Estimação Pontual

Luan Fiorentin

UFPR - DEST

2019-05-04

Sumário

Introdução

- Seja X uma variável aleatória com **função densidade (ou de probabilidade)** denotada por $f(x;\theta)$, em que θ é um **parâmetro desconhecido**.

- Chama-se de **inferência estatística** o problema que consiste em especificar um ou mais valores para θ, baseado em um conjunto de valores X.

- A **inferência estatística** pode ser feita por meio de:
  - Estimativa pontual.
  - Estimativa intervalar.

- Um experimentador usa as informações em uma amostra aleatória $X_1, X_2, ..., X_n$ para fazer **inferências sobre θ**.

- Normalmente o tamanho da amostra é bastante grande e fica **inviável** tirar conclusões baseadas em uma grande quantidade de dados.

- Assim, um dos objetivos da inferência estatística é **resumir as informações** de uma amostra, da maneira mais compacta possível, mas que ao mesmo tempo também seja informativa.

- Normalmente, esse resumo é feito por meio de **estatísticas**:
    - Média amostral.
    - Mediana amostral.
    - Variância amostral.
    - ...

Conceitos de Inferência Estatística

- **População**: conjunto de valores (ou itens) de uma característica associada a uma coleção de indivíduos ou objetos de interesse.

- **Amostra**: conjunto de dados coletados e/ou selecionados de uma população por um procedimento estatístico. Os elementos de uma amostra são conhecidos como pontos amostrais, unidades amostrais ou observações.

"**Amostra aleatória** é uma sequência $X_1, X_2, ..., X_n$ de n variáveis aleatórias **independentes e identicamente distribuídas** (iid) com função densidade (ou de probabilidade) $f(x;\theta)$".

- Normalmente $n > 1$, então fdp ou fp conjunta será

$$f(\boldsymbol{x};\boldsymbol{\theta}) = f(x_1, x_2, ..., x_n;\theta) = \prod_{i=1}^{n} f(x_i;\theta).$$

- **Parâmetro**:

População → censo → parâmetro

Uma medida numérica que descreve alguma **característica da população** é usualmente representada por letras gregas: $\theta, \mu, \sigma, ...$

Exemplo: Média populacional (μ).

- **Estatística**:

População → amostra → estatística

Uma medida numérica que descreve alguma **característica da amostra** é usualmente denotada pela letra grega do respectivo parâmetro com um acento circunflexo: $\hat{\theta}$, $\hat{\mu}$, $\hat{\sigma}$,. . . , ou letras do alfabeto comum.

Exemplo: Média amostral ($\bar{x}$).

- **Estatística**:

Qualquer **função da amostra** que não depende de parâmetros desconhecidos, denotada por $T(\boldsymbol{X}) = T(X_1, X_2, ..., X_n)$

Exemplos:

$T_1(\boldsymbol{X}) = \sum_{i=1}^{n} X_i = X_1 + X_2 + ... + X_n$
$T_2(\boldsymbol{X}) = \sum_{i=1}^{n} \frac{X_i}{n}$
$T_3(\boldsymbol{X}) = \prod_{i=1}^{n} X_i = X_1 \cdot X_2 \cdot ... \cdot X_n$
$T_4(\boldsymbol{X}) = X_{min}$
$T_5(\boldsymbol{X}) = \sum_{i=1}^{n} (X_i - \mu)^2$

Verificamos que T_1, T_2, T_3, T_4 são **estatísticas**, mas T_4 não.

Como as demais é uma função da amostra, então uma estatística também é uma variável aleatória. Logo, tem uma **distribuição amostral**.

- **Espaço paramétrico**: é o conjunto Θ em que θ pode assumir valores.

- **Estimador**: qualquer estatística que assume valores em Θ é um estimador para θ.

