Inferência para CS
Tópico 09 - Outros Métodos de Estimação

Renato Martins Assunção

DCC - UFMG

2013

Métodos de Estimação

- Aprendemos um método (MLE) para estimar o vetor de parâmetros $\boldsymbol{\theta}$ de um modelo estatístico para os dados $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$.
- O método é intuitivo e pode ser usado sempre que pudermos escrever a verossimilhança $L(\boldsymbol{\theta})$.
- Mas o MLE não é o único método para estimar parâmetros.
- Existem vários outros, todos intuitivos. *Nem sempre eles coincidem nas suas estimativas de* $\boldsymbol{\theta}$.
- Como escolher um bom método de estimação?
- Antes de responder a isto, vamos ver exemplos de diferentes métodos de estimação.

Método de Momentos

- A ideia é igualar os momentos teóricos com momentos amostrais correspondentes e resolver um sistema de equações para obter uma estimativa para $\boldsymbol{\theta}$.
- Caso inicial simples: $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ é formado por variáveis i.i.d.
- Cada uma delas segue a distribuição de Y.
- A distribuição de Y depende de parâmetro $\boldsymbol{\theta} \in \mathbb{R}^k$.
- Momentos:

Momento Teórico	Momento Amostral
$\mathbb{E}(Y)$	$m_1 = \overline{Y} = \sum_{i=1}^{n} Y_i/n$
$\mathbb{E}(Y^2)$	$m_2 = \sum_i Y_i^2/n$
$\mathbb{E}(Y^3)$	$m_3 = \sum_i Y_i^3/n$
$\vdots$	$\vdots$

Método de Momentos

- Se $\boldsymbol{\theta} \in \mathbb{R}^k$, igualamos os k primeiros momentos teóricos com os momentos amostrais correspondentes e resolvemos o sistema de equações para $\boldsymbol{\theta}$.
- Por exemplo, suponha que $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ sejam v.a.'s i.i.d. com distribuição gama com parâmetros α e β

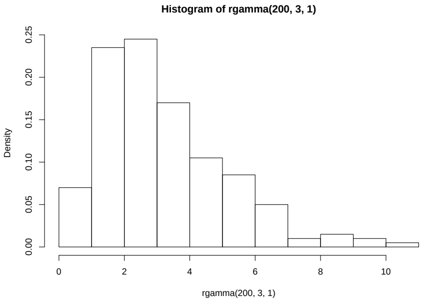


Figura: Amostra de $n = 200$ valores de Gama($\alpha = 3, \beta = 1$)

Densidade de Gama(α, β)

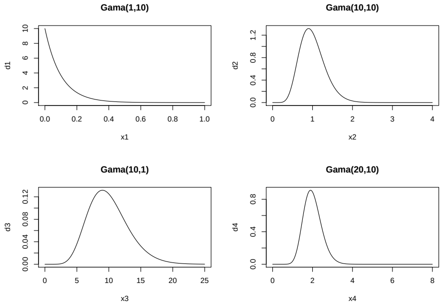


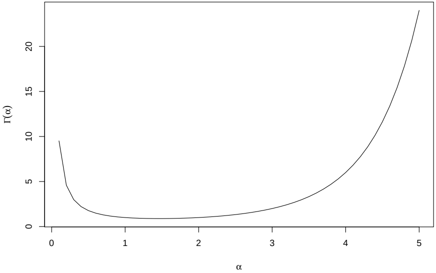
Figura: Gama(1,10); Gama(10,10); Gama(10,1); Gama(20,10)

Método de Momentos: gama

- Se $Y \sim \text{Gama}(\alpha, \beta)$, então a densidade de UMA delas é

$$f(y; \alpha, \beta) = \beta^{\alpha}/\Gamma(\alpha)\ y^{\alpha-1} \exp(-\beta y)$$

- $\Gamma(\alpha)$ é a função gama
- Quando $\alpha = k$, inteiro positivo, temos $\Gamma(k) = (k-1)!$
- Para α geral, $\Gamma(\alpha)$ "interpola" os fatoriais k!



