CE225 - Modelos Lineares Generalizados

Cesar Augusto Taconeli

16 de agosto, 2018

Aula 4 - Família exponencial de distribuições

Componente aleatório de um modelo linear generalizado

- O componente aleatório de um modelo linear generalizado consiste em uma variável aleatória y, por meio de um conjunto de observações independentes $y_1, y_2, ..., y_n$, com distribuição pertencente à *família exponencial*.
- Mais especificamente, assumimos que a função (densidade) de probabilidades de y possa ser expressa na forma:

$$f(y_i; \theta_i, \phi) = exp\left\{\frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i; \phi)\right\}, \tag{1}$$

sendo usualmente chamada de *forma canônica da família exponencial*, ou *família exponencial de dispersão*.

Componente aleatório de um modelo linear generalizado

- O parâmetro θ_i é chamado *parâmetro natural* (ou *parâmetro canônico*) e ϕ o *parâmetro de dispersão* da distribuição.

- Em geral, temos $a(\phi) = \phi$ ou $a_i(\phi) = \frac{\phi}{\omega_i}$, sendo ω_i um peso particular a cada observação.

- A família exponencial de dispersão contempla diversas distribuições uni e bi-paramétricas pertencentes à família exponencial, por exemplo as distribuições binomial, poisson, normal, gama e normal inversa.

Algumas propriedades da família exponencial de dispersão

- Para distribuições pertencentes à família exponencial de dispersão, expressões para $E(y_i)$ e $Var(y_i)$ são dadas por:

$$E(y_i) = \mu_i = b'(\theta_i) = \frac{\partial b(\theta_i)}{\partial \theta_i} \tag{2}$$

e

$$Var(y_i) = a(\phi) \times b''(\theta_i) = a(\phi) \times \frac{\partial \mu_i}{\partial \theta_i}. \tag{3}$$

Algumas propriedades da família exponencial de dispersão

- Assim, a variância de y_i pode ser fatorada em dois componentes:
  - O primeiro ($a(\phi)$) é função de um parâmetro (ϕ) que está associado exclusivamente à dispersão de y_i (não à sua média);
  - O segundo, usualmente denotado por $V(\mu_i) = b''(\theta_i)$ e chamado *função de variância*, é função da média da distribuição, e exprime a relação média-variância de y.
- Cada distribuição pertencente à família exponencial de dispersão tem sua particular função de variância e vice-versa (unicidade).

Algumas propriedades da família exponencial de dispersão

- Uma vez que a distribuição conjunta de $y_1, y_2, ..., y_n$ é dada por:

$$f(\boldsymbol{y}; \boldsymbol{\theta}, \phi) = \prod_{i=1}^{n} f(y_i; \theta_i, \phi) = \exp\left\{ \frac{\sum_{i=1}^{n} y_i\theta_i - \sum_{i=1}^{n} b(\theta_i)}{a(\phi)} \right\} \exp \sum_{i=1}^{n} c(y_i; \phi), \tag{4}$$

pelo teorema da fatoração de Neyman-Fisher, tem-se que $\sum_{i=1}^{n} y_i$ é uma estatística suficiente para θ_i se ϕ for conhecido.

- Na sequência são ilustradas algumas distribuições pertencentes à família exponencial de dispersão.

Distribuição binomial

- Uma variável aleatória x_i tem distribuição binomial se sua função de probabilidades é dada por:

$$f(x_i; n_i, \pi_i) = \binom{n_i}{x_i} \pi_i^{x_i} (1 - \pi_i)^{n_i - x_i}; \quad x_i = 0, 1, 2, ..., n_i; \quad 0 < \pi_i < 1, \quad (5)$$

em que x_i corresponde à contagem de *sucessos* em n_i observações independentes de um experimento binário.

