Instituto Tecnológico de Aeronáutica
Divisão de Engenharia Mecânica

MOQ-13

Probabilidade e Estatística

Prof. Mauri Aparecido de Oliveira

mauri@ita.br

2019

Estimação de Parâmetros

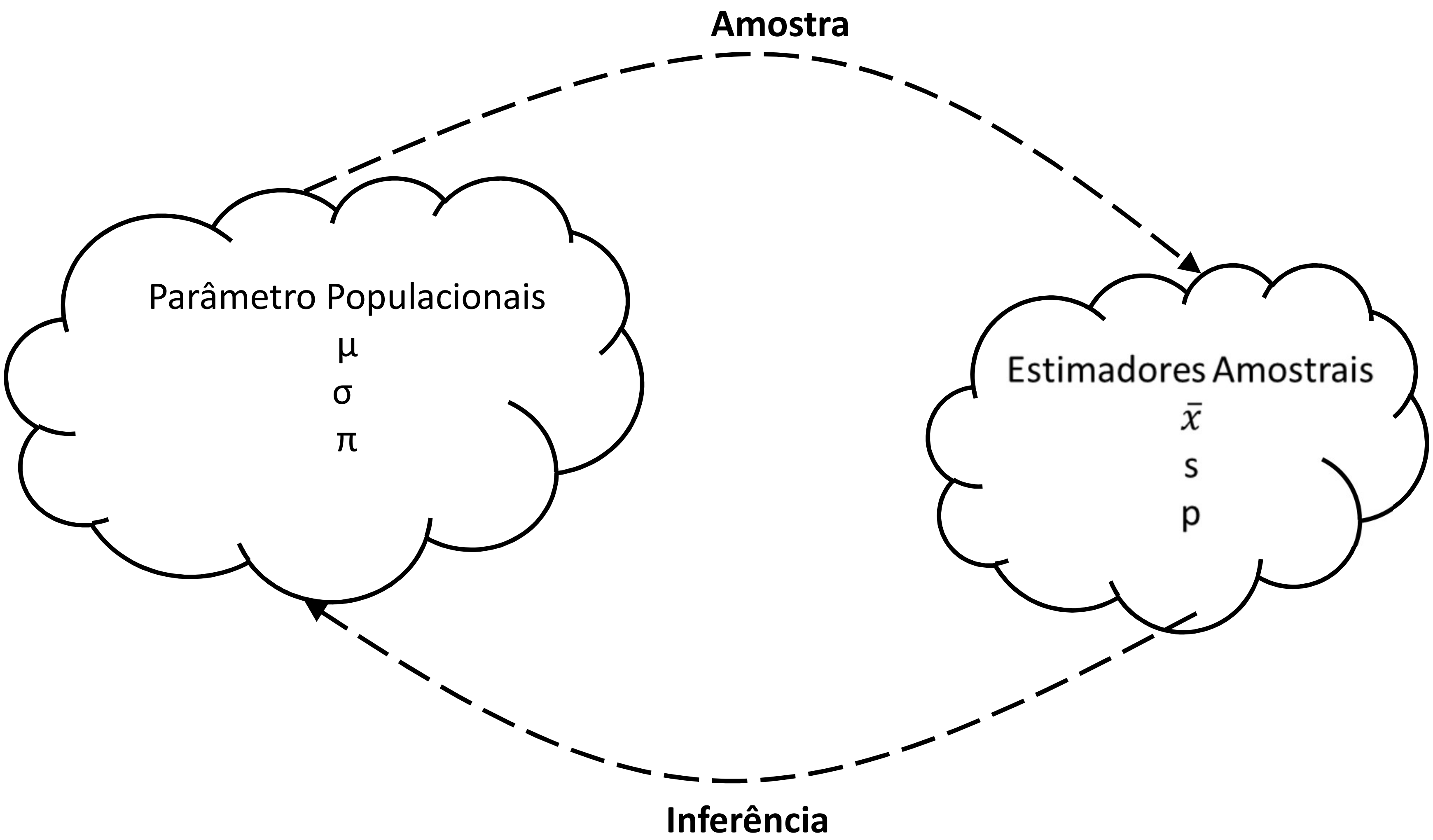
SEMANAS 08/09: Roteiro

- Inferência Estatística: Estimação de Parâmetros
  – Estimação pontual
  – Estimação por intervalos

Estimadores

Um **estimador** é uma estatística derivada de uma amostra para inferir o valor de um parâmetro populacional. Uma **estimativa** é o valor do estimador em uma amostra particular. A Tabela e a Figura a seguir mostram alguns estimadores comuns. Usualmente, denotamos um parâmetro populacional por uma letra grega (por exemplo, μ, σ ou π). O correspondente estimador amostral é usualmente denotado por uma letra latina (por exemplo, $\bar{x}$, s ou p) ou uma letra grega com um "acento circunflexo" (por exemplo, $\hat{\mu}$, $\hat{\sigma}$ ou $\hat{\pi}$).

Estimador	Definição	Parâmetro Populacional
Média Amostral	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, em que x_i é o i-ésimo valor dos dados e n é o tamanho da amostra.	μ
Proporção Amostral	$p = \frac{x}{n}$, em que x é o número de sucessos na amostra e n é o tamanho da amostra.	π
Desvio-padrão Amostral	$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$, em que x_i é o i-ésimo valor dos dados e n é o tamanho da amostra.	σ


Amostra
Parâmetro Populacionais
μ
σ
π
Estimadores Amostrais
$\bar{x}$
s
p
Inferência

Erro Amostral

Na melhor das hipóteses, poderíamos encontrar um estimador $\hat{\theta}$ para o qual $\hat{\theta} = \theta$, sempre. Entretanto, $\hat{\theta}$ é uma função da amostra dos X_i, portanto, é uma variável aleatória. Para algumas amostras, $\hat{\theta}$ produzirá um valor um valor maior que θ, enquanto, para outras, $\hat{\theta}$ subestimará θ. Se escrevermos

$$\hat{\theta} = \theta + erro\ de\ estimação,$$

então, um estimador acurado seria um resultante de um erro de estimação pequeno, tal que os valores estimados estariam próximos ao valor real.

Erro Amostral

Amostras aleatórios variam, portanto um estimador é uma variável aleatória. O erro amostral é a diferença entre uma estimativa e o parâmetro de população correspondente. Por exemplo, para a média da população:

$$\text{Erro Amostral} = \bar{x} - \mu.$$

O erro amostral existe porque amostras diferentes fornecerão valores diferentes para $\bar{x}$, dependendo de quais itens da população foram incluídos na amostra.

Erro Quadrado Médio [Mean Squared Error – MSE]

A qualidade de um estimador está relacionada a quão próximas suas estimativas estão do parâmetro verdadeiro. A diferença entre um estimador, por exemplo, T e o parâmetro desconhecido θ é chamado de erro amostral. Como essa quantidade pode ser positiva ou negativa, é comum usarmos o quadrado do erro para que vários estimadores T_1, T_2, ..., possam ser comparados usando uma medida de erro não negativa.

