Fundação Getúlio Vargas
Inferência Estatística
Wellington José

Abstract of Titica
Due: October 1, 2021

Resumo Titica

Aula 1 - O que é e para que serve Inferência Estatística?

Definição 1 (Modelo estatístico: informal) *Um **modelo estatístico** consiste na identificação de variáveis aleatórias de interesse (observáveis e potencialmente observáveis), na especificação de uma distribuição conjunta para as variáveis aleatórias observáveis e na identificação dos parâmetros* (θ) *desta distribuição conjunta. Às vezes é conveniente assumir que os parâmetros são variáveis aleatórias também, mas para isso é preciso especificar uma distribuição conjunta para* θ.

Definição 2 (Modelo estatístico: formal) *Seja* $\mathcal{X}$ *um espaço amostral qualquer,* Θ *um conjunto não-vazio arbitrário e* $\mathcal{P}(\mathcal{X})$ *o conjunto de todas as distribuições de probabilidade em* $\mathcal{X}$. *Um modelo estatístico **paramétrico** é uma função* $P : \Theta \to \mathcal{P}(\mathcal{X})$ *que associa a cada* $\theta \in \Theta$ *uma distribuição de probabilidade* P_θ *em* $\mathcal{X}$.

Definição 3 (Afirmação probabilística) *Dizemos que uma afirmação é **probabilística** quando ela utiliza conceitos da teoria de probabilidade para falar de um objeto.*

Definição 4 (Inferência Estatística) *Uma **inferência estatística** é uma **afirmação probabilística** sobre uma ou mais partes de um modelo estatístico.*

Definição 5 (Estatística) *Suponha que temos uma coleção de variáveis aleatórias* $X_1, X_2, \dots, X_n \subseteq \mathbf{R}^n$ *e uma função* $r : \mathbf{X} \to R^m$. *Dizemos que a variável aleatória* $T = r(X_1, X_2, \dots, X_n)$ *é uma **estatística**.*

Definição 6 (Permutabilidade) *Uma coleção finita de variáveis aleatórias* $X_1, X_2, \dots, X_n$ *com densidade conjunta* f *é dita **permutável** se*

$$f(x_1, x_2, \dots, x_n) = f(x_{\pi(1)}, x_{\pi(2)}, \dots, x_{\pi(n)}) \tag{1}$$

para qualquer permutação $\pi = \{\pi(1), \pi(2), \dots, \pi(n)\}$ *dos seus elementos. Uma coleção finita é permutável se qualquer subconjunto finito é permutável.*

Aula 2 - Distribuição a priori e a posteriori

Definição 7 (Distribuição a priori) *Se tratamos o parâmetro* θ *como uma variável aleatória, então a **distribuição a priori** é a distribuição que damos a* θ *antes de observarmos as outras variáveis aleatórias de interesse. Vamos denotar a função de densidade/massa de probabilidade da priori por* $\xi(\theta)$.

Definição 8 (Distribuição a posteriori) *Considere o problema estatístico com parâmetros* θ *e variáveis aleatórias observáveis* $X_1, X_2, \dots, X_n$. *A distribuição condicional de* θ *dados os valores observados das variáveis aleatórias,* $\boldsymbol{x} := \{x_1, x_2, \dots, x_n\}$ *é a **distribuição a posteriori** de* θ*, denotamos por* $\xi(\theta \mid \boldsymbol{x})$ *a f.d.p./f.m.p. condicional a* $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$.

Teorema 1 (Distribuição a posteriori: derivação) *Considere a amostra aleatória $X_1, X_2, \ldots, X_n$ de uma distribuição com f.d.p./f.m.p. $f(x \mid \theta)$. Se a distribuição a priori é $\xi(\theta)$, temos*

$$\xi(\theta \mid x) = \frac{\xi(\theta)\Pi_{i=1}^{n} f(x_i \mid \theta)}{g_n(x)}, \ \theta \in \Omega \tag{2}$$

Chamamos $g_n(x)$ de distribuição marginal de $X_1, X_2, \ldots, X_n$.

