Modelos de Regressão para Dados de Contagem com R

Prof. Dr. Walmes M. Zeviani
Eduardo E. Ribeiro Jr
Prof. Dr. Cesar A. Taconelli

Laboratório de Estatística e Geoinformação
Departamento de Estatística
Universidade Federal do Paraná

7 de maio de 2016

edujrrib@gmail.com

Disponibilização

https://github.com/leg-ufpr/MRDCr

https://gitlab.c3sl.ufpr.br/leg/MRDCr

Modelos de Regressão para Dados de Contagem com r - MRDCr

Conteúdo

1 Introdução

2
Modelos Lineares Generalizados

3 Modelo de Regressão Poisson

4

Modelo de Quase-Verossimilhança

5

Modelos Paramétricos Alternativos

5.1

Modelos Paramétricos Alternativos
Modelo Binomial Negativa

5.2

Modelos Paramétricos Alternativos

Modelo Poisson-Generalizada

5.3

Modelos Paramétricos Alternativos

Modelo COM-Poisson

Distribuiçao COM-Poisson I

- Nome COM-Poisson, advém de seus autores **CO**nway e **M**axwell (também é chamada de distribuição Conway-Maxwell-Poisson).
- Proposta em um contexto de filas [?], essa distribuição generaliza a Poisson com a adição de um parâmetro.

Razão de probabilidades consecutivas

- **Distribuição Poisson**

$$\frac{P(Y = y - 1)}{P(Y = y)} = \frac{y}{\lambda}$$

- **Distribuição COM-Poisson**

$$\frac{P(Y = y - 1)}{P(Y = y)} = \frac{y^{\nu}}{\lambda}$$

Distribuiçao COM-Poisson II

Densidade de probabilidade

$$\Pr(Y = y \mid \lambda, \nu) = \frac{\lambda^y}{(y!)^\nu Z(\lambda, \nu)}, \quad \text{em que } Z(\lambda, \nu) = \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\nu}; \text{ e } \quad \lambda > 0,\ \nu \geqslant 0$$

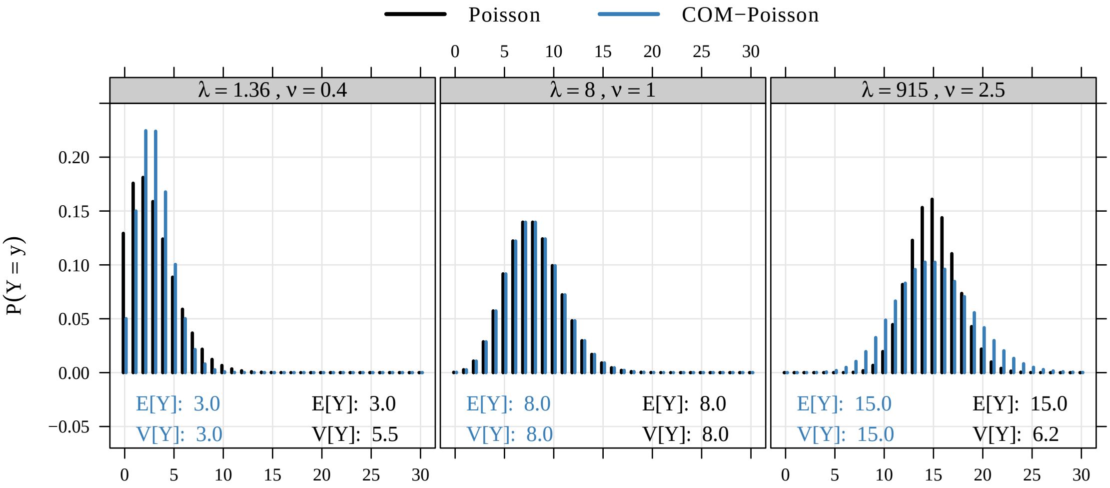
Propriedades

- $\frac{P(Y=y-1)}{P(Y=y)} = \frac{y^\nu}{\lambda}$
- $E(Y) \approx \lambda^{\frac{1}{\nu}} - \frac{\nu-1}{2\nu}$
- $V(Y) \approx \frac{1}{\nu} E(Y)$

