

Resumo

Nesta dissertação introduzimos a distribuição Poisson-Lindley zero modificada (ZMPL - *Zero Modified Poisson-Lindley*) e obtemos algumas propriedades. Este modelo é indexado por dois parâmetros, sendo que o parâmetro adicional, ν , tem uma interpretação natural em termos tanto da deflação como da inflação de zeros. Com o intuito de reduzirmos os vieses dos estimadores de máxima verossimilhança, consideramos o mecanismo de correção de viés via *bootstrap*. Apresentamos resultados das simulações de Monte Carlo, as quais foram utilizadas a fim de verificarmos o comportamento dos estimadores de máxima verossimilhança dos parâmetros da distribuição ZMPL, bem como os de suas versões corrigidas, em amostras pequenas e grandes. O modelo foi aplicado a dois conjuntos de dados reais e considerado melhor que seus concorrentes.

Abstract

In this work we present the distribution modified Poisson-Lindley (ZMPL), with some properties and study of simulation. The additional parameter ν of the ZMPL has a natural interpretation in terms of either zero-deflated/inflated proportion. Inference is dealt with by using the likelihood approach. In particular the maximum likelihood estimators of the distribution's parameter are compared in small and large samples. We also consider an alternative bias-correction mechanism based on Efron's bootstrap resampling. The model is applied to real data sets and found to perform better than other competing models.

Universidade Federal de Campina Grande
Centro de Ciências e Tecnologia
Programa de Pós-Graduação em Matemática
Curso de Mestrado em Matemática

Distribuição Poisson-Lindley Zero Modificada

por

Danillo Magalhães Xavier Assunção [†]

sob orientação do

Prof. Dr. Manoel Ferreira dos Santos Neto

Dissertação apresentada ao Corpo Docente do Programa de Pós-Graduação em Matemática - CCT - UFCG, como requisito parcial para obtenção do título de Mestre em Matemática.

[†]Este trabalho contou com apoio financeiro da CAPES

Distribuição Poisson-Lindley Zero Modificada

por

Danillo Magalhães Xavier Assunção

Dissertação apresentada ao Corpo Docente do Programa de Pós-Graduação em Matemática - CCT - UFCG, como requisito parcial para obtenção do título de Mestre em Matemática.

Área de Concentração: Probabilidade e Estatística

Aprovada por:

Prof. Dr. Alexsandro Bezerra Cavalcanti - UFCG

Prof. Dr. Davis Matias de Oliveira - UEPB

Prof. Dr. Manoel Ferreira dos Santos Neto - UFCG

Orientador

Universidade Federal de Campina Grande

Centro de Ciências e Tecnologia

Programa de Pós-Graduação em Matemática

Curso de Mestrado em Matemática

Fevereiro/2018

Agradecimentos

- A Deus, por mais uma conquista. Por ele ter me dado forças para seguir em frente sem desistir.
- A minha família e namorada, que nos momentos difíceis me apoiaram bastante.
- Ao professor Manoel Ferreira dos Santos Neto, pela orientação, dedicação e compreensão, contribuindo para o desenvolvimento desta dissertação.
- Aos professores Manoel Ferreira dos Santos Neto e Michelli Karinne Barros da Silva, pois acima de tudo eles foram grandes amigos, que nos deram aprendizados importantes, coragem e sabedoria.
- Aos colegas do mestrado, que foram companheiros de alegrias e angustias.
- À Andrezza, por sua competência, compreensão e presteza aos alunos do mestrado.
- Aos participantes da banca examinadora, pelas sugestões.
- À CAPES, pelo apoio financeiro.

Dedicatória

A todos aqueles que me apoiaram, amigos, meus pais e irmãos.

Conteúdo

Introdução	6
1 Distribuição Poisson-Lindley zero modificada	8
1.1 Distribuição Poisson-Lindley	8
1.2 Distribuição Poisson-Lindley zero modificada	8
1.2.1 Função de Probabilidade	8
1.2.2 Propriedades	9
1.3 Estimação pontual dos parâmetros	14
1.3.1 Método dos Momentos	15
1.3.2 Método de Máxima Verossimilhança	15
1.3.3 Correção de viés	17
1.4 Estimação Intervalar	18
1.4.1 Intervalo de Confiança Assintótico (ICa)	18
1.4.2 Intervalo de Confiança Percentil (ICp)	19
1.5 Teste de Hipóteses	19
1.5.1 Teste gradiente	19
2 Simulação	21
3 Aplicações	24
3.1 Contagem de lesões de dosimetria citogenética	24
3.2 Contagem da ocorrência de greves	25
4 Conclusões	28

A	Demonstrações	29
A.1	Relação de recorrência da função geradora de probabilidade	29
A.2	Momentos de primeira e segunda ordem	30
A.3	Coefficiente de assimetria	32
A.4	Coefficiente de curtose	35
A.5	Estimadores de Momentos	39

Introdução

A distribuição Poisson-Lindley (PL), uni-paramétrica, foi inicialmente proposta por Lindley (1958). Sankaran (1970) utilizou este modelo para modelar o número de erros na cópia de grupos de dígitos aleatório e o número de acidentes sofridos por mulheres durante o trabalho. A distribuição PL também é conhecida, na literatura, como distribuição Poisson generalizada (CONSUL, 1989). Posteriormente, Ghitany e Al-Mutairi (2009) estudaram algumas propriedades deste modelo. Por fim, Ghitany, Al-Mutairi e Nadarajah (2008) propõem a versão truncada, no zero, da distribuição Poisson-Lindley e Borah e Nath (2001) introduziram a distribuição Poisson-Lindley inflacionada de zeros.

A distribuição Poisson-Lindley inflacionada de zeros não é capaz de modelar dados com deflação de zeros, pois seu parâmetro de “modificação” está definido no intervalo $(0, 1)$. Desta forma, o nosso primeiro objetivo é estender o espaço paramétrico do parâmetro de “modificação” possibilitando desta maneira o estudo de dados truncados, inflacionados e deflacionados, no zero, através de um único modelo probabilístico que chamaremos de distribuição Poisson-Lindley zero modificada.

Uma importante área de pesquisa é o estudo do comportamento dos estimadores de máxima verossimilhança. No cenário de grande amostras o viés destes estimadores não representam um problema, pois em geral é da ordem de $O(n^{-1})$. No entanto, no caso de pequenas amostras a existência de viés pode ser problemática. Sendo assim, é necessária a utilização de algum método para correção de viés. Existem na literatura uma grande quantidade de trabalhos que utilizam os mais diferentes métodos de correção de viés. Especificamente, para este trabalho iremos analisar o comportamento do viés dos parâmetros que indexam a Poisson-Lindley zero modificada,

com a utilização de correção baseada no método *bootstrap* (EFRON, 1979).

Além do estudo das estimativas pontuais dos parâmetros iremos utilizar intervalos de confiança para associar um grau de confiabilidade a estas estimativas. Podemos obter intervalos de confiança através da distribuição assintótica do estimadores de máxima verossimilhança. Além disso, o método *bootstrap* pode ser utilizado para obtenção de estimativas intervalares. Portanto, nesta dissertação, além de analisar o comportamento dos estimadores de máxima verossimilhança corrigidos também iremos estudar o comportamento dos intervalos de confiança dos parâmetros que indexam a distribuição Poisson-Lindley zero modificada através de intervalos de confiança do tipo assintótico, *bootstrap-t* e *bootstrap* percentil.

Esta dissertação está dividida em quatro capítulos e contém um apêndice. No Capítulo 1, estudamos a distribuição Poisson-Lindley zero modificada, apresentando suas principais características e propriedades. Além disso, apresentamos a metodologia *bootstrap* e como utilizá-la para obter estimativas do viés de ordem $O(n^{-1})$. No Capítulo 2, realizamos simulações de Monte Carlo, a fim de verificarmos o desempenho dos estimadores de máxima verossimilhança dos parâmetros da distribuição Poisson-Lindley zero modificada, bem como o de suas versões corrigidas, em amostras de tamanho finito. No Capítulo 3, aplicamos a distribuição Poisson-Lindley zero modificada a dois conjuntos de dados reais. Por fim, no último capítulo, apresentamos as conclusões deste trabalho.

Capítulo 1

Distribuição Poisson-Lindley zero modificada

1.1 Distribuição Poisson-Lindley

Segundo Ghitany e Al-Mutairi (2009), uma variável aleatória X segue uma distribuição Poisson-Lindley discreta se sua função massa de probabilidade (f.m.p) é dada por

$$P(X = x) = \frac{\theta^2(x + \theta + 2)}{(\theta + 1)^{x+3}}, \quad x = 0, 1, 2, \dots, \quad \theta > 0,$$

com média e variância, dados, respectivamente por

$$E(X) = \frac{\theta + 2}{\theta(\theta + 1)} \quad \text{e} \quad \text{Var}(X) = \frac{\theta^3 + 4\theta^2 + 6\theta + 2}{\theta^2(\theta + 1)^2}.$$

1.2 Distribuição Poisson-Lindley zero modificada

1.2.1 Função de Probabilidade

Se uma variável aleatória discreta X tem distribuição Poisson-Lindley zero modificada, então sua f.m.p é dada por

$$P(X = x) = \begin{cases} \nu + (1 - \nu) \frac{\theta^2(\theta+2)}{(\theta+1)^3}, & k = 0, \\ (1 - \nu) \frac{\theta^2(x+\theta+2)}{(\theta+1)^{x+3}}, & k = 1, 2, \dots \end{cases} \quad (1.1)$$

Denotaremos esta distribuição como $ZMPL(\theta, \nu)$. Para o parâmetro ν , pressupomos que $-\frac{\theta^2(\theta+2)}{\theta^2+3\theta+1} \leq \nu \leq 1$. O parâmetro ν é chamado de *parâmetro de modificação no zero* e diferentes valores levam a diferentes modificações nesta distribuição.

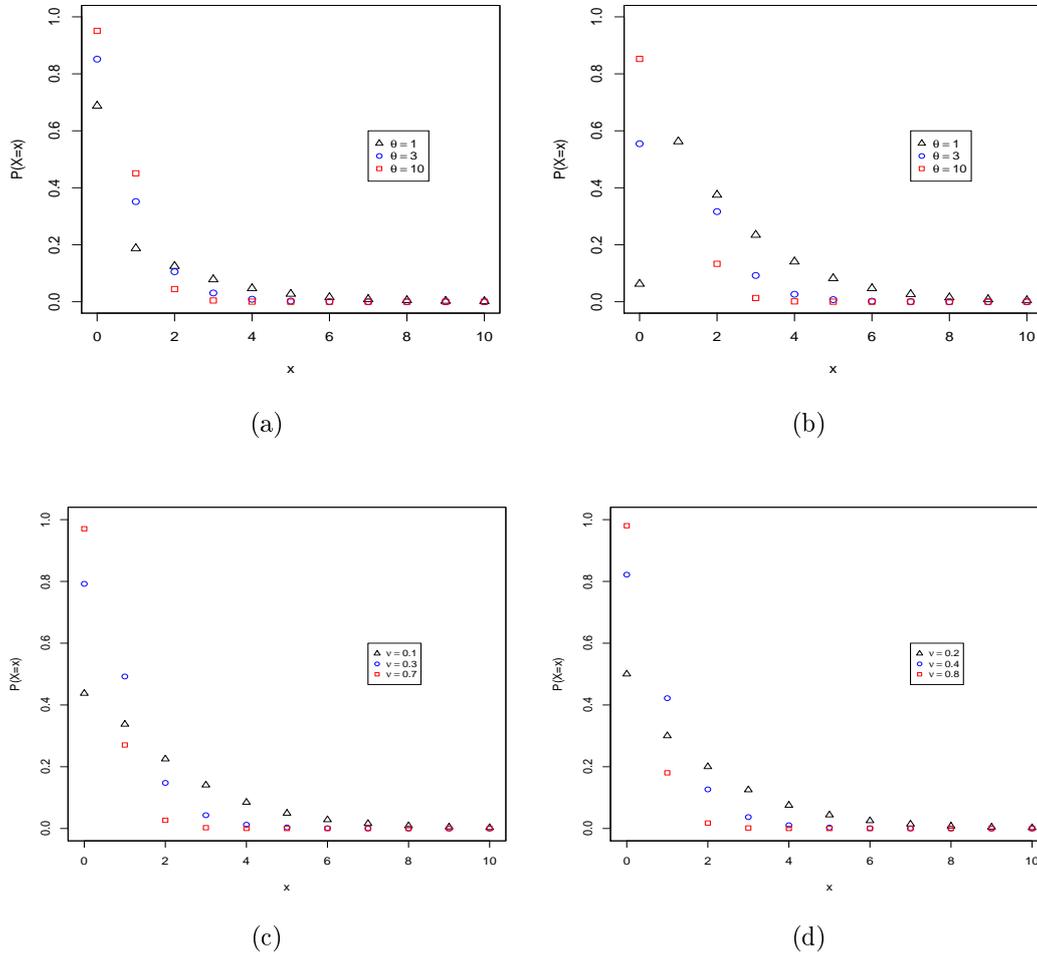
- Se $\nu = -\frac{\theta^2(\theta+2)}{\theta^2+3\theta+1}$, a distribuição (1.1) resulta na distribuição Poisson-Lindley truncada no zero (GHITANY; AL-MUTAIRI; NADARAJAH, 2008), em que o parâmetro ν é nulo e portanto, não há chance de obter uma observação zero na amostra;
- Se $\nu \in \left(-\frac{\theta^2(\theta+2)}{\theta^2+3\theta+1}, 0\right)$, isso produz uma distribuição Poisson-Lindley deflacionada de zeros, ou seja, ocorrem menos zero do que o esperado na distribuição Poisson-Lindley. Este modelo é conhecido como distribuição Poisson-Lindley deflacionada de zeros. Vale destacar que deflação de zeros raramente surge na prática;
- Se $\nu = 0$, temos a distribuição Poisson-Lindley (LINDLEY, 1958);
- Se $\nu \in (0, 1)$, temos como resultado a distribuição de Poisson-Lindley inflacionada de zeros (BORAH; NATH, 2001);
- Se $\nu = 1$, temos uma distribuição degenerada no zero.

