Inferência Estatística

Luiz Max de Carvalho[lmax.fgv@gmail.com]

Disciplina da graduação em Matemática Aplicada
Escola de Matemática Aplicada (EMAp/FGV), Rio de Janeiro.

5 de Agosto de 2020

Bem-vindas (os)!

Este é um curso de 60 (sessenta) horas sobre Inferência Estatística.

Princípios:

- △ Em uma palavra: <u>Liberdade</u>;
- △ Construção conjunta do conhecimento;
- △ Pontualidade na entrega das tarefas;
- △ Participação em aula;

Burocracia:

- □ Horário de atendimento: Segundas e quartas de 14:30h a 15:30h.
  - ◇ Por favor, mandar e-mail com antecedência de 24h para marcar;
- □ Podem escrever por e-mail (ou carta) quando quiserem;
- □ Teremos duas avaliações (A1 e A2) e 4 (quatro) trabalhos (T_i, $i = 1, 2, 3, 4$). Trabalhos valerão 20% do grau final.
- □ Nota final será NF $:= \frac{3A_1 + 7A_2}{10} + \sum_{i=1} T_i$.

- Desigualdade de Markov;
- Desigualdade de Chebychev;
- Convergência;
- Lei(s) dos grandes números;
- Teorema(s) Central(is) do Limite;

Desigualdade de Markov[1]

Seja X uma variável aleatória não-negativa e $t > 0$. Então

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

Prova: Assumindo que X é absolutamente contínua, escrever $E[X]$ explicitamente e usar linearidade e monotonicidade da integral. Para $n = 1$ e X discreta, ver De Groot, página 349, Teorema 6.2.1 $\square$

[1] Em homenagem a Andrey Andreyevich Markov (1856–1922).

Desigualdade de Chebychev[2]

Seja Y uma variável aleatória com média $E[Y] =: \mu$ e variância $\text{Var}(Y) =: \sigma^2$, ambas finitas. Mais uma vez, $t > 0$. Então

$$\Pr(|Y - \mu| \geq t) \leq \frac{\text{Var}(Y)}{t^2}. \tag{2}$$

Prova: Notar que $E[(Y - \mu)^2] = \sigma^2$ e aplicar Markov. Ver De Groot, página 349, Teorema 6.2.2 $\square$

[2] Em homenagem a Pafnuty Lvovich Chebyshev (1821–1894).

A média amostral

Considere uma **amostra aleatória** $X_1, X_2, \ldots, X_n$, $n \in \mathbb{N}$ de variáveis aleatórias de uma mesma distribuição com média $E[X_i] = \mu$ e variância $\mathrm{Var}(X_i) = \sigma^2$.

Definição 1

Média amostral. *A média amostral de $X_1, X_2, \ldots, X_n$ é*

$$\bar{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i. \tag{3}$$

Teorema 1

Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias independentes e identicamente distribuídas, com média μ e variância σ^2. Temos que (i) $E[\bar{X}_n] = \mu$ e (ii) $\mathrm{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$.

Prova: Para (i), usar a linearidade da esperança e o fato de as variáveis serem identicamente distribuídas – note a falta de menção à independência. Para (ii), usar a soma das variâncias de variáveis idependentes, além do fato de serem identicamente distribuídas. Ver De Groot, página 350, Teorema 6.2.3.

Exemplo: determinando o tamanho de amostra (1)

Vamos estudar o exemplo 6.2.3 de De Groot. Suponha que uma moeda justa é lançada n vezes. Seja X_i a variável aleatória que é 1 se o i-ésimo lançamento dá cara e 0 caso contrário. Considere $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Pergunta 1

Quantos lançamentos devemos fazer para que

$$\Pr(0.4 \leq \bar{X}_n \leq 0.6) \geq 0.7\,?$$

Resolução: Primeiro, faça $S_n = \sum_{i=1}^{n} X_i$ e deduza que $E[S_n] = np = n/2$ e $\mathrm{Var}(S_n) = np(1-p) = n/4$. Agora:

$$\Pr(0.4 \leq \bar{X}_n \leq 0.6) = \Pr\left(\frac{4n}{10} \leq S_n \leq \frac{6n}{10}\right).$$