Casos particulares

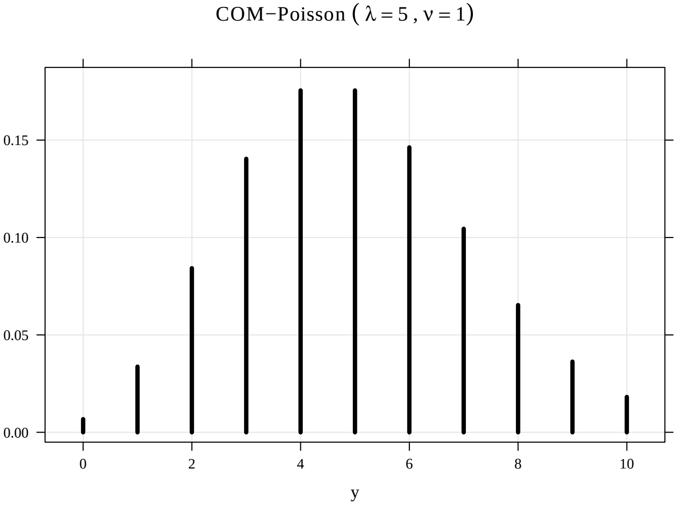
- Distribuição Poisson, quando $\nu = 1$
- Distribuição Bernoulli, quando $\nu \to \infty$
- Distribuição Geométrica, quando $\nu = 0,\ \lambda < 1$

Distribuiçao COM-Poisson III



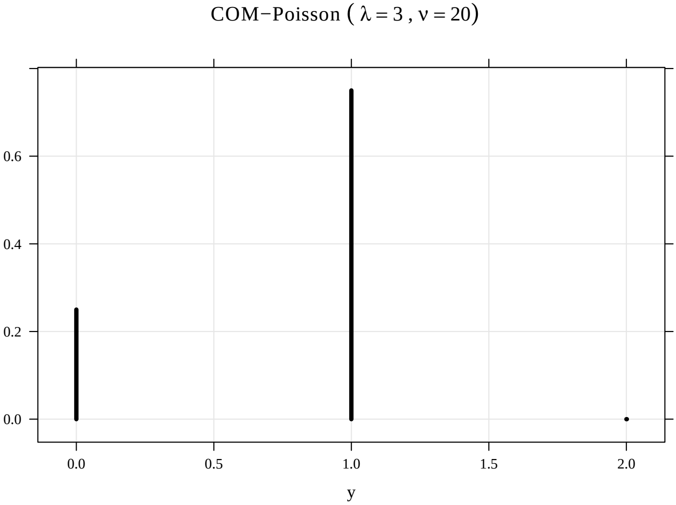
Casos Particulares

- Poisson $\nu = 1$
- Bernoulli $\nu \to \infty$
- Geométrica $\nu = 0$, $\lambda < 1$



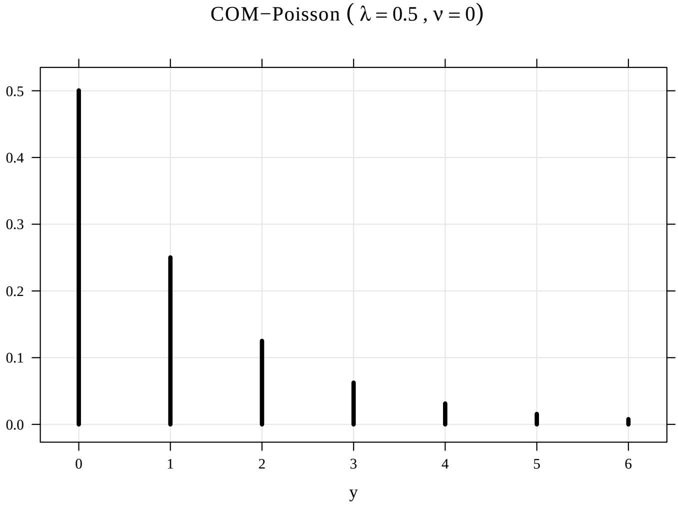
Casos Particulares

- Poisson $\nu = 1$
- Bernoulli $\nu \to \infty$
- Geométrica $\nu = 0$, $\lambda < 1$



Casos Particulares

- Poisson $\nu = 1$
- Bernoulli $\nu \to \infty$
- Geométrica $\nu = 0$, $\lambda < 1$



Modelo de Regressão COM-Poisson

- Incorporando covariáveis em λ da forma $\lambda_i = \exp(X_i \beta)$, em que X_i é o vetor de covariáveis do i-ésimo indivíduo e β o vetor de parâmetros.

