

Resumo de Inferência Estatística

Content

- Aula 0: Revisão de Probabilidade
- Aula 1: O que é e para que serve Inferência Estatística?
- Aula 2: Distribuição a priori e a posteriori
- Aula 3: Prioris conjugadas e função de perda
- Aula 4: Estimadores de Bayes e EMV
- Aula 5: EMV
- Aula 6: Método dos momentos e suficiência
- Aula 7: Suficiência conjunta e mínima, teorema de Rao-Blackwell
- Aula 8: Admissibilidade e viés
- Aula 9: Eficiência
- Aula 10: Distribuição de uma estatística amostral e qui-quadrado
- Aula 11: Distribuição da média e variância amostrais
- Aula 12: Distribuição t de Student e intervalos de confiança
- Aula 13: Intervalos de confiança e Quantidades Pivotalis
- Aula 14: Testes de hipótese I
- Aula 15: Testes de hipótese II
- Aula 16: Testes de hipótese III
- Aula 17: Testes e conjuntos de confiança
- Aula 18: Teste t I
- Aula 19: Teste t II
- Aula 20: Teste f
- Aula 21: Regressão Linear I
- Aula 22: Regressão Linear II

Aula 0: Revisão de Probabilidade

Teorema 1 (Desigualdade de Markov) *Seja X uma variável aleatória não-negativa e $t > 0$. Então*

$$P(X \geq t) \leq \frac{E[X^n]}{t^n}. \quad (1)$$

Teorema 2 (Desigualdade de Chebychev) *Seja uma variável aleatória com média $E[Y] := \mu$ e variância $Var(Y) := \sigma^2$, ambas finitas. Mas uma vez, $t > 0$. Então*

$$P(|Y - \mu| \geq t) \leq \frac{Var(Y)}{t^2}. \quad (2)$$

Teorema 3 (Média e variância em uma amostra i.i.d.) *Sejam X_1, X_2, \dots, X_n variáveis aleatórias i.i.d., com média μ e variância σ^2 . Temos que*

$$(i) \ E[\bar{X}_n] = \mu$$

$$(ii) \ Var(\bar{X}_n) = \frac{\sigma^2}{n}$$

Definição 1 (Convergência em probabilidade) *Dizemos que uma sequência de variáveis aleatórias converge em probabilidade para b se, $\forall \varepsilon > 0$, temos*

$$\lim_{n \rightarrow \infty} P(|Z_n - b| < \varepsilon) = 1$$

Neste caso, escrevemos $Z_n \xrightarrow{p} b$.

Teorema 4 (Lei Fraca dos Grandes Números) *Sejam X_1, X_2, \dots, X_n variáveis aleatórias i.i.d., com média μ e variância σ^2 . Então*

$$\bar{X}_n \xrightarrow{p} \mu$$

Definição 2 (Convergência quase certa) *Dizemos que uma sequência de variáveis aleatórias $(Z_n)_{n \leq 1}$ converge quase certamente para b se*

$$P(\lim_{n \rightarrow \infty} Z_n = b) = 1$$

Teorema 5 (Lei Forte dos Grandes Números) *Sejam X_1, X_2, \dots, X_n variáveis aleatórias i.i.d., com média μ . Então*

$$P(\lim_{n \rightarrow \infty} \bar{X}_n = \mu) = 1$$

Teorema 6 (Teorema Central do Limite (Lindenberg e Lévy)) *Sejam X_1, X_2, \dots, X_n variáveis aleatórias i.i.d., com média μ e variância σ^2 . Então, para cada x , temos*

$$\lim_{n \rightarrow \infty} P\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq x\right) = \Phi(x),$$

onde,

$$\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_0^x \exp\left(-\frac{t^2}{2}\right) dt$$

é a função de distribuição (cumulativa) normal padrão.

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

Definição 3 (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 4 (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 $P : \Theta \rightarrow \mathcal{P}(\mathcal{X})$ que associa a cada $\theta \in \Theta$ uma distribuição de probabilidade P_θ em \mathcal{X} .

Definição 5 (Afirmiação probabilística) Dizemos que uma afirmação é *probabilística* quando ela utiliza conceitos da teoria de probabilidade para falar de um objeto.

Definição 6 (Inferência Estatística) Uma *inferência estatística* é uma *afirmação probabilística* sobre uma ou mais partes de um modelo estatístico.

Definição 7 (Estatística) Suponha que temos uma coleção de variáveis aleatórias $X_1, X_2, \dots, X_n \subseteq \mathbf{R}^n$ e uma função $r : \mathbf{X} \rightarrow \mathbf{R}^m$. Dizemos que a variável aleatória $T = r(X_1, X_2, \dots, X_n)$ é uma *estatística*.

Definição 8 (Permutabilidade) Uma coleção finita de variáveis aleatórias X_1, X_2, \dots, X_n com densidade conjunta f é dita *permutável* se

$$f(x_1, x_2, \dots, x_n) = f(x_{\pi(1)}, x_{\pi(2)}, \dots, x_{\pi(n)}) \quad (3)$$

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 9 (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 10 (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, $\mathbf{x} := \{x_1, x_2, \dots, x_n\}$ é a *distribuição a posteriori* de θ , denotamos por $\xi(\theta | \mathbf{x})$ a f.d.p./f.m.p. condicional a $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$.

