



Plano Aula 03 e 04

Markus Stein

Inferência Estatística

Essa semana veremos resultados e extensões de Probabilidade que terão aplicação nas próximas semanas.

- Estatística descritiva \times inferência estatística;
 - população e amostra: parâmetros $(\mu, \sigma^2, \pi, \dots)$ \times estatísticas (\bar{x}, s^2, p, \dots) .

Definição **Estatística**: é qualquer valor obtido em função da amostra. Exemplo, \bar{x}, s^2, p, \dots

Distribuição amostral (Bussab e Morettin - Seção 10.7)

“Toda função de variáveis aleatórias (v.a.s) é uma v.a.”

Definição **Distribuição Amostral**: é a distribuição de probabilidade de uma estatística.

Exemplo 1: Seja X a v.a. que denota o número de livros que a população de monitores do curso ‘Probabilidade e Estatística’ lêem por semestre. Suponha que no último semestre foram lidos 5, 7, 4. Se não soubéssemos essa informação e decidíssemos observar uma amostra de tamanho $n = 2$ para saber a média de livros lidos \bar{X} dessa população.

- a. Quais as possíveis amostras? (Cada amostra pode gerar um \bar{x} diferente)
- b. Os valores de média calculados com cada amostra formam a distribuição amostral de \bar{X} .

Lembrando: Amostra aleatória simples (a.a.s.) = v.a. independentes e identicamente distribuídas (i.i.d.)

Definição **A.A.S**: Seja X_1, X_2, \dots, X_n uma a.a.s. de tamanho n de $X \sim f(x; \theta)$, então $X_1 \sim f(x; \theta), \dots, X_n \sim f(x; \theta)$ e X_i e X_j são independentes para todo $i \neq j$.

Definição **Erro padrão**: é o desvio padrão de uma estatística. Exemplo, erro padrão da *média amostral* é $\sigma_{\bar{X}} = \sqrt{\text{Var}(\bar{X})}$.

Exemplo 2: Proporção amostral $p = \frac{1}{n} \sum_{i=1}^n X_i$, para $X_i \in \{0, 1\}$

Teorema central do limite - TCL (Bussab e Morettin - Seção 10.8)

“Garante que uma média amostral se aproxima do seu valor esperado à medida que o tamanho da amostra aumenta (dadas algumas condições...)”

- Teorema 10.2 e Corolário 10.1
- Aplicativo que ilustra o TCL - https://brunamdalmoro.shinyapps.io/TCL_medias/



Estimação Pontual (Bussab e Morettin - Capítulo 11)

- Estatísticas: Estimador *versus* Estimativa.

(cont.) **Exemplo 1:** Média amostral, $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, em que X_1, \dots, X_n uma amostra aleatória de $X_i \sim Normal(\mu, \sigma^2)$ e σ^2 conhecido:

- Qual a distribuição amostral de \bar{X} ?
- \bar{X} é um bom estimador para a média populacional μ ?
- Como usar $Var(\bar{X})$ para fornecer um grau de certeza sobre usarmos \bar{X} para representar/estimar μ ?

- se $X \sim Normal$ então $\bar{X} \sim Normal\left(\mu, \frac{\sigma^2}{n}\right)$! Suposições?

Definição (**Estimador**): Um estimador T do parâmetro θ é qualquer função das observações da amostra, $T = g(X_1, \dots, X_n)$.

Definição (**Estimativa**): Uma estimativa é um particular valor do estimador. Para uma amostra observada x_1, \dots, x_n uma estimativa t do parâmetro θ é dada por $t = g(x_1, \dots, x_n)$.

(cont.) **Exemplo 1:** E para a média amostral $\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ se σ^2 desconhecido?

- vale o resultado $\bar{X} \sim Normal\left(\mu, \frac{\sigma^2}{n}\right)$? (Suposições?)
 - se n for “pequeno” e $X \sim Normal$, então X não segue uma $Normal$;
 - se n for “grande”, então pelo TCL teremos que $\bar{X} \sim Normal$.

(cont.) **Exemplo 2:** E para a proporção amostral $\hat{\pi} = P = \frac{1}{n} \sum_{i=1}^n X_i$?

- quando $P \sim Normal\left(\pi, \frac{\pi(1-\pi)}{n}\right)$? Suposições?

Exemplo 3: Simulação de distribuições de estimadores (*estatísticas*)

- Para a variância amostral $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - E(X))^2$? E para $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - E(X))^2$?
- E para outras estatísticas como a mediana, Md , o máximo, $X_{(n)}$, ou o mínimo, $X_{(1)}$, ...?

