GET00211 - Modelos Lineares 2

Rafael Erbisti

Instituto de Matemática e Estatística

Universidade Federal Fluminense

Aula 05

Inferência

- Geralmente, um procedimento estatístico envolve três passos:
  1. Especificação do modelo
  2. Estimação dos parâmetros
  3. Inferência ("propriamente dita"): teste de hipóteses, intervalos de confiança, adequação do modelo etc.

- Hoje, iniciaremos o passo 3.

- Para realizar inferência é necessário conhecer a distribuição amostral dos estimadores e das estatísticas de qualidade de ajuste.

Inferência

- No caso Normal, sabe-se que as distribuições amostrais podem ser determinadas exatamente.
- Para outras distribuições, é preciso basear-se em resultados assintóticos para grandes amostras, com base no Teorema Central do Limite.
- Esses resultados requerem atenção e a verificação de certas condições de regularidade.
- Para amostras independentes de variáveis com distribuição na família exponencial e, em particular, para os Modelos Lineares Generalizados, essas condições são satisfeitas.

Inferência

- A ideia é que, sob certas condições, se S é uma estatística de interesse, então, para amostras grandes, tem-se assintoticamente que:

$$\frac{S - E(S)}{\sqrt{V(S)}} \;\dot{\sim}\; N(0,1)$$

Ou, de forma equivalente:

$$\frac{[S - E(S)]^2}{V(S)} \;\dot{\sim}\; \chi^2(1)$$

- Assumindo que **S** é um vetor de estatísticas de interesse de dimensão p, com:
  - $E(\mathbf{S})$ a esperança de **S**,
  - **V** a matriz de covariâncias de **S**.

Então, assintoticamente:

$$Q = [\mathbf{S} - E(\mathbf{S})]' \, \mathbf{V}^{-1} \, [\mathbf{S} - E(\mathbf{S})] \;\dot{\sim}\; \chi^2(p)$$

Exemplo: médias da Exponencial

- Sejam p variáveis aleatórias independentes: $X_1, \ldots, X_p \sim \text{Exp}(\lambda)$, com médias $E(X_j) = \frac{1}{\lambda}$ e variâncias $V(X_j) = \frac{1}{\lambda^2}$, para $j = 1, \ldots, p$.

- Dada amostras de tamanho n, as médias amostrais são:

$$S_j = \bar{X}_j = \frac{1}{n} \sum_{i=1}^{n} X_{ji}, \quad j = 1, \ldots, p.$$

- Pelo Teorema Central do Limite, para n grande:

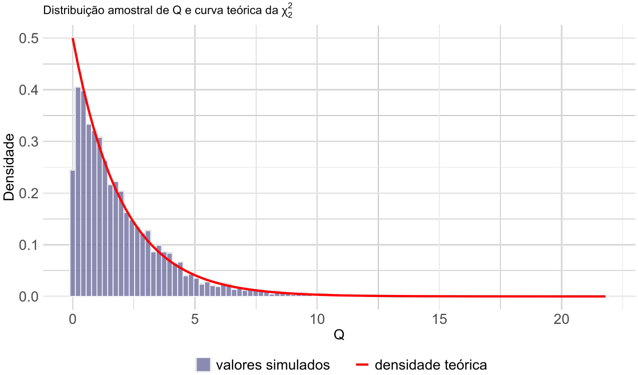
$$\bar{X}_j \;\dot{\sim}\; N\left(\frac{1}{\lambda}, \frac{1}{n\lambda^2}\right)$$

- Como as variáveis são independentes, a matriz de covariância teórica das médias amostrais é:

$$V = \frac{1}{n\lambda^2} \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} = \frac{1}{n\lambda^2} \mathbf{I}_p$$

Exemplo: médias da Exponencial

- Supondo $\lambda = 1$
- Número de estatísticas: $p = 2$
- Tamanho da amostra: $n = 60$ (suficiente para o TCL)
- Número de simulações: 5000



Código R disponível AQUI.

