

Tópicos em Atuária Computacional

Sumário

1	Introdução ao uso do <i>software</i> R	5
2	Geração de números aleatórios	15
3	Simulação de variáveis aleatórias	18
3.1	Método de Transformação Inversa	18
3.1.1	Caso Contínuo	18
3.1.2	Caso Discreto	21
3.2	Método de Aceitação-Rejeição	32
4	Métodos Monte Carlo em Inferência Estatística	38
4.1	Distribuição amostral da média	38
4.2	Estimadores	41
4.2.1	Estimadores de máxima verossimilhança	44
4.2.2	Estimadores de mínimos quadrados	51
4.2.3	Estimação intervalar	55
4.2.4	Integração Monte Carlo	62
4.2.5	Amostragem por importância	70
4.2.6	Testes de hipóteses	73
4.2.6.1	Testes de aderência	79
4.3	Estatística bayesiana	85
4.3.1	Inferência preditiva	90

4.4	Método <i>bootstrap</i>	94
5	Introdução aos Métodos Monte Carlo via Cadeias de Markov	103
5.1	Cadeias de Markov	103
5.2	Integração Monte Carlo via Cadeias de Markov	109
5.3	Algoritmo de Metropolis-Hastings	110
6	Aplicações	112
6.1	Modelo de risco coletivo	112
6.2	Modelo clássico de ruína	114
6.3	Resseguro não proporcional	117
6.4	Cálculo do prêmio utilizando estatística bayesiana	119
6.5	Cálculo do Valor em Risco (VaR)	123
7	Referências	125

Prefácio

Um atuário pode estar interessado numa determinada informação contida em um modelo atuarial que envolve variáveis aleatórias. Na ausência de uma solução analítica de tal modelo ou na ausência de um banco de dados apropriado, pode-se utilizar simulação de variáveis aleatórias.

Muitos problemas em Ciências Atuariais envolvem a área de Estatística Computacional, como na análise de modelos de perdas em seguros. Nesta obra, buscamos apresentar alguns tópicos de Atuária Computacional, abordando métodos estatísticos com simulação de variáveis aleatórias. A simulação de variáveis aleatórias deu origem aos chamados Métodos Monte Carlo (MMC). O termo Monte Carlo se refere ao distrito de Monte Carlo, localizado no Principado de Mônaco na Riviera Francesa, conhecido pelos seus cassinos, sendo as roletas consideradas como mecanismos simples para gerar números aleatórios.

A Atuária Computacional é extremamente importante na construção e avaliação de cenários de modelos atuariais nos ramos vida e não vida. A presente obra se destina aos alunos dos cursos de graduação em Ciências Atuariais, onde poderão conhecer a Estatística Computacional com aplicações em modelos atuariais.

A obra apresenta resoluções de questões presentes nos exames aplicados pela *Casualty Actuarial Society* (CAS) e *Society of Actuaries* (SOA). Os exemplos computacionais foram implementados no *software* livre R para análise de dados, que pode ser obtido pelo site *The Comprehensive R Archive Network* (CRAN), no endereço <https://cran.r-project.org/>.

O Capítulo 1 apresenta uma introdução ao uso do *software* R. O Capítulo 2 apresenta a geração de números aleatórios, abordando o método congruencial misto para gerar uma sequência de números pseudo-aleatórios. No Capítulo 3, é apresentado o método de transformação inversa para simular variáveis aleatórias discretas e contínuas. No Capítulo 4, são abordados Métodos Monte Carlo em inferência estatística. No Capítulo 5, é apresentada uma introdução aos Métodos Monte Carlo via Cadeias de Markov. Por fim, no Capítulo 6, são apresentadas aplicações, abordando: modelo de risco coletivo, modelo clássico de ruína, resseguro não proporcional, cálculo de prêmio utilizando estatística bayesiana e cálculo do valor em risco.

Desejamos que a presente obra seja útil aos estudantes dos cursos de graduação em Ciências Atuariais, contribuindo de maneira significativa para o aprendizado e futuros trabalhos.

Os autores

Capítulo 1

Introdução ao uso do *software* R

O *software* livre R tem código fonte aberto, utilizado para análise e manipulação de dados. O R pode ser obtido pelo site *The Comprehensive R Archive Network* (CRAN), no endereço <https://cran.r-project.org/>.

Um dos símbolos importante na utilização do R é o símbolo `#`. Numa linha de comando, o que for digitado após tal símbolo será ignorado pelo *software*. Os pacotes instalados podem ser mostrados por meio do comando **library()**. Para carregar um pacote em específico, basta digitar **library(nome_do_pacote)**. Para se obter uma descrição detalhada de um determinado comando, basta digitar **help()**.

Um objeto pode ser criado utilizando o sinal de menos e o símbolo `<` ou `>`. Também pode-se utilizar o sinal de igualdade `=`. As funções **mode()** e **length()** mostram o tipo (por exemplo, numérico, caractere e funções) e o tamanho de um objeto, respectivamente. A seguir, serão apresentados

tipos de objetos que podem ser trabalhados no R.

Vetores - a função `c()` é utilizada para criar um vetor a partir de seus argumentos. O exemplo a seguir traz o vetor `x` que contém a sequência de números inteiros de 1 a 5. Tal sequência também poderia ser criada utilizando dois pontos.

```
x<-c(1,2,3,4,5)
x

## [1] 1 2 3 4 5

x<-1:5
x

## [1] 1 2 3 4 5
```

Sequências - a função `seq()` é utilizada para produzir sequência de valores, que tem como argumentos: início, fim e os passos da sequência. O exemplo a seguir traz a sequência de valores entre 5 e 20, com passo igual a 2.

```
seq(5,20,2)

## [1] 5 7 9 11 13 15 17 19
```

Listas - a função `list()` é utilizada para construir listas que podem combinar, por exemplo, vetores, caracteres e matrizes em um mesmo objeto. O exemplo a seguir traz uma lista com informações de um indivíduo. Pode-se acessar cada componente da lista utilizando o símbolo `$`.

```
individuo<-list(sexo = 'masculino',idade = '37')
individuo

## $sexo
## [1] "masculino"
##
## $idade
## [1] "37"

individuo$sexo #acessando o primeiro componente da lista.

## [1] "masculino"
```

Matrizes - a função `matrix()` é utilizada para construir matrizes. Tal função recebe um vetor como argumento e o transforma em uma matriz conforme as dimensões especificadas. Semelhantes as matrizes, têm-se os `data.frames`, onde cada coluna pode armazenar elementos de diferentes tipos. O exemplo a seguir apresenta a matriz `A`, contendo o vetor `a`, com número de colunas igual a 4. Observe que a matriz é preenchida ao longo das colunas. O preenchimento pode ser realizado por linhas, adicionando o argumento `byrow=TRUE`. Também são apresentadas as funções `dim()`, `summary()`, `rbind()` e `cbind()`.


```
a<-1:20
A<-matrix(a,ncol=4)
A

##      [,1] [,2] [,3] [,4]
## [1,]    1    6   11   16
## [2,]    2    7   12   17
## [3,]    3    8   13   18
## [4,]    4    9   14   19
## [5,]    5   10   15   20

A<-matrix(a,ncol=4,byrow=TRUE)
A

##      [,1] [,2] [,3] [,4]
## [1,]    1    2    3    4
## [2,]    5    6    7    8
## [3,]    9   10   11   12
## [4,]   13   14   15   16
## [5,]   17   18   19   20

#dimensões da matriz A.
dim(A)

## [1] 5 4
```

#função summary() - medidas descritivas das colunas.

summary(A)

```
##           V1           V2           V3           V4
##  Min.      : 1    Min.      : 2    Min.      : 3    Min.      : 4
## 1st Qu.: 5      1st Qu.: 6      1st Qu.: 7      1st Qu.: 8
## Median : 9      Median :10      Median :11      Median :12
## Mean   : 9      Mean   :10      Mean   :11      Mean   :12
## 3rd Qu.:13      3rd Qu.:14      3rd Qu.:15      3rd Qu.:16
## Max.    :17      Max.    :18      Max.    :19      Max.    :20
```

#adicionando uma nova coluna com a função cbind.

B<-cbind(A,1:5)

B

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]    1    2    3    4    1
## [2,]    5    6    7    8    2
## [3,]    9   10   11   12    3
## [4,]   13   14   15   16    4
## [5,]   17   18   19   20    5
```

#adicionando uma nova linha com a função rbind.

C<-rbind(A,c(3,7,5,6))

C

```
##      [,1] [,2] [,3] [,4]
```

```
## [1,] 1 2 3 4
## [2,] 5 6 7 8
## [3,] 9 10 11 12
## [4,] 13 14 15 16
## [5,] 17 18 19 20
## [6,] 3 7 5 6

#extraíndo a quarta linha da matriz C.
C[4,]

## [1] 13 14 15 16

#extraíndo as duas primeiras colunas de C.
C[,c(1,2)]

##      [,1] [,2]
## [1,] 1 2
## [2,] 5 6
## [3,] 9 10
## [4,] 13 14
## [5,] 17 18
## [6,] 3 7
```

No R, uma função pode ser criada com base nas seguintes atribuições:

nomefunção < –function(argumento 1,...,argumento n).

Com a implementação da função, ela pode ser utilizada, fazendo:

nomefunção(argumento 1,...,argumento n).

A seguir, é apresentado um exemplo de uma função para calcular a média de um conjunto de dados.

```
#criando a função media
media<-function(x){      #argumento - vetor x (dados).
  somax<-sum(x)           #soma de x.
  n<-length(x)           #tamanho de x.
  mediax<-somax/n        #média de x.
  return(mediax)         #retorna o valor da média de x.
}

x<-c(1,2,5,6,10) #exemplo de um conjunto de dados.
media(x)         #utilizando a função media.

## [1] 4.8

mean(x) #obtendo a média pelo comando mean.

## [1] 4.8
```

Também pode-se criar funções envolvendo estruturas de repetição, com os comandos **while(condição)** e **for(condição)**, e estruturas de seleção utilizando **if()**. A estrutura de seleção pode ser apresentada no seguinte formato:

if(condição){comandos que satisfazem a condição}else{comandos que não satisfazem a condição}.

Vale resaltar que a especificação do **else()** não é obrigatória. A seguir, são apresentados alguns exemplos utilizando estruturas de repetição e seleção.

