

UNIVERSIDADE DE BRASÍLIA
DEPARTAMENTO DE ESTATÍSTICA

A distribuição de Touchard e suas aplicações

Lucas Keniti Nanami

*Projeto apresentado como requisito parcial para
obtenção do título de Bacharel em Estatística.*

Brasília

2016

LUCAS KENITI NANAMI

A distribuição de Touchard e suas aplicações

Orientador:

Prof. Dr. **Raul Yukihiro Matsushita**

Brasília

2016

Sumário

1	Apresentação	5
1.1	Introdução	5
1.2	Objetivos do trabalho	6
1.3	Estrutura do trabalho	6
2	Modelos de Contagem Não Poisson	7
2.1	Introdução	7
2.2	Distribuição Binomial negativa	7
2.3	Distribuição hiper-Poisson	11
2.4	Distribuição de Poisson generalizada de Consul e Jain	13
2.5	Distribuição e modelo de regressão de Poisson inflado com zeros	15
2.6	Distribuição de Poisson generalizada de Chandra e Roy	17
2.7	Distribuição de Poisson-Lindley	19
2.8	Distribuição de Conway-Maxwell-Poisson	20
3	A distribuição de Touchard	23
3.1	Introdução	23
3.2	Distribuição de probabilidade	23
3.3	Momentos	24
3.4	Estatísticas suficientes e máxima verossimilhança	26
3.5	Teste da razão de verossimilhança	27
3.6	Critérios de informação	28
3.7	Procedimentos para estimação dos parâmetros	28
4	Resultados	31
5	Conclusão	41
	Referências Bibliográficas	43
	Anexo	45

Capítulo 1

Apresentação

1.1 Introdução

Um dos propósitos da estatística é estudar e descrever fenômenos reais através de ferramentas probabilísticas. Ao longo dos anos, diversos modelos teóricos foram propostos na tentativa de representar matematicamente o comportamento destes fenômenos.

Denominamos de distribuições de probabilidade os modelos que atribuem probabilidades aos possíveis resultados de um experimento aleatório. Um modelo bastante utilizado para ajustar dados de contagens é a distribuição de Poisson, introduzido por Siméon Denis Poisson em 1837. A variável aleatória de Poisson tem aplicações em diversas disciplinas, como teoria de filas, teoria de estoques e engenharia de confiabilidade. Essa distribuição possui um parâmetro λ que representa a média e a variância do modelo. Na prática, porém, existem situações em que a taxa varia no tempo, o que pode causar superdispersão ou subdispersão nos dados, e também o excesso de zeros. Nesses casos, a distribuição de Poisson pode não ser apropriada, já que ela não permite controlar a média e a variância de forma independente.

Várias extensões da distribuição de Poisson foram propostas na literatura. Esses modelos alternativos possuem um número extra de parâmetros na tentativa de explicar a relação entre a média e a variância. Por exemplo, uma distribuição utilizada para o excesso de zeros é a Poisson inflada de zeros, que propõe que os zeros provêm de um processo de Bernoulli independente da parte Poisson. Mas o excesso de zeros é um problema difícil de resolver com um modelo de estado único.

A proposta deste trabalho é estudar a distribuição de Touchard como uma nova generalização da Poisson. Essa distribuição possui dois parâmetros, λ e δ , que permitem melhor ajuste para dados com superdispersão ou subdispersão, e também para o excesso de zeros. Acredita-se que ela seja útil por sua flexibilidade e simplicidade. Por exemplo, esse modelo não requer mistura entre dois processos estocásticos distintos, e a forma do modelo é relativamente mais simples comparada a outras extensões propostas na literatura. A constante de normalização da função de distribuição de probabilidade se relaciona com os polinômios

de Touchard (Touchard, 1939; Chrysaphinou, 1985), o que justifica o nome a distribuição. Outras propriedades e características da distribuição serão apresentadas neste trabalho.

Será apresentado um resumo de alguns modelos de contagem de dados não Poisson propostos na literatura, com representações gráficas das distribuições e suas propriedades, para posteriormente comparar a qualidade de ajuste destas distribuições com a Touchard. Na parte final do trabalho, é realizada uma aplicação das distribuições no banco de dados de acidentes de veículos motorizados no condado de New York nos anos de 2011 a 2013, disponibilizada pelo website oficial do estado de New York.

1.2 Objetivos do trabalho

O objetivo geral do trabalho é estudar a distribuição de Touchard, suas propriedades e aplicações, e avaliar se esse modelo permite descrever dados de contagem não Poisson em situações práticas.

Outro objetivo é realizar um estudo sobre modelos alternativos de contagem propostos na literatura, e comparar seus desempenhos com a Touchard. Implementaremos rotinas de estimação dos parâmetros das distribuições para o cálculo do AIC e BIC. Por último, utilizaremos a base de dados dos acidentes de veículos motorizados do condado de New York para aplicação prática das distribuições.

1.3 Estrutura do trabalho

O capítulo 2 tratará de modelos utilizados para contagens não Poisson e suas características. No capítulo 3, será apresentado a distribuição de Touchard e suas propriedades. O capítulo também apresenta o teste de razão de verossimilhança, os critérios de informação de Akaike e bayesiano, e os procedimentos computacionais para a estimação dos parâmetros. O capítulo 4 traz as análises e os resultados do trabalho feitas sobre a base de dados escolhida. Por último, o capítulo 5 apresenta a conclusão do trabalho.

Capítulo 2

Modelos de Contagem Não Poisson

2.1 Introdução

Neste capítulo serão apresentados alguns modelos para contagem de dados não Poisson, apresentando algumas características e propriedades de cada distribuição. Exibiremos as equações de máxima verossimilhança para estimação dos parâmetros dos modelos. Para equações sem forma fechada, será necessário o uso de técnicas de otimização numérica como o algoritmo EM e o método de Newton-Raphson.

O capítulo é dividido em seções, cada uma apresentando um modelo. A seção 2.2 falará sobre a distribuição binomial negativa. Na seção 2.3 será apresentada a distribuição hiper-Poisson. A seção 2.4 apresenta a distribuição de Poisson generalizada, proposta por Consul e Jain. A seção 2.5 traz a distribuição e o modelo de regressão de Poisson inflada com zeros. Na seção 2.6 é apresentada a distribuição de Poisson generalizada de Chandra e Roy. A seção 2.7 trata da distribuição de Poisson-Lindley. Por último, na seção 2.8, temos a distribuição de Conway-Maxwell-Poisson.

2.2 Distribuição Binomial negativa

A distribuição binomial negativa, conhecida também como distribuição de Pascal, é o modelo de probabilidade do número n de sucessos de ensaios independentes de Bernoulli. A distribuição é também resultado da soma de n distribuições geométricas de parâmetro p .

O modelo da binomial negativa pode ser originada de uma distribuição de probabilidade composta, em que assume-se que o parâmetro λ é variável aleatória com distribuição Gama(α, β).

Seja $Y \sim \text{Poisson}(\lambda)$ e $\lambda \sim \text{Gama}(\alpha, \beta)$, então a distribuição de probabilidade composta

é dada por

$$\begin{aligned} p(y|\alpha, \beta) &= \int_0^\infty P_P(y|\lambda)P_G(\lambda|\alpha, \beta)d\lambda = \int_0^\infty \frac{e^{-\lambda}\lambda^y}{y!} \frac{\lambda^{\alpha-1}e^{-\lambda/\beta}}{\Gamma(\alpha)\beta^\alpha} d\lambda \\ &= \frac{1}{\Gamma(\alpha)\beta^\alpha y!} \int_0^\infty \lambda^{\alpha+y-1} e^{-\lambda(1+1/\beta)} d\lambda = \frac{\Gamma(\alpha+y)}{\Gamma(\alpha)\beta^\alpha y!} \left(\frac{\beta}{\beta+1}\right)^{\alpha+y} \\ &= \frac{\Gamma(\alpha+y)\beta^y}{\Gamma(\alpha)y!(\beta+1)^{\alpha+y}}, \quad y = 0, 1, 2, \dots, \end{aligned}$$

em que Y tem distribuição Gama-Poisson com parâmetros α e β , ambos positivos. Para $n = \alpha$ e $p = 1/(\beta + 1)$, X segue uma distribuição binomial negativa de parâmetros n e p .

$$\begin{aligned} p(x|n, p) &= \frac{\Gamma(n+x)}{\Gamma(n)x!} \left(\frac{p}{1-p}\right)^n (1-p)^{n+x} = \frac{\Gamma(n+x)}{\Gamma(n)\Gamma(x+1)} p^n (1-p)^x \\ &= \binom{n+x-1}{x} p^n (1-p)^x, \quad x = 0, 1, 2, \dots, \end{aligned}$$

em que $n > 0$ e $0 < p < 1$. A distribuição de Poisson é um caso particular da distribuição binomial negativa, se $p = \lambda/(n + \lambda)$ e $n \rightarrow \infty$, $p(x|n, p)$ converge para uma $\text{Poisson}(\lambda)$.

A seguir são dadas as expressões da esperança e variância de X .

$$E[X] = n \left(\frac{1-p}{p}\right), \quad \text{Var}[X] = n \left(\frac{1-p}{p^2}\right).$$

A distribuição de Poisson possui média e variância iguais, que são dependentes exclusivamente de λ , o que dificulta a modelagem de dados com a presença de subdispersão ou superdispersão. O modelo da binomial negativa se ajusta melhor a problemas com superdispersão, já que $\text{Var}[X] > E[X]$ para $n > 0$ e $0 < p < 1$. E com um parâmetro a mais, o controle sobre a média e a variância é maior.

A distribuição binomial negativa possui a propriedade da convolução. A soma de n variáveis binomiais negativas independentes, de parâmetro (n_i, p) , tem distribuição binomial negativa com parâmetro $(\sum n_i, p)$.

Segue abaixo mais algumas características da distribuição:

Função geradora de momentos:

$$M_X(t) = E(e^{tx}) = \frac{p^r}{(1 - (1-p)e^t)^r}, \quad t < -\ln(1-p).$$

Função característica:

$$\varphi_x(t) = E(e^{itx}) = \frac{p^r}{(1 - (1-p)e^{it})^r}, \quad t \in \mathbb{R}.$$

Função distribuição acumulada:

$$F(x) = P(X \leq k) = 1 - I_p(k+1, n-k).$$

em que I_p é a função beta incompleta regularizada.

A seguir são dados exemplos de distribuições binomiais negativas.

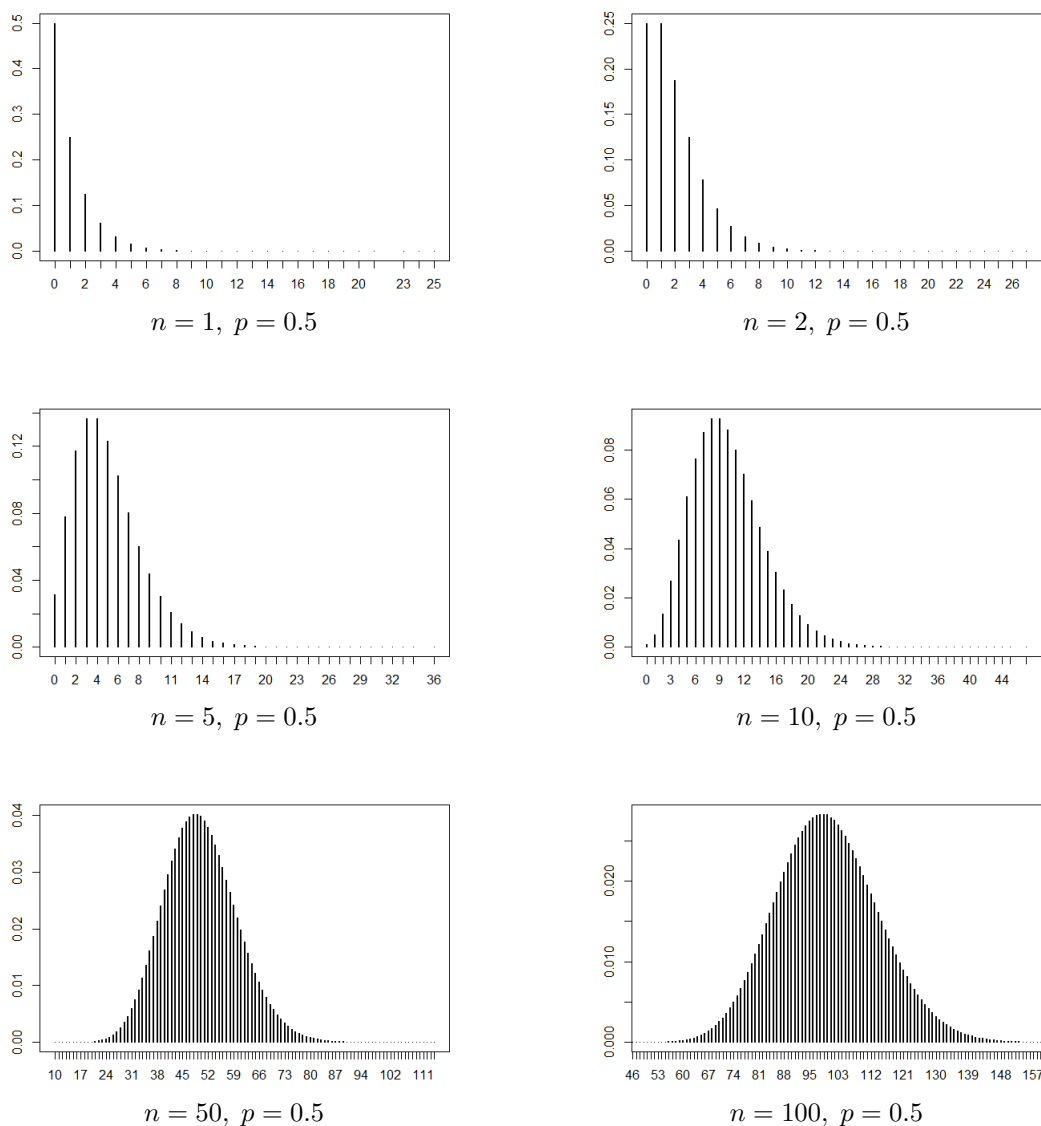


Figura 2.1: Gráficos da distribuição Binomial negativa, com $p = 0.5$ e n variando de 1 a 100.