Propriedades dos estimadores (Bussab e Morettin - Seção 11.2)

(cont.) **Exemplo 1:** ...

- Viés e o Erro Quadrático Médio (EQM)
- Constistência
- Eficiência



Ler slides das aulas 3 a 4

Fazer exercícios lista 1-2

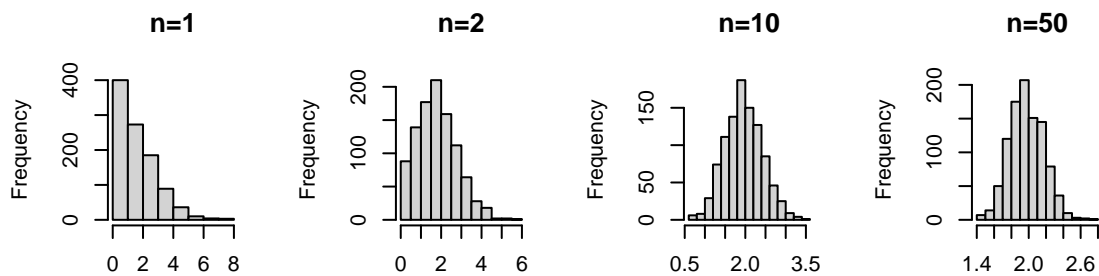
Fazer avaliação pontual 1 da área 1 - vale nota!!!

Ilustração do TCL no R

No R é possível gerar amostras, calcular a média de cada amostra e plotar o histograma:
(usamos *replicate* para gerar 1000 amostras de tamanho $n = 1, 2, 10$ e 50)

- V.a. $X \sim \text{Poisson}(2)$. (para $n = 1$ geramos amostra da **distribuição real** de X)

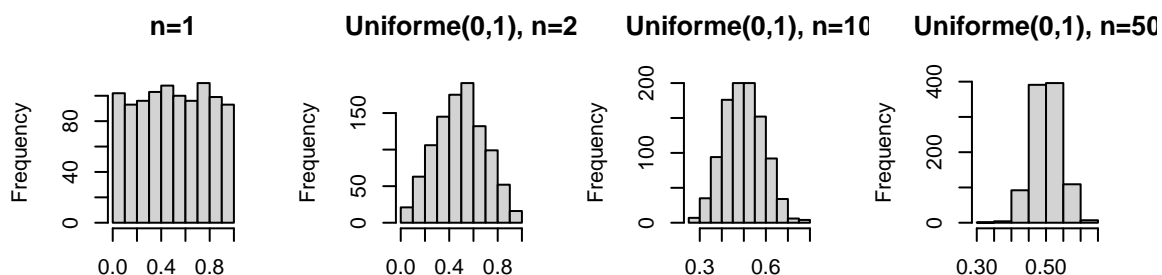
```
par(mfrow=c(1,4))
hist( rpois( n = 1000, lambda = 2), main = "n=1")
hist( colMeans( replicate( n = 1000, rpois( n = 2, lambda = 2))), main="n=2")
hist( colMeans( replicate( n = 1000, rpois( n = 10, lambda = 2))), main="n=10")
hist( colMeans( replicate( n = 1000, rpois( n = 50, lambda = 2))), main="n=50")
```



`rpois(n = 1000, lambda = 2)`
`replicate(n = 1000, rpois(n = 2))`
`replicate(n = 1000, rpois(n = 10))`
`replicate(n = 1000, rpois(n = 50))`

- V.a. de $X \sim \text{Uniforme}(0,1)$.

```
par(mfrow=c(1,4))
hist( replicate( n = 1000, runif( n = 1, min = 0, max = 1)), main="n=1")
hist( colMeans( replicate( n = 1000, runif( n = 2, min = 0, max = 1))), main="Uniforme(0,1), n=2")
hist( colMeans( replicate( n = 1000, runif( n = 10, min = 0, max = 1))), main="Uniforme(0,1), n=10")
hist( colMeans( replicate( n = 1000, runif( n = 50, min = 0, max = 1))), main="Uniforme(0,1), n=50")
```



`runif(n = 1000, min = 0, max = 1)`
`replicate(n = 1000, runif(n = 2, min = 0, max = 1))`
`replicate(n = 1000, runif(n = 10, min = 0, max = 1))`
`replicate(n = 1000, runif(n = 50, min = 0, max = 1))`