Aula 15: Método Monte Carlo em Inferência

Prof. Dr. Eder Angelo Milani e Érika Soares Machado

26/06/2023

Método Monte Carlo em Inferência

- Os métodos de Monte Carlo abrangem um vasto conjunto de ferramentas computacionais na estatística.
- Podem se referir a qualquer método em inferência estatística ou análise numérica onde a simulação é usada.
- Podem ser aplicados para estimar parâmetros da distribuição amostral de uma estatística, erro quadrático médio (EQM), percentis ou outras quantidades de interesse.
- Podem ser projetados para avaliar a probabilidade de cobertura para intervalos de confiança, para estimar o poder do teste e para comparar o desempenho de diferentes procedimentos para um determinado problema.

Método de Monte Carlo para Estimação

Suponha $X_1, \ldots, X_n$ v.a's com distribuição X. Um estimador $\hat{\theta}$ para o parâmetro θ é uma função n variada da amostra, podendo ser escrita como

$$\hat{\theta} = \hat{\theta}(x_1, ..., x_n).$$

Por simplicidade, considere $x = (x_1, ..., x_n)^T \in \mathbb{R}^n$ uma amostra aleatória e seja $x^{(1)}, x^{(2)}, ...$ a sequência de amostras aleatórias independentes geradas da distribuição de X. A distribuição amostral de $\hat{\theta}$ pode ser gerada por repetidas retiradas independentes de amostras aleatórias $x^{(j)}$ e então calcular

$$\hat{\theta}^{(j)} = \hat{\theta}(x_1^{(j)}, ..., x_n^{(j)})$$

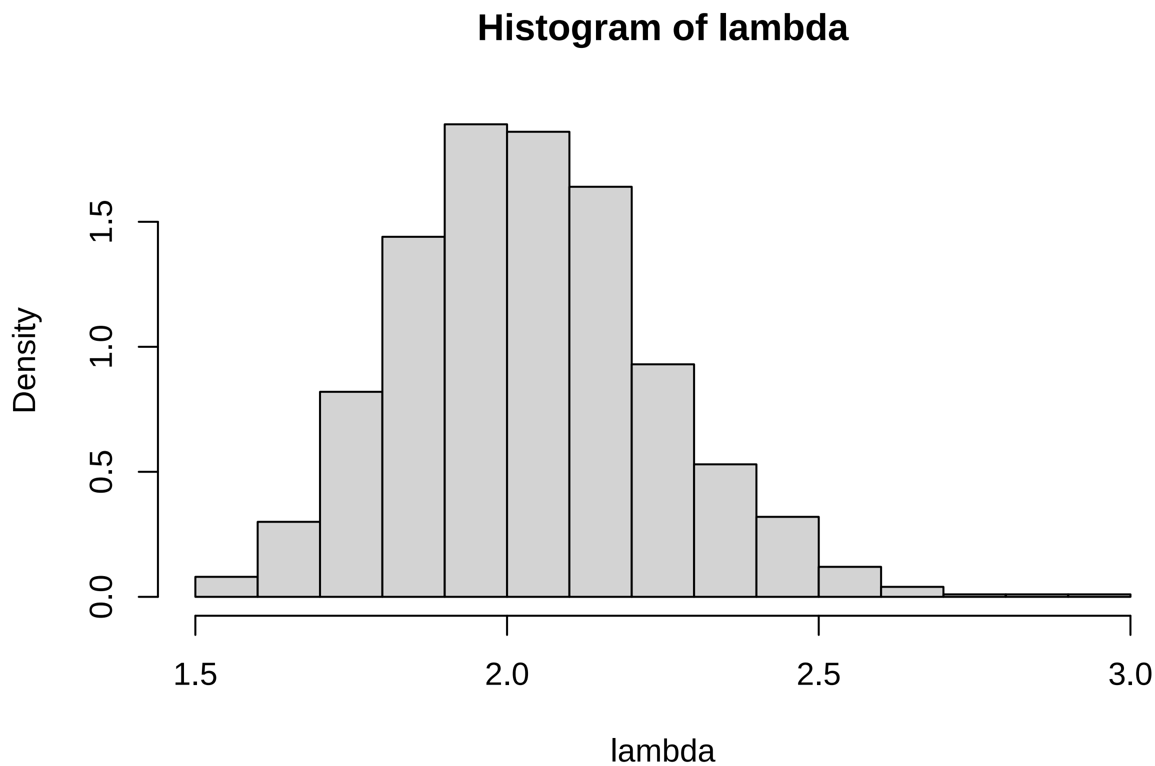
para cada amostra.