Para esse fim, o erro quadrático médio de um estimador, denominado $MSE[T]$, é definido como

$$MSE[T] = E[T - \theta]^2.$$

Os estimadores com pequenos MSE's terão uma distribuição tal que os valores na distribuição estarão próximos do parâmetro verdadeiro.
De fato, o MSE consiste de dois componentes não-negativos, a variância do estimador T e o viés (vício, ou bias) ao quadrado do estimador T, onde o viés é definido como $E[T] - \theta$, uma vez que

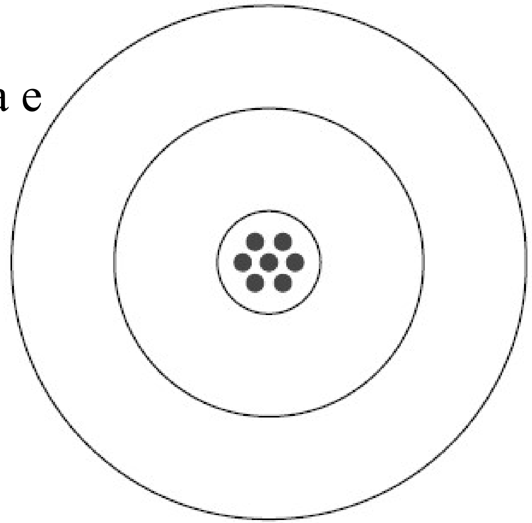
$$\begin{aligned}
MSE &= E\left[\left(T - E[T] + E[T] - \theta\right)^2\right] \\
&= E\left[\left(T - E[T]\right)^2\right] + E\left[\left(E[T] - \theta\right)^2\right] + 2E\left[\left(T - E[T]\right)\left(E[T] - \theta\right)\right] \\
&= Var[T] + E\left[\left(E[T] - \theta\right)^2\right] + 2E\left[TE[T] - T\theta - E[T]E[T] + E[T]\theta\right] \\
&= Var[T] + E\left[E^2[T] - 2E[T]\theta + \theta^2\right] + 0 \\
&= Var[T] + \left(E[T] - \theta\right)^2
\end{aligned}$$

$$MSE = Var[T] + \left(Bias[T]\right)^2$$

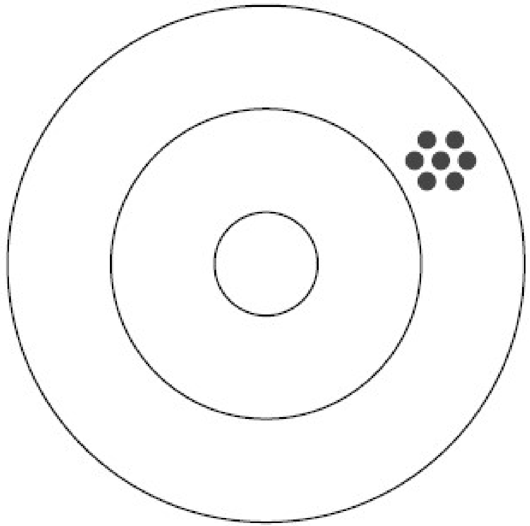
Nesse ponto, podemos apresentar os conceitos de acurácia e precisão. A **acurácia** mede a proximidade de cada observação do valor alvo que se procura atingir. A **precisão** mede a proximidade de cada observação da média de todas as observações.

Baixa Variância, Baixo Viés

Não-viesada, muito acurada e muito precisa

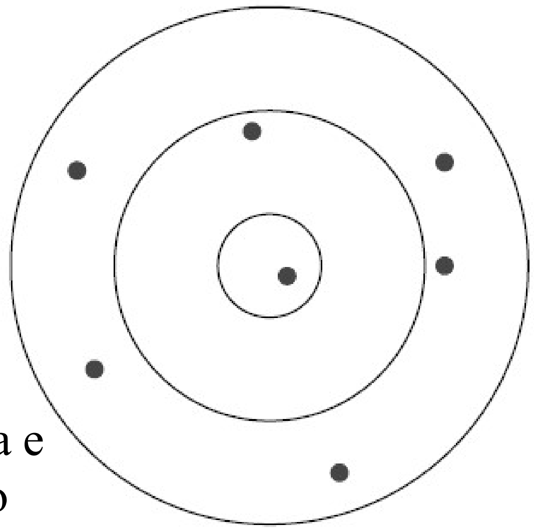


Baixa Variância, Alto Viés



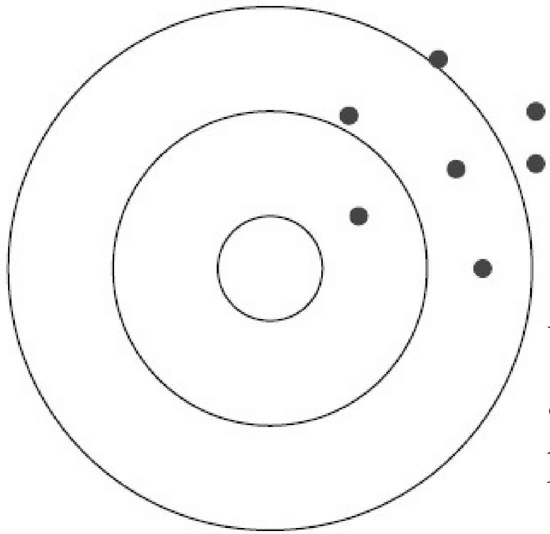
Viesada, pouco acurada e alta precisão

Alta Variância, Baixo Viés



Não-viesada, pouco acurada e baixa precisão

Alta Variância, Alto Viés



Viesada, pouco acurada e baixa precisão

Estimação de Parâmetros

De maneira geral, se θ for o parâmetro populacional de interesse, então:

Estimador ($\hat{\Theta}$) é uma função (v.a.) dos valores não-observados de uma a.a.

$$\hat{\Theta} = g(X_1, X_2, \ldots, X_n)$$

Estimativa ($\hat{\theta}$) é um valor numérico calculado a partir dos valores efetivamente observados de uma a.a.

$$\hat{\theta} = g(x_1, x_2, \ldots, x_n)$$

$\hat{\theta}$ é uma *realização* de $\hat{\Theta}$.

A ideia é escolher $\hat{\Theta}$ capaz de nos dar uma boa indicação a respeito do valor "real" do parâmetro θ...

Estimação de Parâmetros
Motivação

1. Para estimar λ:

Sabemos que, para $X \sim \text{Pois}(\lambda)$, temos:

$$\lambda = E[X] \qquad \text{e} \qquad \lambda = \text{Var}[X]$$

- As características da distribuição de Poisson sugerem que temos dois estimadores naturais para λ:

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \text{e} \qquad S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

- outros estimadores são possíveis...

Estimação de Parâmetros

Motivação

2. Para estimar p_o:

Temos que

$$p_o = P[X = 0] = e^{-\lambda}$$

- Um estimador natural para p_o é a frequência relativa de zeros na a.a.:

$$\frac{\text{número de observações em que } X_i = 0}{n}$$

- Por outro lado, se usamos $\bar{X}_n$ para estimar μ, podemos estimar p_o utilizando:

$$e^{-\bar{X}_n}$$

- aqui também é possível escolher outros estimadores...

Estimação de Parâmetros

Motivação

Temos inúmeras opções de estimadores para os parâmetros de interesse (ou funções deles)!

Como escolher um estimador?

Quando um estimador é melhor que outro?

Existe um estimador que seja o "melhor" de todos?

Estimação de Parâmetros

Motivação

Podemos determinar qual das estimativas ...

$$\bar{x}_n \quad \text{ou} \quad s_n^2$$

... se aproxima mais do valor "real" do parâmetro λ?

A resposta é NÃO!

As observações que compõe uma a.a., bem como as estimativas obtidas a partir dela, são sujeitas à aleatoriedade...

... portanto, não é possível afirmar com certeza qual dos valores estimados é mais próximo de λ!

Características Desejáveis de Estimadores

– INVESTIGANDO O COMPORTAMENTO DE UM ESTIMADOR –

Estimação de Parâmetros

Propriedades Desejáveis de Estimadores

Temos um conjunto de dados, ou uma "base de dados".

O conjunto de dados é uma realização de uma a.a. $X_1, X_2, \ldots, X_n$, de tamanho $n = 30$ de uma população Pois(λ).

Com base neste conjunto de dados, vamos proceder com a estimação dos parâmetros de interesse:

$$\lambda \qquad \text{e} \qquad p_o$$

Estimação de Parâmetros

Propriedades Desejáveis de Estimadores

Vamos começar estimando λ.

Temos dois estimadores candidatos: $\bar{X}_n$ e S_n^2.

Qual deles devemos escolher?

