Aplicações à Inferência Estatística

Celso Cabral

08/11/2019

O Vício de um Estimador

Seja $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} f(\cdot|\theta)$, uma amostra aleatória, onde $f(\cdot|\theta)$ é uma densidade ou função de probabilidade indexada pelo parâmetro $\theta \in \Theta$. Defina $\mathbf{X} = (X_1, \ldots, X_n)$ e seja $T(\mathbf{X})$ um estimador para θ.

Definição

O vício do estimador $T(\mathbf{X})$ é definido por

$$\mathrm{bias}_\theta(T) = \mathrm{E}_\theta(T(\mathbf{X})) - \theta, \;\; \theta \in \Theta.$$

Observe que:

- $\mathrm{E}_\theta(T(\mathbf{X}))$ é a esperança calculada supondo que θ é o verdadeiro valor do parâmetro. Ou seja,

$$\mathrm{E}_\theta(T(\mathbf{X})) = \int T(\mathbf{x}) f(\mathbf{x}|\theta) d\mathbf{x};$$

O Vício de um Estimador

- $\mathrm{bias}_\theta(T)$ depende de θ, ou seja, para cada $\theta \in \Theta$ temos um valor para o vício. Assim, na realidade, o vício é uma função de $\theta \in \Theta$. Desta forma, é importante que tenhamos uma ideia do comportamento desta função.
- Além disso, o cálculo do vício depende do cálculo da integral $\mathrm{E}_\theta(T(\mathbf{X}))$. Já vimos que em algumas situações estas integrais não podem ser obtidas de forma analítica.
- Neste caso, podemos aproximar $\mathrm{E}_\theta(T(\mathbf{X}))$ usando Monte Carlo. Basta gerar N observações do estimador T e depois aproximar esta integral pela média aritmética. Assim, a proposta é usar o seguinte algoritmo para aproximar o vício do estimador em θ:

O Vício de um Estimador

Algoritmo 1

1. **For** $j = 1, \ldots, N$, gere uma amostra $X_1^{(j)}, \ldots, X_n^{(j)} \sim f(\cdot|\theta)$
2. Calcule $T^{(j)} = T(X_1^{(j)}, \ldots, X_n^{(j)})$
3. **end for**
4. Retorne $\frac{1}{N} \sum_{j=1}^{N} T^{(j)} - \theta$

O Vício de um Estimador - Exemplo 3.34 de Voss (2014)

Sejam X e Z variáveis aleatórias independentes com distribuição normal padrão e seja $\rho \in [-1, 1]$. Defina

$$Y = \rho X + Z\sqrt{1 - \rho^2}.$$

É possível mostrar que $\mathrm{Cov}(X, Y) = \rho$ e que $\mathrm{Var}(Y) = 1$ (mostre isso). Assim,

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = \rho.$$

O Vício de um Estimador - Exemplo 3.34 de Voss (2014)

Para amostras $\mathbf{X} = (X_1, \ldots, X_n)$ de X e $\mathbf{Y} = (Y_1, \ldots, Y_n)$ de Y, um estimador para ρ é a chamada *correlação amostral*, definida por

$$\widehat{\rho}(\mathbf{X}, \mathbf{Y}) = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2 \sum_{i=1}^{n}(Y_i - \overline{Y})^2}},$$

onde $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ e $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$.

O Vício de um Estimador - Exemplo 3.34 de Voss (2014)

- Vamos aproximar o valor do vício do estimador $\widehat{\rho}(\mathbf{X}, \mathbf{Y})$ para um determinado ρ fixado. Para isso vamos usar o algoritmo 1.
- Observações de X e Y são obtidas primeiramente gerando X e Z independentemente com distribuição $\mathrm{N}(0, 1)$ e em seguida calculando $Y = \rho X + Z\sqrt{1 - \rho^2}$.

O Vício de um Estimador - Exemplo 3.34 de Voss (2014)

```
# Cálculo do vício do coeficiente
# de correlação amostral
set.seed(1)
bias.cor <- function(n,N,rho){
# n: tamanho das amostras de x e y
# N: número de amostras geradas
# rho: verdadeiro valor do
#      coeficiente de correlação
#      de Pearson
rho.hat <- vector()
for (j in 1:N){
x <- rnorm(n)
z <- rnorm(n)
y <- rho*x+sqrt(1-rho^2)*z
rho.hat[j] <- cor(x,y) #Coeficiente
                       #de correlação
                       #para a amostra j
}
return(list(bias=mean(rho.hat)-rho,rho.hat=rho.hat))
}
bias <- bias.cor(20,5000,0.7)
bias$bias #Vício
```