Teorema 7 (Distribuição a posteriori: derivação) Considere a amostra aleatória X_1, X_2, \dots, X_n de uma distribuição com f.d.p./f.m.p. $f(x | \theta)$. Se a distribuição a priori é $\xi(\theta)$, temos

$$\xi(\theta | \mathbf{x}) = \frac{\xi(\theta) \prod_{i=1}^n f(x_i | \theta)}{g_n(\mathbf{x})}, \quad \theta \in \Omega \quad (4)$$

Chamamos $g_n(\mathbf{x})$ de distribuição marginal de X_1, X_2, \dots, X_n .

Definição 11 (Função de verossimilhança) Quando encaramos a f.d.p./f.m.p. $f(x_1, x_2, \dots, x_n | \theta)$ como uma função do parâmetro θ , chamamos esta função de *função de verossimilhança*, e podemos denotá-la como $L(\theta; \mathbf{x})$ ou, quando a notação não criar ambiguidade, simplesmente $L(\theta)$.

Aula 3: Prioris conjugadas e função de perda

Definição 12 (Hiper-parâmetros) Seja $\xi(\theta | \phi)$ 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 13 (Priori conjugada) Suponha que X_1, X_2, \dots sejam condicionalmente independentes dado θ , com f.d.p./f.m.p. $f(x | \theta)$. Defina

$$\Psi = \left\{ f : \Omega \rightarrow (0, \infty), \int_{\Omega} f dx = 1 \right\} \quad (5)$$

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

$$\frac{f(\mathbf{x} | \theta)f(\theta)}{\int_{\Omega} f(\mathbf{x} | \theta)f(\theta)d\theta} \in \Psi \quad (6)$$

Teorema 8 (Distribuição a posteriori da média de uma normal) Suponha que X_1, X_2, \dots, X_n formam uma amostra aleatória com distribuição normal e com média desconhecida θ e variância $\sigma^2 > 0$, conhecida e fixa. Suponha que $\theta \sim \text{Normal}(\mu_0, v_0^2)$ a priori. Então

$$\xi(\theta | x, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\theta - \mu_1)^2}{2v_1^2}\right), \quad (7)$$

onde

$$\mu_1 := \frac{\sigma^2\mu_0 + nv_0^2\bar{x}_n}{\sigma^2 + nv_0^2} \text{ e } v_1^2 := \frac{\sigma^2v_0^2}{\sigma^2 + nv_0^2} \quad (8)$$

Definição 14 (Priori imprópria) Seja $\xi : \Lambda \rightarrow (0, \infty), \Omega \subseteq \Lambda$, 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 15 (Estimador) Sejam X_1, X_2, \dots, X_n variáveis aleatórias com distribuição conjunta indexada por θ . Um **estimador** de θ é qualquer função real $\delta : X_1, X_2, \dots, X_n \rightarrow \mathbb{R}^d, d \geq 1$.

Definição 16 (Estimativa) Dizemos que o valor de δ avaliado nas realizações de X_1, X_2, \dots, X_n , $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$, $\delta(\mathbf{x})$ é uma **estimativa** de θ .

Definição 17 (Função de perda) Uma função de perda é uma função real em duas variáveis

$$L : \Omega \times \mathbb{R}^d \rightarrow \mathbb{R}, \quad (9)$$

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

Famílias Conjugadas

Se X_1, \dots, X_n são iid e seguem a distribuição da coluna “Dados” na tabela 1.

Notações: $\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$; $y = \sum_{i=1}^n x_i$

¹p.d.f. - “probability density function” ou função de densidade de probabilidade

| Dados | Priori | Posteriori |
|---------------------------|--------------------------|--|
| Bernoulli(θ) | Beta(α, β) | Beta($\alpha + y, \beta + n - y$) |
| Poisson(θ) | Gama(α, β) | Gama($\alpha + y, \beta + n$) |
| Normal(μ, σ^2) | Normal(μ_0, v_0^2) | Normal($\frac{\sigma^2 \mu_0 + n v_0^2 \bar{x}_n}{\sigma^2 + n v_0^2}, \frac{\sigma^2 v_0^2}{\sigma^2 + n v_0^2}$) |
| Exp(θ) | Gama(α, β) | Gama($\alpha + n, \beta + y$) |

Table 1: Famílias Conjugadas

Aula 4: Estimadores de Bayes e EMV

Definição 18 (Estimador de Bayes) Considere a perda esperada a posteriori:

$$E_{\theta|x}[L(\theta, a)] = E[L(\theta, a) | x] = \int_{\Omega} L(\theta, a) \xi(\theta | x) d\theta \quad (10)$$

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

$$E[L(\theta, \delta^*(x)) | x] = \min_{a \in A} E[L(\theta, a) | x]. \quad (11)$$

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

$$\delta^*(x) = E[\theta | X = x] = \int_{\Omega} \theta \xi(\theta | x) d\theta \quad (12)$$

Teorema 10 (δ^* sob perda absoluta) Suponha que a função de perda é dada por

$$L(\theta, \delta^*) = |\theta - \delta^*|. \quad (13)$$

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,

$$\int_{-\infty}^{\delta^*(x)} \xi(\theta | x) d\theta = \frac{1}{2} \quad (14)$$

Definição 19 (Estimador consistente) Seja $\delta_1, \delta_2, \dots, \delta_n$ uma sequência de estimadores de θ . Se quando $n \rightarrow \infty$ a sequência convergente para θ , dizemos que esta é uma **sequência consistente de estimadores**.