A seguir, na Figura 1.1, mostramos os gráficos da função de probabilidade $ZMPL(\nu, \theta)$ fixando o valor de ν e variando o valor de θ , e vice versa.

1.2.2 Propriedades

Nesta seção apresentamos algumas propriedades importantes da distribuição Poisson-Lindley zero modificada. Alguns destes resultados são similares aos obtidos por Borah e Nath (2001).

Figura 1.1: Gráficos da função massa de probabilidade da ZMPL: (a) $\nu = 0.5$, (b) $\nu = -0.5$ (c) $\theta = 1.0$ e (d) $\theta = 1.0$.



Fonte: Próprio autor.

1.2.2.1 Função Geradora de Momentos

Nesta seção apresentaremos a função geradora de probabilidade (f.g.p) de X denotada por

$$\varphi_X(s) = E(s^X) = \nu + (1 - \nu) \frac{\theta^2}{(\theta + 1)} \cdot \frac{(2 + \theta - s)}{(\theta + 1 - s)^2}. \quad (1.2)$$

Consequentemente, a função geradora de momentos (f.g.m), denotado por $M_X(s) = \varphi_X(e^s)$, é dada por

$$M_X(s) = \nu + (1 - \nu) \frac{\theta^2}{\theta + 1} \cdot \frac{(2 + \theta - e^s)}{(\theta + 1 - e^s)^2},$$

e a função geradora de momentos fatorial (f.g.m.f) pode ser expressa por

$$M_{[X]}(s) = E(1+s)^X = \varphi_X(1+s) = \nu + (1-\nu) \frac{\theta^2}{(\theta+1)} \cdot \frac{(1+\theta-s)}{(\theta-s)^2}.$$

O r -ésimo momento fatorial, da distribuição Poisson Lindley Modificada no zero, é dado pela r -ésima derivada de $M_{[X]}(s)$ com relação a s e tomando $s = 0$. Assim, os r -ésimo momento fatorial da ZMPL (SHANKER; FESSHAYE, 2015) são dados por

$$E[(X)_r] = (1-\nu) \frac{r!(\theta+r+1)}{\theta^r(\theta+1)}, \quad r = 1, 2, 3, \dots$$

Em particular, os quatros primeiros momentos de X são:

$$E(X) = (1-\nu) \frac{\theta+2}{\theta(\theta+1)},$$

$$E(X^2) = (1-\nu) \frac{(\theta+2)^2+2}{\theta^2(\theta+1)},$$

$$E(X^3) = (1-\nu) \left[\frac{\theta+2}{\theta(\theta+1)} + \frac{6(\theta+3)}{\theta^2(\theta+1)} + \frac{6(\theta+4)}{\theta^3(\theta+1)} \right],$$

e

$$E(X^4) = (1-\nu) \left[\frac{\theta+2}{\theta(\theta+1)} + \frac{14(\theta+3)}{\theta^2(\theta+1)} + \frac{36(\theta+4)}{\theta^3(\theta+1)} + \frac{24(\theta+5)}{\theta^4(\theta+1)} \right].$$

O k -ésimo momento central, da distribuição ZMPL é definido por

$$\mu_k = E[(X - E(X))^k] = \sum_{j=0}^k \binom{k}{j} E(X^j) [-E(X)]^{k-j}, \quad k = 1, 2, 3, \dots,$$

e como consequência temos que

$$\begin{aligned}\mu_2 &= \frac{(1-\nu)(\theta^3 + 4\theta^2 + 6\theta + 2 + \nu(\theta + 2)^2)}{\theta^2(\theta + 1)^2}, \\ \mu_3 &= \frac{(1-\nu)}{\theta^3(\theta + 1)^2} \left[\frac{(\theta^5 + 10\theta^4 + 41\theta^3 + 80\theta^2 + 72\theta + 24)}{(\theta + 1)} - \right. \\ &\quad \left. (1-\nu)(\theta^3 + 18\theta^2 + 42\theta + 36) + \frac{2(1-\nu)^2(\theta^3 + 6\theta^2 + 12\theta + 8)}{(\theta + 1)} \right], \\ \mu_4 &= \frac{(1-\nu)}{\theta^4(\theta + 1)^4} \left[(\theta^4 + 16\theta^3 + 78\theta^2 + 168\theta + 120) - \right. \\ &\quad \frac{4(1-\nu)(\theta^6 + 12\theta^5 + 61\theta^4 + 162\theta^3 + 232\theta^2 + 168\theta + 48)}{(\theta + 1)^3} + \\ &\quad \frac{6(1-\nu)^2(\theta^4 + 8\theta^3 + 26\theta^2 + 40\theta + 24)}{(\theta + 1)^2} - \\ &\quad \left. \frac{3(1-\nu)^3(\theta^4 + 8\theta^3 + 24\theta^2 + 32\theta + 16)}{(\theta + 1)^3} \right].\end{aligned}$$

1.2.2.2 Função Geradora de Probabilidades

A f.g.p, $\varphi_X(s)$, da distribuição ZMPL, dada em (1.2), pode ser escrita como

$$\varphi_X(s) = \nu + (1-\nu)\varphi_Y(s), \quad (1.3)$$

em que $\varphi_Y(s) = \frac{\theta^2(\theta+2-t)}{(\theta+1)(\theta+1-t)^2}$ é a f.g.p da distribuição Poisson-Lindley. Diferenciando (1.3) com relação a t e igualando os coeficientes de t^r em ambos os lados, temos

$$P_r = \frac{(\theta + 2 + r)}{(\theta + 1)(\theta + 1 + r)} P_{r-1}, \quad r > 1,$$

em que $P_0 = \nu + \frac{(1-\nu)\theta^2(\theta+2)}{(\theta+1)^3}$ e $P_1 = \frac{(1-\nu)\theta^2(\theta+3)}{(\theta+1)^4}$. A demonstração deste resultado pode ser encontrada no Apêndice A.

1.2.2.3 Momentos

Na estatística, o n -ésimo momento ou momento de ordem n de uma variável aleatória X é calculado através da expressão: $E(X^n)$. Os momentos são muito importantes, pois através deles conseguimos caracterizar as distribuições de probabilidades. O primeiro e segundo momentos, da distribuição Poisson-Lindley modificado no zero, são dados,

respectivamente, por

$$\mu'_1 = \frac{(1-\nu)(\theta+2)}{\theta(\theta+1)} \quad \text{e} \quad \mu'_2 = \frac{(1-\nu)(\theta^2+4\theta+6)}{\theta^2(\theta+1)}.$$

Sabemos que $\text{Var}(X) = \mu'_2 - [\mu'_1]^2$, portanto, temos que a variância é dada por

$$\mu_2 = \frac{(1-\nu)[\theta^3+4\theta^2+6\theta+2+\nu(\theta+2)^2]}{\theta^2(\theta+1)^2}.$$

A demonstração deste resultado pode ser encontrada no Apêndice A.

1.2.2.4 Coeficientes de Assimetria e Curtose

A expressão para o coeficiente de assimetria apresentada em Borah e Nath (2001) apresenta um pequeno erro. Portanto, a expressão correta, escrita em função de θ e ν , é dada por

$$\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}} = \frac{G}{F},$$

em que $G = \nu^2(2\theta^3+12\theta^2+24\theta+16) + \nu(3\theta^4+17\theta^3+36\theta^2+30\theta+4) + \theta^5+7\theta^4+22\theta^3+18\theta+4$ e $F = \sqrt{(1-\nu)[\theta^3+4\theta^2+6\theta+\nu(\theta+2)^3]}$. Já a expressão para o coeficiente de curtose pode ser escrita em função de θ e ν da seguinte forma

$$\gamma_2 = \frac{\mu_4}{\mu_2} - 3 = \frac{A + B\nu + 3\nu^2C + 3\nu^3D}{(1-\nu)\{\theta^3+4\theta^2+6\theta+2+\nu(\theta+2)^2\}^2}.$$

As demonstrações para os coeficientes de assimetria e curtose são apresentadas no Apêndice A.

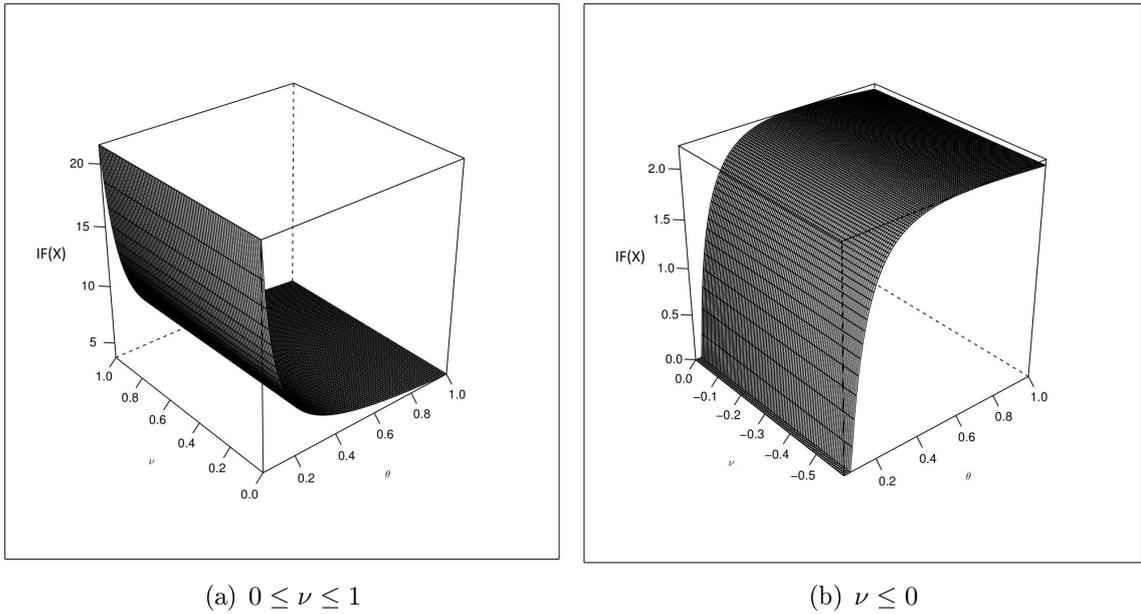
1.2.2.5 Índice de Dispersão de Fisher

O índice de dispersão de Fisher (IF) é frequentemente utilizado como uma medida de variação para uma variável aleatória X . Definido por $\text{IF}(X) = \text{Var}(X)/\text{E}(X)$, este índice é uma medida de agregação ou desagregação; para maiores detalhes ver Johnson, Kemp e Kotz (2005). O IF para a distribuição Poisson-Lindley modificada no zero é dado por

$$\text{IF}(X) = \left\{ \frac{\text{E}(X^2)}{\text{E}(X)} - \text{E}(X) \right\} = \nu \mu_{PL} + \frac{(\theta^3+4\theta^2+6\theta+2)}{\theta(\theta+1)(\theta+2)} = \nu \mu_{PL} + \text{IF}(Y),$$

em que μ_{PL} e $IF(Y)$ são, respectivamente, a média e o índice de Fisher da distribuição Poisson-Lindley. Assim a distribuição ZMPL apresenta subdispersão quando $\theta > \sqrt{2}$ e $\nu \in \left[-\frac{\theta^2(\theta+2)}{\theta^2+3\theta+1}, 0\right)$; e superdispersão quando $\nu \in [0, 1)$. A Figura 1.2 mostra a forma do Índice de dispersão de Fisher quando variamos os parâmetros θ e ν . Note na Figura 1.3(a) que a ZMPL é um modelo superdisperso. Já na Figura 1.3(b) nota-se que a distribuição ZMPL pode ser um modelo sub- ou super-disperso.