Definição 9 (Função de verossimilhança) *Quando encaramos a f.d.p./f.m.p. $f(x_1, x_2, \ldots, x_n \mid \theta)$ como uma função do parâmetro θ, chamamos esta função de **função de verossimilhança**, e podemos denotá-la como $L(\theta; x)$ ou, quando a notação não criar ambiguidade, simplesmente $L(\theta)$.*

Aula 3 - Prioris conjugadas e função de perda

Definição 10 (Hiper-parâmetros) *Seja $\xi(\theta \mid \phi)$ a distribuição a priori para o parâmetro θ, indexada por $\phi \in \Phi$. Dizemos que ϕ é(são) o(s) **hiper-parâmetro(s)** da priori de θ.*

Definição 11 (Priori conjugada) *Suponha que $X_1, X_2, \ldots$ sejam condicionalmente independentes dado θ, com f.d.p./f.m.p. $f(x \mid \theta)$. Defina*

$$\Psi = \left\{ f : \Omega \to (0, \infty), \int_{\Omega} f dx = 1 \right\} \tag{3}$$

*onde Ω é o espaço de parâmetros. Dizemos que Ψ é uma **família de distribuições conjugadas** para $f(x \mid \theta)$ se $\forall f \in \Psi$ e toda realização $\boldsymbol{x}$ de $X = X_1, X_2, \ldots, X_n$*

$$\frac{f(\boldsymbol{x} \mid \theta) f(\theta)}{\int_{\Omega} f(\boldsymbol{x} \mid \theta) f(\theta) d\theta} \in \Psi \tag{4}$$

Teorema 2 (Distribuição a posteriori da média de uma normal) *Suponha que $X_1, X_2, \ldots, X_n$ formam uma amostra aleatória com distribuição normal e com média desconhecida θ e variância $\sigma^2 > 0$, conhecida e fixa. Suponha que $\theta \backsim Normal(\mu_0, v_0^2)$ a priori. Então*

$$\xi(\theta \mid x, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(\frac{(\theta - \mu_1)^2}{2v_1^2} \right), \tag{5}$$

onde

$$\mu_1 := \frac{\sigma^2 \mu_0 + n v_0^2 \overline{x}_n}{\sigma^2 + n v_0^2} \ e \ v_1^2 := \frac{\sigma^2 v_0^2}{\sigma^2 + n v_0^2} \tag{6}$$

Definição 12 (Priori imprópria) *Seja $\xi : \Lambda \to (0, \infty), \Omega \subseteq \Lambda$, uma função tal que $\int_{\Omega} \xi(\theta) d\theta = \infty$. Se utilizamos ξ como uma p.d.f.* [1] *para θ, dizemos que ξ é uma **priori imprópria** para θ.*

Definição 13 (Estimador) *Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias com distribuição conjunta indexada por θ. Um **estimador** de θ é qualquer função real δ: $X_1, X_2, \ldots, X_n \to \mathbb{R}^d, d \geq 1$.*

Definição 14 (Estimativa) *Dizemos que o valor de δ avaliado nas realizações de $X_1, X_2, \ldots, X_n$, $\boldsymbol{x} = \{x_1, x_2, \ldots, x_n\}$, $\delta(\boldsymbol{x})\}$ é uma **estimativa** de θ.*

Definição 15 (Função de perda) *Uma função de perda é uma função real em duas variáveis*

$$L : \Omega \times \mathbb{R}^d \to \mathbb{R}, \tag{7}$$

em que dizemos que o estatístico perde $L(\theta, a)$ se o parâmetro vale θ e a estimativa dada vale a.

[1] p.d.f. - "probability density function" ou função de densidade de probabilidade

Aula 4 - Estimadores de Bayes e EMV

Definição 16 (Estimador de Bayes) *Considere a perda esperada a posteriori:*

$$E_{\theta|x}[L(\theta,a)] = E[L(\theta,a) \mid x] = \int_{\Omega} L(\theta,a)\xi(\theta \mid x)d\theta \tag{8}$$

*Dizemos que δ^* é um **estimador de Bayes** se, para toda realização $X = x$,*

$$E[L(\theta,\delta^*(x)) \mid x] = \min_{a \in A} E[L(\theta,a) \mid x]. \tag{9}$$

Em outras palavras, um estimador de Bayes é uma função real dos dados que minimiza a perda esperada com respeito à posteriori dos parâmetros.

