Departamento de Estatística e Matemática Aplicada da UFC
CC0288- Inferência Estatística I.
Estimador Consistente -19/04/2023
Professor: Maurício Mota

1. **Consistência**

   O Vladison (turma 22.2) na última aula me pediu para resolver um exercício sobre a consistência de um estimador.

   Esta pergunta gerou a seguinte aula. Vamos aproveitar?

   Definição: Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de uma variável aleatória de X que depende de um parâmetro θ

   Dizemos que o estimador $T_n = h(x_1, , X_2, \ldots, X_n)$ é consistente para θ se para todo $\epsilon > 0$ temos:

$$\lim_{n \to \infty} P\left(|T_n - \theta| > \epsilon\right) = 0.$$

   Dizemos também que T_n converge em probabilidade para θ.

   Exemplo 1: Seja X uma variável aleatória X com $E(X) = \theta$ e $V(X) = \sigma^2$. Considere $X_1, X_2, \ldots, X_n$ uma amostra aleatória de uma variável aleatória de X.
   Mostre que $\bar{X}$ é um consistente para θ.

   Devemos provar que

$$\lim_{n \to \infty} P\left(|\bar{X} - \theta| > \epsilon\right) = 0.$$

   Sabemos que:

$$E(\bar{X}) = \theta \quad e \quad Var(\bar{X}) = \frac{\sigma^2}{n}.$$

   Vamos aplicar a desigualdade de Chebyshev:

$$P\left(|\bar{X}-\theta|>\epsilon\right) \leq \frac{V(\bar{X})}{\epsilon^2} = \frac{\frac{\sigma^2}{n}}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$$

$$\lim_{n\to\infty} P\left(|\bar{X}-\theta|>\epsilon\right) = \sigma^2 \lim_{n\to\infty} \frac{1}{n} = 0.$$

Provando assim a consistência de $\bar{X}$.

Vamos provar que a consistência de T_n também pode definida como:

i. T_n é um estimador assintoticamente não viciado para θ, isto é,

$$\lim_{n\to\infty} E(T_n) = \theta.$$

ii.

$$\lim_{n\to\infty} Var(T_n) = 0.$$

Vamos justificar:

$$P\left(|T_n-\theta|>\epsilon\right) = P\left(|T_n-\theta|^2>\epsilon^2\right) = P\left((T_n-\theta)^2>\epsilon^2\right) \leq \frac{E(T_n-\theta)^2}{\epsilon^2}.$$

Note que:

$$E(T_n-\theta)^2 = EQM(T_n) = Var(T_n) + \left[E(T_n)-\theta\right]^2.$$

Note que se as duas condições são satisfeitas então:

$$\lim_{n\to\infty} E(T_n-\theta)^2 = 0$$

provando assim a consistência de T_n.

Vamos verificar a consistência por simulação:

2. Vamos reproduzir o capítulo 22 da apostila do professor Paulo Justiniano.

   Um estimador é consistente quando seu valor se aproxima do verdadeiro valor do parâmetro à medida que aumenta-se o tamanho da amostra. Vejamos como podemos ilustrar este resultado usando simulação.

A ideia básica é a seguinte:

1. escolher uma distribuição e seus parâmetros,
2. definir o estimador,
3. definir uma sequência crescente de valores de tamanho de amostras,
4. obter uma amostra de cada tamanho,
5. calcular a estatística para cada amostra,
6. fazer um gráfico dos valores das estimativas contra o tamanho de amostra, indicando neste gráfico valor verdadeiro do parâmetro.

Média da distribuição normal

Seguindo os passos acima vamos:

Passo 1: tomar a distribuição Normal de média 10 e variância 4.

Passo 2: definir o estimador

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

Passo 3: escolhemos os tamanhos de amostra $n = 2, 5, 10, 15, 20, \ldots, 1000, 1010, 1020, \ldots, 5000$,

Passo 4: fazemos os cálculos e produzimos um gráfico como mostrado na 27 com os comandos a seguir.

Veja com carinho o gráfico:

Médias de amostras de diferentes tamanhos.

Este gráfico é obtido no **R** através dos comandos:

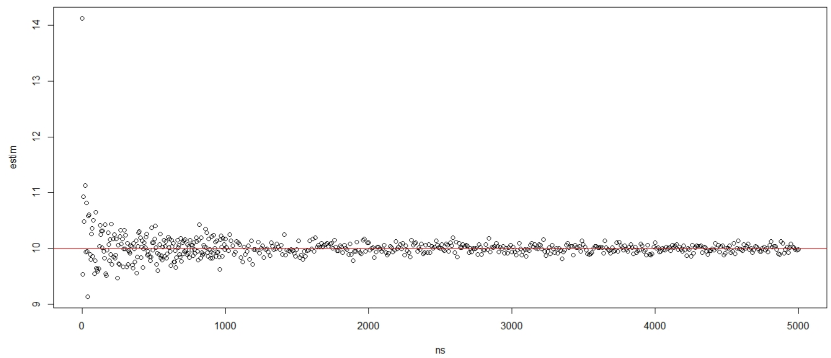

Figura 1:

```
ns <- c(2, seq(5, 1000, by=5), seq(1010, 5000, by=10))
estim <- numeric(length(ns))
for (i in 1:length(ns)){
amostra <- rnorm(ns[i], 10, 4)
estim[i] <- mean(amostra)
}
plot(ns, estim)
abline(h=10,col="red")

########Vamos Comentar!!!!!
```