Método de Momentos: gama

- Precisamos dos dois primeiros momentos *teóricos*:
  - $\mathbb{E}(Y) = \alpha/\beta$
  - $\mathbb{E}(Y^2) = \alpha(\alpha+1)/\beta^2$
- Precisamos também dos dois primeiros momentos *amostrais*, inteiramente baseados nos dados:
  - $m_1 = \sum_i Y_i/n$
  - $m_2 = \sum_i Y_i^2/n$

Método de Momentos: gama

- Igualamos os momentos correspondentes, teóricos e amostrais:
  - $\mathbb{E}(Y) = m_1$
  - $\mathbb{E}(Y^2) = m_2$
- Isto é:

$$\begin{cases} \frac{\alpha}{\beta} = m_1 \\ \frac{\alpha(\alpha+1)}{\beta^2} = m_2 \end{cases}$$

- Em seguida, resolvemos para α e β. Encontramos a solução:
  - $\widehat{\alpha} = m_1^2/(m_2 - m_1^2)$
  - $\widehat{\beta} = m_1/(m_2 - m_1^2)$
- Estes são os estimadores de momentos de $\boldsymbol{\theta} = (\alpha, \beta)$.
- O que seria o MLE neste caso de v.a.'s iid gama?
- MLE não dá o mesmo resultado que o método de momentos.

MLE: gama

- Log-verossimilhança:

$$\ell(\alpha, \beta) = n\beta \log(\alpha) - n \log(\Gamma(\alpha)) + (\alpha - 1) \sum_i \log(y_i) - \beta \sum_i y_i$$

- Derivando:

$$\begin{cases} \frac{\partial \ell}{\partial \beta} & = & \frac{n\alpha}{\beta} - \sum_i y_i \\ \frac{\partial \ell}{\partial \alpha} & = & n \log(\beta) - n \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} + \sum_i \log(y_i) \end{cases}$$

- Igualando a zero a derivada em β produz $\widehat{\beta} = \widehat{\alpha}/\overline{y}$

MLE: gama

- Igualando a zero a derivada em α produz

$$n \log(\beta) - n\Gamma'(\alpha)/\Gamma(\alpha) + \sum_i \log(y_i) = 0$$

- Substituindo $\beta = \alpha/\bar{y}$ teremos

$$n \log(\alpha) - n \log(\bar{y}) - n\Gamma'(\alpha)/\Gamma(\alpha) + \sum_i \log(y_i) = 0$$

- Precisa de solução numérica.
- MLE difere do estimador de momentos neste caso.

Justificativa do método de momentos

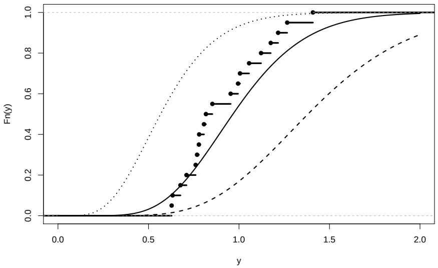
- Pela lei dos grandes números, $m_1 = \sum_i Y_i/n$ converge para $\mathbb{E}(Y)$ quando $n \to \infty$.
- Assim, se o tamanho n da amostra é grande, podemos esperar $m_1 \approx \mathbb{E}(Y)$.
- Pela mesma lei dos grandes números, $m_k = \sum_i Y_i^k/n$ converge para $\mathbb{E}(Y^k)$.
- Assim, $m_2 \approx \mathbb{E}(Y^2)$
- $m_3 \approx \mathbb{E}(Y^3)$, etc.
- Trocamos a aproximação por uma igualdade para obter estimativa de momentos.

Vários outros métodos

- Um terceiro método é baseado na estatística de Kolmogorov.
- Considere o modelo de v.a.'s i.i.d. Gama(α, β)
- Construa a função acumulada empírica com os dados observados.
- Para cada α e β fixos, podemos calcular a distribuição acumulada empírica teórica de uma Gama(α, β).

Baseado em Kolmogorov

- Gráfico com acumulada empírica $\mathbb{F}_n(y)$ de amostra com $n = 20$ dados
- E três acumuladas teóricas de distribuição gama com diferentes parâmetros α e β:
- Gama(10,10), Gama(10, 7) e Gama(6,10)



Minimizando distância de Kolmogorov

- Para cada $\boldsymbol{\theta} = (\alpha, \beta)$, obtenha $F(y; \boldsymbol{\theta})$.
- Calcule a distância de Kolmogorov:

$$D_n(\boldsymbol{\theta}) = \max_y |\mathbb{F}_n(y) - F(y; \boldsymbol{\theta})|$$

- $D_n(\boldsymbol{\theta})$ é função dos dados e de $\boldsymbol{\theta}$. Com os dados fixos, é função apenas de $\boldsymbol{\theta}$.
- Obtenha agora o valor de $\boldsymbol{\theta} = (\alpha, \beta)$ que minimiza $D_n(\boldsymbol{\theta})$:

$$\widehat{\boldsymbol{\theta}} = (\widehat{\alpha}, \widehat{\beta}) = \arg\min_y D_n(\boldsymbol{\theta})$$

- Intuitivo também, não?
- Vai dar um resultado diferente de momentos e de MLE.

