

Modelos de regressão para dados de contagem: além do modelo Poisson.

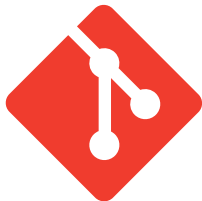
Prof. PhD. Wagner Hugo Bonat
Prf. Dr. Walmes M. Zeviani
Eduardo E. Ribeiro Jr

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

21 de março de 2017

{wbonat,walmes,eduardo.jr}@ufpr.br

Disponibilização



<https://github.com/leg-ufpr/rmcd>

Conteúdo

1. Introdução
2. Distribuições de contagem: propriedades e modelos de regressão
 - 2.1 Distribuição Poisson
 - 2.2 Distribuição Gamma-Count
 - 2.3 Distribuição Poisson-Tweedie
 - 2.4 Distribuição COM-Poisson
 - 2.5 Comparando distribuições para contagens
3. Método de máxima verossimilhança
4. Modelos especificados por suposições de momentos
5. Estudos de casos
6. Discussão

1

Introdução

Dados de contagens

Alguns exemplos de problemas envolvendo contagens:

- ▶ Número de acidentes em uma rodovia por semana;
- ▶ Número de automóveis vendidos por dia;
- ▶ Número de gols marcados por times de futebol por partida;
- ▶ Número de falhas por metro de fio de cobre produzido;
- ▶ Número de colônias de bactérias por $0,01mm^2$ de uma dada cultura ...

Modelos probabilísticos para dados de contagens

- ▶ Modelos probabilísticos para variáveis aleatórias discretas, com suporte no conjunto de números inteiros não-negativos, são potenciais candidatos para a análise de dados de contagens.
- ▶ Algumas alternativas: Distribuição Binomial, Poisson e generalizações; distribuições geradas por misturas, como a beta-binomial, binomial negativa; distribuições fundamentadas na modelagem do tempo entre eventos, na razão de probabilidades sucessivas ...

Regressão para dados de contagens

- ▶ Modelos de regressão são utilizados para modelar a distribuição de uma variável aleatória Y condicional aos valores de um conjunto de variáveis explicativas x_1, x_2, \dots, x_p .
- ▶ Métodos para inferência e modelos de regressão para dados de contagem estão aquém, em quantidade e diversidade, em relação ao verificado para dados contínuos.
- ▶ A aplicação de modelos de regressão com erros normais na análise de contagens, embora frequente, em geral é desaconselhável.

Regressão com erros normais na análise de dados de contagens

- ▶ O modelo linear com erros normais não considera a natureza discreta dos dados;
- ▶ Associa probabilidade nula a qualquer possível contagem;
- ▶ Admite probabilidades não nulas a valores negativos da variável;

Regressão com erros normais na análise de dados de contagens

- ▶ O uso de transformações dificulta a interpretação dos resultados;
- ▶ O uso da transformação logarítmica apresenta problemas para contagens iguais a zero;
- ▶ Não se contempla a relação não constante entre variância e média, característica de dados de contagens.

A distribuição de Poisson

- ▶ A distribuição de Poisson é a principal referência para a análise de dados de contagens.
- ▶ Função de probabilidades:

$$P(Y = k) = \frac{e^{-\mu} \mu^k}{k!}, \quad k = 0, 1, 2, \dots; \mu > 0.$$

- ▶ Se os eventos sob contagem ocorrem independentemente e sujeitos a uma taxa constante $\mu > 0$, sob o modelo Poisson, para um intervalo de exposição de tamanho t tem-se:

$$P(Y_t = k) = \frac{e^{-\mu t} (\mu t)^k}{k!}, \quad k = 0, 1, 2, \dots$$

Propriedades da distribuição de Poisson

Dentre as principais propriedades da distribuição de Poisson, tem-se:

- ▶ Média: $E(Y) = \mu$;
- ▶ Variância: $Var(Y) = \mu$ (equidispersão);
- ▶ Razão de probabilidades sucessivas: $\frac{P(X=k)}{P(X=k-1)} = \frac{\lambda}{k}$, gerando a relação de recorrência:

$$P(Y = k)k = P(Y = k - 1)\lambda;$$

- ▶ Se Y_1, Y_2, \dots, Y_n são v.a.s independentes com $Y_i \sim \text{Poisson}(\lambda_i)$, e $\sum \mu_i < \infty$, então $\sum Y_i \sim \text{Poisson}(\sum \mu_i)$.

Motivações para a distribuição de Poisson

- ▶ Se o tempo decorrido entre sucessivos eventos é uma variável aleatória com distribuição exponencial de média $\lambda = 1/\mu$, então o número de eventos ocorridos em um intervalo t de tempo tem distribuição de Poisson com média μt .
 - ▶ A dualidade entre as distribuições Poisson e exponencial implica que a taxa de ocorrência do evento, definida por:

$$\mu(t) = \lim_{\Delta t \rightarrow 0} \frac{P \{ \text{evento ocorrer em } (t, t + \Delta t) \}}{\Delta t},$$

dado que o evento não ocorreu até o tempo t , **é constante** para todo $t > 0$.

