Estatística - Capítulo 7

Thursday, September 10, 2020 9:24 PM

Itens importantes capítulo 6
- Markov: $\mathbb{P}(X \geq t) \leq E[X^n]/t^n$, quando X é não negativa.
- Chebyshev: $\mathbb{P}(|X - E[X]| \geq t) \leq Var[X]/t^2$
- Lei dos Grandes Números: $\forall \varepsilon > 0$, $\lim_{n\to\infty} \mathbb{P}[|\bar{X}_n - \mu| < \varepsilon] = 1$ — média amostral
- Teorema Central do Limite: $\lim_{n\to\infty} \mathbb{P}\left[\frac{\bar{X}_n - \mu}{\sigma/n^{1/2}} \leq x\right] = \Phi(x)$

Chamamos de Convergência em probabilidade.

Distribuição da Normal

Vamos usar sempre que $\bar{X}_n = 1/n \sum_{i=1}^n X_i$, que é a média amostral.

O que é um modelo estatístico?
Variáveis aleatórias + Distribuições + Parâmetros → Bayesiano adota distribuição para os parâmetros

Espaço de Parâmetros
Denotamos como Ω. Conjunto onde os parâmetros residem. $\mathbb{R}$

Estatística
Função da amostra aleatória, $T = r(X) = r(X_1, \dots, X_n)$

Variáveis aleatórias independentes e identicamente distribuídas de tamanho n } Mesma distribuição

Bayesianismo: parâmetros são variáveis aleatórias, isto é, tem uma distribuição

- Priori: Distribuição do parâmetro com conhecimento ANTES dos dados.
- Posteriori: " " " " DEPOIS " "

Se $X_1, \dots, X_n$ é amostra aleatória de $f(x|\theta)$, com priori para θ $\xi(\theta)$, distribuição de θ

$$\xi(\theta|x) = \frac{f(x_1|\theta) \dots f(x_n|\theta)\, \xi(\theta)}{g_n(x)}, \quad \theta \in \Omega$$

$$\int_\Omega \xi(\theta|x)\, d\theta = 1 \qquad g_n(x) = \int_\Omega f_n(x|\theta)\xi(\theta)\, d\theta$$

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A)\,\mathbb{P}(A)}{\mathbb{P}(B)}$$

$x_1 \dots x_n$ θ θ $x_1, \dots, x_n$

Obs.: Nos exercícios, $\xi(\theta|x) \propto f_n(x|\theta)\xi(\theta)$ e depois calculamos $g_n(x)$ — proporcionalidade

- Verossimilhança: $L: \Omega \to \mathbb{R}^+$, $\theta \mapsto L(\theta|x) = f_n(x|\theta)$ → Distribuição Conjunta: $f(x_1|\theta) \cdot f(x_2|\theta) \cdot \ldots \cdot f(x_n|\theta)$ $= L(\theta|x_1, \dots, x_n)$

Isto é, distribuição conjunta como função de θ.

- Famílias Conjugadas: Quando temos uma amostra $X_1, \dots, X_n$ com distribuição $f_n(x|\theta)$, se $\xi(\theta|x)$ e $\xi(\theta)$ são pdfs da mesma família de distribuições, ela é dita família conjugada. Gamma, Normal, Uniforme, etc.

$\tilde{\mathcal{F}} = \{N(\mu, \sigma^2),\ \mu \in \mathbb{R},\ \sigma^2 > 0\}$ — $\xi(\theta)$ — $\xi(\theta|x)$

Exemplos para lembrar: ① ② — $Beta(1,1) \equiv Unif([0,1])$

- $X_1, \dots, X_n \sim Be(\theta)$, $\theta \sim Beta(\alpha, \beta) \Rightarrow \theta|X \sim Beta(\alpha + \sum_{i=1}^n x_i, \beta + n - \sum_{i=1}^n x_i)$
- $X_1, \dots, X_n \sim Poisson(\theta)$, $\theta \sim Gamma(\alpha, \beta) \Rightarrow \theta|X \sim Gamma(\alpha + \sum_{i=1}^n x_i, \beta + n)$
- $X_1, \dots, X_n \sim Normal(\theta, \sigma^2)$, $\theta \sim N(\mu_0, v_0^2) \Rightarrow \theta|X \sim N((\sigma^2\mu_0 + nv_0^2\bar{x}_n)/(\sigma^2 + nv_0^2),\ \sigma^2 v_0^2/(\sigma^2 + nv_0^2))$
- $X_1, \dots, X_n \sim Exp(\theta)$, $\theta \sim Gamma(\alpha, \beta) \Rightarrow \theta|X \sim Gamma(\alpha + n, \beta + \sum_{i=1}^n x_i)$

Obs.: Priori pode ser imprópria, isto é, $\int \xi(\theta)d\theta \neq 1$ desde que $\int \xi(\theta|x)\,d\theta = 1$

Estimador: Uma estatística com objetivo de estimar θ.