Distribuição Amostral para Estatística Escore

- Já vimos que a estatística escore é dada por:

$$U_j = \frac{\partial \ell}{\partial \beta_j}, \quad j = 1, \ldots, p$$

- Além disso, a esperança do escore é zero:

$$E(U_j) = 0.$$

- A matriz de covariâncias da estatística escore é a matriz informação de Fisher, dada por

$$\boldsymbol{\mathcal{I}} = E(\mathbf{U}\mathbf{U}')$$

com elementos

$$\mathcal{I}_{jk} = E(U_j U_k).$$

Distribuição Amostral da Estatística Escore

Portanto, para grandes amostras, pelo Teorema Central do Limite:

- Caso de um único parâmetro β:

$$\frac{U}{\sqrt{\mathcal{I}}} \;\dot{\sim}\; N(0,1) \quad \text{ou, equivalentemente,} \quad \frac{U^2}{\mathcal{I}} \;\dot{\sim}\; \chi^2(1)$$

- Caso de vetor de parâmetros $\boldsymbol{\beta}$ de dimensão $p \times 1$:

$$\boldsymbol{U} = (U_1, \ldots, U_p)' \;\dot{\sim}\; N_p(\boldsymbol{0}, \boldsymbol{\mathcal{I}})$$

e a estatística quadrática Q é dada por

$$\boldsymbol{U}'\boldsymbol{\mathcal{I}}^{-1}\boldsymbol{U} \;\dot{\sim}\; \chi^2_p$$

Aproximação por série de Taylor: Conceito Básico

A série de Taylor aproxima uma função $f(x)$ usando um polinômio em torno de um ponto a:

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f^{(3)}(a)}{3!}(x-a)^3 + \dots$$

- $f(a) \to$ valor da função no ponto a
- $f'(a), f''(a), \dots \to$ derivadas no ponto a
- $(x-a) \to$ distância até o ponto de expansão
- $n! \to$ fatorial que divide cada termo

Se $a = 0$, chamamos de série de Maclaurin (caso particular da série de Taylor).

Aproximação por série de Taylor: Exemplo

Passo 1: Derivadas

$$f(x) = e^x, \quad f'(x) = e^x, \quad f''(x) = e^x, \quad f^{(n)}(x) = e^x$$

No ponto $a = 0$:

$$f^{(n)}(0) = e^0 = 1$$

Passo 2: Montar a série

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \dots$$

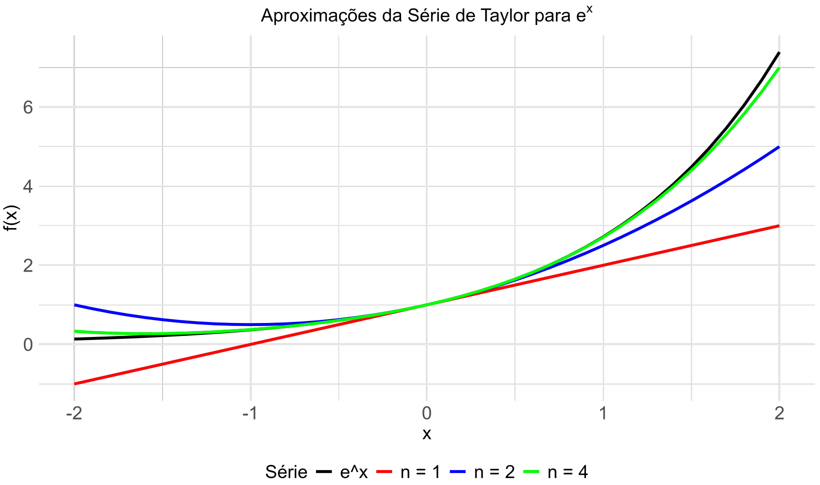
Passo 3: Aproximação numérica

Para $x = 1$:

$$e^1 \approx 1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24} = 2,7084$$

Valor real: $e \approx 2,7183$. Com apenas quatro termos, já temos um bom resultado.

Aproximação por série de Taylor: Exemplo



Distribuição Amostral para o EMV

- Suponha que a log-verossimilhança possua apenas um máximo $\hat{\boldsymbol{\beta}}$ e que esse valor estimado esteja próximo de $\boldsymbol{\beta}$.

- A aproximação de primeira ordem da expansão de Taylor para o vetor de escores $\boldsymbol{U}(\boldsymbol{\beta})$ em torno de $\hat{\boldsymbol{\beta}}$ é dada por:

$$\boldsymbol{U}(\boldsymbol{\beta}) \approx \boldsymbol{U}(\hat{\boldsymbol{\beta}}) + \boldsymbol{H}(\hat{\boldsymbol{\beta}})\,(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})$$

  onde $\boldsymbol{H}(\hat{\boldsymbol{\beta}})$ denota a matriz da segunda derivada da log-verossimilhança avaliada em $\hat{\boldsymbol{\beta}}$.