Figura 1.2: Gráfico do Índice de dispersão de Fisher da distribuição ZMPL.



Fonte: Próprio autor.

1.3 Estimação pontual dos parâmetros

Seja X_1, X_2, \dots, X_n uma amostra aleatória de tamanho n da distribuição ZMPL, ou seja, os X_i 's são variáveis aleatórias independentes com função massa de probabilidade dada pela equação (1.1). Nesta seção estamos interessados na estimação dos parâmetros que indexam a distribuição ZMPL e para isso iremos utilizar o método de máxima verossimilhança. No entanto, para utilizar como chute inicial no processo iterativo, utilizado para maximizar o logaritmo da função de verossimilhança, iremos primeiro apresentar as expressões dos estimadores de momentos para os parâmetros θ e ν .

1.3.1 Método dos Momentos

Considere X_1, X_2, \dots, X_n uma amostra aleatória de tamanho n oriunda da distribuição ZMPL. A média amostral e quadrática são definidas, respectivamente, por

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{e} \quad s_{x^2} = \frac{1}{n} \sum_{i=1}^n x_i^2.$$

Os dois primeiros momentos populacionais são dados por $E(X) = (1 - \nu)(\theta + 2)/[\theta(\theta + 1)]$ e $E(X^2) = (1 - \nu)[(\theta + 2)^2 + 2]/[\theta^2(\theta + 1)]$. Desta forma, podemos montar o seguinte sistema de equações

$$E(X) = \bar{x} \quad \text{e} \quad E(X^2) = s_{x^2}.$$

Substituindo os respectivos valores da média amostral e quadrática nas equações acima, e resolvendo com relação a θ e ν temos que as estimativas de momentos para os parâmetros θ e ν , da distribuição ZMPL, são dadas por

$$\tilde{\theta} = \frac{\left[(2\bar{x} - s_{x^2}) + \sqrt{(s_{x^2} - 2\bar{x} + 2\bar{x}s_{x^2})} \right]}{(s_{x^2} - \bar{x})} \quad \text{e} \quad \tilde{\nu} = 1 - \frac{\tilde{\theta}(\tilde{\theta} + 1)\bar{x}}{\tilde{\theta} + 2}.$$

Note que $s_{x^2} = s^2 - \bar{x}^2$, conseqüentemente nós podemos reescever os estimadores dos momentos em função da média e variância. Portanto, os estimadores para θ e ν são dados, respectivamente, por

$$\tilde{\theta} = \frac{\left[2\bar{x} - (s^2 - \bar{x}^2) + \sqrt{(s^2 - \bar{x}^2)(1 + 2\bar{x}) - 2\bar{x}} \right]}{s^2 - \bar{x}^2 - \bar{x}},$$

e

$$\tilde{\nu} = 1 - \frac{\left[2\bar{x} - (s^2 - \bar{x}^2) + \sqrt{(s^2 - \bar{x}^2)(1 + 2\bar{x}) - 2\bar{x}} \right] \left[\bar{x} + \sqrt{(s^2 - \bar{x}^2)(1 + 2\bar{x}) - 2\bar{x}} \right] \bar{x}}{\left[s^2 - \bar{x}^2 + \sqrt{(s^2 - \bar{x}^2)(1 + 2\bar{x}) - 2\bar{x}} \right]}.$$

em que $s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$.

1.3.2 Método de Máxima Verossimilhança

Considere X_1, X_2, \dots, X_n uma amostra aleatória proveniente da distribuição ZMPL. Assim, a função de máxima verossimilhança correspondente para $\boldsymbol{\lambda} = (\theta, \nu)^\top$ é dada

por

$$\begin{aligned}
L(\lambda; x) &= \prod_{i=1}^n \left[\nu + (1 - \nu) \frac{\theta^2(\theta + 2)}{(\theta + 1)^3} \right]^{I_{\{0\}}(x_i)} \left[(1 - \nu) \frac{\theta^2(x_i + \theta + 2)}{(\theta + 1)^{x_i+3}} \right]^{1 - I_{\{0\}}(x_i)} \\
&= \left[\nu + (1 - \nu) \frac{\theta^2(\theta + 2)}{(\theta + 1)^3} \right]^{\sum_{i=1}^n I_{\{0\}}(x_i)} \prod_{i=1}^n \left[(1 - \nu) \frac{\theta^2(x_i + \theta + 2)}{(\theta + 1)^{x_i+3}} \right]^{1 - I_{\{0\}}(x_i)} \\
&= \left[\nu + (1 - \nu) \frac{\theta^2(\theta + 2)}{(\theta + 1)^3} \right]^{n_0} \left[\frac{(1 - \nu)\theta^2}{(\theta + 1)^3} \right]^{n - n_0} \prod_{i=1}^n \left[\frac{(x_i + \theta + 2)}{(\theta + 1)^{x_i}} \right]^{1 - I_{\{0\}}(x_i)}
\end{aligned}$$

em que $n_0 = \sum_{i=1}^n I_{\{0\}}(x_i)$ é a quantidade de número de zeros na amostra e $I_A(\cdot)$ é a função indicador do conjunto A . Portanto, a log-verossimilhança obtida pela equação acima, pode ser expressada como

$$\begin{aligned}
\ell(\lambda; x) &= n_0 \log \left[\nu + (1 - \nu) \frac{\theta^2(\theta + 2)}{(\theta + 1)^3} \right] + (n - n_0) \log \left[\frac{(1 - \nu)\theta^2}{(\theta + 1)^3} \right] \\
&\quad + \sum_{i=1}^n [1 - I_{\{0\}}(x_i)] \log(x_i + \theta + 2) - \log(\theta + 1) \sum_{i=1}^n [1 - I_{\{0\}}(x_i)] x_i.
\end{aligned}$$

Calculando as derivadas com relação aos parâmetros θ e ν , obtemos

$$\begin{aligned}
\frac{\partial \ell(\lambda; x)}{\partial \nu} &= \frac{n_0 \left[1 - \frac{(\theta+2)\theta^2}{(\theta+1)^3} \right]}{\nu + (1 - \nu) \frac{(\theta+2)\theta^2}{(\theta+1)^3}} - \frac{n - n_0}{1 - \nu}, \\
\frac{\partial \ell(\lambda; x)}{\partial \theta} &= \frac{n_0(1 - \nu) \left[\frac{(3\theta^2+4\theta)}{(\theta+1)^3} - \frac{3(\theta+2)\theta^2}{(\theta+1)^4} \right]}{\nu + (1 - \nu) \frac{(\theta+2)\theta^2}{(\theta+1)^3}} - (n - n_0) \frac{\theta - 2}{\theta(\theta + 1)} \\
&\quad + \sum_{i=1}^n \frac{1 - I_{\{0\}}(x_i)}{x_i + \theta + 2} - \frac{1}{\theta + 1} \sum_{i=1}^n [1 - I_{\{0\}}(x_i)] x_i. \tag{1.4}
\end{aligned}$$

Da equação (1.4), temos que o estimador de máxima verossimilhança para ν é $\hat{\nu} = 1 - \left(1 - \frac{n_0}{n}\right) \frac{(\hat{\theta}+1)^3}{\hat{\theta}^2+3\hat{\theta}+1}$. Para θ , vê-se que não é possível encontrar uma expressão de forma fechada para o estimador de máxima verossimilhança. Portanto, devemos utilizar procedimentos de otimização não-linear objetivando maximizar a função de log-verossimilhança. Alguns métodos que podem ser utilizados são: Newton-Raphson, BHHH, Escore de Fisher e BFGS.

A matriz de Informação de Fisher para o modelo ZMPL é dada por

$$i(\lambda) = \begin{bmatrix} i_{\theta\theta} & i_{\theta\nu} \\ i_{\nu\theta} & i_{\nu\nu} \end{bmatrix}, \tag{1.5}$$

com

$$\begin{aligned}
i_{\nu\nu} &= n \left\{ \frac{\left[1 - \frac{\theta^2(\theta+2)}{(\theta+1)^3}\right]^2}{\left[\nu + \frac{(1-\nu)\theta^2(\theta+2)}{(\theta+1)^3}\right]} + \frac{\left[1 - \left(\nu + \frac{(1-\nu)\theta^2(\theta+2)}{(\theta+1)^3}\right)\right]}{(1-\nu)^2} \right\}, \\
i_{\theta\nu} = i_{\nu\theta} &= n \left[\frac{3\theta^2 - 4\theta}{(\theta+1)^3} - \frac{3\theta^2(\theta+2)}{(\theta+1)^4} \right] + \frac{n \left[1 - \frac{\theta^2(\theta+2)}{(\theta+1)^3}\right] \left[\frac{(1-\nu)(3\theta^2+4\theta)}{(\theta+1)^3} - \frac{3(1-\nu)\theta^2(\theta+2)}{(\theta+1)^4} \right]}{\left(\nu + \frac{(1-\nu)\theta^2(\theta+2)}{(\theta+1)^3}\right)}, \\
i_{\theta\theta} &= \frac{n(1-\nu)^2 \left[\frac{3\theta^2+4\theta}{(\theta+1)^3} - \frac{3\theta^2(\theta+2)}{(\theta+1)^4} \right]^2}{\left(\nu + \frac{(1-\nu)\theta^2(\theta+2)}{(\theta+1)^3}\right)} \\
&\quad + n \left\{ \left[\frac{3\theta^3 - 7\theta^2 + 4\theta + 2}{\theta^2(\theta+1)^2} \right] \left\{ 1 - \left[\nu + \frac{(1-\nu)\theta^2(\theta+2)}{(\theta+1)^3} \right] \right\} \right\} \\
&\quad - n(1-\nu) \left[\frac{[\theta^7 + 7\theta^6 + 19\theta^5 + 23\theta^4 + 10\theta^3 - 12\theta^2 - 16\theta + 8]}{(\theta+2)(\theta+1)^5} - \frac{\theta^2}{(\theta+1)} \Phi \left(\frac{1}{\theta+1}; 1; \theta \right) \right],
\end{aligned}$$

em que $\Phi\left(\frac{1}{\theta+1}; 1; \theta\right)$ é uma função zeta de Lerch. Uma representação em termos de integral é dada por $\Phi\left(\frac{1}{\theta+1}; 1; \theta\right) = (\theta+1) \int_0^1 \frac{u^{\theta+1}}{(\theta+1-u)} du$.

Uma maneira de estimar o viés é através do método de reamostragem *bootstrap*, proposto por Efron (1979). O método consiste em obter um grande número de subamostras provenientes da amostra original e desta forma, coletar informações destas subamostras que possam ser de interesse para a realização de inferências. Existem dois tipos de método *bootstrap*: paramétrico ou não-paramétrico. Nesse presente trabalho iremos utilizar o *bootstrap* paramétrico.

Seja $\mathbf{x} = (x_1, \dots, x_n)$ uma amostra aleatória de uma variável aleatória X com função de distribuição $F = F(\cdot)$. Se λ é um parâmetro que indexa a distribuição F , então ele pode ser visto como um funcional de F , ou seja, $\lambda = g(F)$. Considere que $\hat{\lambda} = h(\mathbf{x})$ é um estimador para λ obtido utilizando a amostra acima. Como estamos considerando o *bootstrap* paramétrico, então as pseudo-amostras $\mathbf{x}^* = (x_1^*, \dots, x_n^*)$ são obtidas a partir de $F(\lambda)$.