O parâmetro n é um fator multiplicativo da esperança e variância. A média cresce quando o valor de n aumenta, e a variância é inflada na mesma proporção. Os gráficos para $n = 50$ e $n = 100$ indicam modelos com dispersão mais alta.

Os gráficos a seguir mostram que a média e a variância crescem a medida que a probabilidade de falha aumenta, dado um n fixo.

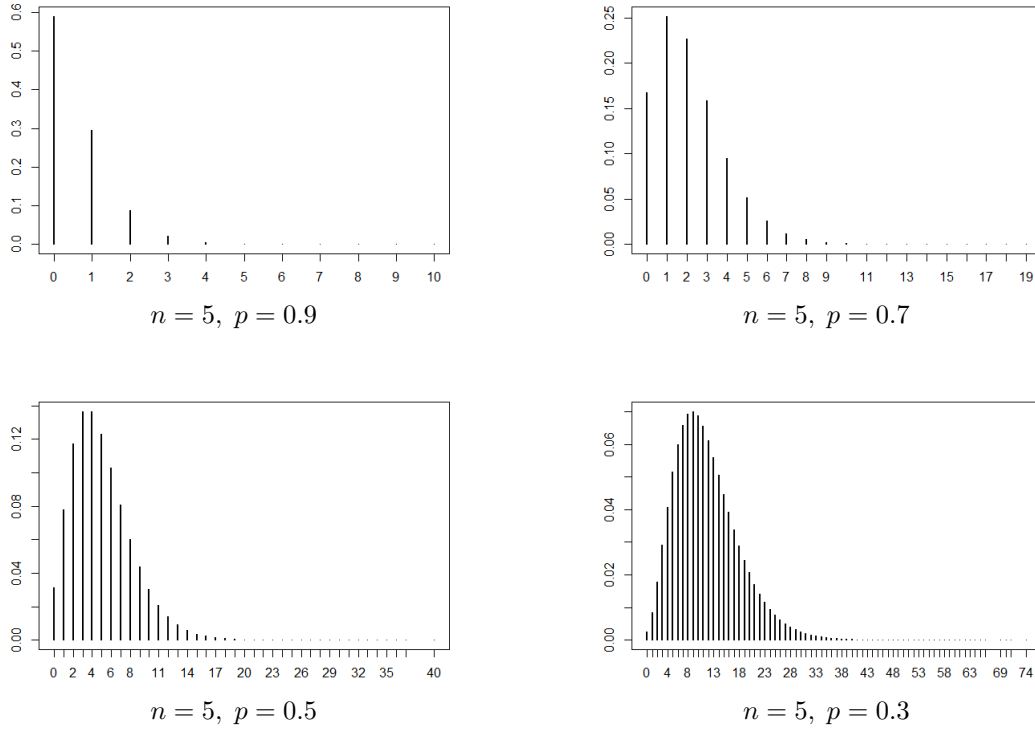


Figura 2.2: Gráficos da distribuição Binomial negativa, com $n = 5$ e p variando de 0.3 a 0.9.

Dada uma amostra de N realizações independentes da distribuição, a função de log-verossimilhança é dada por

$$l(n, p) = \sum_{i=1}^N \ln(\Gamma(x_i + n)) - \sum_{i=1}^N \ln(x_i!) - N \ln(\Gamma(n)) + \sum_{i=1}^N x_i \ln(1 - p) + Nn \ln(p).$$

A estimativa de p por máxima verossimilhança é dada por

$$p = \frac{Nn}{Nn + \sum_{i=1}^N x_i}.$$

Para o parâmetro n , a equação não tem forma fechada, e é dada por

$$N \ln \left(\frac{Nn}{Nn + \sum_{i=1}^N x_i} \right) - N\psi(n) + \sum_{i=1}^N \psi(n + x_i) = 0,$$

em que $\psi(y) = \partial/\partial y(\ln \Gamma(y))$ é a função digama de y .

Para encontrar os parâmetros de máxima verossimilhança, utilizaremos o método de otimização de Newton-Raphson. As derivadas de segunda ordem são dadas a seguir.

$$\frac{\partial^2 l(n, p)}{\partial n^2} = -N\varphi(n) + \sum_{i=1}^N \varphi(n + x_i),$$

$$\frac{\partial^2 l(n, p)}{\partial p^2} = -\frac{Nn}{p^2} - \frac{1}{(1-p)^2} \sum_{i=1}^N x_i,$$

$$\frac{\partial^2 l(n, p)}{\partial n \partial p} = \frac{N}{p},$$

em que $\varphi(y)$ é a função trigama de y .

2.3 Distribuição hiper-Poisson

Em 1964, Bardwell e Crow propuseram uma nova generalização da distribuição de Poisson de dois parâmetros, que é um caso particular da distribuição de série hipergeométrica confluyente. A distribuição foi estudada por diversos autores como Nisida (1962), Roohi e Ahmad (2003), Kumar e Nair (2011), e Kemp (2002). Uma variável aleatória X , $x = 0, 1, 2, \dots$, tem distribuição hiper-Poisson se possui a seguinte distribuição de probabilidade:

$$p(x|\lambda, \theta_2) = \frac{\Gamma(\lambda)\theta_2^x}{{}_1F_1[1, \lambda, \theta_2]\Gamma(\lambda + x)}, \quad x = 0, 1, 2, \dots,$$

em que $\lambda = \theta_2 - \theta_1 + 1 > 0$, $\theta_2 > 0$ e

$${}_1F_1[1, \lambda, \theta_2] = 1 + \frac{\theta_2}{\lambda} + \frac{\theta_2^2}{\lambda(\lambda + 1)} + \frac{\theta_2^3}{\lambda(\lambda + 1)(\lambda + 2)} + \dots$$

é a função hipergeométrica confluyente. A média e variância de X são:

$$E[X] = \mu = \theta_1 + \frac{(\theta_2 - \theta_1)}{{}_1F_1[1, \lambda, \theta_2]}, \quad \text{Var}[X] = \theta_2 + (\theta_1 - \mu)\mu.$$

Se $\lambda = 1$, isto é $\theta_1 = \theta_2$, a distribuição hiper-Poisson se reduz a Poisson. O modelo é superdisperso quando $\lambda > 1$, ou seja $\theta_2 > \theta_1$, e subdisperso se $0 < \lambda < 1$, em que $\theta_1 > \theta_2$.

A seguir são apresentadas alguns gráficos da distribuição.

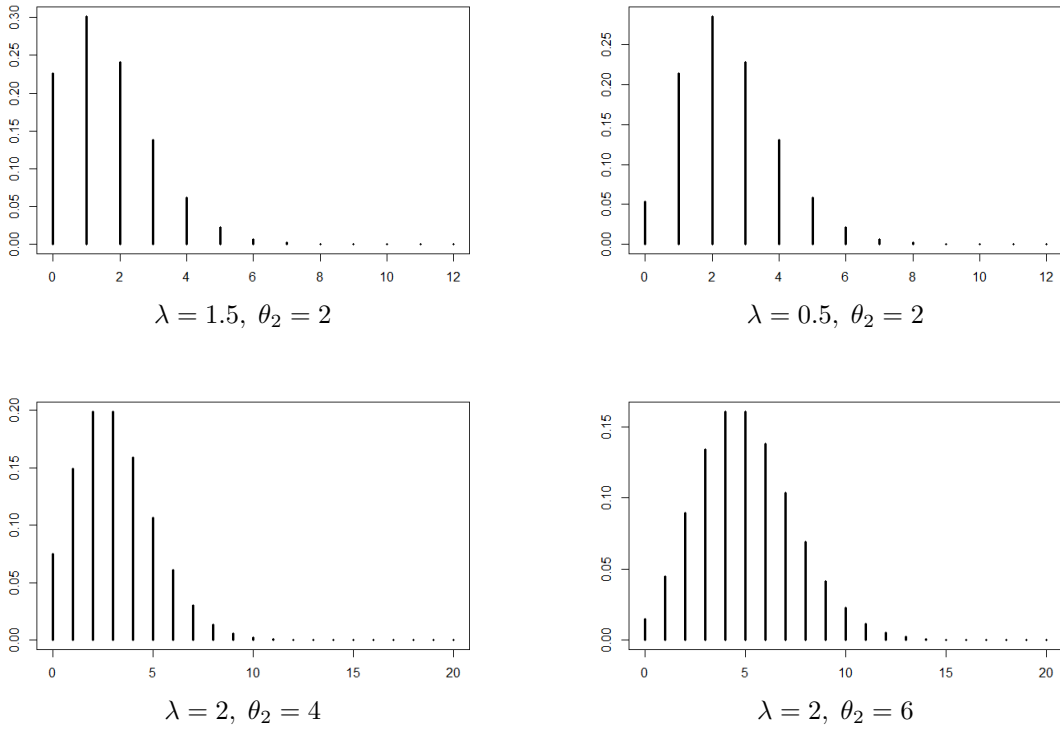


Figura 2.3: Gráficos da distribuição hiper-Poisson.

O gráfico no canto superior direito da figura representa a distribuição hiper-Poisson de parâmetros $\lambda = 0.5$ e $\theta_2 = 2$. O modelo é subdisperso, já que $\lambda = 0.5$, enquanto que a distribuição com $\lambda = 1.5$ é superdispersa. Note que a média é significativamente alterada quando qualquer um dos parâmetros varia.

Dada uma amostra de n realizações independentes da distribuição, a função de log-verossimilhança é dada por

$$l(\lambda, \theta_2) = n\bar{x} \ln(\theta_2) - n \ln {}_1F_1[1, \lambda, \theta_2] + n \ln \Gamma(\lambda) - \sum_{i=1}^n \ln \Gamma(\lambda + x_i).$$

Bardwell e Crow demonstraram que as estimativas por máxima verossimilhança quando os dois parâmetros são desconhecidos são dadas pela solução da seguinte equação:

$$\begin{cases} \frac{\theta_2}{{}_1F_1[1, \lambda, \theta_2]} \frac{\partial {}_1F_1[1, \lambda, \theta_2]}{\partial \theta_2} = \bar{x} \\ \frac{1}{{}_1F_1[1, \lambda, \theta_2]} \frac{\partial {}_1F_1[1, \lambda, \theta_2]}{\partial \lambda} - \psi(\lambda) + \frac{1}{n} \sum_{i=1}^n \psi(\lambda + x_i) = 0 \end{cases}$$

em que $\psi(y) = \partial/\partial y(\ln \Gamma(y))$ é a função digama de y .

Iremos resolver estas equações e encontrar os estimadores de máxima verossimilhança

através do método de Newton-Raphson. As derivadas de segunda ordem são dadas a seguir.

$$\begin{aligned} \frac{\partial^2 l(\lambda, \theta_2)}{\partial \lambda^2} &= \frac{n}{{}_1F_1(1, \lambda, \theta_2)} \sum_{k=0}^{\infty} \left(\frac{\theta_2^k}{(\lambda)_k} \left((\varphi(k + \lambda) - \varphi(\lambda)) - (\psi(k + \lambda) - \psi(\lambda))^2 \right) \right) + \\ &+ \frac{n}{{}_1F_1(1, \lambda, \theta_2)^2} \left(\sum_{k=0}^{\infty} \left(\frac{\theta_2^k}{(\lambda)_k} (\psi(k + \lambda) - \psi(\lambda)) \right) \right)^2 + n\varphi(\lambda) - \sum_{i=1}^n \varphi(\lambda + x_i). \end{aligned}$$

$$\frac{\partial^2 l(\lambda, \theta_2)}{\partial \theta^2} = \frac{-n\bar{x}}{\theta_2^2} + \frac{n}{\lambda^2} - \frac{2n}{\lambda(\lambda + 1)} \frac{{}_1F_1(3, \lambda + 2, \theta_2)}{{}_1F_1(1, \lambda, \theta_2)}.$$

$$\begin{aligned} \frac{\partial^2 l(\lambda, \theta_2)}{\partial \lambda \partial \theta_2} &= \frac{n}{{}_1F_1(1, \lambda, \theta_2)} \sum_{k=0}^{\infty} \frac{k(\theta_2^{k-1})}{(\lambda)_k} (\psi(k + \lambda) - \psi(\lambda)) \\ &- n \frac{{}_1F_1(2, \lambda + 1, \theta_2)}{\lambda({}_1F_1(1, \lambda, \theta_2))^2} \sum_{k=0}^{\infty} \frac{\theta_2^k}{(\lambda)_k} (\psi(k + \lambda) - \psi(\lambda)). \end{aligned}$$

em que $(\lambda)_k = \lambda(\lambda + 1) \dots (\lambda + k - 1)$ e $\varphi(y) = \partial/\partial y(\psi(y))$ é a função trigama de y .

2.4 Distribuição de Poisson generalizada de Consul e Jain

Em 1973, Consul e Jain apresentaram a distribuição de Poisson generalizada com a justificativa de que o modelo é capaz de se ajustar a uma grande variedade de padrões encontrados, incluindo uma boa aproximação para dados Poisson, binomial e binomial negativa.

A distribuição é derivada de um caso particular da distribuição binomial negativa generalizada. Seja X uma variável aleatória com distribuição de Poisson generalizada, sua distribuição de probabilidade é dada por

$$p(x|\lambda_1, \lambda_2) = \frac{\lambda_1(\lambda_1 + x\lambda_2)^{x-1}}{x!} e^{-\lambda_1 - x\lambda_2}, \quad x = 0, 1, 2, \dots,$$

para $\lambda_1 > 0$ e $|\lambda_2| < 1$. E temos que

$$p(x|\lambda_1, \lambda_2) = 0 \quad \text{para } x \geq m \quad \text{se } \lambda_1 + m\lambda_2 \leq 0.$$

A média e a variância da distribuição são dadas a seguir.

$$\mu = E[X] = \frac{\lambda_1}{1 - \lambda_2}, \quad \text{Var}[X] = \frac{\mu}{(1 - \lambda_2)^2} = \frac{\lambda_1}{(1 - \lambda_2)^3}.$$

Note que a média da distribuição será menor que a variância se $\lambda_2 > 0$, maior se $\lambda_2 < 0$ e igual se $\lambda_2 = 0$, em que temos a distribuição clássica de Poisson. Uma variação pequena em λ_2 causa alterações grandes na variância por causa do termo cúbico presente na equação, enquanto a média é afetada em menor proporção. Exemplificaremos o comportamento da distribuição com gráficos dados a seguir.

