Estatística computacional - Trabalho 2

Arthur Sonntag Kuchenbecker

July 2019

1 Introdução

O objetivo do trabalho é demonstrar computacionalmente alguns conceitos fundamentais da teoria da probabilidade, bem como a relacionados à inferência estatística.

Isso será feito por meio de quatro exercícios específicos no software `R`, conforme explicado nas seções seguintes.

2 Lei dos Grandes Números

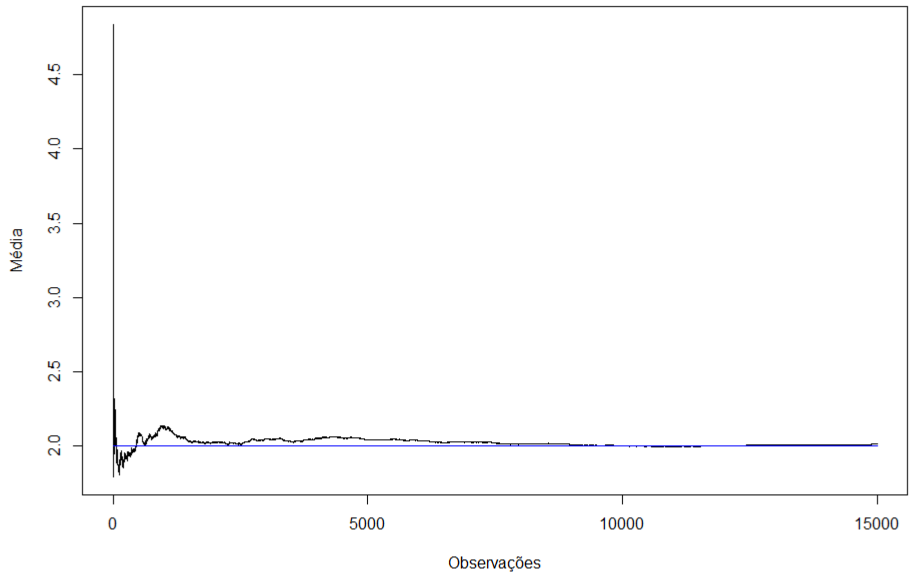
O primeiro exercício consiste em demonstrar empiricamente a conhecida Lei dos Grandes Números.

De acordo com este teorema, a média aritmética ($\bar{x}$) dos resultados da realização da mesma experiência repetidas vezes tende a se aproximar do valor esperado à medida que mais tentativas se sucederem.

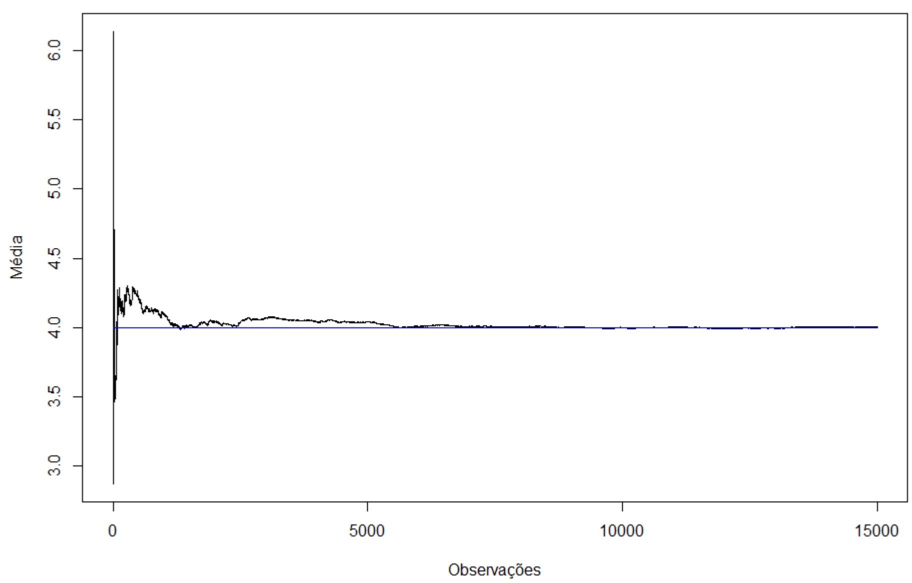
No experimento realizado, uma amostra de 15 mil observações será criada a partir de uma distribuição exponencial com $\lambda = \frac{1}{2}$.

Espera-se que quanto maior o número de observações utilizado para compor a média ($\bar{x}$), mais próximo o valor obtido seja de 2 (valor esperado de uma distribuição com as características apresentadas).

O resultado pode ser observado no gráfico abaixo:



Testando também para o caso mais geral da distribuição gamma (com parâmetros $\alpha = 2$ e $\beta = \frac{1}{2}$) temos que $\bar{x}$ deve convergir para 4.