Estimação de Parâmetros

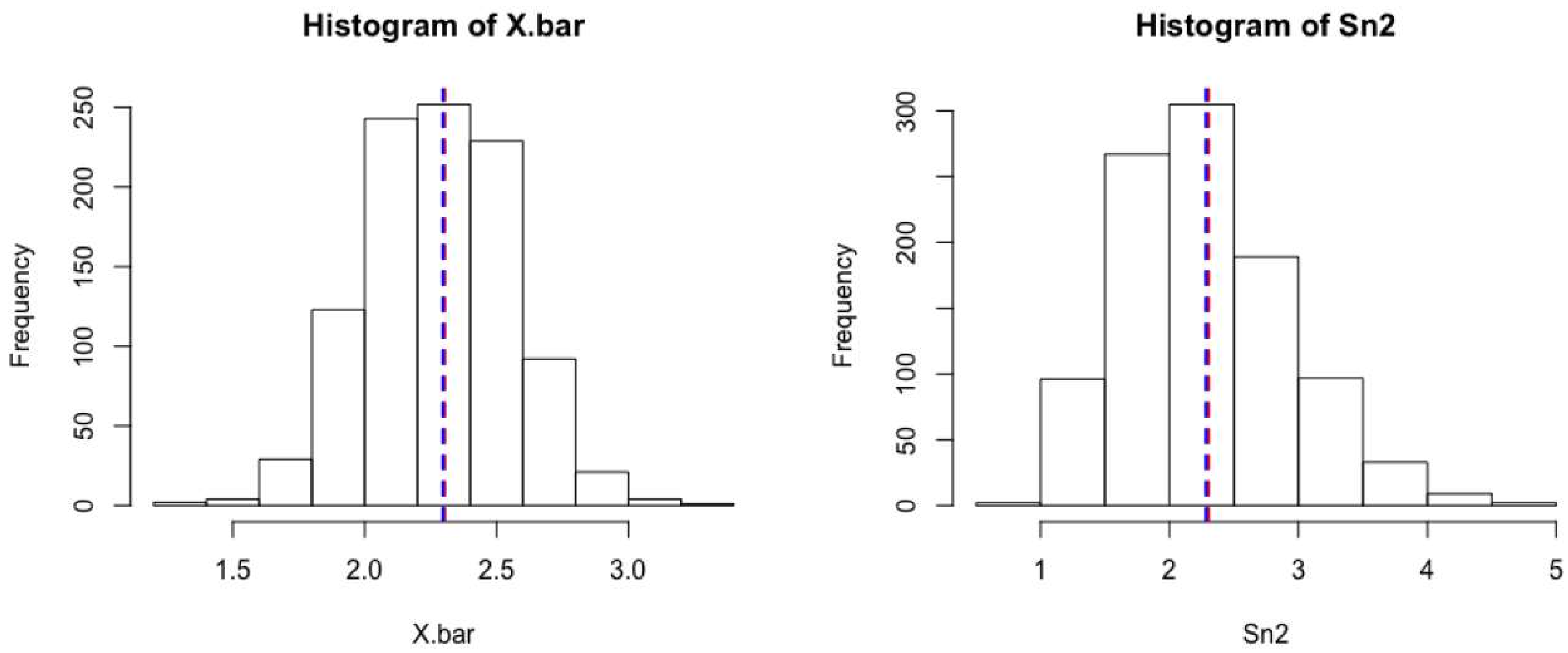
Propriedades Desejáveis de Estimadores

Vamos simular:

– Fingimos que conhecemos o valor de λ $(= \ln 10)$ tq. $p_o = 0,1$;
– Amostramos $n = 30$ observações a partir de Pois $(\lambda = \ln 10)$;
– Calculamos os valores que $\bar{X}_n$ e S_n^2 assumem para as observações da a.a.;
– Repetimos esse procedimento um grande número de vezes, $n \quad = 1000$;
– Construímos a distribuição amostral para cada estimador:

Estimação de Parâmetros

Propriedades Desejáveis de Estimadores



Os valores estimados para λ parecem flutuar em torno do valor "real" ($\ln 10 \approx 2,3$).

Estimação de Parâmetros

Propriedades Desejáveis de Estimadores

De fato, é possível demonstrar que :

$$E[\bar{X}_n] = \lambda \qquad \text{e} \qquad E[S_n^2] = \lambda$$

Ou seja, as distribuições amostrais de ambos os estimadores estão centradas no parâmetro λ que se quer estimar.

Isto significa que os estimadores $\bar{X}_n$ e S_n^2 não apresentam uma tendência sistemática de produzir estimativas nem maiores (superestimadas) nem menores (subestimadas) do parâmetro de interesse, λ.

Esta é uma característica muito desejável em um estimador.

Estimação de Parâmetros

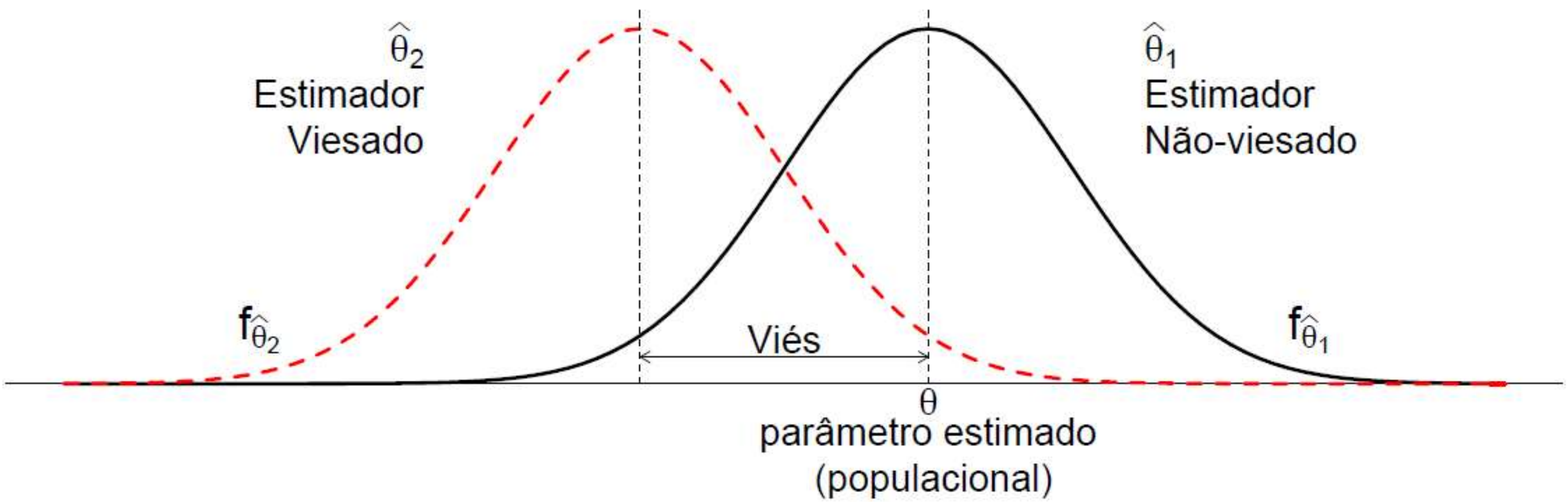
Propriedades Desejáveis de Estimadores

Não-tendenciosidade (exatidão):

- Um estimador não-viesado é aquele que não tende a subestimar ou superestimar o valor do parâmetro real da população, θ.
- A distribuição amostral de $\hat{\Theta}$ está centrada em θ:

$$E[\hat{\Theta}] = \theta \quad \Longrightarrow \quad Viés = b(\hat{\Theta}) = E[\hat{\Theta}] - \theta = 0$$

Distribuições Amostrais de Estimadores



Estimação de Parâmetros

Propriedades Desejáveis de Estimadores

Ótimo, encontramos dois estimadores não-tendenciosos para o parâmetro λ.

Isto significa que é indiferente escolher um dos dois?