- **Estimador pontual**: um estimador pontual para θ é qualquer estatística que possa ser usada para estimar esse parâmetro, ou seja,

$$\hat{\theta} = T(\boldsymbol{X}).$$

Observação:

- Todo estimador é uma estatística, mas nem toda estatística é um estimador.
- O valor assumido pelo estimador pontual é chamado de **estimativa pontual**,

$$\theta = T(\boldsymbol{X}) = T(X_1, X_2, ..., X_n) = t.$$

- Então, o estimador é uma **função da amostra**, e a estimativa é o **valor observado** de um estimador (um número) de uma amostra particular.

Métodos de Estimação Pontual

- **Método da Máxima Verossimilhança**

- Método dos Momentos

- Método dos Mínimos Quadrados

Método da Máxima Verossimilhança

- Consiste em encontrar valor para os parâmetros de maneira a **maximizar a probabilidade dos dados observados** (isto é, busca parâmetros que maximizem a função de verossimilhança).

- Considere uma população e uma variável aleatória X, com fp $p(x, \theta)$ (se X é V.A. discreta) ou fdp $f(x, \theta)$ (se X é V.A. contínua), sendo θ o **parâmetro desconhecido**.

- Se X for variável aleatória **contínua**, a **função de verossimilhança** L é definida por

$$L(\theta; x_1, \ldots, x_n) = f(x_1; \theta) \times \ldots \times f(x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta).$$

- Se X for variável aleatória **discreta**, a **função de verossimilhança** L é definida por

$$L(\theta; x_1, \ldots, x_n) = p(x_1; \theta) \times \ldots \times p(x_n; \theta) = \prod_{i=1}^{n} p(x_i; \theta).$$

- O estimador de máxima verosssimilhança é equivalente ao **estimador de máxima log-verossimilhança**:
  - Melhor desempenho computacional.
  - Cálculos se tornam mais "fáceis".
- A **função de log-verossimilhança** l é definida por

$$l(\theta, x_1, \ldots, x_n) = lnL(\theta, x_1, \ldots, x_n) = \sum_{i=1}^{n} ln[f(\theta, x_1, \ldots, x_n)].$$

Em muitos casos, o estimador de máxima verossimilhança pode ser encontrado seguindo os passos abaixo:

1. Encontrar a função de verossimilhança.
2. Aplicar a função log.
3. Derivar em relação ao parâmetro θ.
4. Igualar o resultado a zero.
5. Isolar o parâmetros de interesse θ.

Método do Momentos

- Consiste em igualar os **momentos populacionais** (definidos através da amostra) com os **momentos da distribuição**.

- Seja $X_1, ..., X_n$ uma amostra de uma população com distribuição de probabilidade $f(x)$. O **k-ésimo momento da distribuição** é definido por

$$E(X^k).$$

- O momento **k-ésimo populacional** correspondente é dado por

$$\frac{1}{n}\sum_{i=1}^{n} X_i^k,$$

onde $k = 1, 2, 3, ...$

- Assim, o método dos momentos é definido igualando-se o **momento populacional com o momento da distribuição**.

Para ilustrar, o primeiro momento da distribuição é

$$E(X^1) = E(X) = \mu,$$

e o primeiro momento populacional é

$$\frac{1}{n}\sum_{i=1}^{n} X_i^1 = \bar{x}.$$

Portanto, o **estimador de momentos** para a média é $\mu = \bar{x}$.

Exercício 1

Dado o modelo abaixo, encontre o estimador para o parâmetro desconhecido λ pelo método de máxima verossimilhança (MMV) e pelo método dos momentos (MM).

$$f(x_i;\lambda) = \lambda e^{\ \lambda x_i},$$

onde $\lambda > 0$ e $x_i = 1, 2, ..., n$.

Estimação Pontual

- Quando a **amostragem** é feita a partir de uma população descrita por uma função $f(x, \theta)$, o conhecimento de θ a partir da amostra, gera todo o **conhecimento para a população**.
- Assim, é natural que se procure um método para se achar um **bom estimador** para θ.
- Existem algumas **propriedades** que definem o que é um bom estimador, ou o "melhor" estimador entre uma série de candidatos.