Diferentes comportamentos para $\mu(t)$

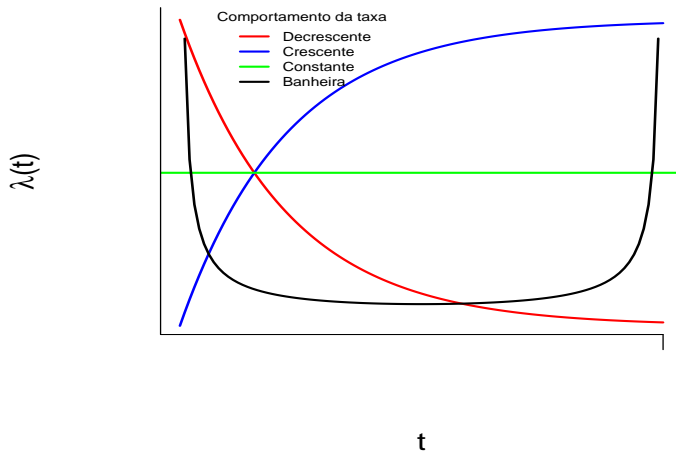


Figura : Diferentes comportamentos para $\mu(t)$

O processo de Poisson

O Processo de Poisson configura um processo de contagem em que $Y(t), t \geq 0$, representa o número de eventos que ocorrem até t , satisfazendo:

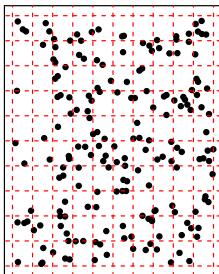
- 1 $Y(t)$ é inteiro e não negativo;
- 2 Para $s < t$, $Y(s) \leq Y(t)$;
- 3 $Y(t) - Y(s)$ é o número de eventos que ocorrem no intervalo $(s, t]$;
- 4 O processo é estacionário:

$$Y(t_2 + s) - Y(t_1 + s) \stackrel{i.d.}{\sim} Y(t_2) - Y(t_1), \forall s > 0$$

- 5 O processo tem incrementos independentes, ou seja, os números de eventos verificados em intervalos disjuntos são independentes.

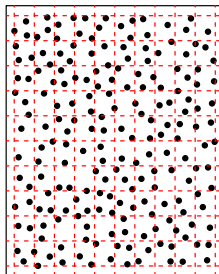
Diferentes padrões em processos de contagens

Padrão aleatório



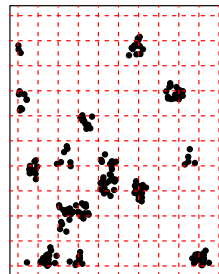
Equidispersão
 $\text{Var}(Y) = E(Y)$

Padrão uniforme



Subdispersão
 $\text{Var}(Y) < E(Y)$

Padrão agregado



Superdispersão
 $\text{Var}(Y) > E(Y)$

Figura : Ilustração de diferentes tipos de processos pontuais

O desafio de dados de contagem

- ▶ Poisson implica equidispersão, ou seja, $E(Y) = \text{var}(Y) = \mu$.
- ▶ Na prática podemos ter
 - ▶ Sub dispersão $E(Y) > \text{var}(Y)$;
 - ▶ Super dispersão $E(Y) < \text{var}(Y)$.
- ▶ Desvios da equidispersão implicam:
 - ▶ Mais ou menos zeros e
 - ▶ Caldas mais leves ou mais pesadas que o modelo Poisson.

Causas da não equidispersão

- ▶ Desvios do processo Poisson;
- ▶ Heterogeneidade entre unidades amostrais.

- ▶ O que acontece caso o modelo Poisson seja usada para dados não equidispersos?
 - ① Super dispersão: erros padrões associados aos coeficientes de regressão serão sub estimados.
 - ② Sub dispersão: erros padrões associados aos coeficientes de regressão serão super estimados.

- ▶ Ambos os casos o modelo Poisson resulta em erros padrões não-confiáveis o que implica em inferências incorretas.

Como lidar com a não equidispersão

- ▶ Mudar a distribuição dos tempos entre eventos: Ex Gamma-Count.
- ▶ Incluir efeitos aleatórios ao nível das observações. Ex Poisson-Tweedie.
- ▶ Modificar a distribuição de Poisson incluindo um parâmetro extra de dispersão. Ex COM-Poisson.