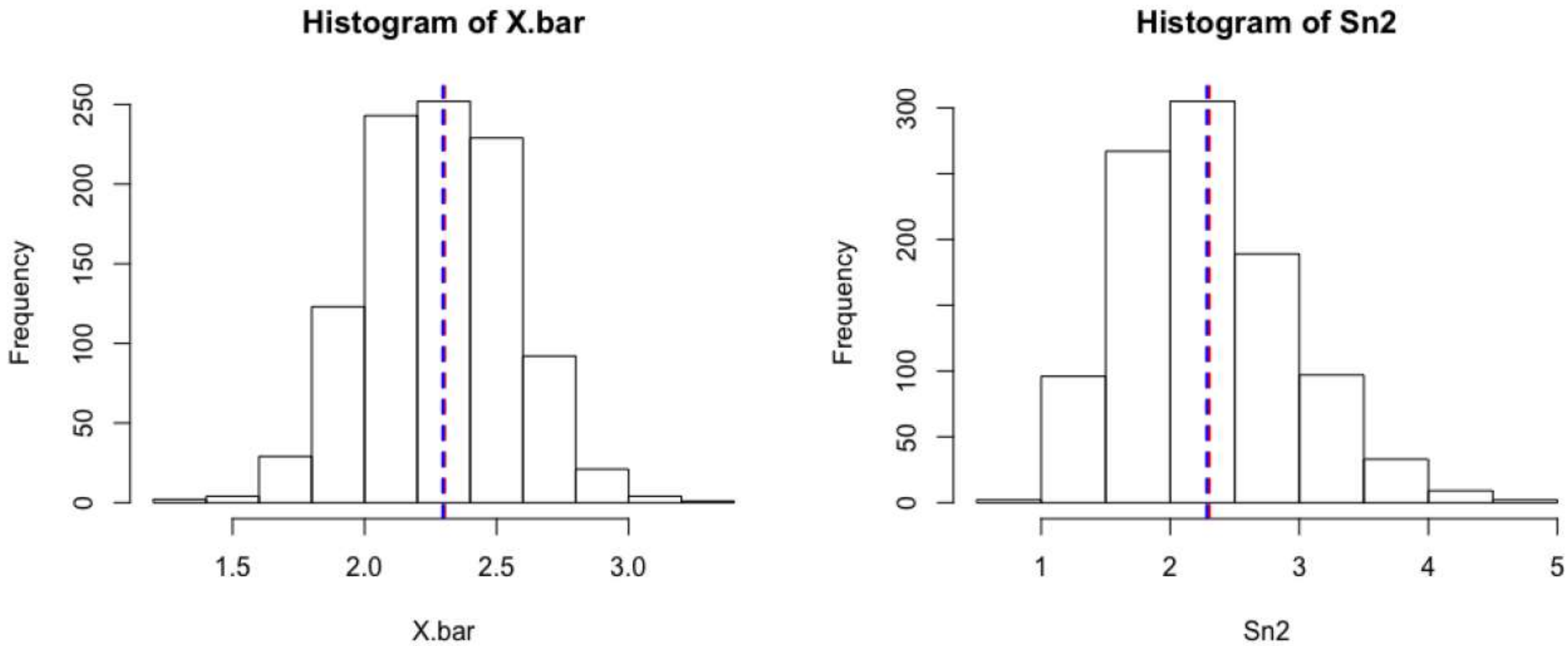
A resposta é NÃO!

Vamos observar novamente os histogramas para os valores obtidos para os dois estimadores...

Qual a diferença entre eles?

Estimação de Parâmetros

Propriedades Desejáveis de Estimadores



Perceba que a variação nos valores de $\bar{X}_n$ é *menor* que a variação nos valores de S_n^2.

Isto indica que o estimador $\bar{X}_n$ estima λ mais eficientemente que S_n^2, já que produz estimativas, em geral, mais próximas do valor real do parâmetro que estamos estimando.

```r
dev.off(dev.list()["RStudioGD"])
rm(list=ls())
cat("\f")


# Definindo os parâmetros da simulação e obtendo as amostras

lambda = log(10)
n = 30
linhas = 1000

par(mfrow=c(1,2))

sim <- rpois(n*linhas, lambda)
#sim
m <- matrix(sim, linhas)
#m


# Considerando a média como estimador de lambda
medias.amostras <- rowMeans(m)
#medias.amostras

media.teorica = lambda
media.amostral = mean(medias.amostras)

hist(medias.amostras)
abline(v=media.teorica,col="red", lwd=3, lty=2)
abline(v=media.amostral,col="blue", lwd=3, lty=2)
expr.media <- vector("expression", 2)
expr.media[[1]] <- bquote("Média Teórica"==.(media.teorica))
expr.media[[2]] <- bquote("Média da Simulação"==.(media.amostral))
legend("topright",
       bty = "n",
       legend = expr.media,,cex=0.7)      # bty = box type, pode ser "o" (default) e "n"

# Comparando a média teórica com a média amostral
estats.media <- matrix(c(media.teorica,media.amostral),ncol=1)
rownames(estats.media) <- c("Média Teórica:","Média da Simulação:")
colnames(estats.media) <- c("Valores")
estats.media
```

```r
# Considerando a variância como estimador de lambda
rowVar <- function(x, ...) {
  rowSums((x - rowMeans(x, ...))^2, ...)/(dim(x)[2] - 1)
                            }
variancias.amostras <- rowVar(m)
#variancias.amostras

variancia.teorica = lambda
variancia.amostral = mean(variancias.amostras)

hist(variancias.amostras)
abline(v=variancia.teorica,col="red", lwd=3, lty=2)
abline(v=variancia.amostral,col="blue", lwd=3, lty=2)
expr.var <- vector("expression", 2)
expr.var[[1]] <- bquote("Variância Teórica"==.(variancia.teorica))
expr.var[[2]] <- bquote("Variância da Simulação"==.(variancia.amostral))
legend("topright",
       bty = "n",
       legend = expr.var,cex=0.7)      # bty = box type, pode ser "o" (default) e "n"

# Comparando a variância teórica com a variância amostral
estats.var <- matrix(c(variancia.teorica,variancia.amostral),ncol=1)
rownames(estats.var) <- c("Variância Teórica:","Variância da Simulação:")
colnames(estats.var) <- c("Valores")
estats.var
```

Estimação de Parâmetros

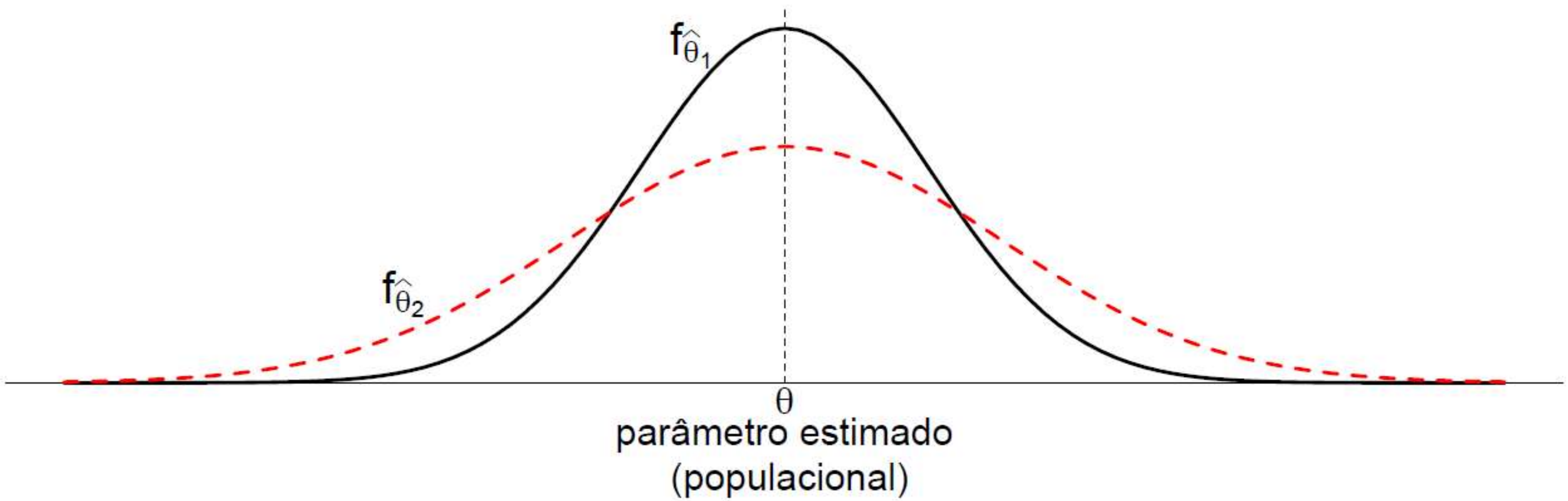
Propriedades Desejáveis de Estimadores

Eficiência:

– Dados dois estimadores não-tendenciosos, $\hat{\Theta}_1$ e $\hat{\Theta}_2$, com variâncias amostrais dadas por S_1^2 e S_2^2:

$$S_1^2 < S_2^2 \quad \Longrightarrow \quad \hat{\Theta}_1 \text{ é mais eficiente que } \hat{\Theta}_2$$

Distribuições Amostrais de Estimadores



Estimação de Parâmetros

Propriedades Desejáveis de Estimadores

Agora, vamos estimar:
$p_o =$ Probabilidade de $X = 0$.