3 Teorema Central do Limite

O Teorema Central do Limite diz que quando o tamanho da amostra aumenta, a distribuição amostral da sua média aproxima-se cada vez mais da distribuição normal.

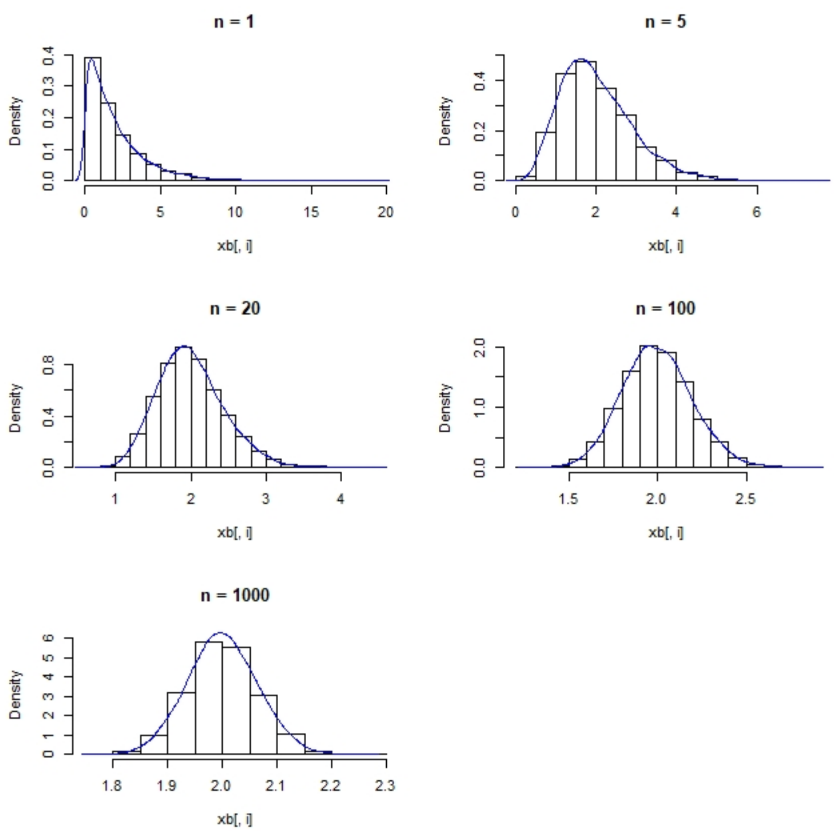
Para mostrar computacionalmente o resultado, serão criadas 10 mil amostras de uma distribuição exponencial (com $\lambda = \frac{1}{2}$) para cada um dos tamanhos (n_i) contidos no vetor n, definido abaixo.

$$n = (1, 5, 20, 100, 1000) \tag{1}$$

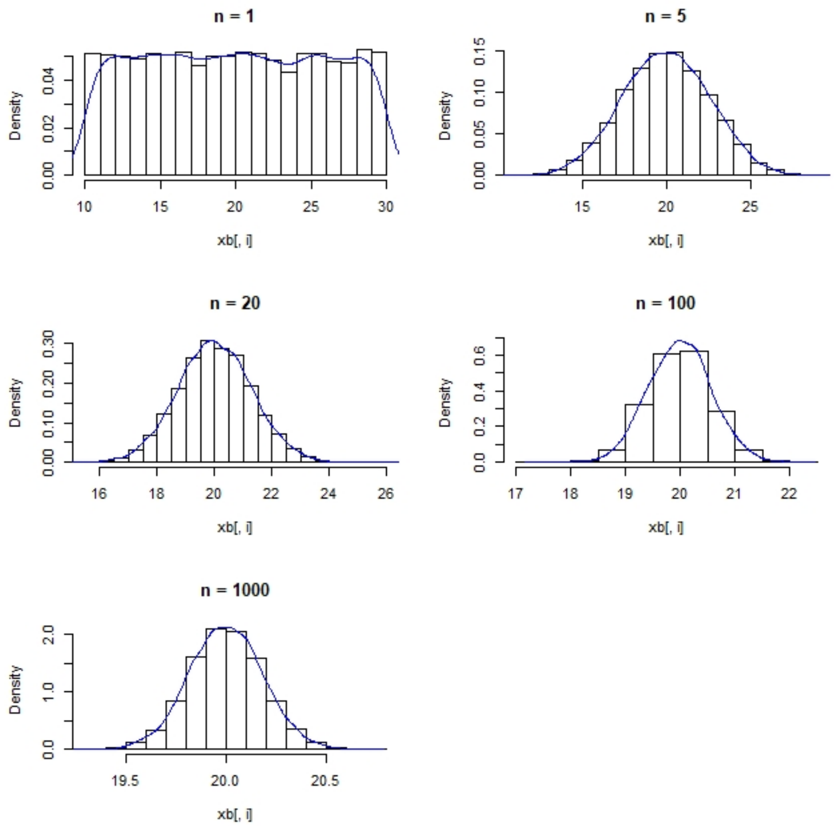
Para cada uma das 10 mil amostras, será computada a média aritmética das observações ($\bar{x}$), de modo que será possível mostrar graficamente as características da distribuição de $\bar{x}$.