Definição 20 (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, $L(\theta) \propto f(x | \theta)$ ², atinge o máximo. Dizemos que $\hat{\theta} = \delta(\mathbf{X})$ é o **estimador de máxima verossimilhança** de θ (Fisher, 1922)³. Quando observamos $\mathbf{X} = x$, dizemos que $\delta(x)$ é uma estimativa de θ . Dito de outra forma:

$$\max_{\theta \in \Omega} f(\mathbf{X} | \theta) = f(\mathbf{X} | \hat{\theta}). \quad (15)$$

² \propto - é um operador matemático binário que indica que o valor esquerdo é proporcional ao valor direito.

³Ronald Aylmer Fisher (1890-1962), biólogo e estatístico inglês.

Aula 5: EMV

Teorema 11 (Invariância do EMV) Considere uma função $\phi : \Omega \rightarrow \mathbb{R}$. Se $\hat{\theta}$ é um EMV para θ , então $\phi(\hat{\theta})$ é um EMV para $\omega = \phi(\theta)$.

Teorema 12 (Consistência do EMV) Defina $l(\theta) := \log f_n(x | \theta)$ e assumamos que $X_1, \dots, X_n \sim f(\theta_0)$, isto é, que θ_0 é o valor verdadeiro do parâmetro. Denote $E_{\theta_0}[g] := \int_{\mathcal{X}} g(x, \theta_0) f(x | \theta_0) dx$. Suponha que

- $f(x_i | \theta)$ 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

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

Aula 6: Método dos momentos e suficiência

Definição 21 (Método dos momentos) Suponha que X_1, \dots, X_n formam uma sequência aleatória com distribuição conjunta $f_n(X_1, \dots, X_n | \theta)$, $\theta \in \Omega \subseteq \mathbb{R}^k$ e que o k -ésimo momento existe. Defina $\mu_j(\theta) = E[X_1^j | \theta]$ e suponha que $\mu : \Omega \rightarrow \mathbb{R}^k$ 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 13 (Consistência do EMM) Suponha que X_1, \dots, X_n formam uma amostra aleatória com distribuição conjunta $f_n(X_1, \dots, X_n | \theta)$, $\theta \in \Omega \subseteq \mathbb{R}^k$ e que o k -ésimo momento existe. Suponha que a inversa M existe e é contínua. Então o EMM é **consistente** para θ .

Definição 22 (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

$$f(X_1, \dots, X_n | T = t, \theta) = f(X_1, \dots, X_n | T = t, \theta'), \forall \theta, \theta' \in \Omega, \quad (16)$$

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

Definição 23 (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

$$f(X_1, \dots, X_n | \theta) = f(X'_1, \dots, X'_n | \theta), \forall \theta \in \Omega, \quad (17)$$

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

Teorema 14 (Teorema de fatorização) Suponha que X_1, \dots, X_n perfazem uma amostra aleatória com f.d.p./f.m.p. $f(x | \theta)$, $\theta \in \Omega$. 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

$$f_n(x | \theta) = u(x)v[r(x), \theta]. \quad (18)$$

Definição 24 (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 25 (Estatísticas de ordem) Seja $\mathbf{X} = X_1, \dots, X_n$ uma amostra aleatória. Dizemos que Y_1, \dots, Y_n são **estatísticas de ordem** se Y_1 é o menor valor de \mathbf{X} , Y_2 é o segundo menor valor e assim sucessivamente.

Teorema 15 (Estatísticas de ordem são suficientes conjuntas) Seja X_1, \dots, X_n uma amostra aleatória com f.d.p./f.m.p. $f(x | \theta)$. As estatísticas de ordem Y_1, \dots, Y_n são suficientes conjuntas para θ .

Definição 26 (Suficiência mínima) Uma estatística \mathbf{T} é dita **mínima suficiente** se \mathbf{T} é suficiente e é função de qualquer outra estatística suficiente. Um vetor $\mathbf{T} = \{T_1, \dots, T_n\}$ é dito **minimamente suficiente conjunto** se é função de qualquer outro valor de estatísticas suficientes conjuntas.

Teorema 16 (EMV e Bayes são suficientes) Se a função de verossimilhança admite fatorização pelo Teorema 14, os estimadores de Bayes e de máxima verossimilhança são estatísticas minimamente suficientes.

Definição 27 (Notação conveniente) É conveniente definir que para $g : \mathcal{X}^n \rightarrow \mathbb{R}$, escrevemos

$$E_\theta[g] = \int_{\mathcal{X}} \cdots \int_{\mathcal{X}} g(\mathbf{x}) f_n(\mathbf{x} | \theta) dx_1 \cdots dx_n = \int_{\mathcal{X}} g(\mathbf{x}) f_n(\mathbf{x} | \theta) d\mathbf{x} \quad (19)$$

Definição 28 (Erro quadrático médio)

$$R(\theta, \delta) := E_\theta [\{\delta(\mathbf{X}) - \theta\}^2]. \quad (20)$$

Definição 29 (Estimador condicionado)

$$\delta_0(\mathbf{T}) := E_\theta [\delta(\mathbf{X}) | \mathbf{T}]. \quad (21)$$

Teorema 17 (Teorema de Rao-Blackwell) Seja $\delta(\mathbf{X})$ um estimador, \mathbf{T} uma estatística suficiente para θ e seja $\delta_0(\mathbf{T})$ como na Definição 29. Então vale que

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

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

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

Aula 8: Admissibilidade e viés

Definição 30 (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 31 (Estimador não-viesado) Um estimador $\delta(\mathbf{X})$ de uma função $g(\theta)$ é dito **não-viesado** se $E_\theta[\delta(\mathbf{X})] = g(\theta), \forall \theta \in \Omega$. Um estimador que não atende a essa condição é dito **viesado**. E o **viés** de δ é definido como $B_\delta(\theta) := E_\theta[\delta(\mathbf{X})] - g(\theta)$.