Distribuição binomial

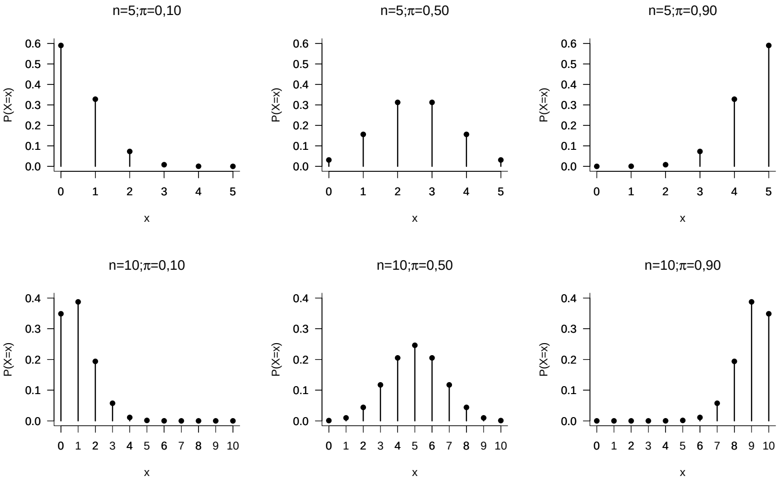

Figura 1: Ilustração - distribuição binomial

Distribuição binomial

- Podemos expressar a distribuição binomial, de maneira alternativa, pela variável $y_i = \frac{x_i}{n_i}$, a fração amostral de sucessos, com função de probabilidades:

$$f(y_i; n_i, \pi_i) = \binom{n_i}{n_i y_i} \pi_i^{n_i y_i} (1 - \pi_i)^{n_i - (n_i y_i)}; y_i = 0, \frac{1}{n_i}, \frac{2}{n_i}, ..., 1; 0 < \pi_i < 1. \quad (6)$$

Exercício 1

Verifique que a distribuição binomial pode ser expressa na forma da família exponencial de dispersão. Identifique $a(\phi)$, θ_i, $b(\theta_i)$ e $c(y_i, \phi)$. Deduza a média e a variância de y_i e identifique a função de variância.

Distribuição binomial

- Algumas notas sobre o modelo binomial binomial:

  * O modelo binomial é usado, principalmente, na modelagem de dados binários ou de proporções discretas;

  * É bem aproximado pela distribuição $Normal(\pi, \frac{\pi(1-\pi)}{m})$ quando $m\pi > 0,5$ e $0,1 \leq \pi \leq 0,9$ ou $m\pi > 25$, para qualquer valor de π.

Distribuição Poisson

- Uma variável aleatória discreta y_i tem distribuição de Poisson se sua função de probabilidades é dada por:

$$f(y_i; \mu_i) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!}, \tag{7}$$

com $y_i = 0, 1, 2, ...$ e $\mu_i > 0$.

Exercício 2

Verifique que a distribuição Poisson pode ser expressa na forma da família exponencial de dispersão. Identifique $a(\phi)$, θ_i, $b(\theta_i)$ e $c(y_i, \phi)$. Deduza a média e a variância de y_i e identifique a função de variância.

Distribuição Poisson

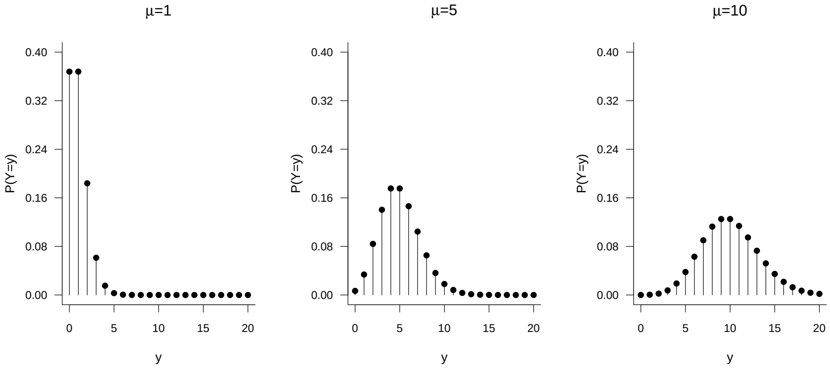

Figura 2: Ilustração - distribuição de Poisson

Distribuição Poisson

- Algumas notas sobre o modelo de Poisson:
  - Se eventos ocorrem independente e aleatoriamente no tempo (ou espaço), com taxa média de ocorrência constante, o modelo atribui probabilidades ao número de eventos por intervalo de tempo (ou região do espaço);
  - Proporciona, em geral, uma descrição satisfatória de dados cuja variância é proporcional à média;
  - Surge como caso limite para a distribuição binomial quando $n \to \infty$ e $\pi \to 0$ (matendo fixo $\mu = n\pi$);
  - É bem aproximada pela distribuição $Normal(\mu, \mu)$ para μ suficientemente grande.

