmente, apresentamos o modelo Poisson bivariado proposto por Holgate (1964). Em seguida considerando situações em que existem sobredispersão ou uma quantidade excessiva de zeros nos dados, discutimos duas generalizações da distribuição proposta por Holgate: PGB e a BZIP. A primeira tem distribuições marginais que apresentam sobredispersão, enquanto a segunda contempla a situação em que há excesso de zeros.

Todos os modelos foram ajustados sob enfoque frequentista, enquanto que os modelos PB e BZIP também foram ajustados sob enfoque Bayesiano a partir de distribuições a priori não informativas para os

parâmetros. A principal contribuição deste trabalho foi a dedução de expressões analíticas para a implementação de algoritmos de estimação iterativos (algoritmo EM e Gibbs sampling com dados aumentados) para os modelos PB e BZIP. Estes algoritmos de estimação utilizam variáveis latentes e simplificam a inferência sobre os parâmetros.

Do ponto de vista frequentista, a comparação entre os ajustes dos modelos acima é realizada por meio dos critérios de seleção AIC e BIC, enquanto que a verificação da associação entre as distribuições marginais dos modelos PB e PGB é realizada por meio do teste de razão de verossimilhança.

1.5 Organização do trabalho

No Capítulo 2, será apresentada uma revisão do modelo PB e métodos inferenciais. Para ilustrar aplicação do modelo, será utilizado um conjunto de dados reais em que a estimação será determinada através dos métodos dos momento e máxima verossimilhança.

No Capítulo 3, será descrito o modelo PGB, suas propriedades e testes de hipóteses. Em seguida, os métodos de momentos e de máxima verossimilhança serão usados a partir de um conjunto de dados reais.

Para melhorar o ajuste dos modelos ao caso em que os dados apresentam uma quantidade de zeros excessiva, no Capítulo 4 será apresentado o modelo BZIP. Para ilustrar a aplicação do modelo, será utilizado um conjunto de dados reais em que a estimação será determinada através do método máxima verossimilhança via algoritmo EM.

No Capítulo 5, será apresentada uma metodologia Bayesiana para os modelo PB e BZIP, baseada em dados aumentados. Finalmente, no Capítulo 6, serão apresentadas as conclusões e as propostas futuras.

Capítulo 2

Distribuição Poisson Bivariada de Holgate

2.1 Distribuição Poisson univariado

O modelo Poisson univariado é amplamente utilizado na modelagem de dados de contagem. Este modelo foi escrito por Siméon Denis Poisson em 1837 com o objetivo de estudar a probabilidade de julgamentos criminais e civis. A distribuição de Poisson é aplicada em diversas áreas. Por exemplo: na medicina na contagem de recorrências de um determinado tipo de doença e na ecologia na contagem de peixes em uma região demarcada.

Segundo Johnson *et al.* (2005), uma variável aleatória X tem distribuição de Poisson com parâmetro $\lambda > 0$ se sua função de probabilidades é dada por

$$f(x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2, \dots \quad (2.1)$$

A média e a variância da distribuição Poisson são iguais a λ . A igualdade entre a média e a variância caracteriza a distribuição de Poisson entre as demais distribuições discretas univariadas. Esta distribuição satisfaz a relação de recorrência univariada dada por

$$xf(x|\lambda) = \lambda f(x-1|\lambda), \quad x \geq 1.$$

A f.g.p. da distribuição de Poisson com média λ é

$$g_X(t) = E[t^X] = \sum_{x=0}^{\infty} t^x f(x|\lambda) = \sum_{x=0}^{\infty} t^x \frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda t)^x}{x!} = e^{\lambda(t-1)}, \quad t \in \mathbb{R}.$$

2.2 Distribuição Poisson bivariada de Holgate

O modelo Poisson bivariado proposto por Holgate (1964) é construído pelo método da redução trivariada. Este método é muito usado na construção de distribuições bivariadas. Basicamente, a ideia é criar um vetor de variáveis aleatórias dependentes a partir de três ou mais variáveis aleatórias independen-

tes. Um importante aspecto deste método é que as funções que conectam estas variáveis são geralmente funções elementares; maiores detalhes ver, Lai (1995).

Supondo V_1 , V_2 , U variáveis aleatórias independentes com distribuições Poisson univariadas com parâmetros $\lambda_1 > 0$, $\lambda_2 > 0$ e $\lambda_3 > 0$, respectivamente, definimos o vetor (X, Y) como

$$\begin{aligned} X &= V_1 + U, \\ Y &= V_2 + U. \end{aligned}$$

Portanto, a f.p. de (X, Y) da distribuição PB é dada por

$$\begin{aligned} p(x, y) &= \mathbb{P}(X = x, Y = y) \\ &= \mathbb{P}(V_1 + U = x, V_2 + U = y) \\ &= \sum_{u=0}^{\infty} \mathbb{P}(V_1 + U = x, V_2 + U = y | U = u) \mathbb{P}(U = u) \\ &= \sum_{u=0}^{\min(x, y)} \mathbb{P}(V_1 = x - u, V_2 = y - u | U = u) \mathbb{P}(U = u) \\ &= \sum_{u=0}^{\min(x, y)} \mathbb{P}(V_1 = x - u) \mathbb{P}(V_2 = y - u) \mathbb{P}(U = u) \\ &= e^{-(\lambda_1 + \lambda_2 + \lambda_3)} \sum_{u=0}^{\min(x, y)} \frac{\lambda_1^{x-u}}{(x-u)!} \frac{\lambda_2^{y-u}}{(y-u)!} \frac{\lambda_3^u}{u!}, \quad x, y = 0, 1, \dots \end{aligned} \tag{2.2}$$

Note que, quando $\lambda_3 = 0$, temos que a variável U será degenerada no ponto 0, isto é, $f(u|0) = I_{\{0\}}(u)$, o que implica que as variáveis X e Y são independentes.

Como em Kocherlakota & Kocherlakota (1992) a f.p. de (X, Y) , dada em (2.2), pode ser obtida através da f.g.p. De fato, a f.g.p de (X, Y) , é

$$\begin{aligned} g_{X, Y}(t_1, t_2) &= \mathbb{E}[t_1^X t_2^Y] \\ &= \mathbb{E}[t_1^{V_1+U} t_2^{V_2+U}] \\ &= \mathbb{E}[t_1^{V_1} t_2^{V_2} (t_1 t_2)^U] \\ &= \mathbb{E}[t_1^{V_1}] \mathbb{E}[t_2^{V_2}] \mathbb{E}[(t_1 t_2)^U] \\ &= g_{V_1}(t_1) g_{V_2}(t_2) g_U(t_1 t_2) \\ &= e^{(\lambda_1[t_1-1] + \lambda_2[t_2-1] + \lambda_3[t_1 t_2 - 1])}, \quad t_1, t_2 \in \mathbb{R}. \end{aligned} \tag{2.3}$$

Expandindo a expressão (2.3) em série de potências de t_1 e t_2 , temos

$$\begin{aligned}
g_{X,Y}(t_1, t_2) &= e^{(\lambda_1[t_1-1]+\lambda_2[t_2-1]+\lambda_3[t_1t_2-1])} \\
&= \left[\sum_{v_1=0}^{\infty} \frac{t_1^{v_1} e^{-\lambda_1} \lambda_1^{v_1}}{v_1!} \right] \left[\sum_{v_2=0}^{\infty} \frac{t_2^{v_2} e^{-\lambda_2} \lambda_2^{v_2}}{v_2!} \right] \left[\sum_{u=0}^{\infty} \frac{(t_2t_1)^u e^{-\lambda_3} \lambda_3^u}{u!} \right] \\
&= e^{-(\lambda_1+\lambda_2+\lambda_3)} \sum_{v_1=0}^{\infty} \sum_{v_2=0}^{\infty} \sum_{u=0}^{\infty} \frac{t_1^{v_1+u} t_2^{v_2+u} \lambda_1^{v_1} \lambda_2^{v_2} \lambda_3^u}{v_1! v_2! u!}.
\end{aligned}$$

Fazendo $x = v_1 + u$ e $y = v_2 + u$,

$$\begin{aligned}
&= e^{-(\lambda_1+\lambda_2+\lambda_3)} \sum_{x=0}^{\infty} \sum_{y=0}^{\infty} \sum_{u=0}^{\min(x,y)} \frac{\lambda_1^{x-u} \lambda_2^{y-u} \lambda_3^u}{(x-u)!(y-u)!u!} t_1^x t_2^y \\
&= \sum_{x=0}^{\infty} \sum_{y=0}^{\infty} \underbrace{e^{-(\lambda_1+\lambda_2+\lambda_3)} \sum_{u=0}^{\min(x,y)} \frac{\lambda_1^{x-u} \lambda_2^{y-u} \lambda_3^u}{(x-u)!(y-u)!u!}}_{p(x,y)} t_1^x t_2^y,
\end{aligned}$$

obtém-se a f.p. (2.2) no ponto (x, y) como o coeficiente de $t_1^x t_2^y$, para $x, y = 0, 1, \dots$

As variáveis aleatórias X e Y possuem a f.p. marginal dadas, respectivamente, por

$$P(X = x) = \frac{e^{-(\lambda_1+\lambda_3)} (\lambda_1 + \lambda_3)^x}{x!}, \quad x = 0, 1, 2, \dots,$$

e

$$P(Y = y) = \frac{e^{-(\lambda_2+\lambda_3)} (\lambda_2 + \lambda_3)^y}{y!}, \quad y = 0, 1, 2, \dots$$

Note que a distribuição marginal de X e Y é Poisson, com parâmetros $\lambda_1 + \lambda_3$ e $\lambda_2 + \lambda_3$, respectivamente.

Logo, a média e a variância das variáveis são dadas por

$$E[X] = V[X] = \lambda_1 + \lambda_3 \quad e$$

$$E[Y] = V[Y] = \lambda_2 + \lambda_3.$$

A covariância entre X e Y é dada por

$$\begin{aligned}
\text{Cov}[X, Y] &= \text{Cov}[V_1 + U, V_2 + U] \\
&= \text{Cov}[V_1 + U, V_2] + \text{Cov}[V_1 + U, U] \\
&= \text{Cov}[V_1, V_2] + \text{Cov}[V_1, U] + \text{Cov}[U, V_2] + \text{Cov}[U, U] \\
&= \text{Cov}[U, U] \\
&= V[U] \\
&= \lambda_3,
\end{aligned}$$

ou seja, λ_3 é uma medida de dependência entre as variáveis aleatórias X e Y . Logo, o coeficiente de correlação entre X e Y é

$$\text{Corr}[X, Y] = \frac{\lambda_3}{\sqrt{(\lambda_1 + \lambda_3)(\lambda_2 + \lambda_3)}}.$$

Note que o coeficiente de correlação assume somente valores positivos pois $\lambda_1, \lambda_2, \lambda_3 > 0$. Além disso, para o modelo PB, se $\lambda_3 = \text{Cov}[X, Y] = 0$, então as variáveis X e Y são independentes e, neste caso, a f.p. (2.2) é

$$\begin{aligned} p(x, y) &= \text{P}[X = x, Y = y] \\ &= \text{P}[X = x] \text{P}[Y = y] \\ &= \frac{\lambda_1^x \lambda_2^y e^{-(\lambda_1 + \lambda_2)}}{x! y!}, \quad x, y = 0, 1, \dots, \end{aligned} \quad (2.4)$$

a qual é referida em Karlis & Ntzoufras (2003) como distribuição Poisson dupla.

Teicher (1954) e Kawamura (1985) obtiveram as relações de recorrência mostradas no Lema 2.2.1.

Lema 2.2.1 *A f.p. $p(x, y)$ associada com a distribuição PB satisfaz as relações de recorrência:*

$$i) \quad xp(x, y) = \lambda_1 p(x-1, y) + \lambda_3 p(x-1, y-1) \quad e$$

$$ii) \quad yp(x, y) = \lambda_2 p(x, y-1) + \lambda_3 p(x-1, y-1),$$

em que $p(-1, y) = p(x, -1) = 0$ para todos os inteiros $x, y = 0, 1, 2, \dots$

Demonstração: A f.p. da distribuição PB é dada por

$$p(x, y) = \sum_{u=0}^{\min(x, y)} f(x-u|\lambda_1) f(y-u|\lambda_2) f(u|\lambda_3) \quad x, y = 0, 1, \dots,$$

em que $f(k|\lambda)$, para $k = 0, 1, \dots$ é a f.p. da distribuição Poisson univariada.

Supondo $x \leq y$, tem-se

$$\begin{aligned} p(x, y) &= \sum_{u=0}^x f(x-u|\lambda_1) f(y-u|\lambda_2) f(u|\lambda_3) \\ &= f(x|\lambda_1) f(y|\lambda_2) f(0|\lambda_3) + f(x-1|\lambda_1) f(y-1|\lambda_2) f(1|\lambda_3) \\ &\quad + f(x-2|\lambda_1) f(y-2|\lambda_2) f(2|\lambda_3) + f(x-3|\lambda_1) f(y-3|\lambda_2) f(3|\lambda_3) \\ &\quad + \dots \\ &\quad + f(0|\lambda_1) f(y-x|\lambda_2) f(x|\lambda_3). \end{aligned} \quad (2.5)$$

Multiplicando ambos os lados de (2.5) por x , com $1 \leq x \leq y$, tem-se

$$\begin{aligned}
xp(x, y) &= xf(x|\lambda_1)f(y|\lambda_2)f(0|\lambda_3) \\
&+ (x-1)f(x-1|\lambda_1)f(y-1|\lambda_2)f(1|\lambda_3) + 1f(x-1|\lambda_1)f(y-1|\lambda_2)f(1|\lambda_3) \\
&+ (x-2)f(x-2|\lambda_1)f(y-2|\lambda_2)f(2|\lambda_3) + 2f(x-2|\lambda_1)f(y-2|\lambda_2)f(2|\lambda_3) \\
&+ \dots \\
&+ 1f(1|\lambda_1)f(y-x+1|\lambda_2)f(x-1|\lambda_3) + (x-1)f(1|\lambda_1)f(y-x+1|\lambda_2)f(x-1|\lambda_3) \\
&+ xf(0|\lambda_1)f(y-x|\lambda_2)f(x|\lambda_3), \tag{2.6}
\end{aligned}$$

utilizando $kf(k|\lambda) = \lambda f(k-1|\lambda)$, para $k \geq 1$, em (2.6), temos que

$$\begin{aligned}
xp(x, y) &= \lambda_1 f(x-1|\lambda_1)f(y|\lambda_2)f(0|\lambda_3) \\
&+ \lambda_1 f(x-2|\lambda_1)f(y-1|\lambda_2)f(1|\lambda_3) + \lambda_3 f(x-1|\lambda_1)f(y-1|\lambda_2)f(0|\lambda_3) \\
&+ \lambda_1 f(x-3|\lambda_1)f(y-2|\lambda_2)f(2|\lambda_3) + \lambda_3 f(x-2|\lambda_1)f(y-2|\lambda_2)f(1|\lambda_3) \\
&+ \dots \\
&+ \lambda_1 f(0|\lambda_1)f(y-x+1|\lambda_2)f(x-1|\lambda_3) + \lambda_3 f(1|\lambda_1)f(y-x+1|\lambda_2)f(x-2|\lambda_3) \\
&+ \lambda_3 f(0|\lambda_1)f(y-x|\lambda_2)f(x-1|\lambda_3) \\
&= \lambda_1 f(x-1|\lambda_1)f(y|\lambda_2)f(0|\lambda_3) + \lambda_1 f(x-2|\lambda_1)f(y-1|\lambda_2)f(1|\lambda_3) \\
&+ \lambda_1 f(x-3|\lambda_1)f(y-2|\lambda_2)f(2|\lambda_3) + \dots + \lambda_1 f(0|\lambda_1)f(y-x+1|\lambda_2)f(x-1|\lambda_3) \\
&+ \lambda_3 f(x-1|\lambda_1)f(y-1|\lambda_2)f(0|\lambda_3) + \lambda_3 f(x-2|\lambda_1)f(y-2|\lambda_2)f(1|\lambda_3) \\
&+ \dots + \lambda_3 f(1|\lambda_1)f(y-x+1|\lambda_2)f(x-2|\lambda_3) + \lambda_3 f(0|\lambda_1)f(y-x|\lambda_2)f(x-1|\lambda_3) \\
&= \lambda_1 \sum_{u=0}^{\min(x-1, y)} f(x-1-u|\lambda_1)f(y-u|\lambda_2)f(u|\lambda_3) \\
&+ \lambda_3 \sum_{u=0}^{\min(x-1, y-1)} f(x-1-u|\lambda_1)f(y-1-u|\lambda_2)f(u|\lambda_3) \\
&= \lambda_1 p(x-1, y) + \lambda_3 p(x-1, y-1).
\end{aligned}$$

Similarmente, se $x > y$, então

$$\begin{aligned}
p(x, y) &= \sum_{u=0}^y f(x-u|\lambda_1)f(y-u|\lambda_2)f(u|\lambda_3) \\
&= f(x|\lambda_1)f(y|\lambda_2)f(0|\lambda_3) + f(x-1|\lambda_1)f(y-1|\lambda_2)f(1|\lambda_3) \\
&+ \dots \\
&+ f(x-y|\lambda_1)f(0|\lambda_2)f(y|\lambda_3). \tag{2.7}
\end{aligned}$$

Multiplicando por x ambos os lados de (2.7), com $1 \leq y < x$, tem-se

$$\begin{aligned}
xp(x, y) &= xf(x|\lambda_1)f(y|\lambda_2)f(0|\lambda_3) \\
&+ (x-1)f(x-1|\lambda_1)f(y-1|\lambda_2)f(1|\lambda_3) + 1f(x-1|\lambda_1)f(y-1|\lambda_2)f(1|\lambda_3) \\
&+ \dots \\
&+ (x-y)f(x-y|\lambda_1)f(0|\lambda_2)f(y|\lambda_3) + yf(x-y|\lambda_1)f(0|\lambda_2)f(y|\lambda_3).
\end{aligned} \tag{2.8}$$

Usando $kf(k|\lambda) = \lambda f(k-1|\lambda)$, para $k \geq 1$, em (2.8) temos

$$\begin{aligned}
&= \lambda_1 f(x-1|\lambda_1)f(y|\lambda_2)f(0|\lambda_3) \\
&+ \lambda_1 f(x-2|\lambda_1)f(y-1|\lambda_2)f(1|\lambda_3) + \lambda_3 f(x-1|\lambda_1)f(y-1|\lambda_2)f(0|\lambda_3) \\
&+ \dots \\
&+ \lambda_1 f(x-y-1|\lambda_1)f(0|\lambda_2)f(y|\lambda_3) + \lambda_3 f(x-y|\lambda_1)f(0|\lambda_2)f(y-1|\lambda_3) \\
&= \lambda_1 f(x-1|\lambda_1)f(y|\lambda_2)f(0|\lambda_3) + \lambda_1 f(x-2|\lambda_1)f(y-1|\lambda_2)f(1|\lambda_3) \\
&+ \dots + \lambda_1 f(x-y-1|\lambda_1)f(0|\lambda_2)f(y|\lambda_3) \\
&+ \lambda_3 f(x-1|\lambda_1)f(y-1|\lambda_2)f(0|\lambda_3) + \lambda_3 f(x-1|\lambda_1)f(y-1|\lambda_2)f(0|\lambda_3) \\
&+ \dots + \lambda_3 f(x-y|\lambda_1)f(0|\lambda_2)f(y-1|\lambda_3) \\
&= \lambda_1 \sum_{u=0}^{\min(x-1, y)} f(x-1-u|\lambda_1)f(y-u|\lambda_2)f(u|\lambda_3) \\
&+ \lambda_3 \sum_{u=0}^{\min(x-1, y-1)} f(x-1-u|\lambda_1)f(y-1-u|\lambda_2)f(u|\lambda_3) \\
&= \lambda_1 p(x-1, y) + \lambda_3 p(x-1, y-1).
\end{aligned}$$

Portanto, a f.p. associada com a distribuição PB satisfaz a relação de recorrência (i). Analogamente a relação de recorrência (ii) é obtida. ■

2.3 Estimação pontual

Na literatura são mencionados os seguintes métodos de estimação para os parâmetros do modelo PB: método dos momentos, método de máxima verossimilhança, método do duplo zero e o método dos pontos pares. Alguns métodos foram especificamente desenvolvidos para estimar o parâmetro associado com a covariância como o método do duplo zero; maiores detalhes ver, Kocherlakota & Kocherlakota (1992). Neste trabalho utilizamos os métodos dos momentos e de máxima verossimilhança para estimar os parâmetros do modelo sob o ponto de vista clássico.