1.3.3 Correção de viés

Seja $\mathbf{x} = (x_1, \dots, x_n)$ uma amostra aleatória de tamanho n de uma função de distribuição, no nosso caso conhecida. Iremos denotar o viés do estimador $\hat{\lambda} = h(\mathbf{x})$ por

$V_F(\hat{\lambda}, \lambda)$, definido como

$$V_F(\hat{\lambda}, \lambda) = E_F(h(\mathbf{x})) - g(F), \quad (1.6)$$

em que \mathbf{x} é obtido da variável aleatória X , que possui função de distribuição acumulada F . As estimativas de viés *bootstrap* na versão paramétrica são definidas substituindo F , a distribuição que gerou a amostra original, por $F_{\hat{\lambda}}$ na expressão (1.6).

Das B amostras *bootstrap*, $\mathbf{x}^{*1}, \dots, \mathbf{x}^{*B}$ e após obter as estimativas *bootstrap* $\hat{\lambda}^{*1}, \dots, \hat{\lambda}^{*B}$, em que $\hat{\lambda}^{*b} = h(\mathbf{x}^{*b})$ para $b = 1, 2, \dots, B$, é possível aproximar o valor esperado $E_{F_{\hat{\lambda}}}(g(\mathbf{x}^*))$ a partir de

$$\hat{\lambda}^{*(\cdot)} = \frac{1}{B} \sum_{b=1}^B \lambda^{*(b)} = \frac{1}{B} \sum_{b=1}^B h(\mathbf{x}^{*(b)}).$$

Conseqüentemente, as estimativas do viés baseadas nas B réplicas *bootstrap* de $\hat{\lambda}$ são dadas por

$$\hat{V}_{F_{\hat{\lambda}}}(\hat{\lambda}, \lambda) = \hat{\lambda}^{*(\cdot)} - h(\mathbf{x}).$$

A partir destas estimativas pode-se definir o estimador corrigido, que é dado por

$$\bar{\lambda} = h(\mathbf{x}) - \hat{V}_{F_{\hat{\lambda}}}(\hat{\lambda}, \lambda) = 2h(\mathbf{x}) - \hat{\lambda}^{*(\cdot)}.$$

1.4 Estimação Intervalar

1.4.1 Intervalo de Confiança Assintótico (ICa)

Sob algumas condições de regularidades, $\hat{\lambda}$ é um estimador consistente para λ e esta quantidade tem uma distribuição assintótica normal. A quantidade $\sqrt{n}(\hat{\lambda} - \lambda) \rightarrow N_2(0, j(\lambda)^{-1})$ quando $n \rightarrow \infty$, em que $j(\lambda) = \lim_{n \rightarrow \infty} \frac{1}{n} i(\lambda)$, com $i(\lambda)$ sendo a matriz esperada de informação de Fisher dado em (1.5) e \rightarrow denota convergência em distribuição. Note que $i(\lambda)^{-1}$ é um estimador consistente para a matriz de variâncias e covariâncias assintótica de $\hat{\lambda}$. Na prática, pode-se aproximar a matriz de informação de Fisher esperada pela sua versão observada, enquanto que os elementos da diagonal da inversa desta matriz podem ser usados para aproximar os erros padrão correspondentes (EFRON; HINKLEY, 1978).

Com o resultado acima, os intervalos assintóticos com $100(1 - \alpha)\%$ de confiança

para θ e ν , são dados, respectivamente, por

$$\hat{\theta} \pm z_{\alpha/2} \widehat{\text{ep}}(\hat{\theta}) \quad \text{e} \quad \hat{\nu} \pm z_{\alpha/2} \widehat{\text{ep}}(\hat{\nu}),$$

em que $\widehat{\text{ep}}(\hat{\theta})$ e $\widehat{\text{ep}}(\hat{\nu})$ são os erros padrões assintóticos estimados (ep) dos estimadores de máxima verossimilhança de θ e ν , respectivamente, e $z_{\alpha/2}$ é o quantil ($\alpha/2$) da distribuição normal padrão.

1.4.2 Intervalo de Confiança Percentil (ICp)

Outra maneira de construir intervalo de confiança para λ é utilizar os intervalos de confiança *bootstrap* percentil. Este método funciona da seguinte maneira: a partir da distribuição empírica \hat{F} obtemos as B réplicas *bootstrap*. Os limites inferior e superior do intervalo de confiança, para λ , são definidos pelos percentis $\alpha/2$ e $1 - \alpha/2$, da forma

$$\left(\hat{F}^{-1}(\alpha/2), \hat{F}^{-1}(1 - \alpha/2) \right),$$

em que $\hat{F}^{-1}(\alpha/2) = \hat{\lambda}^{*(\alpha/2)}$ e $\hat{F}^{-1}(1 - \alpha/2) = \hat{\lambda}^{*(1-\alpha/2)}$. Portanto, podemos obter o intervalo de confiança para λ da seguinte maneira

$$\left(\hat{\lambda}^{*(\alpha/2)}, \hat{\lambda}^{*(1-\alpha/2)} \right),$$

em que ordenamos as B réplicas *bootstrap* de $\hat{\lambda}$ e posteriormente consideramos, como limites inferior e superior, os valores das estimativas que estão na posição $\alpha/2$ e $1 - \alpha/2$.

1.5 Teste de Hipóteses

Considere a hipótese $\mathcal{H}_0 : \nu = 0$ e $\mathcal{H}_1 : \nu \neq 0$. O interesse é teste a hipótese nula (distribuição PL) contra a hipótese alternativa (distribuição ZMPL).

1.5.1 Teste gradiente

Seja $X = (X_1, X_2, \dots, X_n)^\top$ uma amostra aleatória de tamanho n da distribuição Poisson-Lindley zero modificada, em que cada $X_i, i = 1, 2, \dots, n$, tem uma f.m.p dado

pela Equação 1.1. A estatística gradiente, S_g , é dada por:

$$S_g = n \hat{\nu}^2 (\hat{\theta}^2 + 3\hat{\theta} + 1),$$

em que $\hat{\boldsymbol{\lambda}} = (\hat{\theta}, \hat{\nu})^\top$ é o estimador de máxima verossimilhança irrestrito de $\boldsymbol{\lambda} = (\theta, \nu)^\top$. Assintoticamente, S_g tem uma distribuição χ_1^2 sob \mathcal{H}_0 .

Capítulo 2

Simulação

Neste capítulo iremos apresentar os resultados das simulações de Monte Carlo realizadas para estudar o comportamento em amostras finitas dos estimadores de máxima verossimilhança dos parâmetros que indexam a distribuição ZMPL e de suas versões com viés corrigido. As simulações foram realizadas utilizando a linguagem de programação **R** (R Core Team, 2017).

Foram usadas algumas medidas para a avaliação da qualidade das estimativas pontuais. As medidas utilizadas foram: viés relativo e erro quadrático médio (EQM). Para a avaliação das estimativas intervalares calculamos a probabilidade de cobertura empírica dos intervalos, a probabilidade de cobertura empírica à direita e a probabilidade de cobertura empírica à esquerda.

No estudo de simulação consideramos 5 000 réplicas de Monte Carlo e 1 000 réplicas de bootstrap. Os tamanhos amostrais considerados foram $n = 35, 60, 90, 120$. Enquanto que os valores considerados para o parâmetro θ foram 1, 5 e 2, 0. Por fim, para o parâmetro ν foram considerados os valores $-0.10, 0.00$ e 0.10 .

Na Tabela 2.1 apresentamos os vieses relativos e EQM's dos estimadores de máxima verossimilhança e de suas versões com viés corrigido. Podemos observar que para todos os tamanhos amostrais e para os diferentes valores de θ e ν , as versões corrigidas dos estimadores de máxima verossimilhança trabalham de forma eficiente no que diz respeito à redução de viés e de EQM. Nota-se também que os EQM's de $\bar{\nu}$ não apresentam grande variação em relação aos resultados apresentados pela versão não corrigida.

Tabela 2.1: Medidas de precisão (com EQM em parênteses) dos estimadores dos parâmetros que indexam a distribuição ZMPL.

n	ν	$\theta = 1,5$				$\theta = 2,0$			
		$\hat{\theta}$	$\bar{\theta}$	$\hat{\nu}$	$\bar{\nu}$	$\hat{\theta}$	$\bar{\theta}$	$\hat{\nu}$	$\bar{\nu}$
35	-0,10	0,163 (0,426)	-0,029 (0,216)	-0,080 (0,119)	0,075 (0,113)	0,322 (1,388)	-0,103 (1,008)	-0,139 (0,292)	0,073 (0,205)
	0,00	0,198 (0,487)	-0,024 (0,219)	-0,082 (0,114)	0,076 (0,097)	0,367 (1,527)	-0,045 (1,228)	-0,146 (0,275)	0,058 (0,199)
	0,10	0,230 (0,572)	-0,013 (0,271)	-0,092 (0,121)	0,072 (0,104)	0,371 (1,511)	-0,001 (1,244)	-0,131 (0,236)	0,056 (0,155)
60	-0,10	0,090 (0,162)	0,005 (0,100)	-0,042 (0,052)	0,018 (0,045)	0,193 (0,690)	-0,042 (0,553)	-0,080 (0,144)	0,027 (0,120)
	0,00	0,096 (0,179)	-0,002 (0,107)	-0,043 (0,050)	0,023 (0,046)	0,212 (0,617)	-0,042 (0,385)	-0,085 (0,107)	0,022 (0,070)
	0,10	0,116 (0,228)	0,001 (0,134)	-0,046 (0,051)	0,028 (0,047)	0,217 (0,717)	-0,037 (0,541)	-0,069 (0,099)	0,032 (0,071)
90	-0,10	0,056 (0,091)	0,010 (0,066)	-0,029 (0,031)	0,003 (0,028)	0,120 (0,289)	-0,018 (0,201)	-0,048 (0,064)	0,017 (0,051)
	0,00	0,062 (0,105)	0,009 (0,076)	-0,028 (0,031)	0,010 (0,029)	0,127 (0,317)	-0,025 (0,218)	-0,050 (0,059)	0,015 (0,046)
	0,10	0,077 (0,126)	0,014 (0,088)	-0,032 (0,029)	0,011 (0,027)	0,164 (0,428)	-0,005 (0,311)	-0,057 (0,061)	0,009 (0,046)
120	-0,10	0,045 (0,066)	0,018 (0,050)	-0,023 (0,023)	0,000 (0,022)	0,090 (0,195)	-0,004 (0,152)	-0,039 (0,044)	0,006 (0,037)
	0,00	0,048 (0,074)	0,015 (0,056)	-0,024 (0,023)	0,003 (0,022)	0,111 (0,240)	0,004 (0,179)	-0,044 (0,043)	0,002 (0,036)
	0,10	0,057 (0,087)	0,019 (0,064)	-0,021 (0,022)	0,013 (0,023)	0,106 (0,250)	-0,013 (0,187)	-0,038 (0,037)	0,008 (0,031)

Fonte: Próprio autor.

Por fim, podemos notar que para todos os cenários os valores do EQM são reduzidos com o aumento dos valores amostrais.

Na Tabela 2.2 apresentamos as probabilidades de cobertura empíricas dos intervalos de confiança assintótico e bootstrap percentil. Podemos observar que, de modo geral, os ICp para θ obtiveram uma proximidade maior aos níveis nominais quando comparados aos ICa. Nota-se que para os intervalos de confiança para ν , o intervalo que apresentou melhor desempenho em relação a probabilidade de cobertura foi o ICp. Por fim, podemos observar que os intervalos tendem a apresentar comportamento simétrico quando consideramos amostras grandes.

Tabela 2.2: Probabilidades de cobertura empíricas e probabilidades de cobertura empíricas das caudas (Esquerda, Direita) para $\theta = 1, 5$.