Distribuição normal

- Uma variável aleatória contínua y_i tem distribuição normal se sua função densidade de probabilidade é dada por:

$$f(y_i;\mu_i,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(y_i-\mu_i)^2}{2\sigma^2}\right\}, \tag{8}$$

com $-\infty < y_i < \infty;\ \ -\infty < \mu_i < \infty;\ \ \sigma > 0.$

Exercício 3

Verifique que a distribuição normal pode ser expressa na forma da família exponencial de dispersão. Identifique $a(\phi)$, θ_i, $b(\theta_i)$ e $c(y_i,\phi)$. Deduza a média e a variância de y_i e identifique a função de variância.

Distribuição normal

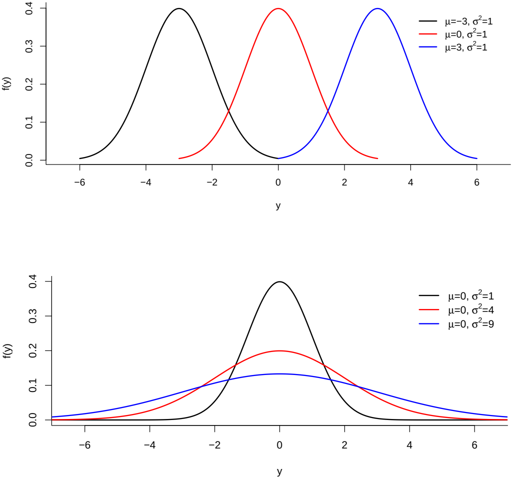

Figura 3: Ilustração - distribuição normal

Distribuição gama

- Uma variável aleatória contínua y_i tem distribuição gama se sua função densidade de probabilidade é dada por:

$$f(y_i; \mu_i, \nu) = \frac{\left(\frac{\nu}{\mu_i}\right)^{\nu}}{\Gamma(\nu)} y_i^{\nu-1} \exp\left\{-\frac{y_i \nu}{\mu_i}\right\}, \tag{9}$$

com $y_i > 0$, $\mu_i > 0$, $\nu > 0$ e $\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt$.

- Uma das parametrizações alternativas da distribuição gama é a seguinte:

$$f(y; \alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} y^{\alpha-1} exp\left\{-\beta y\right\}, \tag{10}$$

tal que a equivalência das duas parametrizações decorre de $\mu = \frac{\alpha}{\beta}$ e $\nu = \alpha$.

Distribuição gama

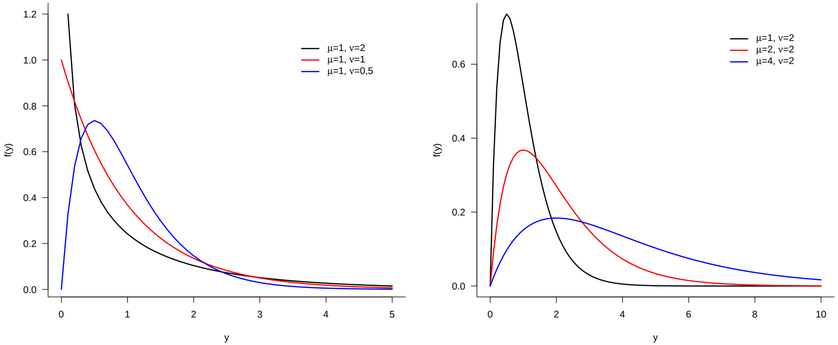

Figura 4: Ilustração - distribuição gama

Distribuição gama

- O modelo gama é usado na análise de dados contínuos não negativos em que a variância aumenta conforme a média, particularmente no caso em que o coeficiente de variação é aproximadamente constante.

Exercício 4

Verifique que a distribuição gama pode ser expressa na forma da família exponencial de dispersão (use a primeira parametrização apresentada). Identifique $a(\phi)$, θ_i, $b(\theta_i)$ e $c(y_i, \phi)$. Deduza a média e a variância de y_i e identifique a função de variância.