Teorema 3 (δ^* sob perda quadrática) *Seja θ um parâmetro tomando valores reais. Sob perda quadrática,*

$$\delta^*(x) = E[\theta \mid X = x] = \int_{\Omega} \theta\xi(\theta \mid x)d\theta \tag{10}$$

Teorema 4 (δ^* sob perda absoluta) *Suponha que a função de perda é dada por*

$$L(\theta,\delta^*) = |\theta - \delta^*|. \tag{11}$$

*Dizemos que a função de perda é **absoluta**. Seja θ um parâmetro tomando valores na reta. Sob perda absoluta, $\delta^*(x)$ é a **mediana** a posteriori, isto é,*

$$\int_{\infty}^{\delta^*(x)} \xi(\theta \mid x)d\theta = \frac{1}{2} \tag{12}$$

Definição 17 (Estimador consistente) *Seja $\delta_1, \delta_2, \ldots, \delta_n$ uma sequência de estimadores de θ. Se quando $n \to \infty$ a sequência convergente para θ, dizemos que esta é uma **sequência consistente de estimadores**.*

Definição 18 (Estimador de máxima verossimilhança) *Para cada possível vetor (de observações) x, seja $\delta(x) \in \Omega$ um valor de $\theta \in \Omega$ de modo que a função de verossimilhança, $L(\theta) \propto f(x \mid \theta)$ [2], atinge o máximo. Dizemos que $\hat{\theta} = \delta(\boldsymbol{X})$ é o **estimador de máximo verossimilhança** de θ (Fisher, 1922)[3]. Quando observamos $\boldsymbol{X} = x$, dizemos que $\delta(x)$ é uma estimativa de θ. Dito de outra forma:*

$$\max_{\theta \in \Omega} f(\boldsymbol{X} \mid \theta) = f(\boldsymbol{X} \mid \hat{\theta}). \tag{13}$$

Famílias Conjugadas

Se $X_1, \ldots, X_n$ são iid e seguem a distribuição da coluna "Dados" na tabela 1.
Notações: $\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i; \quad y = \sum_{i=1}^{n} x_i$

Dados	Priori	Posteriori
$Bernoulli(\theta)$	$Beta(\alpha,\beta)$	$Beta(\alpha+y,\beta+n-y)$
$Poisson(\theta)$	$Gama(\alpha,\beta)$	$Gama(\alpha+y,\beta+n)$
$Normal(\mu,\sigma^2)$	$Normal(\mu_0,v_0^2)$	$Normal\left(\sigma^2\mu_0 + nv_0^2\bar{x}_n\sigma^2 + nv_0^2, \sigma^2 v_0^2 \sigma^2 + nv_0^2\right)$
$Exp(\theta)$	$Gama(\alpha,\beta)$	$Gama(\alpha+n,\beta+y)$

Table 1: Famílias Conjugadas

[2] $\propto$ - é um operador matemático binário que indica que o valor esquerdo é proporcional ao valor direito.
[3] Ronald Aylmer Fisher (1890-1962), biólogo e estatístico inglês.