		θ						ν					
n	ν	ICa(-; 0,90)	ICp(-; 0,90)	ICa(-; 0,95)	ICa(-; 0,95)	ICa(-; 0,99)	ICp(-; 0,99)	ICa(-; 0,90)	ICp(-; 0,90)	ICa(-; 0,95)	ICp(-; 0,95)	ICa(-; 0,99)	ICp(-; 0,99)
35	-0,10	0,925	0,891	0,949	0,939	0,980	0,976	0,924	0,916	0,955	0,959	0,985	0,988
		(0,001;0,074)	(0,089;0,020)	(0,000;0,050)	(0,054;0,007)	(0,000;0,020)	(0,024;0,000)	(0,065;0,010)	(0,005;0,080)	(0,042;0,003)	(0,000;0,041)	(0,015;0,000)	(0,000;0,012)
	0,00	0,929	0,881	0,955	0,932	0,983	0,976	0,919	0,910	0,955	0,956	0,982	0,988
	(0,000;0,071)	(0,102;0,016)	(0,000;0,045)	(0,064;0,004)	(0,000;0,017)	(0,024;0,000)	(0,072;0,010)	(0,007;0,082)	(0,044;0,002)	(0,000;0,044)	(0,018;0,000)	(0,000;0,012)	
	0,10	0,929	0,886	0,953	0,933	0,982	0,974	0,923	0,910	0,955	0,954	0,985	0,987
		(0,000;0,071)	(0,101;0,013)	(0,000;0,047)	(0,064;0,004)	(0,000;0,018)	(0,026;0,000)	(0,067;0,010)	(0,009;0,081)	(0,043;0,002)	(0,001;0,045)	(0,015;0,000)	(0,000;0,013)
60	-0,10	0,924	0,892	0,957	0,942	0,987	0,983	0,919	0,910	0,960	0,959	0,988	0,991
		(0,01;0,066)	(0,084;0,024)	(0,000;0,043)	(0,051;0,006)	(0,000;0,013)	(0,017;0,000)	(0,061;0,019)	(0,023;0,067)	(0,037;0,003)	(0,005;0,036)	(0,012;0,000)	(0,000;0,009)
	0,00	0,925	0,900	0,958	0,945	0,986	0,982	0,917	0,904	0,956	0,956	0,988	0,990
	(0,007;0,068)	(0,079;0,021)	(0,000;0,042)	(0,047;0,009)	(0,000;0,014)	(0,017;0,001)	(0,063;0,020)	(0,022;0,073)	(0,040;0,004)	(0,004;0,040)	(0,012;0,000)	(0,000;0,010)	
	0,10	0,922	0,896	0,953	0,942	0,984	0,982	0,912	0,906	0,955	0,960	0,983	0,989
		(0,005;0,073)	(0,081;0,023)	(0,000;0,047)	(0,051;0,006)	(0,000;0,016)	(0,018;0,000)	(0,069;0,019)	(0,022;0,072)	(0,041;0,004)	(0,002;0,038)	(0,016;0,000)	(0,000;0,011)
90	-0,10	0,915	0,896	0,956	0,948	0,986	0,985	0,917	0,901	0,958	0,952	0,987	0,992
		(0,017;0,068)	(0,075;0,029)	(0,003;0,041)	(0,040;0,012)	(0,000;0,014)	(0,014;0,001)	(0,058;0,025)	(0,033;0,067)	(0,035;0,007)	(0,014;0,035)	(0,012;0,000)	(0,000;0,008)
	0,00	0,913	0,893	0,958	0,943	0,986	0,982	0,908	0,898	0,957	0,955	0,988	0,991
	(0,020;0,067)	(0,077;0,030)	(0,002;0,040)	(0,046;0,011)	(0,000;0,014)	(0,017;0,001)	(0,064;0,028)	(0,029;0,073)	(0,036;0,007)	(0,007;0,038)	(0,012;0,000)	(0,000;0,009)	
	0,10	0,916	0,896	0,957	0,939	0,987	0,983	0,915	0,910	0,959	0,961	0,987	0,990
		(0,018;0,066)	(0,078;0,026)	(0,003;0,040)	(0,050;0,011)	(0,000;0,013)	(0,016;0,001)	(0,058;0,027)	(0,021;0,068)	(0,034;0,008)	(0,002;0,036)	(0,012;0,000)	(0,000;0,010)
120	-0,10	0,908	0,895	0,953	0,943	0,987	0,986	0,910	0,903	0,953	0,958	0,989	0,990
		(0,027;0,064)	(0,073;0,031)	(0,007;0,040)	(0,043;0,014)	(0,000;0,013)	(0,013;0,002)	(0,057;0,033)	(0,031;0,066)	(0,036;0,012)	(0,007;0,035)	(0,011;0,000)	(0,000;0,010)
	0,00	0,909	0,893	0,954	0,944	0,987	0,986	0,903	0,913	0,952	0,962	0,989	0,991
	(0,024;0,067)	(0,075;0,032)	(0,007;0,039)	(0,042;0,014)	(0,000;0,013)	(0,013;0,001)	(0,063;0,034)	(0,020;0,067)	(0,036;0,012)	(0,002;0,037)	(0,010;0,001)	(0,000;0,009)	
	0,10	0,907	0,891	0,954	0,943	0,986	0,984	0,905	0,918	0,948	0,962	0,984	0,990
		(0,023;0,070)	(0,076;0,032)	(0,005;0,041)	(0,044;0,013)	(0,000;0,014)	(0,014;0,002)	(0,063;0,032)	(0,018;0,064)	(0,041;0,011)	(0,003;0,036)	(0,016;0,001)	(0,000;0,010)

Próprio autor.

Capítulo 3

Aplicações

3.1 Contagem de lesões de dosimetria citogenética

Os dados, estudados nesta seção, foram obtidos através de Shanker e Fesshaye (2015) e eles são referentes à contagem de lesões de dosimetria citogenética em coelhos. Na Tabela 3.1 apresentamos algumas medidas descritivas. Nela vemos que o coeficiente de assimetria foi 2,42, ou seja, este valor implica que a distribuição dos dados é positivamente assimétrica. O valor do coeficiente de curtose foi de 10,70, isto é, a distribuição dos dados é leptocúrtica (a distribuição têm caudas pesadas). O índice de dispersão de Fisher, igual a 1,56, mostra que os dados devem ser modelados por um modelo superdisperso. Temos também que a proporção de zeros, no conjunto de dados, é de 69% e assim, temos evidências de que o conjunto de dados é inflacionado de zeros. Portanto, parece razoável o uso da distribuição ZMPL para modelar os dados.

Na Tabela 3.2 temos as estimativas de máxima verossimilhança dos parâmetros e os valores da estatística de teste qui-quadrado. A estatística de teste gradiente para testar $\mathcal{H}_0 : \nu = 0$ é $S_g = 114.49$ e o nível descritivo foi < 0.001 , ou seja, o parâmetro ν é estatisticamente diferente de zero.

As diferenças padronizadas (DP) podem ser calculadas da seguinte maneira

$$\Delta_{sd_i} = \frac{\delta_{sd_i}}{\max \{ \delta_{(\cdot)d_1}, \dots, \delta_{(\cdot)d_m} \}}, \quad i = 1, \dots, m; s \in \mathcal{S},$$

em que $\delta_{sd_i} = (\text{observado}_{sd_i} - \text{esperado}_{sd_i})$, $\delta_{(\cdot)d_i} = (\delta_{s_1d_i}, \dots, \delta_{s_{|S|}d_i})$; m é o número de

Tabela 3.1: Medidas descritivas

Medidas	Mínimo	Máximo	Média	Variância	Coef. de Assimetria	Coef. de Curtose	IF
Valores	0,00	6,00	0,47	0,74	2,42	10,70	1,56

Fonte: Próprio autor.

Tabela 3.2: Distribuição do número de lesões de dosimetria citogenética em coelhos, com frequência esperada obtida pelo ajuste das distribuições Poisson, Poisson zero modificada, Poisson-Lindley e ZMPL.

Nº de lesões de dosimetria citogenética de mamíferos	Frequência Observada	Frequência Esperada			
		Poisson	Poisson Zero-modificada	Poisson-Lindley	Poisson Lindley Modificada no zero (ZMPL)
0	413	374,0	413,0	405,7	413,0
1	124	177,4	116,0	133,6	123,4
2	42	42,1	52,1	42,6	42,9
3	15	6,6	15,6	13,3	14,5
4	5	0,8	3,5	4,1	4,8
5	0	0,1	0,6	1,2	1,6
6	2	0,0	0,2	0,5	0,5
Total	601	60	601	601	601
Estimativa do parâmetro		$\hat{\theta} = 0,47421$	$\hat{\theta} = 0,8989$ $\hat{\nu} = 0,4725$	$\hat{\theta} = 2,6854$	$\hat{\theta} = 2,4098$ $\hat{\nu} = 0,1165$
Intervalo de confiança		aCI($\theta; 0,95$) = (0,4192; 0,5293)	aCI($\theta; 0,95$) = (0,7304; 1,0675) aCI($\nu; 0,95$) = (0,3852; 0,5599)	aCI($\theta; 0,95$) = (2,3619; 3,0088)	aCI($\theta; 0,95$) = (1,8904; 2,9290) aCI($\nu; 0,95$) = (-0,0649; 0,2979)
χ^2		726,2816	42,19	9,6880	5,9064
d.f		6	6	6	6
p-value		< 0,0001	< 0,0001	0,1384	0,4338

Fonte: Próprio autor.

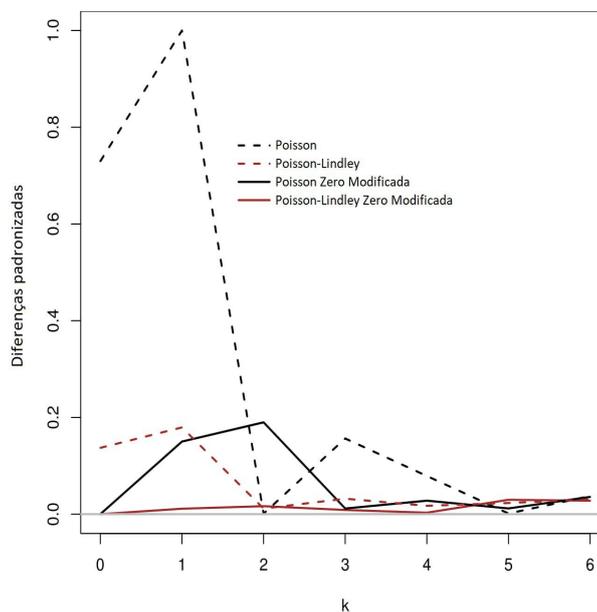
modelos estudados; \mathcal{S} é o suporte (comum) das distribuições; e $|A|$ é a cardinalidade de um conjunto A . O gráfico DP é construído plotando as diferenças padronizadas versus os valores observados. Este gráfico é útil para saber qual modelo se ajusta melhor aos dados.

O gráfico das diferenças padronizadas para os modelos ZMPL, Poisson-Lindley, Poisson zero modificada e Poisson é apresentado na Figura 3.1. Baseados nos valores da Tabela 3.2 e do comportamento observado na Figura 3.1, podemos concluir que a distribuição ZMPL se ajusta melhor aos dados quando comparado com os outros modelos ajustados.

3.2 Contagem da ocorrência de greves

Este conjunto de dados se refere aos número de ocorrência de greves na indústria de mineração de carvão no Reino Unido em períodos sucessivos de quatro semanas entre os anos de 1948 a 1959. Estes dados podem ser obtidos em Ridout e Besbeas (2004). Analisando os valores apresentados na Tabela 3.3 podemos notar que os dados apresenta uma leve subdispersão, pois o valor do índice de dispersão de Fisher é 0,75. Além disso, temos que o coeficiente de curtose foi de 3,5, ou seja, a distribuição é leptocúrtica e o

Figura 3.1: Gráfico das diferenças padronizadas para os diferentes modelos ajustados aos dados.



Fonte: Próprio autor.

coeficiente de assimetria foi de 0.8, isto é, os dados possuem assimetria positiva.

Tabela 3.3: Medidas descritivas

Medidas	Mín.	Máx.	Média	Variância	Coef.Assimetria	Coef.Curtose	FI
Values	0,00	4,00	0,99	0,74	0,80	3,5	0,75

Fonte: Próprio autor.

Na Tabela 3.4 são apresentadas as estimativas dos parâmetros dos modelos estudados e as estimativas da estatística de teste qui-quadrado. A estatística de teste gradiente para testar $\mathcal{H}_0 : \nu = 0$ é $S_g = 5275,1$ e nível descritivo foi $< 0,001$, ou seja, o parâmetro ν é estatisticamente diferente de zero.

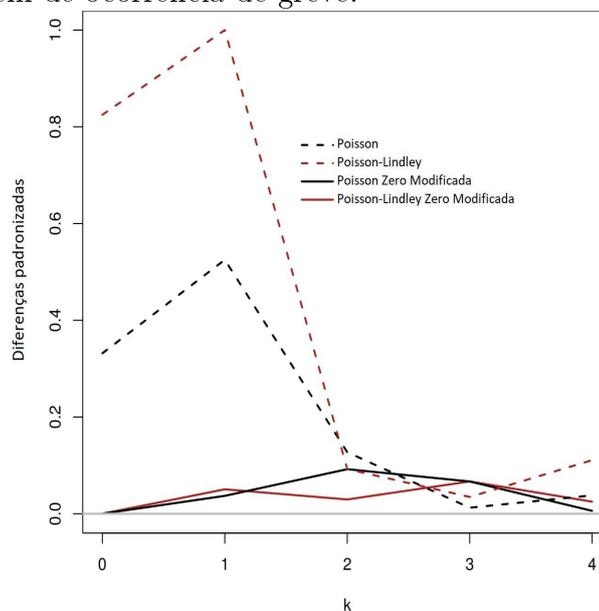
A Figura 3.2 mostra as diferenças padronizadas para os diferentes modelos ajustados aos dados de contagem de ocorrência de greve (RIDOUT; BESBEAS, 2004). Portanto, após a análise da Tabela 3.4 e da Figura 3.2, nota-se que as distribuições Poisson zero modificada e ZMPL apresentam um ajuste similar.