```
t<-7
if(t>10) t else 2*t

## [1] 14

ifelse(t>10,t,2*t)

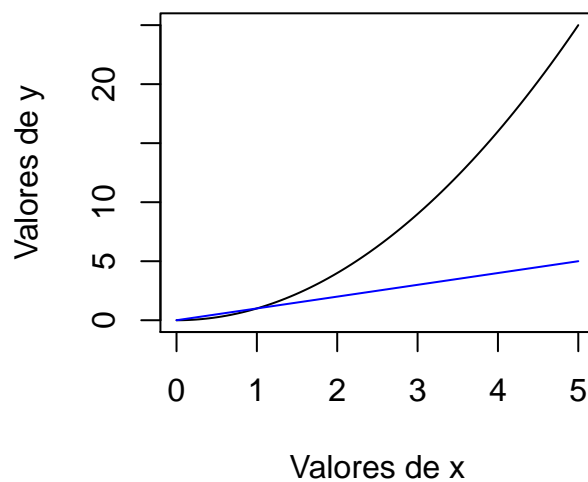
## [1] 14

y<-1:10
for(i in 1:5){
  if(y[i]<=3){
    print("Menor ou igual a 3")
  } else{
    print("Maior do que 3")
  }
}
```

```
    }  
  }  
  
## [1] "Menor ou igual a 3"  
## [1] "Menor ou igual a 3"  
## [1] "Menor ou igual a 3"  
## [1] "Maior do que 3"  
## [1] "Maior do que 3"  
  
n<-0  
soma<-0  
while(soma<=5){  
  n<-n+1  
  soma<-soma+n  
}  
soma  
  
## [1] 6
```

A construção de gráficos no R pode ser realizada de diferentes maneiras. A função **plot()** cria um gráfico com os valores de coordenadas x e y. Nos gráficos, pode-se adicionar pontos e linhas utilizando os comandos **points()** e **lines()**, alterando os tipos de pontos e de linhas. Também pode-se definir os intervalos dos eixos e adicionar textos. A seguir, é apresentado um exemplo de um gráfico utilizando o comando **plot()**.

```
x<-seq(0,5,0.01)
y<-x**2
#plotando os pares (x,y)
plot(x,y,type="l",      #argumento type="l" - linha suave.
      xlab="Valores de x",ylab="Valores de y",  #nomes dos eixos.
      xlim=c(0,5),ylim=c(0,25)) #definindo os intervalos.
lines(x,sqrt(y),col=4) #adicionando linha azul.
```



Capítulo 2

Geração de números aleatórios

A simulação de variáveis aleatórias deu origem aos Métodos Monte Carlo (MMC), que tem como objetivo, encontrar realizações aleatórias de um modelo probabilístico. Os MMCs utilizam uma sequência de números pseudo-aleatórios, que são considerados como números aleatórios, sendo que o método mais utilizado é o congruencial linear.

O método congruencial linear gera uma sequência de números pseudo-aleatórios a partir de um valor inicial U_0 , denominado de semente, por meio de uma relação matemática recursiva.

Com base no valor U_0 , são gerados os valores U_1, U_2, \dots por meio da relação

$$U_{i+1} = (aU_i + c) \bmod m,$$

em que U_0 , a , m são inteiros não negativos e $i = 0, 1, 2, \dots, m - 1$. a é denominado de multiplicador, c é o incremento e m é o módulo. A sequência

de números pseudo-aleatórios entre 0 e 1 é obtida fazendo $\frac{U_{i+1}}{m}$, que representa as realizações de uma variável aleatória uniformemente distribuída no intervalo $(0, 1)$.

Vale ressaltar que a operação módulo (mod) é usada para encontrar o resto da divisão de um número por outro. Por exemplo: $16 \bmod 5$ é o resto da divisão de 16 por 5, que é igual a 1.

No *software* R, uma sequência de números aleatórios entre a e b é gerada pelo comando **runif(n,a,b)**, utilizando o método congruencial linear, sendo que **n** é o tamanho da amostra. Ainda no *software* R, para especificar a semente, deve-se usar o comando **set.seed**, estabelecendo assim uma sequência reproduzível de números aleatórios. No caso, o argumento do comando deve ser um número inteiro.

EXEMPLO 2.1. Gere uma sequência de números aleatórios, considerando $U_0 = 2$, $a = 5$, $c = 0$ e $m = 32$.

Solução:

Como $c = 0$, tem-se o método congruencial linear multiplicativo, em que $U_{i+1} = aU_i \bmod m$. A tabela a seguir apresenta uma sequência de números aleatórios a partir da semente $U_0 = 2$.

i	0	1	2	3	4	5	...
U_i	2	10	18	26	2	10	...

No caso, tem-se que:

$$U_0 = 2$$

$$U_1 = 10 \bmod 32 = 10$$

$$U_2 = 50 \bmod 32 = 18$$

$$U_3 = 90 \bmod 32 = 26$$

$$U_4 = 130 \bmod 32 = 2$$

$$U_5 = 10 \bmod 32 = 10$$

$$\vdots$$

Pode-se observar que a partir de $i = 4$, os valores da sequência se repetem. Dessa maneira, o ciclo ou período do gerador de números aleatórios, denotado por h , é igual a 4. Os números aleatórios correspondentes ao intervalo de 0 a 1 são: 0,0625; 0,3125; 0,5625; 0,8125; 0,0625; 0,3125; ...

Capítulo 3

Simulação de variáveis aleatórias

Neste capítulo, será apresentado o método de transformação inversa, utilizado para obter realizações de variáveis aleatórias de determinadas distribuições de probabilidade discretas ou contínuas, a partir de números aleatórios uniformes. Tal método aplica o teorema da transformação integral de probabilidade. Neste capítulo, também será apresentado o método de aceitação-rejeição para geração de variáveis aleatórias.

3.1 Método de Transformação Inversa

3.1.1 Caso Contínuo

O método de transformação inversa é baseado no teorema da transformação integral de probabilidade, apresentado a seguir.

Teorema 1. *Sendo X uma variável aleatória contínua com função de distribuição $F_X(x)$, então a variável aleatória $U = F_X(X)$ é uniformemente*

distribuída no intervalo $[0, 1]$.

Demonstração: Sabendo que X tem função densidade de probabilidade $f_X(x)$ e função de distribuição $F_X(x)$ invertível, considere que $G_U(u) = P(U \leq u)$. Dessa maneira, segue que:

$$\begin{aligned} G_U(u) &= P(U \leq u) = P(F_X(X) \leq u), \\ &= P(X \leq F_X^{-1}(u)) = F_X(F_X^{-1}(u)) = u. \end{aligned} \tag{3.1}$$

Portanto, $G_U(u) = u$, na qual é a função de distribuição de uma uniforme no intervalo $[0, 1]$. Assim, $U \sim \text{Uniforme}(0, 1)$.

□

Pode-se resumir, o método de transformação inversa nos dois passos a seguir:

1. Encontrar $F_X^{-1}(u)$;
2. Para cada variável aleatória:
 - a) Gerar u da distribuição $\text{Uniforme}(0, 1)$;
 - b) Encontrar $x = F_X^{-1}(u)$.

EXEMPLO 3.1. Gere uma variável aleatória com função densidade de probabilidade (f.d.p)

$$f_X(x) = \begin{cases} 0, & \text{se } x < 0, \\ e^{-x}, & \text{se } x \geq 0. \end{cases}$$

Solução:

Sabendo que

$$F_X(x) = \int_{-\infty}^x f_S(s)ds = \int_0^x e^{-s}ds = [-e^{-s}]_0^x = 1 - e^{-x},$$

pelo Teorema 1, obtém-se $x = F_X^{-1}(u)$:

$$\begin{aligned} F_X(x) &= 1 - e^{-x}, \\ u &= 1 - e^{-x}, \\ e^{-x} &= 1 - u, \\ x &= -\ln(1 - u) = F_X^{-1}(u), x \geq 0. \end{aligned}$$

Algoritmo:

1. Gerar u da distribuição $Uniforme(0, 1)$;
2. Encontrar $x = F_X^{-1}(u) = -\ln(1 - u)$.

EXEMPLO 3.2. Para uma seguradora, o valor do sinistro segue uma distribuição com função densidade dada por

$$f_X(x) = 57800x^{-3}, x \geq 170.$$

Considere uma franquia igual a 400 e um valor máximo de cobertura igual a 1200. Os seguintes números aleatórios de uma distribuição $Uniforme(0, 1)$

são considerados: 0,87, 0,41 e 0,98. Encontre os pagamentos realizados pela seguradora.

Solução:

Pelo método da transformação integral, obtém-se:

$$F_X(x) = \int_{170}^x 57800s^{-3}ds = -\frac{28900}{x^2} + 1, x \geq 170.$$

$$u = -\frac{28900}{x^2} + 1,$$

$$x = \frac{1}{\sqrt{\frac{1-u}{28900}}},$$

em que u vem da distribuição $Uniforme(0, 1)$.

Considerando a franquia igual a 400, valor máximo de cobertura igual a 1200 e os valores gerados de U , obtém-se os pagamentos realizados pela seguradora:

u	x	Pagamento
0,87	471,50	71,50
0,41	221,32	0
0,98	1202,08	800

3.1.2 Caso Discreto

O método da transformação integral pode ser aplicado para distribuições discretas. Se X é uma variável aleatória discreta, então:

$$\cdots < x_{i-1} < x_i < x_{i+1} < \cdots$$

são os pontos de descontinuidade de $F_X(x)$, onde a transformação é $F_X^{-1}(u) = x_i$, sendo $F_X(x_{i-1}) \leq u < F_X(x_i)$.

De maneira geral, pode-se obter X conforme os valores simulados de U , sendo:

$$X = \begin{cases} x_0, & \text{se } u < p_0, \\ x_1, & \text{se } p_0 \leq u < p_0 + p_1, \\ x_2, & \text{se } p_0 + p_1 \leq u < p_0 + p_1 + p_2, \\ \vdots & \vdots \\ x_i, & \text{se } \sum_{k=0}^{i-1} p_k \leq u < \sum_{k=0}^i p_k, \\ \vdots & \vdots \end{cases}$$

em que $p_i = P(X = x_i) = P(\sum_{k=0}^{i-1} p_k \leq U < \sum_{k=0}^i p_k) = F_U(\sum_{k=0}^i p_k) - F_U(\sum_{k=0}^{i-1} p_k) = \sum_{k=0}^i p_k - \sum_{k=0}^{i-1} p_k$, onde $U \sim \text{Uniforme}(0, 1)$.

EXEMPLO 3.3 Gere um valor de uma variável aleatória que possui a seguinte distribuição de probabilidade.

$$X = \begin{cases} 2, & \text{com } p_0 = 0,125, \\ 7, & \text{com } p_1 = 0,125, \\ 11, & \text{com } p_2 = 0,375, \\ 25, & \text{com } p_3 = 0,375. \end{cases}$$

Solução:

Conforme apresentado anteriormente, deve-se primeiramente gerar $U \sim \text{Uniforme}(0, 1)$ e escolher X conforme o intervalo em que U pertencer. Dessa maneira, tem-se que:

$$X = \begin{cases} 2, & \text{se } u < 0,125, \\ 7, & \text{se } 0,125 \leq u < 0,250, \\ 11, & \text{se } 0,250 \leq u < 0,625, \\ 25, & \text{se } 0,625 \leq u < 1,000. \end{cases}$$

EXEMPLO 3.4 (CAS/SOA) Com o objetivo de simular a distribuição do sinistro agregado, considere que:

- o número de sinistros, N , tem distribuição binomial com $m = 3$ e média igual a 1,8;
- o valor do sinistro, X , tem distribuição uniforme discreta, podendo assumir os seguintes valores: 1, 2, 3, 4 ou 5;
- os valores dos sinistros são independentes entre si e independentes do número de sinistros;
- os números aleatórios gerados de uma distribuição $\text{Uniforme}(0, 1)$ em sequência são: 0,7, 0,1, 0,3, 0,1, 0,9, 0,5, 0,5, 0,7, 0,3 e 0,1;
- ao simular o valor de N , obtém-se os valores de X , como por exemplo, se $N = 3$, tem-se X_1 , X_2 e X_3 , e o sinistro agregado é dado por $X_1 + X_2 + X_3$.

Calcule o sinistro agregado associado ao terceiro valor simulado de N .

Solução:

Como N tem distribuição binomial com $m = 3$ e média igual a 1,8, ou seja, $E(N) = mp = 1,8$, a probabilidade de ocorrência de sinistro, p , é igual a 0,6. Dessa maneira, a distribuição de N é:

n	0	1	2	3
$P_N(n)$	0,064	0,288	0,432	0,216
$F_N(n)$	0,064	0,352	0,784	1,000

Para gerar um valor de N , deve-se gerar $U \sim \text{Uniforme}(0, 1)$ e escolher N conforme o intervalo em que U pertencer. Dessa maneira, tem-se que:

$$N = \begin{cases} 0, & \text{se } u < 0,064, \\ 1, & \text{se } 0,064 \leq u < 0,352, \\ 2, & \text{se } 0,352 \leq u < 0,784, \\ 3, & \text{se } 0,784 \leq u < 1,000. \end{cases}$$

Como X tem distribuição uniforme discreta, podendo assumir os valores 1, 2, 3, 4 ou 5, tem-se que:

x	1	2	3	4	5
$P_X(x)$	0,2	0,2	0,2	0,2	0,2
$F_X(x)$	0,2	0,4	0,6	0,8	1,0

Para gerar um valor de X , deve-se também gerar $U \sim \text{Uniforme}(0, 1)$ e escolher X conforme o intervalo em que U pertencer, obtendo:

$$X = \begin{cases} 1, & \text{se } u < 0, 2, \\ 2, & \text{se } 0, 2 \leq u < 0, 4, \\ 3, & \text{se } 0, 4 \leq u < 0, 6, \\ 4, & \text{se } 0, 6 \leq u < 0, 8, \\ 5, & \text{se } 0, 8 \leq u < 1, 0. \end{cases}$$

Considerando a sequência de números aleatórios gerados de uma distribuição $\text{Uniforme}(0, 1)$, tem-se os seguintes resultados para o sinistro agregado:

Simulação		n		x	Sinistro agregado
1	$u_1 = 0, 7$	2	$u_2 = 0, 1$	1	1+2=3
			$u_3 = 0, 3$	2	
2	$u_4 = 0, 1$	1	$u_5 = 0, 9$	5	5
3	$u_6 = 0, 5$	2	$u_7 = 0, 5$	3	3+4=7
			$u_8 = 0, 7$	4	

Na simulação 1, para $u_1 = 0, 7$, N é igual a 2, resultando em $x = 1$ para $u_2 = 0, 1$ e $x = 2$ para $u_3 = 0, 3$. Assim, o sinistro agregado é igual a 3. Na simulação 2, para $u_4 = 0, 1$, N é igual a 1, resultando em $x = 5$. O sinistro agregado associado ao terceiro valor simulado de N ($N = 2$) é igual a 7, onde $x = 3$ refere-se a $u_7 = 0, 5$ e $x = 4$ refere-se a $u_8 = 0, 7$.

EXEMPLO 3.5 (CAS/SOA) Considere que N representa o número de acidentes em um período, sendo distribuído com função de probabilidade

$$P_N(n) = 0,9(0,1)^{n-1}, n = 1, 2, \dots$$

Considere ainda que X_i representa o valor do sinistro referente ao i -ésimo acidente, e são independentes e identicamente distribuídas (i.i.d) com função densidade de probabilidade

$$f_X(x) = 0,01e^{-0,01x}, x > 0.$$

Suponha que U e V_1, V_2, \dots são variáveis independentes $Uniforme(0, 1)$. U é usado para simular N e V_i é usado para simular o número requerido de X_i , todos usando o método da transformação integral.

Considere os seguintes valores gerados: $u = 0,05$, $v_1 = 0,3$, $v_2 = 0,22$, $v_3 = 0,52$ e $v_4 = 0,46$. Determine o valor total dos sinistros (sinistro agregado) simulados, ou seja, $X_1 + X_2 + \dots + X_N$, para o período.

Solução:

Sabendo que $P_N(n) = 0,9(0,1)^{n-1}$ e $F_N(n) = P(N \leq n)$, tem-se que:

$$N = \begin{cases} 1, u < F_N(1), \\ 2, F_N(1) \leq u < F_N(2), \\ 3, F_N(2) \leq u < F_N(3), \\ \vdots \end{cases}$$

Encontrando $F_N(1) = P_N(1) = 0,9$, tem-se que $u = 0,05 < 0,9$ e $n = 1$, ou seja, para o valor gerado de u , tem-se apenas um sinistro.

Pelo método da transformação integral, considerando $v = 0,3$, referente a um sinistro, obtém-se:

$$\begin{aligned} F_X(x) &= \int_0^x 0,01e^{-0,01s} ds = [-e^{-0,01s}]_0^x = -e^{-0,01x} + 1, \\ v &= -e^{-0,01x} + 1, \\ x &= -\frac{\ln(1-v)}{0,01}, \\ x &= -\frac{\ln(1-0,3)}{0,01} = 35,6675. \end{aligned}$$

EXEMPLO COMPUTACIONAL 3.1

Considere que um modelo de perda em seguro é dado pela função de distribuição

$$F_X(x) = 1 - \frac{1+\alpha}{1+\alpha e^{\frac{x}{\beta}}}, x > 0,$$

em que α e β são parâmetros positivos. Escreva um programa que gere m simulações de X .

Solução:

Pelo método da transformação integral, tem-se que:

$$\begin{aligned} F_X(x) &= 1 - \frac{1+\alpha}{1+\alpha e^{\frac{x}{\beta}}}, \\ u &= 1 - \frac{1+\alpha}{1+\alpha e^{\frac{x}{\beta}}}, \\ (1-u)(1+\alpha e^{\frac{x}{\beta}}) &= 1+\alpha, \\ x &= \beta \ln\left(\frac{\alpha+u}{\alpha(1-u)}\right), \end{aligned}$$

em que u vem da distribuição $Uniforme(0, 1)$.

A seguir, são apresentados os comandos para gerar m simulações de X . No R, existem funções específicas para gerar números aleatórios. No caso da distribuição uniforme, usa-se **runif(m,a,b)**, em que **m** é o tamanho da amostra, **a** e **b** são os parâmetros da distribuição uniforme, correspondentes ao valor mínimo e valor máximo, respectivamente.