Teorema 18 (Estimador não-viesado da variância) Seja $\mathbf{X} = \{X_1, \dots, X_n\}$ uma amostra aleatória, com $E[X_1] = m$ e $Var(X_1) = v < \infty$. Então

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

é um estimador não-viesado de v .

Aula 9: Eficiência

Definição 32 (Informação de Fisher) Seja X uma variável aleatória com f.d.p./f.m.p. $f(x | \theta)$, $\theta \in \Omega \subseteq \mathbb{R}$. Suponha que $f(x | \theta)$ é duas vezes diferenciável com respeito a θ . Defina $\lambda(x | \theta) = \log f(x | \theta)$ e

$$\lambda'(x | \theta) = \frac{\partial \lambda(x | \theta)}{\partial \theta} \quad \text{e} \quad \lambda''(x | \theta) = \frac{\partial^2 \lambda(x | \theta)}{\partial \theta^2} \quad (22)$$

Definimos a **informação de Fisher** como

$$I(\theta) = E_{\theta} [\{\lambda'(x | \theta)\}^2] \stackrel{(1)}{=} -E_{\theta} [\lambda''(x | \theta)] = \text{Var}_{\theta} (\lambda'(x | \theta)). \quad (23)$$

Teorema 19 (Informação de Fisher em uma amostra aleatória) Seja $\mathbf{X} = \{X_1, \dots, X_n\}$ uma amostra aleatória e seja $I_n = E_{\theta}[-\lambda''_n(\mathbf{X} | \theta)]$ a **informação de Fisher** da amostra. Então

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

Teorema 20 (Teorema de Cramér-Rao) Seja $\mathbf{X} = \{X_1, \dots, X_n\}$ uma amostra aleatória, onde f.d.p./f.m.p. tem as mesmas premissas da Definição 32. Supondo que $T = r(\mathbf{X})$ é uma estatística com variância finita. Seja $m(\theta) = E_{\theta}(T)$ uma função diferenciável de θ . Então,

$$\text{Var}_{\theta}(T) \geq \frac{[m'(\theta)]^2}{nI(\theta)}, \quad (24)$$

com igualdade apenas se existem u e v tal que

$$T = u(\theta)\lambda'_n(\mathbf{X} | \theta) + v(\theta).$$

Definição 33 (Estimador eficiente) Um estimador $\delta(\mathbf{X})$ é dito **eficiente** de (sua esperança) $m(\theta)$ se

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

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

Definição 34 (Distribuição qui-quadrado) Dizemos que uma variável aleatória Y tem distribuição **qui-quadrado** com m graus de liberdade quando

$$f_Y(y) = \frac{1}{2^{m/2}\Gamma(m/2)} y^{m/2-1} e^{-y/2}, y > 0 \quad (25)$$

Vemos que Y tem função geradora de momentos:

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

Teorema 21 (Soma de variáveis aleatórias qui-quadrado) Se X_1, \dots, 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 22 (Distribuição do quadrado de uma variável aleatória Normal padrão) Se

$$X \sim \text{Normal}(0, 1), Y = X^2$$

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

Aula 11: Distribuição da média e variância amostrais

Teorema 23 (Independência da média e variância amostrais na Normal) Seja X_1, \dots, X_n uma amostra aleatória de uma distribuição Normal com parâmetros μ e σ^2 , \bar{X}_n e a variância amostral \bar{S}_n^2 , são independentes. Ademais, $\bar{X}_n \sim \text{Normal}(\mu, \sigma^2)$ e $\bar{S}_n^2 \sim \text{Gama}(\frac{n-1}{2}, \frac{n}{2n^2})$

Aula 12: Distribuição t de Student e intervalos de confiança

Definição 35 (A distribuição t de Student) Tome, $Y \sim \text{Qui-quadrado}(m)$ e $Z \sim \text{Normal}(0, 1)$ e defina a variável aleatória

$$X = \frac{Z}{\sqrt{\frac{Y}{m}}}.$$

Dizemos que X tem distribuição **t de Student com m graus de liberdade**. E sabemos que

$$f_X = \frac{\Gamma(\frac{m+1}{2})}{\sqrt{m\pi}\Gamma(\frac{m}{2})} \left(1 + \frac{x^2}{m}\right)^{-\frac{m+1}{2}}, \quad x \in (-\infty, +\infty).$$

Teorema 24 (Distribuição amostral do estimador não-viesado da variância) Considere o estimador

$$\hat{\sigma}' = \sqrt{\frac{\Delta^2}{n-1}},$$

onde $\Delta^2 = \sum_{i=1}^n (X_i - \bar{X}_n)^2$. Então, vale que

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\hat{\sigma}'} \sim \text{Student}(n-1)$$

Teorema 25 (Intervalo de confiança) Seja $\mathbf{X} = \{X_1, \dots, X_n\}$ uma amostra aleatória, onde cada uma tem p.d.f. $f(x | \theta)$, e considere uma função real $g(\theta)$. Sejam $A(\mathbf{X})$ e $B(\mathbf{X})$ duas estatísticas de modo de valha

$$P(A(\mathbf{X}) < g(\theta) < B(\mathbf{X})) \geq \gamma. \quad (26)$$

Dizemos que $I(\mathbf{X}) = (A(\mathbf{X}), B(\mathbf{X}))$ é um **intervalo de confiança** de $100\gamma\%$ para $g(\theta)$. Se a desigualdade for uma igualdade para todo $\theta \in \Omega$, dizemos que o intervalo é **exato**.

