Instituto de Educação Superior de Brasília
A Teoria e a Prática Juntas

INFERÊNCIA ESTATÍSTICA CONCEITOS INTERMEDIÁRIOS (Unidade 2)

Prof. Natália Ribeiro S. Evangelista

natalia.evangelista@iesb.br

ESTATÍSTICA SUFICIENTE

- Uma estatística T=T(Y) é suficiente para um parâmetro θ quando resume toda informação sobre esse parâmetro contida na amostra Y. Se T é suficiente para θ, então, a distribuição condicional Y dada a estatística T(Y) é independente de θ, isto é:

$$\mathrm{P}(\mathbf{Y} = \mathbf{y} | T = t, \theta) = \mathrm{P}(\mathbf{Y} = \mathbf{y} | T = t).$$

- Ou seja, se X e Y são duas amostras tais que T(x)=T(y), então a inferência sobre θ é a mesma independente da observação X ou Y.

Prof. Natália Ribeiro S. Evangelista

natalia.evangelista@iesb.br

Inferência Estatística (Conceitos Intermediários)

Critério da fatoração e estatísticas conjuntamente suficientes

- O critério da fatoração é uma forma conveniente de caracterizar uma estatística suficiente. Uma condição necessária e suficiente para T ser suficiente para um parâmetro θ é que a função (densidade ou de probabilidade) $f_Y(y; \theta)$ possa ser decomposta como

$$f_{\mathbf{Y}}(\mathbf{y};\theta) = h(\mathbf{y})g(t,\theta),$$

- Em que $t=T(y)$ e $h(y)$ não dependem de θ. Esse resultado é válido para o caso contínuo e discreto.
- O conceito de estatística suficiente é importante em estatística no sentido de se poder concentrar a informação sobre o parâmetro. Formalmente tem: se Y_1, Y_2, ..., Y_n é uma amostra de uma densidade $f(y; \theta)$, em que θ é um vetor de parâmetros, as estatísticas S_1, S_2, ..., S_r são <u>conjuntamente suficientes</u> se, e somente se, a distribuição conjunta de Y_1, Y_2, ..., Y_n, dadas $S_1 = s_1$, $S_2 = s_2$, ..., $S_r = s_r$ não depende de θ.

Prof. Natália Ribeiro S. Evangelista
natalia.evangelista@iesb.br

Família exponencial

- Família exponencial de distribuições: conjunto de distribuições conhecidas inseridas em uma família paramétrica.
- Exemplos de distribuições:
  - Normal;
  - Binomial;
  - Binomial Negativa;
  - Gama;
  - Poisson;
  - Normal Inversa;
  - Multinomial;
  - Beta;
  - Logarítmica, etc.
- A importância da família exponencial de distribuições teve maior destaque, na área dos modelos de regressão, a partir do trabalho pioneiro de Nelder e Wedderburn (1972) que definiram os modelos lineares generalizados (MLG). Na teoria da Inferência Estatística auxilia nas diversas demonstrações das propriedades dos estimadores.

Prof. Natália Ribeiro S. Evangelista
natalia.evangelista@iesb.br

Família exponencial Uniparamétrica

- A família exponencial uniparamétrica é caracterizada por uma função (de probabilidade ou densidade) da seguinte forma:

$$f(x;\theta) = h(x)\ \exp\left[\ \eta(\theta)\ t(x) - b(\theta)\ \right]. \quad (1.1)$$

onde: $\eta(\theta)$, $b(\theta)$, $t(x)$ e $h(x)$ pertencem aos subconjuntos dos reais.

- Diversas distribuições conhecidas e importantes podem ser estruturadas de acordo com a função supra. Exemplos:
  - Distribuição de Poisson
  - Distribuição Binomial
  - Distribuição Normal

Prof. Natália Ribeiro S. Evangelista
natalia.evangelista@iesb.br

- Exemplo da distribuição de Poisson:

A distribuição de Poisson $P(\theta)$ de parâmetro $\theta > 0$, usada para análise de dados na forma de contagens, tem função de probabilidade

$$f(x;\theta) = \frac{e^{-\theta}\theta^x}{x!} = \frac{1}{x!}\exp[x\log(\theta) - \theta]$$

e, portanto, é um membro da família exponencial (1.1) com $\eta(\theta) = \log(\theta)$, $b(\theta) = \theta$, $t(x) = x$ e $h(x) = 1/x!$.