```
simulaX<-function(m,alfa,beta)
{
  set.seed(123)
  u<-runif(m)  #a=0 e b=1.
  x<-beta*log((alfa+u)/(alfa*(1-u)))
  return(x)
}

# 5 valores simulados de X para alfa=1 e beta=2.

simulaX(5,1,2)

## [1] 1.183693 4.267755 1.737528 5.557211 6.968316
```

EXEMPLO COMPUTACIONAL 3.2

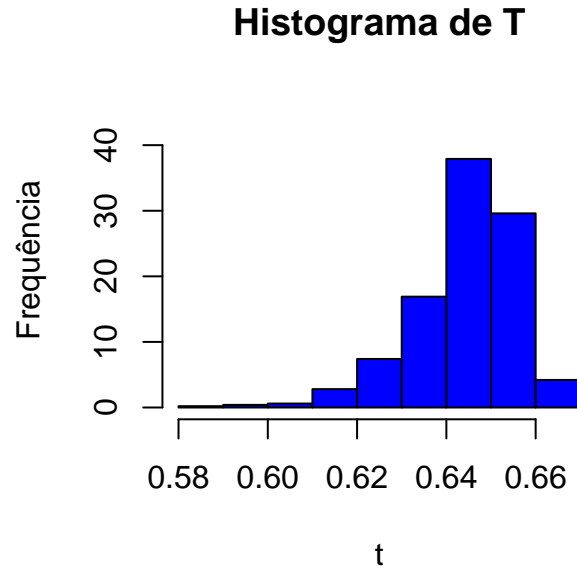
Suponha que $T = \mu e^{-\sigma^2/2 + \sigma\varepsilon}$ em que $\varepsilon \sim Normal(0, 1)$. Obtenha $m = 1000$ simulações quando $\mu = 0,7$ e $\sigma = 0,4$, encontre sua média e apresente o histograma dos valores simulados de T .

Solução:

No caso da distribuição normal, usa-se **rnorm(m,a,b)**, em que **m** é o tamanho da amostra, **a** e **b** são os parâmetros da distribuição normal, correspondentes a média e desvio padrão, respectivamente.