2

Distribuições de contagem: propriedades e modelos de regressão

2.1

Distribuições de contagem: propriedades e
modelos de regressão
Distribuição Poisson

Distribuição Poisson

- Função de probabilidade

$$\begin{aligned} p(y; \mu) &= \frac{\mu^y}{y!} \exp\{-\mu\} \\ &= \frac{1}{y!} \exp\{\phi y - \exp\{\phi\}\}, \quad y \in \mathbb{N}_0, \end{aligned} \tag{1}$$

onde $\phi = \log\{\mu\} \in \mathbb{R}$ e $\kappa(\phi) = \exp\{\phi\}$ denota a função cumulante.

- $E(Y) = \kappa'(\phi) = \exp\{\phi\} = \mu$.
- $\text{var}(Y) = \kappa''(\phi) = \exp\{\phi\} = \mu$.
- Em R temos `dpois()`.

Regressão Poisson

- ▶ Considere $(y_i, x_i), i = 1, \dots, n$, onde y_i 's são iid realizações de Y_i de acordo com a distribuição Poisson.
- ▶ Modelo de regressão Poisson

$$Y_i \sim P(\mu_i), \quad \text{sendo} \quad \mu_i = g^{-1}(x_i^\top \beta),$$

onde x_i and β são vetores $(p \times 1)$ de covariáveis conhecidas e parâmetros de regressão, respectivamente.

- ▶ Em R temos `glm(..., family = poisson)`.

2.2

Distribuições de contagem: propriedades e
modelos de regressão
Distribuição Gamma-Count

© 1995 American Statistical Association

Journal of Business & Economic Statistics, October 1995, Vol. 13, No. 4

Duration Dependence and Dispersion in Count-Data Models

Rainer WINKELMANN

Department of Economics, University of Canterbury, Christchurch, New Zealand

This article explores the relation between nonexponential waiting times between events and the distribution of the number of events in a fixed time interval. It is shown that within this framework the frequently observed phenomenon of overdispersion—that is, a variance that exceeds the mean—is caused by a decreasing hazard function of the waiting times, whereas an increasing hazard function leads to underdispersion. Using the assumption of iid gamma-distributed waiting times, a new count-data model is derived. Its use is illustrated in two applications: the number of births and the number of doctor consultations.

KEY WORDS: Gamma distribution; Negative binomial distribution; Overdispersion; Poisson process; Renewal theory.

**WINKELMANN, R.**

Duration Dependence and Dispersion in Count-Data Models. *Journal of Business & Economic Statistics*, v.13, n.4, p.467–474, 1995.

Duração dependência

- ▶ Considere um processo estocástico definido pela sequência da v.a. τ_i , intervalo de tempo entre eventos.
- ▶ Se $\{\tau_1, \tau_2, \dots\}$ são independentes e identicamente distribuídos, todos com densidade $f(\tau)$, esse processo é chamado de *renewal process*.
- ▶ Defina a variável de contagem N_T como o número de eventos no intervalo $[0, T)$.
- ▶ Defina $\vartheta_n = \sum_{i=1}^n \tau_i$ o tempo até o n -ésimo evento.
- ▶ A distribuição de ϑ_n determina a distribuição de N_T , mas é baseada em covolução.
- ▶ São distribuições fechadas para convolução: normal, Poisson, binomial e gama.
- ▶ Destas, apenas a gama é contínua e positiva.

Duração dependência

- ▶ Denote $E(\tau) = \mu$, $V(\tau) = \sigma^2$ e $CV(\tau) = \sigma/\mu$.
- ▶ Defina $\lambda(\tau) = \frac{f(\tau)}{1-F(\tau)}$ como a função de risco e assuma que é monótona.
- ▶ Existe relação entre o tipo de duração dependência e o coeficiente de variância

$$\left. \begin{array}{c} \frac{d\lambda(t)}{dt} < \\ = \\ > \end{array} \right\} 0 \Rightarrow v = \left. \begin{array}{c} < \\ = \\ > \end{array} \right\} 1 \quad (2)$$

Relação entre número de eventos e intervalo entre eventos

- ▶ Intervalos entre tempo $\tau \sim \text{Gama}(\alpha, \gamma)$,

$$f(\tau, \alpha, \gamma) = \frac{\gamma^\alpha}{\Gamma(\alpha)} \cdot \tau^{\alpha-1} \cdot \exp\{-\gamma\tau\},$$

$$E(\tau) = \frac{\alpha}{\gamma}, \quad V(\tau) = \frac{\alpha}{\gamma^2}.$$

- ▶ Tempo até o n -ésimo evento

$$\vartheta_n = \tau_1 + \cdots + \tau_n \sim \text{Gama}(n\alpha, \gamma),$$

$$f_n(\vartheta, \alpha, \gamma) = \frac{\gamma^{n\alpha}}{\Gamma(n\alpha)} \cdot \vartheta^{n\alpha-1} \cdot \exp\{-\gamma\vartheta\},$$

$$E(\vartheta) = \frac{n\alpha}{\gamma}, \quad V(\vartheta) = \frac{n\alpha}{\gamma^2}.$$