Qual o **melhor** estimador para a amostra de alturas (em metros) dada a seguir?

$$\mathbf{x} = \{1,65; 1,57; 1,72; 1,66; 1,71; 1,74; 1,81; 1,68; 1,60; 1,77\}$$

1. Estimador μ_1: média artimética entre os valores mínimo e máximo da amostra.
2. Estimador μ_2: primeiro valor sorteado da amostra.
3. Estimador μ_3: média artimética dos valores da amostra.

$$\hat{\mu}_1 = \frac{1,57 + 1,81}{2} = 1,69$$

$$\hat{\mu}_2 = 1,65$$

$$\hat{\mu}_3 = \frac{1,57 + ... + 1,81}{2} = 1,69$$

- O “**bom**” estimador para o parâmetro deve apresentar as seguintes propriedades:
  - **Não viciado**.
  - **Consistente**.
  - **Eficiente**

Erro Quadrático Médio

O Erro Quadrático Médio (EQM) de um estimador $\hat{\theta}$ de θ é dado por

$$EQM[\hat{\theta}] = E[(\hat{\theta} - \theta)^2],$$

$$EQM[\hat{\theta}] = Var[\hat{\theta}] + B[\hat{\theta}]^2,$$

em que

$$B[\hat{\theta}]^2 = E[\hat{\theta}] - \theta$$

é denominado de vício do estimador $\hat{\theta}$. Portanto, dizemos que um estimador é não viciado quando

$$B[\hat{\theta}] = 0 \Rightarrow E[\hat{\theta}] = \theta.$$

Vício

Seja $(X_1, X_2, ..., X_n)$ uma amostra aleatória de uma variável aleatória com fdp ou fp $f(x, \theta)$, em que $\theta \in \Theta$, dizemos que o estimador $\hat{\theta} = T(X)$ é não viciado para θ se

$$E[\hat{\theta}] = E[T(\boldsymbol{X})] = \theta.$$

Um estimador $\hat{\theta}$ é dito assintoticamente não viciado se

$$\lim_{n \to \infty} E[\hat{\theta}] = \theta.$$

Isso quer dizer que para as amostras suficientemente grande, $\hat{\theta}$ passa a ser imparcial.

Consistência

Seja $(X_1, X_2, ...; X_n)$ uma amostra aleatória de uma variável aleatória com fdp ou fp $f(x, \theta)$, em que $\theta \in \Theta$, o estimador $\hat{\theta} = T(X)$ é consistente para θ se satisfaz simultaneamente

$$\lim_{n \to \infty} E[\hat{\theta}] = \theta$$

$$\lim_{n \to \infty} Var[\hat{\theta}] = 0.$$

Exemplo

- considere a média amostral como um estimador da média populacional μ:

$$E[\bar{x}] = E\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \mu$$

$$Var[\bar{x}] = Var\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{\sigma^2}{n}$$

Logo, $\bar{x}$ é um estimador não viciado e consistente para μ.

Eficiência

Sejam $\hat{\theta}_1 = T_1(\boldsymbol{X})$ e $\hat{\theta}_2 = T_2(\boldsymbol{X})$ dois estimadores pontuais não viciados para θ. A eficiência relativa de $\hat{\theta}_1$ em relação a $\hat{\theta}_2$ é

$$ER[\hat{\theta}_1; \hat{\theta}_2] = \frac{Var[\hat{\theta}_1]}{Var[\hat{\theta}_2]}.$$

Logo, se
$ER[\hat{\theta}_1; \hat{\theta}_2] > 1$, então $\hat{\theta}_2$ é mais eficiente.
$ER[\hat{\theta}_1; \hat{\theta}_2] < 1$, então $\hat{\theta}_1$ é mais eficiente.