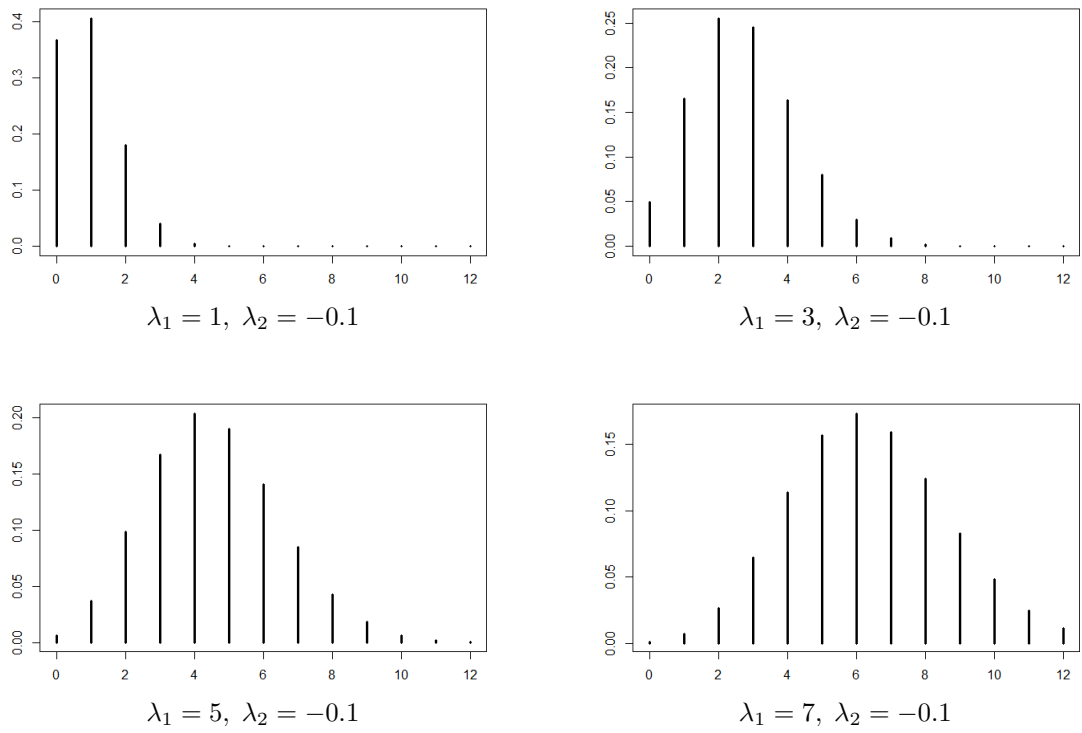


Figura 2.4: Gráficos da distribuição de Poisson generalizada, para $\lambda_2 = -0.1$ e λ_1 variando de 1 a 7.

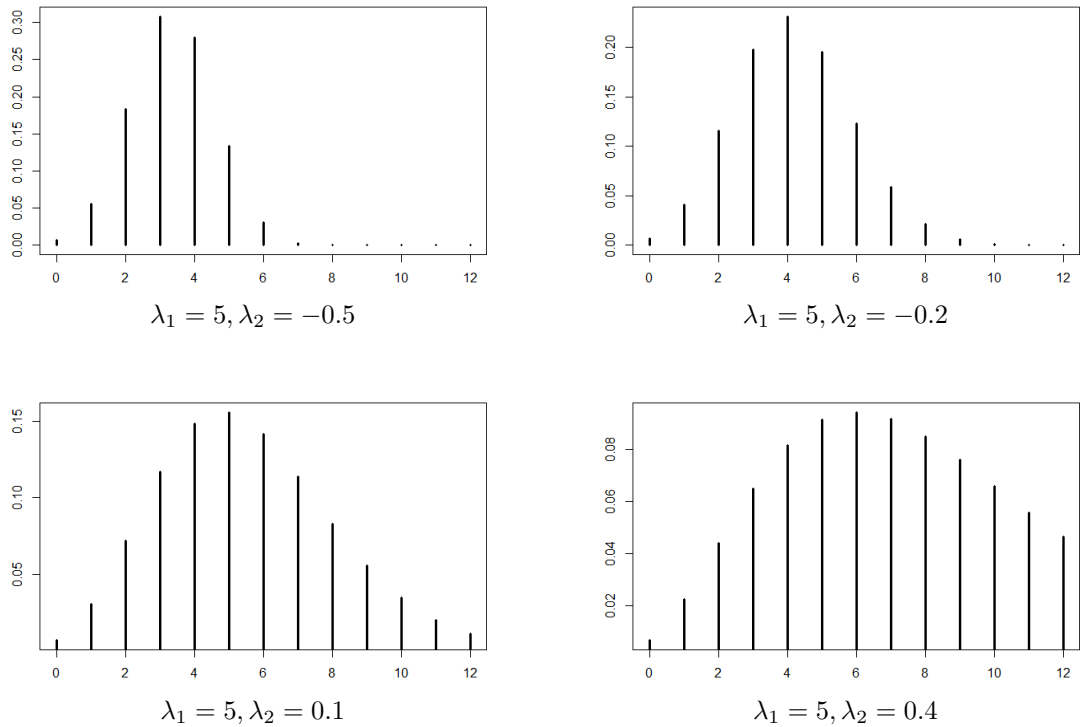


Figura 2.5: Gráficos da distribuição de Poisson generalizada, para $\lambda_1 = 5$ e λ_2 variando de -0.5 a 0.4.

A figura 2.4 apresenta exemplos em que o parâmetro λ_2 é fixo e próximo de zero. Assim, o modelo tem média e variância próximas, e um comportamento similar ao modelo de Poisson.

Quando λ_1 aumenta, a variância e a média crescem na mesma proporção, o que causa o acúmulo de probabilidade em valores menores quando a média é baixa e o formato mais simétrico para valores maiores.

Os gráficos da figura 2.5 mostram o quanto a variância do modelo cresce a uma pequena variação de λ_2 , dado λ_1 fixo.

Uma propriedade interessante da distribuição é a propriedade de convolução. Consul e Jain mostraram que a soma de n variáveis independentes com parâmetros $(\lambda_{1,r}, \lambda_2)$ tem distribuição de Poisson Generalizada de parâmetro $(\sum \lambda_{1,r}, \lambda_2)$. Também foi demonstrado que o k -ésimo momento da distribuição pode ser expresso como

$$E[X^k] = \sum_{x=j}^{\infty} \sum_{j=1}^k S(k, j) \lambda_1 (\lambda_1 + x \lambda_2)^{x-1} e^{-(\lambda_1 + x \lambda_2)} / (x - j)!,$$

em que $S(k, j)$ é o número de Stirling de segunda ordem, dado por

$$S(k, j) = \frac{1}{j!} \sum_{i=0}^j (-1)^{j-i} \binom{j}{i} i^k.$$

Tripathi e Gupta (1984) apresentaram as equações de máxima verossimilhança para estimativa dos dois parâmetros,

$$\begin{cases} \hat{\lambda}_1 = \frac{\bar{x}}{1 + (\lambda_2/\lambda_1)\bar{x}} \\ \sum_{i=1}^n \frac{x_i(x_i - 1)}{(1 + (\lambda_2/\lambda_1)x_i)} - n\hat{\lambda}_1 = 0 \end{cases},$$

que pode ser resolvida pelo método de otimização de Newton-Raphson.

2.5 Distribuição e modelo de regressão de Poisson inflado com zeros

Em 1992, Lambert publicou uma aplicação do modelo de regressão de Poisson inflado com zeros (zero inflated Poisson - ZIP) com covariáveis em dados sobre defeitos de fabricação. O modelo é utilizado em dados com excesso de zeros, com a ideia de que a contagem de zeros provém de um processo separado da subpopulação com distribuição Poisson, e que pode ser modelado independentemente. O modelo de regressão ZIP sem covariáveis foi estudado por Cohen (1963) e Johnson e Kotz (1969).

A distribuição ZIP tem a seguinte função de probabilidade.

$$P(Y = k) = \begin{cases} p + (1 - p)e^{-\lambda}, & k = 0 \\ (1 - p) \frac{\lambda^k e^{-\lambda}}{k!}, & k = 1, 2, \dots \end{cases}$$

em que $k = 1, 2, \dots$, para $\lambda > 0$ e $0 \leq p \leq 1$. O modelo sugere que as observações provém de duas subpopulações Quando $p = 0$, Y tem distribuição de Poisson.

Na regressão ZIP (Lambert, 1992), dado Y_i variáveis respostas, temos que

$$Y_i = \begin{cases} 0, & \text{com probabilidade } p_i + (1 - p_i)e^{-\lambda_i} \\ k, & \text{com probabilidade } (1 - p_i)\frac{e^{-\lambda_i}\lambda_i^k}{k!} \end{cases}$$

A média e variância do modelo de regressão ZIP são dadas pelas seguintes expressões:

$$E[Y_i] = (1 - p_i)\lambda_i, \quad \text{Var}[Y_i] = (1 - p_i)(1 + \lambda_i p_i)\lambda_i.$$

Note que

$$\frac{\text{Var}[Y]}{E[Y]} = 1 + \lambda p,$$

então o modelo de regressão de dois estados apresentado é superdisperso para todos os valores dos parâmetros, já que o excesso de zeros representa de certa maneira uma forma de superdispersão. Lambert considerou o modelo com as seguintes funções de ligação:

$$\ln(\lambda) = X\beta, \quad \ln\left(\frac{p}{1-p}\right) = Z\gamma,$$

em que X e Z são matrizes de covariáveis, e β e γ os vetores dos parâmetros.

As estimativas de máxima verossimilhança para os parâmetros do modelo de regressão ZIP são dadas pela solução que maximiza a seguinte equação:

$$\sum_{i=1}^n (u_i Z_i \gamma - \ln(1 + e^{Z_i \gamma})) + \sum_{i=1}^n (1 - u_i)(y_i X_i \beta - e^{X_i \beta}) - \sum_{i=1}^n (1 - u_i) \ln(y_i!) = 0,$$

em que X_i e Z_i representam a i -ésimo linha de X e Z , e u_i indica se y_i pertence ao estado de zeros, $u_i = 1$, ou ao estado de Poisson, $u_i = 0$. A equação pode ser resolvida via algoritmo EM.

A seguir são apresentadas algumas representações gráficas da distribuição ZIP:

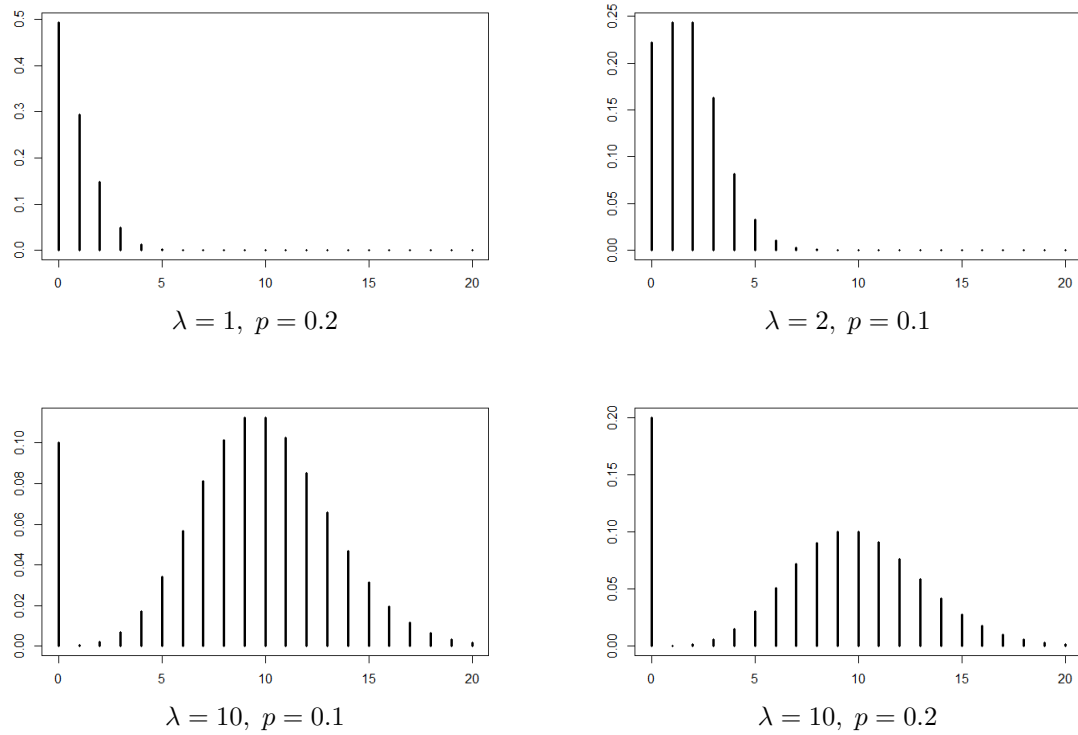


Figura 2.6: Gráficos da distribuição ZIP.

O parâmetro p determina a massa de probabilidade distribuída entre os dois estados. Quando λ assume valores tal que $(1-p)e^{-\lambda} \approx 0$, temos que $P(Y=0) \approx p$, como observado nas distribuições de parâmetros $\lambda = 10$ com $p = 0.1$ e $p = 0.2$.

2.6 Distribuição de Poisson generalizada de Chandra e Roy

Em 2013, Chandra, Roy e Ghosh proporem uma extensão da Poisson, nomeada de Poisson generalizada. A distribuição possui dois parâmetros, λ e α , e provém de sucessivas integrais parciais da distribuição gama. Seja X uma variável aleatória com distribuição de Poisson generalizada, sua distribuição de probabilidade é dada por

$$P(X = k) = \begin{cases} \frac{1}{\Gamma(1 + \alpha)} \int_{\lambda}^{\infty} z^{\alpha} e^{-z} dz, & \text{se } k = 0 \\ \frac{e^{-\lambda} \lambda^{k+\alpha}}{\Gamma(k + \alpha + 1)}, & \text{se } k = 1, 2, \dots \end{cases}$$

em que $\lambda \geq 0$ e $0 \leq \alpha < 1$. A distribuição se reduz a Poisson se $\alpha = 0$.