Relação entre número de eventos e intervalo entre eventos

- ▶ A distribuição acumulada do tempo até ϑ_n é

$$F_n(T) = \Pr(\vartheta_n \leq T) = \int_0^T \frac{\gamma^{n\alpha}}{\Gamma(n\alpha)} \cdot t^{n\alpha-1} \cdot \exp\{-\gamma t\} dt.$$

- ▶ Seja $[0, T)$ um intervalo e N_T a v.a. número de eventos neste intervalo.
- ▶ Segue que $N_T < n$ se e somente se $\vartheta_n \geq T$. Assim

$$\Pr(N_T < n) = \Pr(\vartheta_n \geq T) = 1 - F_n(T);$$

- ▶ Já que $\Pr(N_T = n) = \Pr(N_T < n + 1) - \Pr(N_T < n)$, então

$$\Pr(N_T = n) = F_n(T) - F_{n+1}(T).$$

Relação entre número de eventos e intervalo entre eventos

- ▶ Portanto, distribuição de N_T é resultado da diferença de acumuladas da distribuição Gama, pois

$$F_n(T) = G(n\alpha, \gamma T) = \int_0^T \frac{\gamma^{n\alpha}}{\Gamma(n\alpha)} t^{n\alpha-1} \cdot \exp\{-\gamma t\} dt. \quad (3)$$

- ▶ Assim

$$\begin{aligned} \Pr(N_T = n) &= G(n\alpha, \gamma T) - G((n+1)\alpha, \gamma T) \\ &= \left[\int_0^T \frac{\gamma^{n\alpha}}{\Gamma(n\alpha)} t^{n\alpha-1} \cdot \exp\{-\gamma t\} dt \right] \\ &\quad - \left[\int_0^T \frac{\gamma^{(n+1)\alpha}}{\Gamma((n+1)\alpha)} t^{(n+1)\alpha-1} \cdot \exp\{-\gamma t\} dt \right] \end{aligned}$$

Função de probabilidade

- Em R temos

```
dgc <- function(y, gamma, alpha, log = FALSE) {  
  p <- pgamma(q = 1,  
              shape = y * alpha,  
              rate = alpha * gamma) -  
    pgamma(q = 1,  
          shape = (y + 1) * alpha,  
          rate = alpha * gamma)  
  if(log == TRUE) {p <- log(p)}  
  return(p)  
}
```

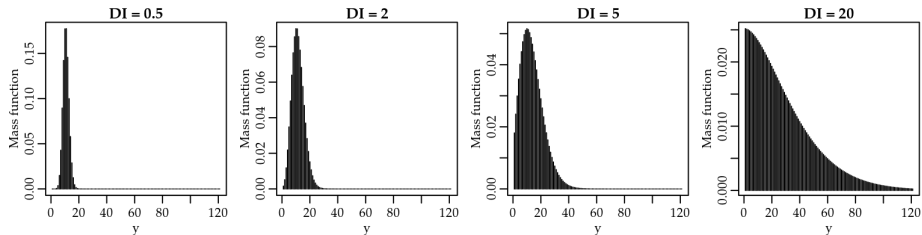


Figura : Função de probabilidade de acordo com valores do índice de dispersão - Gamma-Count.

- Índice de dispersão - $DI = E(Y)/var(Y)$

Parametrização para modelo de regressão

- ▶ A média da variável aleatória N_T é resultado de

$$\begin{aligned} E(N) &= \sum_{i=0}^{\infty} i \cdot \Pr(i) \\ &= \sum_{i=1}^{\infty} i \cdot \Pr(i) \\ &= \sum_{i=1}^{\infty} G(i\alpha, \gamma T). \end{aligned}$$

- ▶ Para um T cada vez maior, tem-se que

$$N(T) \sim \text{Normal} \left(\frac{\gamma}{\alpha}, \frac{\gamma}{\alpha^2} \right).$$

Parametrização para modelo de regressão

- Considere que

$$\frac{\gamma}{\alpha} = \exp\{x^\top \beta\} \Rightarrow \gamma = \alpha \exp\{x^\top \beta\}.$$

Essa parametrização produz um modelo de regressão para a média do tempo entre eventos definida por

$$E(\tau|x) = \frac{\alpha}{\gamma} = \exp\{-x^\top \beta\}.$$

- O modelo de regressão é para o tempo entre eventos (τ) e não diretamente para contagem porque, a menos que $\alpha = 1$, não é certo que $E(N_i|x_i) = [E(\tau_i|x_i)]^{-1}$.
- Em R temos `MrDCR::gcnt(formula, data)`.

2.3

Distribuições de contagem: propriedades e
modelos de regressão
Distribuição Poisson-Tweedie

Distribuição Tweedie

- ▶ Distribuição Tweedie (Jørgensen, 1997)

$$f(z; \mu, \phi, p) = a(z, \phi, p) \exp\{(z\psi - k(\psi))/\phi\},$$

onde $\mu = E(Z) = k'(\psi)$ é a média.