Exemplo

- Considere a média amostral $\hat{\mu}_1 = \bar{X}$ e a mediana amostral $\hat{\mu}_2 = mediana(X_1, ..., X_n)$ como estimadores não viciados, e as respectivas variâncias dadas por $Var(\hat{\mu}_1 = \sigma^2/n)$ e $Var(\hat{\mu}_2 = (\pi/2)\sigma^2/n)$. Qual é mais eficiente?

$$ER(\hat{\mu}_1, \hat{\mu}_2) = \frac{Var(\hat{\mu}_1)}{Var(\hat{\mu}_2)} = \frac{\sigma^2/n}{(\pi/2)\sigma^2/n} = \frac{2}{\pi} = 0,63.$$

Logo, como $0,63 < 1$, então $Var(\hat{\mu}_1) < Var(\hat{\mu}_2)$. Assim, conclui-se que $\hat{\mu}_1$ é mais eficiente que $\hat{\mu}_2$.

- O **erro padrão** de um estimador dá uma ideia da precisão da estimativa.

- O erro padrão (EP) de um estimador é seu desvio-padrão (raíz quadrada da variância), ou seja

$$EP[\hat{\theta}] = \sqrt{Var[\hat{\theta}]}.$$

Exemplo:

Sabemos que a distribuição de $\bar{x}$ tem média μ e variância σ^2/n. Então, o erro padrão de $\bar{x}$ é

$$EP[\hat{\theta}] = \sqrt{Var[\hat{\theta}]} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}.$$

Distribuições Amostrais

- Uma amostra de tamaho n é descrita pelos valores $x_1, x_2, ..., x_n$ das variáveis aleatórias $X_1, X_2, ..., X_n$, configurando uma **amostra aleatória**.

- No caso de uma **Amostragem Aleatória Simples** (AAS) com reposição, $X_1, X_2, ..., X_n$ serão variáveis aleatórias **independentes e identicamente distribuídas** (iid) com **função de probabilidade** (fp) ou **função densidade de probabilidade** (fdp) **conjunta** dada por

$$f(x_1, x_2, ..., x_n; \boldsymbol{\theta}) = f(x_1; \boldsymbol{\theta}) \cdot f(x_2; \boldsymbol{\theta}) \cdot ... \cdot f(x_n; \boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i; \boldsymbol{\theta}),$$

sendo que o mesmo valor do parâmetro θ é utilizado em cada um dos termos no produto.

- Quando uma **amostra** $X_1, X_2, ..., X_n$ é obtida, geralmente estamos interessados em um resumo destes valores, que pode ser expresso matematicamente pela estatística $T(x_1, x_2, ..., x_n)$.

- Dessa forma, $Y = T(x_1, x_2, ..., x_n)$ é também uma **variável aleatória**.

- Se Y é uma V. A., então ela possui uma **distribuição de probabilidade**.

- Uma vez que a distribuição de Y é derivada da amostra $X_1, X_2, ..., X_n$, vamos denominá-la de **distribuição amostral** de Y.

- A **distribuição de probabilidade de uma estatística** qualquer $Y = T(x_1, x_2, ..., x_n)$ é denominada de **distribuição amostral** de Y.

- Assim, uma estatística também é uma variável aleatória, pois seus valores mudam conforme a **amostra aleatória**.

- Duas estatísticas comumente utilizadas para o resumo de uma amostra aleatória são a **média amostral** $(\bar{x})$ e a **proporção amostral** $(\bar{p})$. Cada uma delas também possui uma distribuição amostral.

Distribuição Amostral da Média

Para estudarmos a distribuição amostral da estatística $\bar{x}$, considere uma população identificada pela V. A. X, com parâmetros de média ($E[X] = \mu$) e variância ($Var[X] = \sigma^2$) conhecidos.

Em seguida, realiza-se os seguintes passos:

- Retira-se m amostras aleatórias (com reposição) de tamanho n dessa população.
- Para cada uma das m amostras, calcula-se a média amostral $\bar{x}$.
- Verifica-se a distribuição das m médias amostrais e estudamos suas propriedades.

Seja $X \sim N(10, 16)$, como se comporta $\bar{x}$ para $n = 10, 30, 50, 100$?