No que segue vamos considerar uma amostra aleatória $(X_1, Y_1), \dots, (X_n, Y_n)$ de tamanho n da distribuição PB. Denotamos $n_{ij} = \sum_{\kappa=1}^n I_{\{(i, j)\}}(X_\kappa, Y_\kappa)$ como a frequência de cada par (i, j) , para $i, j = 0, 1, \dots$

e $D = \{(i, j) : n_{ij} > 0\}$ como o conjunto dos pares (i, j) observados. Consequentemente, o tamanho da amostra é dado por

$$n = \sum_{(i,j) \in D} n_{ij}.$$

Denotaremos por $\mathbf{x} = (x_1, x_2, \dots, x_n)$ e $\mathbf{y} = (y_1, y_2, \dots, y_n)$ os vetores observados de contagens de X e Y , respectivamente.

2.3.1 Método dos momentos

Os momentos populacionais $\mu'_{1,0}, \mu'_{0,1}$ e $\mu_{1,1}$ da distribuição PB são, respectivamente,

$$\begin{aligned} E[X] &= \lambda_1 + \lambda_3, \\ E[Y] &= \lambda_2 + \lambda_3, \\ \text{Cov}[X, Y] &= \lambda_3. \end{aligned}$$

e os seus respectivos momentos amostrais são dados por

$$\begin{aligned} \bar{X} &= \frac{1}{n} \sum_{(i,j) \in D} in_{ij}, \\ \bar{Y} &= \frac{1}{n} \sum_{(i,j) \in D} jn_{ij}, \\ m_{1,1} &= \frac{1}{n} \sum_{(i,j) \in D} (i - \bar{X})(j - \bar{Y})n_{ij}. \end{aligned}$$

Denotando $\tilde{\lambda}_1, \tilde{\lambda}_2$ e $\tilde{\lambda}_3$ os estimadores de λ_1, λ_2 e λ_3 , respectivamente, e igualando os momentos populacionais a seus respectivos momentos amostrais, temos que

$$\begin{aligned} \tilde{\lambda}_1 + \tilde{\lambda}_3 &= \bar{X}, \\ \tilde{\lambda}_2 + \tilde{\lambda}_3 &= \bar{Y}, \\ \tilde{\lambda}_3 &= m_{1,1}, \end{aligned}$$

o que implica

$$\begin{aligned} \tilde{\lambda}_1 &= \bar{X} - m_{1,1}, \\ \tilde{\lambda}_2 &= \bar{Y} - m_{1,1}, \\ \tilde{\lambda}_3 &= m_{1,1}. \end{aligned}$$

2.3.2 Método de máxima verossimilhança

Para estimar os parâmetros de interesse pelo método de máxima verossimilhança de uma forma mais simples, Holgate (1964) reescreveu a f.p. (2.2) em termos

$$\begin{aligned}\lambda_1^* &= \lambda_1 + \lambda_3 = E[X] \quad e \\ \lambda_2^* &= \lambda_2 + \lambda_3 = E[Y],\end{aligned}$$

ou seja,

$$p(x, y) = e^{-(\lambda_1^* + \lambda_2^* - \lambda_3)} \sum_{u=0}^{\min(x, y)} \frac{(\lambda_1^* - \lambda_3)^{x-u} (\lambda_2^* - \lambda_3)^{y-u} \lambda_3^u}{(x-u)! (y-u)! u!}, \quad x, y = 0, 1, \dots \quad (2.9)$$

onde $0 < \lambda_1^*, \lambda_2^* < \infty$ e $0 \leq \lambda_3 < \min(\lambda_1^*, \lambda_2^*)$.

Assim, a função de verossimilhança do vetor de parâmetros $\boldsymbol{\lambda}^* = (\lambda_1^*, \lambda_2^*, \lambda_3)$ é dada por

$$L(\boldsymbol{\lambda}^*) = e^{-n(\lambda_1^* + \lambda_2^* - \lambda_3)} \prod_{i=1}^n \sum_{u=0}^{\min(x_i, y_i)} \frac{(\lambda_1^* - \lambda_3)^{x_i - u} (\lambda_2^* - \lambda_3)^{y_i - u} \lambda_3^u}{(x_i - u)! (y_i - u)! u!},$$

e o logaritmo da função de verossimilhança é

$$l(\boldsymbol{\lambda}^*) = \log \prod_{i=1}^n p(x_i, y_i) = \sum_{i=1}^n \log \{p(x_i, y_i)\}. \quad (2.10)$$

Derivando o logaritmo da função de verossimilhança (2.10) em relação aos parâmetros λ_1^*, λ_2^* e λ_3 , obtém-se as respectivas escores dadas por

$$\frac{\partial l(\boldsymbol{\lambda}^*)}{\partial \lambda_1^*} = \sum_{i=1}^n \frac{\partial p(x_i, y_i)}{\partial \lambda_1^*} \frac{1}{p(x_i, y_i)}, \quad (2.11)$$

$$\frac{\partial l(\boldsymbol{\lambda}^*)}{\partial \lambda_2^*} = \sum_{i=1}^n \frac{\partial p(x_i, y_i)}{\partial \lambda_2^*} \frac{1}{p(x_i, y_i)}, \quad (2.12)$$

$$\frac{\partial l(\boldsymbol{\lambda}^*)}{\partial \lambda_3} = \sum_{i=1}^n \frac{\partial p(x_i, y_i)}{\partial \lambda_3} \frac{1}{p(x_i, y_i)}. \quad (2.13)$$

As derivadas da f.p. (2.9) com respeito a λ_1^*, λ_2^* e λ_3 , são respectivamente dadas por

$$\begin{aligned}\frac{\partial p(x, y)}{\partial \lambda_1^*} &= e^{-(\lambda_1^* + \lambda_2^* - \lambda_3)} \sum_{u=0}^{\min(x-1, y)} \frac{(\lambda_1^* - \lambda_3)^{x-1-u} (\lambda_2^* - \lambda_3)^{y-u} \lambda_3^u}{(x-1-u)! (y-u)! u!} \\ &\quad - e^{-(\lambda_1^* + \lambda_2^* - \lambda_3)} \sum_{u=0}^{\min(x, y)} \frac{(\lambda_1^* - \lambda_3)^{x-u} (\lambda_2^* - \lambda_3)^{y-u} \lambda_3^u}{(x-u)! (y-u)! u!} \\ &= p(x-1, y) - p(x, y),\end{aligned} \quad (2.14)$$

$$\begin{aligned}
\frac{\partial p(x, y)}{\partial \lambda_2^*} &= e^{-(\lambda_1^* + \lambda_2^* - \lambda_3)} \sum_{u=0}^{\min(x, y-1)} \frac{(\lambda_1^* - \lambda_3)^{x-u} (\lambda_2^* - \lambda_3)^{y-1-u} \lambda_3^u}{(x-u)! (y-1-u)! u!} \\
&\quad - e^{-(\lambda_1^* + \lambda_2^* - \lambda_3)} \sum_{u=0}^{\min(x, y)} \frac{(\lambda_1^* - \lambda_3)^{x-u} (\lambda_2^* - \lambda_3)^{y-u} \lambda_3^u}{(x-u)! (y-u)! u!} \\
&= p(x, y-1) - p(x, y),
\end{aligned} \tag{2.15}$$

$$\begin{aligned}
\frac{\partial p(x, y)}{\partial \lambda_3} &= e^{-(\lambda_1^* + \lambda_2^* - \lambda_3)} \sum_{u=0}^{\min(x, y)} \frac{(\lambda_1^* - \lambda_3)^{x-u} (\lambda_2^* - \lambda_3)^{y-u} \lambda_3^u}{(x-u)! (y-u)! u!} \\
&\quad - e^{-(\lambda_1^* + \lambda_2^* - \lambda_3)} \sum_{u=0}^{\min(x-1, y)} \frac{(\lambda_1^* - \lambda_3)^{x-1-u} (\lambda_2^* - \lambda_3)^{y-u} \lambda_3^u}{(x-1-u)! (y-u)! u!} \\
&\quad - e^{-(\lambda_1^* + \lambda_2^* - \lambda_3)} \sum_{u=0}^{\min(x, y-1)} \frac{(\lambda_1^* - \lambda_3)^{x-1-u} (\lambda_2^* - \lambda_3)^{y-1-u} \lambda_3^u}{(x-u)! (y-1-u)! u!} \\
&\quad + e^{-(\lambda_1^* + \lambda_2^* - \lambda_3)} \sum_{r=0}^{\min(x-1, y-2)} \frac{(\lambda_1^* - \lambda_3)^{x-1-r} (\lambda_2^* - \lambda_3)^{y-1-r} \lambda_3^r}{(x-1-r)! (y-1-r)! r!} \\
&= p(x, y) - p(x-1, y) - p(x, y-1) + p(x-1, y-1),
\end{aligned} \tag{2.16}$$

em que $p(-1, y) = p(x, -1) = p(-1, -1) = 0$, para $x, y = 0, 1, \dots$

Utilizando o Lema (2.2.1) temos as relações de recorrência

$$\begin{aligned}
xp(x, y) &= (\lambda_1^* - \lambda_3)p(x-1, y) + \lambda_3 p(x-1, y-1) \quad e \\
yp(x, y) &= (\lambda_2^* - \lambda_3)p(x, y-1) + \lambda_3 p(x-1, y-1),
\end{aligned}$$

para todos os inteiros $x, y = 0, 1, 2, \dots$, das quais temos

$$\frac{p(x-1, y)}{p(x, y)} = \frac{x}{(\lambda_1^* - \lambda_3)} - \frac{p(x-1, y-1)}{p(x, y)} \frac{\lambda_3}{(\lambda_1^* - \lambda_3)}, \tag{2.17}$$

$$\frac{p(x, y-1)}{p(x, y)} = \frac{y}{(\lambda_2^* - \lambda_3)} - \frac{p(x-1, y-1)}{p(x, y)} \frac{\lambda_3}{(\lambda_2^* - \lambda_3)}, \tag{2.18}$$

com $x, y = 0, 1, \dots$

Substituindo (2.14) e (2.17) na escora (2.11), temos

$$\sum_{i=1}^n \frac{x_i}{(\lambda_1^* - \lambda_3)} - \frac{\lambda_3}{(\lambda_1^* - \lambda_3)} \sum_{i=1}^n \frac{p(x_i-1, y_i-1)}{p(x_i, y_i)} - n = \frac{n\bar{x}}{(\lambda_1^* - \lambda_3)} - \frac{\lambda_3}{(\lambda_1^* - \lambda_3)} n\bar{R} - n, \tag{2.19}$$

em que

$$\bar{R} = \frac{1}{n} \sum_{i=1}^n \frac{p(x_i - 1, y_i - 1)}{p(x_i, y_i)} \quad \text{e} \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Substituindo (2.15) e (2.18) na equação escore (2.12), temos

$$\sum_{i=1}^n \frac{y_i}{(\hat{\lambda}_2^* - \lambda_3)} - \frac{\lambda_3}{(\lambda_2^* - \lambda_3)} \sum_{i=1}^n \frac{p(x_i - 1, y_i - 1)}{p(x_i, y_i)} - n = \frac{n\bar{y}}{(\lambda_2^* - \lambda_3)} - \frac{\lambda_3}{(\lambda_2^* - \lambda_3)} n\bar{R} - n, \quad (2.20)$$

onde

$$\bar{R} = \frac{1}{n} \sum_{i=1}^n \frac{p(x_i - 1, y_i - 1)}{p(x_i, y_i)} \quad \text{e} \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i.$$

Usando (2.16) e as equações (2.17) e (2.18) na escore (2.13), temos

$$n - \frac{n\bar{x}}{\hat{\lambda}_1^* - \lambda_3} - \frac{n\bar{y}}{\lambda_2^* - \lambda_3} - \left[\frac{\lambda_3}{\hat{\lambda}_1^* - \lambda_3} + \frac{\lambda_3}{\lambda_2^* - \lambda_3} + 1 \right] n\bar{R}. \quad (2.21)$$

Igualando os escores dados em (2.19), (2.20) e (2.21) a zero, obtemos o sistema de equações

$$\left\{ \begin{array}{l} \frac{\bar{X}}{\hat{\lambda}_1^* - \hat{\lambda}_3} - \frac{\hat{\lambda}_3 \bar{R}}{\hat{\lambda}_1^* - \hat{\lambda}_3} - 1 = 0, \end{array} \right. \quad (2.22)$$

$$\left\{ \begin{array}{l} \frac{\bar{Y}}{\hat{\lambda}_2^* - \hat{\lambda}_3} - \frac{\hat{\lambda}_3 \bar{R}}{\hat{\lambda}_2^* - \hat{\lambda}_3} - 1 = 0, \end{array} \right. \quad (2.23)$$

$$\left\{ \begin{array}{l} \frac{\bar{X}}{\hat{\lambda}_1^* - \hat{\lambda}_3} + \frac{\bar{Y}}{\hat{\lambda}_2^* - \hat{\lambda}_3} - \left[1 + \frac{\hat{\lambda}_3}{\hat{\lambda}_1^* - \hat{\lambda}_3} + \frac{\hat{\lambda}_3}{\hat{\lambda}_2^* - \hat{\lambda}_3} \right] \bar{R} - 1 = 0, \end{array} \right. \quad (2.24)$$

em que $\hat{\lambda}_1^*$, $\hat{\lambda}_2^*$ e $\hat{\lambda}_3$ são os estimadores de máxima verossimilhança de λ_1 , λ_2 e λ_3 , respectivamente.

De (2.22) e (2.23), segue que

$$\frac{\bar{X}}{\hat{\lambda}_1^* - \hat{\lambda}_3} = 1 + \frac{\hat{\lambda}_3 \bar{R}}{\hat{\lambda}_1^* - \hat{\lambda}_3} \quad \text{e}$$

$$\frac{\bar{Y}}{\hat{\lambda}_2^* - \hat{\lambda}_3} = 1 + \frac{\hat{\lambda}_3 \bar{R}}{\hat{\lambda}_2^* - \hat{\lambda}_3}.$$

Substituindo em (2.24), temos que

$$2 + \frac{\hat{\lambda}_3 \bar{R}}{\hat{\lambda}_1^* - \hat{\lambda}_3} + \frac{\hat{\lambda}_3 \bar{R}}{\hat{\lambda}_2^* - \hat{\lambda}_3} - \left[1 + \frac{\hat{\lambda}_3}{\hat{\lambda}_1^* - \hat{\lambda}_3} + \frac{\hat{\lambda}_3}{\hat{\lambda}_2^* - \hat{\lambda}_3} \right] \bar{R} - 1 = 0 \Rightarrow 1 - \bar{R} = 0 \Rightarrow \bar{R} = 1.$$

Logo, de (2.22) e (2.23), segue que

$$\frac{\bar{X}}{\hat{\lambda}_1^* - \hat{\lambda}_3} = 1 + \frac{\hat{\lambda}_3}{\hat{\lambda}_1^* - \hat{\lambda}_3} = \frac{\hat{\lambda}_1^*}{\hat{\lambda}_1^* - \hat{\lambda}_3} \Rightarrow \bar{X} = \hat{\lambda}_1^*,$$

$$\frac{\bar{Y}}{\hat{\lambda}_2^* - \hat{\lambda}_3} = 1 + \frac{\hat{\lambda}_3}{\hat{\lambda}_2^* - \hat{\lambda}_3} = \frac{\hat{\lambda}_2^*}{\hat{\lambda}_2^* - \hat{\lambda}_3} \Rightarrow \bar{Y} = \hat{\lambda}_2^*.$$

Para obter o estimador de máxima verossimilhança $\hat{\lambda}_3$, devemos encontrar a raiz de $\bar{R} = 1$, em relação a variável λ_3 , para valores fixados de $\hat{\lambda}_1^* = \bar{X}$ e $\hat{\lambda}_2^* = \bar{Y}$ no intervalo $[0, \min(\bar{X}, \bar{Y})]$, o que implica

$$\hat{\lambda}_1 = \bar{X} - \hat{\lambda}_3,$$

$$\hat{\lambda}_2 = \bar{Y} - \hat{\lambda}_3,$$

$$\hat{\lambda}_3 \in [0, \min(\bar{X}, \bar{Y})].$$

2.4 Teste de hipótese

Nesta seção, iremos utilizar o teste da razão de verossimilhança para testar a hipótese de interesse.

2.4.1 Teste de independência

As variáveis aleatórias X e Y são independentes se $\lambda_3 = 0$, já que neste caso a f.p. conjunta de X e Y é dada por (2.4). Portanto, para testar a hipótese de independência, devemos considerar as hipóteses

$$H_0 : \lambda_3 = 0$$

versus $H_1 : \lambda_3 \neq 0,$

e a estatística da razão de verossimilhanças dada por

$$\Psi = \frac{\sup_{\boldsymbol{\lambda} \in \Lambda_0} L(\boldsymbol{\lambda})}{\sup_{\boldsymbol{\lambda} \in \Lambda} L(\boldsymbol{\lambda})},$$

onde $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3) \in \Lambda = (0, \infty)^3$ e $\Lambda_0 = (0, \infty)^2 \times \{0\}$. A estatística $-2 \log(\Psi)$ segue, assintoticamente, uma distribuição qui-quadrado com $3 - 2 = 1$ grau de liberdade (diferença entre a dimensão do espaço paramétrico irrestrito e restrito). A região crítica do teste é, portanto, $C = \{-2 \log(\Psi) \geq \chi_{1, \alpha}^2\}$, onde $\chi_{1, \alpha}^2$ é o ponto crítico da χ_1^2 correspondente ao nível de significância α .

2.5 Aplicação

Nesta seção, aplicaremos a teoria descrita neste capítulo através de um conjunto de dados reais utilizado por Cresswell & Froggatt (1963). O dados referem-se ao número de acidentes de ônibus ocorridos com 708 motoristas pesquisados no norte da Irlanda durante os períodos 1952 – 53 e 1954 – 55. Estes dados são mostrados na Tabela 2.1.

Tabela 2.1: Distribuição das contagens do número de acidentes

Nro de acidentes em 1952-53 (X)	Nro de acidentes em 1954-55 (Y)							
	0	1	2	3	4	5	6	7
0	117	96	55	19	2	2	0	0
1	61	69	47	27	8	5	1	0
2	34	42	31	13	7	2	3	0
3	7	15	16	7	3	1	0	0
4	3	3	1	1	2	1	1	1
5	2	1	0	0	0	0	0	0
6	0	0	0	0	1	0	0	0
7	0	0	0	1	0	0	0	0

A variável X representa o número de acidentes no período 1952 – 53 e a variável Y o número de acidentes no período 1954 – 55. Toda a programação utilizada neste trabalho foi desenvolvida em linguagem de programação R, veja www.r-project.org.