Quanto maior o tamanho da amostra da distribuição exponencial, mais perto a distribuição de $\bar{x}$ deve estar de uma distribuição normal.

Seguem os resultados:



Testando também para uma distribuição uniforme (com valores entre 10 e 30), observamos o seguinte:



É importante observar que para $n = 1$ a distribuição da média amostral ($\bar{x}$) segue a distribuição original (exponencial ou uniforme, para os casos estudados).

Isso acontece porque quando a amostra possui apenas uma observação, a média amostral assume valor igual ao desta única observação, gerando - portanto - uma amostra de tamanho 10 mil da distribuição original.

4 Coeficiente de Confiança

Neste exercício, o objetivo é verificar o coeficiente de confiança na construção de Intervalos de Confiança.

Em termos práticos, serão criadas 10 mil amostras de uma distribuição normal $\sim N(0, 1)$, e serão estimados os limites superiores e inferiores do intervalo de confiança para a média da distribuição.

Com isso, será possível verificar quantas vezes a média da distribuição fica fora do intervalo de confiança estimado. Considerando que o coeficiente de confiança escolhido é de 0.9, espera-se que a média da distribuição esteja contida no intervalo de confiança para aproximadamente 9 mil repetições.

Abaixo segue o código utilizado, o qual realiza o experimento citado para dois tamanhos de amostra diferentes: 10 e 100.

```
nr = 10e3
nv = c(10,100)
nvt = length(nv)
mi = 0
de = 1
ga = 0.90

al2 = (1-ga)/2
ta = rep(0,nvt)

for (n in 1:nvt) {
  for (i in 1:nr){
    x = rnorm(nv[n],mi,de)
    xb = mean(x)
    sn = sd(x)

    t = qt(1-al2,nv[n]-1)
    li = xb-t*(sn/(nv[n]^.5))
    ls = xb+t*(sn/(nv[n]^.5))

    if (mi>=li & mi<=ls) {
      ta[n]=ta[n]+1
    }
  }
}
print(ta)
```

O resultado da função `print` segue abaixo:

```
> print(ta)
[1] 8994 9037
```

O que está de acordo com o esperado.

5 Teste de Hipóteses

No último exercício proposto, o objetivo é calcular o poder do teste t, que faz inferência a respeito da média da distribuição populacional.

Definimos aqui o poder do teste como $(1 - p)$ em que p é a probabilidade de cometer erro do tipo II.

O erro do tipo II é cometido ao aceitar a hipótese nula, quando esta é, na verdade, falsa. Para este exercício, serão criadas 10 mil amostras de uma distribuição normal $\sim N(0, 1)$, e serão realizados testes de hipótese com a hipótese nula de que a média da distribuição é diferente de 0.

O poder do teste sera calculado computacionalmente, já que será avaliada a quantidade de vezes em que a hipótese nula é aceita neste conjunto de 10 mil amostras.

Segue o código (para tamanhos de amostra 10 e 20, coeficiente de confiança 0.9, e $H0 : \mu = 0.5$).

```
nr = 10e3
nv = c(10,20)
nvt = length(nv)
mi = 0
de = 1
ga = 0.90
al2 = (1-ga)/2

ta = rep(0,nvt)

# Hipotese nula -> mi = mic
mic = 0.5

for (n in 1:nvt) {
  for (i in 1:nr){
    x = rnorm(nv[n],mi,de)
    xb = mean(x)
    sn = sd(x)

    # Estatistica do teste T
    tobs = (xb-mic)/(sn/(nv[n]^.5))
    t = qt(1-al2,nv[n]-1)

    # Quantas vezes aceitamos a hipotese nula
    if (tobs<=t & tobs>=(-t)) {
      ta[n]=ta[n]+1
    }
  }
}
```

```
print(nr-ta)
```

O resultado da função `print` é:

```
> print(nr-ta)
[1] 4276 6970
```

Ou seja, o poder do teste é de 0.4276 para a amostra de tamanho 10 e 0.697 para a amostra de tamanho 20.

Espera-se que quanto mais longe de 0 for o μ da hipótese nula, maior seja o poder do teste, na medida em que é cada vez menos provável que a hipótese nula seja aceita.

Abaixo seguem os resultados da função `print` para as seguintes hipóteses nulas:

$H0 : \mu = 1$

```
> print(nr-ta)
[1] 8976 9965
```

$H0 : \mu = 1.5$

```
> print(nr-ta)
[1]  9955 10000
```