Consideremos os estimadores:

$$S = \frac{\text{número de observações em que } X_i = 0}{n} \qquad \text{e} \qquad T = e^{-\bar{X}_n}$$

O que podemos esperar deles?

Estimação de Parâmetros

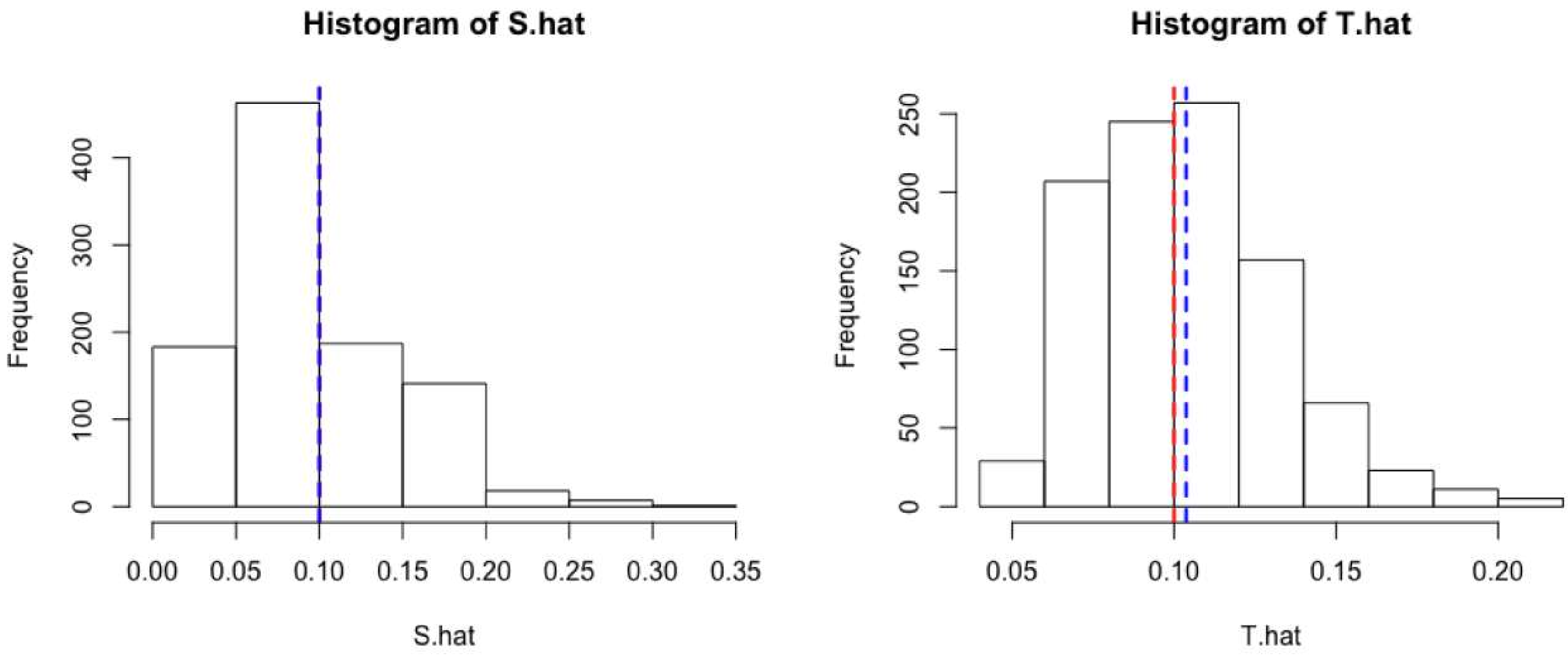
Propriedades Desejáveis de Estimadores

Assim como no caso anterior, vamos simular:

– Fingimos que conhecemos o valor de λ $(= \ln 10)$ tq. $p_o = 0,1$;
– Amostramos $n = 30$ observações a partir de Pois $(\lambda = \ln 10)$,
– Calculamos os valores que S e T assumem para as observações da a.a.;
– Repetimos esse procedimento um grande número de vezes, $nS = 1000$;
– Construímos a distribuição amostral para cada estimador:

Estimação de Parâmetros

Propriedades Desejáveis de Estimadores



Os valores estimados para p_o parecem flutuar em torno do valor "real" (0,1). Será verdade?

```r
dev.off(dev.list()["RStudioGD"])
rm(list=ls())
cat("\f")


# Definindo os parâmetros da simulação e obtendo as amostras

lambda = log(10)
p.zero = 0.1
n = 30
linhas = 1000

par(mfrow=c(1,2))

sim <- rpois(n*linhas, lambda)
sim
m <- matrix(sim, linhas)
m


# Conta a quantidade de zeros em cada linha
zeros.na.linha <- function(x, ...) {
  rowSums(m=='0', ...)
                                    }

z <- zeros.na.linha(m)
#z

# Calcula e plota a estatistica S
S <- z/n
#S

prob.teorica = p.zero
prob.amostral = mean(S)

hist(S)
abline(v=prob.teorica,col="red", lwd=3, lty=2)
abline(v=prob.amostral,col="blue", lwd=3, lty=2)
expr.media <- vector("expression", 2)
expr.media[[1]] <- bquote("Prob. Teórica"==.(prob.teorica))
expr.media[[2]] <- bquote("Prob. da Simulação"==.(prob.amostral))
legend("topright",
       bty = "n",
       legend = expr.media,cex=0.7)      # bty = box type, pode ser "o" (default) e "n"
```

```r
# Calcula e plota a estatística T
medias.amostras <- rowMeans(m)

T <- exp(-medias.amostras)
#T

prob.teorica = p.zero
prob.amostral = mean(T)

hist(T)
abline(v=prob.teorica,col="red", lwd=3, lty=2)
abline(v=prob.amostral,col="blue", lwd=3, lty=2)
expr.media <- vector("expression", 2)
expr.media[[1]] <- bquote("Prob. Teórica"==.(prob.teorica))
expr.media[[2]] <- bquote("Prob. da Simulação"==.(prob.amostral))
legend("topright",
       bty = "n",
       legend = expr.media,cex=0.7)      # bty = box type, pode ser "o" (default) e "n"
```

Estimação de Parâmetros

Propriedades Desejáveis de Estimadores

Definindo:

Y = número de observações em que $X_i = 0$ (sucesso), então

$$Y \sim Binom(n, p_o).$$

É possível demonstrar que :

$$S = \frac{Y}{n} \qquad \Rightarrow \qquad E[S] = p_o$$

Portanto, o estimador S é não-viesado.

Estimação de Parâmetros

Propriedades Desejáveis de Estimadores

Agora, o estimador $T = e^{-\bar{X}_n}$:

Podemos ter a impressão de que T é também um estimador não-viesado.
Isto não é verdade!

É possível mostrar que :

$$E[T] = E[e^{-\bar{X}_n}] = e^{-\ \lambda(1-e^{-1/n})} \quad \underset{n\to\infty}{\longrightarrow} e^{-\lambda} = p_o$$

Embora o estimador seja positivamente viesado (tende a superestimar o parâmetro real), o viés tende a zero à medida em que o tamanho da amostra aumenta.