Tabela 3.4: Frequências esperadas, estimativas dos parâmetros e estatística qui-quadrado referentes ao conjunto de dados de contagem de ocorrência de greves.

Nº de ocorrência	Frequência Observada	Frequência esperada			
		Poisson	Zero-modified Poisson	Poisson-Lindley	Zero-modified Poisson-Lindley
0	46	57,76	46,00	75,24	46,00
1	76	57,39	74,69	40,55	77,79
2	24	28,51	27,27	20,73	22,95
3	9	9,44	6,64	10,23	6,63
≥4	1	2,35	1,21	4,93	1,89
Total	156	156	156	156	156
Estimativa dos parâmetros		$\hat{\theta} = 0,9936$	$\hat{\theta} = 0,7301$ $\hat{\nu} = -0,3609$	$\hat{\theta} = 1,4010$	$\hat{\theta} = 2,9579$ $\hat{\nu} = -1,3475$
Intervalo de confiança		aCI($\theta, 0,95$) = (0,8372; 1,1500)	aCI($\theta, 0,95$) = (0,5271; 0,9331) aCI($\nu, 0,95$) = (-0,6526; -0,0691)	aCI($\theta, 0,95$) = (1,1478; 1,6542)	aCI($\theta, 0,95$) = (2,0436; 3,8721) aCI($\nu, 0,95$) = (-1,9923; -0,7028)
χ^2		9,8986	1,2916	44,748	1,3404
d.f		4	4	4	4
p-value		0,04217	0,8628	< 0,0001	0,8545

Fonte: Próprio autor.

Figura 3.2: Gráfico das diferenças padronizadas para os diferentes modelos ajustados aos dados de contagem de ocorrência de greve.



Fonte: Próprio autor.

Capítulo 4

Conclusões

Neste trabalho apresentamos uma generalização da distribuição Poisson-Lindley inflacionada de zeros chamada de Poisson-Lindley zero modificada. Com esta generalização, além de modelar dados com inflação zeros, conseguimos modelar dados com deflação de zeros. Caracterizamos este modelo apresentando algumas de suas propriedades, como por exemplo, média e variância. Além disso, apresentamos funções geradora de momentos e de probabilidade. Também estudamos, via simulações de Monte Carlo, os estimadores de máxima verossimilhança e suas versões corrigidas de viés via *bootstrap* e conseguimos notar que as versões corrigidas apresentaram um melhor desempenho em relação ao viés e EQM. Estudamos o comportamento das estimativas intervalares considerando os intervalos de confiança assintótico e percentil. Tais intervalos apresentaram bom desempenho quando comparado aos níveis nominais. Por fim, utilizamos duas aplicações em dados reais que mostram a aplicabilidade e flexibilidade do modelo proposto.

Apêndice A

Demonstrações

A.1 Relação de recorrência da função geradora de probabilidade

Prova. Pela equação (1.3) temos que $G'(t)$ é dado por:

$$\begin{aligned} G'(t) &= \frac{\partial}{\partial t} \left[\nu + (1 - \nu) \frac{\theta^2(\theta + 2 - t)}{(\theta + 1)(\theta + 1 - t)^2} \right]; \\ G'(t = 0) &= \frac{\theta^2(1 - \nu)}{(\theta + 1)} \frac{(\theta + 3)}{(\theta + 1)^3}; \end{aligned}$$

Além disso, tem-se que:

$$\begin{aligned} G''(t = 0) &= \frac{\theta^2(1 - \nu)}{(\theta + 1)} \frac{2(\theta + 4)}{(\theta + 1)^4}; \\ G'''(t = 0) &= \frac{\theta^2(1 - \nu)}{(\theta + 1)} \frac{6(\theta + 5)}{(\theta + 1)^5}; \\ &\vdots \\ G^{(r-1)}(t = 0) &= \frac{\theta^2(1 - \nu)(r - 1)!(\theta + r + 1)}{(\theta + 1)^{r+2}}; \\ G^{(r)}(t = 0) &= \frac{\theta^2(1 - \nu)r!(\theta + r + 2)}{(\theta + 1)^{r+3}}. \end{aligned}$$

Temos que:

$$P_r = \frac{G^{(r)}(t=0)}{r!} \quad \text{e} \quad P_{r-1} = \frac{G^{(r-1)}(t=0)}{(r-1)!}.$$

Então, segue que $P_r/P_{(r-1)}$ é dado por:

$$\begin{aligned} \frac{P_r}{P_{r-1}} &= \frac{\frac{\theta^2(1-\nu)r!(\theta+r+2)}{(\theta+1)^{r+3}}}{\frac{\theta^2(1-\nu)(r-1)!(\theta+r+1)}{(\theta+1)^{r+2}}} = \frac{\theta^2(1-\nu)r!(\theta+r+2)}{r!(\theta+1)^{r+3}} \\ &= \frac{\theta^2(1-\nu)(\theta+r+2)(\theta+1)^{r+2}}{(\theta+1)^{r+3}\theta^2(1-\nu)(\theta+r+1)} \\ \frac{P_r}{P_{r-1}} &= \frac{(\theta+r+2)}{(\theta+1)(\theta+r+1)} \\ P_r &= \frac{(\theta+r+2)}{(\theta+1)(\theta+r+1)}P_{r-1}, \quad r > 1. \end{aligned}$$

■

A.2 Momentos de primeira e segunda ordem

Prova. Como

$$\mu'_1 = E(X) = \sum_{x=1}^{\infty} xP_x = \sum_{x=1}^{\infty} x(1-\nu)\frac{\theta^2(\theta+2+x)}{(\theta+1)^{x+3}},$$

fazendo $j = x - 1$ e substituindo na equação acima, temos:

$$\begin{aligned} \sum_{j=0}^{\infty} (j+1)(1-\nu)\frac{\theta^2(\theta+2+j+1)}{(j+1+1)^{j+1+3}} &= \sum_{j=0}^{\infty} (j+1)(1-\nu)\frac{\theta^2(\theta+j+3)}{(j+2)^{j+4}} \\ &= \sum_{j=0}^{\infty} j(1-\nu)\frac{\theta^2(\theta+j+3)}{(j+2)^{j+4}} + \\ &\quad \sum_{j=0}^{\infty} (1-\nu)\frac{\theta^2(\theta+j+3)}{(j+2)^{j+4}}. \end{aligned}$$

Calculando, computacionalmente, a convergência das séries, ambas são convergentes pelo teste da razão, se $|1/(\theta + 1)| < 1$, como o parâmetro θ é sempre maior do que 0, a condição é satisfeita. Então:

$$\begin{aligned}
 \mu'_1 &= (1 - \nu) \left[\frac{\theta^2(\theta^2 + 4\theta + 2)}{\theta^3(\theta + 1)^3} + \frac{\theta^2(\theta^2 + 3\theta + 1)}{\theta^2(\theta + 1)^3} \right] \\
 &= \frac{(1 - \nu)}{(\theta + 1)^3} \left[\frac{(\theta^2 + 4\theta + 2)}{\theta} + (\theta^2 + 3\theta + 1) \right] \\
 &= \frac{(1 - \nu)}{(2\theta + 1)^3} (4\theta^2 + \theta^3 + 5\theta + 2) \\
 &= \frac{(1 - \nu)(\theta + 2)(\theta + 1)^2}{\theta(\theta + 1)^3} \\
 &= \frac{(1 - \nu)(\theta + 2)}{\theta(\theta + 1)}.
 \end{aligned}$$

Já μ'_2 pode ser calculado pela seguinte expressão:

$$\mu'_2 = E(X^2) = \sum_{x=1}^{\infty} x^2 P = \sum_{x=1}^{\infty} x^2 (1 - \nu) \frac{\theta^2(\theta + 2 + x)}{(\theta + 1)^{x+3}} = (1 - \nu)\theta^2 \sum_{x=1}^{\infty} x^2 \frac{(\theta + 2 + x)}{(\theta + 1)^{x+3}},$$

de modo que, fazendo $j = x - 1$ e substituindo na equação acima, temos:

$$\begin{aligned}
 (1 - \nu)\theta^2 \sum_{j=0}^{\infty} (j + 1)^2 \frac{(\theta + 2 + j + 1)}{(\theta + 1)^{j+1+3}} &= (1 - \nu)\theta^2 \left[\sum_{j=0}^{\infty} \frac{j^2(\theta + j + 3)}{(\theta + 1)^{j+4}} \right. \\
 &\quad \left. + \sum_{j=0}^{\infty} \frac{2j(\theta + j + 3)}{(\theta + 1)^{j+4}} + \sum_{j=0}^{\infty} \frac{(\theta + j + 3)}{(\theta + 1)^{j+4}} \right].
 \end{aligned}$$

Calculando, computacionalmente, a convergência das séries, ambas são convergentes pelo teste da razão, se $|1/(\theta + 1)| < 1$, como o parâmetro θ é sempre maior do que 0, a condição é satisfeita. Então:

$$\begin{aligned}
\mu_2' &= (1 - \nu)\theta^2 \left[\frac{\theta^3 + 6\theta^2 + 2\theta + 6}{\theta^4(\theta + 1)^3} + \frac{2\theta^2 + 8\theta + 4}{\theta^3(\theta + 1)^3} + \frac{\theta^2 + 3\theta + 1}{\theta^2(\theta + 1)^3} \right] \\
&= \frac{(1 - \nu)}{(\theta + 1)^3} \left[\frac{\theta^3 + 6\theta^2 + 2\theta + 6}{\theta^2} + \frac{2\theta^2 + 8\theta + 4}{\theta} + \theta^2 + 3\theta + 1 \right] \\
&= \frac{(1 - \nu)}{(\theta + 1)^3\theta^2} [\theta^4 + 6\theta^3 + 15\theta^2 + 6\theta + 6] \\
&= \frac{(1 - \nu)}{(\theta + 1)^3\theta^2} [(\theta + 1)^2(\theta^2 + 4\theta + 6)] \\
&= \frac{(1 - \nu)(\theta^2 + 4\theta + 6)}{\theta^2(\theta + 1)}.
\end{aligned}$$

■

A.3 Coeficiente de assimetria

Prova.

$$\mu_3 = E(X - \mu)^3 = E(-\mu^3 + X^3 - 3\mu X^2 + 3\mu^2 X) = E(X^3) - 3\mu E(X^2) + 2\mu^3.$$

Primeiro, calculamos o valor de $E(X^3)$, e posteriormente os valores de $-\mu E(X^2)$ e de $-2\mu^3$, respectivamente.

$$E(X^3) = \sum_{x=0}^{\infty} X_i^3 P_i \sum_{x=0}^{\infty} x_i^3 (1 - \nu) \frac{\theta^2(\theta + 2 + x)}{(\theta + 1)^{x+3}} = (1 - \nu)\theta^2 \sum_{x=1}^{\infty} x_i^3 \frac{(\theta + 2 + x)}{(\theta + 1)^{x+3}}.$$

Fazendo $j = x - 1$ e substituindo na equação acima, temos:

$$\begin{aligned}
E(X^3) &= (1 - \nu)\theta^2 \left[\sum_{j=0}^{\infty} j^3 \frac{(\theta + j + 3)}{(\theta + 1)^{j+4}} + \sum_{j=0}^{\infty} 3j^2 \frac{(\theta + j + 3)}{(\theta + 1)^{j+4}} + \right. \\
&\quad \left. \sum_{j=0}^{\infty} 3j \frac{(\theta + j + 3)}{(\theta + 1)^{j+4}} + \sum_{j=0}^{\infty} \frac{(\theta + j + 3)}{(\theta + 1)^{j+4}} \right].
\end{aligned}$$