Distribuição normal inversa

- Uma variável aleatória contínua tem distribuição normal inversa se sua função densidade de probabilidade é dada por:

$$f(y_i; \mu_i, \lambda) = \sqrt{\frac{1}{2\pi\phi y_i^3}} exp\left\{-\frac{(y_i - \mu_i)^2}{2\mu^2\phi y_i}\right\}, \tag{11}$$

com $y_i > 0$, $\mu_i > 0$, $\phi > 0$.

Distribuição normal inversa

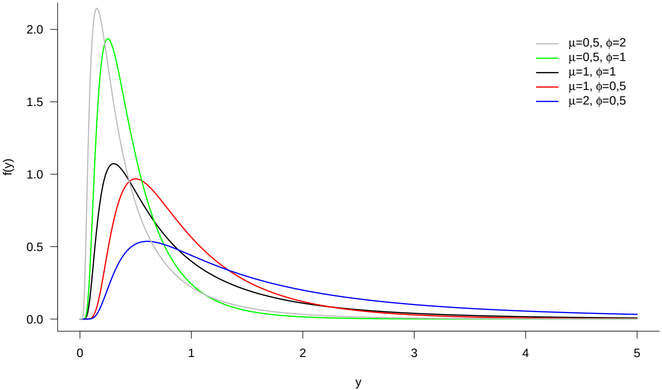

Figura 5: Ilustração - distribuição normal inversa

Distribuição normal inversa

- O modelo normal inverso se aplica a análise de dados contínuos, não negativos com distribuição acentuadamente assimétrica.

Exercício 5

Verifique que a distribuição normal inversa pode ser expressa na forma da família exponencial de dispersão. Identifique $a(\phi)$, θ_i, $b(\theta_i)$ *e* $c(y_i, \phi)$. *Deduza a média e a variância de* y_i *e identifique a função de variância.*

Distribuição binomial negativa

- Uma variável aleatória discreta Y tem distribuição binomial negativa se a sua função de probabilidades é dada por:

$$f(y_i; \mu_i, k) = \frac{\Gamma(k + y_i)}{\Gamma(k) y_i!} \frac{\mu_i^{y_i} k^k}{(\mu_i + k)^{k+y_i}}, \tag{12}$$

com $y_i = 0, 1, 2, ...;$ $\mu_i > 0;$ $k > 0$.

Distribuição binomial negativa

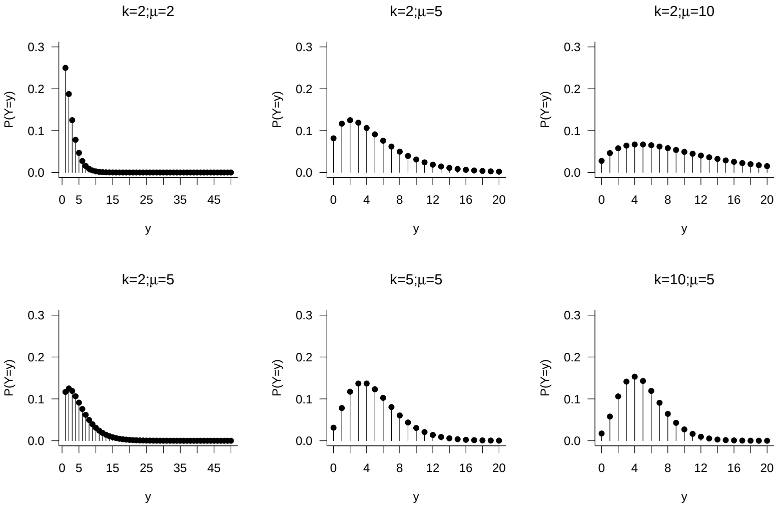

Figura 6: Ilustração - distribuição binomial negativa

Distribuição binomial negativa

- O modelo binomial negativo é uma alternativa ao de Poisson em situações em que a variância dos dados aumenta mais rapidamente que a média;
- Para valores inteiros de k, usa-se também a denominação *modelo de Pascal*;
- Para $k = 1$, temos como caso particular a distribuição geométrica.

Exercício 6

Verifique que a distribuição binomial negativa pode ser expressa na forma da família exponencial de dispersão (k fixo). Identifique $a(\phi)$, θ_i, $b(\theta_i)$ e $c(y_i, \phi)$. Deduza a média e a variância de y_i e identifique a função de variância.