Estimação de Parâmetros

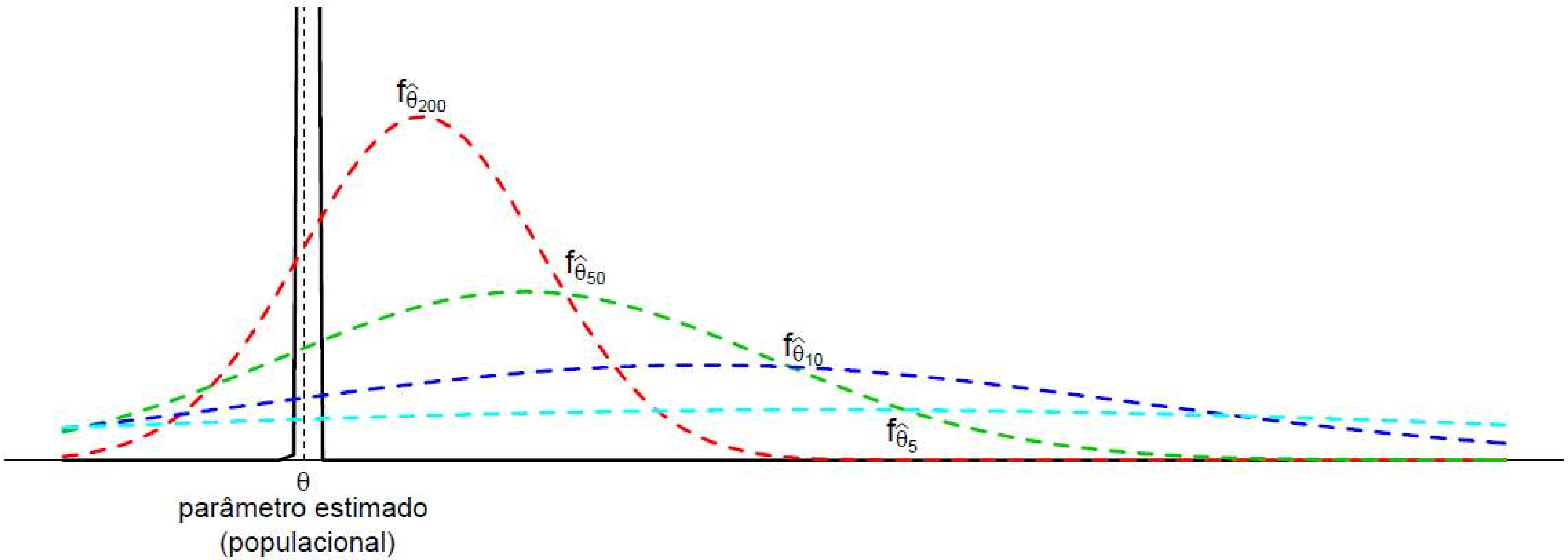
Propriedades Desejáveis de Estimadores

Consistência:

– À medida que o tamanho da amostra n aumenta, o valor esperado do estimador consistente tende para o parâmetro populacional e a variância do estimador cai a zero:

$$\lim_{n\to\infty} P[|\hat{\Theta}_n - \theta| < \epsilon] = 1, \quad \{\hat{\Theta}_n, n \geq 0\}$$

Distribuições Amostrais de Estimadores



Estimação de Parâmetros

Propriedades Desejáveis de Estimadores

Exemplo:

Seja $X_1, X_2, \ldots, X_n$ uma a.a. de uma distribuição com média μ e variância σ^2.
Responda:

1. A estatística $\overline{X}_n$ é um estimador não-tendencioso de μ?
2. A estatística S_n^2 é um estimador não-viesado de σ^2?
3. A estatística S_n é um estimador não-viciado de σ?

OBS:
Se $\hat{\Theta}$ é um estimador não-viesado de θ, $g(\hat{\Theta})$ não-necessariamente é um estimador não-viesado de $g(\theta)$.

Para o caso particular em que $g(\hat{\Theta}) = a\hat{\Theta} + b$, a não-tendenciosidade é transmitida.

Estimação Pontual

Propriedades Desejáveis de Estimadores

Exemplo – Solução (1):

$X_1, X_2, \ldots, X_n \overset{ind.}{\sim} f_X : E[X_i] = \mu$ e $\text{Var}[X_i] = \sigma^2$, $i = 1, 2, \ldots, n.$

1. A estatística $\bar{X}_n$ é um estimador não-tendencioso de μ?

$$E[\bar{X}_n] = E\left[\frac{1}{n}\sum X_i\right] = \frac{1}{n}\sum E[X_i] = \frac{1}{n}\sum \mu = \frac{1}{n}(n \cdot \mu) = \mu$$

Portanto, $\bar{X}_n$ é um estimador não tendencioso de μ.

Estimação Pontual

Propriedades Desejáveis de Estimadores

Exemplo – Solução (2):

$X_1, X_2, \ldots, X_n \overset{ind.}{\sim} f_X : E[X_i] = \mu$ e $\text{Var}[X_i] = \sigma^2$, $i = 1, 2, \ldots, n$.

2. A estatística S_n^2 é um estimador não-viesado de σ^2?

Para qualquer v.a. Y, temos: $\text{Var}[Y] = E[Y^2] - (E[Y])^2 \Rightarrow E[Y^2] = \text{Var}[Y] - (E[Y])^2 = \sigma^2 + \mu^2$

$$E[S_n^2] = E\left[\frac{1}{n-1}\sum(X_i - \bar{X})^2\right] = E\left[\frac{1}{n-1}\left(\sum X_i^2 - \frac{(\sum X_i)^2}{n}\right)\right] = \frac{1}{n-1}\left[\sum E[X_i^2] - \frac{E[(\sum X_i)^2]}{n}\right]$$

$$= \frac{1}{n-1}\left[\sum(\sigma^2 + \mu^2) - \frac{\text{Var}[\sum X_i] + (E[\sum X_i])^2}{n}\right] = \frac{1}{n-1}\left[n(\sigma^2 + \mu^2) - \frac{1}{n}(n\sigma^2 + (n\mu)^2)\right]$$

$$= \frac{1}{n-1}\left[n\sigma^2 + n\mu^2 - \sigma^2 - n\mu^2\right] = \frac{1}{n-1}\left[(n-1)\sigma^2\right] = \sigma^2$$

Portanto, S_n^2 é um estimador não tendencioso de σ^2.

Já S_n é um estimador tendencioso de σ: $E[S_n] \neq \sigma$ (viés $\downarrow$ quando $n \uparrow$).