Calculando, computacionalmente, a convergência das séries, ambas são convergentes pelo teste da razão, se $|1/(\theta + 1)| < 1$, como o parâmetro θ é sempre maior do

que 0, a condição é satisfeita. Então:

$$\begin{aligned}
E(X^3) &= (1-\nu)\theta^2 \left[\frac{\theta^4 + 10\theta^3 + 38\theta^2 + 54\theta + 24}{\theta^5(\theta+1)^3} + \right. \\
&\quad \left. \frac{3\theta^3 + 18\theta^2 + 36\theta + 18}{\theta^4(\theta+1)^3} + \frac{3\theta^2 + 12\theta + 6}{\theta^3(\theta+1)^3} + \frac{\theta^2 + 3\theta + 1}{\theta^2(\theta+1)^3} \right] \\
&= \frac{(1-\nu)}{(\theta+1)^3} \left[\frac{\theta^4 + 10\theta^3 + 38\theta^2 + 54\theta + 24}{\theta^3} + \right. \\
&\quad \left. \frac{3\theta^3 + 18\theta^2 + 36\theta + 18}{\theta^2} + \frac{3\theta^2 + 12\theta + 6}{\theta} + \theta^2 + 3\theta + 1 \right] \\
&= \frac{(1-\nu)}{(\theta+1)^3} \left[\frac{\theta^4 + 10\theta^3 + 38\theta^2 + 54\theta + 24}{\theta^3} + \right. \\
&\quad \left. \frac{\theta(3\theta^3 + 18\theta^2 + 36\theta + 18)}{\theta^3} + \frac{\theta^2(3\theta^2 + 12\theta + 6) + \theta^3(\theta^2 + 3\theta + 1)}{\theta^3} \right] \\
&= \frac{(1-\nu)}{\theta^3(\theta+1)^3} (\theta^5 + 10\theta^4 + 41\theta^3 + 80\theta^2 + 72\theta + 24).
\end{aligned}$$

A seguir, calculamos o valor de $3\mu E(X^2)$

$$\begin{aligned}
-3\mu E(X^2) &= \frac{-3(1-\nu)(\theta+2)}{\theta(\theta+1)} \frac{(1-\nu)(\theta^2 + 4\theta + 6)}{\theta^2(\theta+1)} \\
&= \frac{-3(1-\nu)^2(\theta+2)(\theta^2 + 4\theta + 6)}{\theta^3(\theta+1)^2} \\
&= \frac{(1-\nu)^2(-3\theta^3 - 18\theta^2 - 42\theta - 36)}{\theta^3(\theta+1)^2}.
\end{aligned}$$

Em consequência, calculamos o valor de $E(2\mu^3)$

$$E(2\mu^3) = 2 \left(\frac{(1-\nu)(\theta+2)}{\theta(\theta+1)} \right)^3 = \frac{2(1-\nu)^3(\theta^3 + 6\theta^2 + 12\theta + 8)}{\theta^3(\theta+1)^3}.$$

Anteriormente calculamos separadamente os termos que compõe o valor de $\mu_3 = E(X - \mu)^3$. A seguir, uniremos todas as partes:

$$\begin{aligned}
\mu_3 &= \frac{1-\nu}{\theta^3(\theta+1)^2} \left[\frac{\theta^5 + 10\theta^4 + 41\theta^3 + 80\theta^2 + 72\theta + 24}{\theta+1} + \right. \\
&\quad \left. (1-\nu)(-3\theta^3 - 18\theta^2 - 42\theta - 36) + \frac{(1-\nu)^2(2\theta^3 + 12\theta^2 + 24\theta + 16)}{\theta+1} \right].
\end{aligned}$$

Para simplificar as expressões, chamemos:

$$p(\theta) = \left[\frac{\theta^5 + 10\theta^4 + 41\theta^3 + 80\theta^2 + 72\theta + 24}{\theta + 1} + (1 - \nu)(-3\theta^3 - 18\theta^2 - 42\theta - 36) + \frac{(1 - \nu)^2(2\theta^3 + 12\theta^2 + 24\theta + 16)}{\theta + 1} \right].$$

Segue que o coeficiente de assimetria é dado por:

$$\begin{aligned} \gamma_1 &= \frac{\frac{(1 - \nu)}{\theta^3(\theta + 1)^2} p(\theta)}{\left[\frac{(1 - \nu)(\theta^3 + 4\theta^2 + 6\theta + 2 + \nu(\theta + 2)^2)}{\theta^2(\theta + 1)^2} \right]^{3/2}} \\ &= \frac{(1 - \nu)p(\theta)}{\theta^3(\theta + 1)^2} \frac{\theta^3(\theta + 1)^3}{(1 - \nu)^{3/2} \{ \theta^3 + 4\theta^2 + 6\theta + 2 + \nu(\theta + 2)^2 \}^{3/2}} \\ &= \frac{p(\theta)(\theta + 1)}{(1 - \nu)^{1/2} \{ \theta^3 + 4\theta^2 + 6\theta + 2 + \nu(\theta + 2)^2 \}^{3/2}} = \frac{p(\theta)(\theta + 1)}{F}. \end{aligned}$$

Expandindo novamente a parte $[p(\theta)](\theta + 1)$ e assim obtemos o resultado final.

Então calculando a parte do numerador da fração anterior, temos que:

$$\begin{aligned} p(\theta)(\theta + 1) &= (\theta + 1) \left[\frac{\theta^5 + 10\theta^4 + 41\theta^3 + 80\theta^2 + 72\theta + 24}{\theta + 1} + \right. \\ &\quad \left. (1 - \nu)(-3\theta^3 - 18\theta^2 - 42\theta - 36) + \frac{(1 - \nu)^2(2\theta^3 + 12\theta^2 + 24\theta + 16)}{\theta + 1} \right] \\ &= [\theta^5 + 10\theta^4 + 41\theta^3 + 80\theta^2 + 72\theta + 24 + \\ &\quad (\theta + 1)(1 - \nu)(-3\theta^3 - 18\theta^2 - 42\theta - 36) + (1 - \nu)^2(2\theta^3 + 12\theta^2 + 24\theta + 16)] \\ &= [\theta^5 + 10\theta^4 + 41\theta^3 + 80\theta^2 + 72\theta + 24 + \nu(3\theta^4 + 21\theta^3 + 60\theta^2 + 78\theta + 36) + \\ &\quad -3\theta^4 - 21\theta^3 - 60\theta^2 - 78\theta - 36 + \nu^2(2\theta^3 + 12\theta^2 + 24\theta + 16) + \\ &\quad \nu(-4\theta^3 - 24\theta^2 - 48\theta - 32) + 2\theta^3 + 12\theta^2 + 24\theta + 16] \\ &= \nu^2(2\theta^3 + 12\theta^2 + 24\theta + 16) + \nu(3\theta^4 + 17\theta^3 + 36\theta^2 + 30\theta + 4) + \\ &\quad \theta^5 + 7\theta^4 + 22\theta^3 + 32\theta^2 + 18\theta + 4 \\ &= G. \end{aligned}$$



A.4 Coeficiente de curtose

Prova.

$$\begin{aligned}\mu_4 = E(X - \mu)^4 &= E(X^4) + \mu^4 - 4\mu E(X^3) + 6\mu^2 E(X^2) - 4\mu^3 E(X) \\ &= E(X^4) + 6\mu^2 E(X^2) - 3\mu^4 - 4\mu E(X^3).\end{aligned}$$

A seguir calculamos separadamente o valor de cada expressão que compõe o valor de μ_4 , calculando primeiro o valor de $E(X^4)$ e em seguida o de $6\mu^2 E(X^2)$, $-3\mu^4$ e de $-4\mu E(X^3)$, respectivamente.

$$E(X^4) = \sum_{x=0}^{\infty} x P_x = \sum_{x=1}^{\infty} x^4 (1 - \nu) \frac{\theta^2 (x + 2 + \theta)}{(\theta + 1)^{x+3}} = \theta^2 (1 - \nu) \sum_{x=1}^{\infty} \frac{x^4 (x + 2 + \theta)}{(\theta + 1)^{x+3}}.$$

Calculando, computacionalmente, a convergência das séries, ambas são convergentes pelo teste da razão, se $|1/(\theta + 1)| < 1$, como o parâmetro θ é sempre maior do que 0, a condição é satisfeita. Então:

$$\begin{aligned}\theta^2 (1 - \nu) \sum_{x=1}^{\infty} \frac{x^4 (x + 2 + \theta)}{(\theta + 1)^{x+3}} &= \theta^2 (1 - \nu) \frac{\theta^4 + 16\theta^3 + 78\theta^2 + 168\theta + 120}{\theta^6 (\theta + 1)} \\ &= (1 - \nu) \frac{\theta^4 + 16\theta^3 + 78\theta^2 + 168\theta + 120}{\theta^4 (\theta + 1)}.\end{aligned}$$

A seguir calculamos o valor de $6\mu^2 E(X^2)$:

$$\begin{aligned}6\mu^2 E(X^2) &= \frac{6(1 - \nu)^2 (\theta + 2)^2 (1 - \nu) (\theta^2 + 4\theta + 6)}{\theta^2 (\theta + 1)^2 \theta^2 (\theta + 1)} \\ &= \frac{(1 - \nu)^3 (6\theta^4 + 48\theta^3 + 156\theta^2 + 240\theta + 144)}{\theta^4 (\theta + 3)^3}.\end{aligned}$$

Agora, calculamos o valor de $-3\mu^4$:

$$-3\mu^4 = -3 \frac{(1 - \nu)^4 (\theta + 2)^4}{\theta^4 (\theta + 1)^4} = \frac{(1 - \nu)^4 (-3\theta^4 - 24\theta^3 - 72\theta^2 - 96\theta - 48)}{\theta^4 (\theta + 1)^4}.$$

A seguir calculamos o valor de $-4\mu E(X^3)$:

$$\begin{aligned}
-4\mu E(X^3) &= -4 \frac{(1-\nu)(\theta+2)}{\theta(\theta+1)} \frac{(1-\nu)(\theta^5 + 10\theta^4 + 41\theta^3 + 80\theta^2 + 72\theta + 24)}{\theta^3(\theta+1)^3} \\
&= \frac{(1-\nu)^2(-4\theta^6 - 48\theta^5 - 244\theta^4 - 648\theta^3 - 928\theta^2 - 672\theta - 192)}{\theta^4(\theta+1)^4}.
\end{aligned}$$

Anteriormente calculamos separadamente os termos que compõe o valor de $\mu_4 = E(X - \mu)^4$, a seguir uniremos todas as partes:

$$\begin{aligned}
\mu_4 &= \left[(1-\nu) \frac{\theta^4 + 16\theta^3 + 78\theta^2 + 168\theta + 120}{\theta^4(\theta+1)} \right. \\
&\quad + \frac{(1-\nu)^3(6\theta^4 + 48\theta^3 + 156\theta^2 + 240\theta + 144)}{\theta^4(\theta+3)^3} + \\
&\quad + \frac{(1-\nu)^4(-3\theta^4 - 24\theta^3 - 72\theta^2 - 96\theta - 48)}{\theta^4(\theta+1)^4} + \\
&\quad \left. \frac{(1-\nu)^2(-4\theta^6 - 48\theta^5 - 244\theta^4 - 648\theta^3 - 928\theta^2 - 672\theta - 192)}{\theta^4(\theta+1)^4} \right] \\
&= \frac{(1-\nu)}{\theta^4(\theta+1)} \left[(\theta^4 + 16\theta^3 + 78\theta^2 + 168\theta + 120) + \right. \\
&\quad \frac{(1-\nu)^2(6\theta^4 + 48\theta^3 + 156\theta^2 + 240\theta + 144)}{(\theta+3)^2} + \\
&\quad \frac{(1-\nu)^3(-3\theta^4 - 24\theta^3 - 72\theta^2 - 96\theta - 48)}{(\theta+1)^3} + \\
&\quad \left. \frac{(1-\nu)(-4\theta^6 - 48\theta^5 - 244\theta^4 - 648\theta^3 - 928\theta^2 - 672\theta - 192)}{(\theta+1)^3} \right].
\end{aligned}$$

Para simplificar as expressões, chamemos:

$$\begin{aligned}
U(\theta) &= (\theta^4 + 16\theta^3 + 78\theta^2 + 168\theta + 120) + \frac{(1-\nu)^2(6\theta^4 + 48\theta^3 + 156\theta^2 + 240\theta + 144)}{(\theta+3)^2} + \\
&\quad + \frac{(1-\nu)(-4\theta^6 - 48\theta^5 - 244\theta^4 - 648\theta^3 - 928\theta^2 - 672\theta - 192)}{(\theta+1)^3} \\
&\quad + \frac{(1-\nu)^3(-3\theta^4 - 24\theta^3 - 72\theta^2 - 96\theta - 48)}{(\theta+1)^3}.
\end{aligned}$$