3. **Momentos das distribuições amostrais de estimadores**

Para Inferência Estatística é necessário conhecer a distribuição amostral dos estimadores. Em alguns casos estas distribuições são derivadas analiticamente. Isto se aplica a diversos resultados vistos em um curso de Inferência Estatística. Por exemplo seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de

$$X \sim N(\mu, \sigma^2).$$

A distribuição da média amostral $\bar{X}$ é:

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

Além disto este procedimento permite investigar distribuições amostrais que são complicadas ou não podem ser obtidas analiticamente.

Vamos ver um exemplo: Suponha que temos interesse no parâmetro:

$$\theta = \frac{\mu}{\sigma^2}.$$

Para estimar θ vamos usar o seguinte estimador:

$$T = \frac{\bar{X}}{S^2},$$

em que $\bar{X}$ é a média amostral e S^2 a variância amostral.

Para obter por simulação a esperança e variância do estimador T devemos seguir os seguintes passos.

1. escolher uma distribuição e seus parâmetros, no caso vamos escolher uma

$$N(\mu = 180, \sigma^2 = 64),$$

2. definir um tamanho de amostra, no caso escolhemos $n = 20$,

3. obter por simulação um número N de amostras, vamos usar $N = 1000$,

4. calcular a estatística de interesse para cada amostra,

5. usar as amostras para obter as estimativas $E[T]$ e $Var[T]$.

Vamos ver agora comandos do **R**:

```
amostras <- matrix(rnorm(20*1000, mean=180, sd=8), nc=1000)

Tvals = apply(amostras, 2, function(x) {mean(x)/var(x)})

ET_est = mean(Tvals)
```

```
ET_est
VarT_est=  var(Tvals)
VarT_est
```

Nestes comandos primeiro obtemos 1000 amostras de tamanho 20 que armazenamos em uma matriz de dimensão 20×1000, onde cada coluna é uma amostra. A seguir usamos a função **apply** para calcular a quantidade desejada que definimos com

$$function(x)\ \ mean(x)/var(x).$$

No caso anterior foi obtido:

Se você rodar os comandos acima deverá obter resultados um pouco diferentes (mas não muito!) pois nossas amostras da distribuição normal não são as mesmas.

Veja este fato aqui.

```
> ###Vamos rodar mais uma vez!!!!
>
>
>
> amostras <- matrix(rnorm(20*1000, mean=180, sd=8), nc=1000)
>
>  Tvals = apply(amostras, 2, function(x) {mean(x)/var(x)})
>
>  ET_est = mean(Tvals)
>  ET_est
[1] 3.105506
>  VarT_est=  var(Tvals)
>  VarT_est
[1] 1.259352
>
>
> ####Mais outra vez!!!!
>
>
> amostras <- matrix(rnorm(20*1000, mean=180, sd=8), nc=1000)
```

```
>
> Tvals = apply(amostras, 2, function(x) {mean(x)/var(x)})
>
> ET_est = mean(Tvals)
> ET_est
[1] 3.14434
> VarT_=  var(Tvals)
> VarT_est
[1] 1.259352
>
```

4. **Exercícios:**

   1. Estude a consistência do estimador

   $$\hat{\lambda} = \frac{1}{\bar{X}},$$

   para o parâmetro λ de $X \sim Exp(\lambda)$.

   2. No exemplo dos momentos das distribuições de estimadores visto anteriormente ilustramos a obtenção dos momentos para um tamanho fixo de amostra $n = 20$. Repita o procedimento para vários tamanhos de amostra e faça um gráfico mostrando o comportamento de $\hat{E}[T]$ e $\hat{V}[T]$ em função de n.

3. Estime por simulação a esperança e a variância do estimador

$$\hat{\lambda} = \bar{X},$$

do parâmetro λ de uma Poisson para um tamanho de amostra $n = 30$. Compare com os valores obtidos analiticamente. Mostre em um gráfico como os valores de $\hat{E}[\hat{\lambda}]$ e $\hat{V}[\hat{\lambda}]$ variam em função de de n.

4. Crie um exemplo para ilustrar a não tendenciosidade de estimadores.
Sugestão: compare os estimadores

$$S^2 = \frac{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2}{n-1} \quad , \quad \hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2}{n} = \frac{n-1}{n}\, S^2,$$

da variância σ^2 de uma distribuição normal.

5. Crie um exemplo para comparar a variância de dois estimadores. Por exemplo compare por simulação as variâncias dos estimadores

$$T_1 = \bar{X} \quad e \quad T_2 = \frac{Y_n + Y_1}{2},$$

do parâmetro μ de uma $N(\mu, \sigma^2)$ em
que

$$Y_1 = min(X_1, X_2, \ldots, X_n) \quad e \quad Y_n = max(X_1, X_2, \ldots, X_n).$$