Aula 5 - EMV (parte II)

Teorema 5 (Invariância do EMV) *Considere uma função $\phi : \Omega \to \mathbb{R}$. Se $\hat{\theta}$ é um EMV para θ, então $\phi(\hat{\theta})$ é um EMV para $\omega = \phi(\theta)$.*

Teorema 6 (Consistência do EMV) *Defina $l(\theta) := \log f_n(x \mid \theta)$ e assuma que $X_1, \dots, X_n \backsim f(\theta_0)$, isto é, que θ_0 é o valor verdadeiro do parâmetro. Denote $E_{\theta_0}[g] := \int_{\mathcal{X}} g(x, \theta_0) f(x \mid \theta_0)\, dx$. Suponha que*

- *$f(x_i \mid \theta)$ tem o mesmo suporte;*
- *θ_0 é o ponto inferior de Ω;*
- *$I(\theta)$ é diferenciável;*
- *$\hat{\theta}_{EMV}$ é única solução de $I'(\theta) = 0$.*

Então

$$\hat{\theta}_{EMV} \to \theta$$

Aula 6 - Método dos momentos e suficiência

Definição 19 (Método dos momentos) *Suponha que $X_1, \dots, X_n$ formam uma sequânica aleatória com distribuição conjunta $f_n(X_1, \dots, X_n \mid \theta), \theta \in \Omega \subseteq \mathbb{R}^k$ e que o k-ésimo momento existe. Defina $\mu_j(\theta) = E[X_1^j \mid \theta]$ e suponha que $\mu : \Omega \to \mathbb{R}^k$ é biunívoca, de modo que sua inversa é*

$$\theta = M(\mu_1(\theta), \dots, \mu_k(\theta)).$$

*Dados os momentos amostrais $m_j := \frac{1}{n} \sum_{i=1}^{n} X_i^j, j = 1, \dots, k$ o **estimador de momentos** (EMM) de θ é*

$$\hat{\theta}_{EMM} = M(m_1, \dots, m_k).$$

Teorema 7 (Consistência do EMM) *Suponha que $X_1, \dots, X_n$ formam uma amostra aleatória com distribuição conjunta $f_n(X_1, \dots, X_n \mid \theta), \theta \in \Omega \subseteq \mathbb{R}^k$ e que o k-ésimo momento existe. Suponha que a inversa M existe e é continua. Então o EMM é **consistente** para θ.*

Definição 20 (Estatística suficiente) *Seja $X_1, \dots, X_n$ uma amostra aleatória de uma distribuição indexada pelo parâmetro θ. Seja $T = r(X_1, \dots, X_n)$ uma estatística. Dizemos que T é uma **estatística suficiente** para θ se e somente se*

$$f(X_1, \dots, X_n \mid T = t, \theta) = f(X_1, \dots, X_n \mid T = t, \theta'), \forall \theta, \theta' \in \Omega, \tag{14}$$

isto é, se a distribuição condicional da amostra dado o valor da estatística não depende de θ.

Definição 21 (Aleatorização auxiliar) *Suponha que T é suficiente para θ. O processo de simular $X_1', \dots, X_n'$ dado que $T = r(X_1, \dots, X_n)$ de modo que*

$$f(X_1, \dots, X_n \mid \theta) = f(X_1', \dots, X_n' \mid \theta), \forall \theta \in \Omega, \tag{15}$$

*é chamado de **aleatorização auxiliar** (em inglês, auxiliary randomisation).*

Teorema 8 (Teorema de fatorização) *Suponha que $X_1, \dots, X_n$ perfazem uma amostra aleatória com f.d.p./f.m.p.$f(x \mid \theta), \theta \in \Omega$. Uma estatística $T = r(X_1, \dots, X_n)$ é suficiente para θ se, e somente se, para todo $x \in \mathcal{X}$ e $\theta \in \Omega$ existem u e v não negativos tal que*

$$f_n(x \mid \theta) = u(x) v[r(x), \theta]. \tag{16}$$

Definição 22 (Suficiência conjunta) *Dizemos que um conjunto de estatísticas $T = \{T_1, \dots, T_n\}$ é **suficiente** (conjuntamente) se que a distribuição condicional conjunta de $X_1, \dots, X_n$ dado $T_1 = t_1, \dots, T_n = t_n$ não dependentes de θ.*

Aula 7 - Suficiência conjunta e mínima, teorema de Rao-Blackwell

Definição 23 (Estatísticas de ordem) *Seja $\boldsymbol{X} = X_1, \dots, X_n$ uma amostra aleatória. Dizemos que $Y_1, \dots, Y_n$ são **estatísticas de ordem** se Y_1 é o menor valor de $\boldsymbol{X}$, Y_2 é o segundo menor valor e assim sucessivamente.*

