



Estatística Computacional II

Trabalho 2 - Incerteza em estimação de probabilidade

Luiz Francisco - GRR20213026

Mateus Souza - GRR20207154

novembro/2023

1.2 Incerteza em estimação de probabilidade

Foi coletado um conjunto de dados e julga-se que a distribuição exponencial $f(t) = \lambda e^{-\lambda t}$ com $t > 0, \lambda > 0$ é adequada para descrevê-lo. Utilize *bootstrap* para obter a distribuição amostral para:

```
dados <- c(2.77, 8.46, 0.83, 0.81, 6.86, 0.9, 0.62, 13.77, 0.15, 0.26, 8.36, 10.39)
```

a. O parâmetro λ da distribuição,

Para uma distribuição exponencial, temos que o MLE $\hat{\lambda}$ é dado por $\frac{n}{\sum_{i=1}^n x_i}$ que é a inversa da média dos dados.

```
require(boot)
library(ggplot2)
set.seed(2023)

mle_lambda <- function(dados, indices) {
  amostra <- dados[indices]
  return(1/mean(amostra))}

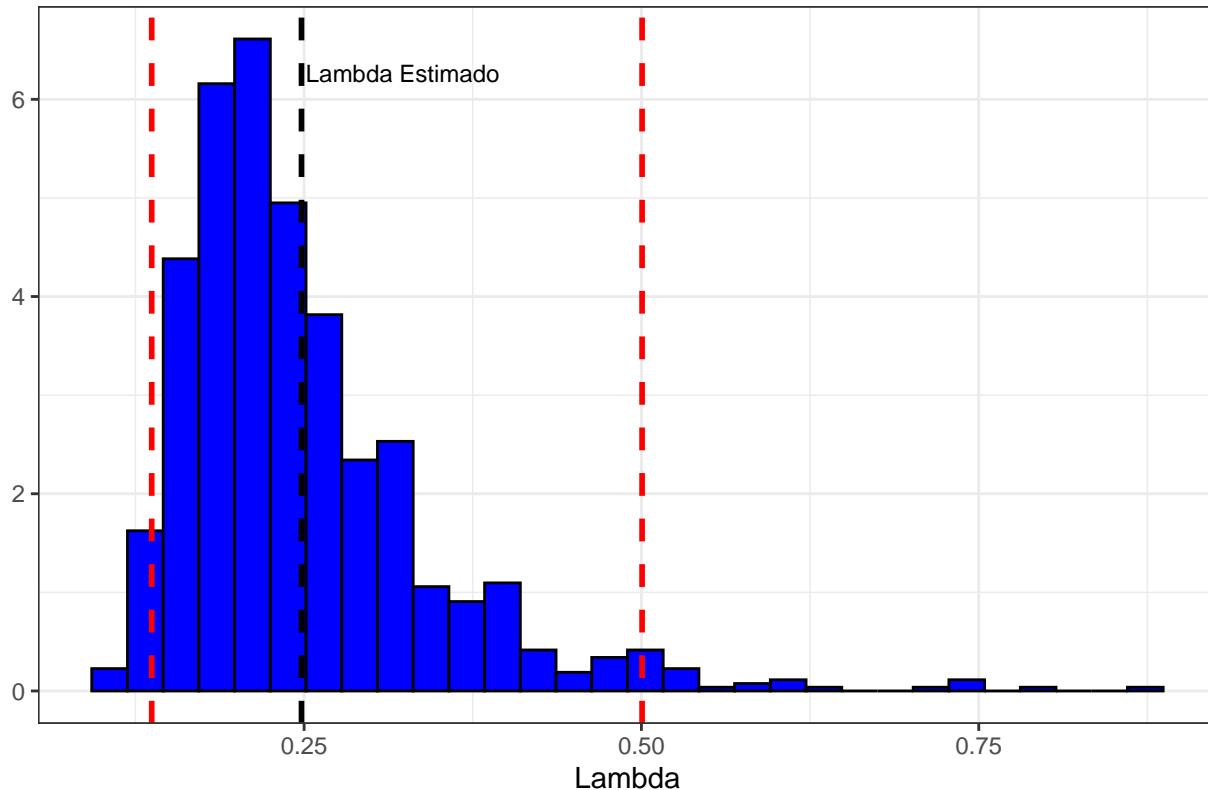
bootstrap <- boot(data = dados, statistic = mle_lambda, R = 1000)
cat("Lambda Estimado =", bootstrap$t0)

## Lambda Estimado = 0.2214839
ic <- quantile(bootstrap$t, c(0.025, 0.975))
cat("IC(95%) = ", ic[1], ic[2])

## IC(95%) = 0.1370615 0.5004905
df <- data.frame(x = bootstrap$t)

gg <- ggplot(df) +
  geom_histogram(aes(x = x, y = after_stat(density)), color = "black", fill = "blue") +
  geom_vline(aes(xintercept = mean(x)), color = "black", linewidth = 1, lty = "dashed") +
  geom_vline(aes(xintercept = ic[1]), color = "red", linewidth = 1, lty = "dashed") +
  geom_vline(aes(xintercept = ic[2]), color = "red", linewidth = 1, lty = "dashed") +
  theme_bw() +
  labs(title = "Distribuição Amostral de Lambda") +
  xlab("Lambda") +
  ylab(element_blank())
gg + annotate("text", x = mean(df$x) + 0.075, y = 6, label = "Lambda Estimado",
              color = "black", vjust = -1, size = 3)
```