- ▶ $\phi > 0$ e ψ são os parâmetros de dispersão e canônico. $k(\psi)$ é a função cumulante e $a(z, \phi, p)$ é a constante normalizadora.
- ▶ $\text{Var}(Z) = \phi\mu^p$ onde $p \in (-\infty, 0] \cup [1, \infty)$ é um index determinando a distribuição.
- ▶ Casos especiais: Normal ($p = 0$), Poisson ($p = 1$), não-central gamma ($p = 1.5$), gamma ($p = 2$), normal inversa ($p = 3$) and distribuições estáveis ($p > 2$).
- ▶ Notação $Z \sim Tw_p(\mu, \phi)$.

Função de probabilidade

- Especificação hierárquica:

$$\begin{aligned}Y|Z &\sim \text{Poisson}(Z) \\ Z &\sim \text{Tw}_p(\mu, \phi).\end{aligned}$$

- Função de probabilidade ($p > 1$)

$$f(y; \mu, \phi, p) = \int_0^\infty \frac{z^y \exp^{-z}}{y!} a(z, \phi, p) \exp\{(z\psi - k(\psi))/\phi\} dz.$$

- Forma fechada está disponível apenas no caso especial - binomial negativa.
- Pode ser aproximada por integração Monte Carlo e/ou integração Gauss-Laguerre.

Função de probabilidade

► Em R temos

```
require(tweedie)
# Integrand Poisson X Tweedie distributions
integrand <- function(x, y, mu, phi, power) {
  int = dpois(y, lambda = x)*dtweedie(x, mu = mu,
                                       phi = phi, power = power)

  return(int)
}

# Computing the pmf using Monte Carlo
dptw <- function(y, mu, phi, power, control_sample) {
  pts <- control_sample$pts
  norma <- control_sample$norma
  integral <- mean(integrand(pts, y = y, mu = mu, phi = phi,
                             power = power)/norma)

  return(integral)
}
```

Função de probabilidade

► Exemplo

```
set.seed(123)
pts <- rtweedie(n = 1000, mu = 10, phi = 1, power = 2)
norma <- dtweedie(pts, mu = 10, phi = 1, power = 2)
control_sample <- list("pts" = pts, "norma" = norma)
dptw(y = c(0, 5, 10, 15), mu = 10, phi = 1, power = 2,
      control_sample = control_sample)

## [1] 0.09374152 0.05902132 0.03539478 0.02171625

dnbinom(x = c(0, 5, 10, 15), mu = 10, size = 1)

## [1] 0.09090909 0.05644739 0.03504939 0.02176291
```

Momentos e casos especiais

- ▶ Média e variância marginal são facilmente obtidos

$$\begin{aligned}E(Y) &= \mu \\ \text{Var}(Y) &= \mu + \phi\mu^p.\end{aligned}$$

- ▶ Casos especiais: Hermite ($p = 0$), Neyman-Type A ($p = 1$), Pólya-Aeppli ($p = 1.5$), binomial negativa ($p = 2$) e Poisson inversa-Normal ($p = 3$).
- ▶ Cuidado! - Hermite é um caso limite.
- ▶ p é um índice que distingue entre importantes distribuições.
- ▶ Espaço paramétrico de p é não trivial $p \in 0 \cup [1, \infty)$.
- ▶ Estimação de p funciona como uma seleção automática de distribuições.
- ▶ Notação $Y \sim PTw_p(\mu, \phi)$.

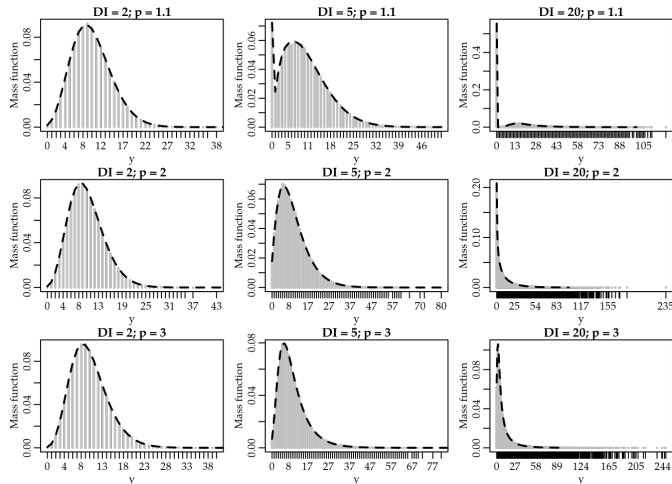


Figura : Distribuição de probabilidade empírica (cinza) e função de probabilidade aproximada (preta) por valores do índice de dispersão e valores do parâmetro de potência: Poisson-Tweedie.