Um método sem nome

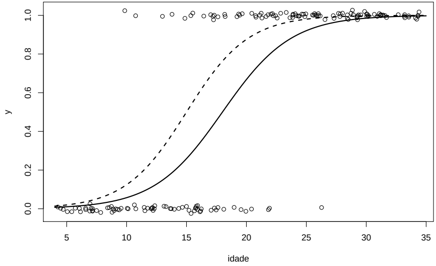
- Lembre do modelo de regressão logística com uma única covariável e $\boldsymbol{\theta} = (\beta_0, \beta_1)$

$$\mathbb{P}(Y_i = 1) = \frac{1}{1 + \exp(-(\beta_0 + \beta_1 x_i))} = p_i(\boldsymbol{\theta})$$

- Vimos o MLE para $\boldsymbol{\theta} = (\beta_0, \beta_1)$
- Um outro método, simples e intuitivo, é o seguinte.
- Chute um valor qualquer para $\boldsymbol{\theta} = (\beta_0, \beta_1)$ e construa a curva logística.
- Compare com os dados e ache $\boldsymbol{\theta}$ que seja ótimo

Duas curvas logísticas

- Dados (x_i, y_i) de $n = 173$ crianças e duas possíveis curvas logísticas com $\boldsymbol{\theta} = (-6.3, -0.35)$ e $\boldsymbol{\theta} = (-5.85, -0.39)$



- Calcule $D(\boldsymbol{\theta}) = \sum_i |y_i - p_i(\boldsymbol{\theta})|$
- Ache $\boldsymbol{\theta}$ que minimize esta função.

Outro critério para escolher curva logística

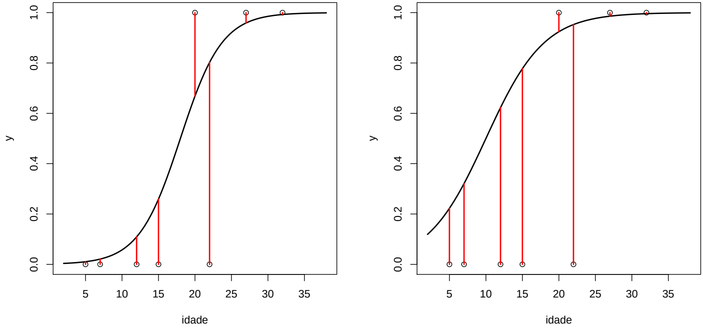


Figura: Duas possíveis curvas com os valores de $|y_i - p_i(\theta)|$

Ainda outro critério

- Podemos inventar muitos critérios razoáveis.
- Por exemplo:
  - Encontre $\boldsymbol{\theta}$ que minimize $D(\boldsymbol{\theta}) = \sum_i (y_i - p_i(\boldsymbol{\theta}))^2$
  - Encontre $\boldsymbol{\theta}$ que minimize $D(\boldsymbol{\theta}) = \sum_i \left(p_i(\boldsymbol{\theta})(1 - p_i(\boldsymbol{\theta})) \, |y_i - p_i(\boldsymbol{\theta})|\right)$. Veja que $p(\boldsymbol{\theta})(1 - p(\boldsymbol{\theta})) \approx 0$ se $p(\boldsymbol{\theta}) \approx 0$ ou ≈ 1. Assim, damos mais peso aos pontos no "centro", aqueles em que $p(\boldsymbol{\theta}) \approx 1/2$.
- Podemos inclusive misturar estimadores obtidos por métodos diferentes.
  - Suponha que $\widehat{\boldsymbol{\theta}}_{MLE}$ seja o estimador MLE na regressão logística
  - Seja $\widehat{\boldsymbol{\theta}}_D$ o estimador que minimiza $D(\boldsymbol{\theta}) = \sum_i |y_i - p_i(\boldsymbol{\theta})|$
  - Defina então seu estimador preferido como $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\theta}}_{MLE} + \widehat{\boldsymbol{\theta}}_D)/2$, a média aritmética de $\widehat{\boldsymbol{\theta}}_{MLE}$ e $\widehat{\boldsymbol{\theta}}_D$.
  - Podemos usar outra média ponderada qualquer: $\widehat{\boldsymbol{\theta}} = 0.75 \, \widehat{\boldsymbol{\theta}}_{MLE} + 0.25 \, \widehat{\boldsymbol{\theta}}_D$.
- Lição: podemos pensar em INFINITOS estimadores, todos aparentemente razoáveis. Como escolher?