Exemplo 1

Considerando uma amostra de tamanho 100 da distribuição Exponencial com parâmetro $\lambda = 2$, obtenha um esboço da distribuição amostral do parâmetro λ.

```
set.seed(2023)
n <- 100
lambda <- numeric()
repeticao <- 1000
for(i in 1:repeticao){
  lambda[i] <- 1/mean(rexp(n, 2))
}

hist(lambda, freq = F)
```



Exemplo 2

Suponha que X_1 e X_2 são v.a.i.i.d. da distribuição normal padrão. Estime a diferença absoluta média, ou seja, $E(|X_1 - X_2|)$.

Solução:

- Por integração:

$$E(|X_1 - X_2|) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |x_1 - x_2| \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{x_1^2}{2}\Big) \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{x_2^2}{2}\Big) dx_1 dx_2,$$

após alguns cálculos, obtemos que $E(|X_1 - X_2|) = \frac{2}{\sqrt{(\pi)}} = 1,128379$

- Usando o método de Monte Carlo

Para obter um estimador de Monte Carlo de

$$\theta = E(g(X_1, X_2)) = E(|X_1 - X_2|)$$

baseado em m repetições, geramos amostras aleatórias $x^{(j)} = (x_1^{(j)}, x_2^{(j)})$ de tamanho 2 da distribuição normal padrão, $j = 1, ..., m$. Então, calcule as réplicas

$$\hat{\theta}^{(j)} = g_j(x_1, x_2) = |x_1^{(j)} - x_2^{(j)}|, j = 1, ..., m$$

e a média das réplicas

$$\hat{\theta} = \frac{1}{m}\sum_{i=1}^{m}\hat{\theta}^{(j)} = \frac{1}{m}\sum_{i=1}^{m}|x_1^{(j)} - x_2^{(j)}|$$

```
set.seed(2023)
m <- 10000
g <- numeric(m)
for(i in 1:m){
  x <- rnorm(2)
  g[i] <- abs(x[1]-x[2])
}
theta_chapeu <- mean(g)
cat("O valor esperado é ", theta_chapeu,"\n")
```

```
## O valor esperado é  1.120249
```

```
### Outra forma de calcular

set.seed(2023)
m <- 10000
x1 <- rnorm(m)
x2 <- rnorm(m)
g2 <- abs(x1-x2)
theta_chapeu2 <- mean(g2)
cat("O valor esperado é ", theta_chapeu2,"\n")
```

```
## O valor esperado é  1.125256
```

Exemplo 3 - Estimando o erro padrão da média

O erro padrão da média $\overline{X}$ de uma amostra de tamanho n é $\sqrt{\frac{Var(X)}{n}}$. O estimador da variância de X é

$$s^2 = \hat{Var}(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2.$$

A estimativa do erro padrão de $\overline{x}$ é

$$\hat{s}(\overline{x}) = \sqrt{\frac{Var(X)}{n}} = \frac{1}{\sqrt{n}}\Big[\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2\Big]^{1/2},$$

ou usando um estimador não-viesado para $Var(X)$, temos

$$\hat{s}(\overline{x}) = \frac{1}{\sqrt{n}}\Big[\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2\Big]^{1/2},$$

mas caso o tamanho da amostra seja grande, as duas estimativas são próximas.

Voltando ao Exemplo 2, queremos encontrar o desvio padrão de θ. Logo,

$$\hat{s}(\hat{\theta}) = \frac{1}{m}\Big[\sum_{i=1}^{m}(\theta^{(j)} - \overline{\theta})^2\Big]^{1/2},$$

com m = quantidade de repetições e $\overline{\theta} = \frac{1}{m}\sum \hat{\theta}^{(j)}$.

```
set.seed(2023)
m <- 10000
g <- numeric(m)
for(i in 1:m){
  x <- rnorm(2)
  g[i] <- abs(x[1]-x[2])
}
theta_chapeu <- mean(g)
cat("O valor esperado é ", theta_chapeu,"\n")
```

```
## O valor esperado é  1.120249
```

```
s <- sqrt(sum((g-mean(g))^2))/m
cat("O erro padrão obtido via MC é", s, "\n")
```

```
## O erro padrão obtido via MC é 0.008514285
```

```
cat("O erro padrão exato é", sqrt((2-4/pi)/m), "\n")
```

```
## O erro padrão exato é 0.008525025
```

Exercícios

1) Considerando amostras de tamanho 100 da distribuição Poisson($\lambda = 3$), obtenha um esboço da distribuição amostral do parâmetro λ. Utilize o fato que $\hat{\lambda} = \overline{x}$.

2) Considerando amostras de tamanho 100 da distribuição Berolli($p = 0,3$), obtenha um esboço da distribuição amostral do parâmetro p. Utilize o fato que $\hat{p}$ = proporção de sucesso.

3) Suponha que X tenha distribuição normal padrão. Estime a quantidade $E(\exp(X))$. Compare com o valor exato.

Obs.: Se $Y = \exp(X)$, com $X \sim N(0,1)$, então Y tem distribuição log-normal. Neste caso, $E(Y) = \exp(E(X) + 0.5Var(X))$.