Distribuição Amostral de Lambda



No gráfico acima temos a distribuição amostral λ pelo método *bootstrap*. Em preto está destacado o λ estimado e em vermelho os limites do intervalo de confiança de 95%.

b) A quantidade de interesse $\psi = P[t > 15]$

```
set.seed(2023)

psi <- function(dados, indices) {
  amostra <- dados[indices]
  lambda_estimado <- 1/mean(amostra)
  return(1 - pexp(15, rate = lambda_estimado))}

bootstrap_psi <- boot(data = dados, statistic = psi, R = 1000)
cat("Estimativa de Psi =", bootstrap_psi$t0, "\n")

## Estimativa de Psi = 0.03607125
ic_psi <- quantile(bootstrap_psi$t, c(0.025,0.975))
cat("IC(95%) para Psi =", ic_psi[1], ic_psi[2], "\n")

## IC(95%) para Psi = 0.000549043 0.1279749
df_psi <- data.frame(x = bootstrap_psi$t)

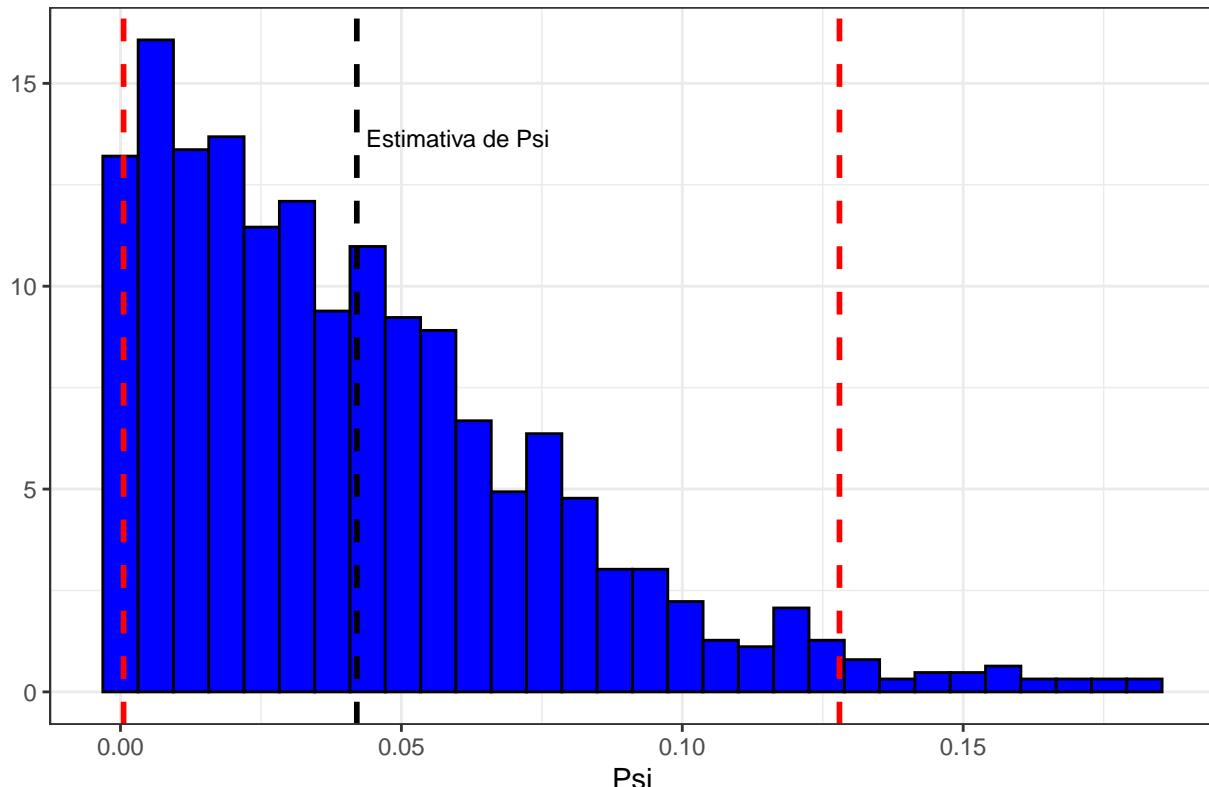
gg_psi <- ggplot(df_psi) +
  geom_histogram(aes(x = x, y = after_stat(density)), color = "black", fill = "blue") +
  geom_vline(aes(xintercept = mean(x)), color = "black", linewidth = 1, lty = "dashed") +
  geom_vline(aes(xintercept = ic_psi[1]), color = "red", linewidth = 1, lty = "dashed") +
```

```

geom_vline(aes(xintercept = ic_psi[2]), color = "red", linewidth = 1, lty = "dashed")+
theme_bw() +
labs(title = "Distribuição Amostral de Psi") +
xlab("Psi") +
ylab(element_blank())
gg_psi + annotate("text", x = mean(df_psi$x) + 0.018, y = 13, label = "Estimativa de Psi",
color = "black", vjust = -1, size = 3)

```

Distribuição Amostral de Psi



No gráfico acima temos a distribuição amostral para a probabilidade $\psi = P[t > 15]$ utilizando o método *bootstrap*. Em preto está destacado o valor estimado e em vermelho os limites do intervalo de confiança de 95%.