Estimação de Parâmetros

Propriedades Desejáveis de Estimadores

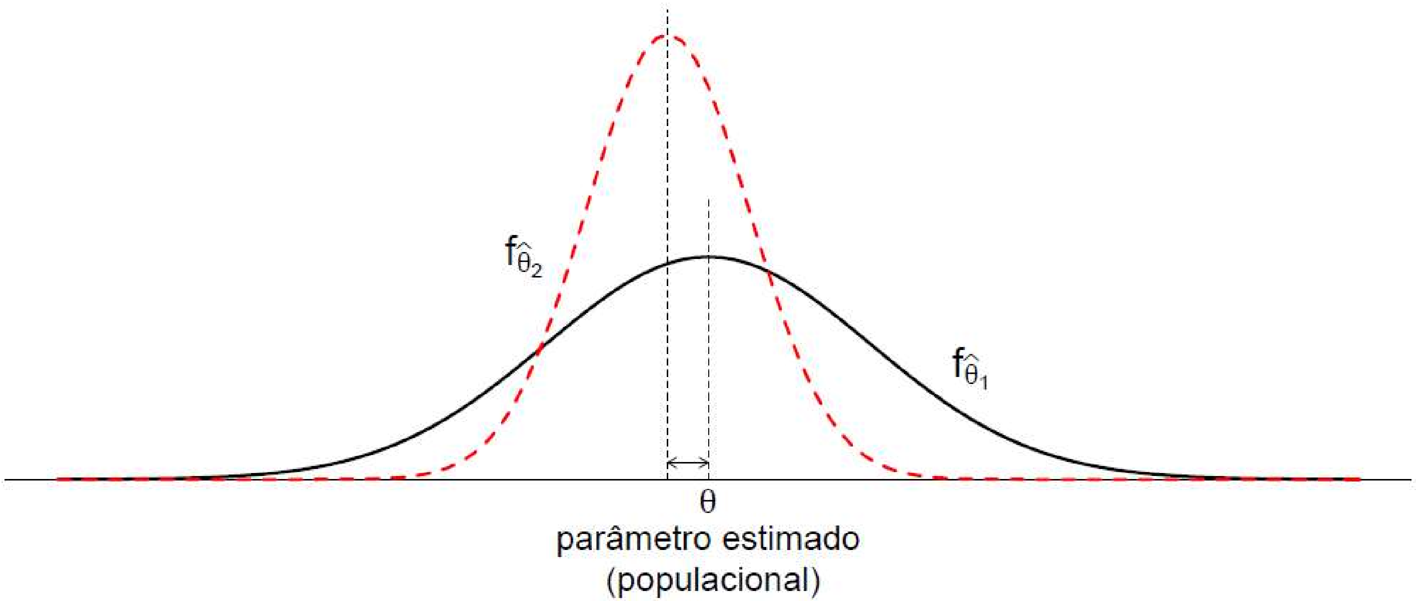
Erro Médio Quadrático

Nem sempre é possível obter um estimador que seja, ao mesmo tempo,

não-tendencioso e eficiente.

Neste caso, é preferível escolher um estimador que minimize o erro médio quadrático (MSE), que corresponde ao valor médio dos desvios quadráticos entre $\hat{\Theta}$ e θ.

$$MSE[\hat{\Theta}] = E[(\hat{\Theta} - \theta)^2] = Var[\hat{\Theta}] + \left(b[\hat{\Theta}]\right)^2$$



Estimação Pontual

Propriedades Desejáveis de Estimadores

Exemplo:

Sejam Θ_1 e Θ_2 estimadores independentes não-tendenciosos de θ, com variâncias conhecidas dadas por σ_1^2 e σ_2^2.

1. Mostre que qualquer estimador na forma

$$\Theta^\star = \lambda\Theta_1 + (1-\lambda)\Theta_2$$

   é também não tendencioso.

2. Qual o valor de λ que confere a $\Theta^\star$ a menor variância?

$$\theta^* = \lambda\theta_1 + (1-\lambda)\theta_2$$

$$E\left[\theta^*\right] = E\left[\lambda\theta_1 + (1-\lambda)\theta_2\right]$$

$$E\left[\theta^*\right] = E\left[\lambda\theta_1\right] + E\left[(1-\lambda)\theta_2\right]$$

$$E\left[\theta^*\right] = \lambda E\left[\theta_1\right] + (1-\lambda)E\left[\theta_2\right]$$

$$E\left[\theta^*\right] = \lambda\theta_1 + (1-\lambda)\theta_2$$

$$E\left[\theta^*\right] = \lambda\theta + \theta - \lambda\theta$$

$$E\left[\theta^*\right] = \theta$$

$$\theta^* = \lambda\theta_1 + \left(1-\lambda\right)\theta_2$$

$$Var\left[\theta^*\right] = Var\left[\lambda\theta_1 + \left(1-\lambda\right)\theta_2\right]$$

$$Var\left[\theta^*\right] = Var\left[\lambda\theta_1\right] + Var\left[\left(1-\lambda\right)\theta_2\right]$$

$$Var\left[\theta^*\right] = \lambda^2 Var\left[\theta_1\right] + \left(1-\lambda\right)^2 Var\left[\theta_2\right]$$

$$Var\left[\theta^*\right] = \lambda^2\sigma_1^2 + \left(1-\lambda\right)^2\sigma_2^2$$

$$Var\left[\theta^*\right] = \lambda^2\sigma_1^2 - \lambda^2\sigma_2^2 + \sigma_2^2$$

$$Var\left[\theta^*\right] = \lambda^2\left(\sigma_1^2 - \sigma_2^2\right) + \sigma_2^2$$

$$\lambda^2\left(\sigma_1^2 - \sigma_2^2\right) + \sigma_2^2 = 0$$

$$\lambda^2 = \frac{\sigma_2^2}{\sigma_2^2 - \sigma_1^2} \Rightarrow \lambda = \sqrt{\frac{\sigma_2^2}{\sigma_2^2 - \sigma_1^2}}$$

Métodos de Estimação de Parâmetros

Métodos de Estimação de Parâmetros

1. Método dos Momentos (Karl Pearson, 1894)
2. Método da Máxima Verossimilhança (R. A. Fisher, 1922)
3. Método dos Mínimos Quadrados (C. F. Gauss, 1810)

Método dos Momentos

A ideia por trás do método dos momentos é igualar os momentos da população sobre a origem aos seus correspondentes momentos de amostra, em que o r-ésimo momento da amostra sobre a origem, denotado m_r, é definido como

$$m_r = \frac{1}{n}\sum_{i=1}^{n} X_i^r$$

e, posteriormente, resolve-se para os estimadores dos parâmetros desconhecidos. O momento da população sobre a origem de uma variável aleatória X, denominada α_r, é definido como

$$E\left[X^r\right] = \alpha_r .$$

Segue que

$$\alpha_r = E\left[X^r\right] = \sum_{i=1}^{\infty} x_i^r P\left(X = x_i\right), \quad \text{para } X \text{ discreta e}$$

$$\alpha_r = E\left[X^r\right] = \int_{-\infty}^{\infty} x^r f\left(x\right) dx, \quad \text{para } X \text{ contínua.}$$

Especificamente, dada uma amostra aleatória X_1, X_2, ..., X_n de uma população com fdp $f\,(x|\ \theta_1, \theta_2, \ldots, \theta_n)$, os estimadores do Método do Momentos, denotados $\tilde{\theta}_i$ para i = 1, 2, ..., k, são encontrados igualando os k primeiros momentos da população sobre a origem aos momentos de amostra correspondentes e resolvendo o sistema resultante de equações simultâneas:

$$\begin{cases} \alpha_1\left(\theta_1, \theta_2, \ldots, \theta_k\right) = m_1 \\ \alpha_2\left(\theta_1, \theta_2, \ldots, \theta_k\right) = m_2 \\ \qquad\qquad \vdots \qquad\qquad\quad \vdots \\ \alpha_k\left(\theta_1, \theta_2, \ldots, \theta_k\right) = m_k \end{cases}$$

O método dos momentos é uma técnica interessante para derivar estimadores devido à sua simplicidade e ao fato de que os estimadores de método de momentos são consistentes.
De fato, a justificativa teórica para equacionar os momentos da amostra aos momentos da população é que, sob certas condições, pode-se mostrar que os momentos da amostra convergem em probabilidade para os momentos da população e que os momentos da amostra em torno da origem são estimadores momentos correspondentes da população.

Exemplo

Dada uma amostra aleatória de tamanho n de uma população $N(\mu, \sigma^2)$, encontre pelo método dos momentos estimadores de μ e σ^2.

O primeiro e o segundo momentos amostrais m_1 e m_2 são $\bar{X}$ e $\frac{1}{n}\sum_{i=1}^{n} X_i^2$, respectivamente.

O primeiro e segundo momentos da população, em torno de zero para uma variável aleatória normal, são $\alpha_1 = E\left[X^1\right] = \mu$ e $\alpha_2 = E\left[X^2\right] = \sigma^2 + \mu^2$.

Ao equacionar os dois primeiros momentos da população para os dois primeiros momentos da amostra,

$$\begin{cases} \alpha_1\left(\mu,\sigma^2\right) = \mu = \bar{X} = m_1 \\ \alpha_2\left(\mu,\sigma^2\right) = \sigma^2 + \mu^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 = m_2. \end{cases}$$

Resolvendo o sistema de equações, produz

$$\tilde{\mu} = \overline{X}$$

$$\tilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = S^2.$$

como estimadores pelo método dos momentos para μ e σ^2, respectivamente.