Aula 13: Intervalos de confiança e Quantidades Pivotalis

Definição 36 (Intervalo de confiança unilateral) Seja $\mathbf{X} = \{X_1, \dots, X_n\}$ uma amostra aleatória, onde cada uma tem p.d.f. $f(x | \theta)$, e considere uma função real $g(\theta)$. Seja $A(\mathbf{X})$ uma estatística que

$$P(A(\mathbf{X}) < g(\theta)) \geq \gamma, \quad \forall \theta \in \Omega$$

dizemos que o intervalo aleatório $(A(\mathbf{X}), \infty)$ é chamado de intervalo de confiança **unilateral** de $100\gamma\%$ para $g(\theta)$ (ou ainda, de intervalo de confiança **inferior** de $100\gamma\%$ para $g(\theta)$). O intervalo $(-\infty, B(\mathbf{X}))$, com

$$P(g(\theta) < B(\mathbf{X})) \geq \gamma, \quad \forall \theta \in \Omega$$

é definido de forma análoga, e é chamado de intervalo de confiança **superior** de $100\gamma\%$ para $g(\theta)$. Se a desigualdade é uma igualdade para todo $\theta \in \Omega$, os intervalos são chamados **exatos**.

Definição 37 (Quantidade pivotal) Seja $\mathbf{X} = \{X_1, \dots, X_n\}$ uma amostra aleatória com p.d.f. $f(x | \theta)$. Seja $V(\mathbf{X}, \theta)$ uma variável aleatória cuja distribuição é a mesma para todo $\theta \in \Omega$. Dizemos que $V(\mathbf{X}, \theta)$ é uma **quantidade pivotal**.

Teorema 26 (Intervalo de confiança unilateral) Seja $\mathbf{X} = \{X_1, \dots, X_n\}$ uma amostra aleatória com p.d.f. $f(x | \theta)$. Suponha que existe uma quantidade pivotal V , com c.d.f.⁴ continua G . Assuma que existe $r(v, \mathbf{x})$ estritamente crescente em v para todo \mathbf{x} . Finalmente, tome $0 < \gamma < 1$ e $\gamma_1 < \gamma_2$ de modo que $\gamma_2 - \gamma_1 = \gamma$. Então as estatísticas

$$A(\mathbf{X}) = r(G^{-1}(\gamma_1), \mathbf{X}),$$

$$B(\mathbf{X}) = r(G^{-1}(\gamma_2), \mathbf{X}),$$

são os limites de um intervalo de confiança de $100\gamma\%$ para $g(\theta)$.

Aula 14: Testes de hipótese I

Definição 38 (Hipótese nula e hipótese alternativa) Considere o espaço de parâmetros Ω e defina $\Omega_0, \Omega_1 \subset \Omega$ de modo que $\Omega_0 \cup \Omega_1 = \Omega$ e $\Omega_0 \cap \Omega_1 = \emptyset$. Definimos

$$H_0 := \theta \in \Omega_0,$$

$$H_1 := \theta \in \Omega_1,$$

E dizemos que H_0 é a **hipótese nula** e H_1 é a **hipótese alternativa**. Se $\theta \in \Omega_1$, então dizemos que rejeitamos a hipótese nula. Por outro lado, se $\theta \in \Omega_0$, então dizemos que não rejeitamos ou falhamos em rejeitar H_0 .

Definição 39 (Hipótese simples e hipótese composta) Dizemos que uma hipótese H_i , é **simples**, se $\Omega_i = \{\theta_i\}$, isto é, se a partição correspondente é um único ponto. Uma hipótese é dita **composta** se não é simples.

Definição 40 (Hipótese unilateral e hipótese bilateral) Uma hipótese da forma $H_0 : \theta \leq \theta_0$ ou $H_0 : \theta \geq \theta_0$ é dita **unilateral** (“one-sided”), enquanto hipóteses da forma $H_0 : \theta \neq \theta_0$ são ditas **bilaterais** (“two-sided”).

Aula 15: Testes de hipótese II

Definição 41 (Região crítica) O conjunto

$$S_1 := \{\mathbf{x} : |\bar{X}_n - \mu_0| \geq c\}$$

é chamado de **região crítica** do teste.

Definição 42 (Região de rejeição) Se $R \subseteq \mathbb{R}$ é tal que “rejeitamos H_0 se $T \in R$ ”, então R é chamada uma **região de rejeição** para a estatística T e o teste associado.