A função geradora de momentos pode ser escrita como

$$M_X(t) = p_0 + \frac{e^{-\alpha t - \lambda(1-e^t)}}{\Gamma(1 + \alpha)} \int_0^{\lambda e^t} z^{\alpha} e^{-z} dz,$$

em que $p_0 = P(X = 0)$. Assim, obtemos a esperança e variância da distribuição,

$$E[X] = (\lambda - \alpha)(1 - p_0) + (1 + \alpha)p_1,$$

$$\begin{aligned} \text{Var}(X) = & [(\lambda - \alpha)^2 + \lambda](1 - p_0) + (1 + \alpha)(\lambda - \alpha + 1)p_1 - (\lambda - \alpha)^2(1 - p_0)^2 \\ & - (1 + \alpha)^2 p_1^2 - 2(1 + \alpha)(\lambda - \alpha)(1 - p_0)p_1. \end{aligned}$$

As fórmulas da esperança e variância são complexas, impossibilitando identificar pontualmente em que condições o modelo seria subdisperso ou superdisperso. Os autores da distribuição testaram valores de λ até 30 e α até 0.9, e concluíram que a variância do modelo seria sempre menor que λ , que é a variância do modelo de Poisson.

A estimação dos parâmetros por máxima verossimilhança é difícil devido a presença da função gama envolvendo o parâmetro α . Chandra, Roy e Ghosh apresentaram uma alternativa de estimação baseada em relações entre funções de probabilidade. Neste trabalho utilizaremos a função *nlm* do software R para realizar a estimação dos parâmetros. A seguir é apresentado alguns gráficos da distribuição.

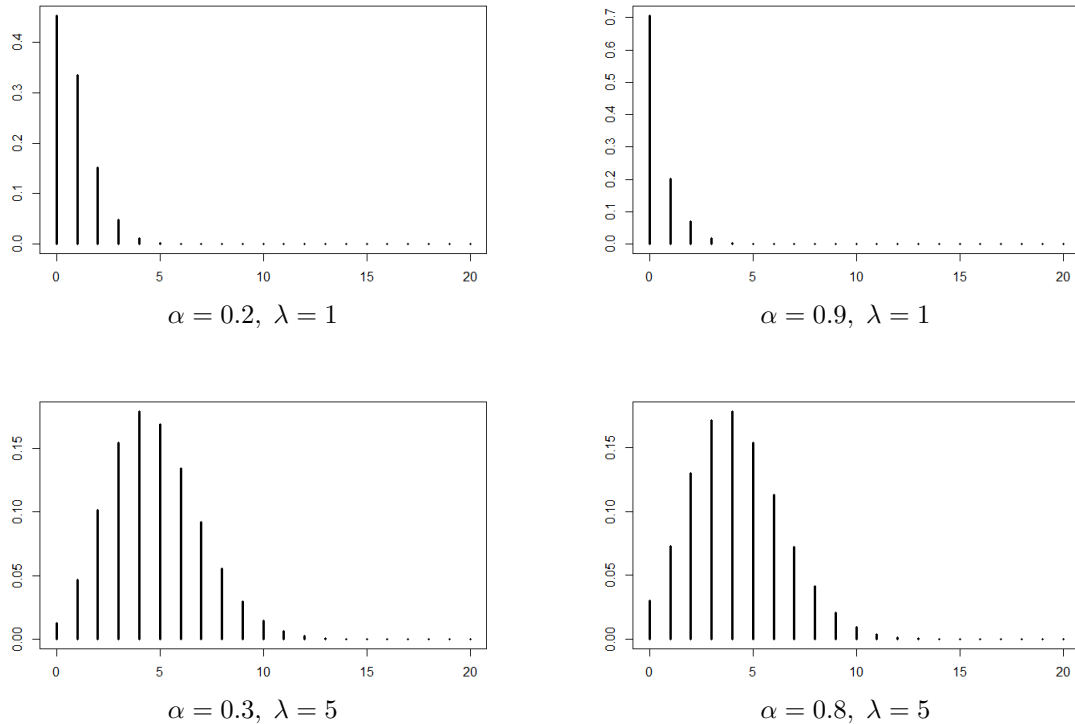


Figura 2.7: Gráficos da distribuição Poisson generalizada de Chandra e Roy.

Observando os gráficos da figura 2.7 notamos a influência dos parâmetros sobre o modelo. Quando a distribuição está mais próxima da origem, uma variação em α causa alteração significativa sobre a concentração de zeros no modelo. Entretanto, quando a distribuição está afastada da origem, uma variação em α pouco altera a forma do modelo. O parâmetro λ tem influência maior sobre a média do modelo. A medida que λ cresce a média é deslocada para valores maiores.

2.7 Distribuição de Poisson-Lindley

Em 1970, Sankaran propôs o modelo de Poisson-Lindley com a ideia de atribuir a distribuição de Lindley (1958) ao parâmetro λ da distribuição de Poisson. Seja λ uma variável aleatória com esta distribuição de Lindley, então λ possui a seguinte distribuição de probabilidade,

$$p(\lambda|\theta) = \frac{\theta^2}{\theta + 1}(\lambda + 1)e^{-\lambda\theta},$$

em que $\lambda, \theta > 0$. Da composição entre as distribuições obtemos o modelo de Poisson-Lindley. Seja X uma variável aleatória do modelo de Poisson-Lindley, sua distribuição de probabilidade é definida por

$$p(x|\theta) = \frac{\theta^2(\theta + 2 + x)}{(\theta + 1)^{x+3}}, \quad x = 0, 1, 2, \dots$$

A função geradora de momentos da distribuição pode ser escrita como

$$M_X(t) = \frac{\theta^2}{\theta + 1} \frac{2 + \theta - t}{(\theta + 1 - t)^2}.$$

E assim obtemos a esperança e a variância, dadas por

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

Seja uma amostra aleatória de n observações independentes da distribuição Poisson-Lindley, x_1, x_2, \dots, x_n , a função log-verossimilhança é dada por

$$l(\theta) = 2n \ln(\theta) - n(\bar{x} + 3) \ln(\theta + 1) + \sum_{i=1}^n \ln(x_i + \theta + 2).$$

Os estimadores de máxima verossimilhança são soluções que maximizam esta equação. Para encontrar o máximo da função utilizaremos a função `nlnmb()` do software R. A seguir, é apresentado gráficos da distribuição para alguns valores do parâmetro.

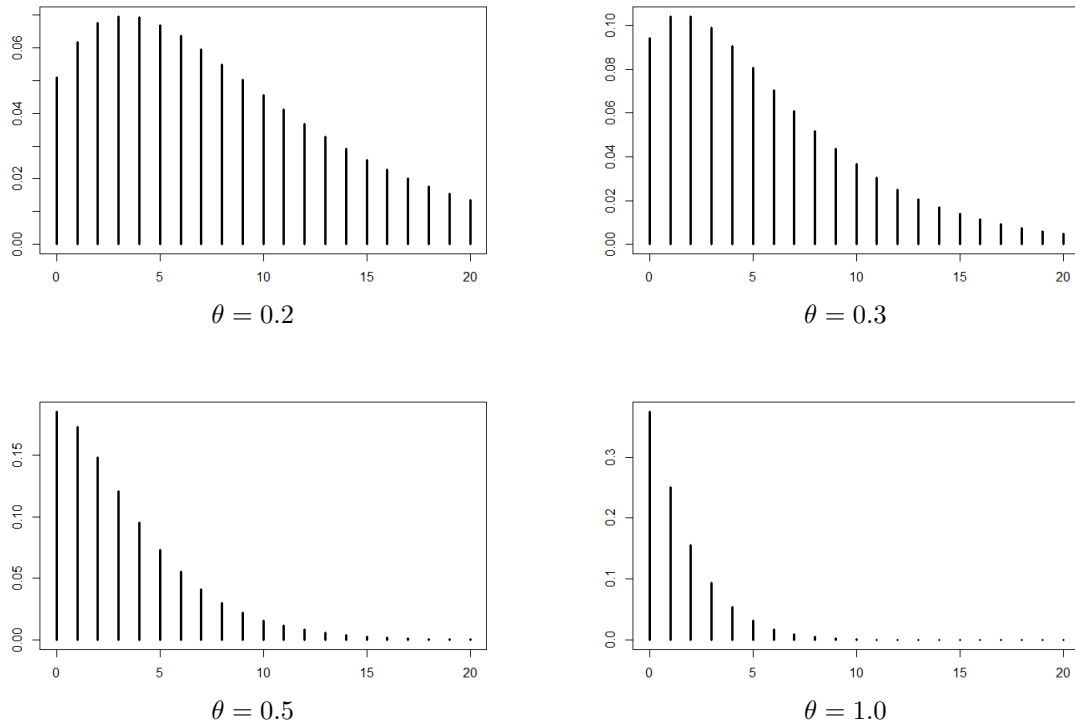


Figura 2.8: Gráficos da distribuição Poisson-Lindley.

Pelo fato do modelo ter apenas um parâmetro, o controle da média e da variância é limitada. O comportamento da distribuição é similar a da Poisson, quando a média cresce a variância também aumenta. Observando os gráficos, notamos que com o aumento de θ a distribuição tende a se concentrar no zero. Na medida que θ diminui, a média e a variância crescem.

2.8 Distribuição de Conway-Maxwell-Poisson

A distribuição foi introduzida por Conway e Maxwell em 1962. A distribuição possui um parâmetro extra, ν , que adiciona flexibilidade, e possui uma estrutura bem mais simples que outros modelos sugeridos na literatura. A distribuição tem bom desempenho em muitas situações, inclusive em casos de excesso de zeros. Seja X uma variável aleatória com distribuição COM-Poisson, sua função de probabilidade é

$$p(x|\lambda, \nu) = \frac{\lambda^x}{(x!)^\nu} \frac{1}{Z(\lambda, \nu)}, \quad x = 0, 1, 2, \dots,$$

em que $\lambda > 0$, $\nu \geq 0$ e $Z(\lambda, \nu) = \sum_{j=0}^{\infty} \lambda^j / (j!)^\nu$ é a constante normalizadora do modelo. Uma característica interessante da distribuição é o decaimento não linear da razão de probabilidades sucessivas, dada por

$$\frac{P(X = x - 1)}{P(X = x)} = \frac{x^\nu}{\lambda}.$$

Note que ν tem grande influência neste decaimento. Quando $\nu = 1$, a distribuição se simplifica na Poisson. Outros modelos são derivados da COM-Poisson: se $\nu \rightarrow \infty$ e $Z(\lambda, \nu) \rightarrow 1 + \lambda$ a distribuição tende a Bernoulli, e quando $\nu = 0$ e $\lambda < 1$, $Z(\lambda, \nu)$ é uma soma geométrica.

Os momentos da distribuição podem ser expressos usando a seguinte fórmula recursiva

$$E(X^{r+1}) = \begin{cases} \lambda E(X + 1)^{1-\nu} & r = 0, \\ \lambda \frac{\partial}{\partial \lambda} E(X^r) + E(X)E(X^r) & r > 0, \end{cases}$$

e assim, a esperança da distribuição é aproximadamente

$$E(X) \approx \lambda^{1/\nu} - \frac{\nu - 1}{2\nu}.$$

Seja uma amostra de n observações independentes da distribuição da COM-Poisson, x_1, x_2, \dots, x_n , a função de log-verossimilhança é dada por

$$l(\lambda, \nu) = \ln(\lambda) \sum_{i=1}^n x_i - \nu \sum_{i=1}^n \ln(x_i!) - n \ln(Z(\lambda, \nu)).$$

Para encontrar a solução ótima desta equação utilizaremos a função `nlminb()` do software R. A seguir segue alguns exemplos da distribuição de COM-Poisson.

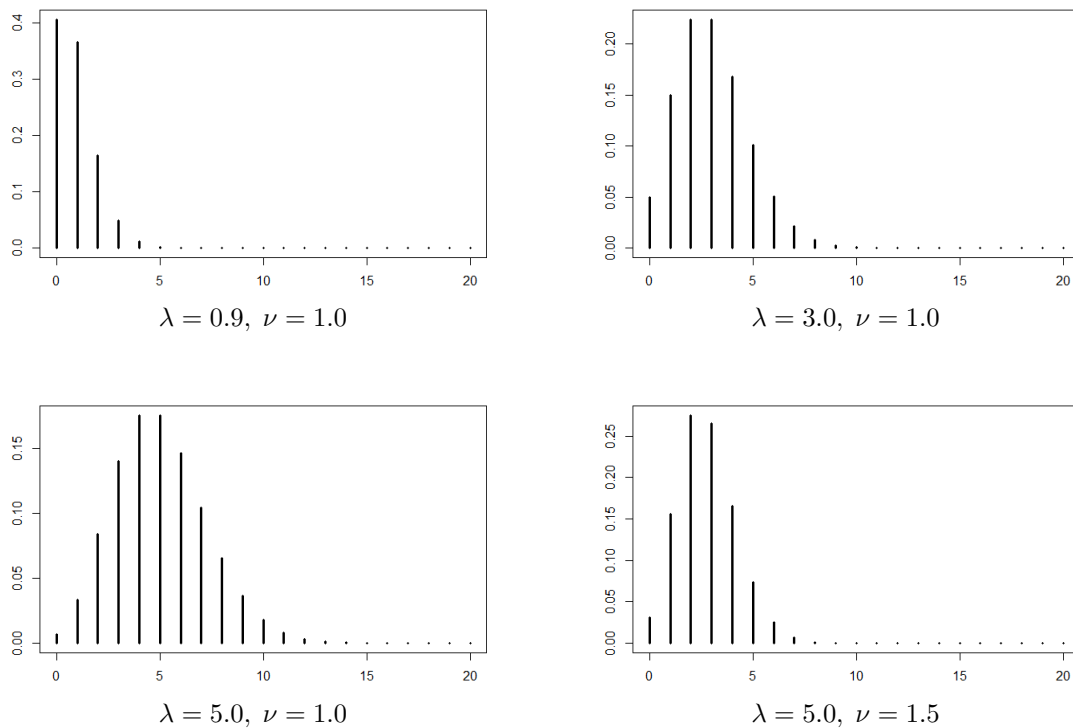


Figura 2.9: Gráficos da distribuição de Conway-Maxwell-Poisson.

Através dos gráficos da figura 2.9 notamos que o aumento que ocorre na média quando variamos λ . Sobre a variância do modelo, o parâmetro ν tem grande controle, já que deter-

mina a potência de x no decaimento das razões de probabilidade. Observando os gráficos, o aumento de ν , para $\lambda = 5$ fixo, faz com que as diferenças entre as probabilidades sucessivas aumente, proporcionando um modelo de menor dispersão.