Prof. Natália Ribeiro S. Evangelista
natalia.evangelista@iesb.br

- A família exponencial multiparamétrica de dimensão k é caracterizada por uma função (de probabilidade ou densidade) da forma

$$f(\mathbf{x};\boldsymbol{\theta}) = h(\mathbf{x}) \exp\left[\sum_{i=1}^{k} \eta_i(\boldsymbol{\theta}) t_i(\mathbf{x}) - b(\boldsymbol{\theta})\right]$$ equivalente à $$f(\mathbf{x};\boldsymbol{\theta}) = h(\mathbf{x}) \exp\left[\sum_{i=1}^{k} \theta_i t_i(\mathbf{x}) - b(\boldsymbol{\theta})\right]$$

em que θ é um vetor de parâmetros, usualmente, de dimensão k, e as funções $\eta_i(\theta)$, $b(\theta)$, $t_i(x)$ e $h(x)$ têm valores em subconjuntos dos reais.

- A forma abaixo é um caso especial:

$$f(x;\theta) = h(x) \; \exp\left[\, \eta(\theta)\; t(x) - b(\theta) \,\right].$$

- Diversas distribuições conhecidas são pertencentes à família exponencial multiparamétrica, dentre elas: Normal, Gama Beta, Weibull e Multinomial.

- Definição: a obtenção de estimadores pelo método conhecido por Máxima Verossimilhança (M.V) parte do princípio de que devemos escolher o(s) valor(es) do(s) parâmetro(s) desconhecido(s) que maximiza a probabilidade de obter a amostra particular observada, ou seja, o(s) valor(es) do(s) parâmetro(s) que torna aquela amostra a "mais provável".
- Exemplo: consideremos a função de probabilidade de uma v.a. Binomial a qual é dada por

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \ldots, n.$$

Quando nós passamos a observar esta função não como dependente da amostra x(total de sucessos) mas sim como função do parâmetro p, temos o que denominamos função de verossimilhança, L(p), a qual deve ser maximizada para assim obtermos o estimador de máxima verossimilhança. Em geral, e por facilidade, nós maximizamos a log-verossimilhança a qual é o logaritmo neperiano da função de verossimilhança denotada por $l(p) = \ln L(p)$.

Prof. Natália Ribeiro S. Evangelista
natalia.evangelista@iesb.br

- Seja $(X_1, X_2, \ldots, X_n)$ uma amostra aleatória da variável aleatória $X \sim \text{Poisson}(\theta)$, $\theta > 0$. Encontre o Estimador de Máxima Verossimilhança (E.M.V.) de θ.

Nesse caso a função de verossimilhança é dada por

$$L(\theta|x) = \prod_{i=1}^{n} f(x_1, \ldots, x_n|\theta) = \prod_{i=1}^{n} \frac{\theta^{x_i} e^{-\theta}}{x_i!} = \frac{\theta^{\sum x_i} e^{-n\theta}}{\prod x_i!}$$

O logaritmos da função de verossimilhança é dado por:

$$l(\theta|x) = -n\theta + \sum_{i=1}^{n} x_i \ln(\theta) - \sum_{i=1}^{n} ln x_i!$$

Prof. Natália Ribeiro S. Evangelista
natalia.evangelista@iesb.br

MÉTODO DE MÁXIMA VEROSSIMILHANÇA

O estimador de máxima verossimilhança é obtido aplicando a 1ª derivada e igualando a 0 (zero). Isto é:

$$l(\theta|x) = -n + \frac{\sum_{i=1}^{n} x_i}{\hat{\theta}} = 0 \rightarrow \hat{\theta} = \bar{x}$$

Agora é preciso verificar se o resultado encontrado é um ponto de máximo:

$$\frac{d^2 l(\theta|x)}{\theta} = \frac{-\sum_{i=1}^{n} x_i}{\hat{\theta}^2} \vdots \, \theta = \bar{x} < 0$$

Logo, $\hat{\theta} = \bar{x}$ é um EMV.