Acima calculamos o valor de μ_4 , que representa o numerador do coeficiente de curtose que é dado por:

$$\begin{aligned}
\gamma_2 &= \frac{\mu_4}{\mu_2^2} - 3 = \frac{\frac{(1-\nu)U(\theta)}{\theta^4(\theta+1)}}{(1-\nu)^2\{\theta^3+4\theta^2+6\theta+2+\nu(\theta+2)^2\}^2} - 3 \\
&= \frac{(1-\nu)U(\theta)}{\theta^4(\theta+1)} \frac{\theta^4(\theta+1)^4}{(1-\nu)^2\{\theta^3+4\theta^2+6\theta+2+\nu(\theta+2)^2\}^2} - 3 \\
&= \frac{U(\theta)(\theta+1)^3}{(1-\nu)\{\theta^3+4\theta^2+6\theta+2+\nu(\theta+2)^2\}^2} - 3.
\end{aligned}$$

Expandindo novamente a parte $U(\theta)(1-\theta)^3$, chegamos no resultado final. Então, calculando a parte do numerador da fração anterior, temos que:

$$\begin{aligned}
U(\theta)(1-\theta)^3 &= (1-\theta)^3 \left[\frac{(\theta^4+16\theta^3+78\theta^2+168\theta+120)}{1} + \right. \\
&\quad \frac{(1-\nu)^2(6\theta^4+48\theta^3+156\theta^2+240\theta+144)}{(\theta+3)^2} + \\
&\quad \frac{(1-\nu)^3(-3\theta^4-24\theta^3-72\theta^2-96\theta-48)}{(\theta+1)^3} + \\
&\quad \left. \frac{(1-\nu)(-4\theta^6-48\theta^5-244\theta^4-648\theta^3-928\theta^2-672\theta-192)}{(\theta+1)^3} \right] \\
&= [\nu^3(3\theta^4+24\theta^3+72\theta^2+96\theta+48) + \nu^2(6\theta^5+45\theta^4+132\theta^3+180\theta^2+96\theta) \\
&\quad \nu(4\theta^6+36\theta^5+145\theta^4+312\theta^3+352\theta^2+192\theta+48) + \\
&\quad \theta^7+15\theta^6+87\theta^5+258\theta^4+406\theta^3+338\theta^2+144\theta+24] = \mu_4.
\end{aligned}$$

Então temos que o coeficiente de curtose é:

$$\begin{aligned}
\gamma_2 &= \frac{\mu_4}{(1-\nu)[\theta^3+4\theta^2+6\theta+2+\nu(\theta+2)^2]^2} - 3 \\
&= \frac{\mu_4 - 3(1-\nu)(\theta^3+4\theta^2+6\theta+2+\nu(\theta+2)^2)^2}{(1-\nu)\{\theta^3+4\theta^2+6\theta+2+\nu(\theta+2)^2\}^2}.
\end{aligned}$$

Calculando o numerador da equação acima temos que:

$$\begin{aligned}
\gamma_2 &= \mu_4 - 3(1 - \nu)[(\theta^3 + 4\theta^2 + 6\theta + 2 + \nu(\theta + 2))^2]^2 \\
&= [\nu^3(3\theta^4 + 24\theta^3 + 72\theta^2 + 96\theta + 48) + \\
&\quad \nu^2(6\theta^5 + 45\theta^4 + 132\theta^3 + 180\theta^2 + 96\theta) + \\
&\quad \nu(4\theta^6 + 36\theta^5 + 145\theta^4 + 312\theta^3 + \\
&\quad 352\theta^2 + 192\theta + 48) + \theta^7 + 15\theta^6 + 87\theta^5 + \\
&\quad 258\theta^4 + 406\theta^3 + 338\theta^2 + 144\theta + 24] + \\
&\quad (-3)(1 - \nu)\{\theta^3 + 4\theta^2 + 6\theta + 2 + \nu(\theta + 2)\}^2 \\
&= [\nu^3(3\theta^4 + 24\theta^3 + 72\theta^2 + 96\theta + 48) + \\
&\quad \nu^2(6\theta^5 + 45\theta^4 + 132\theta^3 + 180\theta^2 + 96\theta) + \\
&\quad \nu(4\theta^6 + 36\theta^5 + 145\theta^4 + \\
&\quad 312\theta^3 + 352\theta^2 + 192\theta + 48) + \\
&\quad \theta^7 + 15\theta^6 + 87\theta^5 + 258\theta^4 + 406\theta^3 + \\
&\quad 338\theta^2 + 144\theta + 24] + \\
&\quad [\nu^3(3\theta^4 + 24\theta^3 + 72\theta^2 + 96\theta + 48) + \\
&\quad \nu^2(6\theta^5 + 45\theta^4 + 132\theta^3 + 180\theta^2 + 96\theta) \\
&\quad \nu(3\theta^6 + 18\theta^5 + 36\theta^4 + 132\theta^3 - \\
&\quad 96\theta^2 - 120\theta - 36) \\
&\quad - 3\theta^6 - 24\theta^5 - 84\theta^4 - \\
&\quad 156\theta^3 - 156\theta^2 - 72\theta - 12] \\
&= 3\nu^3(2\theta^4 + 16\theta^3 + 48\theta^2 + 64\theta + 32) \\
&\quad 3\nu^2(4\theta^5 + 30\theta^4 + 88\theta^3 + 120\theta^2 + 64\theta) \\
&\quad \nu(7\theta^6 + 54\theta^5 + 181\theta^4 + \\
&\quad 444\theta^3 + 256\theta^2 + 72\theta + 12) \\
&\quad \theta^7 + 12\theta^6 + 63\theta^5 + 174\theta^4 + \\
&\quad 250\theta^3 + 182\theta^2 + 72\theta + 12 \\
&= A + B\nu + C3\nu^2 + D3\nu^3.
\end{aligned}$$

Finalmente o coeficiente de curtose, calculado, é dado por:

$$\gamma_2 = \frac{\mu_4}{\mu_2} - 3 = \frac{A + B\nu + 3\nu^2C + 3\nu^3D}{(1 - \nu)\{\theta^3 + 4\theta^2 + 6\theta + 2 + \nu(\theta + 2)^2\}^2},$$

em que:

$$A = \theta^7 + 12\theta^6 + 63\theta^5 + 174\theta^4 + 250\theta^3 + 182\theta^2 + 72\theta + 12$$

$$B = 7\theta^6 + 54\theta^5 + 181\theta^4 + 444\theta^3 + 256\theta^2 + 72\theta + 12$$

$$C = 4\theta^5 + 30\theta^4 + 88\theta^3 + 120\theta^2 + 64\theta$$

$$D = 2\theta^4 + 16\theta^3 + 48\theta^2 + 64\theta + 32.$$



A.5 Estimadores de Momentos

Prova. A seguir a demonstração para o estimador de ν , logo temos que:

$$\begin{aligned} \mu_1' &= \frac{(1 - \nu)(\theta + 2)}{\theta(\theta + 1)} \\ \mu_1'[\theta(\theta + 1)] &= (\theta + 2) - \nu(\theta + 2) \\ \nu(\theta + 2) &= (\theta + 2) - \mu_1'[\theta(\theta + 1)] \\ \hat{\nu} &= \frac{(\theta + 2) - \mu_1'[\theta(\theta + 1)]}{(\theta + 2)} \\ \tilde{\nu} &= 1 - \frac{\mu_1'[\theta(\theta + 1)]}{\theta + 2}. \end{aligned}$$

A seguir um estimador para θ é dado por:

$$\mu_2' = \frac{(1 - \tilde{\nu})(\theta^2 + 4\theta + 6)}{\theta^2(\theta + 1)}.$$

Substituindo o valor de $\tilde{\nu}$, na equação acima, por $1 - \frac{\mu_1'(\theta(\theta + 1))}{\theta + 2}$, temos:

$$\begin{aligned}
\mu'_2 &= \frac{\left\{1 - \left[1 - \frac{\mu'_1 \theta (\theta + 1)}{\theta + 2}\right]\right\}}{\theta^2 (\theta + 1)} \{\theta^2 + 4\theta + 6\} \\
\mu'_2 &= \frac{\left[\frac{\mu'_1 \theta (\theta + 1)}{\theta + 2}\right] (\theta^2 + 4\theta + 6)}{\theta^2 (\theta + 1)} \\
\mu'_2 &= \frac{\theta (\theta + 1) \mu'_1 (\theta^2 + 4\theta + 6)}{(\theta + 2) \theta^2 (\theta + 1)} \\
\mu'_2 &= \frac{\mu'_1 (\theta^2 + 4\theta + 6)}{\theta (\theta + 2)} \\
\mu'_2 (\theta^2 + 2\theta) &= \mu'_1 \theta^2 + 4\theta \mu'_1 + 6\mu'_1 \\
\mu'_2 \theta^2 + 2\mu'_2 \theta - \mu'_1 \theta^2 - 4\theta \mu'_1 - 6\mu'_1 &= 0 \\
\theta^2 (\mu'_2 - \mu'_1) - \theta (4\mu'_1 - 2\mu'_2) - 6\mu'_1 &= 0 \\
\Delta &= (4\mu'_1 - 2\mu'_2)^2 + 4(\mu'_2 - \mu'_1)6\mu'_1 \\
\Delta &= 16\mu_1'^2 - 16\mu'_1 \mu'_2 + 4\mu_2'^2 + 24\mu'_1 \mu'_2 - 24\mu_1'^2 \\
\Delta &= -8\mu_1'^2 + 4\mu_2'^2 + 8\mu'_2 \mu'_1 \\
\tilde{\theta} &= \frac{(4\mu'_1 - 2\mu'_2) \pm \sqrt{-8\mu_1'^2 + 4\mu_2'^2 + 8\mu'_2 \mu'_1}}{2(\mu'_2 - \mu'_1)} \\
\tilde{\theta} &= \frac{(\mu'_1 - \mu'_2) + \sqrt{-2\mu_1'^2 + \mu_2'^2 + 2\mu'_2 \mu'_1}}{(\mu'_2 - \mu'_1)},
\end{aligned}$$

em que μ'_1 e μ'_2 são o primeiro e segundo momentos centrais, respectivamente. ■

Bibliografia

BORAH, M.; NATH, A. D. A study on the inflated poisson lindley distribution. *Jour. Ind. Soc. Ag. Statistics*, v. 54, n. 3, p. 317–323, 2001.

CONSUL, P. *Generalized Poisson Distribution Properties and Applications*. New York and Basel: Marcel Dekker, Inc., 1989.

EFRON, B. Bootstrap methods: Another look at the jackknife. *Ann. Statist.*, The Institute of Mathematical Statistics, p. 1–26, 1979. Disponível em: <<<http://projecteuclid.org/euclid.aos/1176344552>>>.

EFRON, B.; HINKLEY, D. Assessing the accuracy of the maximum likelihood estimator: Observed versus expected fisher information. *Biometrika*, [Oxford University Press, Biometrika Trust], v. 65, n. 3, p. 457–482, 1978. Disponível em: <<<http://www.jstor.org/stable/2335893>>>.

GHITANY, M.; AL-MUTAIRI, D. K. Estimation methods for the discrete poisson-lindley distribution. *Journal of Statistical Computation and Simulation*, p. 1–9, 2009.

GHITANY, M. E.; AL-MUTAIRI, D. K.; NADARAJAH, S. Zero-truncated poisson-lindley distribution and its application. *Mathematics and Computers in Simulation*, v. 79, p. 279–287, 2008.

JOHNSON, N.; KEMP, A.; KOTZ, S. *Univariate Discrete Distribution*. New York: Wiley, 2005.

LINDLEY, D. Fiducial distributions and baye's theorem. *J. Roy. Statist. Soc. B*, v. 20, p. 102–107, 1958.

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2017.

RIDOUT, M.; BESBEAS, P. An empirical model for underdispersed count data. *Statistical Modelling*, SAGE Publications, v. 4, n. 1, p. 77–89, 2017/06/29 2004. Disponível em: <<<http://journals.sagepub.com/doi/abs/10.1191/1471082X04st0640a>>>.

SANKARAN, M. The discrete poisson-lindley distribution. *Biometrics*, v. 26, p. 145–149, 1970.

SHANKER, R.; FESSHAYE, H. On poisson-lindley distribution and its applications to biological sciences. *Biometrics & Biostatistics International Journal*, v. 2, p. 1–5, 2015.