Capítulo 3

A distribuição de Touchard

3.1 Introdução

Este capítulo inicia apresentando a distribuição de Touchard e alguma de suas propriedades, e segue definindo o teste da razão de verossimilhança e os critérios de informação que serão utilizados posteriormente no trabalho. A seção 3.2 introduz o modelo, com algumas representações gráficas. A seção 3.3 trata dos momentos da variável aleatória de Touchard e a seção 3.4 das estimativas por máxima verossimilhança dos parâmetros. A seção 3.5 apresenta o teste da razão de verossimilhança e a seção 3.6 os critérios de informação. Por último, a seção 3.7 apresenta os procedimentos computacionais utilizados para a estimativa dos parâmetros das distribuições.

3.2 Distribuição de probabilidade

Seja X uma variável aleatória, que assume somente valores inteiros não negativos, $k \in \mathbb{N}$, com distribuição de probabilidade definida por

$$p_k = P(X = k) = \frac{\lambda^k (k+1)^\delta}{k! \tau(\lambda, \delta)},$$

em que $\lambda > 0$ e $\delta \in \mathbb{R}$ são os parâmetros da distribuição e a função

$$\tau(\lambda, \delta) = \sum_{j \in \mathbb{N}} \frac{\lambda^j (j+1)^\delta}{j!}$$

normaliza a distribuição de probabilidade. A distribuição ganhou o nome do matemático francês Touchard, do século XIX, porque esta função está relacionada com os polinômios de Touchard. O modelo proposto pode ser considerado como uma generalização da distribuição de Poisson. Se $\delta = 0$, X segue uma distribuição de Poisson de parâmetro λ .

A figura 3.1 mostra exemplos da distribuição com $\lambda = 10$ e δ variando de -5.0 a -1.0.

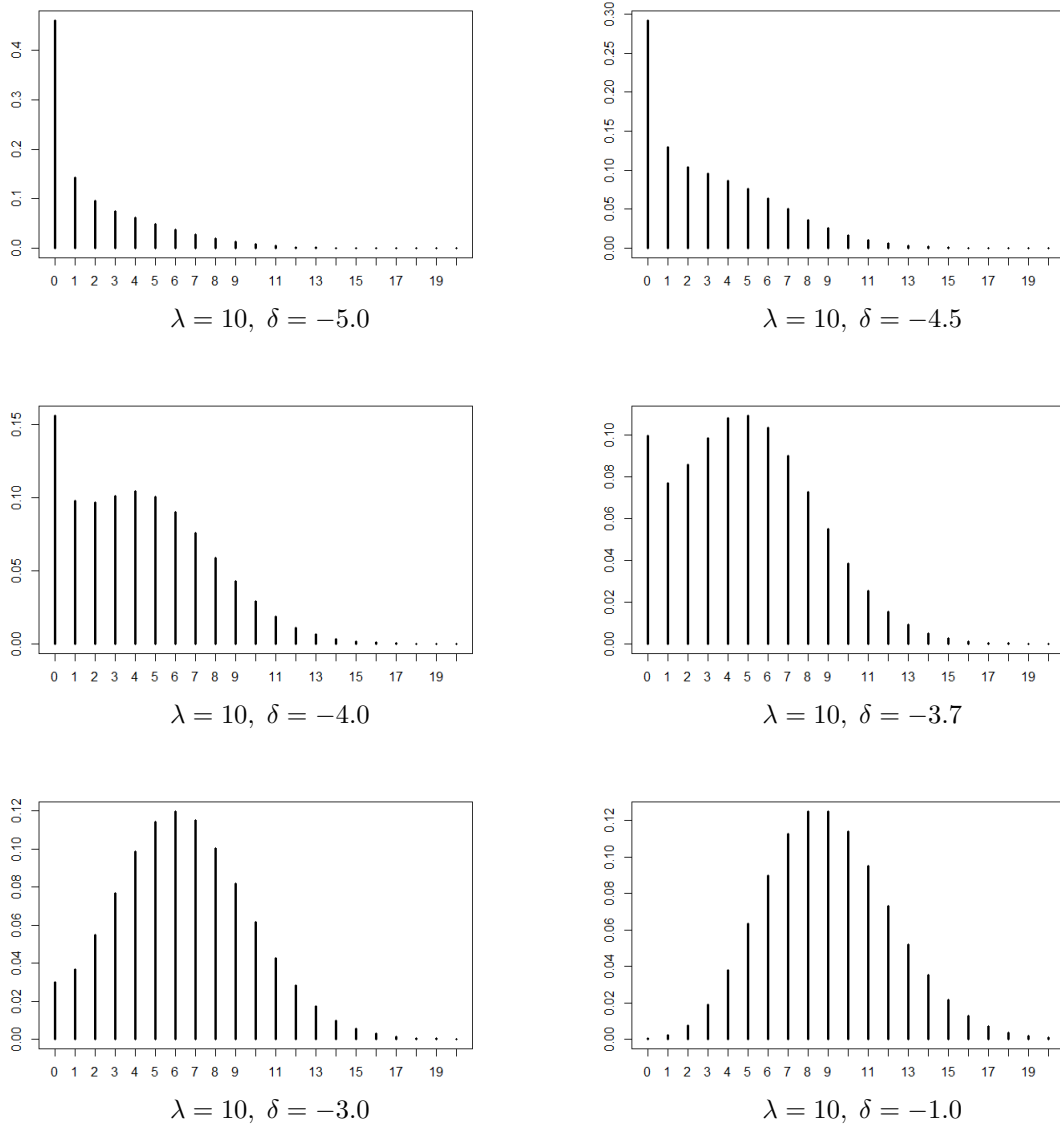


Figura 3.1: Gráficos da distribuição de Touchard, com $\lambda = 10$ e δ variando de -5.0 a -1.0.

Observando os gráficos da figura 3.1, notamos que a distribuição de Touchard tem a característica de acumular zeros de forma semelhante a ZIP, em que o modelo considera uma alta probabilidade no zero e tem uma curva mais suave para o restante do domínio. A vantagem da Touchard é que ela não precisa considerar o zero como um processo independente. Outra vantagem sobre a ZIP seria o fato da Touchard poder se ajustar a dados subdispersos, enquanto a ZIP é sempre um modelo superdisperso.

3.3 Momentos

O r -ésimo momento de uma variável aleatória com distribuição de Touchard, em que o termo com potência r é dado por uma expansão binomial, pode ser expresso da seguinte

forma

$$\begin{aligned} \mathbb{E}[X^r] &= \sum_{k=0}^{\infty} k^r \frac{\lambda^k (k+1)^\delta}{k! \tau(\lambda, \delta)} \\ &= \sum_{k=0}^{\infty} (k+1-1)^r \frac{\lambda^k (k+1)^\delta}{k! \tau(\lambda, \delta)} \\ &= \sum_{j=0}^r \binom{r}{j} \frac{(-1)^{r-j} \tau(\lambda, \delta + j)}{\tau(\lambda, \delta)}, \end{aligned}$$

com a função geratriz de momentos dada por

$$\begin{aligned} M_X(q) = \mathbb{E}[e^{qX}] &= \sum_{k=0}^{\infty} \frac{e^{qk} \lambda^k (k+1)^\delta}{k! \tau(\lambda, \delta)} \\ &= \frac{1}{\tau(\lambda, \delta)} \sum_{k=0}^{\infty} \frac{(\lambda e^q)^k (k+1)^\delta}{k!} \\ &= \frac{\tau(\lambda e^q, \delta)}{\tau(\lambda, \delta)}, \end{aligned}$$

em que $\delta \in \mathbb{R}$. Pela função geradora de momentos, calculamos a esperança de X , que é dada por

$$\begin{aligned} \mathbb{E}[X] &= \frac{\tau(\lambda, \delta + 1)}{\tau(\lambda, \delta)} - 1 \\ &= \lambda \cdot \mathbb{E} \left[\left(\frac{X+2}{X+1} \right)^\delta \right], \end{aligned}$$

e a variância pode ser escrita como

$$\begin{aligned} \sigma^2 = \text{Var}[X] &= \frac{\tau(\lambda, \delta + 2)}{\tau(\lambda, \delta)} - \left[\frac{\tau(\lambda, \delta + 1)}{\tau(\lambda, \delta)} \right]^2 \\ &= \lambda \cdot \mathbb{E} \left[(X+1) \left(\frac{X+2}{X+1} \right)^\delta \right] - \mu^2. \end{aligned}$$

Como $\mu > 0$, temos que $\mu > \lambda$ para $\delta > 0$ e $\mu < \lambda$ se $\delta < 0$. Para entendermos a relação média/variância, consideramos a razão $r = \sigma^2/\mu$, que pode ser expressa como

$$r = \frac{\mathbb{E} \left[(X+1) \left(\frac{X+2}{X+1} \right)^\delta \right]}{\mathbb{E} \left[\left(\frac{X+2}{X+1} \right)^\delta \right]} - \mu.$$

Para $\delta > 0$, como $(X+1)$ e $((X+2)/(X+1))^\delta$ são inversamente correlacionados, temos que $r < 1$. Se $\delta < 0$, então $r > 1$.

A função de distribuição de probabilidade da Touchard pode ser escrita de forma recursiva pela seguinte fórmula

$$p_{k+1} = \frac{\lambda}{k+1} \left(\frac{k+2}{k+1} \right)^\delta p_k,$$

em que $p_k = P(X = k)$. Com esta fórmula podemos entender em que condições a Touchard

possui a característica de acumular zeros, como apresentada na figura 3.1, em que ela tem alta probabilidade no zero, seguida de uma menor probabilidade em um, e depois cresce até certo valor. Os parâmetros da distribuição devem respeitar as seguintes inequações:

$$\frac{p_1}{p_0} = \lambda 2^\delta < 1,$$

$$\frac{p_{k+1}}{p_k} = \frac{\lambda}{k+1} \left(\frac{k+2}{k+1} \right)^\delta > 1.$$

3.4 Estatísticas suficientes e máxima verossimilhança

A distribuição de Touchard pertence a família exponencial já que sua densidade pode ser escrita como

$$p_k = \frac{1}{k!} \exp\{k \ln(\lambda) + \delta \ln(k+1) - \ln(\tau(\lambda, \delta))\}.$$

Dado uma amostra de n realizações independentes da distribuição, x_1, \dots, x_n , a verossimilhança resultante pode ser escrita como

$$L(\lambda, \delta | \{x_i\}) = \left(\prod_i x_i! \right)^{-1} \lambda^{S_1} e^{\delta S_2} [\tau(\lambda, \delta)]^{-n},$$

com $S_1 = \sum_i x_i$ e $S_2 = \sum_i \ln(x_i + 1)$, que de acordo com o teorema da fatorização são as estatísticas suficientes.

Consideramos a função de log-verossimilhança $l(\lambda, \delta) = \ln L(\lambda, \delta | \{x_i\})$

$$l(\lambda, \delta) = \left(\prod_i x_i! \right)^{-1} \lambda^{S_1} e^{\delta S_2} [\tau(\lambda, \delta)]^{-n},$$

e considerando as derivadas de $\tau(\lambda, \delta)$ com respeito a λ e δ

$$\begin{aligned} \frac{\partial \tau(\lambda, \delta)}{\partial \lambda} &= \frac{\partial}{\partial \lambda} \sum_{k=0}^{\infty} \frac{\lambda^k (k+1)^\delta}{k!} \\ &= \sum_{k=0}^{\infty} \frac{k \lambda^{k-1} (k+1)^\delta}{k!} \\ &= \frac{\tau(\lambda, \delta)}{\lambda} \sum_{k=0}^{\infty} k \frac{\lambda^k (k+1)^\delta}{k! \tau(\lambda, \delta)} \\ &= \frac{\tau(\lambda, \delta)}{\lambda} \mu, \end{aligned}$$

$$\begin{aligned}
\frac{\partial \tau(\lambda, \delta)}{\partial \delta} &= \frac{\partial}{\partial \delta} \sum_{k=0}^{\infty} \frac{\lambda^k (k+1)^\delta}{k!} \\
&= \sum_{k=0}^{\infty} \frac{\lambda^k e^{\delta \ln(k+1)} \ln(k+1)}{k!} \\
&= \tau(\lambda, \delta) \sum_{k=0}^{\infty} \ln(k+1) \frac{\lambda^k (k+1)^\delta}{k! \tau(\lambda, \delta)} \\
&= \tau(\lambda, \delta) E(\ln(X+1)),
\end{aligned}$$

derivamos $l(\lambda, \delta)$ em relação a λ e δ , e obtemos as equações de máxima verossimilhança,

$$\begin{cases} S_1 - n\mu = 0, \\ S_2 - n E\{\ln[X+1]\} = 0. \end{cases}$$

As estimativas de λ e δ que satisfazem estas equações correspondem com as estimativas de máxima verossimilhança. As equações podem ser resolvidas pelo método de Newton-Raphson.

3.5 Teste da razão de verossimilhança

Para comparar dois modelos estimados de diferentes amostras utilizaremos o teste da razão de verossimilhança, com a estatística de teste

$$T = -2 \ln \left(\frac{L_1 L_2}{L} \right),$$

em que $L_1 = \max_{\tilde{\theta}_1} L(\tilde{\theta}_1)$ e $L_2 = \max_{\tilde{\theta}_2} L(\tilde{\theta}_2)$. A verossimilhança L é obtida do conjunto das duas amostras, em que $L = \max_{\tilde{\theta}} L(\tilde{\theta})$.

Pelo teorema de Wilks a estatística T tem distribuição assintótica qui-quadrado, com a diferença do número de parâmetros entre as distribuições como grau de liberdade. Para maiores detalhes sobre o teorema, ver Casella e Berger (2001).

Para análise dos nossos resultados, iremos comparar duas distribuições de Touchard. Então as hipóteses do teste podem ser dadas da seguinte maneira

$$\begin{cases} H_0 : \lambda_1 = \lambda_2 \text{ e } \delta_1 = \delta_2 \\ H_A : \lambda_1 \neq \lambda_2 \text{ ou } \delta_1 \neq \delta_2 \end{cases}$$

considerando $\text{Touchard}(\lambda_1, \delta_1)$ e $\text{Touchard}(\lambda_2, \delta_2)$.

A decisão será tomada com um nível de confiança de 5%, em uma distribuição qui-quadrado com dois graus de liberdade, que é a diferença do número de parâmetros utilizados na verossimilhança do numerador e o denominador da razão presente na estatística do teste.