Regressão Poisson-Tweedie

- ▶ Considere (y_i, x_i) , $i = 1, \dots, n$, onde y_i 's são iid realizações de Y_i de acordo com a distribuição Poisson-Tweedie.
- ▶ Modelo de regressão Poisson-Tweedie

$$Y_i \sim PTw_p(\mu_i, \phi), \quad \text{sendo} \quad \mu_i = g^{-1}(x_i^\top \beta),$$

onde x_i and β são vetores $(p \times 1)$ de covariáveis conhecidas e parâmetros de regressão, respectivamente.

- ▶ Em R temos ??

2.4

Distribuições de contagem: propriedades e
modelos de regressão
Distribuição COM-Poisson

Distribuição COM-Poisson

- ▶ Nome COM-Poisson, advém de seus autores **CO**nway e **MA**xwell (também é chamada de distribuição Conway-Maxwell-Poisson).
- ▶ Proposta em um contexto de filas [1], essa distribuição generaliza a Poisson com a adição de um parâmetro.
- ▶ Modifica a relação entre probabilidades consecutivas.

▶ Distribuição Poisson

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

▶ Distribuição COM-Poisson

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

Distribuição COM-Poisson

Distribuição de probabilidades

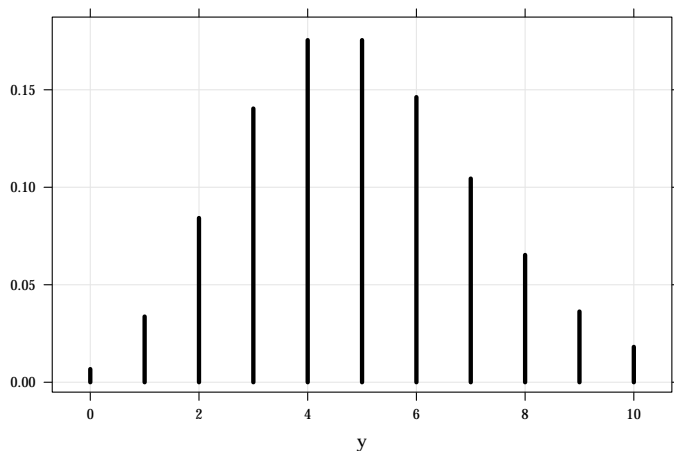
$$\Pr(Y = y \mid \lambda, \nu) = \frac{\lambda^y}{(y!)^\nu Z(\lambda, \nu)}, \quad \text{em que } Z(\lambda, \nu) = \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\nu}; \text{ e } \lambda > 0, \nu \geq 0$$

Casos particulares

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

Distribuição COM-Poisson

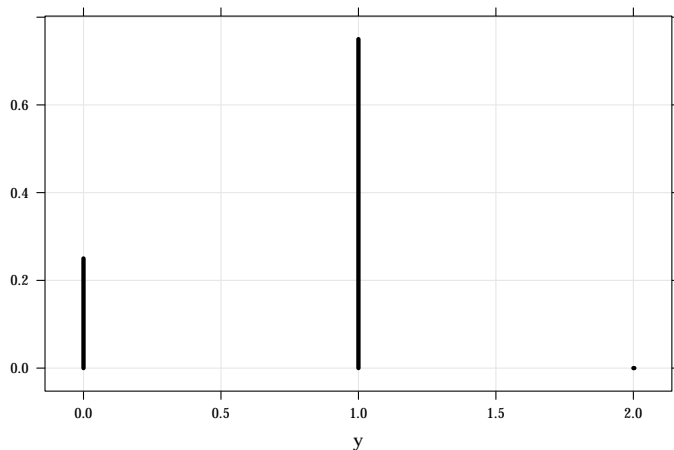
- Poisson $\nu = 1$
- Bernoulli $\nu \rightarrow \infty$
- Geométrica $\nu = 0, \lambda < 1$

COM-Poisson ($\lambda = 5, \nu = 1$)

Distribuição COM-Poisson

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

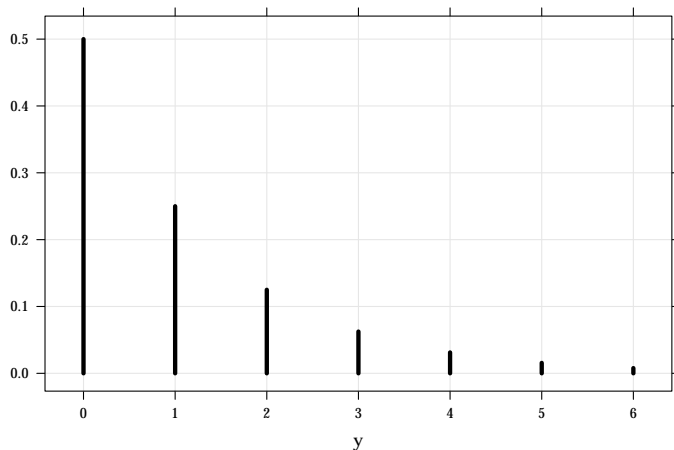
COM-Poisson ($\lambda = 3, \nu = 20$)