A Tabela 2.2 apresenta os resultados descritivos para este conjunto de dados reais.

Tabela 2.2: Resumo descritivo dos dados

\bar{x}	\bar{y}	$m_{2,0}$	$m_{0,2}$	$m_{1,1}$
1,0014	1,2909	1,1934	1,5961	0,3258

Observamos que os dados apresentam uma covariância amostral positiva $m_{1,1} = 0,3258$, o que indica que existe uma associação positiva entre o número de acidentes no primeiro período e o segundo período. Desta forma, o modelo PB pode ser adequado para este conjunto de dados.

Na Tabela 2.3 estão apresentadas as estimativas dos parâmetros do modelo PB, que foram obtidas pelos métodos dos momentos e de máxima verossimilhança, e seus erros-padrão dos estimadores de máxima verossimilhança obtidos através da inversa da matriz de informação observada. Observamos uma proximidade das estimativas segundo ambos os métodos de estimação.

Tabela 2.3: Resultados para o modelo PB

Parâmetro	MM	EMV	EP
λ_1	0,6756	0,7722	0,0469
λ_2	0,9651	1,0617	0,0510
λ_3	0,3258	0,2292	0,0378

Para testar se o modelo com independência é adequado, foi realizado o teste de razão de verossimilhanças descrito na Subseção 2.4.1, obtendo $-2\log(\Psi) = 38,87$. O percentil da χ_1^2 , considerando o nível 0,1% de significância, tem valor crítico do teste igual a 10,82 e, observa-se que o valor obtido é maior que o valor crítico, levando a rejeição da hipótese nula. Assim, rejeitamos a hipótese de independência entre as variáveis X e Y e, desta forma, vale ressaltar a importância de se considerar a associação entre as variáveis aleatórias.

A Tabela 2.4 apresenta as frequências observadas e esperadas de acordo com o modelo PB, cujos parâmetros foram estimados pelo método de máxima verossimilhança. Observamos que as frequências esperadas não estão próximos dos valores observados, notadamente para as frequências dos pares (0,0) e (1,1). Visualmente o modelo PB não se ajusta bem a este conjunto de dados. A falta de um melhor ajuste do modelo PB pode ter ocorrido devido a sobredispersão ou a quantidade excessiva de zeros aos dados. A fim de obter um melhor ajuste dos dados em questão, apresentaremos a seguir duas generalizações da distribuição PB.

Tabela 2.4: Frequências observadas e esperadas para o modelo PB

x	y =	0	1	2	3	4	5	6	7	Total
0	Obs	117	96	55	19	2	0	0	0	291
	PB	89,94	95,50	50,70	17,94	4,76	1,01	0,17	0,02	
1	Obs	61	69	47	27	8	5	1	0	218
	PB	69,46	94,36	61,04	25,47	7,79	1,87	0,37	0,06	
2	Obs	34	42	31	13	7	2	3	0	132
	PB	26,82	44,39	34,38	16,83	5,92	1,61	0,35	0,06	
3	Obs	7	15	16	7	3	1	0	0	49
	PB	6,90	13,47	12,24	6,95	2,81	0,86	0,21	0,04	
4	Obs	3	3	1	1	2	1	1	1	13
	PB	1,33	2,99	3,13	2,04	0,94	0,32	0,09	0,020	
5	Obs	2	1	0	0	0	0	0	0	3
	PB	0,20	0,52	0,62	0,45	0,23	0,09	0,02	0,00	
6	Obs	0	0	0	0	1	0	0	0	1
	PB	0,02	0,07	0,10	0,08	0,04	0,00	0,00	0,00	
7	Obs	0	0	0	1	0	0	0	0	1
	PB	0,00	0,00	0,01	0,01	0,00	0,00	0,00	0,00	

Capítulo 3

Distribuição Poisson Generalizada Bivariada

3.1 Distribuição Poisson generalizada

Quando trabalhamos com dados de contagem, geralmente o modelo Poisson é considerado, sabe-se que a média e a variância são iguais. Na prática são encontradas aplicações que não obedecem essa condição sobre a igualdade entre a variância e a média; uma, por exemplo, é a distribuição do número de organismos em uma colônia de bactérias, onde a taxa de nascimento de qualquer lugar no espaço ou tempo é dependente do número anterior de organismos próximos daquele lugar em questão (devido à disponibilidade de comida, temperatura, comportamento social, etc).

Numerosos investigadores tentaram explicar essa divergência do modelo Poisson, sugerindo algumas generalizações. Em particular, Consul & Jain (1973) propuseram a distribuição de Poisson generalizada, a qual tem um parâmetro adicional que a torna mais flexível a dados observados que tendem a ser sobredispersos. Sobredispersão é um fenômeno comum que ocorre na modelagem de dados discretos e cuja ocorrência é caracterizada quando a variância é maior que a média.

Uma variável aleatória X tem uma distribuição Poisson generalizada se a f.p. é dada por

$$p(x|\lambda, \theta) = \frac{\lambda(\lambda + \theta x)^{x-1} e^{-(\lambda + \theta x)}}{x!}, \quad x = 0, 1, 2, \dots, \quad (3.1)$$

em que $\lambda > 0$ e $0 \leq \theta < 1$. A média e a variância da variável X são, respectivamente,

$$\begin{aligned} E[X] &= \lambda[1 - \theta]^{-1}, \\ V[X] &= \lambda[1 - \theta]^{-3}. \end{aligned}$$

Nota-se que a variância é maior que a média quando $\theta > 0$ e se $\theta = 0$, o modelo Poisson generalizado corresponde ao modelo Poisson clássico com $E(X) = V(X) = \lambda$.

3.2 Distribuição Poisson generalizada bivariada

A distribuição Poisson generalizada bivariada proposta por Famoye & Consul (1995) é construída pelo método de redução trivariada aplicada a distribuição Poisson generalizada.

Supondo V_1 , V_2 e U variáveis aleatórias independentes com distribuições Poisson generalizada com vetor de parâmetros (λ_1, θ_1) , (λ_2, θ_2) , (λ_3, θ_3) , respectivamente definimos o vetor aleatório (X, Y) como

$$\begin{aligned} X &= V_1 + U, \\ Y &= V_2 + U. \end{aligned}$$

Com isso, a f.p. conjunta do vetor (X, Y) é

$$\begin{aligned} p(x, y) &= P[X = x, Y = y] \\ &= P[V_1 + U = x, V_2 + U = y] \\ &= \sum_{k=0}^{\infty} P[V_1 - U = x, V_2 - U = y | U = k] P[U = k] \\ &= \sum_{k=0}^{\min(x, y)} P[V_1 = x - k] P[V_2 = y - k] P[U = k] \\ &= \lambda_1 \lambda_2 \lambda_3 \exp\{-(\lambda_1 + \lambda_2 + \lambda_3 + \theta_1 x + \theta_2 y)\} H(\boldsymbol{\lambda}, \boldsymbol{\theta}), \quad x, y = 0, 1, \dots, \end{aligned} \quad (3.2)$$

em que

$$H(\boldsymbol{\lambda}, \boldsymbol{\theta}) = \sum_{k=0}^{\min(x, y)} \frac{[\lambda_1 + \theta_1(x - k)]^{x-k-1} [\lambda_2 + \theta_2(y - k)]^{y-k-1} [\lambda_3 + \theta_3 k]^{k-1}}{(x - k)! (y - k)! k!} e^{k(\theta_1 + \theta_2 - \theta_3)},$$

com $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3)$, $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$, sendo $(\boldsymbol{\lambda}, \boldsymbol{\theta}) \in B_1 \cup B_2$ onde $B_1 = (0, +\infty)^3 \times [0, 1]^3$ e $B_2 = (0, \infty)^2 \times \{0\} \times [0, 1]^2 \times \{0\}$. No caso em que $(\boldsymbol{\lambda}, \boldsymbol{\theta}) \in B_2$, temos que as variáveis V_1, V_2 tem distribuição Poisson generalizada com respectivos parâmetros (λ_1, θ_1) , (λ_2, θ_2) e $U = 0$ (degenerada em 0), o que implica que as variáveis X e Y são independentes.

Note que se $\theta_1 = \theta_2 = \theta_3 = 0$, a f.p. (3.2) se reduz a distribuição PB.

As f.p. marginais de X e Y são dadas, respectivamente, por

$$P(X = x) = \lambda_1 \lambda_3 e^{-(\lambda_1 + \lambda_3 + \theta_3 x)} \sum_{v_1=0}^x \frac{[\lambda_1 + \theta_1 v_1]^{v_1-1} [\lambda_3 + \theta_3(x - v_1)]^{x-v_1-1} e^{-v_1(\theta_1 - \theta_3)}}{(x - v_1)! v_1!}, \quad x = 0, 1, \dots$$

e

$$P(Y = y) = \lambda_2 \lambda_3 e^{-(\lambda_2 + \lambda_3 + \theta_3 y)} \sum_{v_2=0}^y \frac{[\lambda_2 + \theta_2 v_2]^{v_2-1} [\lambda_3 + \theta_3(y - v_2)]^{y-v_2-1} e^{-v_2(\theta_1 - \theta_3)}}{(y - v_2)! v_2!}, \quad y = 0, 1, \dots$$

A média das variáveis X e Y são dadas, respectivamente, por

$$\begin{aligned} E(X) &= \lambda_1[1 - \theta_1]^{-1} + \lambda_3[1 - \theta_3]^{-1} \quad e \\ E(Y) &= \lambda_2[1 - \theta_2]^{-1} + \lambda_3[1 - \theta_3]^{-1}, \end{aligned}$$

as correspondentes variâncias são respectivamente

$$\begin{aligned} V(X) &= \lambda_1[1 - \theta_1]^{-3} + \lambda_3[1 - \theta_3]^{-3} \quad e \\ V(Y) &= \lambda_2[1 - \theta_2]^{-3} + \lambda_3[1 - \theta_3]^{-3}, \end{aligned}$$

enquanto que a covariância entre X e Y é dada por

$$\text{Cov}[X, Y] = \text{Cov}[V_1 + U, V_2 + U] = V(U) = \lambda_3[1 - \theta_3]^{-3},$$

e o coeficiente de correlação é dado por

$$\begin{aligned} \text{Corr}[X, Y] &= \frac{\text{Cov}[X, Y]}{\sqrt{V(X)V(Y)}} \\ &= \frac{\lambda_3[1 - \theta_3]^{-3}}{\sqrt{(\lambda_1[1 - \theta_1]^{-3} + \lambda_3[1 - \theta_3]^{-3})(\lambda_2[1 - \theta_2]^{-3} + \lambda_3[1 - \theta_3]^{-3})}} \\ &= \lambda_3 \{[\lambda_1(1 - \theta_1)^{-3}(1 - \theta_3)^3 + \lambda_3][\lambda_2(1 - \theta_2)^{-3}(1 - \theta_3)^3 + \lambda_3]\}^{-1/2}. \end{aligned}$$

Note que o coeficiente de correlação assume somente valores positivos pois $\lambda_i > 0$ e $0 \leq \theta_i < 1$, para $i = 1, 2, 3$.

Para o modelo PGB se $\lambda_3 = \theta_3 = 0$, as variáveis X e Y são independentes e a $\text{Corr}[X, Y] = 0$, neste caso, a f.p. (3.2) é o produto de duas distribuição Poisson generalizadas independentes, ou seja,

$$\begin{aligned} p(x, y) &= P[X = x, Y = y] \\ &= P[X = x] P[Y = y] \\ &= \frac{\lambda_1 \lambda_2 e^{-(\lambda_1 + \lambda_2 + \theta_1 x + \theta_2 y)} (\lambda_1 + \theta_1 x)^{x-1} (\lambda_2 + \theta_2 y)^{y-1}}{x! y!}, \quad x, y = 0, 1, \dots, \end{aligned} \quad (3.3)$$

que neste trabalho chamamos de distribuição Poisson generalizada dupla.

3.3 Estimação pontual

Nesta seção apresentaremos dois métodos de estimação para os parâmetros do modelo PGB, os quais foram utilizados em Famoye & Consul (1995). No que segue vamos considerar uma amostra aleatória

$(X_1, Y_1), \dots, (X_n, Y_n)$ de tamanho n tendo distribuição PGB. Assumiremos que a frequência de cada par (i, j) é denotada por $n_{ij} = \sum_{\kappa=1}^n I_{\{(i,j)\}}(X_\kappa, Y_\kappa)$, para $i, j = 0, 1, \dots$ e $D = \{(i, j) : n_{ij} > 0\}$ como o conjunto dos pares (i, j) observados. Logo, o tamanho da amostra é dado por

$$n = \sum_{(i,j) \in D} n_{ij}.$$

3.3.1 Método dos momentos

Os momentos populacionais da distribuição PGB são dados, respectivamente por

$$\begin{aligned} \mu'_{1,0} &= \lambda_1[1 - \theta_1]^{-1} + \lambda_3[1 - \theta_3]^{-1}, \\ \mu'_{0,1} &= \lambda_2[1 - \theta_2]^{-1} + \lambda_3[1 - \theta_3]^{-1}, \\ \mu_{2,0} &= \lambda_1[1 - \theta_1]^{-3} + \lambda_3[1 - \theta_3]^{-3}, \\ \mu_{0,2} &= \lambda_2[1 - \theta_2]^{-3} + \lambda_3[1 - \theta_3]^{-3}, \\ \mu_{2,1} &= \lambda_3(3[1 - \theta_3]^{-1} - 2)[1 - \theta_3]^{-4}, \quad e \\ \mu_{1,1} &= \lambda_3[1 - \theta_3]^{-3}, \end{aligned}$$

enquanto que os correspondentes momentos amostrais são dados por

$$\begin{aligned} m'_{1,0} &= \bar{X} = \frac{1}{n} \sum_{(i,j) \in D} i n_{ij}, \\ m'_{0,1} &= \bar{Y} = \frac{1}{n} \sum_{(i,j) \in D} j n_{ij}, \\ m_{2,0} &= \frac{1}{n} \sum_{(i,j) \in D} (i - m'_{1,0})^2 n_{ij}, \\ m_{0,2} &= \frac{1}{n} \sum_{(i,j) \in D} (j - m'_{0,1})^2 n_{ij}, \\ m_{2,1} &= \frac{1}{n} \sum_{(i,j) \in D} (i - m'_{1,0})^2 (j - m'_{0,1}) n_{ij}, \quad e \\ m_{1,1} &= \frac{1}{n} \sum_{(i,j) \in D} (i - m'_{1,0})(j - m'_{0,1}) n_{ij}. \end{aligned}$$

Igualando os momentos populacionais com seus respectivos momentos amostrais obtemos as seguintes

equações

$$\tilde{\lambda}_1[1 - \tilde{\theta}_1]^{-1} + \tilde{\lambda}_3[1 - \tilde{\theta}_3]^{-1} = \bar{X}, \quad (3.4)$$

$$\tilde{\lambda}_2[1 - \tilde{\theta}_2]^{-1} + \tilde{\lambda}_3[1 - \tilde{\theta}_3]^{-1} = \bar{Y}, \quad (3.5)$$

$$\tilde{\lambda}_1[1 - \tilde{\theta}_1]^{-3} + \tilde{\lambda}_3[1 - \tilde{\theta}_3]^{-3} = m_{2,0}, \quad (3.6)$$

$$\tilde{\lambda}_2[1 - \tilde{\theta}_2]^{-3} + \tilde{\lambda}_3[1 - \tilde{\theta}_3]^{-3} = m_{0,2}, \quad (3.7)$$

$$\tilde{\lambda}_3(3[1 - \tilde{\theta}_3]^{-1} - 2)[1 - \tilde{\theta}_3]^{-4} = m_{2,1}, \quad \text{e} \quad (3.8)$$

$$\tilde{\lambda}_3[1 - \tilde{\theta}_3]^{-3} = m_{1,1}, \quad (3.9)$$

nas quais $\tilde{\lambda}_1, \tilde{\lambda}_2, \tilde{\lambda}_3, \tilde{\theta}_1, \tilde{\theta}_2$ e $\tilde{\theta}_3$ são estimadores de $\lambda_1, \lambda_2, \lambda_3, \theta_1, \theta_2$ e θ_3 , respectivamente.

A partir das equações (3.4), (3.5) e (3.9), obtemos os estimadores $\tilde{\lambda}_1, \tilde{\lambda}_2$ e $\tilde{\lambda}_3$ dados, respectivamente, por

$$\tilde{\lambda}_1 = \frac{\bar{X} - \tilde{\lambda}_3[1 - \tilde{\theta}_3]^{-1}}{[1 - \tilde{\theta}_1]^{-1}}, \quad (3.10)$$

$$\tilde{\lambda}_2 = \frac{\bar{Y} - \tilde{\lambda}_3[1 - \tilde{\theta}_3]^{-1}}{[1 - \tilde{\theta}_2]^{-1}}, \quad \text{e} \quad (3.11)$$

$$\tilde{\lambda}_3 = \frac{m_{1,1}}{[1 - \tilde{\theta}_3]^{-3}}. \quad (3.12)$$

Substituindo (3.9) em (3.6), e utilizando (3.10) temos

$$\begin{aligned} [1 - \tilde{\theta}_1]^{-3} &= \frac{m_{2,0} - m_{1,1}}{\tilde{\lambda}_1} = \frac{(m_{2,0} - m_{1,1})[1 - \tilde{\theta}_1]^{-1}}{\bar{X} - \tilde{\lambda}_3[1 - \tilde{\theta}_3]^{-1}} \Rightarrow \\ [1 - \tilde{\theta}_1]^{-2} &= \frac{m_{2,0} - m_{1,1}}{\bar{X} - \tilde{\lambda}_3[1 - \tilde{\theta}_3]^{-1}} \Rightarrow \\ [1 - \tilde{\theta}_1]^{-1} &= \sqrt{\frac{m_{2,0} - m_{1,1}}{\bar{X} - \tilde{\lambda}_3[1 - \tilde{\theta}_3]^{-1}}}. \end{aligned}$$

Substituindo (3.9) em (3.7), e utilizando (3.11) temos

$$\begin{aligned} [1 - \tilde{\theta}_2]^{-3} &= \frac{m_{0,2} - m_{1,1}}{\tilde{\lambda}_2} = \frac{(m_{0,2} - m_{1,1})[1 - \tilde{\theta}_2]^{-1}}{\bar{Y} - \tilde{\lambda}_3[1 - \tilde{\theta}_3]^{-1}} \Rightarrow \\ [1 - \tilde{\theta}_2]^{-2} &= \frac{m_{0,2} - m_{1,1}}{\bar{Y} - \tilde{\lambda}_3[1 - \tilde{\theta}_3]^{-1}} \Rightarrow \\ [1 - \tilde{\theta}_2]^{-1} &= \sqrt{\frac{m_{0,2} - m_{1,1}}{\bar{Y} - \tilde{\lambda}_3[1 - \tilde{\theta}_3]^{-1}}}. \end{aligned}$$