3.6 Critérios de informação

A escolha do modelo apropriado será feita pelo critério de informação de Akaike (AIC) e o critério de informação bayesiano (BIC), que se baseiam no máximo da função de verossimilhança. Ambos critérios buscam modelos mais parcimoniosos e que expliquem bem o comportamento dos dados.

O valor AIC de um modelo é dado por

$$AIC = 2k - 2 \ln(L),$$

e o BIC por

$$BIC = k \ln(n) - 2 \ln(L),$$

em que L é o máximo da função de verossimilhança e k é o número de parâmetros do modelo. Dado um grupo de modelos candidatos para um conjunto de dados, o modelo preferido será aquele de menor AIC e BIC. Note que, se $\ln(n) > 2$, o BIC penaliza mais rigorosamente os parâmetros extras do que o AIC.

3.7 Procedimentos para estimação dos parâmetros

Esta seção apresentará um detalhamento dos procedimentos computacionais que foram feitos no trabalho e outras alternativas que poderiam resolver o problema de otimização. Os códigos estão em anexo, com todas as programações utilizadas para estimação de parâmetros que foram feitas no R.

O processo de estimação dos parâmetros variou pra cada modelo de probabilidade. Para a distribuição Poisson generalizada de Consul e Jain foi utilizada a função *generalized_poisson_likelihood()* do pacote *GPseq*, que encontra os estimadores de máxima verossimilhança da distribuição.

Para as distribuições de Touchard, hiper-Poisson e binomial negativa foi implementado o método de Newton-Raphson, que funciona de maneira iterativa, através do cálculo da matriz Hessiana. Para determinar o valor inicial dos parâmetros foi testado diversos valores dentro de um intervalo arbitrário até que fosse alcançado a convergência do algoritmo. Como não temos garantia que os parâmetros encontrados representariam o ótimo global da superfície de verossimilhança, tentamos atingir a convergência partindo de diferentes pontos iniciais. Na maioria dos casos as estimativas encontradas coincidiam, então estes parâmetros foram considerados os estimadores de máxima verossimilhança. Os cálculos de funções sem forma fechada, como a função $\tau(\lambda, \delta)$ da distribuição de Touchard, foram feitas recursivamente.

Para a distribuição da Poisson inflada com zeros a estimação foi realizada via algoritmo

EM. O método é útil quando não temos a informação do processo que gerou a observação.

As distribuições de Conway-Maxwell-Poisson, Poisson-Lindley e Poisson generalizada de Chandra e Roy tiveram os parâmetros estimados pela função *nlminb()* do software R. O algoritmo maximiza uma função de entrada, e apesar de ser considerada ultrapassada, obteve tempos de execução satisfatórios. Para resolver o problema de restrição do domínio dos parâmetros foi feito uma reparametrização. Outras funções como *optim()* e *constrOptim()* são mais modernas e poderiam ser utilizadas para aperfeiçoar o procedimento de otimização.

Capítulo 4

Resultados

Utilizaremos o banco de dados de acidentes de veículos motorizados do estado de New York para analisar o comportamento das distribuições. Os dados foram obtidos pelo website oficial do estado, para acidentes registrados nos anos de 2011, 2012 e 2013. Cada observação é representada por um acidente, com informações como o horário do acidente, condado em que ocorreu, condição do tempo. A variável de contagem será o número de acidentes por dia, dada certas restrições.

Consideramos o condado de Nova York, o terceiro mais populoso do estado de acordo com o censo americano de 2010. Dividimos os acidentes pelo horário de sua ocorrência, considerando os intervalos de 00:00 às 00:10, 12:00 às 12:10 e 18:00 às 18:10, para os anos de 2011, 2012 e 2013. O pequeno intervalo de tempo proporciona uma quantidade menor de acidentes registrados por dia, e conseqüentemente maior proporção de zeros.

Queremos avaliar se a distribuição de Touchard é adequada nas condições propostas e descobrir se ao decorrer dos anos houve uma alteração no comportamento do número de acidentes no estado de New York. Para fazer a avaliação entre duas distribuições, utilizaremos o teste de razão de verossimilhança.

A tabela a seguir apresenta a qualidade do ajuste das distribuições para os diferentes casos mencionados, no ano de 2011.

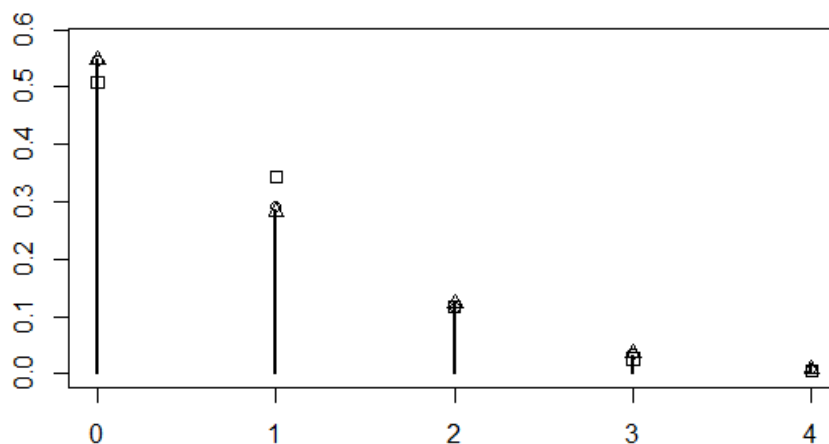
Tabela 4.1: Comparação entre as distribuições, em diferentes horários no condado de New York, em 2011.

	00:00 ~ 00:10		12:00 ~ 12:10		18:00 ~ 18:10	
Modelos	AIC	BIC	AIC	BIC	AIC	BIC
Touchard	698.80	706.28	587.09	594.58	699.13	706.62
Poisson	703.00	706.75	587.98	591.73	697.73	701.48
Binomial Negativa	699.58	707.06	586.67	594.16	699.28	706.77
Poisson Generalizada	699.69	707.18	586.65	594.13	699.29	706.78
Hiper-Poisson	698.84	706.33	587.06	594.55	699.13	706.61
ZIP	696.58	706.06	587.92	595.40	699.12	706.60
PG - Chandra e Roy	698.57	706.05	587.63	595.12	699.11	706.60
Poisson-Lindley	704.80	708.54	598.94	593.68	716.04	719.78
COM-Poisson	699.15	706.64	586.85	594.33	699.18	706.67

Segue abaixo a estimativa dos parâmetros da distribuição de Touchard por máxima verossimilhança, para cada caso.

Tabela 4.2: Estimativas de máxima verossimilhança dos parâmetros da distribuição de Touchard.

	00:00 ~ 00:10	12:00 ~ 12:10	18:00 ~ 18:10
$\hat{\lambda}$	1.5802102	1.402978	1.989452
$\hat{\delta}$	-0.8305442	-0.2096629	-0.6051284

**Figura 4.1:** Comparação entre o observado e o ajuste dos modelos do número de acidentes de veículos motorizados por dia, no horário de 00:00 às 00:10, no estado de New York em 2011 (\circ = Touchard, \square = Poisson, \triangle = ZIP).

A figura 4.8 mostra que das 00:00 às 00:10 os melhores ajustes foram da distribuição ZIP,

da Poisson generalizada de Chandra e Roy e da Touchard. De 12:00 às 12:10, a distribuição binomial negativa e a Poisson generalizada tiveram os menores AIC. A distribuição de Poisson também obteve bom resultado, com o menor BIC. O BIC é mais rigoroso com os modelos multiparamétricos se a amostra é muito grande, dificultando o bom resultado destes modelos. Para o horário de 18:00 às 18:10, a distribuição de Poisson obteve os melhores resultados, indicando a desnecessidade da modelagem por distribuições com parâmetros extras.

Considerando o modelo de Touchard estimado para cada caso, temos que a média de acidentes estimada é de 0.68 por dia no horário de 00:00 às 00:10, de 0.49 acidentes por dia de 12:00 às 12:10, e de 0.70 de 18:00 às 18:10.

Segue abaixo as medidas de qualidade do ajuste das distribuições para o ano de 2012.

Tabela 4.3: Comparação entre as distribuições, em diferentes horários no condado de New York, em 2012.

	00:00 ~ 00:10		12:00 ~ 12:10		18:00 ~ 18:10	
Modelos	AIC	BIC	AIC	BIC	AIC	BIC
Touchard	1025.98	1033.38	598.67	606.17	700.30	707.80
Poisson	1042.07	1045.82	601.94	605.69	700.30	704.10
Binomial Negativa	1032.01	1039.51	599.07	606.57	700.20	707.70
Poisson Generalizada	1032.54	1040.03	599.15	606.65	700.21	707.71
Hiper-Poisson	1026.43	1033.93	598.70	606.20	700.31	707.81
ZIP	1020.87	1028.37	598.80	606.30	707.71	708.21
PG - Chandra e Roy	1023.17	1030.67	598.69	606.19	700.53	708.03
Poisson-Lindley	1054.10	1057.85	600.48	604.23	711.62	715.37
COM-Poisson	1028.92	1036.42	598.86	606.36	700.22	707.72

Segue abaixo a estimativa dos parâmetros da distribuição de Touchard por máxima verossimilhança, para cada caso.

Tabela 4.4: Estimativas de máxima verossimilhança dos parâmetros da distribuição de Touchard.

	00:00 ~ 00:10	12:00 ~ 12:10	18:00 ~ 18:10
$\hat{\lambda}$	2.754923	1.150503	1.027696
$\hat{\delta}$	-1.501969	-1.505677	-0.7843778

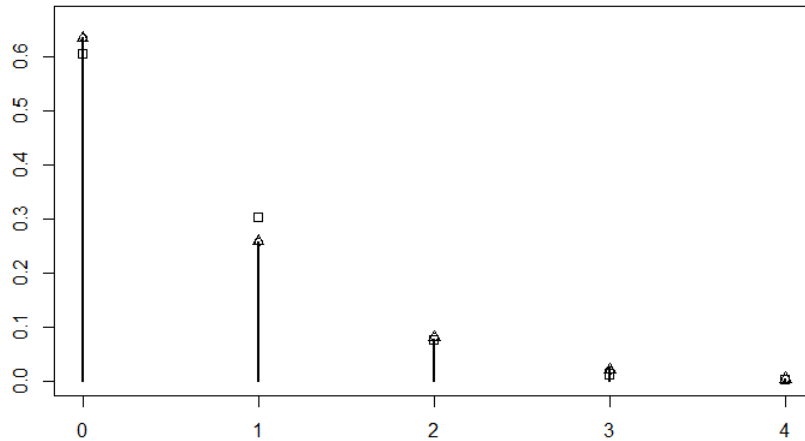


Figura 4.2: Comparação entre o observado e o ajuste dos modelos do número de acidentes de veículos motorizados por dia, no horário de 12:00 às 12:10, no estado de New York em 2012 (○ = Touchard, □ = Poisson, △ = PG - Chandra e Roy).

Em 2012, no horário de 00:00 às 00:10, as distribuições com melhor desempenho foram a ZIP, a Poisson generalizada de Chandra e Roy e a Touchard, o que também ocorreu na análise para o ano de 2011. O modelo de Touchard estima uma média de 1.53 acidentes por dia no ano de 2012. O teste de razão de verossimilhança foi feito para avaliar estatisticamente a diferença entre os modelos de Touchard estimados em ambos os anos.

Tabela 4.5: Teste da razão de verossimilhança para os modelos Touchard estimados nos anos de 2011 e 2012, no horário de 00:00 às 00:10.

T	p-valor	Conclusão
77.55	$< 10^{-15}$	Rejeitar a hipótese nula

Então, rejeitamos a hipótese nula de que os modelos são estatisticamente iguais com uma estatística de 77.55, com nível de confiança de 5%. No ano de 2012, no horário de 00:00 às 00:10, a média foi muito mais alta do que em 2011, com uma estimativa de 1.53 acidentes por dia.

De 12:00 às 12:10, a distribuição de Touchard foi o modelo com menor AIC, seguido da distribuição de Poisson generalizada de Chandra e Roy. A figura 4.2 mostra a similaridade da estimação das duas distribuições. Para verificar a diferença estatística dos modelos nos diferentes anos, um novo teste foi realizado.

Tabela 4.6: Teste da razão de verossimilhança para os modelos Touchard estimados nos anos de 2011 e 2012, no horário de 12:00 às 12:10.

T	p-valor	Conclusão
0.17	0.92	Não rejeitar a hipótese nula

Com uma estatística de 0.17, não rejeitamos a hipótese nula de igualdade dos modelos, concluindo que não houve mudança no comportamento dos acidentes neste horário. Para o ano de 2012, a média estimada foi de 0.50 acidentes por dia.

Para os acidentes ocorridos no período de 18:00 às 18:10, as distribuições tiveram desempenho similar a da Poisson. O modelo de Poisson obteve o menor BIC menor, indicando que um modelo mais simples já seria suficiente para explicar esses dados. Uma conclusão semelhante a esta ocorreu para o ano de 2011. Será feito o teste de razão de verossimilhança para as distribuições de Touchard estimadas nos anos de 2011 e 2012, no horário de 18:00 às 18:10.

Tabela 4.7: Teste da razão de verossimilhança para os modelos Touchard estimados nos anos de 2011 e 2012, no horário de 18:00 às 18:10.

T	p-valor	Conclusão
0.27	0.87	Não rejeitar a hipótese nula

Com uma estatística de 0.27, não rejeitamos a hipótese nula de que as distribuições são iguais. Em 2012, a média estimada de acidentes por dia foi de 0.68, e de 0.70 no ano de 2011.

Notamos uma mudança de comportamento no número de acidentes apenas no horário de 00:00 às 00:10. No período do início da tarde e da noite, os resultados foram muito similares. Agora, as análises serão feitas considerando o ano de 2013. Segue abaixo a tabela com as comparações do desempenho dos modelos.

Tabela 4.8: Comparação entre as distribuições, em diferentes horários no condado de New York, em 2013.