Função de verossimilhança

$$L(\lambda, \nu; \underline{y}) = \prod_i^n \left(\frac{\lambda_i^{y_i}}{(y_i!)^\nu} Z(\lambda_i, \nu)^{-1} \right)$$

$$= \lambda_i^{\sum_i^n y_i} \prod_i^n \frac{Z(\lambda_i, \nu)^{-1}}{(y_i!)^\nu}$$

Função de log-verossimilhança

$$l(\lambda, \nu, \underline{y}) = \log \left(\lambda_i^{\sum_i^n y_i} \prod_i^n \frac{Z(\lambda_i, \nu)^{-1}}{(y_i!)^\nu} \right)$$

$$= \sum_i^n y_i \log(\lambda_i) - \nu \sum_i^n \log(y!) - \sum_i^n \log(Z(\lambda_i, \nu))$$

Estudos de caso

Vignette `compoisson.html`

- `capdesfo` : número de capulhos sob efeito de desfolha (sub)
- `capmosca` : número de capulhos sob exposição à mosca branca (sub)
- `ninfas` : número de ninfas de mosca branca em plantas de soja (super)

5.4

Modelos Paramétricos Alternativos
Modelo Gamma-Count

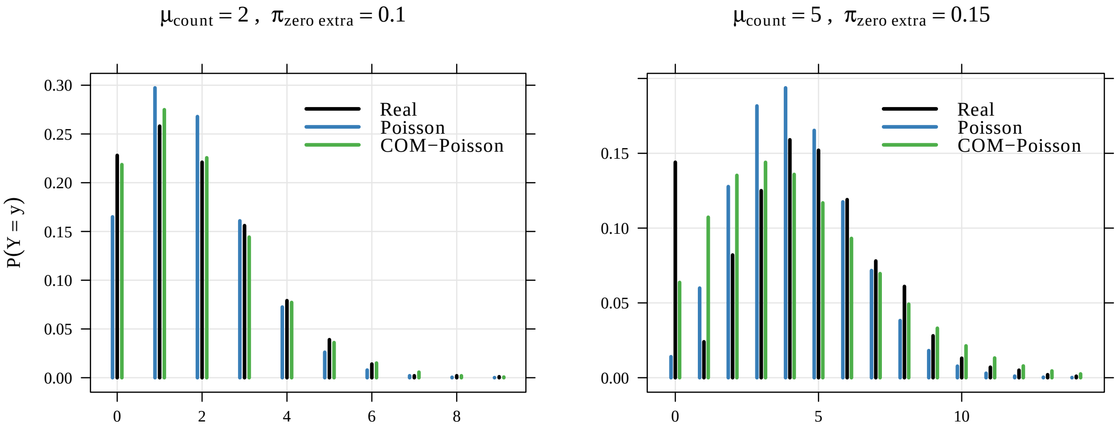
6

Modelos para Excesso de Zeros

Excesso de Zeros I

- Casos em que a proporção de valores nulos na amostra é superior àquela estimada por um modelo de contagem. No caso Poisson $e^{-\lambda}$
- Geralmente contagens com um número excessivo de valores nulos apresentam superdispersão (ocasionada pelo excesso de zeros).
- Os modelos mais flexíveis abordados não capturam esse excesso de zeros e não se ajustam adequadamente.

Excesso de Zeros II



Gerador de excesso de zeros

- Uma limitação das abordagens estudadas é que as contagens nulas e não nulas são provenientes do mesmo processo gerador dos dados.
- Para dados com excesso de zeros, é razoável a suposição da haver mais de um processo gerador atuando na geração dos dados.
- Assim a ocorrência de valores nulos podem ser caracterizada como:
  - **zeros amostrais**: Ocorrem segundo um processo gerador de contagens (e.g Processo Poisson)
  - **zeros estruturais**: Ausência de determinada característica da população.

Modelando contagens com excesso de zeros

Como há dois processos que geram os valores da população, na modelagem deve-se considerar ambos. As principais abordagens nestes casos são via:

- **Modelos de barreira (*Hurdle Models*)**: que desconsidera os zeros amostrais e modela os zeros estruturais e as contagens positivas (seção **??**); e
- **Modelos de mistura (*Zero Inflated Models*)**: que modela os zeros (estruturais e amostrais) em conjunto com as contagens positivas (**??**).

6.1

Modelos para Excesso de Zeros

Modelos de Barreira (Hurdle)

6.2

Modelos para Excesso de Zeros

Modelos de Mistura (Zero Inflated)

7

Modelos com Efeitos Aleatórios

Referências I

- Conway, R. W., Maxwell, W. L. (1962). A queuing model with state dependent service rates. *Journal of Industrial Engineering*, 12, 132–136.
- Paula, G. A. (2013). *Modelos de regressão com apoio computacional*. IME-USP, São Paulo.
- Shmueli, G., Minka, T. P., Kadane, J. B., Borle, S., Boatwright, P. (2005). A useful distribution for fitting discrete data: Revival of the Conway-Maxwell-Poisson distribution. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, 54(1), 127–142.
- Winkelmann, R. (2008). *Econometric analysis of count data* (5th Ed.). Springer Science & Business Media.