Substituindo (3.12) em (3.8), obtemos

$$\begin{aligned} \frac{m_{1,1}(3[1 - \tilde{\theta}_3]^{-1} - 2)[1 - \tilde{\theta}_3]^{-4}}{[1 - \tilde{\theta}_3]^{-3}} &= m_{2,1} \Rightarrow \\ 3[1 - \tilde{\theta}_3]^{-2} - 2[1 - \tilde{\theta}_3]^{-1} &= \frac{m_{2,1}}{m_{1,1}} \Rightarrow \\ 3 \left([1 - \tilde{\theta}_3]^{-1} \right)^2 - 2 \left([1 - \tilde{\theta}_3]^{-1} \right) - a &= 0, \end{aligned} \quad (3.13)$$

sendo

$$a = \frac{m_{2,1}}{m_{1,1}}.$$

Resolvendo a equação quadrática (3.13), obtemos as soluções

$$[1 - \tilde{\theta}_3]^{-1} = \frac{1 \pm \sqrt{1 + 3a}}{3}.$$

Como $0 \leq \theta_3 < 1$, então $0 < [1 - \tilde{\theta}_3]^{-1}$. Logo,

$$[1 - \tilde{\theta}_3]^{-1} = \frac{1 + \sqrt{1 + 3a}}{3}.$$

Com isso, a partir das equações (3.10), (3.11) e (3.12), obtemos os estimadores $\tilde{\theta}_1$, $\tilde{\theta}_2$ e $\tilde{\theta}_3$ dados, respectivamente, por

$$\begin{aligned} \tilde{\theta}_1 &= 1 - \frac{\tilde{\lambda}_1}{\bar{X} - \tilde{\lambda}_3[1 - \tilde{\theta}_3]^{-1}}, \\ \tilde{\theta}_2 &= 1 - \frac{\tilde{\lambda}_2}{\bar{Y} - \tilde{\lambda}_3[1 - \tilde{\lambda}_3]^{-1}}, \quad \text{e} \\ \tilde{\theta}_3 &= 1 - \sqrt[3]{\frac{\tilde{\lambda}_3}{m_{1,1}}}. \end{aligned}$$

3.3.2 Método de máxima verossimilhança

A função de verossimilhança é dada por

$$L(\boldsymbol{\lambda}, \boldsymbol{\theta}) = \prod_{(i,j) \in D} [p(i, j)]^{n_{ij}},$$

onde $(\boldsymbol{\lambda}, \boldsymbol{\theta}) = (\lambda_1, \lambda_2, \lambda_3, \theta_1, \theta_2, \theta_3)$ e seu logaritmo é

$$\begin{aligned} l(\boldsymbol{\lambda}, \boldsymbol{\theta}) &= \log \left\{ \prod_{(i,j) \in D} [p(i,j)]^{n_{ij}} \right\} \\ &= \sum_{(i,j) \in D} n_{ij} \log p(i,j) \\ &= \sum_{(i,j) \in D} n_{ij} \{ \log \lambda_1 + \log \lambda_2 + \log \lambda_3 - (\lambda_1 + \lambda_2 + \lambda_3 + \theta_1 i + \theta_2 j) + \log H(\boldsymbol{\theta}, \boldsymbol{\lambda}) \}. \end{aligned}$$

Calculando as derivadas parciais do logaritmo da função de verossimilhança com respeito aos parâmetros λ_ℓ e θ_ℓ , para $\ell = 1, 2, 3$, e igualando elas a zero, obtemos as equações

$$\begin{aligned} \frac{\partial l(\boldsymbol{\lambda}, \boldsymbol{\theta})}{\partial \lambda_\ell} &= \sum_{(i,j) \in D} n_{ij} \left\{ \frac{1 - \lambda_\ell}{\lambda_\ell} + \frac{\partial \log H(\boldsymbol{\lambda}, \boldsymbol{\theta})}{\partial \lambda_\ell} \right\} = 0, \quad \ell = 1, 2, 3, \\ \frac{\partial l(\boldsymbol{\lambda}, \boldsymbol{\theta})}{\partial \theta_1} &= \sum_{(i,j) \in D} n_{ij} \left\{ -i + \frac{\partial \log H(\boldsymbol{\lambda}, \boldsymbol{\theta})}{\partial \theta_1} \right\} = 0, \\ \frac{\partial l(\boldsymbol{\lambda}, \boldsymbol{\theta})}{\partial \theta_2} &= \sum_{(i,j) \in D} n_{ij} \left\{ -j + \frac{\partial \log H(\boldsymbol{\lambda}, \boldsymbol{\theta})}{\partial \theta_2} \right\} = 0, \\ \frac{\partial l(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \theta_3} &= \sum_{(i,j) \in D} n_{ij} \frac{\partial \log H(\boldsymbol{\lambda}, \boldsymbol{\theta})}{\partial \theta_3} = 0. \end{aligned}$$

As estimativas de máxima verossimilhança dos parâmetros do modelo são obtidas a partir das equações acima. Observamos que as equações não tem solução explícita; no entanto podem ser resolvidas por meio de métodos numéricos. Nos usamos o método de Newton-Raphson neste caso.

3.4 Teste de hipótese

Nesta seção, iremos utilizar o teste da razão de verossimilhanças para testar as hipóteses de interesse.

3.4.1 Teste de independência

As variáveis aleatórias X e Y são independentes quando os parâmetros $\lambda_3 = 0$ e $\theta_3 = 0^1$. Para testar a hipótese de independência, devemos considerar as hipóteses

$$\begin{aligned} H_0 : \lambda_3 = \theta_3 = 0 \\ \text{versus } H_1 : \lambda_3 + \theta_3 > 0, \end{aligned}$$

¹Quando $\lambda_3 = 0$ e $\theta_3 = 0$ a variável U será degenerada no ponto 0, isto é $P[U = 0] = 1$.

e a estatística da razão de verossimilhanças dada por

$$\Psi = \frac{\sup_{(\boldsymbol{\lambda}, \boldsymbol{\theta}) \in B_2} L(\boldsymbol{\lambda}, \boldsymbol{\theta})}{\sup_{(\boldsymbol{\lambda}, \boldsymbol{\theta}) \in B_1 \cup B_2} L(\boldsymbol{\lambda}, \boldsymbol{\theta})},$$

na qual $B_1 = (0, +\infty)^3 \times [0, 1)^3$ e $B_2 = (0, \infty)^2 \times \{0\} \times [0, 1)^2 \times \{0\}$. A estatística $-2 \log(\Psi)$ segue, assintoticamente, uma distribuição qui-quadrado com 2 graus de liberdade. A região crítica do teste é, portanto, $C = \{-2 \log(\Psi) \geq \chi_{2, \alpha}^2\}$, onde $\chi_{2, \alpha}^2$ é o ponto crítico da χ_2^2 correspondente ao nível de significância α .

3.4.2 Teste para a distribuição PB

A distribuição PGB se reduz a distribuição PB quando $\theta_1 = \theta_2 = \theta_3 = 0$. Para avaliar a adequação da distribuição PB aos dados vamos considerar as hipóteses

$$H_0 : \theta_1 = \theta_2 = \theta_3 = 0 \\ \text{versus } H_1 : \theta_1 + \theta_2 + \theta_3 > 0,$$

e a estatística da razão de verossimilhanças dada por

$$\Psi' = \frac{\sup_{\boldsymbol{\lambda} \in \Lambda} L^{\text{PB}}(\boldsymbol{\lambda})}{\sup_{(\boldsymbol{\lambda}, \boldsymbol{\theta}) \in B_1 \cup B_2} L(\boldsymbol{\lambda}, \boldsymbol{\theta})},$$

em que $L^{\text{PB}}(\boldsymbol{\lambda}) = L(\lambda_1, \lambda_2, \lambda_3, 0, 0, 0)$ é a função de verossimilhança para o modelo restrito, PB, com $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3) \in \Lambda$. A estatística $-2 \log(\Psi')$ segue, assintoticamente, uma distribuição qui-quadrado com 3 graus de liberdade. A região crítica do teste é, portanto, $C = \{-2 \log(\Psi') \geq \chi_{3, \alpha}^2\}$, onde $\chi_{3, \alpha}^2$ é o ponto crítico da χ_3^2 correspondente ao nível de significância α .

3.5 Aplicação

Neste seção, utilizaremos o conjunto de dados reais apresentados na Tabela 2.1 da Seção 2.5. Este conjunto de dados apresentou variâncias amostrais maiores que as médias amostrais, e uma covariância amostral positiva o que indica a sobredispersão dos dados e que existe uma associação positiva entre o número de acidentes no primeiro período e o segundo período. A análise descritiva dos dados sugere que o modelo PGB pode ser adequado para este conjunto de dados.

Na Tabela 3.1 estão apresentadas as estimativas dos parâmetros do modelo PGB, que foram obtidas pelos métodos dos momentos e de máxima verossimilhança, e os erros-padrão dos estimadores de máxima verossimilhança obtidos através da inversa da matriz de informação observada. Observamos uma proximidade entre as estimativas obtidas por ambos os métodos de estimação.

Tabela 3.1: Estimativas dos parâmetros do modelo PGB

Parâmetro	MM	EMV	EP
λ_1	0,6206	0,6767	0,0570
λ_2	0,8653	0,9456	0,0628
λ_3	0,2987	0,2310	0,0460
θ_1	0,1057	0,0949	0,0336
θ_2	0,1200	0,0885	0,0313
θ_3	0,0286	0,0892	0,0591

Uma questão de interesse é testar se as variáveis X e Y são independentes, isto é, se o modelo com independência é adequado. Para tanto, utilizaremos os critérios AIC e BIC e realizamos o teste de razão de verossimilhanças descrito na Subseção 3.4.1. A Tabela 3.2 apresenta os valores destes critérios para os modelos considerados, em que se observa um melhor ajuste do modelo PGB.

Tabela 3.2: Critérios de seleção de modelos

Modelo	AIC	BIC
Poisson generalizado bivariado	4005,25	4018,93
Poisson generalizado duplo	4043,66	4061,91

Para este conjunto de dados obtivemos a estatística $-2\log(\Psi) = 36,41$. O percentil da χ_2^2 , considerando o nível 0,1% de significância, tem valor crítico do teste igual a 13,81 e, observa-se que o valor obtido é maior que o valor crítico. Logo, rejeitamos a hipótese de que as variáveis X e Y são independentes e, desta forma, vale ressaltar a importância de considerar a associação entre as variáveis aleatórias.

A Tabela 3.3 apresenta as frequências observadas e esperadas de acordo com os modelos PGB e PB, cujos parâmetros foram estimados pelo método de máxima verossimilhança. Em geral, observamos que as frequências esperadas do modelo PGB estão mais próximas das frequências observadas do que as do modelo PB, notadamente para as frequências dos pares $(0,0)$ e $(0,1)$. Assim, o modelo de PGB se ajusta melhor a este conjunto de dados reais.

Tabela 3.3: Frequências observadas e esperadas para os modelos PGB e PB

x	y =	0	1	2	3	4	5	6	7	Total
0	Obs	117	96	55	19	2	0	0	0	291
	PGB	110,94	96,04	49,34	19,66	6,72	2,08	0,59	0,16	
	PB	89,94	95,50	50,70	17,94	4,76	1,01	0,17	0,02	
1	Obs	61	69	47	27	8	5	1	0	218
	PGB	68,27	82,55	50,66	22,52	8,29	2,70	0,80	0,22	
	PB	69,46	94,36	61,04	25,47	7,79	1,87	0,37	0,06	
2	Obs	34	42	31	13	7	2	3	0	132
	PGB	26,90	37,72	28,84	14,98	6,14	2,15	0,68	0,20	
	PB	26,82	44,39	34,38	16,83	5,92	1,61	0,35	0,06	
3	Obs	7	15	16	7	3	1	0	0	49
	PGB	8,70	13,21	11,49	7,22	3,44	1,34	0,46	0,14	
	PB	6,90	13,47	12,24	6,95	2,81	0,86	0,21	0,04	
4	Obs	3	3	1	1	2	1	1	1	13
	PGB	2,52	4,02	3,77	2,68	1,53	0,70	0,26	0,09	
	PB	6,90	13,47	12,24	6,95	2,81	0,86	0,21	0,04	
5	Obs	2	1	0	0	0	0	0	0	3
	PGB	0,68	1,12	1,11	0,85	0,55	0,30	0,13	0,05	
	PB	0,20	0,52	0,62	0,45	0,23	0,09	0,02	0,00	
6	Obs	0	0	0	0	1	0	0	0	1
	PGB	0,17	0,29	0,30	0,24	0,17	0,10	0,05	0,02	
	PB	0,02	0,07	0,10	0,08	0,04	0,00	0,00	0,00	
7	Obs	0	0	0	1	0	0	0	0	1
	PGB	0,04	0,07	0,07	0,06	0,04	0,03	0,02	0,01	
	PB	0,00	0,00	0,01	0,01	0,00	0,00	0,00	0,00	

Uma alternativa para a comparação dos modelos propostos PB e PGB é utilizar os critérios de seleção AIC e BIC. A Tabela 3.4 apresenta os valores destes critérios para os modelos considerados, em que se observa um melhor ajuste do modelo PGB.

Tabela 3.4: Critérios de seleção de modelos

Modelo	AIC	BIC
PGB	4005,25	4018,93
PB	4028,40	4037,53

Outra forma de comparação dos modelos é dada pelo teste da razão de verossimilhança discutido na Subseção 3.4.2, que fornece $-2\log(\Psi') = 25,16$. Considerando o nível 0,1% de significancia tem-se valor crítico do teste igual a 16,26 e, observa-se que o valor obtido é maior que o valor crítico, indicando forte evidência contra do modelo PB. Portanto, pode-se concluir que o modelo PGB se ajustou melhor que o modelo PB para este conjunto de dados.

Capítulo 4

Distribuição Poisson Bivariada Inflacionada de Zeros

4.1 Distribuição Poisson inflacionada de zeros

O excesso de zeros em dados de contagem é uma ocorrência muito comum em experimentos de diversas áreas, tais como epidemiologia, medicina, odontologia, agricultura e indústria. Esses zeros podem ter diferentes origens. Ridout *et al.* (1998) mencionaram duas classificações: zeros estruturais, os zeros que são inevitáveis e zeros amostrais os que podem ocorrer ao acaso. Para exemplificar a diferença entre as duas classificações, em um estudo sobre a quantidade de lesões em plantas, Ridout *et al.* (1998) asseguravam que uma planta pode não apresentar a lesão porque é resistente aquela doença (zero estrutural) ou simplesmente porque o agente causador da doença não recaiu sobre aquela planta (zero amostral).

Johnson *et al.* (2005) discutiram uma maneira simples de modificar uma distribuição discreta para modelar os zeros extras, através de uma mistura finita de distribuições. Por exemplo, o modelo Poisson inflacionado de zeros (ZIP, do inglês zero inflated Poisson) é construído como uma mistura da distribuição Poisson e uma distribuição degenerada em zero. Este modelo foi discutido por diversos autores, entre eles, Lambert (1992) quem ajustou o modelo Poisson inflacionado de zeros a dados sobre número de defeitos em itens industriais. Van den Broek (1995) apresentou uma estatística escore para testar se o modelo Poisson inflacionado de zeros se ajusta melhor a um conjunto de dados que um modelo de Poisson clássico.

Segundo Ridout *et al.* (1998), a distribuição Poisson inflacionada de zeros considera que alguns zeros, estruturais ocorrem com probabilidade ω e os zeros amostrais com probabilidade $(1 - \omega)$. Bayarri *et al.* (2008) referem-se a ω como parâmetro de inflação de zero e definem a f.p. da distribuição ZIP como

$$P_{\text{ZIP}}(X = x|\omega, \lambda) = \omega I_{\{0\}}(x) + (1 - \omega) \frac{e^{-\lambda} \lambda^x}{x!}, \quad x = 0, 1, \dots, \quad (4.1)$$

em que

$$I_{\{0\}}(x) = \begin{cases} 1, & \text{se } x = 0, \\ 0, & \text{se } x \neq 0, \end{cases}$$

onde $0 \leq \omega \leq 1$, $\lambda > 0$.

A média e a variância da distribuição Poisson inflacionada de zeros são dadas por

$$\begin{aligned} E[X] &= (1 - \omega)\lambda, \\ V[X] &= E[X](1 + \omega\lambda). \end{aligned}$$

Note que a variância é maior que a média, ou seja, a distribuição ZIP contempla a superdispersão gerada pelo excesso de zeros. Quanto maior a probabilidade de zeros, ω , maior variância. Quando $\omega = 0$ o modelo corresponde ao modelo Poisson usual com $E[X] = V[X] = \lambda$.

4.2 Distribuição Poisson bivariada inflacionada de zeros

Para dados de contagem bivariados com excessos de zeros, a distribuição BZIP pode ser construída como uma mistura da distribuição PB e uma distribuição degenerada em $(0, 0)$. A f.p. da distribuição BZIP para o vetor (X, Y) é dada por

$$P_{\text{BZIP}}(X = x, Y = y | \boldsymbol{\delta}) = \omega I_{\{(0,0)\}}(x, y) + (1 - \omega)p(x, y | \boldsymbol{\lambda}), \quad x, y = 0, 1, \dots, \quad (4.2)$$

em que

$$I_{\{(0,0)\}}(x, y) = \begin{cases} 1, & \text{se } (x, y) = (0, 0), \\ 0, & \text{se } (x, y) \neq (0, 0), \end{cases}$$

sendo $\boldsymbol{\delta} = (\omega, \boldsymbol{\lambda})$, $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3)$, $\boldsymbol{\delta} \in \Delta = [0, 1] \times (0, +\infty)^2 \times \{0\}$ e $p(x, y | \boldsymbol{\lambda})$ é a f.p. da distribuição PB dada por

$$p(x, y | \boldsymbol{\lambda}) = e^{-(\lambda_1 + \lambda_2 + \lambda_3)} \sum_{u=0}^{\min(x,y)} \frac{\lambda_1^{x-u} \lambda_2^{y-u} \lambda_3^u}{(x-u)!(y-u)!u!}.$$

O parâmetro ω é a probabilidade de zeros de forma determinística e $(1 - \omega)$ é a probabilidade de zeros aleatórios gerados pela distribuição PB, a qual se adequaria aos dados caso não houvesse a presença excessiva de zeros nos dados.

Temos os seguintes casos particulares:

1. Para $w = 0$, o modelo (4.2) é reduzido a distribuição PB;
2. Para $\lambda_3 = 0$, o modelo (4.2) é reduzido a distribuição ZIPD;
3. Quando $w = 0$ e $\lambda_3 = 0$, o modelo (4.2) é reduzido a distribuição PD.

A f.g.p. de (X, Y) , com distribuição BZIP é dada por

$$\begin{aligned}
g_{X,Y}(t_1, t_2) &= E[t_1^X t_2^Y] \\
&= \sum_{x=0}^{\infty} \sum_{y=0}^{\infty} t_1^x t_2^y P(X = x, Y = y) \\
&= t_1^0 t_2^0 \left\{ \omega + (1 - \omega) e^{-(\lambda_1 + \lambda_2 + \lambda_3)} \right\} + \sum_{x=1}^{\infty} \sum_{y=1}^{\infty} (1 - \omega) p(x, y | \boldsymbol{\lambda}) \\
&= \omega + (1 - \omega) e^{-(\lambda_1 + \lambda_2 + \lambda_3)} + (1 - \omega) \sum_{x=1}^{\infty} \sum_{y=1}^{\infty} t_1^x t_2^y p(x, y | \boldsymbol{\lambda}) \\
&= \omega + (1 - \omega) \sum_{x=0}^{\infty} \sum_{y=0}^{\infty} t_1^x t_2^y p(x, y | \boldsymbol{\lambda}) \\
&= \omega + (1 - \omega) e^{\lambda_1(t_1-1) + \lambda_2(t_2-1) + \lambda_3(t_1 t_2 - 1)}, \quad t_1, t_2 \in \mathbb{R}.
\end{aligned} \tag{4.3}$$

As f.p. marginais de X e Y são dadas respectivamente por

$$P_{\text{BZIP}}(X = x) = \omega I_{\{0\}}(x) + (1 - \omega) \frac{e^{-(\lambda_1 + \lambda_3)} (\lambda_1 + \lambda_3)^x}{x!}, \quad x = 0, 1, 2, \dots,$$

e

$$P_{\text{BZIP}}(Y = y) = \omega I_{\{0\}}(y) + (1 - \omega) \frac{e^{-(\lambda_1 + \lambda_3)} (\lambda_1 + \lambda_3)^y}{y!}, \quad y = 0, 1, 2, \dots$$

Note que a distribuição marginal de X e Y é ZIP, com parâmetros $\lambda_1 + \lambda_3$ e $\lambda_2 + \lambda_3$, respectivamente. Logo, a média e a variância das variáveis X e Y são dadas por

$$\begin{aligned}
E[X] &= (1 - \omega)[\lambda_1 + \lambda_3], & V[X] &= E[X][1 + \omega(\lambda_1 + \lambda_3)], \\
E[Y] &= (1 - \omega)[\lambda_2 + \lambda_3], & V[Y] &= E[Y][1 + \omega(\lambda_2 + \lambda_3)].
\end{aligned}$$

Para se obter a covariância entre X e Y , é preciso calcular $E[XY]$ utilizando a f.g.p. bivariada, $g_{X,Y}(t_1, t_2) = \sum_{x=0}^{\infty} \sum_{y=0}^{\infty} P(X = x, Y = y) t_1^x t_2^y$ e

$$\frac{\partial^2 g_{X,Y}(t_1, t_2)}{\partial t_1 \partial t_2} = \sum_{x=0}^{\infty} \sum_{y=0}^{\infty} xy P(X = x, Y = y) t_1^{x-1} t_2^{y-1},$$

para $(t_1, t_2) \in I \times J \subseteq \mathbb{R}^2$.