Compare os intervalos de confiança com obtidos por algum outro método que não seja o *bootstrap*.

```

n <- length(dados)
alpha <- 0.05
z <- qnorm(1 - alpha / 2)

ic_analitico <- c(1/mean(dados) - z * sqrt(1/sum(dados^2)),
1/mean(dados) + z * sqrt(1/sum(dados^2)))
cat("IC(95% | Bootstrap) = ", ic, "\nIC(95% | Wald) = ", ic_analitico)

## IC(95% | Bootstrap) = 0.1370615 0.5004905
## IC(95% | Wald) = 0.1335134 0.3094545

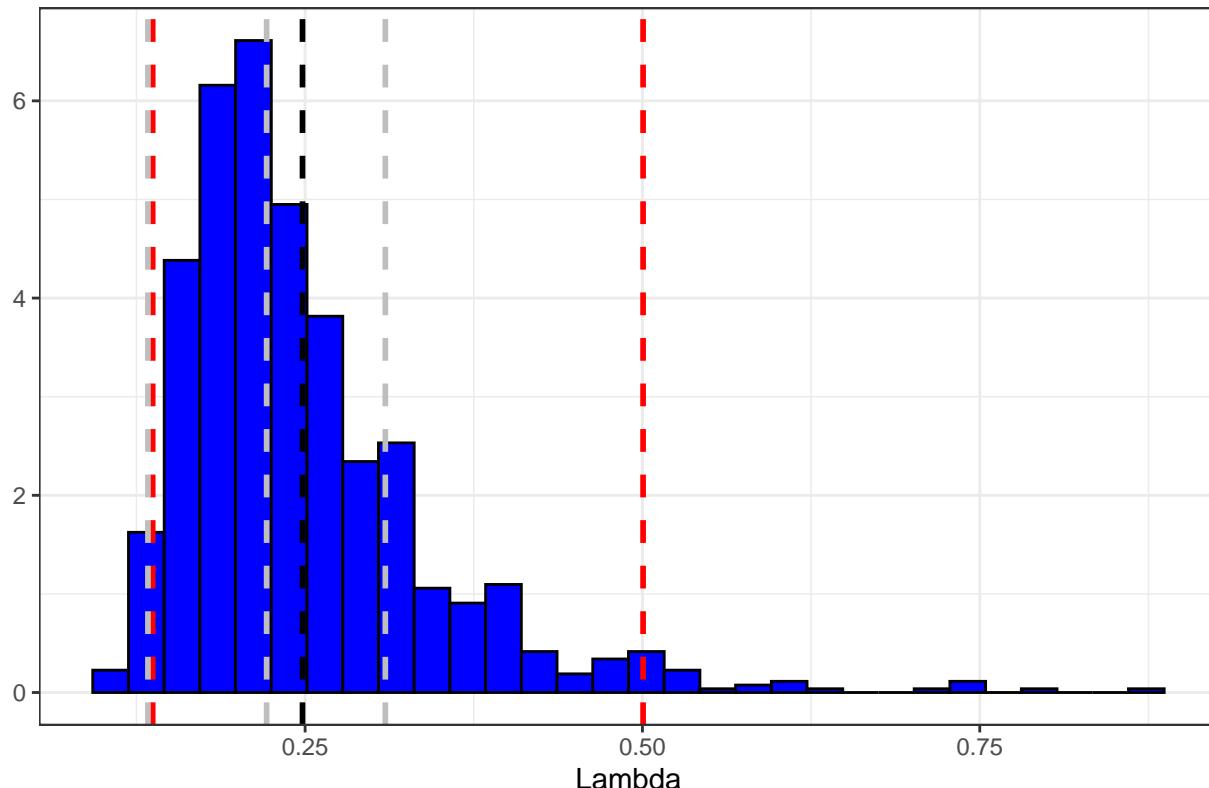
```

```

gg <- gg+
  geom_vline(aes(xintercept = 1/mean(dados)), color = "gray", linewidth = 1, lty = "dashed")+
  geom_vline(aes(xintercept = ic_analitico[1]), color = "gray", linewidth = 1, lty = "dashed")+
  geom_vline(aes(xintercept = ic_analitico[2]), color = "gray", linewidth = 1, lty = "dashed")
gg