Distribuição COM-Poisson

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

COM-Poisson ($\lambda = 0.5, \nu = 0$)



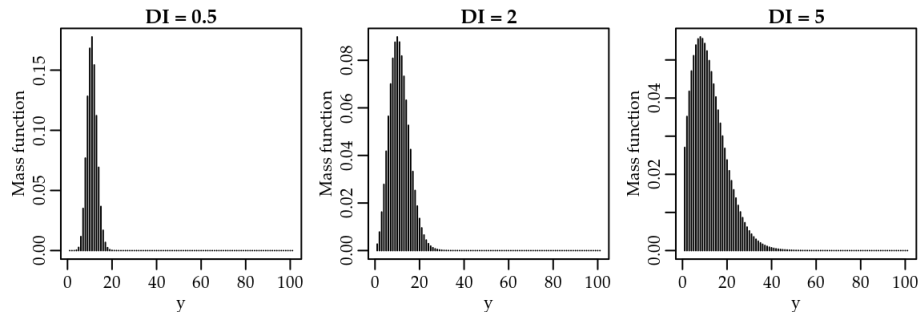
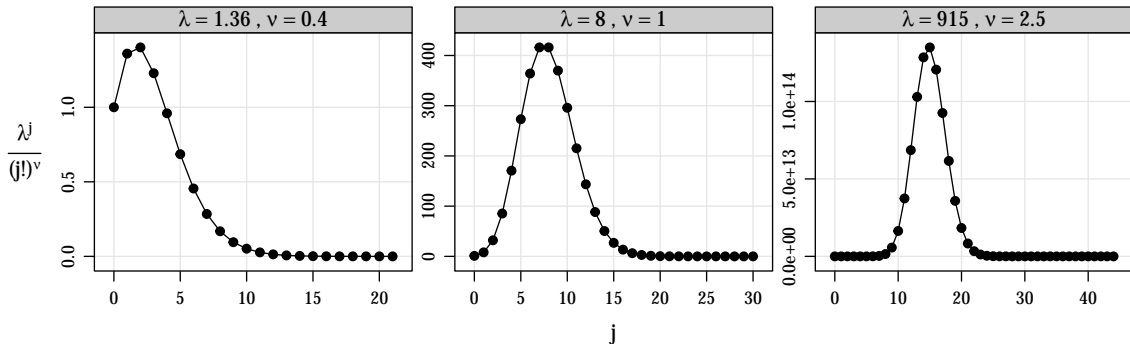


Figura : Distribuição de probabilidade por valores do índice de dispersão: COM-Poisson.

Assintoticidade da função Z

$$Z(\lambda, \nu) = \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\nu}$$



Momentos da distribuição

Não tem expressão analítica, calculamos utilizando a definição de média e variância;

$$\blacktriangleright E(Y) = \sum_{y=0}^{\infty} y \cdot p(y)$$

$$\blacktriangleright V(Y) = \sum_{y=0}^{\infty} y^2 \cdot p(y) - E^2(Y)$$

- \blacktriangleright Regressão COM-Poisson: $\lambda_i = \exp(x_i\beta)$, em que x_i é o vetor de covariáveis do i-ésimo indivíduo e β o vetor de parâmetros.

Aproximação proposta por [3], boa aproximação para $\nu \leq 1$ ou $\lambda > 10^\nu$

$$\blacktriangleright E(Y) \approx \lambda^{\frac{1}{\nu}} - \frac{\nu - 1}{2\nu}$$

$$\blacktriangleright V(Y) \approx \frac{1}{\nu} \cdot E(Y)$$

2.5

Distribuições de contagem: propriedades e
modelos de regressão

**Comparando distribuições para
contagens**

Medindo propriedades das distribuições

- ▶ Índice de dispersão:

$$DI = \frac{\text{var}(Y)}{E(Y)}.$$

$DI < 1$ sub dispersão, $DI = 1$ equidispersão e $DI > 1$ super dispersão.

- ▶ Índice de inflação de zeros:

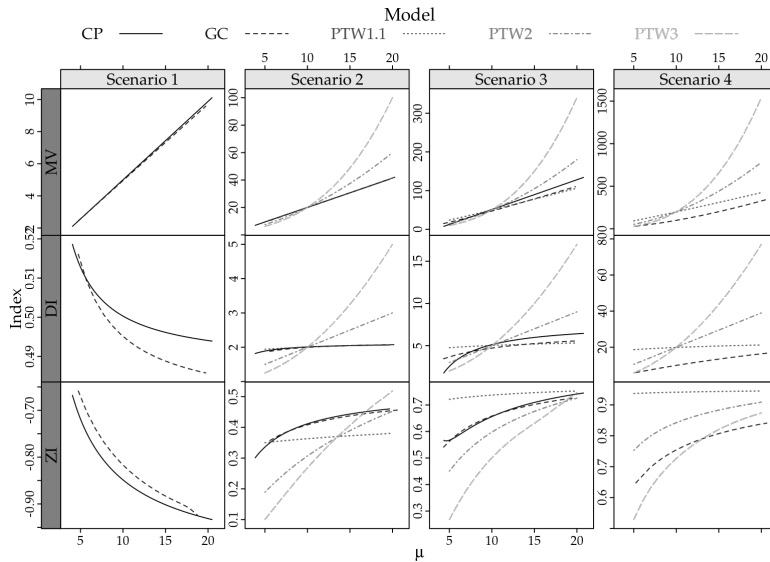
$$ZI = 1 + \frac{\log P(Y = 0)}{E(Y)}.$$

$ZI < 0$ zero deflacionado, $ZI = 1$ não zero inflado e $ZI > 1$ zero inflacionado.

- ▶ Índice de cauda pesada:

$$HT = \frac{P(Y = y + 1)}{P(Y = y)} \quad \text{for } y \rightarrow \infty.$$

$HT \rightarrow 1$ quando $y \rightarrow \infty$ indica cauda pesada.



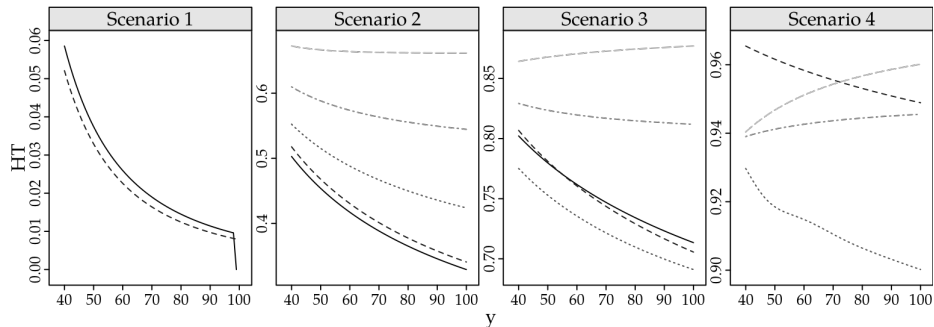


Figura : Índice de cauda pesada para alguns valores extremos da Y .

Flexibilidade

Tabela : Reference models and dominant features by dispersion and power parameter values.

Modelo de referência	Fatos dominantes	Dispersão	Power
Poisson	Equi	$\phi = 0$	—
Gamma-Count	Sub, Equi, Super, deflação de zero	$\alpha \leq 1$	—
COM-Poisson	Sub, Equi, Super, deflação de zero	$\nu \leq 1$	—
Hermite	Super	$\phi \leq 0$	$p = 0$
Neyman Type A	Super, Zero-inflacionado	$\phi > 0$	$p = 1$
<i>Poisson compound Poisson</i>	Super, Zero-inflacionado	$\phi > 0$	$1 < p < 2$
Pólya-Aeppli	Super, Zero-inflacionado	$\phi > 0$	$p = 1.5$
Negative binomial	Super	$\phi > 0$	$p = 2$
<i>Poisson positive stable</i>	Super, cauda pesada	$\phi > 0$	$p > 2$
Poisson-inverse Gaussian	Super, cauda pesada	$\phi > 0$	$p = 3$

3

Método de máxima verossimilhança

4

Modelos especificados por suposições de momentos







5

Estudos de casos

6

Discussão

Referências

-  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.
-  Zeileis, A., Kleiber, C., Jackman, S. (2008). Regression Models for Count Data in R. *Journal of Statistical Software*, 27(8), 1 - 25. doi:<http://dx.doi.org/10.18637/jss.v027.i08>
-  Winkelmann, R. (2008). *Econometric analysis of count data* (5th Ed.). Springer Science & Business Media.
-  SILVA, A. M.; DEGRANDE, P. E.; SUEKANE, R.; FERNANDES, M. G.; ZEVIANI, W. M. Impacto de diferentes níveis de desfolha artificial nos estádios fenológicos do algodoeiro. *Revista de Ciências Agrárias*, v.35, n.1, 2012 (prelo).

Referências



WINKELMANN, R.; ZIMMERMANN, K.

Count data models for demographic data.

Mathematical Population Studies, v.4, n.3, p.205–221, 1994.



WINKELMANN, R.

Duration dependence and dispersion in count-data models.

Journal of Business & Economic Statistics, v.13, n.4, p.467–474, 1995.



CONSUL, P. C. AND G. C. JAIN

A generalization of the Poisson distribution. **Technometrics**, v.15, n.4, p.791–799, 1973.



CONSUL, P. C

Generalized Poisson Distributions: Properties and Applications. **Statistics: Textbooks and Monographs**, New York: Marcel Dekker Inc. 1989.