Teorema 9 (Estatísticas de ordem são suficientes conjuntas) *Seja $X_1, \dots, X_n$ uma amostra aleatória com f.d.p./f.m.p. $f(x \mid \theta)$. As estatísticas de ordem $Y_1, \dots, Y_n$ são suficientes conjuntas para θ.*

Definição 24 (Suficiência mínima) *Uma estatística $\boldsymbol{T}$ é dita **mínima suficiente** se $\boldsymbol{T}$ é suficiente e é função de qualquer outra estatística suficiente. Um vetor $\boldsymbol{T} = \{T_1, \dots, T_n\}$ é dito **minimamente suficiente conjunto** se é função de qualquer outro valor de estatísticas suficientes conjuntas.*

Teorema 10 (EMV e Bayes são suficientes) *Se a função de verossimilhança admite fatorização pelo Teorema 8, os estimadores de Bayes e de máxima verossimilhança são estatísticas minimamente suficientes.*

Definição 25 (Notação conveniente) *É conveniente definir que para $g : \mathcal{X}^n \to \mathbb{R}$, escrevemos*

$$E_\theta[g] = \int_{\mathcal{X}} \cdots \int_{\mathcal{X}} g(\boldsymbol{x}) f_n(\boldsymbol{x} \mid \theta) dx_1 \cdots dx_n = \int_{\mathcal{X}} g(\boldsymbol{x}) f_n(\boldsymbol{x} \mid \theta) d\boldsymbol{x} \tag{17}$$

Definição 26 (Erro quadrático médio)

$$R(\theta, \delta) := E_\theta \left[\{\delta(\boldsymbol{X}) - \theta\}^2 \right]. \tag{18}$$

Definição 27 (Estimador condicionado)

$$\delta_0(\boldsymbol{T}) := E_\theta \left[\delta(\boldsymbol{X}) \mid \boldsymbol{T} \right]. \tag{19}$$

Teorema 11 (Teorema de Rao-Blackwell) *Seja $\delta(\boldsymbol{X})$ um estimador, $\boldsymbol{T}$ uma estatística suficiente para θ e seja $\delta_0(\boldsymbol{T})$ como na Definição 27. Então vale que*

$$R(\theta, \delta_0) \leq R(\theta, \delta)$$

Além disso, se $R(\theta, \delta) < \infty$ e $\delta(\boldsymbol{X})$ não é função de $\boldsymbol{T}$, vale a desigualdade estrita:

$$R(\theta, \delta_0) < R(\theta, \delta)$$

Aula 8 - Admissibilidade e viés

Definição 28 (Admissibilidade) *Um estimador δ é dito **inadmissível** se existe outro estimador δ_0 tal que $R(\theta, \delta_0) \leq R(\theta, \delta), \forall \theta \in \Omega$ e existe $\theta' \in \Omega$ tal que $R(\theta', \delta_0) < R(\theta', \delta)$. Nesse caso, dizemos que δ_0 domina δ. O estimador δ_0 é **admissível** se (e somente se) não há nenhum estimador que o domine.*

Definição 29 (Estimador não-viesado) *Um estimador $\delta(\boldsymbol{X})$ de uma função $g(\theta)$ é dito **não-viesado** se $E_\theta[\delta(\boldsymbol{X})] = g(\theta)$, $\forall \theta \in \Omega$. Um estimador que não atende a essa condição é dito viesado. E o **víes** de δ é definido como $B_\delta(\theta) := E_\theta[\delta(\boldsymbol{X})] - g(\theta)$.*

Teorema 12 (Estimador não-viesado da variância) *Seja $\boldsymbol{X} = \{X_1, \dots, X_n\}$ uma amostra aleatória, com $E[X_1] = m$ e $Var(X_1) = v < \infty$. Então*

$$\delta_1(\boldsymbol{X}) = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$$

é um estimador não-viesado de v.