```

Distribuição Amostral de Lambda



```

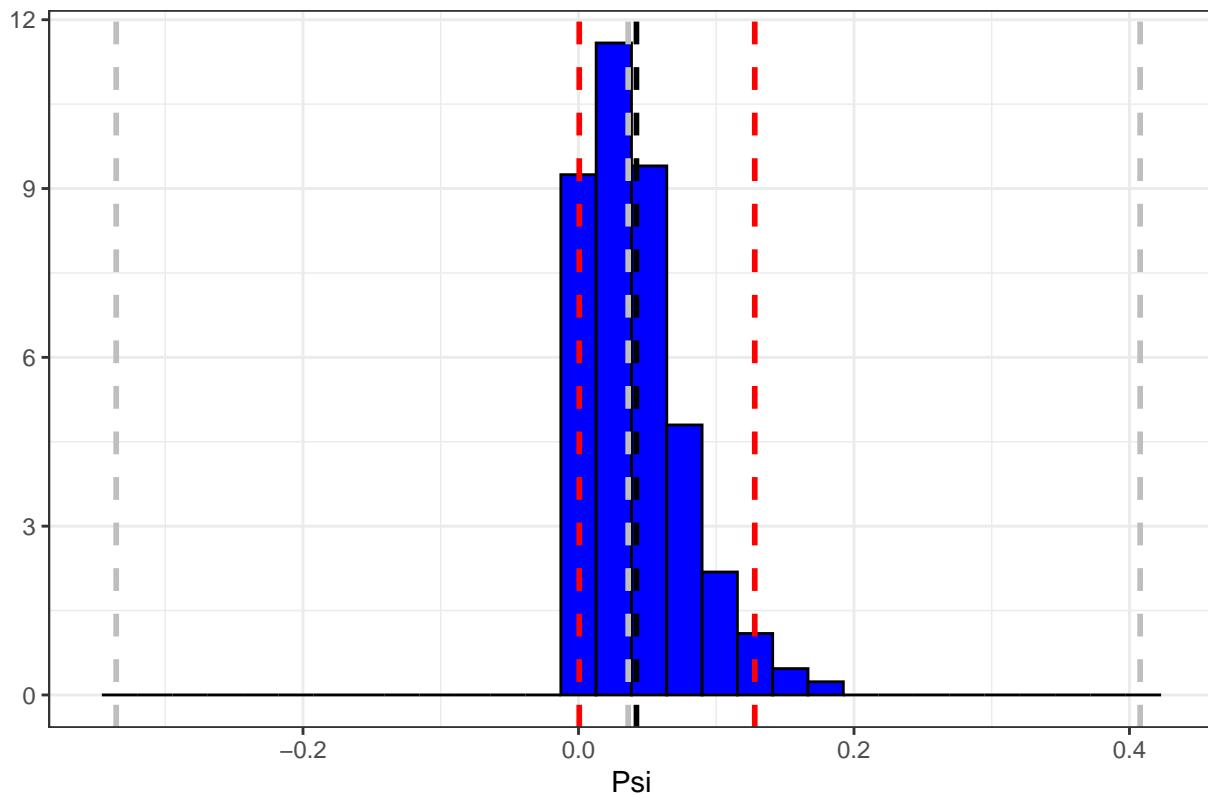
ic_analitico_psi <- c(psi(dados) - z * sqrt(psi(dados) * (1-psi(dados)/n)),
                      psi(dados) + z * sqrt(psi(dados) * (1-psi(dados)/n)))
cat("IC(95% Psi | Bootstrap) = ", ic_psi, "\nIC(95% Psi | Wald) = ", ic_analitico_psi)

## IC(95% Psi | Bootstrap) =  0.000549043 0.1279749
## IC(95% Psi | Wald) =  -0.3356137 0.4077562

gg_psi <- gg_psi+
  geom_vline(aes(xintercept = psi(dados)), color = "gray", linewidth = 1, lty = "dashed")+
  geom_vline(aes(xintercept = ic_analitico_psi[1]), color = "gray", linewidth = 1, lty = "dashed")+
  geom_vline(aes(xintercept = ic_analitico_psi[2]), color = "gray", linewidth = 1, lty = "dashed")
gg_psi

```

Distribuição Amostral de Psi



Ao gráficos mostrados anteriormente foram adicionadas as linhas cinzas que representam a estimativa e os limites do intervalo de confiança, calculados de forma analítica, para λ e ψ , respectivamente. O intervalo utilizado foi o de Wald.

Como podemos perceber λ é subestimado e esse fato leva a um intervalo de confiança muito mais largo, que é explicado pela própria estrutura do intervalo de Wald, que tem uma taxa de cobertura menor do que a taxa nominal para valores muito extremos, como é o caso para $\psi = 0.0361$.

Cabe ressaltar que essas estimativas analíticas foram calculadas em uma distribuição extremamente assimétrica (exponencial) aliada a um valor também extremo para a probabilidade ψ , resultando assim nesse intervalo alargado.

A utilização do método *bootstrap* se mostrou então bastante efetiva para contornar esses problemas e oferecer um intervalo menor que resulta em mais segurança para uma tomada de decisão.