Função de Perda: Seja $a \in \mathbb{R}$ um estimador para θ, a função de perda calcula o erro da estimação. Exemplos: $L(\theta, a) = (\theta - a)^2$ (Loss, Quadrática), $L(\theta, a) = |\theta - a|$ (Absoluta), etc.

Estimador de Bayes:

- Dado uma função de perda $L(\theta, a)$ e uma priori $\xi(\theta)$ para θ,

$$E[L(\theta, a)] = \int_\Omega L(\theta, a)\xi(\theta)\,d\theta \quad \text{Priori}$$

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

- $\delta^*(x) := \arg\min_{a \in \mathbb{R}} E[L(\theta, a)|x]$, isto é, é o valor de a que minimiza a perda média.
- Se $L(\theta, a) = (\theta - a)^2$, $\delta^*(x) = E[\theta|X]$ é estimador de Bayes — média a posteriori
- Se $L(\theta, a) = |\theta - a|$, $\delta^*(x) = $ mediana de $\theta|X$, isto é, $\int_{-\infty}^{mediana} \xi(\theta|x)\,d\theta = \frac{1}{2}$ — $\mathbb{P}(\theta \leq mediana) = \frac{1}{2}$

Definição (Estimador Consistente): Seja $\{\hat{\theta}_n\}$ estimadores. Se $\hat{\theta}_n \xrightarrow{P} \theta$, $\{\hat{\theta}_n\}$ é dito sequência de estimador consistente

$\forall \varepsilon > 0$, $\mathbb{P}(|\hat{\theta}_n - \theta| < \varepsilon) \xrightarrow{n\to\infty} 1$

$f_n(x|\theta) = \prod_{i=1}^n f(x_i|\theta)$

$\log f_n(x|\theta) = \sum_{i=1}^n \log f(x_i|\theta)$

Estimador de Máxima Verossimilhança: Se $f_n(x|\theta)$ é a verossimilhança (dist. conjunta), $\hat{\theta}_{MLE} = \arg\max_{\theta \in \Omega} f_n(x|\theta)$, isto é, queremos maximizar $f_n(x|\theta)$. Fazemos $\lambda_n(x|\theta) = \log f_n(x|\theta)$ e, então: $\lambda_n'(x|\theta) = 0 \Rightarrow \hat{\theta}$ é MLE para θ

CUIDADO: Queremos maximizar $f_n(x|\theta)$. Nem sempre derivar dá! Então tem que tentar ver com cuidado! O MLE pode não existir ou não ser único também.

- Invariância: Se $\hat{\theta}$ é MLE para θ, então $g(\hat{\theta})$ é MLE para $g(\theta)$
- Consistência: $\hat{\theta}_{MLE} \xrightarrow{P} \theta$ se $\hat{\theta}_n$ for único para cada n e + condições de regularidade.

Se conhecemos $\hat{\theta}$ e queremos estimar outro parâmetro, como a mediana, basta encontrar g que relacione ele com θ.

Método dos Momentos: $\mu_j = E[X^j]$, $\mu_1 = E[X]$, $\mu_2 = E[X^2]$, $Var(X) = E[X^2] - E[X]^2 = \mu_2 - \mu_1^2$

$X_1, \dots, X_n$ e θ k-dimensional. Suponha $E[X_i^j] < \infty$, $1 \leq j \leq k$.
O estimador do método de momentos é aquele que resolve $m_j = E[X_1^j|\theta]$, $j = 1, \dots, k$, onde $m_1 = \bar{X}_n$, $m_2 = 1/n\sum_{i=1}^n X_i^2$, $m_3 = 1/n \sum_{i=1}^n X_i^3, \dots$

momento amostral — $\mu_j|\theta$ — Dimensão do parâmetro

Estatística Suficiente: Seja $T = r(X)$ uma estatística. Suponha que $\forall \theta \in \Omega$ e $\forall t$, $T = t$, $f_n(x|\theta, T = t) = f_n(x|T = t)$, isto é, não depende de θ.