```
simulaT<-function(m,mi,sigma)
{
  set.seed(123)
  eps<-rnorm(m) #a=0 e b=1.
  t<-mi*exp((-sigma^2)/(2+sigma*eps))
  media<-mean(t)
  hist(t,main="Histograma de T",
  ylim=c(0,45),xlab="t",ylab="Frequência",col="blue",
  freq=FALSE)
  return(media)
}

simulaT(1000,0.7,0.4)
```



```
## [1] 0.6442081
```

EXEMPLO COMPUTACIONAL 3.3

Referente a cada perda individual Z , num resseguro não proporcional, a seguradora paga no máximo o valor a , chamado de prioridade, dedutível ou nível de retenção, enquanto cede à resseguradora o valor $Z^{RE} = \max(Z - a, 0)$. Considere que o prêmio de resseguro (Π^{RE}) é igual a perda agregada $S^{RE} = \sum_{i=1}^N Z_i^{RE} = E(Z^{RE})E(N)$, em que $Z \sim \text{Pareto}(\alpha, \theta)$ e o número de sinistros N segue uma distribuição de Poisson com parâmetro λ . Encontre o prêmio de resseguro para $a = 4$, quando $\lambda = 10$, $\alpha = 2$ e $\theta = 1$, usando $m = 10000$ simulações.

Solução:

Como $Z \sim \text{Pareto}(\alpha, \theta)$, tem-se pelo método da transformação integral:

$$\begin{aligned} F_X(x) &= 1 - \left(\frac{\theta}{\theta+x}\right)^\alpha, \\ u &= 1 - \left(\frac{\theta}{\theta+x}\right)^\alpha, \\ (1-u)^{\frac{1}{\alpha}} &= \frac{\theta}{\theta+x}, \\ x &= \theta \left(\frac{1}{(1-u)^{\frac{1}{\alpha}}} - 1 \right), \end{aligned}$$

em que u vem da distribuição $\text{Uniforme}(0, 1)$.

Como $N \sim \text{Poisson}(\lambda)$, tem-se que $E(N) = \lambda$. A seguir, são apresentados os comandos para encontrar o prêmio de resseguro, sendo que as m simulações são utilizadas para estimar a média de Z^{RE} .

```
premioRE<-function(m,lambda,alfa,teta,a){
  set.seed(123)
  u<-runif(m)
  z<-teta*((1/(1-u)^(1/alfa))-1)
  zre<-pmax(z-a,0)
  premio<-lambda*mean(zre)
  return(premio)
}

premioRE(10000,10,2,1,4)

## [1] 1.921732
```

3.2 Método de Aceitação-Rejeição

O método de transformação integral é considerado um tipo de método direto, pois utiliza diretamente a função de distribuição acumulada. O método de aceitação-rejeição é um método indireto, podendo ser aplicado quando os métodos diretos falham, como no caso em que a função de distribuição acumulada não pode ser obtida analiticamente.

Considere $f_X(x)$ como uma f.d.p de X , onde $F_X(x)$ não é facilmente invertida, e uma f.d.p $g_X(x)$, com maior facilidade de manipulação. Assuma que existe uma constante M tal que $Mg_X(x) \geq f_X(x)$ para $x \in [a, b]$. Como $g_X(x)$ pode gerar valores impossíveis de X , deve-se verificar quais valores de X são aceitos e quais são rejeitados. Para tanto, considere que $t(x) = \frac{f_X(x)}{Mg_X(x)}$, sendo que se $t(x) \approx 1$ ($f_X(x) \approx Mg_X(x)$), aceita-se x e, $t(x) \approx 0$, rejeita-se x .

Considere o lançamento de uma moeda com probabilidade de sair cara igual a $t(x) = \frac{f_X(x)}{Mg_X(x)}$. Para um determinado valor de x , se jogar a moeda e sair cara, tal valor de x será considerado como valor vindo de $f_X(x)$. Dessa maneira, tem-se a variável aleatória Z que segue uma distribuição Bernoulli com probabilidade de sucesso igual a $p = t(x)$, sendo que $P(Z = 1) = p$ (probabilidade de sair cara) e $P(Z = 0) = 1 - p$. Considerando que $U \sim Uniforme(0, 1)$, tem-se:

$$P(U \in (0, p]) = \int_0^p dy = p = P(Z = 1).$$

Assim, considera-se que $z = 0$, se $u > p$ e $z = 1$, se $u \leq p$, pois $P(Z = 1) = P(U \in (0, p]) = p$. Dessa maneira, se $u \leq p = t(x) = \frac{f_X(x)}{Mg_X(x)}$, aceita-se x como vindo de $f_X(x)$.

O algoritmo pode ser simplificado nos passos a seguir:

1. Gerar um número y com distribuição $g_Y(y)$;
2. Gerar um número u com distribuição $Uniforme(0, 1)$;
3. Se $u \leq \frac{f_Y(y)}{Mg_Y(y)}$, aceitar x e $x = y$, caso contrário, rejeitar y e retornar ao passo 1.

Pode-se observar que a probabilidade de aceitação, dado $Y = y$, é dada por:

$$P\left(U \leq \frac{f_Y(y)}{Mg_Y(y)} | Y = y\right) = \int_0^{\frac{f_Y(y)}{Mg_Y(y)}} du = \frac{f_Y(y)}{Mg_Y(y)}.$$

A probabilidade de aceitação, independente do valor simulado de Y , é dada por:

$$\begin{aligned} P\left(U \leq \frac{f_Y(y)}{Mg_Y(y)}\right) &= \int_{-\infty}^{\infty} P\left(U \leq \frac{f_Y(y)}{Mg_Y(y)} | Y = y\right) g_Y(y) dy = \\ &= \int_{-\infty}^{\infty} \frac{f_Y(y)}{Mg_Y(y)} g_Y(y) dy = \frac{1}{M} \int_{-\infty}^{\infty} f_Y(y) dy = \frac{1}{M}. \end{aligned}$$

Ainda pode ser demonstrada a função de distribuição acumulada de Y , dado $U \leq \frac{f_Y(y)}{Mg_Y(y)}$, sendo:

$$P\left(Y \leq y | U \leq \frac{f_Y(y)}{Mg_Y(y)}\right) = \frac{P\left(Y \leq y, U \leq \frac{f_Y(y)}{Mg_Y(y)}\right)}{P\left(U \leq \frac{f_Y(y)}{Mg_Y(y)}\right)} = \frac{\int_{-\infty}^y \frac{f_Y(y)}{Mg_Y(y)} g_Y(y) dy}{\frac{1}{M}} =$$

$$= \int_{-\infty}^y f_Y(y) dy = F_Y(y).$$

Considerando que N represente o número de iterações necessárias antes de obter o primeiro valor aceito, tem-se que N segue uma distribuição geométrica com parâmetro $p = P\left(U \leq \frac{f_Y(y)}{Mg_Y(y)}\right) = \frac{1}{M}$. Sabendo que $E(N) = \frac{1}{p}$, tem-se que, em média, o número de iterações necessárias para obter o primeiro valor aceito é M . De todos os valores possíveis para M , deve-se escolher o menor valor, para se ter eficiência computacional.

EXEMPLO COMPUTACIONAL 3.4

Apresente o algoritmo de aceitação-rejeição para gerar uma variável aleatória X que possui a seguinte f.d.p

$$f_X(x) = \frac{3}{8}x^2, 0 \leq x \leq 2.$$

Para tanto, considere $g_X(x) = \frac{1}{2}$, que representa a f.d.p de uma variável aleatória que segue a distribuição uniforme no intervalo $[0, 2]$.

Solução:

Como $Mg_X(x) \geq f_X(x)$ para $x \in [0, 2]$, o valor mínimo de M deve ser igual a 3, pois o valor máximo de $f_X(x)$ é 1,5, ou seja,

$$Mg_X(x) \geq f_X(x),$$

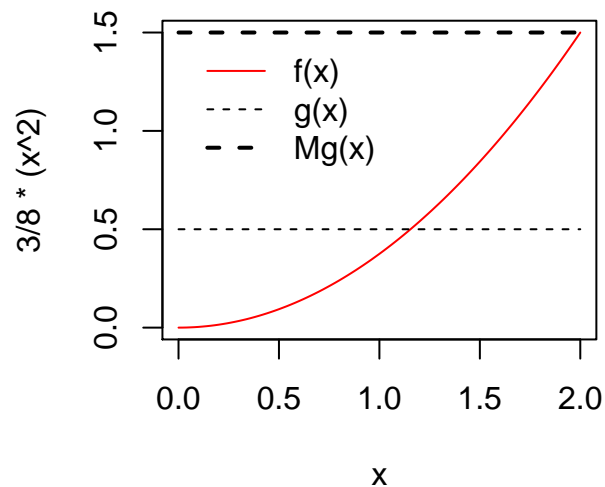
$$\frac{M}{2} \geq 1,5,$$

$$M \geq 3.$$

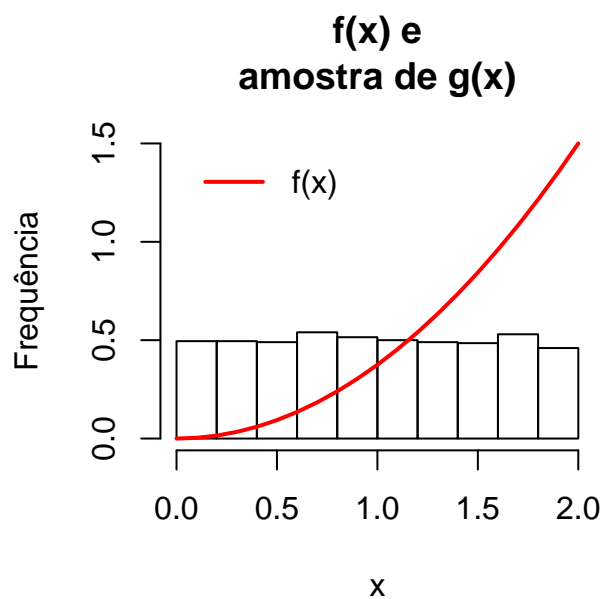
Considerando $M = 3$, o algoritmo pode ser escrito da seguinte maneira:

1. Gerar um número y com distribuição $g_Y(y) = \frac{1}{2}$, $Y \sim Uniforme(0, 2)$;
2. Gerar um número u com distribuição $Uniforme(0, 1)$;
3. Se $u \leq \frac{\frac{3}{8}y^2}{3(\frac{1}{2})} = \frac{1}{4}y^2$, aceitar x , caso contrário, rejeitar x e retornar ao passo 1.

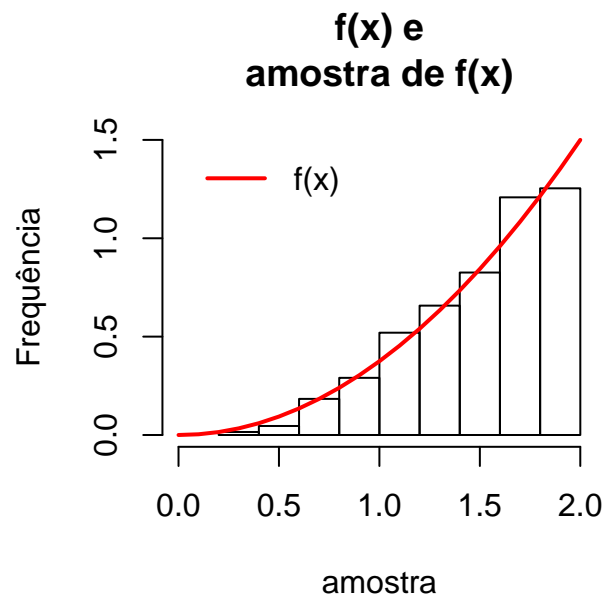
```
curve(3/8* (x^2),0,2,col=2)
curve(1/2+0*x,add=TRUE,lty=2)
curve(3*1/2+0*x,add=TRUE,lty=2,lwd=2)
legend(0,1.5,legend=c("f(x)", "g(x)", "Mg(x)"),lty=c(1, 2, 2),
col=c(2, 1, 1),lwd=c(1, 1, 2),bty="n")
```