Fazendo $t_1 = t_2 = 1$, temos

$$\left[\frac{\partial^2 g_{X,Y}(t_1, t_2)}{\partial t_1 \partial t_2} \right]_{t_1=t_2=1} = \sum_{x=0}^{\infty} \sum_{y=0}^{\infty} xy P(X = x, Y = y) = E[XY].$$

Assim, usando a f.g.p. dada em (4.3), temos

$$E[XY] = \left[\frac{\partial^2 g_{X,Y}(t_1, t_2)}{\partial t_1 \partial t_2} \right]_{t_1=t_2=1} = (1 - \omega) \{ \lambda_3 + (\lambda_1 + \lambda_3)(\lambda_2 + \lambda_3) \}.$$

Desta forma, a covariância entre X e Y é dada por

$$\begin{aligned} \text{Cov}(X, Y) &= E[XY] - E[X]E[Y] \\ &= (1 - \omega) \{ \lambda_3 + (\lambda_1 + \lambda_3)(\lambda_2 + \lambda_3) \} - (1 - \omega)^2 (\lambda_1 + \lambda_3)(\lambda_2 + \lambda_3) \\ &= (1 - \omega) \{ \lambda_3 + \omega(\lambda_1 + \lambda_3)(\lambda_2 + \lambda_3) \}. \end{aligned}$$

4.3 Estimação pontual

4.3.1 Método de máxima verossimilhança

Considere uma amostra aleatória $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ de tamanho n da distribuição BZIP. A função de verossimilhança é dada por

$$\begin{aligned} L(\boldsymbol{\delta}) &= \prod_{i=1}^n [\text{P}_{\text{BZIP}}(x_i, y_i | \boldsymbol{\delta})]^{I_{\{(0,0)\}}(x_i, y_i)} [\text{P}_{\text{BZIP}}(x_i, y_i | \boldsymbol{\delta})]^{1 - I_{\{(0,0)\}}(x_i, y_i)} \\ &= \prod_{i=1}^n \left[\omega + (1 - \omega)e^{-\lambda_{\bullet}} \right]^{I_{\{(0,0)\}}(x_i, y_i)} [(1 - \omega)p(x_i, y_i | \boldsymbol{\lambda})]^{1 - I_{\{(0,0)\}}(x_i, y_i)} \\ &= \left[\omega + (1 - \omega)e^{-\lambda_{\bullet}} \right]^{m_0} (1 - \omega)^{n - m_0} \prod_{i=1}^n [p(x_i, y_i | \boldsymbol{\lambda})]^{1 - I_{\{(0,0)\}}(x_i, y_i)}, \end{aligned} \quad (4.4)$$

sendo $x_i, y_i = 0, 1, \dots$, $m_0 = \sum_{i=1}^n I_{\{(0,0)\}}(x_i, y_i)$ e $\lambda_{\bullet} = (\lambda_1 + \lambda_2 + \lambda_3)$.

Aplicando o logaritmo na função de verossimilhança temos

$$l(\boldsymbol{\delta}) = m_0 \log \left\{ \omega + (1 - \omega)e^{-\lambda_{\bullet}} \right\} + (n - m_0) \log(1 - \omega) + \sum_{i=1}^n [1 - I_{\{(0,0)\}}(x_i, y_i)] \log p(x_i, y_i | \boldsymbol{\lambda}). \quad (4.5)$$

Derivando o logaritmo da função de verossimilhança (4.5) em relação aos parâmetros π , λ_1 , λ_2 e λ_3 e

igualando a zero, obtêm-se as equações

$$\begin{aligned}\frac{\partial l(\boldsymbol{\delta})}{\partial \omega} &= \frac{m_0[1 - e^{-\lambda \bullet}]}{\omega + (1 - \omega)e^{-\lambda \bullet}} - \frac{n - m_0}{1 - \omega} = 0, \\ \frac{\partial l(\boldsymbol{\delta})}{\partial \lambda_1} &= \frac{-m_0[(1 - \omega)e^{-\lambda \bullet}]}{\omega + (1 - \omega)e^{-\lambda \bullet}} + \sum_{i=1}^n (1 - I_{\{(0,0)\}}(x_i, y_i)) \frac{\partial p(x_i, y_i | \boldsymbol{\lambda})}{\partial \lambda_1} \frac{1}{p(x_i, y_i | \boldsymbol{\lambda})} = 0, \\ \frac{\partial l(\boldsymbol{\delta})}{\partial \lambda_2} &= \frac{-m_0[(1 - \omega)e^{-\lambda \bullet}]}{\omega + (1 - \omega)e^{-\lambda \bullet}} + \sum_{i=1}^n (1 - I_{\{(0,0)\}}(x_i, y_i)) \frac{\partial p(x_i, y_i | \boldsymbol{\lambda})}{\partial \lambda_2} \frac{1}{p(x_i, y_i | \boldsymbol{\lambda})} = 0, \\ \frac{\partial l(\boldsymbol{\delta})}{\partial \lambda_3} &= \frac{-m_0[(1 - \omega)e^{-\lambda \bullet}]}{\omega + (1 - \omega)e^{-\lambda \bullet}} + \sum_{i=1}^n (1 - I_{\{(0,0)\}}(x_i, y_i)) \frac{\partial p(x_i, y_i | \boldsymbol{\lambda})}{\partial \lambda_3} \frac{1}{p(x_i, y_i | \boldsymbol{\lambda})} = 0.\end{aligned}$$

As estimativas de máxima verossimilhança dos parâmetros do modelo poderiam ser obtidas a partir da solução das equações acima. Entretanto, as equações não tem solução analítica, o que nos leva a adotar um método numérico para obter as estimativas de máxima verossimilhança para os parâmetros. A seguir utilizaremos o algoritmo EM descrito na Seção 1.2.3 para a obtenção das estimativas de máxima verossimilhança para os parâmetros do modelo.

4.3.2 Função de verossimilhança baseada nos dados aumentados

A distribuição BZIP pode ser explicada a partir de um mecanismo probabilístico que consiste em duas etapas. Na primeira etapa, observa-se uma variável aleatória Z com distribuição de Bernoulli com parâmetro ω . Em seguida, observa-se (X, Y, U) tal que

$$P(X = x, Y = y, U = u | Z = z, \boldsymbol{\lambda}) = \begin{cases} I_{\{(0,0,0)\}}(x, y, u), & \text{se } z = 1, \\ p(x, y, u | \boldsymbol{\lambda}), & \text{se } z = 0, \end{cases}$$

sendo $x, y = 0, 1, \dots, u = 0, 1, \dots, \min(x, y)$, $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3)$ e

$$p(x, y, u | \boldsymbol{\lambda}) = \frac{\lambda_1^{x-u} \lambda_2^{y-u} \lambda_3^u e^{-(\lambda_1 + \lambda_2 + \lambda_3)}}{(x-u)!(y-u)!u!} I_{\{(0, \dots, \min(x, y))\}}(u).$$

Isto é, dado $z = 1$, (X, Y, U) tem distribuição degenerada em $(0, 0, 0)$ e dado $z = 0$, (X, Y, U) tem função de probabilidades $p(x, y, u | \boldsymbol{\lambda})$ que equivale assumir que $X - U, Y - U$ e U , tem distribuição Poisson independente com parâmetros $\lambda_1, \lambda_2, \lambda_3$, respectivamente.

Assim, f.p. conjunta de (X, Y, U, Z) é dada por

$$P(X = x, Y = y, U = u, Z = z | \boldsymbol{\delta}) = \omega I_{\{(0,0,0)\}}(x, y, u) I_{\{1\}}(z) + (1 - \omega) p(x, y, u | \boldsymbol{\lambda}) I_{\{0\}}(z), \quad (4.6)$$

em que

$$I_{\{(0,0,0)\}}(x, y, u) = \begin{cases} 1, & \text{se } (x, y, u) = (0, 0, 0), \\ 0, & \text{se } (x, y, u) \neq (0, 0, 0), \end{cases}$$

onde $x, y = 0, 1, \dots$, $u = 0, \dots, \min(x, y)$, $z = 0, 1$, $\boldsymbol{\delta} = (\omega, \boldsymbol{\lambda})$, $0 \leq \omega \leq 1$ e $\lambda_i > 0$, para $i = 1, 2$ e $\lambda_3 \geq 0$.

As variáveis U e Z são denominadas variáveis latentes, pois não são observadas diretamente, e a f.p. conjunta dada em (4.6) será dita f.p. ampliada. Note que a f.p. marginal do vetor (X, Y) obtida por meio de (4.6) é a mesma f.p. do modelo BZIP dada em (4.2).

A seguir, deduziremos a f.p. condicional de (Z, U) dado (X, Y) para obtermos as esperanças condicionais das variáveis latentes U e Z dado (X, Y) que serão úteis no desenvolvimento dos passos para o algoritmo EM.

A f.p. condicional de (U, Z) dado $X = x$, $Y = y$ e $\boldsymbol{\delta}$ é dada por

$$\begin{aligned} P(U = u, Z = 1 | X = x, Y = y, \boldsymbol{\delta}) &= \frac{P(X = x, Y = y, U = u, Z = 1 | \boldsymbol{\delta})}{P(X = x, Y = y | \boldsymbol{\delta})} \\ &= \frac{\omega I_{\{(0,0,0)\}}(x, y, u)}{\omega I_{\{(0,0)\}}(x, y) + (1 - \omega)p(x, y | \boldsymbol{\lambda})} \\ &= \begin{cases} \frac{\omega}{\omega + (1 - \omega)e^{-(\lambda_1 + \lambda_2 + \lambda_3)}}, & \text{se } (x, y) = (0, 0), \\ 0, & \text{se } (x, y) \neq (0, 0), \end{cases} \end{aligned}$$

e

$$\begin{aligned} P(U = u, Z = 0 | X = x, Y = y, \boldsymbol{\delta}) &= \frac{P(X = x, Y = y, U = u, Z = 0 | \boldsymbol{\delta})}{P(X = x, Y = y | \boldsymbol{\delta})} \\ &= \frac{(1 - \omega)p(x, y, u | \boldsymbol{\lambda})}{\omega I_{\{(0,0)\}}(x, y) + (1 - \omega)p(x, y | \boldsymbol{\lambda})} \\ &= \begin{cases} \frac{(1 - \omega)e^{-(\lambda_1 + \lambda_2 + \lambda_3)}}{\omega + (1 - \omega)e^{-(\lambda_1 + \lambda_2 + \lambda_3)}}, & \text{se } (x, y) = (0, 0), \\ \frac{1}{c(x, y, \boldsymbol{\lambda})(x - u)!(y - u)!u!} \left(\frac{\lambda_3}{\lambda_1 \lambda_2} \right)^u, & \text{se } (x, y) \neq (0, 0), \end{cases} \end{aligned}$$

sendo $u = 0, 1, \dots, \min(x, y)$ e $c(x, y, \boldsymbol{\lambda}) = \sum_{r=0}^{\min(x, y)} \frac{1}{(x - r)!(y - r)!r!} \left(\frac{\lambda_3}{\lambda_1 \lambda_2} \right)^r$.

Assim, definimos as esperanças condicionais como

$$s(x, y; \boldsymbol{\delta}) = E[U | X = x, Y = y, \boldsymbol{\delta}] \text{ e}$$

$$r(x, y; \boldsymbol{\delta}) = E[Z | X = x, Y = y, \boldsymbol{\delta}].$$

As expressões $s(x, y; \boldsymbol{\delta})$ e $r(x, y; \boldsymbol{\delta})$ serão obtidas em dois casos: (i) $(x, y) = (0, 0)$; (ii) $(x, y) \neq (0, 0)$.

Assim,

$$\begin{aligned}
s(0, 0; \boldsymbol{\delta}) &= 0, \\
s(x, y; \boldsymbol{\delta}) &= \sum_{z=0}^1 \sum_{u=0}^{\min(x,y)} u \mathbb{P}(U = u, Z = z | X = x, Y = y, \boldsymbol{\delta}) \\
&= \sum_{u=0}^{\min(x,y)} u \mathbb{P}(U = u, Z = 0 | X = x, Y = y, \boldsymbol{\delta}) \\
&= \frac{1}{c(x, y, \boldsymbol{\lambda})} \sum_{u=0}^{\min(x,y)} \frac{u}{(x-u)!(y-u)!u!} \left(\frac{\lambda_3}{\lambda_1 \lambda_2} \right)^u
\end{aligned}$$

e

$$\begin{aligned}
r(0, 0; \boldsymbol{\delta}) &= \sum_{z=0}^1 \sum_{u=0}^{\min(x,y)} z \mathbb{P}(U = u, Z = z | X = x, Y = y, \boldsymbol{\delta}) \\
&= \sum_{u=0}^{\min(x,y)} \mathbb{P}(U = u, Z = 1 | X = x, Y = y, \boldsymbol{\delta}) \\
&= \frac{\omega}{\omega + (1 - \omega)e^{-(\lambda_1 + \lambda_2 + \lambda_3)}}, \\
r(x, y; \boldsymbol{\delta}) &= 0,
\end{aligned}$$

com $x, y = 0, 1, \dots$, e $u = 0, 1, \dots, \min(x, y)$.

Considere as observações de uma amostra aleatória $(x_1, y_1, u_1, z_1), \dots, (x_n, y_n, u_n, z_n)$ proveniente da f.p. ampliada dada em (4.6) e defina $D_0 = \{i \in \{1, 2, \dots, n\} : (x_i, y_i, u_i) = (0, 0, 0)\}$ o conjunto de zeros na amostra. É conveniente notar que $z_i = 0$ sempre que $i \notin D_0$, pois caso $(x_i, y_i, u_i) \neq (0, 0, 0)$ então a i -ésima observação não é proveniente da distribuição degenerada em $(0, 0, 0)$. A função de verossimilhança ampliada para $\boldsymbol{\delta}$ é

$$\begin{aligned}
L_{\mathcal{A}}(\boldsymbol{\delta}) &= \prod_{i=1}^n \mathbb{P}(X_i = x_i, Y_i = y_i, U_i = u_i, Z_i = z_i | \boldsymbol{\delta}) \\
&= \prod_{i \in D_0} \mathbb{P}(X_i = x_i, Y_i = y_i, U_i = u_i, Z_i = z_i | \boldsymbol{\delta}) \prod_{i \notin D_0} \mathbb{P}(X_i = x_i, Y_i = y_i, U_i = u_i, Z_i = 0 | \boldsymbol{\delta}) \\
&= \prod_{i \in D_0} \omega^{z_i} \left[(1 - \omega)e^{-\lambda \bullet} \right]^{1-z_i} \prod_{i \notin D_0} (1 - \omega) \frac{\lambda_1^{x_i - u_i} \lambda_2^{y_i - u_i} \lambda_3^{u_i} e^{-\lambda \bullet}}{(x_i - u_i)!(y_i - u_i)!u_i!} \\
&= \omega^{\sum_{i \in D_0} z_i} (1 - \omega)^{n - \sum_{i \in D_0} z_i} e^{-\lambda \bullet (n - \sum_{i \in D_0} z_i)} \frac{\prod_{i \notin D_0} \lambda_1^{x_i - u_i} \prod_{i \notin D_0} \lambda_2^{y_i - u_i} \prod_{i \notin D_0} \lambda_3^{u_i}}{\prod_{i \notin D_0} (x_i - u_i)!(y_i - u_i)!u_i!}, \tag{4.7}
\end{aligned}$$

em que $x_i, y_i = 0, 1, \dots, u_i = 0, \dots, \min(x_i, y_i), z_i = 0, 1, n_0 = \sum_{i=1}^n I_{\{(0,0,0)\}}(x_i, y_i, u_i)$ e $\lambda_{\bullet} = (\lambda_1 + \lambda_2 + \lambda_3)$.

Note que a função de verossimilhança ampliada (4.7) é muito mais simples do que a função de verossimilhança original (4.4), pois não envolve um produtório de somatórios.

O logaritmo da função de verossimilhança é dado por

$$\begin{aligned}
l_{\mathcal{A}}(\boldsymbol{\delta}) &= \sum_{i \in D_0} z_i \log \omega + [n - \sum_{i \in D_0} z_i] \log(1 - \omega) - \lambda_{\bullet} [n - \sum_{i \in D_0} z_i] + \sum_{i \notin D_0} (x_i - u_i) \log \lambda_1 + \sum_{i \notin D_0} (y_i - u_i) \log \lambda_2 \\
&\quad + \sum_{i \notin D_0} u_i \log \lambda_3 - \sum_{i \notin D_0} \log \{(x_i - u_i)!(y_i - u_i)!u_i!\} \\
&= [\log \omega - \log(1 - \omega) + \lambda_{\bullet}] \sum_{i \in D_0} z_i + n [\log(1 - \omega) - \lambda_{\bullet}] + \log \lambda_1 \sum_{i \notin D_0} x_i + \log \lambda_2 \sum_{i \notin D_0} y_i \\
&\quad - [\log \lambda_1 + \log \lambda_2 - \log \lambda_3] \sum_{i \notin D_0} u_i - \sum_{i \notin D_0} \log \{(x_i - u_i)!(y_i - u_i)!u_i!\}. \tag{4.8}
\end{aligned}$$

4.3.3 Estimação via algoritmo EM

A esperança do $l_{\mathcal{A}}(\boldsymbol{\delta})$ com respeito à distribuição condicional (\mathbf{Z}, \mathbf{U}) dado $\mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y}$ e o parâmetro $\boldsymbol{\delta}^*$ é dada pela função

$$\begin{aligned}
Q(\boldsymbol{\delta}, \boldsymbol{\delta}^*) &= E_{\mathbf{Z}, \mathbf{U} | \mathbf{X}=\mathbf{x}, \mathbf{Y}=\mathbf{y}, \boldsymbol{\delta}^*} [l_{\mathcal{A}}(\boldsymbol{\delta})] \\
&= [\log \omega - \log(1 - \omega) + \lambda_{\bullet}] \sum_{i \in D_0} E[Z_i | X_i = x_i, Y_i = y_i, \boldsymbol{\delta}^*] + n [\log(1 - \omega) - \lambda_{\bullet}] \\
&\quad + \log \lambda_1 \sum_{i \notin D_0} x_i + \log \lambda_2 \sum_{i \notin D_0} y_i - [\log \lambda_1 + \log \lambda_2 - \log \lambda_3] \sum_{i \notin D_0} E[U_i | X_i = x_i, Y_i = y_i, \boldsymbol{\delta}^*] \\
&\quad - \sum_{i \notin D_0} K(x_i, y_i, \boldsymbol{\delta}^*) \\
&= [\log \omega - \log(1 - \omega) + \lambda_{\bullet}] \sum_{i \in D_0} r(x_i, y_i; \boldsymbol{\delta}^*) + n [\log(1 - \omega) - \lambda_{\bullet}] + \log \lambda_1 \sum_{i \notin D_0} x_i \\
&\quad + \log \lambda_2 \sum_{i \notin D_0} y_i - [\log \lambda_1 + \log \lambda_2 - \log \lambda_3] \sum_{i \notin D_0} s(x_i, y_i; \boldsymbol{\delta}^*) - \sum_{i \notin D_0} K(x_i, y_i, \boldsymbol{\delta}^*)
\end{aligned}$$

onde $K(x_i, y_i, \boldsymbol{\delta}^*) = E[\log \{(x_i - U_i)!(y_i - U_i)!U_i!\} | X_i = x_i, Y_i = y_i, \boldsymbol{\delta}^*]$.