■

Métodos de Estimação de Parâmetros

Método da Máxima Verossimilhança (R. A. Fisher, 1922)

A função de verossimilhança para uma a.a. é dada por:

$$L(\theta) = f(X_1, X_2, \ldots, X_n; \theta)$$

- O método da máxima verossimilhança consiste em maximizar $L(\theta)$ com relação ao parâmetro θ.
- O estimador de máxima verossimilhança maximiza a probabilidade de obter os dados observados.

Método da Máxima Verossimilhança

Quando a amostragem de uma população descrita por uma fdp $f(x \mid \theta)$, o conhecimento de θ fornece conhecimento de toda a população. A ideia por trás da máxima verossimilhança é selecionar o valor para θ que torna os dados observados mais prováveis sob o modelo de probabilidade assumido.

Quando x_1, x_2, ..., x_n são os valores observados de uma variável aleatória X de uma população com o parâmetro θ, a notação $L(\theta|\mathbf{x}) = f(\mathbf{x}|\theta)$ será usada para indicar que a distribuição depende do parâmetro θ, e $\mathbf{x}$ para indicar que a distribuição depende dos valores observados da amostra.

$L(\theta|\mathbf{x})$ é a função de verossimilhança de θ para $\mathbf{x}$ e é denotada por

$$L(\theta|\mathbf{x}) = f(\mathbf{x}|\theta) = \prod_{i=1}^{n} f(x_i|\theta) = f(x_1|\theta) \times f(x_2|\theta) \times \ldots \times f(x_n|\theta).$$

O valor de θ que maximiza L(θ|x) é chamado de estimativa da máxima verossimilhança (EMV) de θ.
A estimativa de máxima verossimilhança é designada como $\hat{\theta}(\mathbf{x})$, e o estimador de máxima verossimilhança (MLE – Maximum Likelihood Estimator), uma estatística, como $\hat{\theta}(\mathbf{X})$.

Em geral, a função de verossimilhança pode ser difícil de manipular, e geralmente é mais conveniente trabalhar com o logaritmo neperiano de $L(\theta|\mathbf{x})$, chamado de função log-verossimilhança, uma vez que converte produtos em somas.

Encontrar o valor θ que maximiza a função de log-verossimilhança [lnL $(\theta|\mathbf{x})$] é equivalente a encontrar o valor de θ que maximiza $L(\theta|\mathbf{x})$, uma vez que o logaritmo natural é uma função monotonicamente crescente.

Se L(θ|**x**) é diferenciável em relação a θ, um EMV possível é a solução para

$$\frac{\partial\left[\ln L\left(\theta|\mathbf{x}\right)\right]}{\partial\theta}=0.$$

Lembre-se que, pontos estacionários onde

$$\left.\frac{\partial^2\left[\ln L\left(\theta|\mathbf{x}\right)\right]}{\partial\theta^2}\right|_{\theta=\hat{\theta}(\mathbf{x})}<0,$$

indicam algum tipo de máximo, local ou global.

Além disso, a solução para a igualdade não inclui pontos nos limites do espaço de parâmetros. Consequentemente, ao avaliar o máximo de L(θ|**x**), os limites do espaço de parâmetros Θ, assim como as soluções para a igualdade da derivada de L(θ|**x**) a zero, devem ser avaliados.

Exemplo

Seja X_1, X_2, ... , X_m uma amostra aleatória de uma população Poisson(λ). Calcule o estimador de máxima verossimilhança e a estimativa da máxima verossimilhança para o parâmetro λ. Verifique sua resposta com uma simulação gerando 1000 valores aleatórios de uma população Pois($\lambda = 5$).

A função de verossimilhança é

$$L(\lambda|\mathbf{x}) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i\,!} = e^{-n\lambda}\prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i\,!},$$

e a função log-verossimilhança é

$$\ln L(\lambda|\mathbf{x}) = \ln\left[e^{-n\lambda}\prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i\,!}\right] = -n\lambda + \sum_{i=1}^{n} x_i \ln\lambda - \sum_{i=1}^{n} \ln(x_i\,!).$$

Em seguida, procure o valor que maximiza a função de log-verossimilhança, tomando a derivada parcial de primeira ordem de $\ln L$ $(\lambda|\mathbf{x})$ e igualando a resposta a zero:

$$\frac{\partial \ln L(\lambda|\mathbf{x})}{\partial\lambda} = -n + \frac{\sum_{i=1}^{n} x_i}{\lambda} = 0.$$

A solução para a equação acima é

$$\lambda = \frac{\sum_{i=1}^{n} x_i}{n} = \overline{x}.$$

Para $\lambda = \bar{x}$ ser um máximo, a derivada parcial de segunda ordem da função log-verossimilhança deve ser negativa em $\lambda = \bar{x}$. A derivada parcial de segunda ordem é

$$\frac{\partial^2 \ln L(\lambda|\mathbf{x})}{\partial \lambda^2} = -\frac{\sum_{i=1}^{n} x_i}{\lambda^2}.$$

Avaliando a derivada de segunda ordem em $\lambda = \bar{x}$ produz

$$\frac{\partial^2 \ln L(\lambda|\mathbf{x})}{\partial \lambda^2} = -\frac{n\overline{x}}{\overline{x}^2} = -\frac{n}{\overline{x}} < 0.$$

Finalmente, como os valores da função de verossimilhança nos limites do espaço de parâmetros, λ = 0 e λ = ∞, são 0, segue-se que $\lambda = \bar{x}$ é o valor que maximiza a função de verossimilhança.
A estimativa da máxima verossimilhança é $\hat{\lambda}(\boldsymbol{x}) = \bar{x}$ e o estimador de máxima verossimilhança $\hat{\lambda}(\boldsymbol{X}) = \bar{X}$.

Para simular $\lambda = \bar{x}$, gere 1000 valores aleatórios de uma população Pois(λ = 5):

```
rm(list=ls())
cat("\f")

#set.seed(1000)
mean(rpois(1000, 5))
```

```
> #set.seed(1000)
> mean(rpois(1000, 5))
[1] 5.006
```

$\blacksquare$

Métodos de Estimação de Parâmetros

Método da Máxima Verossimilhança (R. A. Fisher, 1922)

Exemplo: Distribuição Normal

Sejam $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$. A função de verossimilhança para esta a.a. é:

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} N(X_i; \mu, \sigma^2) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2\right]$$

Tomando as derivadas parciais do logaritmo de L ($l = \ln L$) com respeito a μ e σ^2:

$$\frac{\partial l}{\partial \mu} = \frac{1}{\sigma^2}\sum(X_i - \mu) = 0$$

$$\frac{\partial l}{\partial \sigma^2} = \frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum(X_i - \mu)^2 = 0$$

Resolvendo para μ e σ^2, obtemos:

$$\hat{\mu} = \bar{X} = \frac{1}{n}\sum X_i \quad \text{e} \quad \hat{\sigma}^2 = S'^2 = \frac{1}{n}\sum(X_i - \bar{X})^2$$

... iguais aos estimadores obtidos pelo métodos dos momentos (nem sempre é o caso).

Métodos de Estimação de Parâmetros

OBS:

- Os estimadores de máxima verossimilhança em geral coincidem com os estimadores obtidos pelo método dos momentos.
- O conjunto de equações simultâneas do método da máxima verossimilhança em geral é mais difícil de resolver que aquele obtido para o método dos momentos.
- Estimadores de máxima verossimilhança costumam ter melhores propriedades estatísticas que aqueles obtidos pelo método dos momentos.
  – Princípio da invariância
  – Não-tendenciosidade assimptótica
  – Assimptoticamente BUE ("best unbiased estimators")

Métodos de Estimação de Parâmetros

Método dos Mínimos Quadrados (C. F. Gauss, 1810)

OBS:

- Método muito utilizado em análise multivariada:
  – análise de regressão, ANOVA, RSM etc.
- Raramente utilizado para estimar parâmetros de distribuições.