	00:00 ~ 00:10		12:00 ~ 12:10		18:00 ~ 18:10	
Modelos	AIC	BIC	AIC	BIC	AIC	BIC
Touchard	1038.14	1045.64	593.94	601.43	658.40	665.90
Poisson	1036.70	1040.45	592.38	596.12	656.42	660.17
Binomial Negativa	1038.89	1046.38	594.48	665.94	658.45	665.94
Poisson Generalizada	1037.24	1044.73	593.72	601.21	658.40	665.90
Hiper-Poisson	1038.30	1045.79	593.97	601.46	658.40	665.89
ZIP	1038.70	1046.20	594.38	601.87	658.42	665.91
PG - Chandra e Roy	1038.70	1046.20	594.38	601.87	658.42	665.91
Poisson-Lindley	1142.60	1146.35	613.43	617.17	678.87	682.61
COM-Poisson	1037.81	1045.31	593.88	601.37	658.40	665.90

Segue abaixo a estimativa dos parâmetros da distribuição de Touchard por máxima verossimilhança, para cada caso.

Tabela 4.9: Estimativas de máxima verossimilhança dos parâmetros da distribuição de Touchard.

	00:00 ~ 00:10	12:00 ~ 12:10	18:00 ~ 18:10
$\hat{\lambda}$	1.726151	0.400085	0.6021046
$\hat{\delta}$	0.2856446	0.4753412	0.08175635

De acordo com a tabela, a distribuição de Poisson teve o melhor desempenho em todos os períodos. O modelo de Touchard teve resultado semelhante a das outras distribuições. Faremos os testes de verossimilhança comparando os modelos estimados de Touchard entre os anos de 2012 e 2013. Segue abaixo o teste no horário de 00:00 às 00:10.

Tabela 4.10: Teste da razão de verossimilhança para os modelos Touchard estimados nos anos de 2012 e 2013, no horário de 00:00 às 00:10.

T	p-valor	Conclusão
23.86	$< 10^{-5}$	Rejeitar a hipótese nula

O teste rejeitou a hipótese nula de igualdade entre os modelos com uma estatística de 23.86. Novamente, no horário da madrugada, o modelo estimado do ano anterior foi considerado para o ano seguinte. O modelo estimado de Touchard possui média de 1.91, a maior média de acidentes por dia em todos os casos considerados no condado de New York.

Tabela 4.11: Teste da razão de verossimilhança para os modelos Touchard estimados nos anos de 2012 e 2013, no horário de 12:00 às 12:10.

T	p-valor	Conclusão
4.47	0.11	Não rejeitar a hipótese nula

No horário de 12:00 às 12:10, com nível de confiança de 5%, o teste não rejeitou a hipótese nula de igualdade entre modelos, o que ocorreu também no teste entre os anos de 2011 e 2012. O modelo de Touchard para o ano de 2013 tem média 0.52 de acidentes por dia.

Tabela 4.12: Teste da razão de verossimilhança para os modelos Touchard estimados dos anos de 2012 e 2013, no horário de 18:00 às 18:10.

T	p-valor	Conclusão
1.73	0.42	Não rejeitar a hipótese nula

Para o período de 18:00 às 18:10, não foi rejeitado a hipótese nula de igualdade entre os modelos estimados. A distribuição de Touchard estimada teve média de 0.63, semelhante a dos anos de 2011 e 2012.

A distribuição de Touchard obteve bons resultados em geral, sendo o modelo mais adequado em 2012, no horário de 12:00 às 12:10, por exemplo.

Notamos que em muitos casos, principalmente no período de 18:00 às 18:10, os valores de AIC e BIC da distribuição de Poisson foram os mais baixos, indicando a desnecessidade de um modelo mais complexo para explicar esses dados.

No período da madrugada, seguindo as estimativas do modelo de Touchard, houve um aumento significativo na média de acidentes por dia ao passar dos anos. A média estimada foi de 0.68 em 2011, de 1.53 em 2012, chegando a 1.91 em 2013. Um aumento muito expressivo, considerando o breve intervalo de tempo utilizado nas análises.

Nos horários da tarde e da noite, os testes de razão de verossimilhança apontaram indiferença entre as distribuições estimadas, com a média dos modelos variando pouco ao passar dos anos.

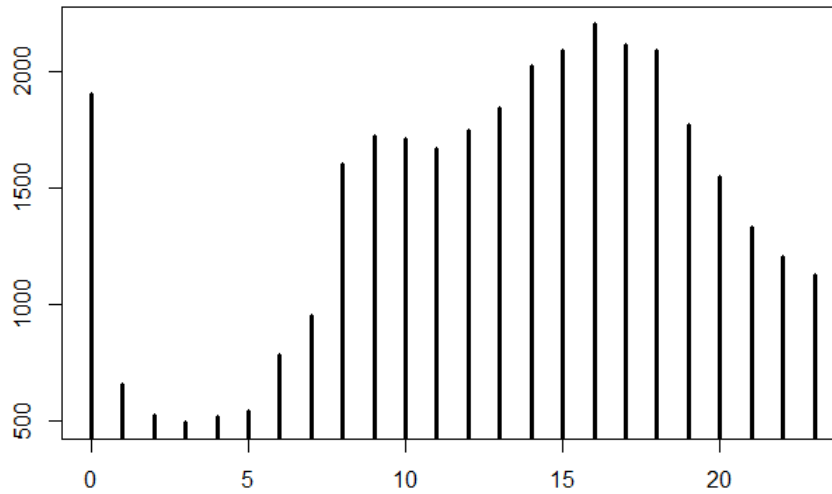


Figura 4.3: Gráfico do número total de acidentes registrados dividido em horas do dia, nos anos de 2011 a 2013.

O gráfico 4.3 mostra o número total de acidentes ocorridos nos anos de 2011 a 2013, separados em intervalos de uma hora. Como esperado, em horários de alto trânsito de veículos o número de acidentes tende a ser maior. Uma observação a se destacar é o pico que ocorre no período de 00:00 às 00:59, o que pode ser resultado de uma imprecisão do registro de horários. Somente no horário exato de 00:00 foram registrados 1101 acidentes nos três anos.

Agora, analisaremos o comportamento do número de acidentes em diferentes janelas de tempo. Consideramos o horário de 12:00 às 12:10, depois de 12:00 às 12:20, de 12:00 às 12:30 e de 12:00 às 13:00, no ano de 2012.

Segue abaixo a tabela com as medidas de ajuste para os horários considerados.

Tabela 4.13: Comparação entre as distribuições, em diferentes intervalos de tempo no condado de New York, em 2012.

Modelos	12:00 ~ 12:10		12:00 ~ 12:20		12:00 ~ 12:30		12:00 ~ 13:00	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
Touchard	598.67	606.17	761.77	769.27	884.27	891.77	1130.55	1138.05
Poisson	601.94	605.69	764.05	767.80	883.91	887.66	1132.01	1135.76
Binomial Negativa	599.07	606.57	762.78	770.28	884.19	891.68	1130.82	1138.32
Poisson Generalizada	599.15	606.65	762.03	769.53	883.51	891.01	1130.71	1138.21
Hiper-Poisson	598.70	606.20	761.77	769.27	884.36	891.86	1130.57	1138.07
ZIP	598.80	606.30	761.74	769.23	885.33	892.82	1131.40	1138.90
GP - Chandra e Roy	598.69	606.19	761.68	769.17	884.86	892.36	1130.71	1138.21
Poisson-Lindley	600.48	604.23	773.82	777.57	915.25	919.00	1203.43	1207.18
COM-Poisson	598.86	606.36	761.83	769.33	883.90	891.40	1130.49	1137.99

Quando consideramos períodos de tempos mais longos, o número de acidentes registrados por dia é maior, o que proporciona também maiores valores dos critérios de informação, como visto na tabela 4.13.

Queremos entender como os resultados dos modelos se comportam nos diferentes cenários considerados. Por exemplo, a distribuição de Poisson obteve seu melhor desempenho no intervalo de 12:00 às 12:30, em que obteve o terceiro menor AIC, atrás apenas da Poisson Generalizada e da COM-Poisson, e o menor BIC. Para os casos em que a proporção de zeros tende a ser maior e também no intervalo de 12:00 às 13:00, que possui o menor acúmulo de zeros, a Poisson não manteve o bom desempenho de acordo com os valores de AIC.

A distribuição de Touchard teve bons resultados em geral, sendo a mais adequada para o caso com maior excesso de zeros, de acordo com o valor de AIC. No período de 12:00 às 13:00, somente a COM-Poisson obteve resultado melhor que a Touchard. A distribuição de COM-Poisson também obteve bons resultados, sendo competitiva em todos os períodos.

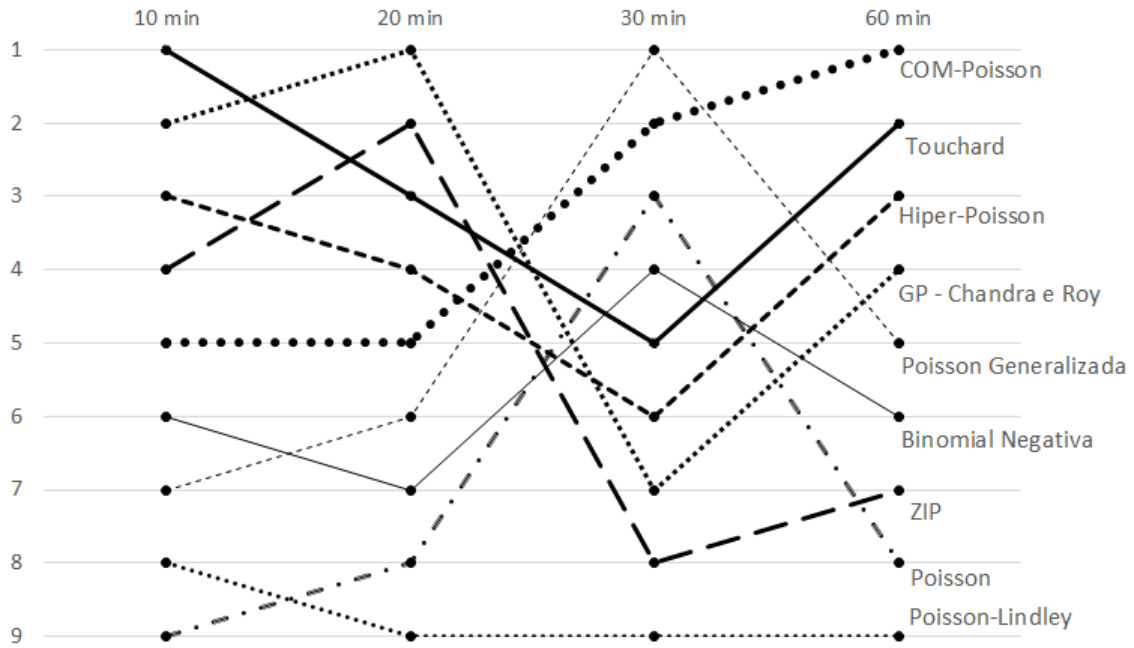


Figura 4.4: Classificação das distribuições baseadas no AIC da tabela 4.13.

A figura 4.4 apresenta a classificação das distribuições pra cada intervalo, baseada no AIC da tabela anterior. Com esta representação gráfica podemos identificar similaridades no comportamento que ocorreram entre algumas distribuições. Por exemplo, a ZIP e a Poisson generalizada de Chandra e Roy tiveram boas colocações para o intervalo de 10 e 20 minutos, mas caíram de posição para os outros intervalos. Outra observação é a mudança de comportamento que acontece do intervalo de 30 minutos para o de 60 minutos. As distribuições de Poisson e binomial negativa caem de posição, enquanto as distribuições de Touchard, COM-Poisson e hiper-Poisson melhoram a classificação. Um fator que pode ter influenciado essa mudança é que quase um terço dos acidentes foram registrados no horário exato de 12:00 ou 13:00. Então, no intervalo de 60 minutos os acidentes registrados no horário de 13:00 são incluídos, o que pode ter alterado o comportamento da distribuição do número de acidentes pelo entrada de um alto número de observações na amostra.

Capítulo 5

Conclusão

O trabalho teve início com o estudo da distribuição de Touchard, e em seguida, um levantamento de modelos de contagem não Poisson foi realizado, com objetivo de entender as motivações e as características dos modelos.

Utilizamos o software R para implementar a estimação dos parâmetros de cada distribuição. Em alguns casos, utilizamos rotinas já criadas, e para outras foi criada algoritmos próprios. Parte dos algoritmos foi simplificada pelo uso da função *nlminb()* do software R. Também foi necessário elaborar uma função para a leitura da base de dados.

Comparamos o desempenho das distribuições através dos critérios de informação de Akaike e bayesiano em diversas situações. A distribuição de Touchard obteve bons resultados, mostrando ser um modelo competitivo, sendo capaz de propor boas estimações tanto com excesso de zeros como também em casos com valores de contagem mais altos. A partir destes resultados, uma análise simples foi feita sobre o número de acidentes de veículos motorizados no condado de New York, considerando diferentes horários em diferentes anos. Para obter resultados mais completos sobre o comportamento dos acidentes seria necessário um estudo mais específico.

O trabalho proporcionou a prática de conceitos aprendidos ao longo do curso, como dos estimadores de máxima verossimilhança e o teste da razão de verossimilhança, e o uso de um ambiente computacional como o R para o manuseio de uma grande base de dados e a implementação de métodos iterativos de otimização. Foi proveitoso também para o aprendizado de novas distribuições, tanto de propostas publicadas recentemente quanto de modelos estudados anos atrás.

Referências Bibliográficas

BARDWELL, G. E.; EDWIN, L. C. A two-parameter family of hiper-Poisson distributions. *Journal of the American Statistical Association*, v. 59, n. 305, p. 133-141, 1964.

CASELLA, G.; BERGER, R. L. *Statistical Inference*. 2. ed. Pacific Grove: Duxbury Press, 2001, 660 p.

CHANDRA, N. K.; ROY, D.; GHOSH, T. A generalized Poisson distribution. *Communications in Statistics - Theory and Methods*, v. 42, n. 15, p. 2786-2797, 2013.

CHRYSAPHINO, O. On Touchard polynomials. *Discrete Mathematics and Applications*, v. 54, p. 143-152, 1985.

COHEN, A.C. Estimation in mixtures of discrete distributions. In: PROCEEDINGS OF THE INTERNATIONAL SYMPOSIUM ON DISCRETE DISTRIBUTIONS. Montreal, p. 373-378, 1963.

CONSUL, P. C.; JAIN, G. C. A generalization of the Poisson distribution. *Technometrics*, v. 15, n. 4, p. 791-799, 1973.

CONWAY, R. W.; MAXWELL, W. L. A queuing model with state dependent services rates. *Journal of Industrial Engineering*, v. 12, p. 132-136, 1962.