Subtraia $E[S_n]$ dos dois lados da desigualdade para obter

$$\Pr\left(\frac{4n}{10} \leq S_n \leq \frac{6n}{10}\right) = \Pr\left(\left|S_n - \frac{n}{2}\right| \leq \frac{n}{10}\right).$$

Exemplo: determinando o tamanho de amostra (2)

Note que, usando a desigualdade de Chebychev, temos uma cota superior para

$$\Pr\left(\left|S_n - \frac{n}{2}\right| \geq \frac{n}{10}\right) = 1 - \Pr\left(\left|S_n - \frac{n}{2}\right| \leq \frac{n}{10}\right).$$

Portanto

$$\Pr\left(\left|S_n - \frac{n}{2}\right| \geq \frac{n}{10}\right) \leq \frac{100}{4n}$$

e então

$$\Pr(0.4 \leq \bar{X}_n \leq 0.6) = \Pr\left(\left|S_n - \frac{n}{2}\right| \leq \frac{n}{10}\right) \geq 1 - \frac{25}{n}.$$

Resolvendo $1 - 25/n = 0.7$ obtemos $n \geq 84$.

Exemplo: determinando o tamanho de amostra (3)

Agora, vamos usar o que sabemos **especificamente** sobre este problema. Usando uma tabela de probabilidades binomiais ou rodando um programa como:

```
calcula_p <- function(n){
  prob <- pbinom(q = round(0.6*n), size = n, p = .5)
  - pbinom(q = round(0.4*n)-1, size = n, p = .5)
  return(prob)
}
n <- 10
p <- calcula_p(n)
alvo <- 0.7
erro <- (p-alvo)^2
while(erro > .001){
  n <- n + 1
  p <- calcula_p(n)
  erro <- (p-alvo)^2
  if(n > 10000) break
}
```

Exemplo: determinando o tamanho de amostra (4)

obtemos `n = 15` e `p = 0.6982422`.
Conclusão: a desigualdade de Chebychev é frouxa, isto é, ela dá uma cota superior para a probabilidade de interesse, mas essa cota pode ser muito maior do que o valor exato. Por outro lado, a desigualdade é válida para qualquer variável aleatória cuja variância exista e seja finita.

Ideia 1

Sem almoço grátis. *Se uma técnica ou resultado é muito geral, isto é, se aplica a muitas situações, há grandes chances de não fornecer uma resposta muito precisa. O contrário também é verdadeiro: se desenvolvemos uma técnica elaborada para uma classe restrita de problemas, geralmente vamos obter respostas precisas, mas nossa técnica não será aplicável a muitos tipos de problemas. Em Estatística não existe almoço grátis.*

Definição 2

Convergência em probabilidade. *Dizemos que uma sequência de variáveis aleatórias converge em probabilidade para b se, para todo $\epsilon > 0$, temos*

$$\lim_{n \to \infty} \Pr\left(|Z_n - b| < \epsilon\right) = 1.$$

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

Em algumas situações, chamamos a convergência em probabilidade de convergência fraca.

Lei(s) dos Grandes Números (LGN)

A lei fraca dos grandes números é um resultado fundamental da Teoria de Probabilidade, extremamente útil em Estatística.

Teorema 2

Lei Fraca dos Grandes Números. *Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias independentes e identicamente distribuídas, com média μ e variância σ^2. Então*