Para aplicar o algoritmo EM atribui-se um valor inicial ao vetor de parâmetros $\boldsymbol{\delta}, \boldsymbol{\delta}^{(0)}$ e repete-se o passo E e o passo M, até obter a convergência.

Passo E : Dado $\boldsymbol{\delta}^{(k)} = (\omega^{(k)}, \lambda_1^{(k)}, \lambda_2^{(k)}, \lambda_3^{(k)})$, o valor de $\boldsymbol{\delta}$ na k -ésima iteração, calcule $r(x_i, y_i, \boldsymbol{\delta}^{(k)})$ e $s(x_i, y_i, \boldsymbol{\delta}^{(k)})$.

Para $i \in D_0$, definimos

$$r(x_i, y_i; \boldsymbol{\delta}^{(k)}) = \frac{\omega^{(k)}}{\omega^{(k)} + (1 - \omega^{(k)})e^{-(\lambda_1^{(k)} + \lambda_2^{(k)} + \lambda_3^{(k)})}}, \quad (4.9)$$

e para $i \notin D_0$

$$s(x_i, y_i; \boldsymbol{\delta}^{(k)}) = \frac{1}{c(x_i, y_i; \boldsymbol{\lambda}^{(k)})} \sum_{u=0}^{\min(x_i, y_i)} \frac{u}{(x_i - u)!(y_i - u)!u!} \left(\frac{\lambda_3^{(k)}}{\lambda_1^{(k)} \lambda_2^{(k)}} \right)^u, \quad (4.10)$$

onde

$$c(x_i, y_i; \boldsymbol{\lambda}^{(k)}) = \sum_{r=0}^{\min(x_i, y_i)} \frac{1}{(x_i - r)!(y_i - r)!r!} \left(\frac{\lambda_3^{(k)}}{\lambda_1^{(k)} \lambda_2^{(k)}} \right)^r.$$

Passo M: Maximize $Q(\boldsymbol{\delta}, \boldsymbol{\delta}^{(k)})$ em relação a $\boldsymbol{\delta}$, obtendo as seguintes soluções analíticas

$$\begin{aligned} \hat{\omega}^{(k+1)} &= \frac{\sum_{i \in D_0} r(x_i, y_i; \boldsymbol{\delta}^{(k)})}{n}, \\ \hat{\lambda}_1^{(k+1)} &= \frac{\sum_{i \notin D_0} [x_i - s(x_i, y_i; \boldsymbol{\delta}^{(k)})]}{n - \sum_{i \in D_0} r(x_i, y_i; \boldsymbol{\delta}^{(k)})}, \\ \hat{\lambda}_2^{(k+1)} &= \frac{\sum_{i \notin D_0} [y_i - s(x_i, y_i; \boldsymbol{\delta}^{(k)})]}{n - \sum_{i \in D_0} r(x_i, y_i; \boldsymbol{\delta}^{(k)})}, \\ \hat{\lambda}_3^{(k+1)} &= \frac{\sum_{i \notin D_0} s(x_i, y_i; \boldsymbol{\delta}^{(k)})}{n - \sum_{i \in D_0} r(x_i, y_i; \boldsymbol{\delta}^{(k)})}. \end{aligned}$$

4.4 Teste de hipótese

Nesta seção, iremos utilizar o teste da razão de verossimilhanças para constatar as hipóteses de interesse.

4.4.1 Teste de hipótese para o parâmetro de inflação

A distribuição BZIP se reduz a distribuição PB quando $\omega = 0$. Desta forma, para avaliar adequação da distribuição PB aos dados, vamos considerar o teste de hipóteses

$$H_0 : \omega = 0$$

versus $H_1 : \omega > 0$.

e a estatística da razão de verossimilhanças dada por

$$\Psi = \frac{\sup_{\lambda \in \Lambda} L^{\text{PB}}(\lambda)}{\sup_{\theta \in \Delta} L(\theta)},$$

em que $L^{\text{PB}}(\lambda) = L(\lambda_1, \lambda_2, \lambda_3)$ é a função de verossimilhança para o modelo restrito, PB, com $\lambda = (\lambda_1, \lambda_2, \lambda_3) \in \Lambda$. A estatística $-2 \log(\Psi)$ segue, assintoticamente, uma distribuição qui-quadrado com 1 graus de liberdade. A região crítica do teste é, portanto, $C = \{-2 \log(\Psi) \geq \chi_{1,\alpha}^2\}$, onde $\chi_{1,\alpha}^2$ é o ponto crítico da χ_1^2 correspondente ao nível de significância α .

4.4.2 Teste para a distribuição ZIPD

A distribuição BZIP se reduz à distribuição ZIPD quando $\lambda_3 = 0$. Desta forma, para avaliar adequação da distribuição ZIPD aos dados, vamos considerar o teste de hipóteses

$$H_0 : \lambda_3 = 0$$

versus $H_1 : \lambda_3 > 0$.

e a estatística da razão de verossimilhanças dada por

$$\Psi^* = \frac{\sup_{\delta_1 \in \Delta_1} L^{\text{ZIPD}}(\delta_1)}{\sup_{\delta \in \Delta} L(\delta)},$$

em que $L^{\text{ZIPD}}(\delta_1) = L(\omega, \lambda_1, \lambda_2, 0)$ é a função de verossimilhança para o modelo restrito, ZIPD, $\delta_1 \in \Delta_1 = [0, 1] \times (0, +\infty)^2 \times \{0\}$. A estatística $-2 \log(\Psi^*)$ segue, assintoticamente, uma distribuição qui-quadrado com 1 graus de liberdade. A região crítica do teste é, portanto, $C = \{-2 \log(\Psi^*) \geq \chi_{1,\alpha}^2\}$, onde $\chi_{1,\alpha}^2$ é o ponto crítico da χ_1^2 correspondente ao nível de significância α .

4.4.3 Teste para a distribuição PD

A distribuição BZIP se reduz a distribuição PD dada por (2.4), quando $\omega = 0$ e $\lambda_3 = 0$. Desta forma, para avaliar adequação da distribuição PD aos dados, vamos considerar o teste de hipóteses

$$H_0 : \omega = \lambda_3 = 0$$

versus $H_1 : \omega = \lambda_3 \neq 0$.

e a estatística da razão de verossimilhanças dada por

$$\Psi' = \frac{\sup_{\delta_0 \in \Delta_0} L^{\text{PD}}(\delta_0)}{\sup_{\delta \in \Delta} L(\delta)},$$

em que $L^{\text{PD}}(\delta_0) = L(0, \lambda_1, \lambda_2, 0)$ é a função de verossimilhança para o modelo restrito, PD, com $\delta_0 \in \Delta_0 = \{0\} \times (0, +\infty)^2 \times \{0\}$. A estatística $-2 \log(\Psi')$ segue, assintoticamente, uma distribuição qui-quadrado com 2 graus de liberdade. A região crítica do teste é, portanto, $C = \{-2 \log(\Psi') \geq \chi_{2,\alpha}^2\}$, onde $\chi_{2,\alpha}^2$ é o ponto crítico da χ_2^2 correspondente ao nível de significância α .

4.5 Aplicação

Nesta seção, utilizaremos o conjunto de dados reais apresentados na Tabela 2.1 da Seção 2.5, referentes ao número de acidentes de ônibus ocorridos com 708 motoristas no norte da Irlanda durante os períodos 1952 – 53 e 1954 – 55.

Na Tabela 4.1 estão apresentadas as estimativas dos parâmetros do modelo BZIP, que foram obtidas pelo método de máxima verossimilhança via algoritmo EM.

Tabela 4.1: Estimativas dos parâmetros do modelo BZIP

parâmetros	ω	λ_1	λ_2	λ_3
EMV	0,072	0,900	1,212	0,177

Utilizando o teste da razão de verossimilhança discutido na Subseção 4.4.1, a fim de determinar se há excesso de zeros, dado que a proporção de zeros estimados pelo modelo BZIP é um critério importante para a qualidade do ajuste. A estatística da razão de verossimilhança é $-2 \log(\Psi) = 791,44$. O percentil da χ_1^2 , considerando o nível 0,1% de significancia, tem valor crítico do teste igual a 10,82 e, observa-se que o valor obtido é maior que o valor crítico, indicando forte evidência contra do modelo de PB. Portanto, pode-se concluir que existe uma quantidade excessiva de zeros nos dados e o modelo BZIP é mais apropriado nesta situação.

Notamos que o modelo ZIPD é um caso particular do modelo BZIP, quando $\lambda_3 = 0$, para verificar a adequação deste modelo aos dados recorreremos ao teste da razão de verossimilhança desenvolvido na Subseção 4.4.2. A estatística da razão de verossimilhança é $-2 \log(\Psi^*) = 791,19$. Considerando o nível

0,1% de significância, o valor obtido é maior que o valor crítico, de modo que rejeitamos a hipótese nula $H_0 : \lambda_3 = 0$. Pode-se concluir que o modelo BZIP é mais apropriado do que o modelo ZIPD.

Outro caso particular é o modelo PD. Quando ω e λ_3 são nulos, foi calculado o valor da estatística da razão de verossimilhança do teste definido na Subseção 4.4.3, que fornece $-2\log(\Psi') = 311,95$ sendo maior que valor crítico, levando a rejeição da hipótese nula. O que significa que o modelo mais adequado para este conjunto de dados é o modelo BZIP.

A Tabela 4.2 apresenta as frequências observadas e esperadas de acordo com os modelos BZIP, PGB e PB, cujos parâmetros dos modelos foram estimados pelo método de máxima verossimilhança. Observamos que as frequências esperadas dos modelos BZIP e PGB estão mais próximas das frequências observadas do que o modelo PB. Nota-se o modelo BZIP é o modelo que mais se ajusta ao número de zeros, (0, 0) do conjunto de dados.

Tabela 4.2: Frequências observadas e esperadas para os modelos BZIP, PGB e PB

x /	y	0	1	2	3	4	5	6	7	Total
0	Obs	117	96	55	19	2	2	0	0	291
	BZIP	116,99	80,66	48,91	19,77	5,99	1,45	0,29	0,05	
	PGB	110,94	96,04	49,34	19,66	6,72	2,08	0,59	0,16	
	PB	89,94	95,50	50,70	17,94	4,76	1,01	0,17	0,02	
1	Obs	61	69	47	27	8	5	1	0	218
	BZIP	59,92	84,47	58,37	26,48	8,90	2,37	0,52	0,09	
	PGB	68,27	82,55	50,66	22,52	8,29	2,70	0,80	0,22	
	PB	69,46	94,36	61,04	25,47	7,79	1,87	0,37	0,06	
2	Obs	34	42	31	13	7	2	3	0	132
	BZIP	26,99	43,36	33,78	17,10	6,36	1,85	0,44	0,09	
	PGB	26,90	37,72	28,84	14,98	6,14	2,15	0,68	0,20	
	PB	26,82	44,39	34,38	16,83	5,92	1,61	0,35	0,06	
3	Obs	7	15	16	7	3	1	0	0	49
	BZIP	8,10	14,62	12,71	7,13	2,92	0,93	0,24	0,05	
	PGB	8,70	13,21	11,49	7,22	3,44	1,34	0,46	0,14	
	PB	6,90	13,47	12,24	6,95	2,81	0,86	0,21	0,04	
4	Obs	3	3	1	1	2	1	1	1	13
	BZIP	1,82	3,65	3,51	2,17	0,97	0,34	0,09	0,02	
	PGB	2,52	4,02	3,77	2,68	1,53	0,70	0,26	0,09	
	PB	6,90	13,47	12,24	6,95	2,81	0,86	0,21	0,04	
5	Obs	2	1	0	0	0	0	0	0	3
	BZIP	0,32	0,72	0,76	0,51	0,25	0,09	0,02	0,00	
	PGB	0,68	1,12	1,11	0,85	0,55	0,30	0,13	0,05	
	PB	0,20	0,52	0,62	0,45	0,23	0,09	0,02	0,00	
6	Obs	0	0	0	0	1	0	0	0	1
	BZIP	0,04	0,11	0,13	0,09	0,05	0,021	0,00	0,00	
	PGB	0,17	0,29	0,30	0,24	0,17	0,10	0,05	0,02	
	PB	0,02	0,07	0,10	0,08	0,04	0,00	0,00	0,00	
7	Obs	0	0	0	1	0	0	0	0	1
	BZIP	0,00	0,01	0,02	0,01	0,00	0,00	0,00	0,00	
	PGB	0,04	0,07	0,07	0,06	0,04	0,03	0,02	0,01	
	PB	0,00	0,00	0,01	0,01	0,00	0,00	0,00	0,00	

Uma alternativa para a comparação dos modelos discutidos PB, PGB e BZIP é utilizar os critérios de seleção de modelos AIC e BIC. A Tabela 4.3 apresenta os valores destes critérios para os modelos considerados, em que se observa um melhor ajuste do modelo BZIP.

Tabela 4.3: Critérios de seleção de modelos

Modelo	AIC	BIC
BZIP	1624,51	1624,76
PGB	4005,25	4018,93
PB	4028,40	4037,53

Portanto, conclui-se que o modelo BZIP se ajustou melhor que os modelos PB e PGB para este conjunto de dados.

Capítulo 5

Abordagem Bayesiana dos Modelos PB e BZIP

5.1 Introdução

A abordagem Bayesiana dos modelos PB e BZIP tem como principal obstáculo a função de verossimilhança pois esta é analiticamente difícil, devido a presença de produtos de somas. Buscando solucionar os problemas apontados com relação a função de verossimilhança, Tanner & Wong (1987) introduziram variáveis latentes aos dados observados reduzindo a complexidade da função de verossimilhança. De acordo com Diebolt & Robert (1994) e Tsionas (1999) o acréscimo de variáveis latentes pode simplificar a implementação do algoritmo Gibbs sampling (Cassela & George (1992)). Albert & Chib (1993) foram um dos primeiros a incluir variáveis latentes para obtenção das distribuições condicionais completas conhecidas no contexto do algoritmo Gibbs sampling. Neste trabalho, vamos ilustrar estas técnicas para o problema de estimação Bayesiana dos modelos PB e BZIP.

O algoritmo Gibbs sampling com dados ampliados é essencialmente um esquema iterativo de amostragem de uma cadeia de Markov formada a partir das distribuições condicionais completas. Sob certas condições de regularidade, a distribuição limite é a densidade de interesse. Uma discussão mais geral sobre este algoritmo pode ser encontrada em Diebolt & Robert (1994). O algoritmo Gibbs sampling com dados ampliados é resumido por:

Passo 1: inicialize o contador $t = 0$;

Passo 2: especifique valores iniciais $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_s^{(0)})$;

Passo 3: faça $t = t + 1$;

Passo 4: gere a variável latente $u^{(t)} \sim f(u|\mathbf{x}, \boldsymbol{\theta}^{(t-1)})$,

Passo 5: a partir do valor de $u^{(t)}$ obtenha o vetor $\boldsymbol{\theta}^{(t)}$ como

$$\begin{aligned}\theta_1^{(t)} &\sim \pi(\theta_1 | \mathbf{x}, u^{(t)}, \theta_2^{(t-1)}, \dots, \theta_s^{(t-1)}), \\ &\vdots \\ \theta_s^{(t)} &\sim \pi(\theta_s | \mathbf{x}, u^{(t)}, \theta_1^{(t+1)}, \dots, \theta_{s-1}^{(t+1)}),\end{aligned}$$

Passo 6: retorne para o passo 3 até obter a convergência.

5.2 Distribuição a posteriori do modelo PB a partir de dados ampliados

Pela construção do modelo PB, como foi visto na Seção 2.2, temos que (X, Y) são as variáveis de interesse e V_1, V_2, U são as variáveis latentes, pois não são observadas diretamente. Note que a variável U é comum aos dados de interesse. Desta forma, (X, Y, U) é o vetor ampliado. A f.p. de (X, Y, U) é dada por

$$\begin{aligned}p(x, y, u | \boldsymbol{\lambda}) &= \text{P}(X = x, Y = y, U = u | \boldsymbol{\lambda}) \\ &= \text{P}(V_1 = x - u, V_2 = y - u, U = u | \boldsymbol{\lambda}) \\ &= \frac{\lambda_1^{x-u} \lambda_2^{y-u} \lambda_3^u e^{-(\lambda_1 + \lambda_2 + \lambda_3)}}{(x-u)!(y-u)!u!},\end{aligned}\tag{5.1}$$

sendo $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3)$, $x, y = 0, 1, \dots$ e $u = 0, \dots, \min(x, y)$.

Supondo que $(x_1, y_1, u_1), \dots, (x_n, y_n, u_n)$ sejam as observações de uma amostra aleatória proveniente da f.p. (5.1). A função de verossimilhança baseada nos dados ampliados $\mathcal{D} = (\mathbf{x}, \mathbf{y}, \mathbf{u})$ é

$$\begin{aligned}L_{\mathcal{D}}(\boldsymbol{\lambda}) &= \prod_{i=1}^n \text{P}(X_i = x_i, Y_i = y_i, U_i = u_i | \boldsymbol{\lambda}) \\ &= \prod_{i=1}^n \frac{\lambda_1^{x_i - u_i} \lambda_2^{y_i - u_i} \lambda_3^{u_i} e^{-(\lambda_1 + \lambda_2 + \lambda_3)}}{(x_i - u_i)!(y_i - u_i)!u_i!} \\ &\propto \lambda_1^{\sum_{i=1}^n (x_i - u_i)} e^{-n\lambda_1} \lambda_2^{\sum_{i=1}^n (y_i - u_i)} e^{-n\lambda_2} \lambda_3^{\sum_{i=1}^n u_i} e^{-n\lambda_3}.\end{aligned}\tag{5.2}$$

Note que a função de verossimilhança torna-se mais simples com o acréscimo da variável latente.

Para estimar os parâmetros λ_1, λ_2 e λ_3 do modelo PB, dentro de um contexto bayesiano, é necessário especificar a distribuição a priori de λ_1, λ_2 e λ_3 .