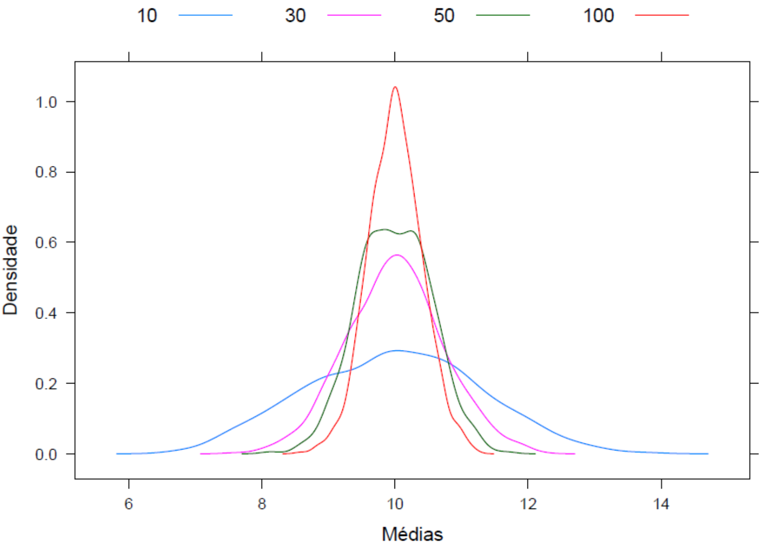

Figura 1: Distribuição amostral da média

- Através do estudo da **distribuição da média amostral** chegamos em um dos resultados mais importantes da inferência estatística.

Distribuição Amostral da Média

$E[\bar{X}] = \mu_{\bar{X}} = \mu$
$Var[\bar{X}] = \sigma^2_{\bar{X}} = \sigma^2/n$
Portanto, se

$$X \sim N(\mu, \sigma^2), \quad \text{então} \quad \bar{X} \sim N(\mu_{\bar{X}}, \sigma^2_{\bar{X}}).$$

Mas como $\mu_{\bar{X}} = \mu$ e $\sigma^2_{\bar{X}} = \sigma^2/n$,
então a distribuição amostral da média amostral $\bar{X}$ é

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

- Para amostras suficientemente grandes, a **média amostral $\bar{X}$ converge para o verdadeiro valor da média populacional** μ (é um estimador não viesado de μ).
- Além disso, a variância das médias amostrais $\sigma^2_{\bar{X}}$ tende a diminuir conforme $n \to \infty$ (é um estimador consistente).
- Esses resultados sugerem que, quando o tamanho da amostra aumenta, independente do formato da distribuição da população original, a **distribuição amostral de $\bar{X}$ aproxima-se cada vez mais de uma distribuição Normal**.
- Esse é um resultado fundamental na teoria de probabilidade, conhecido como **Teorema Central do Limite**.

Teorema Central do Limite (TCL)

Para amostras aleatórias $(X_1, X_2, ..., X_n)$, retiradas de uma população com média μ e variância σ^2, a distribuição amostral da média $\bar{X}$, terá forma dada por

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

no limite, quando $n \to \infty$, $Z \sim N(0, 1)$.

Se a população apresentar distribuição Normal, então $\bar{X}$ terá distribuição exata normal.
A rapidez da convergência para a normal depende da distribuição da população da qual as amostras foram geradas.

- O **teorema TCL** garante que, para n grande, a **distribuição da média amostral**, devidamente padronizada, se **comporta segundo um modelo normal com média 0 e variância 1**.

- Pelo teorema, temos que **quanto maior o tamanho da amostra**, melhor é a aproximação.

- Estudos envolvendo simulações mostram que, em muitos casos, **valores de n ao redor de 30 fornecem aproximações bastante boas** para as aplicações práticas.

Distribuição amostral da proporção

- Muitas vezes, o interesse é conhecer uma **proporção**, e não a média de uma população.

Suponha que uma amostra de tamanho n foi obtida de uma população, e que $x < n$ observações nessa amostra pertençam a uma classe de interesse (ex.: pessoas do sexo masculino).

Dessa forma, a **proporção amostral** é dada pelo **número de sucessos** (x) pelo total de tentativas (n),

$$\hat{p} = \frac{x}{n},$$

onde $\hat{p}$ o melhor estimador para a proporção populacional p.