Prof. Natália Ribeiro S. Evangelista

natalia.evangelista@iesb.br

Método dos momentos

- Seja uma a.a. X' = $[X_1, X_2, \ldots, X_n]$ de uma população com f.p. ou f.d.p. dependendo de k parâmetros $\theta_1, \theta_2, \ldots \theta_k$. Os momentos ordinários da população $m_j = \int_{-\infty}^{\infty} x^j f(x, \theta_1, \ldots, \theta_k) dx$, se existirem, são funções dos k parâmetros $m_j = f(\theta_1, \theta_2, \ldots \theta_k)$, j = 1,2,3, ... Considere, também, os momentos ordinários da amostra, $M_j = \frac{\sum x^j}{n}$, j = 1,2,3, ..., forme o sistema de equações:

$$M_j = m_j = f(\theta_1, \theta_2, \ldots \theta_k) \text{ , } j = 1,2,3, \ldots$$

e admita que tem solução única, $\hat{\theta}_j(X_1, X_2, \ldots, X_n)$, j = 1,2,3, ...,k. Então, estes k estimadores, solução do sistema de equações, são os estimadores dos parâmetros pelo Método dos Momentos.

Os estimadores obtidos pelo Método dos Momentos são em geral consistentes e possuem distribuição assintótica Gaussiana, porém não são assintoticamente mais eficientes do que os estimadores de máxima verossimilhança.

Prof. Natália Ribeiro S. Evangelista
natalia.evangelista@iesb.br

Teorema de Rao Blackwell

Instituto de Educação Superior de Brasília
A Teoria e a Prática Juntas

- Seja $[X_1, X_2, \ldots, X_n]$ uma a.a. da densidade $f(.,\theta)$, e seja $S_1 = s_1(X)$, $S_2 = s_2(X)$, ... ,$S_k = s_k(X)$ um conjunto de estatísticas conjuntamente suficientes. Seja a estatística $T = t(X)$ um estimador não-viciado de $q(\theta)$. Defina T' por $T' = E[T|S_1, S_2, \ldots, S_k]$, então:
  - T' é uma estatística e é uma função de estatísticas suficientes $S_1, S_2, \ldots S_k$;
  - $E_\theta[T'] = q(\theta)$, isto é T' é um estimador não-viciado de $q(\theta)$;
  - $V_\theta[T'] \leq V(T)$ qualquer que seja θ e $V_\theta[T'] < V(T)$ para algum θ a menos que T seja igual a T' com probabilidade 1.

Prof. Natália Ribeiro S. Evangelista

natalia.evangelista@iesb.br

- Se $T(X)$ é uma estatística suficiente e completa e $S(X)$ é um estimador não-viciado de $q(\theta)$, então $T^*(X) = E[S(X) | T(X)]$ é um estimador UMVU de $q(\theta)$. Se $V_\theta(T^*(X)) < \infty$ qualquer que seja θ, $T^*(X)$ é o único estimador UMVU de $q(\theta)$.
- Podemos usar o Teorema de Lehmann-Scheffé na busca de estimadores UMVU de dois modos:
  - Se nós podemos achar uma estatística da forma $h[T(X)]$ tal que $h[T(X)]$ seja um estimador não-viciado de $q(\theta)$, então $h[T(X)]$ é UMVU para $q(\theta)$.
  - Se nós podemos achar algum estimador não-viciado $S(X)$ de $q(\theta)$, então $E[S(X) | T(X)]$ é UMVU para $q(\theta)$.

UMVU: ESTIMADORES NÃO-VICIADOS UNIFORMEMENTE DE MÍNIMA VARIÂNCIA.

Prof. Natália Ribeiro S. Evangelista

natalia.evangelista@iesb.br

EXERCÍCIO

LISTA DE EXERCÍCIOS 1 – parte II

Data limite de entrega: vide cronograma em “Sistema de avaliação”

Prof. Natália Ribeiro S. Evangelista

natalia.evangelista@iesb.br