- Se aproximarmos a matriz Hessiana em $\hat{\boldsymbol{\beta}}$ pela sua esperança, temos que $\boldsymbol{H}(\hat{\boldsymbol{\beta}}) \approx E[\boldsymbol{H}(\hat{\boldsymbol{\beta}})] = -\boldsymbol{\mathcal{I}}$, então a expansão de Taylor do vetor escore se torna:

$$\boldsymbol{U}(\boldsymbol{\beta}) \approx \boldsymbol{U}(\hat{\boldsymbol{\beta}}) - \boldsymbol{\mathcal{I}}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})$$

- Como $\boldsymbol{U}(\hat{\boldsymbol{\beta}}) = \mathbf{0}$, temos:

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \approx \boldsymbol{\mathcal{I}}^{-1}\boldsymbol{U}(\boldsymbol{\beta})$$

  dado que $\boldsymbol{\mathcal{I}}$ é não singular.

Distribuição Amostral para o EMV

- Se a matriz informação de Fisher $\mathcal{I}$ for considerada constante, temos:

$$E\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) = \mathcal{I}^{-1}\, E(\boldsymbol{U}) = \mathbf{0}$$

então $\hat{\boldsymbol{\beta}}$ é um estimador assintoticamente não viesado de $\boldsymbol{\beta}$.

- A matriz de covariâncias assintótica de $\hat{\boldsymbol{\beta}}$ é:

$$\begin{aligned}
V(\hat{\boldsymbol{\beta}}) &= E\left[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\right] \\
&= E\left[\mathcal{I}^{-1}\boldsymbol{U}\boldsymbol{U}'\mathcal{I}^{-1}\right] \\
&= \mathcal{I}^{-1}\, E\left[\boldsymbol{U}\boldsymbol{U}'\right]\, \mathcal{I}^{-1} \\
&= \mathcal{I}^{-1}
\end{aligned}$$

Distribuição Amostral para o EMV

- Para amostras grandes, a estatística quadrática

$$(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \boldsymbol{\mathcal{I}} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \ \dot{\sim} \ \chi^2_p$$

  é conhecida como Estatística de Wald.

- Para o caso de um único parâmetro ($p = 1$), o estimador EMV tem distribuição aproximadamente normal:

$$\hat{\beta} \ \dot{\sim} \ N\left(\beta, \mathcal{I}^{-1}\right)$$

- Se a variável resposta em um Modelo Linear Generalizado for normal, então os resultados acima são exatos, e não apenas assintóticos.

Intervalos de Confiança para EMV

- A distribuição amostral assintótica do EMV de $\boldsymbol{\beta}$ é: $\hat{\boldsymbol{\beta}} \ \dot{\sim} \ N_p\big(\boldsymbol{\beta}, \boldsymbol{\mathcal{I}}^{-1}\big)$.
- O erro padrão da estimativa $\hat{\beta}_j$ é dado por:

$$\mathrm{SE}(\hat{\beta}_j) = \sqrt{v_{jj}}$$

onde v_{jj} é o j-ésimo elemento da diagonal da matriz $\boldsymbol{\mathcal{I}}^{-1}$.

- Intervalos de confiança individuais para os parâmetros:

$$\hat{\beta}_j \pm z_{1-\alpha/2}\,\mathrm{SE}(\hat{\beta}_j)$$

- A correlação entre dois estimadores $\hat{\beta}_j$ e $\hat{\beta}_k$ pode ser calculada como:

$$\mathrm{corr}(\hat{\beta}_j, \hat{\beta}_k) = \frac{v_{jk}}{\sqrt{v_{jj}\, v_{kk}}}$$

- Observação: a matriz de informação depende de $\boldsymbol{\beta}$. Em aplicações práticas, substituímos $\boldsymbol{\beta}$ por $\hat{\boldsymbol{\beta}}$.

Exemplos

Exemplo 1: Sejam $Y_1, \ldots, Y_n$ variáveis iid com $Y_i \sim N(\mu, \sigma^2)$, com σ^2 conhecido.

a. Obtenha o EMV para μ.
b. Determine a distribuição amostral da estatística escore e verifique que o resultado é exato.

Exemplo 2: Seja $Y \sim \text{Bin}(n, \pi)$. Obtenha a estatística escore e sua distribuição amostral aproximada.

Sob o paradigma Bayesiano, como seria o processo de inferência?