A função de verossimilhança (5.2) sugere prioris independentes para λ_1, λ_2 e λ_3 dadas por

$$\begin{aligned}\pi(\lambda_1) &= \frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)} \lambda_1^{\alpha_1-1} e^{-\beta_1 \lambda_1}, \quad \alpha_1, \beta_1 > 0, \\ \pi(\lambda_2) &= \frac{\beta_2^{\alpha_2}}{\Gamma(\alpha_2)} \lambda_2^{\alpha_2-1} e^{-\beta_2 \lambda_2}, \quad \alpha_2, \beta_2 > 0, \\ \pi(\lambda_3) &= \frac{\beta_3^{\alpha_3}}{\Gamma(\alpha_3)} \lambda_3^{\alpha_3-1} e^{-\beta_3 \lambda_3}, \quad \alpha_3, \beta_3 > 0,\end{aligned}$$

onde $\Gamma(\cdot)$ é a função gamma e α_i e β_i , para $i = 1, 2, 3$, são hiperparâmetros conhecidos.

Assim, considerando as prioris independentes, a densidade a priori conjunta é dada por

$$\pi(\lambda_1, \lambda_2, \lambda_3) \propto \lambda_1^{\alpha_1-1} e^{-\beta_1 \lambda_1} \lambda_2^{\alpha_2-1} e^{-\beta_2 \lambda_2} \lambda_3^{\alpha_3-1} e^{-\beta_3 \lambda_3}. \quad (5.3)$$

Portanto, combinando a função de verossimilhança dada em (5.2) e a densidade a priori conjunta dada em (5.3), obtém-se a densidade a posteriori conjunta para $\boldsymbol{\lambda}$ dado $\mathcal{D} = (\mathbf{x}, \mathbf{y}, \mathbf{u})$ dada por

$$\pi(\boldsymbol{\lambda}|\mathcal{D}) \propto \lambda_1^{\sum_{i=1}^n (x_i - u_i) + \alpha_1 - 1} e^{-(n+\beta_1)\lambda_1} \lambda_2^{\sum_{i=1}^n (y_i - u_i) + \alpha_2 - 1} e^{-(n+\beta_2)\lambda_2} \lambda_3^{\sum_{i=1}^n u_i + \alpha_3 - 1} e^{-(n+\beta_3)\lambda_3}. \quad (5.4)$$

A partir da distribuição a posteriori conjunta (5.4), obtemos as distribuições condicionais completas necessárias para o algoritmo Gibbs sampling com dados ampliados dadas por

$$\begin{aligned}\lambda_1|\lambda_2, \lambda_3, \mathcal{D} &\sim \text{Gamma}\left(\sum_{i=1}^n (x_i - u_i) + \alpha_1, n + \beta_1\right); \\ \lambda_2|\lambda_1, \lambda_3, \mathcal{D} &\sim \text{Gamma}\left(\sum_{i=1}^n (y_i - u_i) + \alpha_2, n + \beta_2\right); \\ \lambda_3|\lambda_1, \lambda_2, \mathcal{D} &\sim \text{Gamma}\left(\sum_{i=1}^n u_i + \alpha_3, n + \beta_3\right).\end{aligned}$$

A f.p. condicional dos dados latentes dado os dados observados é dada por

$$\begin{aligned}
p(\mathbf{u}|\mathbf{x}, \mathbf{y}, \boldsymbol{\lambda}) &= P[U = \mathbf{u} | \mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y}, \boldsymbol{\lambda}] \\
&\propto P[\mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y}, \mathbf{U} = \mathbf{u} | \boldsymbol{\lambda}] \\
&= \prod_{i=1}^n P[X_i = x_i, Y_i = y_i, U_i = u_i | \boldsymbol{\lambda}] \\
&= \prod_{i=1}^n P[V_1^i = x_i - u_i, V_2^i = y_i - u_i, U_i = u_i | \boldsymbol{\lambda}] \\
&= \prod_{i=1}^n \frac{e^{\lambda_1} \lambda_1^{x_i - u_i} e^{\lambda_2} \lambda_2^{y_i - u_i} e^{\lambda_3} \lambda_3^{u_i}}{(x_i - u_i)! (y_i - u_i)! u_i!} \\
&\propto \prod_{i=1}^n \frac{1}{(x_i - u_i)! (y_i - u_i)! u_i!} \left(\frac{\lambda_3}{\lambda_2 \lambda_1} \right)^{u_i},
\end{aligned}$$

com $u_i = \{0, \dots, \min(x_i, y_i)\}$, para $i = 1, \dots, n$.

Assim, o algoritmo Gibbs sampling com dados ampliados é dado por:

Passo 1: inicialize o contador $t = 0$;

Passo 2: especifique valores iniciais $\boldsymbol{\lambda}^{(0)} = (\lambda_1^{(0)}, \lambda_2^{(0)}, \lambda_3^{(0)})$;

Passo 3: faça $t = t + 1$;

Passo 4: gere valores para $\mathbf{u}^{(t)}$ a partir da distribuição condicional

$$\mathbf{U}^{(t)} \sim p(\mathbf{u}|\mathbf{x}, \mathbf{y}, \boldsymbol{\lambda}^{(t-1)}),$$

Passo 5: a partir dos valores de $\mathbf{u}^{(t)}$ obtenha o vetor $\boldsymbol{\lambda}^{(t)}$ como

$$\begin{aligned}
\lambda_1^{(t)} &\sim \text{Gamma} \left(\sum_i^n (x_i - u_i^{(t)}) + \alpha_1, n + \beta_1 \right), \\
\lambda_2^{(t)} &\sim \text{Gamma} \left(\sum_i^n (y_i - u_i^{(t)}) + \alpha_2, n + \beta_2 \right), \\
\lambda_3^{(t)} &\sim \text{Gamma} \left(\sum_i^n u_i^{(t)} + \alpha_3, n + \beta_3 \right);
\end{aligned}$$

Passo 6: retorne para o passo 3 até obter a convergência.

5.3 Distribuição a posteriori do modelo BZIP a partir de dados ampliados

Como foi visto na Subseção 4.3.2, o modelo BZIP pode ser reescrito em termos do vetor ampliado (X, Y, U, Z) , onde (X, Y) são as variáveis de interesse e (U, Z) são as variáveis latentes.

Sejam $(x_1, y_1, u_1, z_1), \dots, (x_n, y_n, u_n, z_n)$ observações de uma amostra aleatória proveniente da f.p. (4.6). Seja $D_0 = \{i \in \{1, 2, \dots, n\} : (x_i, y_i, u_i) = (0, 0, 0)\}$ o conjunto de zeros na amostra. Assim a função de verossimilhança baseada nos dados ampliados $\mathcal{A} = (\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u})$ é

$$L_{\mathcal{A}}(\boldsymbol{\delta}) \propto \omega^S (1 - \omega)^{n-S} \lambda_1^{\sum_{i \notin D_0} (x_i - u_i)} e^{-(n-S)\lambda_1} \lambda_2^{\sum_{i \notin D_0} (y_i - u_i)} e^{-(n-S)\lambda_2} \lambda_3^{\sum_{i \notin D_0} u_i} e^{-(n-S)\lambda_3}, \quad (5.5)$$

sendo $x_i, y_i = 0, 1, \dots, u_i = 0, 1, \dots, \min(x_i, y_i), z_i = 0, 1, S = \sum_{i \in D_0} z_i$ e $\boldsymbol{\delta} = (\omega, \lambda_1, \lambda_2, \lambda_3)$.

A função de verossimilhança (5.5) sugere as seguintes distribuições a priori independentes para $\omega, \lambda_1, \lambda_2$ e λ_3

$$\begin{aligned} \pi(\omega) &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \omega^{a-1} (1-\omega)^{b-1}, \quad a, b > 0, \\ \pi(\lambda_i) &= \frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)} \lambda_i^{\alpha_i-1} e^{-\beta_i \lambda_i}, \quad \alpha_i, \beta_i > 0, \end{aligned}$$

onde $\Gamma(\cdot)$ é a função gamma e α_i, β_i, a, b , para $i = 1, 2, 3$, são hiperparâmetros conhecidos.

Assim, considerando as prioris independentes, a distribuição a priori conjunta é descrita como

$$\pi(\omega, \lambda_1, \lambda_2, \lambda_3) \propto \omega^{a-1} (1-\omega)^{b-1} \lambda_1^{\alpha_1-1} e^{-\beta_1 \lambda_1} \lambda_2^{\alpha_2-1} e^{-\beta_2 \lambda_2} \lambda_3^{\alpha_3-1} e^{-\beta_3 \lambda_3}. \quad (5.6)$$

Combinando a função de verossimilhança dada em (5.5) e a distribuição a priori conjunta (5.6), obtém-se a distribuição a posteriori conjunta para $\boldsymbol{\delta}$ dado $\mathcal{A} = (\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u})$ expressada como

$$\begin{aligned} \pi(\boldsymbol{\delta} | \mathcal{A}) &\propto \omega^{S+a-1} (1-\omega)^{n-S+b-1} \lambda_1^{\sum_{i \notin D_0} (x_i - u_i) + \alpha_1 - 1} e^{-(n-S+\beta_1)\lambda_1} \lambda_2^{\sum_{i \notin D_0} (y_i - u_i) + \alpha_2 - 1} e^{-(n-S+\beta_2)\lambda_2} \\ &\times \lambda_3^{\sum_{i \notin D_0} u_i + \alpha_3 - 1} e^{-(n-S+\beta_3)\lambda_3}. \end{aligned} \quad (5.7)$$

Da distribuição a posteriori conjunta (5.7) temos as distribuições condicionais completas necessárias

para aplicar o algoritmo Gibbs sampling com dados ampliados dadas por

$$\begin{aligned}\omega|\lambda_1, \lambda_2, \lambda_3, \mathcal{A} &\sim \text{Beta}(S + a, n - S + b); \\ \lambda_1|\omega, \lambda_2, \lambda_3, \mathcal{A} &\sim \text{Gamma}\left(\sum_{i \notin D_0} (x_i - u_i) + \alpha_1, n - S + \beta_1\right); \\ \lambda_2|\omega, \lambda_1, \lambda_3, \mathcal{A} &\sim \text{Gamma}\left(\sum_{i \notin D_0} (y_i - u_i) + \alpha_2, n - S + \beta_2\right); \\ \lambda_3|\omega, \lambda_1, \lambda_2, \mathcal{A} &\sim \text{Gamma}\left(\sum_{i \notin D_0} u_i + \alpha_3, n - S + \beta_3\right).\end{aligned}$$

É conveniente notar que $z_i = 0$ sempre que $i \notin D_0$, pois caso $(x_i, y_i, u_i) \neq (0, 0, 0)$ então a observação i não é proveniente da distribuição degenerada em $(0, 0, 0)$. Assim, a f.p. condicional da variável latente U_i , para $i \notin D_0$ é dada por

$$\begin{aligned}p(\mathbf{u}|\mathbf{x}, \mathbf{y}, \mathbf{z} = \mathbf{0}, \boldsymbol{\delta}) &= P(\mathbf{U} = \mathbf{u}|\mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y}, \mathbf{Z} = \mathbf{0}, \boldsymbol{\delta}) \\ &\propto P(\mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y}, \mathbf{U} = \mathbf{u}, \mathbf{Z} = \mathbf{0}|\boldsymbol{\delta}) \\ &= \prod_{i=1}^n P[X_i = x_i, Y_i = y_i, U_i = u_i, Z_i = 0|\boldsymbol{\delta}] \\ &= \prod_{i=1}^n (1 - \omega) \frac{\lambda_1^{x_i - u_i} \lambda_2^{y_i - u_i} \lambda_3^{u_i} e^{-\lambda \bullet}}{(x_i - u_i)!(y_i - u_i)!u_i!} \\ &\propto \prod_{i=1}^n \frac{1}{(x_i - u_i)!(y_i - u_i)!u_i!} \left(\frac{\lambda_3}{\lambda_2 \lambda_1}\right)^{u_i},\end{aligned}$$

com $u_i = 0, \dots, \min(x_i, y_i)$. Quando $i \in D_0$, tem-se que $P(U_i = u_i|X_i = x_i, Y_i = y_i, Z_i = 1, \boldsymbol{\delta}) = I_{\{0\}}(u_i)$.

Por outro lado, para $i \in D_0$, a variável aleatória Z_i condicionada a $X_i = x_i, Y_i = y_i, U_i = u_i$ e $\boldsymbol{\delta}$ tem distribuição de Bernoulli com parâmetro $p(\boldsymbol{\delta})$ dado por

$$\begin{aligned}p(\boldsymbol{\delta}) &= P(Z_i = 1|X_i = x_i, Y_i = y_i, U_i = u_i, \boldsymbol{\delta}) \\ &= \frac{P(X_i = x_i, Y_i = y_i, U_i = u_i, Z_i = 1|\boldsymbol{\delta})}{P(X_i = x_i, Y_i = y_i, U_i = u_i|\boldsymbol{\delta})} \\ &= \frac{\omega}{\omega + (1 - \omega)e^{-\lambda \bullet}}.\end{aligned}$$

Como $S = \sum_{i \in D_0} Z_i$, então a distribuição condicional de S dado $\mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y}, \mathbf{U} = \mathbf{u}$ e $\boldsymbol{\delta}$ é Binomial

com parâmetros n_0 e $p(\boldsymbol{\delta})$, em que $n_0 = \sum_{i=1}^n I_{\{(0,0)\}}(x_i, y_i)$ é o número de zeros na amostra.

Portanto, o algoritmo Gibbs sampling com dados ampliados é dado por:

Passo 1: inicialize o contador $t = 0$;

Passo 2: especifique valores iniciais $\boldsymbol{\delta}^{(0)} = (\omega^{(0)}, \lambda_1^{(0)}, \lambda_2^{(0)}, \lambda_3^{(0)})$;

Passo 3: faça $t = t + 1$;

Passo 4: gere valores para $\mathbf{u}^{(t)}$ e $s^{(t)}$ a partir das distribuições condicionais

$$\begin{aligned} U^{(t)} &\sim p(\mathbf{u}|\mathbf{x}, \mathbf{y}, \mathbf{z} = \mathbf{0}, \boldsymbol{\delta}^{(t-1)}), \\ S^{(t)} &\sim \text{Binomial}\left(n_0, p(\boldsymbol{\delta}^{(t-1)})\right); \end{aligned}$$

Passo 5: gere os valores de $\boldsymbol{\delta}^{(t)}$ como

$$\begin{aligned} \omega^{(t)} &\sim \text{Beta}\left(s^{(t)} + a, n - s^{(t)} + b\right), \\ \lambda_1^{(t)} &\sim \text{Gamma}\left(\sum_{i \notin D_0} (x_i - u_i^{(t)}) + \alpha_1, n - s^{(t)} + \beta_1\right), \\ \lambda_2^{(t)} &\sim \text{Gamma}\left(\sum_{i \notin D_0} (y_i - u_i^{(t)}) + \alpha_2, n - s^{(t)} + \beta_2\right), \\ \lambda_3^{(t)} &\sim \text{Gamma}\left(\sum_{i \notin D_0} u_i^{(t)} + \alpha_3, n - s^{(t)} + \beta_3\right); \end{aligned}$$

Passo 6: retorne para o passo 3 até obter a convergência.

5.4 Aplicação

A fim de ilustrar a análise bayesiana dos modelos PB e BZIP, utilizamos os dados reais apresentados na Tabela 2.1 e discutidos na Seção 2.5. Para obter amostras da distribuição a posteriori de cada modelo e calcular as estimativas bayesianas dos parâmetros, utilizamos os métodos MCMC, particularmente o algoritmo Gibbs sampling com dados ampliados, o qual foi desenvolvido no software R.

Considerando a metodologia apresentada na Seção 5.2, adotamos distribuições a priori não informativas, isto é distribuições com variâncias grandes, para os parâmetros $\lambda_i \sim \text{Gamma}(\alpha_i, \beta_i)$, para $i = 1, 2, 3$ com hiperparâmetros $\alpha_i = \beta_i = 0.005$.

Partindo de diferentes valores iniciais, foram geradas duas cadeias de 200.000 iterações. A convergência das cadeias foi verificada através do critério de convergência de Gelman & Rubin (1992) e pela análise gráfica de suas médias ergódicas dadas na Figura 5.1, onde a trajetória das médias ergódicas apresenta um comportamento estável ao longo das iterações indicando a convergência das cadeias. Destas cadeias foram descartadas as 30.000 primeiras como burn-in e considerando um salto de tamanho 50, para evitar a autocorrelação entre os valores gerados, resultando em uma amostra final de tamanho 3.400 para cada cadeia.

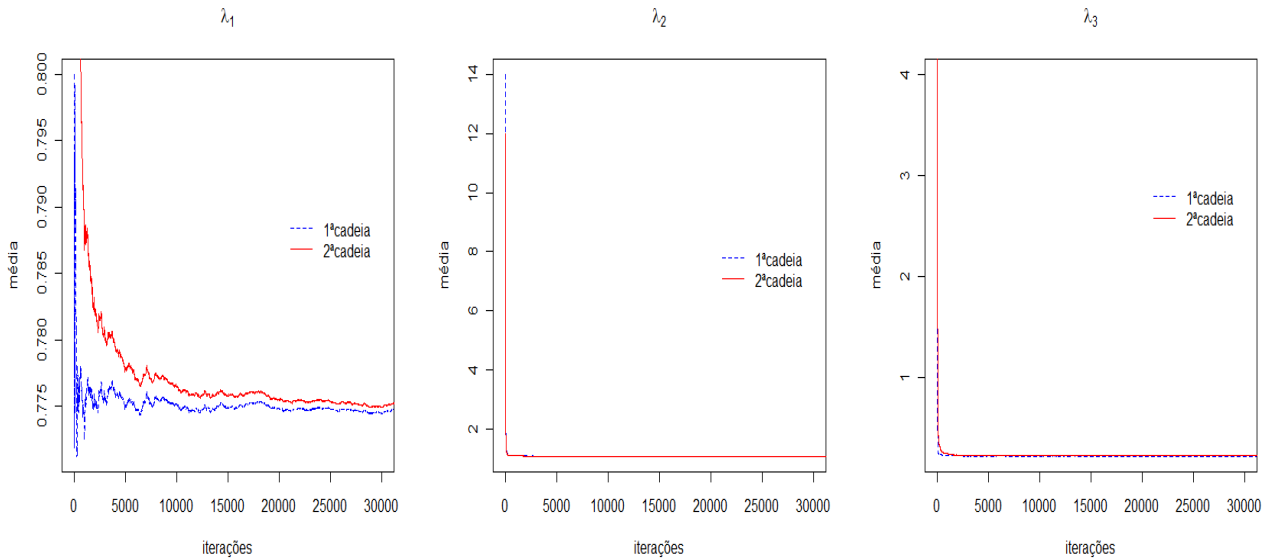


Figura 5.1: Médias ergódicas das cadeias ao longo das iterações

A partir de uma das cadeias, obtém-se o sumário a posteriori dos parâmetros do modelo PB dado na Tabela 5.1 e os gráficos da Figura 5.2. Estes gráficos apresentam as densidades marginais a posteriori do modelo PB e a autocorrelação da cadeia gerada, é observa-se que os valores da cadeia são independentes.