Definição 43 (Função poder) Seja δ um procedimento de aceitação/rejeição como visto anteriormente. A **função poder** é definida como

$$\pi(\theta | \delta) := P(\mathbf{X} \in S_1 | \theta) = P(T \in R | \theta), \quad \theta \in \Omega \quad (27)$$

⁴c.d.f. - cumulative distribution function

Definição 44 (Tipos de erros) *Tipos de erros que podem ser cometidos*

| Nome | Erro cometido |
|--------------|--|
| Erro tipo I | Rejeitar H_0 quando ela é verdadeira . |
| Erro tipo II | Falhar em rejeitar H_0 quando ela é falsa . |

Definição 45 (Tamanho/nível de um teste) *Dizemos que um teste, δ , tem **tamanho** ou **nível de significância** $\alpha(\delta)$, com*

$$\alpha(\delta) := \sup_{\theta \in \Omega_0} \pi(\theta \mid \delta).$$

Aula 16: Testes de hipótese III

Definição 46 (O p-valor) *Para cada t , seja δ_t o teste que rejeita H_0 se $T \geq t$. Então, quando $T = t$, o **p-valor** vale*

$$p(t) := \sup_{\theta \in \Omega_0} \pi(\theta \mid \delta_t) = \sup_{\theta \in \Omega_0} P(T \geq t \mid \theta) \quad (28)$$

*ou seja, o **p-valor** é o tamanho do teste δ_t .*

Aula 17: Testes e conjuntos de confiança

Definição 47 (Intervalos de confiança e testes são equivalentes) *Suponha que dispomos de dados $\mathbf{X} = \{X_1, \dots, X_n\}$ com f.d.p. comum $f(x \mid \theta)$, e estamos interessados em testar as hipóteses:*

$$H_0 : g(\theta) = g_0,$$

$$H_1 : g(\theta) \neq g_0,$$

de modo que existe um teste δ_{g_0} com nível α_0 destas hipóteses. Para cada $\mathbf{X} = \mathbf{x}$, defina

$$w(\mathbf{x}) = \{g_0 : \delta_{g_0} \text{ não rejeita } H_0 \text{ dado que } \mathbf{X} = \mathbf{x}\}.$$

Fazendo o nível de confiança do intervalo $\gamma = 1 - \alpha_0$, temos

$$P(g(\theta_0) \in w(\mathbf{X}) \mid \theta = \theta_0) \geq \gamma, \quad \forall \theta_0 \in \Omega.$$

Definição 48 (Conjunto de confiança) *Se um conjunto aleatório $w(\mathbf{X})$ satisfaz*

$$P(g(\theta_0) \in w(\mathbf{X}) \mid \theta = \theta_0) \geq \gamma,$$

*para todo $\theta_0 \in \Omega$, então chamamos $w(\mathbf{X})$ de um **conjunto de confiança** para $g(\theta)$.*

Teorema 27 (Testando hipóteses a partir de conjuntos de confiança) *Suponha que dispomos de dados $\mathbf{X} = \{X_1, \dots, X_n\}$ com f.d.p. comum $f(x \mid \theta)$ e que $w(\mathbf{X})$ é um conjunto de confiança para uma função de interesse $g(\theta)$. Então para todo valor g_0 assumido por $g(\theta)$ existe um teste δ_{g_0} , de nível α_0 que rejeita $H_0 : g(\theta) = g_0$ se e somente se $g(\theta_0) = g_0 \notin w(\mathbf{X})$.*

Teorema 28 (Teste de razão de verossimilhanças (para o Teo. de Wilks)) *A estatística*

$$\Lambda(\mathbf{x}) = \frac{\sup_{\theta \in \Omega_0} f_n(\mathbf{x} \mid \theta)}{\sup_{\theta \in \Omega} f_n(\mathbf{x} \mid \theta)}$$

*é chamada um **estatística de razão de verossimilhanças**. Um teste de razão de verossimilhanças, δ_k , é um teste que rejeita H_0 se $\Lambda(\mathbf{x}) \leq k$ para uma constante k .*

Teorema 29 (Teorema de Wilks) Suponha que temos um espaço de parâmetros com k coordenadas, $\theta = (\theta_1, \dots, \theta_n)$ e desejamos testar a hipótese (simples) da forma

$$H_0 : \theta_j = \theta_0^j, \quad j = 1, \dots, k,$$

$$H_1 : \theta_j \neq \theta_0^j, \quad j = 1, \dots, k.$$

Então, sob condições de regularidade, temos que, à medida que $n \rightarrow \infty$,

$$-2 \log \Lambda(\mathbf{x}) \xrightarrow{d} X^2(k)$$

Aula 18: Teste t I

Definição 49 (Teste não viesado) Suponha que desejamos testar a hipótese

$$H_0 : \theta \in \Omega_0,$$

$$H_1 : \theta \in \Omega_1,$$

através do teste δ . Dizemos que δ é **não-viesado** se (e somente se) para $\theta \in \Omega_0$ e $\theta' \in \Omega_1$, vale

$$\pi(\theta \mid \delta) \leq \pi(\theta' \mid \delta),$$

ou seja, se a função poder é pelo menos tão grande no espaço onde H_0 é falsa (Ω_1) quando no espaço em que H_0 é verdadeira (Ω_0).

Definição 50 (Teste t) Um teste δ_c que rejeita H_0 se $U \geq c$ (equiv. $U \leq c$), com $c = T^{-1}(1 - \alpha_0; n - 1)$ é chamado de um **teste t** (unicaudal) de tamanho α_0 .

