

Aula 17: Método Monte Carlo em Inferência - Intervalo de Confiança

Prof. Dr. Eder Angelo Milani

03/07/2023

Estimando o Nível de Confiança

Se (LI, LS) é uma estimativa do intervalo de confiança para um parâmetro desconhecido θ , então LI e LS são estatísticas com distribuição que depende da distribuição F_X da população amostrada X . Utilizando inferências Clássica, a construção do intervalo de confiança pode ser obtida via quantidade pivotal (intervalo exato) ou utilizando propriedades assintóticas (intervalo aproximado).

Método de Monte Carlo para Estimação do Nível de Confiança

Suponha que $X \sim F_X$ é a v.a. de interesse e que θ é o parâmetro para ser estimado.

1. Para cada réplica, $j = 1, \dots, m$
 - a. Gerar a j -ésima amostra aleatória, $x_1^{(j)}, \dots, x_n^{(j)}$
 - b. Calcular o intervalo de confiança C_j para a j -ésima amostra
 - c. Calcular $y_j = I(\theta \in C_j)$, p/ j -ésima amostra
2. Estimar o nível de confiança utilizando a expressão

$$\bar{y} = \frac{1}{m} \sum_{i=1}^m y_j.$$

Obs.: O estimador de \bar{y} é uma estimação amostral do verdadeiro nível de confiança $1 - \alpha^*$. Assim,

$$Var(\bar{y}) = \frac{(1 - \alpha^*)\alpha^*}{m},$$

pois a variância da distribuição Bernoulli é $p(1 - p)$, logo a variância da soma de m Bernoulli's resulta em $\frac{p(1-p)}{m}$, e uma estimação do erro padrão é

$$\widehat{se}(\bar{y}) = \sqrt{\frac{(1 - \bar{y})\bar{y}}{m}}.$$

Exemplo 1: Intervalo de confiança para a média

Se X_1, \dots, X_n é uma amostra aleatória da distribuição $N(\mu, \sigma^2)$, com σ^2 conhecido, então $\bar{X} \sim N(\mu, \sigma^2/n)$, logo

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$

O intervalo de confiança $100(1 - \alpha)\%$ é dado por

$$\left(\bar{X} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right),$$

sendo $z_{1-\alpha/2}$ o $(1 - \alpha/2)$ - quantil da distribuição $N(0, 1)$.

Como verificar tal resultado usando Monte Carlo?

A ideia para verificar o resultado acima é repetir a geração de amostra e a construção do IC, após muitas repetições, verificar a proporção de vezes que o verdadeiro valor do parâmetro caiu dentro do intervalo de confiança.

Vamos realizar um estudo de Monte Carlo para estimar o nível de confiança do IC definido acima, considerando uma amostra de tamanho 30 e $\alpha = 0,05$.

```
set.seed(2023)
n <- 30
alpha <- 0.05
rep_ <- 1000
LI <- numeric()
LS <- numeric()

sigma2 <- 4
mu <- 1

for(i in 1:rep_){
  x <- rnorm(n, mean = mu, sd = sqrt(sigma2))
  LI[i] <- mean(x) - qnorm(1-alpha/2)*(sqrt(sigma2)/sqrt(n))
  LS[i] <- mean(x) + qnorm(1-alpha/2)*(sqrt(sigma2)/sqrt(n))
}

cat("A probabilidade de cobertura é ", sum(LI<mu & LS>mu)/rep_, "\n")
```

```
## A probabilidade de cobertura é 0.942
```

Exemplo 2: Intervalo de Confiança para a variância

Se X_1, \dots, X_n é uma amostra aleatória da distribuição $N(\mu, \sigma^2)$, com $n \geq 2$, e S^2 é a variância amostral, então

$$V = \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$$

O intervalo de confiança $100(1 - \alpha)\%$ unilateral é dado por

$$\left(0; \frac{(n-1)S^2}{\chi_{\alpha}^2} \right),$$

sendo X_{α}^2 o α -quantil da distribuição $\chi_{(n-1)}^2$. Se a população amostrada é normal com variância σ^2 , então a probabilidade que o intervalo de confiança contenha σ^2 é $1 - \alpha$.

O cálculo do LS para uma amostra aleatória de tamanho $n = 20$ da $N(0, \sigma^2 = 4)$ é dada por

```
set.seed(2023)
n <- 20
alpha <- 0.05
x <- rnorm(n, mean = 0, sd = 2)
LS <- ((n-1)*var(x))/qchisq(alpha, df = (n-1))
cat("O LS do intervalo de confiança unilateral é dado por", LS, "\n")
```

```
## O LS do intervalo de confiança unilateral é dado por 4.705917
```

No exemplo, a população amostrada é normal com $\sigma^2 = 4$, assim o nível de confiança é exato.

$$P\left(\frac{19 * S^2}{\chi_{0.05}^2(19)} > \sigma^2\right) = P\left(\frac{(n-1)S^2}{\sigma^2} > \chi_{0.05}^2(n-1)\right) = 0.95$$

Se a amostragem e estimação é repetida um número grande de vezes, aproximadamente 95% dos intervalos contém σ^2 , assumindo que a população amostrada é normal com variância σ^2 .

A seguir, geramos 1.000 amostras da distribuição $N(0, \sigma^2 = 4)$ de tamanho amostral 20 e encontramos quantas vezes o verdadeiro valor do parâmetro $\sigma^2 = 4$ pertence ao intervalo de confiança dado por

$$\left(0; \frac{(n-1)S^2}{\chi_{\alpha}^2}\right),$$

com $\alpha = 0.05$.

```
set.seed(2023)
n <- 5
alpha <- 0.05
m <- 1000
LS <- numeric(m)
sigma2 <- 4
x <- matrix(rnorm(n*m, mean = 0, sd = sqrt(sigma2)), ncol=n, nrow=m)
for(i in 1:m){
  LS[i] <- ((n-1)*var(x[i,]))/qchisq(alpha, df = (n-1))
}
sum(LS>sigma2)/m
```

```
## [1] 0.954
```

```
cat("A probabilidade de cobertura é ",sum(LS>sigma2)/m, "\n")
```

```
## A probabilidade de cobertura é 0.954
```

Obs.: nos exemplos anteriores, os intervalos de confiança adotados são oriundos de quantidades pivotais conhecidas. No entanto, na grande maioria das aplicações utiliza-se resultados assintóticos para a obtenção dos intervalos. No entanto, nesse cenário os resultados são verificados para “ n grande”, mas em geral não

conhecemos o valor do n . Assim, um estudo de simulação para verificar o tamanho amostral adequado se faz necessário.

O intervalo de confiança assintótico para uma parâmetro θ é dado por

$$\left(\hat{\theta} - z_{1-\alpha/2} \hat{\sigma}_{\theta}, \hat{\theta} + z_{1-\alpha/2} \hat{\sigma}_{\theta} \right),$$

sendo que $\hat{\sigma}_{\theta}$ pode ser adotado como a raiz quadrada do inverso da função hessian.

Exercícios

1. Considerando amostras de tamanho 10, 20, 30, 50 e 100, realizar o estudo de Monte Carlos considerando o intervalo de confiança assintótico e amostras das distribuições:
 - a. Bernoulli($p = 0,3$) - construir o IC para p ;
 - b. Poisson($\lambda = 4$) - construir o IC para λ
 - c. Exp($\lambda = 0.5$) - construir o IC para λ ;
 - d. Beta($\alpha = 2, \beta = 2$)- construir os ICs para α e β .

Considere 1.000 repetições de geração, construção do intervalo de confiança e verificação.