EFRON, B. Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association*, v. 81, p. 709-721, 1989.

JOHNSON, N.L.; KEMP, A. W.; KOTZ, S. *Univariate Discrete Distributions*. 3. ed. Hoboken: Wiley, 2005, 672 p.

KEMP, C. D. q-analogues of the hyper-Poisson distribution. *Journal of Statistical Planning and Inference*, v. 101, p. 179-183, 2002.

KUMAR, C. S.; NAIR, B. U. A modified version of hyper-Poisson distribution and its applications. *Journal of Statistics and Applications*, v. 6, p. 25-36, 2011.

LAMBERT, D. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, v. 34, p. 1-14, 1992.

LINDLEY, D. V. Fiducial distributions and Bayes' theorem. *Journal of the Royal Statistical Society*, v. 20, n. 1, p. 102-107, 1958.

NISIDA, T. On the multiple exponential channel queuing system with hyper-Poisson arrivals. *Journal of the Operations Research Society*, v. 5, p. 57-66, 1962.

R CORE TEAM. R: A language and environment for statistical computing. 2015. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Disponível em: <<http://www.R-project.org>>. Acesso em: 12 dez. 2016.

ROOHI, A.; AHMAD, M. Estimation of the parameters of hyper-Poisson distribution using negative moments. *Pakistan Journal of Statistics*, v. 19, p. 99-105, 2003.

RSTUDIO. Disponível em: <<http://rstudio.org>>. Acesso em: 13 nov. 2016.

SANKARAN, M. The discrete Poisson-Lindley distribution. *Biometrics*, v. 26, p. 145-149, 1970.

TOUCHARD, J. Sur les cycles des substitutions. *Acta Mathematica*, v. 70, p. 243-279, 1939.

TRIPATHI, R. C.; GUPTA, R. C. Statistical inference regarding the generalized Poisson distribution. *The Indian Journal of Statistics*, v. 46, p. 166-173, 1984.

Anexo

```

EMVBN <- function(x,error=1e-6,maxiter=1000,p0,n0){
  maxerror <- 1
  par <- matrix(c(p0,n0),nrow=2,ncol=1,byrow=TRUE)
  j <- 0
  while(maxerror>error){
    n <- length(x)
    xbar <- mean(x)
    H11 <- -(n*par[2]/(par[1]^2))-(1/((1-par[1])^2))*sum(x)
    H22 <- -n*trigamma(par[2])+sum(trigamma(par[2]+x))
    H12 <- (n/((n*par[2])/(n*par[2]+sum(x))))
    constant <- 1/(H12^2-H11*H22)
    HessianI <- constant*matrix(c(-H22,H12,H12,-H11),nrow=2,ncol=2,byrow=TRUE)
    G1 <- ((par[2])/((par[2]*n)+(sum(x))))-(par[1]/n)
    G2 <- log(((par[2]*n)/((par[2]*n)+(sum(x))))) - digamma(par[2])
      +(sum(digamma(par[2]+x)))/n
    Gradient <- matrix(c(G1,G2),nrow=2,ncol=1,byrow=TRUE)
    Incr <- HessianI %*% Gradient
    par <- par - Incr
    maxerror <- max(abs(Incr))
    j <- j+1
    var1 <- -HessianI[1,1]/n
    var2 <- -HessianI[2,2]/n
    if (j == maxiter) break
  }
  s <- -1*(trunc(par[2]))
  like <- n*par[2]*log(par[1])-n*log(gamma(par[2]))+sum(log(gamma(par[2]+x)))
    +sum(x*(log(1-par[1]))) - sum(log(gamma(x+1)))
  AIC <- 2*2-2*like
  BIC <- 2*log(length(x))-2*like
  result <- c(par[1],par[2],var1,var2,j,like,AIC,BIC)
  return(result)
}

SCTD2 <- function(lambda,theta,errorM=1e-6){
  error <- 1
  E <- 0
  a0 <- 0
  k <- 0
  while(abs(error) > errorM){
    a0 <- ((theta^k)/(gamma(lambda+k)/gamma(lambda)))*((trigamma(k+lambda)
      -trigamma(lambda))-(digamma(k+lambda)-digamma(lambda))^2)
    E <- E+a0
    k <- k+1
    if(E != 0){
      error <- a0/E
    }
  }
  return(E)
}

SCD2 <- function(lambda,theta,errorM=1e-6){
  error <- 1
  E <- 0
  a0 <- 0
  k <- 0
  while(abs(error) > errorM){
    a0 <- ((theta^k)/(gamma(lambda+k)/gamma(lambda)))*(digamma(k+lambda)
      -digamma(lambda))
    E <- E+a0
  }
}

```

```

    k <- k+1
    if(E != 0){
      error <- a0/E
    }
  }
  E <- E^2
  return(E)
}

SCKD <- function(lambda,theta,errorM=1e-6){
  error <- 1
  E <- 0
  a0 <- 0
  k <- 0
  while(abs(error) > errorM){
    a0 <- (k*(theta^(k-1))/(gamma(lambda+k)/gamma(lambda)))
      *(((digamma(k+lambda)-digamma(lambda))))
    E <- E+a0
    k <- k+1
    if(E != 0){
      error <- a0/E
    }
  }
  return(E)
}

SCD <- function(lambda,theta,errorM=1e-6){
  error <- 1
  E <- 0
  a0 <- 0
  k <- 0
  while(abs(error) > errorM){
    a0 <- ((theta^k)/(gamma(lambda+k)/gamma(lambda)))*(digamma(k+lambda)
      -digamma(lambda))
    E <- E + a0
    k <- k + 1
    if(E != 0){
      error <- a0/E
    }
  }
  return(E)
}

EMVHP <- function(x,error=1e-8,maxiter=1000,lambda0,theta0){
  maxerror <- 1
  par <- matrix(c(lambda0,theta0),nrow=2,ncol=1,byrow=TRUE)
  j <- 0
  while(maxerror > error){
    n <- length(x)
    xbar <- mean(x)
    H11 <- ((n/genhypergeo(1,par[1],par[2]))*(SCTD2(par[1],par[2])))
      +((n/((genhypergeo(1,par[1],par[2]))^2))*(SCD2(par[1],par[2])))
      +(n*trigamma(par[1])-sum(trigamma(par[1]+x)))
    H22 <- (-n*xbar/(par[2]^2))+n/(par[1]^2)-((2*n/(par[1]*(par[1]+1)))
      *(genhypergeo(3,par[1]+2,par[2])/genhypergeo(1,par[1],par[2])))
    H12 <- ((n/genhypergeo(1,par[1],par[2]))*(SCKD(par[1],par[2])))
      -(((n/par[1])*(genhypergeo(2,par[1]+1,par[2]))
      /((genhypergeo(1,par[1],par[2]))^2)))*(SCD(par[1],par[2])))
    constant <- 1/(H12^2-H11*H22)
    HessianI <- constant*matrix(c(-H22,H12,H12,-H11),nrow=2,ncol=2,byrow=TRUE)
    G1 <- ((1/genhypergeo(1,par[1],par[2]))*(SCD(par[1],par[2])))
      +digamma(par[1])-(sum(digamma(par[1]+x))/n)
    G2 <- (xbar/par[2])-((1/par[1])*(genhypergeo(2,par[1]+1,par[2]))
      /genhypergeo(1,par[1],par[2]))
    Gradient <- matrix(c(G1,G2),nrow=2,ncol=1,byrow=TRUE)
    Incr <- HessianI %*% Gradient
    par <- par - Incr
    maxerror <- max(abs(Incr))
    j <- j+1
    var1 <- -HessianI[1,1]/n
    var2 <- -HessianI[2,2]/n
    if (j == maxiter) break
  }
}

```



```

}
xbar <- mean(x)
like <- n*xbar*log(par[2])-n*log(genhypergeo(1,par[1],par[2]))
      +n*log(gamma(par[1]))-sum(log(gamma(par[1]+x)))
AIC <- 2*2-2*like
BIC <- 2*log(length(x))-2*like
result <- c(par[1],par[2],var1,var2,j,like,AIC,BIC)
return(result)
}

EMVZIP <- function(x,p0,l0,maxiter=1000){
  sum.x <- sum(x)
  zhat <- rep(0,length(x))
  for (i in 1:maxiter){
    zhat[x==0] <- p0/(p0+(1-p0)*exp(-l0))
    l0 <- sum.x/sum(1-zhat)
    p0 <- mean(zhat)
  }
  zhat[x==0] <- 1
  like <- sum(zhat)*log(p0+(1-p0)*exp(-l0))+sum(1-zhat)*log(1-p0)-sum(1-zhat)
      *l0+log(l0)*sum((1-zhat)*x)-sum((1-zhat)*log(factorial(x)))
  AIC <- 2*2-2*like
  BIC <- 2*log(length(x))-2*like
  return(c(like,p0,l0,AIC,BIC))
}

log.like.GPCR <- function(par,x){
  alpha <- 1/(1+exp(-par[1]))
  lambda <- exp(par[2])
  n.0 <- length(x[x==0])
  x.1 <- x[x!=0]
  n.1 <- length(x.1)
  like <- -n.0*log(pgamma(lambda,alpha+1,lower=FALSE))+n.1*lambda-log(lambda)
      *sum(x.1+alpha)+sum(log(gamma(x.1+alpha+1)))
  return(like)
}

results <- nlminb(c(valorinicial1,valorinicial2),log.like.gpd,x=x)

log.like.PL <- function(alpha,x){
  n <- length(x)
  theta <- exp(alpha)
  like <- -2*n*log(theta)+sum(x+3)*log(theta+1)-sum(log(theta+2+x))
  return(like)
}

results <- nlminb(valorinicial,log.like.PL,x=x)

zeta <- function(lambda,nu,error=.Machine$double.eps){
  a <- 1
  zeta <- 0
  j <- 0
  rerror <- 1
  while(rerror>error){
    zeta <- zeta+a
    j <- j+1
    a <- a*lambda/(j^nu)
    rerror <- abs(a/zeta)
  }
  return(zeta)
}

log.like.COMPOI <- function(par,x){
  n <- length(x)
  lambda <- par[1]
  nu <- par[2]
  like <- -sum(x)*log(lambda)+n*log(zeta(lambda,nu))+nu*sum(log(factorial(x)))
  return(like)
}

results <- nlminb(c(valorinicial1,valorinicial2),log.like.COMPOI,x=x)

```

```

tau <- function(lambda,delta,error=.Machine$double.eps){
  a <- 1
  Touchard <- 0
  j <- 0
  rerror <- 1
  while(rerror>error){
    Touchard <- Touchard+a
    a <- a*(lambda/(j+1))*((j+2)/(j+1))^delta
    j <- j+1
    rerror <- a/Touchard
  }
  return(Touchard)
}

MTouchard <- function(lambda,delta,error=1e-9,errorM=1e-9){
  a <- 1
  a1 <- lambda*(2^delta)
  a2 <- a1
  c1 <- log(2)*a1
  c2 <- log(2)*c1
  b1 <- c1
  Touchard <- 0
  EX <- 0
  EX2 <- 0
  ElnX <- 0
  Eln2X <- 0
  EXlnX <- 0
  j <- 0
  maxerror <- 1
  rerror <- 1
  rerror1 <- 1
  rerror2 <- 1
  rerror3 <- 1
  rerror4 <- 1
  rerror5 <- 1
  while(rerror>error){
    Touchard <-Touchard+a
    a <- a*(lambda/(j+1))*((j+2)/(j+1))^delta
    j <- j+1
    rerror <- a/Touchard
  }
  j <- 1
  while(maxerror>errorM){
    EX <- EX+a1
    EX2 <- EX2+a2
    ElnX <- ElnX+c1
    Eln2X <- Eln2X+c2
    EXlnX <- EXlnX+b1
    a1 <- a1*(lambda/j)*((j+2)/(j+1))^delta
    a2 <- a2*(j+1)*(lambda/(j^2))*((j+2)/(j+1))^delta
    c1 <- c1*(lambda/(j+1))*(log(j+2)/log(j+1))*((j+2)/(j+1))^delta
    c2 <- c2*(lambda/(j+1))*((log(j+2)/log(j+1))^2)*((j+2)/(j+1))^delta
    b1 <- b1*(lambda/j)*(log(j+2)/log(j+1))*((j+2)/(j+1))^delta
    j <- j+1
    rerror1 <- a1/EX
    rerror2 <- a2/EX2
    rerror3 <- c1/ElnX
    rerror4 <- c2/Eln2X
    rerror5 <- b1/EXlnX
    maxerror <- max(rerror1,rerror2,rerror3,rerror4,rerror5)
  }
  return(c(EX,EX2,ElnX,Eln2X,EXlnX)/Touchard)
}

EMVTouch <- function(x,lambda0=mean(x),delta0=0,error=1e-8,maxiter=100){
  maxerror <- 1
  par <- matrix(c(lambda0,delta0),nrow=2,ncol=1,byrow=TRUE)
  j <- 0
  while(maxerror>error){
    n <- length(x)
    xbar <- mean(x)
    moments <- MTouchard(par[1],par[2])
  }
}

```

```

EX      <- moments[1]
VarX    <- moments[2] - EX^2
H11     <- -(VarX+(xbar-moments[1]))/(par[1]^2)
H22     <- -(moments[4]-(moments[3])^2)
H12     <- -(moments[5] - EX*moments[3])/par[1]
constant <- 1/(H12^2-H11*H22)
HessianI <- constant*matrix(c(-H22,H12,H12,-H11),nrow=2,ncol=2,byrow=TRUE)
G1      <- (xbar-EX)/par[1]
G2      <- (mean(log(x+1))-moments[3])
Gradient <- matrix(c(G1,G2),nrow=2,ncol=1,byrow=TRUE)
Incr    <- HessianI %*% Gradient
par     <- par - Incr
maxerror <- max(abs(Incr))
j       <- j+1
var1    <- -HessianI[1,1]/n
var2    <- -HessianI[2,2]/n
if (j == maxiter) break
}
like <- sum(x)*log(par[1]+par[2]*sum(log(x+1))-n*log(tau(par[1],par[2])))
      -sum(log(factorial(x)))
result <- c(par[1],par[2],var1,var2,j,like)
AIC <- 2*2-2*like
BIC <- 2*log(length(x))-2*like
lambda <- par[1]
delta <- par[2]
return(list(lambda,delta,var1,var2,like,AIC,BIC))
}

```