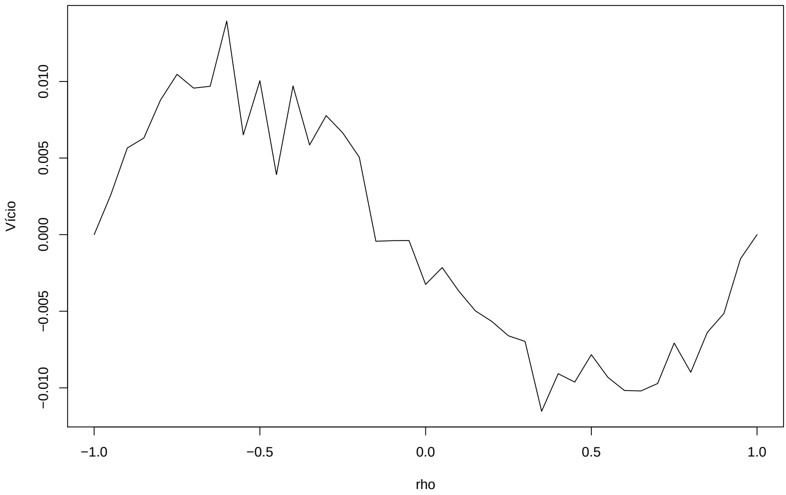
```
## [1] -0.009645586
```

O Vício de um Estimador - Exemplo 3.34 de Voss (2014)

Podemos executar este código para diversos valores de ρ a fim de analisar o vício como função de ρ.

Vício de um Estimador - Exemplo 3.34 de Voss (2014)

```
rho <- seq(-1,1,0.05)
plot(rho,sapply(rho,bias.cor,n=20,N=5000)[1,],
     xlab="rho",ylab="Vício",type="l")
```



O desvio padrão de um Estimador - Exemplo 3.34 de Voss (2014)

Observe que, para um dado valor de θ, podemos aproximar o valor do desvio padrão do estimador calculando

$$\widehat{\mathrm{se}}_\theta(T) = \sqrt{\frac{1}{N-1}\sum_{j=1}^{N}(T^{(j)} - \overline{T})^2},$$

onde $\overline{T} = \frac{1}{N}\sum_{j=1}^{N} T^{(j)}$. No caso do estimador correlação amostral visto acima, temos

Intervalos de Confiança

Sejam $\alpha \in (0,1)$ e $\mathbf{X} = (X_1, \ldots, X_n)$ uma amostra de uma distribuição com parâmetro $\theta \in \Theta$.

Definição

Um intervalo de confiança para θ com coeficiente $1-\alpha$ é um par de estatísticas $(T(\mathbf{X}), U(\mathbf{X}))$ tal que

$$P_\theta(T(\mathbf{X}) \leq \theta \leq U(\mathbf{X})) \geq 1-\alpha \quad \textit{para todo } \theta \in \Theta.$$

Em determinadas situações o cálculo de $P_\theta(T(\mathbf{X}) \leq \theta \leq U(\mathbf{X}))$ pode não ser muito simples. Mas já vimos que podemos aproximar probabilidades usando Monte Carlo. Para isso, precisamos escrever a probabilidade como uma esperança, ou seja, uma integral.

Intervalos de Confiança

- Vamos relembrar como isso é feito. Sejam T uma v.a. e A um evento. Considere a função indicadora de $(T \in A)$, ou seja, $\mathbb{I}_{(T \in A)} = 1$ se $T \in A$ e $\mathbb{I}_{(T \in A)} = 0$ se $T \notin A$. Então,

$$P(T \in A) = \mathrm{E}(\mathbb{I}_{(T \in A)}).$$

- Assim, para uma amostra $(T_1, \ldots, T_n)$ de T, uma aproximação de Monte Carlo é dada por

$$P\widehat{(T \in A)} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{(T_i \in A)}.$$

  Ou seja, basta calcularmos a frequência relativa em que ocorre $T_i \in A$, $i = 1, \ldots, n$.

Intervalos de Confiança

- De volta ao problema de aproximar $P_\theta(T(\mathbf{X}) \leq \theta \leq U(\mathbf{X}))$, vemos que basta aproximar esta probabilidade por

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{(T^{(i)} \leq \theta \leq U^{(i)})}.$$

onde $(T^{(i)}, U^{(i)})$, $i = 1, \ldots, n$ são intervalos simulados supondo que θ é o verdadeiro valor do parâmetro.

Intervalos de Confiança - Exemplo 3.38 de Voss (2014)

- Sejam $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Poisson}(\theta)$. Usando a aproximação normal para a distribuição Poisson, temos que um intervalo de confiança para θ é dado por