Teorema 30 (Propriedades do teste t) Suponha que δ_c rejeita H_0 se $U \geq c$. Então

- $\mu = \mu_0 \implies \pi(\mu, \sigma^2 \mid \delta_c) = \alpha_0$
- $\mu < \mu_0 \implies \pi(\mu, \sigma^2 \mid \delta_c) < \alpha_0$
- $\mu > \mu_0 \implies \pi(\mu, \sigma^2 \mid \delta_c) > \alpha_0$
- $\lim_{\mu \rightarrow -\infty} \pi(\mu, \sigma^2 \mid \delta_c) = 0$
- $\lim_{\mu \rightarrow +\infty} \pi(\mu, \sigma^2 \mid \delta_c) = 1$
- δ_c é não-viesado e tem tamanho α_0 .

Teorema 31 (P-valor para um teste t unicaudal) Suponha que observarmos $U = u$ e seja $T(\cdot; n - 1)$ a f.d.a. de uma distribuição t de Student com $n - 1$ graus de liberdade. Para a hipótese

$$H_0 : \mu \geq \mu_0,$$

$$H_1 : \mu < \mu_0,$$

o p -valor vale $T(u; n - 1)$, enquanto para a hipótese

$$H_0 : \mu \leq \mu_0,$$

$$H_1 : \mu > \mu_0,$$

o p -valor vale $1 - T(u; n - 1)$.

Aula 19: Teste t II

Teorema 32 (Teste pareado) *Sejam amostras \mathbf{X} e \mathbf{Y} (antes e depois), tais que $X_i \sim \text{Normal}(\mu_1, \sigma^2)$ e $Y_i \sim \text{Normal}(\mu_2, \sigma^2)$, a hipótese*

$$H_0 : \mu_1 \leq \mu_2$$

$$H_1 : \mu_1 > \mu_2$$

Pode ser modelada com a variável $Z_i = X_i - Y_i$ ($Z_i \sim \text{Normal}(\mu_Z = \mu_1 - \mu_2, 2\sigma^2)$), então podemos testar hipóteses sobre μ_Z a partir de \mathbf{Z}

$$H_0 : \mu_Z \leq 0$$

$$H_1 : \mu_Z > 0$$

Teorema 33 (Teste t para duas amostras) *Considere $\mathbf{X} = \{X_1, \dots, X_m\}$ e $\mathbf{Y} = \{Y_1, \dots, Y_n\}$, queremos estudar a diferença das médias. Modelando em distribuição normal $X_i \sim \text{Normal}(\mu_1, \sigma_1^2)$, $i = 1, \dots, m$ e $Y_j \sim \text{Normal}(\mu_2, \sigma_2^2)$, $j = 1, \dots, n$. Sob a premissa de homogeneidade $\sigma_1^2 = \sigma_2^2 = \sigma^2$, podemos testar a hipótese*

$$H_0 : \mu_1 \leq \mu_2$$

$$H_1 : \mu_1 > \mu_2$$

computando a estatística

$$U = \frac{\sqrt{m+n-2}(\bar{X}_m - \bar{Y}_n)}{\sqrt{(\frac{1}{m} + \frac{1}{n})(S_X^2 + S_Y^2)}}$$

onde \bar{X}_m e \bar{Y}_n são as médias e S_X^2 e S_Y^2 são a soma das variâncias.

Teorema 34 (Relaxando a premissa de homogeneidade) *Do teorema acima, podemos relaxar a premissa de igualdade das variâncias assumindo que $\sigma_2^2 = k\sigma_1^2$, então a estatística teste vale*

$$U = \frac{\sqrt{m+n-2}(\bar{X}_m - \bar{Y}_n)}{\sqrt{(\frac{1}{m} + \frac{k}{n})(S_X^2 + \frac{S_Y^2}{n})}}$$

Aula 20: Teste f

Definição 51 (A distribuição F) *Sejam $Y \sim \text{Qui-quadrado}(m)$ e $W \sim \text{Qui-quadrado}(n)$. Então*

$$X = \frac{Y/m}{W/n},$$

tem distribuição F com m e n graus de liberdade, com f.d.p.

$$f_X(x) = \frac{\Gamma(\frac{m+n}{2})m^{m/2}n^{n/2}}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})} \cdot \frac{x^{m/2-1}}{(mx+n)^{(m+n)/2}}, \quad x > 0,$$

Teorema 35 (Propriedades da distribuição F) *Propriedades para a distribuição F:*

1. Se $X \sim F(m, n)$, então $\frac{1}{X} \sim F(n, m)$;
2. Se $Y \sim \text{Student}(n)$, então $Y^2 \sim F(1, n)$.

Teorema 36 (Igualdade de duas variâncias) Suponha $X_i \sim \text{Normal}(\mu_1, \sigma_1^2)$, $i = 1, \dots, m$ e $Y_j \sim \text{Normal}(\mu_2, \sigma_2^2)$, $j = 1, \dots, n$. Queremos testar

$$H_0 : \sigma_1^2 \leq \sigma_2^2$$

$$H_1 : \sigma_1^2 > \sigma_2^2$$

Para isso, vamos computar a estatística de teste

$$V = \frac{S_X^2/(m-1)}{S_Y^2/(n-1)}$$

onde $S_X^2 = \sum_{i=1}^m (X_i - \bar{X}_m)^2$ e $S_Y^2 = \sum_{j=1}^n (Y_j - \bar{Y}_n)^2$

Definição 52 (O teste F) O teste F de homogeneidade (igualdade de variâncias) é o teste δ_c que rejeita H_0 de $V \geq c$, para uma constante positiva c .