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

Prova: Usando o teorema 1 e a desigualdade de Chebychev, temos

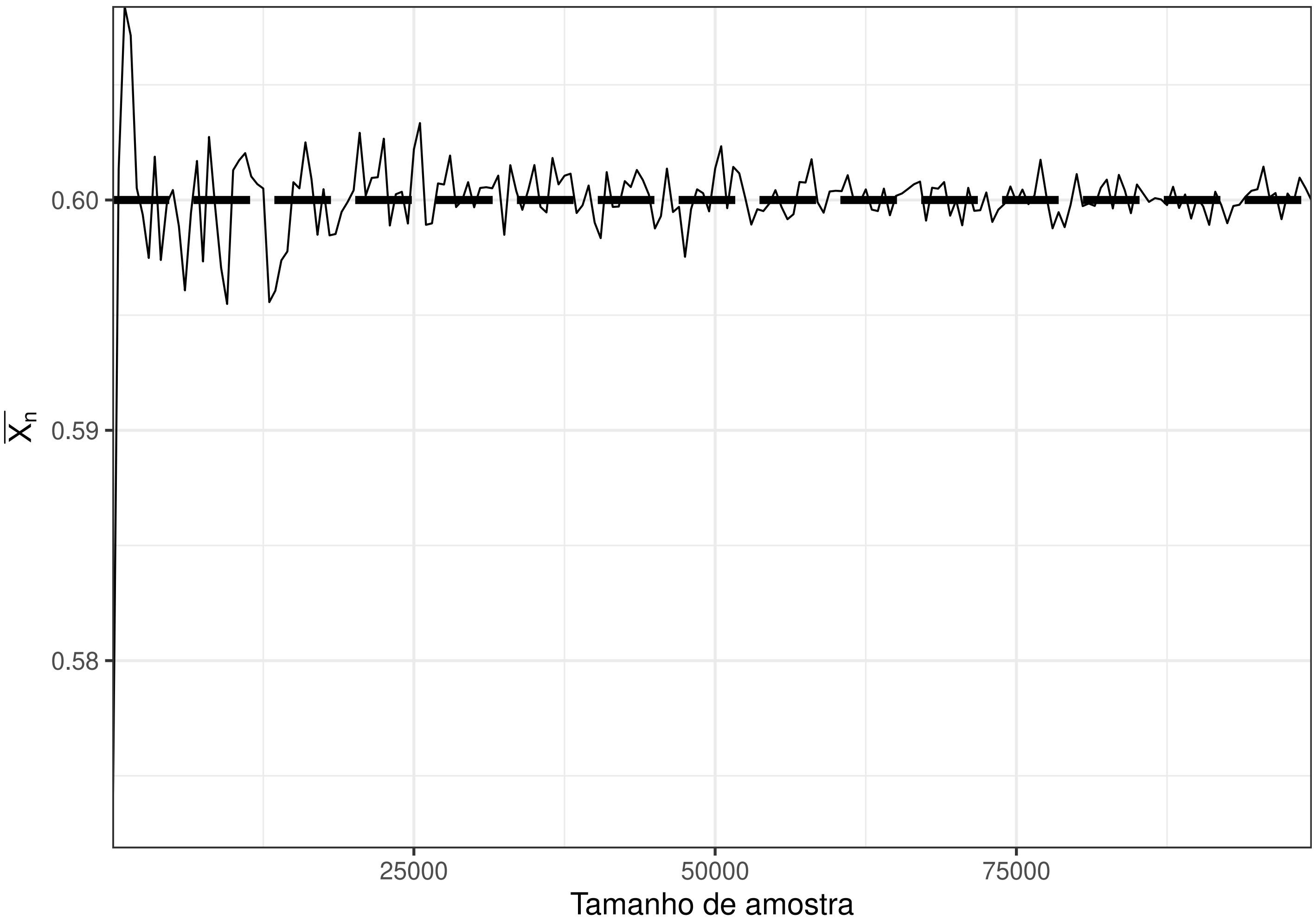
$$\Pr\left(|\bar{X}_n - \mu| < \epsilon\right) \geq 1 - \frac{\sigma^2}{n\epsilon^2},$$

e, portanto,

$$\lim_{n\to\infty} \Pr\left(|\bar{X}_n - \mu| < \epsilon\right) = 1. \quad \square$$

LGN: exemplo

$X_1, \ldots, X_n \sim \text{Beta}(3, 2)$. $E[X] = \alpha/(\alpha + \beta) = 3/5$.