Tabela 5.1: Sumário a posteriori dos parâmetros do modelo PB

Parâmetro	Média	Desvio padrão	Intervalo de credibilidade de 95%
λ_1	0,7805	0,198	[0,68; 0,87]
λ_2	1,0720	0,295	[0,96; 1,17]
λ_3	0,2263	0,123	[0,14; 0,30]

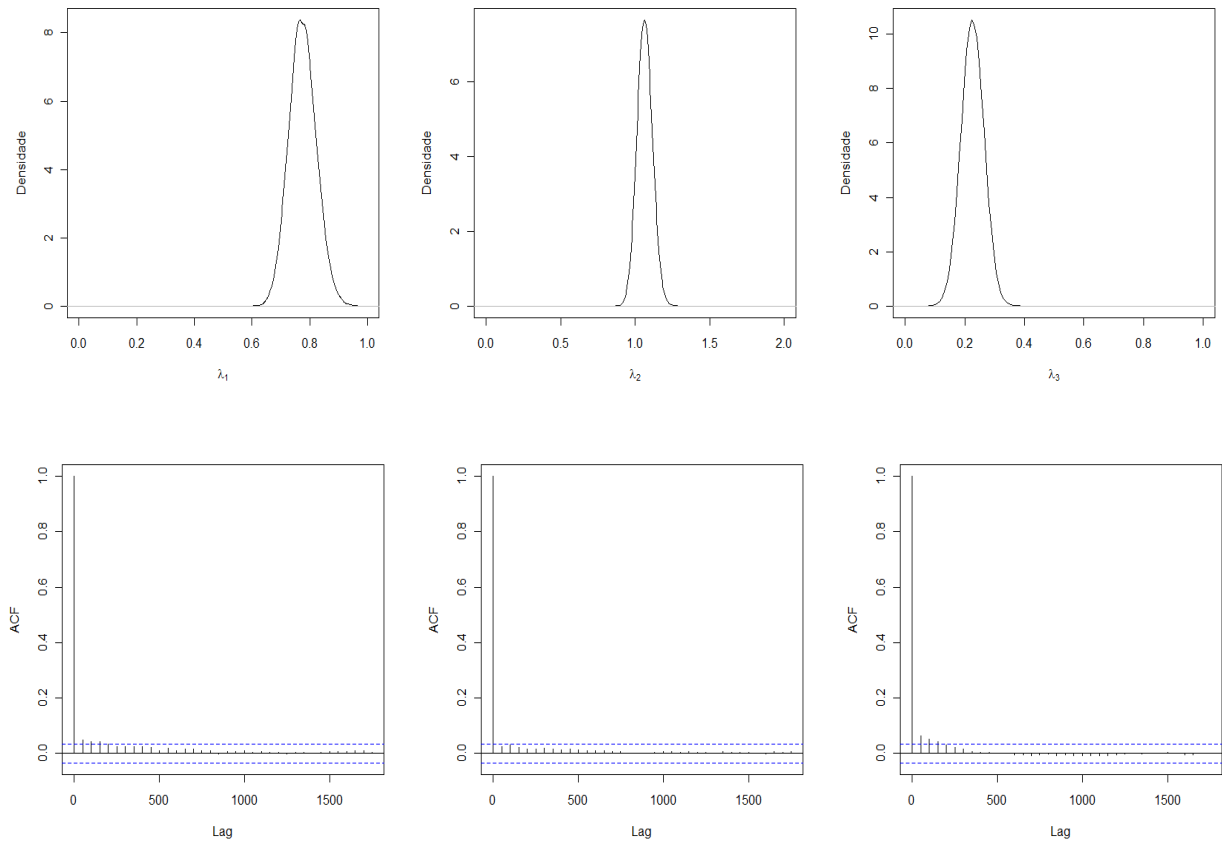


Figura 5.2: Densidades marginais a posteriori do modelo PB e autocorrelação da cadeia gerada

Para a obtenção da distribuição a posteriori do modelo BZIP, adotamos as respectivas prioris não informativas, $\lambda_i \sim \text{Gamma}(\alpha_i, \beta_i)$, para $i = 1, 2, 3$ e $\omega \sim \text{Beta}(a, b)$ com hiperparâmetros $\alpha_i = \beta_i = 0.005$ e $a = b = 1$.

A partir das distribuições condicionais completas da Seção 5.3, geramos duas cadeias de 200.000 iterações, com diferentes valores iniciais. A convergência das cadeias foi verificada através do critério de convergência de Gelman & Rubin (1992) e pela análise gráfica das suas médias ergódicas dadas na Figura 5.3, onde a trajetória das médias ergódicas apresenta um comportamento estável ao longo das iterações indicando a convergência das cadeias. Destas cadeias foram descartadas as 8.000 primeiras como burn-in e considerando um salto de tamanho 160, para evitar a autocorrelação entre os valores gerados, resultando uma amostra final de tamanho 1.200 para cada cadeia.

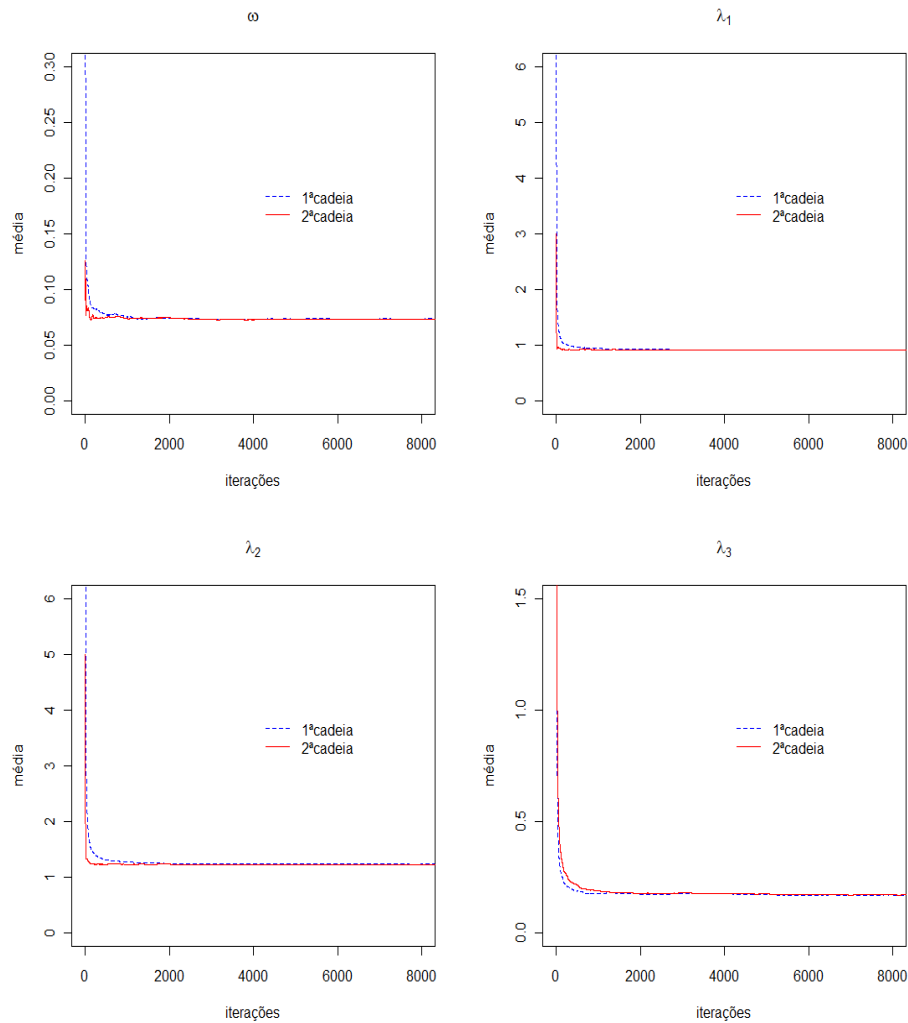


Figura 5.3: Médias ergódicas das cadeias geradas ao longo das iterações

A partir de uma das cadeias, obtém-se o sumário a posteriori dos parâmetros do modelo BZIP dado na Tabela 5.2 e os gráficos da Figura 5.2. Estes gráficos apresentam as densidades marginais a posteriori do modelo BZIP e a autocorrelação da cadeia gerada, em que se observa que os valores da cadeia são independentes.

Tabela 5.2: Sumário a posteriori dos parâmetros do modelo BZIP

Parâmetro	Média	Desvio padrão	Intervalo de credibilidade de 95%
ω	0,07	0,027	[0,03; 0,10]
λ_1	0,92	0,411	[0,79; 1,05]
λ_2	1,25	0,977	[1,09; 1,37]
λ_3	0,17	0,459	[0,06; 0,24]

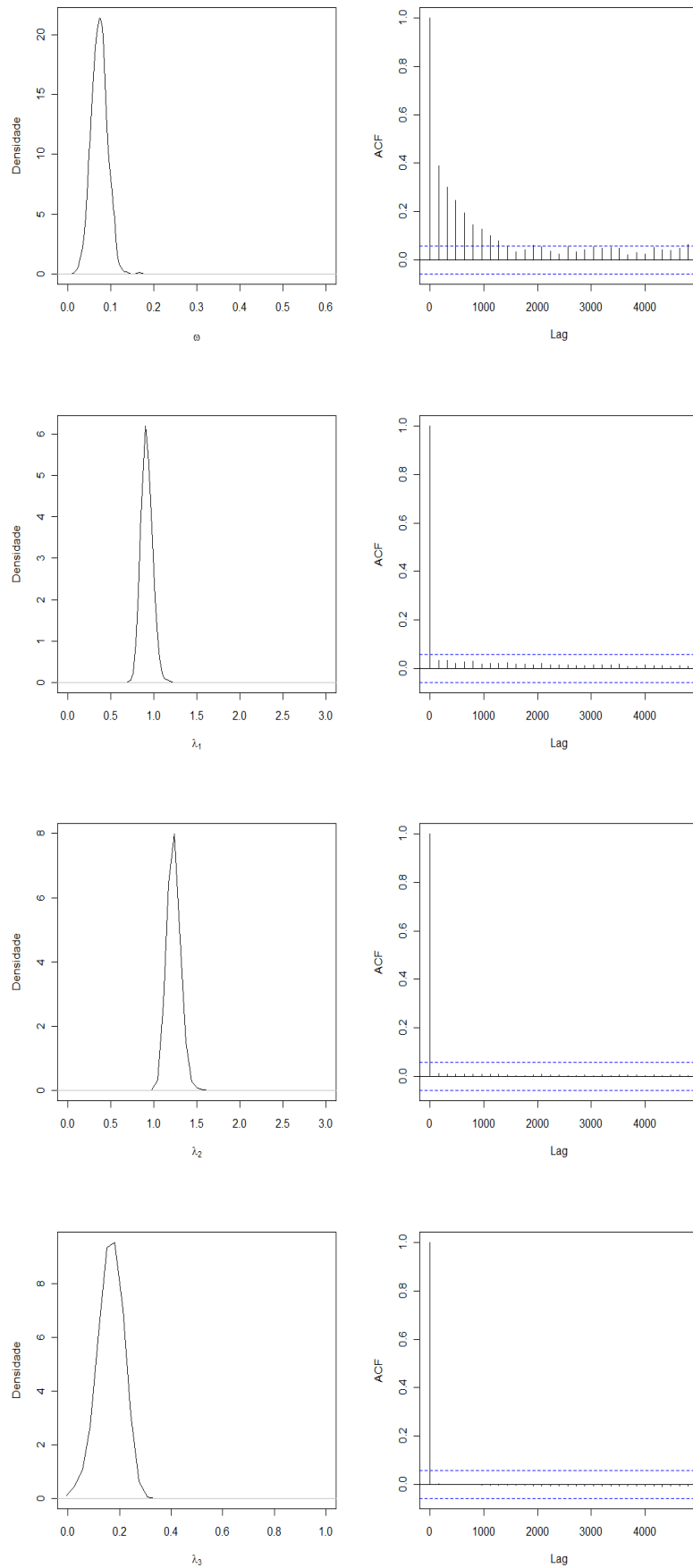


Figura 5.4: Densidades marginais a posteriori do modelo BZIP e autocorrelação da cadeia

Capítulo 6

Conclusões e Propostas Futuras

Neste trabalho foram apresentadas inicialmente as distribuições PB e PGB, suas propriedades básicas, assim como métodos inferenciais para os parâmetros. A construção destas distribuições é baseada no método da redução trivariada, que torna suas distribuições marginais positivamente correlacionadas.

Para ilustrar os métodos, consideramos um conjunto de dados reais sobre acidentes de ônibus ocorridos com 708 motoristas ao norte da Irlanda, durante os períodos 1952 – 53 e 1954 – 55. Os ajustes dos modelos a este conjunto de dados foram ilustrados a partir da comparação entre frequências observadas e esperadas. O teste da razão de verossimilhança e os critérios de seleção AIC e BIC permitiram concluir que os modelos que não incorporam dependência entre as suas distribuições marginais são inadequados para o conjunto de dados em questão. Para a comparação entre os dois modelos discutidos, novamente considerou-se o teste da razão de verossimilhança e os critérios de seleção AIC e BIC, sendo que por todos estes critérios o modelo PGB mostrou-se o mais adequado para este conjunto de dados.

Observamos que o conjunto de dados reais discutido apresenta uma quantidade de zeros excessiva, o que motivou a inclusão da distribuição BZIP neste trabalho. O método de máxima verossimilhança foi utilizado para estimar os parâmetros desta distribuição. No entanto, não foi possível obter os estimadores de máxima verossimilhança explicitamente, devido a complexidade das expressões envolvidas. Consideramos então o algoritmo EM para obter as estimativas dos parâmetros. Neste caso, a obtenção das expressões dos passos E e M do algoritmo foram relativamente simples. Através do teste da razão de verossimilhança, concluímos que modelos particulares a distribuição BZIP não são adequados para o conjunto de dados em questão.

Os ajustes dos modelos PB, PGB e BZIP aos dados foram ilustrados a partir da comparação entre as frequências observadas e esperadas. Para a comparação entre estes três modelos utilizamos os critérios de seleção de modelos AIC e BIC, que sugerem que o modelo BZIP é o mais adequado.

Especificamente, para os modelos PB e BZIP, apresentamos uma metodologia para a realização da inferência bayesiana, conduzida por meio do algoritmo Gibbs sampling com dados ampliados. Portanto, a inclusão de variáveis latentes na modelagem simplificou a tarefa de fazer inferência sobre os parâmetros.

Como sugestão para desenvolvimentos futuros, listamos a seguir algumas possibilidades:

- Os modelos bivariados discutidos tem a limitação de apenas acomodar dependência positiva, sendo

importante, desenvolver métodos capazes de modelar também a dependência negativa;

- Implementar métodos inferenciais para o modelo Poisson generalizado bivariado com inflação de zeros;
- Estudar uma classe de modelos bivariados mais geral, como por exemplo, a classe das distribuições series de potências bivariadas inflacionadas de zeros discutida em Krishna & Tukaram (2011).

Referências

- Aitken, A. (1936). A further note on multivariate selection. *Proceedings of the Edinburgh Mathematical Society (Series 2)*, **5**(1), 37–40.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**(6), 716–723.
- Albert, J. H. & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American statistical Association*, **88**(422), 669–679.
- Arab, A., Holan, S. H., Wikle, C. K. & Wildhaber, M. L. (2011). Semiparametric bivariate zero-inflated Poisson models with application to studies of abundance for multiple species. *ArXiv preprints :1105.3169*.
- Bayarri, M., Berger, J. O. & Datta, G. S. (2008). Objective bayes testing of Poisson versus inflated Poisson models. *Pushing the Limits of Contemporary Statistics: Contributions in honor of Jayanta K. Ghosh. IMS Collections*, **3**, 105–121.
- Campbell, J. (1934). The Poisson correlation function. *Proceedings of the Edinburgh Mathematical Society*, **4**(1), 18–26.
- Cassela, G. & George, I. (1992). Explaining the gibbs sampler. *The American Statistician*, **46**, 167–174.
- Consul, P. C. & Jain, G. C. (1973). A generalization of the Poisson distribution. *Technometrics*, **15**(4), 791–799.
- Cresswell, W. L. & Froggatt, P. (1963). The causation of bus driver accidents; an epidemiological study.
- Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, **39**(1), 1–38.
- Diebolt, J. & Robert, C. P. (1994). Estimation of finite mixture distributions through Bayesian sampling. *Journal of the Royal Statistical Society. Series B (Methodological)*, **56**(2), 363–375.
- Famoye, F. & Consul, P. (1995). Bivariate generalized Poisson distribution with some applications. *Metrika*, **42**(1), 127–138.

-
- Gelman, A. & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, **7**(4), 457–472.
- Ho, L. L. & Singer, J. d. M. (2001). Generalized least squares methods for bivariate Poisson regression. *Communications in Statistics-Theory and Methods*, **30**(2), 263–277.
- Holgate, P. (1964). Estimation for the bivariate Poisson distribution. *Biometrika*, **51**(1-2), 241–245.
- Johnson, N. L., Kemp, A. W. & Kotz, S. (2005). *Univariate discrete distributions*, volume 444. John Wiley & Sons.
- Jung, R. C. & Winkelmann, R. (1993). Two aspects of labor mobility: a bivariate Poisson regression approach. *Empirical Economics*, **18**(3), 543–556.
- Karlis, D. & Ntzoufras, I. (2005). Bivariate Poisson and diagonal inflated bivariate Poisson regression models in R. *Journal of Statistical Software (to appear)*, **14**, 1–36.
- Karlis, D. & Ntzoufras, J. (2003). Bayesian and non-bayesian analysis of soccer data using bivariate Poisson regression models. In *16th Panhellenic Conference in Statistics. Kavala*, volume 20.
- Karlis, D. & Tsiamyrtzis, P. (2008). Exact bayesian modeling for bivariate Poisson data and extensions. *Statistics and Computing*, **18**, 27–40.
- Kawamura, K. (1973). The structure of bivariate Poisson distribution. *Kodai Mathematical Seminar Reports*, **25**, 246–256.
- Kawamura, K. (1985). A note on the recurrent relations for the bivariate Poisson distribution. *Kodai Mathematical Seminar Reports*, **8**, 70–78.
- Kocherlakota, S. & Kocherlakota, K. (1992). *Bivariate discrete distributions*, volume 132. CRC Press.
- Kocherlakota, S. & Kocherlakota, K. (2001). Regression in the bivariate Poisson distribution. **30**, 815–825.
- Krishna, P. M. & Tukaram, S. D. (2011). Bivariate zero-inflated power series distribution. *Applied Mathematics*, **2**(7), 824–829.
- Lai, C. (1995). Construction of bivariate distributions by a generalised trivariate reduction technique. *Statistics & probability letters*, **25**(3), 265–270.
- Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, **34**(1), 1–14.
- Li, C. S., Lu, J. C., Park, J., Kim, K., Brinkley, P. A. & Peterson, J. P. (1999). Multivariate zero-inflated Poisson models and their applications. *Technometrics*, **41**(1), 29–38.
- Majumdar, A. & Gries, C. (2010). Bivariate zero-inflated regression for count data: a bayesian approach with application to plant counts. *The international journal of biostatistics*, **1**, 1–26.

-
- McLachlan, G. & Krishnan, T. (2007). *The EM algorithm and extensions*, volume 382. John Wiley & Sons.
- Ridout, M., Demétrio, C. G. & Hinde, J. (1998). Models for count data with many zeros. In *Proceedings of the XIXth International Biometric Conference*, volume 19, pages 179–192.
- Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, **6**(2), 461–464.
- Stern, J. & Zacks, S. (2002). Testing the independence of Poisson variates under the holgate bivariate distribution: the power of a new evidence test. *Statistics & probability letters*, **60**, 313–320.
- Tanner, M. A. & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American statistical Association*, **82**(398), 528–540.
- Teicher, H. (1954). On the multivariate Poisson distribution. *Scandinavian Actuarial Journal*, **1954**(1), 1–9.
- Tsionas, E. G. (1999). Bayesian analysis of the multivariate Poisson distribution. *Communications in Statistics-Theory and Methods*, **28**(2), 431–451.
- Van den Broek, J. (1995). A score test for zero inflation in a Poisson distribution. *Biometrics*, **51**(2), 738–743.
- Vernic, R. (1997). On the bivariate generalized Poisson distribution. *Astin Bulletin*, **27**, 23–32.
- Walhin, J. F. (2001). Bivariate zip models. *Biometrical Journal*, **43**(2), 147–160.