```
set.seed(123)
M<-3
nsim<-1000
x<-runif(nsim,0,2)
amostra<-x[runif(nsim)<=(3/8)*x^(2)/(M*dunif(x,0,2))]
x1<-seq(0,2,by=0.1)
y1<-(3/8)*x1^(2)
hist(x,prob=T,xlim=c(0,2),ylim=c(0,1.5),main="f(x) e
amostra de g(x)",ylab="Frequência")
lines(x1,y1,lty=1,col=2,lwd=2)
legend(0,1.5,legend = c("f(x)"),lty=1,col=2,lwd=2,bty="n")
```



```
hist(amostra,prob=T,xlim=c(0,2),ylim=c(0,1.5),main="f(x) e  
amostra de f(x)",ylab="Frequência")  
lines(x1,y1,lty=1,col=2,lwd=2)  
legend(0,1.5,legend = c("f(x)"),lty=1,col=2,lwd=2,bty="n")
```



Capítulo 4

Métodos Monte Carlo em Inferência Estatística

4.1 Distribuição amostral da média

A inferência estatística tem como objetivo generalizar informações sobre uma população com base nos dados de uma amostra. Com base numa amostra de tamanho n de uma variável aleatória X , procura-se estimar o parâmetro desconhecido de uma população. Parâmetros são funções de valores populacionais e estatísticas são funções de valores amostrais, não dependendo de parâmetros desconhecidos. Uma estatística bastante conhecida é a média amostral $\bar{X} = \sum_{i=1}^n \frac{X_i}{n}$. Observe que para cada amostra obtida, tem-se um valor da média amostral. Considerando todas as amostras retiradas de uma população conhecida e obtido o valor da média amostral para cada amostra, obtém-se a distribuição amostral de \bar{X} , que é considerada como uma nova população.

De acordo com o Teorema Central do Limite (TCL), para amostras aleatórias

simples (X_1, X_2, \dots, X_n) retiradas de uma população com média μ e variância σ^2 finita, a distribuição amostral de \bar{X} aproxima-se, para n grande, de uma distribuição normal, com média μ e variância $\frac{\sigma^2}{n}$.

EXEMPLO COMPUTACIONAL 4.1

Sabendo que $X \sim \text{Exponencial}(\lambda)$ e que $\mu = E(X) = \frac{1}{\lambda}$ e $\sigma^2 = \text{Var}(X) = \frac{1}{\lambda^2}$, demonstre utilizando simulação que à medida que o tamanho da amostra cresce, a distribuição de \bar{X} aproxima-se de uma distribuição normal com média igual a μ e variância igual a $\frac{\sigma^2}{n}$.

Solução:

Para realizar a simulação, considere $\lambda = \frac{1}{5}$ e os seguintes tamanhos de amostra: $n = 1, 2, 4, 8, 12, 18, 30$ e 40 . Para cada tamanho de amostra, obtenha 250 valores da média amostral que resultará nas distribuições amostrais de \bar{X} .

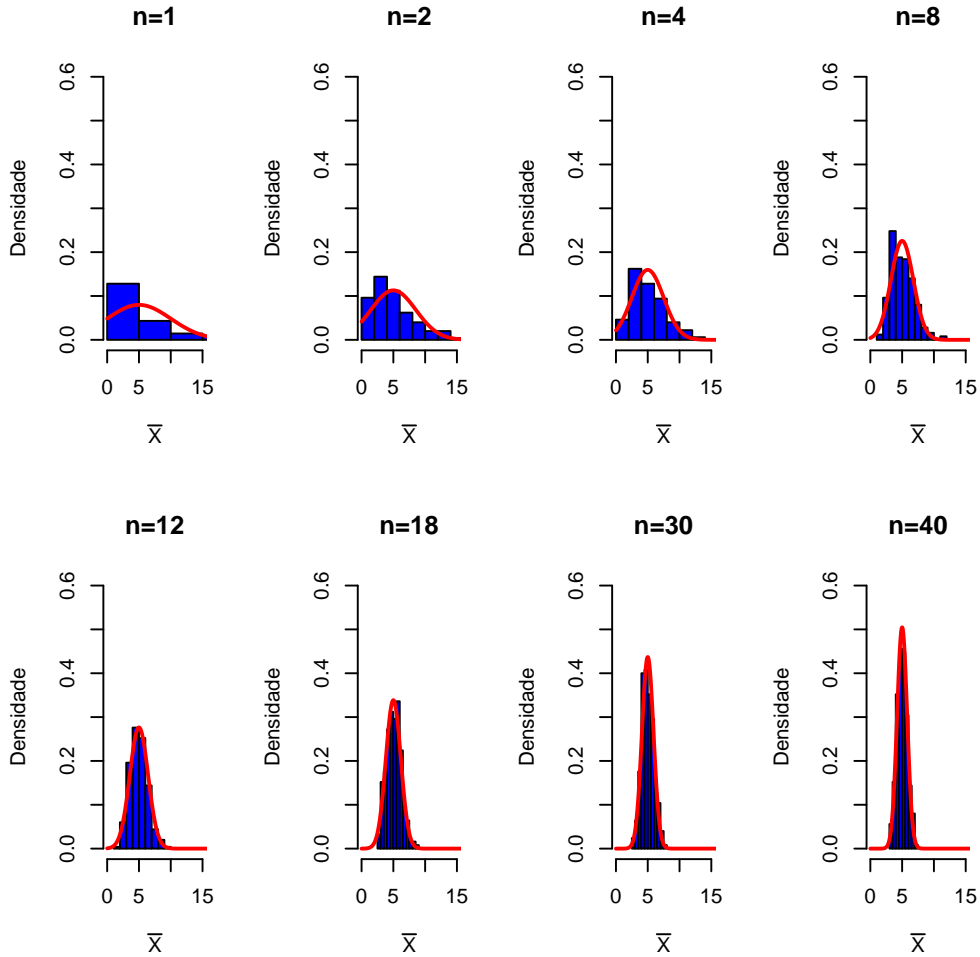
Como $X \sim \text{Exponencial}(\frac{1}{5})$, $\mu = E(X) = 5$ e $\sigma^2 = \text{Var}(X) = 25$, tem-se que a distribuição de \bar{X} aproxima-se de uma distribuição normal com média igual a $\mu = 5$ e variância igual a $\frac{\sigma^2}{n} = \frac{25}{n}$ à medida que o tamanho da amostra cresce. Os comandos a seguir demonstram tal resultado.

```
tcl<-function(n,N,lambda,titulo=" ")
{
  medias<-numeric(N)
  for (i in 1:N)
  {
    x<-rexp(n,lambda)
```



```
medias[i]<-mean(x)
}
hist(medias,freq=F,xlab=expression(bar(X)),ylab="Densidade",
xlim=c(0,15),ylim=c(0,0.6),col="blue",main=titulo)
mexp<-1/lambda #média de X
mmedia<-1/lambda #média da média amostral
vexp<-1/(lambda^2) #variância de X
smedia<-sqrt(vexp/n) #variância da média amostral
xx<-seq(0,20,0.02)
medianormal<-dnorm(xx,mmedia,smedia) #distribuição da
#média amostral
points(xx,medianormal,type="l",lwd=2,col="red")
}

par(mfrow=c(2,4)) #múltiplos gráficos numa mesma página
tcl(1,250,0.2,titulo="n=1") #amostra de tamanho 1
tcl(2,250,0.2,titulo="n=2") #amostra de tamanho 2
tcl(4,250,0.2,titulo="n=4") #amostra de tamanho 4
tcl(8,250,0.2,titulo="n=8") #amostra de tamanho 8
tcl(12,250,0.2,titulo="n=12") #amostra de tamanho 12
tcl(18,250,0.2,titulo="n=18") #amostra de tamanho 18
tcl(30,250,0.2,titulo="n=30") #amostra de tamanho 30
tcl(40,250,0.2,titulo="n=40") #amostra de tamanho 40
```



Observe que com o aumento de n , a variância de \bar{X} vai diminuindo e os histogramas tendem a se concentrar cada vez mais em torno da média de X , se aproximando de uma distribuição normal.

4.2 Estimadores

Considere uma amostra (X_1, X_2, \dots, X_n) de uma variável aleatória vinda de uma população que contém um parâmetro θ desconhecido. Por meio

de tal amostra, pode-se estimar θ , que pode ser, por exemplo, a média populacional ($\mu = E(X)$) ou a variância populacional ($\sigma^2 = Var(X)$). Para estimar θ , utiliza-se um estimador que é considerado como qualquer função das observações da amostra. Por exemplo, a média amostral $\bar{X} = \sum_{i=1}^n \frac{X_i}{n}$ é um estimador associado a média populacional μ . Com base em uma amostra específica, encontra-se o valor assumido pelo estimador. Tal valor é denominado de estimativa. O objetivo é encontrar um estimador $\hat{\theta}$ que seja próximo de θ , conforme algumas características para que seja considerado um bom estimador.

Uma característica é que o estimador $\hat{\theta}$ seja não viesado para θ . Para tanto, deve-se verificar se $E(\hat{\theta}) = \theta$, para todo θ . Caso não seja verificada tal igualdade, $\hat{\theta}$ é considerado viesado e a diferença $E(\hat{\theta}) - \theta$ é o viés de $\hat{\theta}$. Uma outra característica é que $\hat{\theta}$ seja consistente. Para tanto, considere que $\hat{\theta}$ seja calculado para diferentes tamanhos de amostra ($n = 1, 2, 3, \dots$) e obtem-se uma sequência de estimadores $\hat{\theta}_n$. À medida que n aumenta, a distribuição de $\hat{\theta}_n$ deve se tornar mais concentrada ao redor de θ . Com isso, uma sequência $\hat{\theta}_n$ de estimadores de θ é consistente se $\lim_{n \rightarrow \infty} E(\hat{\theta}_n) = \theta$ e $\lim_{n \rightarrow \infty} Var(\hat{\theta}_n) = 0$. Além de verificar se um estimador é não viesado e consistente, pode-se verificar se um determinado estimador é mais eficiente do que outro. Suponha que $\hat{\theta}_1$ e $\hat{\theta}_2$ são estimadores não viesados de θ . Se $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$, então $\hat{\theta}_1$ é mais eficiente do que $\hat{\theta}_2$.

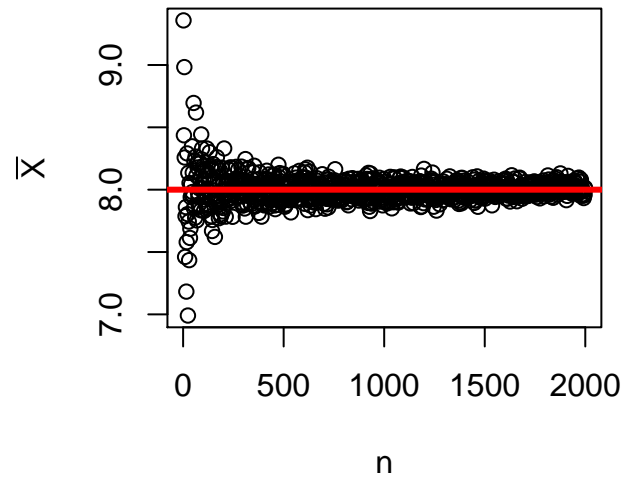
EXEMPLO COMPUTACIONAL 4.2

Considere que $X \sim Normal(\mu, \sigma^2)$. Utilizando simulação, avalie se o estimador $\bar{X} = \sum_{i=1}^n \frac{X_i}{n}$ de μ é consistente.

Solução:

Para verificar se \bar{X} é um estimador consistente para μ , é necessário verificar se à medida que o tamanho da amostra n aumenta, \bar{X} se aproxima de μ . Para tanto, considere $\mu = 8$ e $\sigma^2 = 4$. Os comandos a seguir dão origem a um gráfico que mostra o comportamento de \bar{X} para diferentes tamanhos de amostra.