Comentário: Lei Forte dos Grandes Números

Definição 3

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

$$\Pr\left(\lim_{n\to\infty} Z_n = b\right) = 1.$$

Esse modo de convergência é por vezes chamado de convergência forte.
Observação: convergência quase certa implica convergência em probabilidade.

Teorema 3

Lei forte dos grandes números. *Sejam* $X_1, X_2, \ldots$ *variáveis aleatórias independentes e identicamente distribuídas, com média* μ. *Então*

$$\Pr\left(\lim_{n\to\infty} \bar{X}_n = \mu\right) = 1.$$

Teorema(s) Central(is) do Limite

O Teorema Central do Limite um dos resultados mais importantes da Estatística.

Teorema 4

Teorema Central do Limite *(Lindeberg e Lévy)*[3]. *Sejam* $X_1, X_2, \ldots, X_n$ *variáveis aleatórias independentes e identicamente distribuídas, com média* μ *e variância* σ^2. *Então, para cada* x, *temos*

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

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.

Prova: Ver Casella & Berger (2002), página 237, teorema 5.5.14.

[3] Jarl Waldemar Lindeberg (1876–1932) e Paul Pierre Lévy (1886–1971).

Teorema Central do Limite: interpretação

- Sabemos que a variável aleatória padronizada $Y_n := \left(\bar{X}_n - \mu\right)/\sigma$ tem média 0 e variância 1, por construção;
- O teorema 4 nos diz que se tomamos uma amostra grande de uma distribuição com média μ e variância σ^2, a variável aleatória $\sqrt{n}Y_n$ terá, aproximadamente, distribuição **normal** com média 0 e desvio padrão 1, chamada *distribuição normal padrão*;
- Isto equivale a dizer que $\bar{X}_n \sim \text{normal}(\mu, \sigma^2/n)$;
- Note que o teorema vale para <u>qualquer</u> variável aleatória cujos dois primeiros momentos existam, seja ela discreta ou contínua!

Pergunta 2

Suponha que $X_1, \ldots, X_{12}$ são variáveis aleatórias independentes com distribuição uniforme entre 0 e 1. Defina

$$p := \Pr\left(\left|\bar{X}_n - \frac{1}{2}\right| \leq 0.1\right).$$

Quanto vale p?

Resolução: Lembremos que a variável padronizada $Z = \sqrt{n}(\bar{X}_n - E[X])/\sqrt{\text{Var}(X)}$ terá distribuição aproximadamente normal padrão. Se $X \sim$ uniforme$(0, 1)$, sabemos que $E[X] = 1/2$ e $\text{Var}(X) = 1/12$. Nos aproveitando do fato de que $\sqrt{n}$ e σ coincidem nesse exemplo, escrevemos

$$\Pr\left(\left|\bar{X}_n - \frac{1}{2}\right| \leq 0.1\right) = \Pr\left(12\left|\bar{X}_n - \frac{1}{2}\right| \leq 0.1 \times 12\right) = \Pr(|Z| < 1.2),$$

de modo que $p \approx \Phi(1.2) - \Phi(-1.2) = 0.7698607$. O valor exato, que não discutiremos como obter, é $p = 0.7667213$.

O que aprendemos?

- Desigualdades de Markov e Chebychev: extremamente gerais (mas não muito precisas!);
- Convergência fraca (convergência em probabilidade ou medida), $Z \xrightarrow{\mathbf{p}} b$;
- Lei (fraca) dos grandes números: a média amostral converge para a média populacional à medida que a amostra aumenta, $\bar{X}_n \xrightarrow{\mathbf{p}} \mu$;
- Teorema Central do Limite: para amostras grandes o suficiente,

$$\bar{X}_n \sim \text{normal}(\mu, \sigma^2/n).$$

Leitura recommendada

- De Groot seções 6.2 e 6.3;
- * Casella & Berger, seções 5.2 e 5.5;
- * Nosso repositório (https://github.com/maxbiostat/Statistical_Inference_BSc).