distribuição conjunta

↳ O interessante é o Critério de Fatorização: $f_n(x|\theta) = u(x)v(r(x), \theta)$. Se isso acontece para duas funções u e v, $\forall x \in \mathbb{R}^n$, então T é suficiente

Não depende de θ — Estatísticas

Vale para mais estatísticas, desde que $f_n(x|\theta) = u(x)v(r_1(x), \dots, r_n(x), \theta)$

Obs.: Uma função injetiva de uma estatística suficiente é uma estatística suficiente. T, $T' = h(T)$ e h injetiva

Suficiente mínima: Uma estatística suficiente é mínima se for função de todas as outras estatísticas suficientes.

↳ Difícil de provar no caso geral, logo, temos alguns teoremas interessantes!

* Se $\hat{\theta}$ é MLE e estatística suficiente, então ele será suficiente mínimo.
* Se $\delta^*(x)$ é estimador de Bayes e estatística suficiente, então ele será suf. mínimo.

fatorável

Logo se tivermos MLE ou Bayes, basta provar sua SUFICIÊNCIA que teremos de bandeja a suficiência mínima.

Definição (MSE): $R(\theta, \delta) = E_\theta\{[\delta(X) - h(\theta)]^2\}$ → Erro médio quadrático, $L(\theta, a) = (\theta - a)^2$

estimador para $h(\theta)$

Não depende de θ pois T é suficiente.

Seja T uma estatística suficiente. Então $\delta_0(T) := E[\delta(X)|T]$

Rao-Blackwell (Melhorando um estimador): $R(\theta, \delta_0) \leq R(\theta, \delta)$
Se $R(\theta, \delta) < \infty$, então $R(\theta, \delta_0) < R(\theta, \delta)$ a não ser que $\delta(x) = h(T)$, uma função de T.

Definição (Admissível): δ é inadmissível se existe outro estimador δ^* tal que $R(\theta, \delta^*) \leq R(\theta, \delta)$ e valha a desigualdade estrita em pelo menos um valor de θ. Nesse caso δ^* domina δ. Caso não exista δ^*, δ é admissível!

Exercício (15, 7.10): $X_1, \dots, X_n \sim Pareto(x_0, \alpha)$ — x_0 desconhecido, α conhecido

$$f(x|x_0) = \frac{\alpha x_0^\alpha}{x^{\alpha+1}}, \quad x \in [x_0, \infty)$$

x_0 maior possível

$$f_n(x|x_0) = \prod_{i=1}^n \left(\frac{\alpha x_0^\alpha}{x_i^{\alpha+1}}\right) \mathbb{1}\{x_i \in [x_0, \infty)\} = \frac{\alpha^n x_0^{n\alpha}}{\prod_{i=1}^n (x_i^{\alpha+1})} \mathbb{1}\{x_1, \dots, x_n \in [x_0, \infty)\}$$

$\mathbb{1}$: $\begin{cases} 1, & x \in [x_0, \infty) \\ 0, & x \notin [x_0, \infty) \end{cases}$

$x_0 \leq \min\{x_1, \dots, x_n\}$

$\hat{x}_0 = \min\{x_1, \dots, x_n\}$

$\hat{x}_0$ é uma estatística suficiente mínima? Sim, função de v.a. ? Sim

Critério da Fatorização

(16) $$f_n(x|x_0) = \frac{\alpha^n x_0^{n\alpha}}{\prod_{i=1}^n (x_i^{\alpha+1})} \mathbb{1}\{x_1, \dots, x_n \in [x_0, \infty)\}$$

$x_1 \geq x_0$, $x_2 \geq x_0$, ..., $x_n \geq x_0$ ⟺ $\min\{x_1, \dots, x_n\} \geq x_0$

$\mathbb{1}\{\min\{x_1, \dots, x_n\} \geq x_0\}$ = $\mathbb{1}\{\hat{x}_0 \geq x_0\}$

sem parâmetro

$$u(x) = \frac{1}{\prod_{i=1}^n (x_i^{\alpha+1})}$$

$$v(\hat{x}_0, x_0) = \alpha^n x_0^{n\alpha} \mathbb{1}\{\hat{x}_0 \geq x_0\}$$

suficiência, parâmetro

- Se MLE é suficiente ⇒ MLE é suficiente mínimo