```
#sequência de tamanhos de amostra, com incremento igual a 2
n<-seq(2,2000,by=2)
xbarra<-numeric(length(n))
for(i in 1:length(n)){
  amostrax<-rnorm(n[i],8,2)
  xbarra[i]<-mean(amostrax)
}
plot(n,xbarra,ylab=expression(bar(X)))
#linha horizontal referente ao valor da média populacional
abline(h=8,col="red",lwd=3)
```



Pelo gráfico obtido, pode-se observar que à medida que o tamanho da amostra n aumenta, \bar{X} se aproxima de $\mu = 8$.

4.2.1 Estimadores de máxima verossimilhança

Considere uma sequência X_1, X_2, \dots, X_n de n variáveis aleatórias independentes e identicamente distribuídas com f.d.p ou f.p, $f(x|\theta)$, que depende do parâmetro θ desconhecido. Tal sequência é dita ser uma amostra aleatória de tamanho n da distribuição de X , sendo que a f.d.p ou f.p conjunta é dada por:

$$f(x_1, \dots, x_n|\theta) = \prod_{i=1}^n f(x_i|\theta) = f(x_1|\theta) \cdot f(x_2|\theta) \cdot \dots \cdot f(x_n|\theta).$$

A função de verossimilhança é definida por $L(\theta; x_1, \dots, x_n) = \prod_{i=1}^n f(x_i|\theta)$, que deve ser vista como uma função de possíveis valores de θ . Com base em $L(\theta; x_1, \dots, x_n)$, obtém-se o estimador de máxima verossimilhança (EMV) de θ , que é o valor de θ que maximiza $L(\theta; x_1, \dots, x_n)$. Aqui, o EMV de θ será denotado por $\hat{\theta}_{EMV}$, que representa o valor que considera a amostra observada como a mais verossímil de ocorrer, trazendo assim a melhor informação sobre o parâmetro θ .

EXEMPLO 4.1 (CAS/SOA)

Considere que o objetivo é estimar q_x , que representa a probabilidade de um indivíduo com idade x morrer antes de completar $x+1$ anos. Considere que o tempo de vida de um indivíduo com idade x tem função densidade de probabilidade constante. Em um estudo de mortalidade, 10 indivíduos com idades x foram observados. Dentre eles, 8 sobreviveram o ano completo, um indivíduo morreu antes de completar 1 ano e outro indivíduo foi retirado do estudo na idade $x+0,5$. Determine a estimativa de máxima verossimilhança de q_x .

Solução: O tempo de vida de um indivíduo com idade x é modelado de acordo com a f.d.p $f(t) = c$. Com relação aos 8 indivíduos que sobreviveram o ano completo, a contribuição para a verossimilhança é:

$$P(T > 1) = 1 - P(T \leq 1) = 1 - \int_0^1 c dt = 1 - c.$$

A contribuição para a verossilhança do indivíduo que foi retirado do estudo na idade $x + 0,5$ é dada por:

$$P(T > 0,5) = 1 - P(T \leq 0,5) = 1 - \int_{0,5}^1 c dt = 1 - 0,5c.$$

A contribuição para a verossimilhança do indivíduo que morreu antes de completar 1 ano é dada por:

$$P(T \leq 1) = \int_0^1 c dt = c.$$

Dessa maneira, a função de verossimilhança é obtida, fazendo:

$$L(c; t_1, \dots, t_{10}) = \prod_{i=1}^{10} f(t_i|c) = (1-c)^8(1-0,5c)c.$$

O valor máximo de $L(c; t_1, \dots, t_{10})$ é igual ao valor máximo de $l(c; t_1, \dots, t_{10}) = \ln(L(c; t_1, \dots, t_{10}))$, chamada de log-verossimilhança, na qual é obtida a seguir.

$$l(c; t_1, \dots, t_{10}) = \ln(L(c; t_1, \dots, t_{10})) = 8\ln(1-c) + \ln(1-0,5c) + \ln c.$$

Derivando e igualando a zero, obtém-se:

$$\frac{dl(c; t_1, \dots, t_{10})}{dc} = -\frac{8}{1-c} - \frac{0,5}{1-0,5c} + \frac{1}{c} = 0,$$

$$c = 0,10557.$$

A estimativa de máxima verossimilhança para o parâmetro c é $0,10557 = \hat{q}_x$.

EXEMPLO 4.2 (CAS/SOA)

Uma amostra de 10 valores de sinistros obtidos de uma distribuição gama é apresentada a seguir:

1500 6000 3500 3800 1800 5500 4800 4200 3900 3000

Suponha que $\alpha = 12$. Determine a estimativa de máxima verossimilhança de β .

Solução: O valor do sinistro X segue uma distribuição gama com parâmetros α e β ($X \sim \text{Gama}(\alpha, \beta)$), com $E(X) = \alpha\beta$ e $\text{Var}(X) = \alpha\beta^2$, sendo que sua f.d.p é dada da seguinte maneira:

$$f(x|\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}.$$

em que $\Gamma(\alpha)$ é a função gama, sendo $\Gamma(\alpha) = (\alpha - 1)!$.

A função de verossimilhança é obtida, fazendo:

$$L(\alpha, \beta; x_1, \dots, x_n) = \prod_{i=1}^n \frac{1}{\Gamma(\alpha)\beta^\alpha} x_i^{\alpha-1} e^{-\frac{x_i}{\beta}},$$

$$L(\alpha, \beta; x_1, \dots, x_n) = \frac{1}{[\Gamma(\alpha)]^n \beta^{\alpha n}} (\prod_{i=1}^n x_i)^{\alpha-1} e^{-\frac{\sum_{i=1}^n x_i}{\beta}}.$$

Encontrando $l(\alpha, \beta; x_1, \dots, x_n) = \ln(L(\alpha, \beta; x_1, \dots, x_n))$, tem-se que:

$$l(\alpha, \beta; x_1, \dots, x_n) = -n(\ln \Gamma(\alpha)) - \alpha n(\ln \beta) + (\alpha - 1) \sum_{i=1}^n \ln x_i - \frac{\sum_{i=1}^n x_i}{\beta}.$$

Considerando a amostra de 10 valores de sinistros, obtém-se:

$$l(\alpha, \beta; 1500, \dots, 3000) = -10 \ln \Gamma(\alpha) - 10 \alpha \ln \beta + 81,619(\alpha - 1) - \frac{38000}{\beta}.$$

Derivando em relação a β e igualando a zero, obtém-se a estimativa de máxima verossimilhança de β :

$$\begin{aligned} \frac{dl(\alpha, \beta; 1500, \dots, 3000)}{d\beta} &= 38000\beta^{-2} - 10\alpha\beta^{-1} = 0, \\ \hat{\beta}_{EMV} &= \frac{38000}{120} = 316,67. \end{aligned}$$

EXEMPLO COMPUTACIONAL 4.3

Considere que $X \sim Poisson(\lambda)$. Com base numa amostra aleatória de X , encontre o estimador de máxima verossimilhança para λ .

Solução:

Para encontrar o EMV, é necessário encontrar a função de verossimilhança. Como $X \sim Poisson(\lambda)$, sua f.p é dada por:

$$f(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}.$$

A função de verossimilhança é dada por:

$$L(\lambda; x_1, \dots, x_n) = \prod_{i=1}^n f(x_i|\lambda) = \prod_{i=1}^n \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = \frac{1}{\prod_{i=1}^n x_i!} \lambda^{\sum_{i=1}^n x_i} e^{-n\lambda}.$$

O valor máximo de $L(\lambda; x_1, \dots, x_n)$ é igual ao valor máximo de $l(\lambda; x_1, \dots, x_n) = \ln(L(\lambda; x_1, \dots, x_n))$, chamada de log-verossimilhança, na qual é obtida a seguir.

$$l(\lambda; x_1, \dots, x_n) = \ln(L(\lambda; x_1, \dots, x_n)) = \ln \left(\frac{1}{\prod_{i=1}^n x_i!} \lambda^{\sum_{i=1}^n x_i} e^{-n\lambda} \right),$$

$$l(\lambda; x_1, \dots, x_n) = \ln \left(\frac{1}{\prod_{i=1}^n x_i!} \right) + \sum_{i=1}^n x_i \ln \lambda - n\lambda.$$

Derivando e igualando a zero, obtém-se:

$$\frac{dl(\lambda; x_1, \dots, x_n)}{d\lambda} = \frac{1}{\lambda} \sum_{i=1}^n x_i - n,$$

$$\frac{1}{\hat{\lambda}_{EMV}} \sum_{i=1}^n x_i - n = 0,$$

$$\hat{\lambda}_{EMV} = \frac{\sum_{i=1}^n x_i}{n} = \bar{x}.$$

O EMV para o parâmetro λ é $\hat{\lambda}_{EMV} = \bar{x}$. Utilizando o teste da segunda derivada, é possível verificar que $\hat{\lambda}_{EMV} = \bar{x}$ é um ponto de máximo.

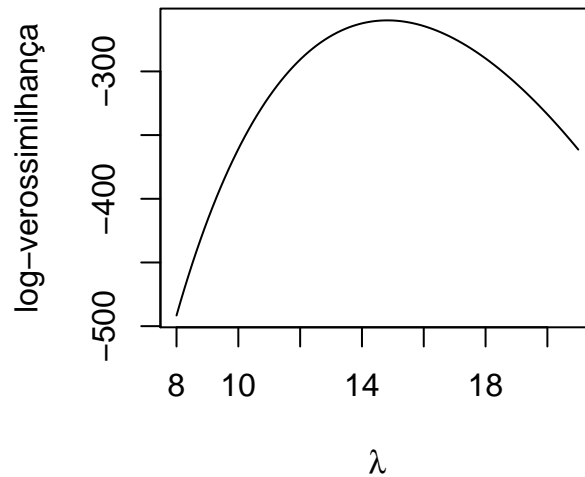
Os comandos a seguir geram o gráfico da função log-verossimilhança para o parâmetro λ . Para tanto, é gerada uma amostra de tamanho 100 de uma distribuição de Poisson com parâmetro $\lambda = 15$.