Teorema 37 (A distribuição de V) Seja $V = \frac{S_X^2/(m-1)}{S_Y^2/(n-1)}$, então:

$$\frac{\sigma_2^2}{\sigma_1^2} V \sim F(m-1, n-1).$$

Além disso, se $\sigma_1^2 = \sigma_2^2$, $V \sim F(m-1, n-1)$.

Aula 21: Regressão Linear I

Teorema 38 (A linha de mínimos quadrados) Sejam $(x_1, y_1), \dots, (x_n, y_n)$ uma coleção de n pontos. Suponha que estamos interessados na reta

$$y_i = \beta_0 + \beta_1 x_i. \quad (29)$$

Os valores dos coeficientes que minimizam a soma de quadrados são

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x},$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

Definição 53 (Modelo linear) Podemos construir um modelo estatístico explícito para a relação entre as variáveis X e Y :

$$E[Y \mid X = x_1, \dots, x_P] = \beta_0 + \beta_1 x_1 + \dots + \beta_P x_P \quad (30)$$

Podemos então idealizar o seguinte modelo

$$Y_i = \beta_0 + \sum_{j=1}^P \beta_j x_{ij} + \varepsilon_i, \quad \varepsilon \sim \text{Normal}(0, \sigma^2).$$

Aula 22: Regressão Linear II

Teorema 39 (EMV para os coeficientes de uma regressão linear (simples)) *Sob as premissas já listadas, os estimadores de máxima verossimilhança para $\theta = (\beta_0, \beta_1, \sigma^2)$ são*

$$\begin{aligned}\hat{\beta}_{0EMV} &= \bar{y} - \hat{\beta}_{1EMV}\bar{x}, \\ \hat{\beta}_{1EMV} &= \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \\ \hat{\sigma}_{EMV}^2 &= \frac{1}{n} \sum_{i=1}^n \left(y_i - (\hat{\beta}_{0EMV} + \hat{\beta}_{1EMV}x_i) \right)^2,\end{aligned}$$

ou seja, os estimadores de máxima verossimilhança dos coeficientes minimizam a soma de quadrados da reta estimada.

Teorema 40 (Distribuição amostral dos estimadores dos coeficientes)

$$\begin{aligned}\hat{\beta}_{0EMV} &\sim \text{Normal} \left(\beta_0, \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{s_X^2} \right) \right), \\ \hat{\beta}_{1EMV} &\sim \text{Normal} \left(\beta_1, \frac{\sigma^2}{s_X^2} \right), \\ \text{Cov}(\hat{\beta}_{0EMV}, \hat{\beta}_{1EMV}) &= -\frac{\bar{x}\sigma^2}{s_X^2},\end{aligned}$$

onde $s_x = \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}$.

Teorema 41 (Intervalos de confiança para os coeficientes de uma regressão linear) *Podemos computar intervalos de confiança para os coeficientes da regressão linear de maneira muito similar ao que já vimos para o caso da média da Normal*

$$\begin{aligned}\hat{\beta}_0 \pm \hat{\sigma}'c\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{s_x^2}} \quad \text{e} \quad \hat{\beta}_1 \pm c\frac{\hat{\sigma}'}{s_x}, \\ \hat{\beta}_0 + \hat{\beta}_1 x_{pred} \pm c\hat{\sigma}'\sqrt{\frac{1}{n} + \frac{(x_{pred} - \bar{x})^2}{s_x^2}}\end{aligned}$$

onde $c = T^{-1}(1 - \frac{\alpha_0}{2}; n - 2)$ e

$$\hat{\sigma}' := \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{n - 2}}.$$

Definição 54 (Testes de hipóteses para o coeficiente angular) *Em geral, estamos interessados em testar a hipótese*

$$\begin{aligned}H_0 : \beta_1 &= \beta^*, \\ H_1 : \beta_1 &\neq \beta^*.\end{aligned}$$

Para tanto, podemos computar a estatística

$$U_1 = s_x \frac{\hat{\beta}_1 - \beta^*}{\hat{\sigma}'}, \tag{31}$$

e computar o p -valor como

$$P(U_1 \geq |u_1|) + P(U_1 \leq -|u_1|). \quad (32)$$

Notando que U_1 tem distribuição t de Student com $n - 2$ graus de liberdade sob H_0 , podemos computar o p -valor exatamente.

Resultados bem similares valem para testar hipóteses sobre β_0 ou \hat{Y} .

Teorema 42 (Predição pontual) Suponha que queremos prever o valor de Y para um certo x_{pred} que não foi observado no experimento. Podemos compor nossa predição (pontual) como

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x_{pred}. \quad (33)$$

Onde a predição tem erro quadrático médio (EQM) igual a

$$E[(\hat{Y} - Y)^2] = \sigma^2 \left(1 + \frac{1}{n} + \frac{(x_{pred} - \bar{x})^2}{s_X^2} \right).$$

Teorema 43 (Intervalos de predição para \hat{Y}) A probabilidade de $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x_{pred}$ estar no intervalo

$$\hat{Y} \pm T^{-1}\left(1 - \frac{\alpha_0}{2}; n - 2\right) \hat{\sigma}' \sqrt{\left[1 + \frac{1}{n} + \frac{(x_{pred} - \bar{x})^2}{s_X^2}\right]},$$

é $1 - \alpha_0$.