Aula 9 - Eficiência

Definição 30 (Informação de Fisher) *Seja X uma variável aleatória com f.d.p./f.m.p. $f(x \mid \theta)$, $\theta \in \Omega \subseteq \mathbb{R}$. Suponha que $f(x \mid \theta)$ é duas vezes diferenciável com respeito a θ. Defina $\lambda(x \mid \theta) = \log f(x \mid \theta)$ e*

$$\lambda'(x \mid \theta) = \frac{\partial \lambda(x \mid \theta)}{\partial \theta} \quad e \quad \lambda''(x \mid \theta) = \frac{\partial^2 \lambda(x \mid \theta)}{\partial \theta^2} \tag{20}$$

*Definimos a **informação de Fisher** como*

$$I(\theta) = E_\theta \left[\{\lambda'(x \mid \theta)\}^2 \right] \overset{(1)}{=} -E_\theta \left[\lambda''(x \mid \theta)\right] = Var_\theta \left(\lambda'(x \mid \theta)\right). \tag{21}$$

Teorema 13 (Informação de Fisher em uma amostra aleatória) *Seja $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ uma amostra aleatória e seja $I_n = E_\theta[-\lambda_n''(\boldsymbol{X} \mid \theta)]$ a **informação de Fisher** da amostra. Então*

$$I_n(\theta) = nI(\theta)$$

Teorema 14 (Teorema de Cramér-Rao) *Seja $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ uma amostra aleatória, onde f.d.p./f.m.p. tem as mesmas premissas da Definição 30. Supondo que $T = r(\boldsymbol{X})$ é uma estatística com variância finita. Seja $m(\theta) = E_\theta(T)$ uma função diferenciável de θ. Então,*

$$Var_\theta(T) \geq \frac{[m'(\theta)]^2}{nI(\theta)}, \tag{22}$$

com igualdade apenas se existem u e v tal que

$$T = u(\theta)\lambda_n'(\boldsymbol{X} \mid \theta) + v(\theta).$$

Definição 31 (Estimador eficiente) *Um estimador $\delta(\boldsymbol{X})$ é dito **eficiente** de (sua esperança) $m(\theta)$ se*

$$Var_\theta(\delta) = \frac{[m'(\theta)]^2}{nI(\theta)}.$$

Aula 10 - Distribuição de uma estatística amostral e qui-quadrado

Definição 32 (Distribuição qui-quadrado) *Dizemos que uma variável aleatória Y tem distribuição **qui-quadrado** com m graus de liberdade quando*

$$f_Y(y) = \frac{1}{2^{m/2}\Gamma(m/2)} y^{m/2-1} e^{-y/2}, y > 0 \tag{23}$$

Vemos que Y tem função geradora de momentos:

$$\psi(t) = \left(\frac{1}{1-2t}\right)^{m/2}, t < 1/2.$$

Teorema 15 (Soma de variáveis aleatórias qui-quadrado) *Se $X_1, \ldots, X_n$ são variáveis aleatórias independentes com graus de liberdade m_i, então $W = \sum_{i=1}^n X_i$ tem distribuição qui-quadrado com graus de liberdade $m = \sum_{i=1}^n m_i$.*

Teorema 16 (Distribuição do quadrado de uma variável aleatória Normal padrão) *Se*

$$X \backsim Normal(0,1), Y = X^2$$

então, tem distribuição qui-quadrado com $m = 1$.

Aula 11 - Distribuição da média e variância amostrais

Teorema 17 (Independência da média e variância amostrais na Normal) *Seja $X_1, \ldots, X_n$ uma amostra aleatória de uma distribuição Normal com parâmetros μ e σ^2, $\overline{X}_n$ e a variância amostral $\overline{S}_n^2$, são independentes. Ademais, $\overline{X}_n \backsim Normal\left(\mu, \sigma^2\right)$ e $\overline{S}_n^2 \backsim Gama\left(\frac{n-1}{2}, \frac{n}{2n^2}\right)$*

Aula 12 - Distribuição t de Student e intervalos de confiança