```
logveros<-function(lambda,x){
  lambda.intervalo<-seq(min(x),max(x),l=101)
  lv<-c()
  for(i in 1:length(lambda.intervalo))
  {
    lv[i]<-sum(dpois(x,lambda.intervalo[i],log=TRUE))
  } #cálculo da função log-verossimilhança
  plot(lambda.intervalo,lv,type="l",xlab=expression(lambda),
        ylab="log-verossimilhança")
  resultado<-data.frame(lambda.intervalo,lv)
  return(resultado[resultado$lv==max(lv),])
}

set.seed(123)
x<-rpois(100,15) #gerando uma amostra de tamanho 100
mean(x)

## [1] 14.82

logveros(15,x)
```



```
##      lambda.intervalo      lv
## 53                14.76 -259.9359
```

O valor de λ que maximiza a função log-verossimilhança é aproximadamente igual a 15 ($\hat{\lambda}_{EMV} = 14,76$), que é aproximadamente igual a média dos dados gerados ($\bar{x} = 14,82$).

4.2.2 Estimadores de mínimos quadrados

Com base numa amostra de pares (X, Y) , pode-se estar interessado em avaliar como Y varia em função de x ($f(x)$). A variável Y é considerada uma variável aleatória, que recebe o nome de variável dependente ou variável resposta. A variável x é uma variável controlada, que recebe o nome de variável independente ou variável explicativa. Como Y é uma variável aleatória, tem-se que $Y = f(x) + \epsilon$, em que ϵ é o erro aleatório, sendo que para

uma amostra de tamanho n , tem-se que $Y_i = f(x_i) + \epsilon_i$ ($i = 1, \dots, n$). Um possível modelo para $f(x_i)$ é o modelo linear $f(x_i) = \beta_0 + \beta_1 x_i$. O modelo $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ é chamado de modelo de regressão linear simples, sendo que ϵ_i segue uma distribuição normal com média zero e variância constante σ_ϵ^2 ($\epsilon_i \stackrel{iid}{\sim} Normal(0, \sigma_\epsilon^2)$). Pelo método de mínimos quadrados, encontram-se valores de β_0 e β_1 que minimizam a soma dos quadrados dos erros (SQE), obtendo os estimadores de mínimos quadrados $\hat{\beta}_0$ e $\hat{\beta}_1$ e o modelo ajustado $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$.

Considerando uma amostra de n pares (x_i, y_i) , a SQE pode ser escrita da seguinte maneira:

$$SQE = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - f(x_i))^2.$$

Aqui, e_i é a realização da variável aleatória ϵ_i e y_i é a realização da variável aleatória Y_i . Para $f(x_i) = \beta_0 + \beta_1 x_i$, tem-se que:

$$SQE = \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2.$$

Minimizando a SQE em relação aos parâmetros β_0 e β_1 , obtém-se:

$$\frac{\partial SQE}{\partial \beta_0} = - \sum_{i=1}^n y_i + n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n x_i = 0,$$

$$\frac{\partial SQE}{\partial \beta_1} = - \sum_{i=1}^n x_i y_i + \hat{\beta}_0 \sum_{i=1}^n x_i + \hat{\beta}_1 \sum_{i=1}^n x_i^2 = 0.$$

Dessa maneira, tem-se os estimadores de mínimos quadrados $\hat{\beta}_0$ e $\hat{\beta}_1$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x},$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2}.$$

EXEMPLO COMPUTACIONAL 4.4

Considere o modelo de regressão linear simples

$$Y_i = 3 + 5x_i + \epsilon_i$$

em que $\epsilon_i \stackrel{iid}{\sim} Normal(0, 2)$, para $i = 1, \dots, 80$. Considere que $X \stackrel{iid}{\sim} Poisson(3)$, sendo que tais valores devem ser considerados fixos. Observe que os verdadeiros valores dos parâmetros são $\beta_0 = 3$ e $\beta_1 = 5$. Com base na amostra, estime os parâmetros β_0 e β_1 utilizando o método de mínimos quadrados.

Solução:

Os comandos a seguir trazem como resultado $\hat{\beta}_0$ e $\hat{\beta}_1$, e o erro com relação aos verdadeiros valores dos parâmetros. Para ajustar o modelo de regressão

linear simples, encontrando as estimativas de mínimos quadrados, foi utilizada a função **lm**.

```
#Gerando a variável x
set.seed(123)
x<-rpois(80,3)

#Gerando o erro e
e<-rnorm(80,0,sqrt(2))

#Encontrando os valores de y
beta0<-3
beta1<-5
y<-beta0+beta1*x+e

#Ajustando o modelo de regressão linear simples
modelo<-lm(y~x)
betas.estimados<-modelo$coefficients
betas.estimados

## (Intercept)          x
##    3.066729    4.977558

#Erro de estimação
erro.estimacao<-betas.estimados-c(beta0,beta1)
erro.estimacao

## (Intercept)          x
```

```
## 0.06672925 -0.02244163
```

Foram encontradas as estimativas $\hat{\beta}_0 = 3,07$ e $\hat{\beta}_1 = 4,98$. Os erros de estimação foram iguais a 0,07 e $-0,02$ referentes aos parâmetros β_0 e β_1 , respectivamente.

4.2.3 Estimação intervalar

Nas seções anteriores, foram apresentados estimadores pontuais, que fornecem um único valor para o estimador ($\hat{\theta}$) do parâmetro (θ). A estimação pontual não apresenta uma medida de precisão quanto a utilizar um único valor para estimar o parâmetro, ou seja, não leva em consideração o erro amostral do estimador. A partir da distribuição amostral do estimador, é possível realizar a estimação intervalar, que consiste em construir intervalos que incluam o verdadeiro valor do parâmetro com certo coeficiente de confiança $(1 - \alpha)$. Tais intervalos são denominados de intervalos de confiança (IC).

Se fosse possível selecionar várias amostras de uma mesma população e para cada uma dessas amostras construir um IC, então em $(1 - \alpha)100\%$ dos ICs construídos iriam conter o verdadeiro valor do parâmetro θ . Os ICs, que são aleatórios, podem ser representados, de maneira geral, pela seguinte expressão, sendo que θ não é uma variável aleatória.

$$P(a < \theta < b) = 1 - \alpha$$

Como exemplo, para a construção de ICs para a média populacional μ , considera-se a distribuição amostral do estimador pontual $\bar{X} = \sum_{i=1}^n \frac{X_i}{n}$. Se for considerada uma população infinita, em que $X \sim Normal(\mu, \sigma^2)$, com σ^2 desconhecida, a variável aleatória $T = \frac{\bar{X} - \mu}{\sqrt{\frac{S^2}{n}}}$ tem distribuição t de *Student* com $v = n - 1$ graus de liberdade. A partir da variável aleatória T e da distribuição t de *Student*, obtém-se uma expressão geral para os ICs para a média populacional μ .

$$\begin{aligned} P(-t_{\frac{\alpha}{2}} < T < t_{\frac{\alpha}{2}}) &= 1 - \alpha, \\ P\left(-t_{\frac{\alpha}{2}} < \frac{\bar{X} - \mu}{\sqrt{\frac{S^2}{n}}} < t_{\frac{\alpha}{2}}\right) &= 1 - \alpha, \\ P\left(-t_{\frac{\alpha}{2}} \sqrt{\frac{S^2}{n}} < \bar{X} - \mu < t_{\frac{\alpha}{2}} \sqrt{\frac{S^2}{n}}\right) &= 1 - \alpha, \\ P\left(\bar{X} - t_{\frac{\alpha}{2}} \sqrt{\frac{S^2}{n}} < \mu < \bar{X} + t_{\frac{\alpha}{2}} \sqrt{\frac{S^2}{n}}\right) &= 1 - \alpha. \end{aligned}$$

Dessa maneira, o IC para a média populacional μ , utilizando o desvio padrão amostral S , é dado pela expressão:

$$IC(\mu; (1 - \alpha)100\%) : \bar{X} \pm t_{\frac{\alpha}{2}} \sqrt{\frac{S^2}{n}},$$

em que $(1 - \alpha)100\%$ é o grau de confiança e $t_{\frac{\alpha}{2}}$ é o quantil de ordem $\frac{\alpha}{2}$ referente a tabela da distribuição t de *Student*, com $v = n - 1$ graus de liberdade.

Pelo Teorema Central do Limite (TCL), considerando uma população qualquer e amostras grandes, tem-se o seguinte IC para μ :

$$IC(\mu; (1 - \alpha)100\%) : \bar{X} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{n}},$$

em que o desvio padrão populacional σ é conhecido e $z_{\frac{\alpha}{2}}$ é o quantil de ordem $\frac{\alpha}{2}$ referente a tabela da distribuição normal padrão.

Como, geralmente, o desvio padrão populacional σ é desconhecido, pode-se substituir σ por S . Como o TCL é utilizado, a amostra será considerada como grande, quando o tamanho amostral for maior ou igual a 30 ($n \geq 30$). Quando a amostra é pequena e a população é aproximadamente normal, pode-se obter intervalos de confiança aproximados considerando a distribuição t de *Student*. Quando a amostra é pequena e a população não é normal, estudos sobre a distribuição da população e distribuições amostrais devem ser realizados.

Também pode-se obter ICs para proporção e variância populacionais, assim como, para diferença dentre médias de duas populações. Tais intervalos não serão abordados no presente texto.

EXEMPLO COMPUTACIONAL 4.5

Uma seguradora considera que o valor de indenização X de um seguro de automóveis segue uma distribuição normal $X \sim Normal(\mu, \sigma^2)$. Com o objetivo de realizar inferência sobre a média populacional, que é desconhecida, e obter um IC, considere os seguintes valores de indenizações (x10000 u.m): 1,7; 2,4; 2,5; 3,1; 4,6; 2,7; 1,5; 3,4; 4,3; 4,1; 1,8; 2,7; 2,4; 6,1; 3,9; 6,5; 3,4; 5,6; 2,6 e 4,3. Encontre o IC para μ , considerando coeficiente de confiança igual a 95%.

Solução:

Como $X \sim Normal(\mu, \sigma^2)$ e a variância σ^2 é desconhecida, o seguinte IC para μ pode ser encontrado:

$$IC(\mu; (1 - \alpha)100\%) : \bar{X} \pm t_{\frac{\alpha}{2}} \sqrt{\frac{S^2}{n}}.$$

Com base na amostra dada, tem-se que $n = 20$, $\bar{X} = 3,48$ e $S^2 = 2,05$. Para coeficiente de confiança igual a 95%, obtem-se:

$$\begin{aligned} IC(\mu; (1 - \alpha)100\%) &: \bar{X} \pm t_{\frac{\alpha}{2}} \sqrt{\frac{S^2}{n}}, \\ IC(\mu; 95\%) &: 3,48 \pm t_{\frac{0,05}{2}} \sqrt{\frac{2,05}{20}}, \\ IC(\mu; 95\%) &: 3,48 \pm t_{0,025} \sqrt{\frac{2,05}{20}}, \\ IC(\mu; 95\%) &: 3,48 \pm 2,09.0,32, \\ IC(\mu; 95\%) &: 3,48 \pm 0,67. \end{aligned}$$

Dessa maneira, com 95% de confiança, pode-se afirmar que o verdadeiro valor médio de indenizações (μ) é um valor entre 28100 u.m. e 41500 u.m. Ou seja, se fossem construídos 100 ICs, 95 deles conteriam o verdadeiro valor médio de indenizações.

Os comandos a seguir fornecem como resultado, IC para μ .