Observação: n e p são os parâmetros da Distribuição Binomial.

- A **distribuição amostral de uma proporção** é a distribuição das proporções de todas as possíveis amostras de tamanho n retiradas de uma população.
  - Considere uma moeda que é lançada $n = 10, 30, 50, 100$ vezes, e a proporção de caras é registrada.
  - Esse processo é repetido $m = 1000$ vezes.

Conclui-se que:

- A média das proporções para $n \to \infty$ tende para a verdadeira proporção populacional $p = 0, 5$.
- A distribuição amostral das proporções é aproximadamente uma distribuição normal.

Seja $P \sim Bin(n, 0.5)$, como se comporta $\bar{p}$ para $n = 10, 30, 50, 100$?

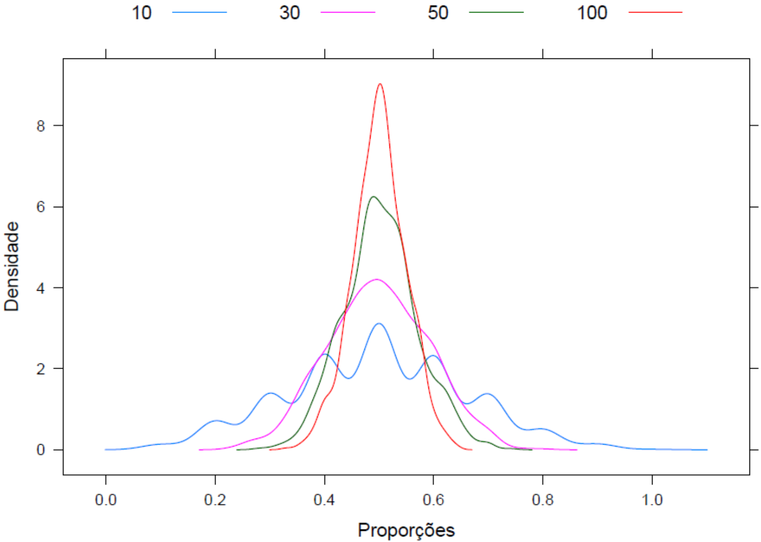

Figura 2: Distribuição amostral da proporção

- Através do estudo da distribuição da média amostral chegamos em um dos resultados mais importantes da inferência estatística.

Distribuição Amostral da Proporção

$E[\hat{p}] = \mu_{\bar{p}} = p$
$Var[\hat{p}] = \sigma^2_{\hat{p}} = \frac{p(1-p)}{n}$
Portanto, a distribuição amostral de $\hat{p}$ é dada

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right).$$

Ainda, $\hat{p}$ um estimador não viciado e consistente para p.

- O **erro padrão** de $\hat{p}$ é dado pelo

$$EP[\hat{p}] = \sqrt{Var[\hat{p}]} = \sqrt{\frac{p(1-p)}{n}}.$$

- Pelo **TCL**, a quantidade

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0, 1).$$

Exercício 2

Suponha que a proporção de peças fora da especificação em um lote é de 40%. Uma amostra de 30 peças foi selecionada. Qual é a probabilidade da proporção de peças defeituosas ser menor do que 0,5?

1. Faça o cálculo considerando a distribuição Binomial.
2. Faça o cálculo considerando uma aproximação pela distribuição Normal.

Exercício 2

1. Distribuição Binomial: $X \sim Bin(n = 30, p = 0,4)$.

$$P(\hat{p} < 0,5) = P(X < 15) = \sum_{x=0}^{14} \binom{30}{x} 0,4^x 0,6^{30-x} = 0,825.$$

2. Aproximação pela distribuição Normal.

$$\hat{p} \sim N\left(0,4, \frac{0,4(1-0,4)}{30}\right)$$

.

$$P(\hat{p} < 0,5) \approx P\left(\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} < \frac{0,5 - 0,4}{\sqrt{\frac{0,4(1-0,4)}{30}}}\right)$$

$$P(\hat{p} < 0,5) \approx P(Z < 1,12) = 0,869.$$