```
ICmedia<-function(x,coef){  
  n<-length(x)  
  mediaX<-mean(x)  
  varianciaX<-var(x)  
  #Quantil da distribuição t através da função qt.  
  quantis<-qt(c((1-coef)/2,1-(1-coef)/2),df=n-1)
```

```
IC<-mediaX+quantis*sqrt(varianciaX/n)
return(IC)
}

x<-c(1.7,2.4,2.5,3.1,4.6,2.7,1.5,3.4,
     4.3,4.1,1.8,2.7,2.4,6.1,3.9,6.5,
     3.4,5.6,2.6,4.3)  #amostra

ICmedia(x,0.95)

## [1] 2.809544 4.150456
```

EXEMPLO COMPUTACIONAL 4.6

Considere que $X \sim Normal(\mu, \sigma^2)$. Encontre uma estimativa Monte Carlo para a proporção de intervalos de confiança empíricos que contém o verdadeiro valor do parâmetro μ . Para tanto, considere $\mu = 2$, $\sigma^2 = 9$, $n = 60$ e 100 simulações Monte Carlo, com coeficiente de confiança teórico igual a 95%.

Solução:

Como $n > 30$, pelo TCL, tem-se que $\bar{X} \approx Normal\left(\mu, \frac{\sigma^2}{n}\right)$, sendo que o seguinte IC para μ pode ser encontrado:

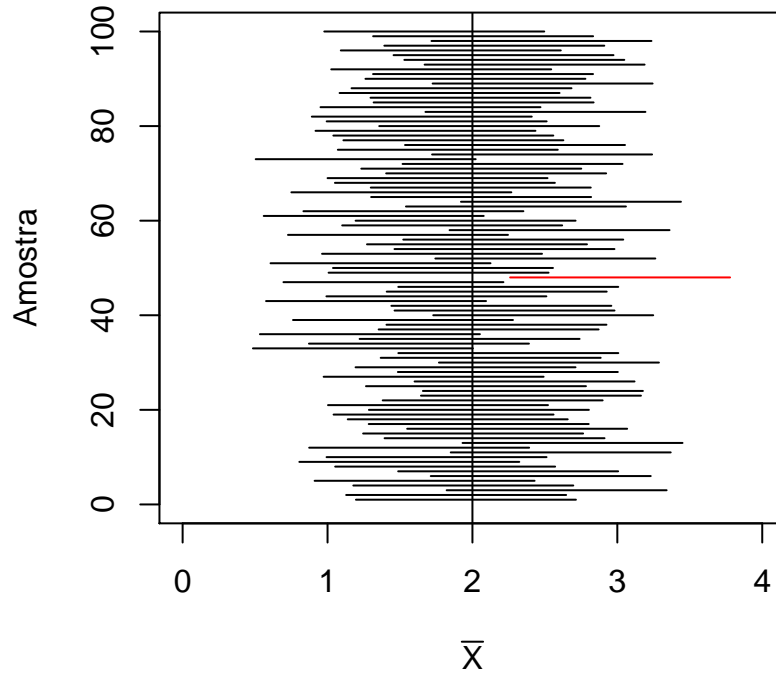
$$IC(\mu; (1 - \alpha)100\%) : \bar{X} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{n}}.$$

Os comandos a seguir fornecem como resultado, os intervalos de confiança empíricos e a proporção de intervalos de confiança que contém o verdadeiro valor do parâmetro μ . O resultado deverá ser próximo ao coeficiente de confiança teórico, que é igual a 95%.

```
IC.proporcao<-function(N,n,mi,sigma,conf)
  #N é número de simulações Monte Carlo
  #n é tamanho da amostra
  #mi é a média de X
  #sigma é o desvio de X
  {
    plot(0,0,type="n",
          xlim=c(0,2*mi),ylim=c(0,N),
          xlab=expression(bar(X)), ylab="Amostra")
    abline(v=mi) #linha vertical em mi
    zt<-qnorm(1-((1-conf)/2)) #valor tabelado
    #pelo TCL, para n grande (n>=30),
    #a média amostral segue
    #uma distribuição que se aproxima da Normal,
    #com média=mi e
    #variância igual a sigma^2/n
    sigma.xbarra<-sigma/sqrt(n)
    t<-c()
    for(i in 1:N){
      #população normal com média mi e
      #variância sigma^2
      x<-rnorm(n,mi,sigma)
      media<-mean(x)
      IClimiteinf<-media-zt*sigma.xbarra
```

```
IClimitesup<-media+zt*sigma.xbarra
plotx<-c(IClimiteinf,IClimitesup)
ploty<-c(i,i)
if(
  IClimiteinf>mi|IClimitesup<mi)lines(
  plotx,ploty,col="red") else (
  lines(plotx,ploty))
if(
  IClimiteinf>mi|IClimitesup<mi
)t[i]=0 else t[i]=1
}
prop<-sum(t)/N
return(prop)
}
```

```
IC.proporcao(100,60,2,3,0.95)
```



```
## [1] 0.99
```

4.2.4 Integração Monte Carlo

Suponha que o objetivo é calcular a integral $I = \int_a^b g(\theta)d\theta$. Para tanto, são realizadas as seguintes manipulações:

$$I = \int_a^b g(\theta)d\theta = \int_a^b (b-a)g(\theta)\frac{1}{b-a}d\theta,$$

$$I = (b-a) \int_a^b g(\theta)\frac{1}{b-a}d\theta,$$

$$I = (b - a) \int_a^b g(\theta) f_{\Theta}(\theta) d\theta,$$

$$I = (b - a) E[g(\Theta)].$$

Observe que agora, θ é visto como a realização da variável aleatória Θ , sendo que $f_{\Theta}(\theta) = \frac{1}{b-a}$ é a f.d.p de uma distribuição uniforme ($\Theta \sim Uniforme(a, b)$), e que $E[g(\Theta)]$ é a esperança da função $g(\Theta)$.

Dessa maneira, tem-se o seguinte algoritmo para resolver I , utilizando simulação Monte Carlo:

1. Gere $\theta_1, \theta_2, \dots, \theta_n$ da distribuição uniforme ($\Theta \sim Uniforme(a, b)$);
2. Obtenha $g(\theta_1), g(\theta_2), \dots, g(\theta_n)$;
3. Encontre $\hat{I} = (b - a) \frac{1}{n} \sum_{i=1}^n g(\theta_i)$.

O estimador \hat{I} é chamado de estimador Monte Carlo de I .

Considerando que $\Theta \sim Uniforme(0, 1)$, tem-se que $I = \int_0^1 g(\theta) d\theta = E[g(\Theta)]$ e $\hat{I} = \frac{1}{n} \sum_{i=1}^n g(\theta_i)$. Pela lei forte dos grandes números, pode-se utilizar simulação Monte Carlo para aproximar $I = E[g(\Theta)]$, através da média aritmética de um grande número de simulações, tendo a seguinte probabilidade para $n \rightarrow \infty$:

$$P \left(\lim_{n \rightarrow \infty} \frac{g(\theta_1) + g(\theta_2) + \dots + g(\theta_n)}{n} = I \right) = 1.$$

Ou seja, a lei forte dos grandes números garante que quanto maior o tamanho da amostra ($n \rightarrow \infty$), \hat{I} converge para I , com probabilidade igual a 1.

Para encontrar $I = \int_a^b g(\theta) d\theta$, pode-se realizar uma mudança de variáveis. Fazendo $y = \frac{\theta-a}{b-a}$, tem-se que $I = \int_a^b g(\theta) d\theta = \int_0^1 g(y(b-a) + a)(b-a) dy$.

EXEMPLO COMPUTACIONAL 4.7

Encontre uma estimativa Monte Carlo de $I = \int_3^9 e^{-x} dx$. Verifique a evolução do erro conforme aumenta o número de simulações.

Solução:

Para encontrar uma estimativa Monte Carlo de I , será implementado no *software* R o seguinte algoritmo:

1. Gere x_1, x_2, \dots, x_n da distribuição uniforme ($X \sim Uniforme(3, 9)$);
2. Obtenha $g(x_1), g(x_2), \dots, g(x_n)$, em que $g(x_i) = e^{-x_i}$;
3. Encontre $\hat{I} = (9 - 3) \frac{1}{n} \sum_{i=1}^n g(x_i) = \frac{6}{n} \sum_{i=1}^n e^{-x_i}$.

```
set.seed(123)
n<-10000
x<-runif(n,3,9) #Gerando n valores de X
i.hat<-6*mean(exp(-x))
print(i.hat) #valor aproximado pela integração Monte Carlo
```

```
## [1] 0.04974616

print(-exp(-9)+exp(-3)) #valor exato da integral

## [1] 0.04966366

#função para utilizar o comando integrate
f<-function(x){exp(-x)}
integrate(f,3,9)

## 0.04966366 with absolute error < 5.5e-16

# função para encontrar uma estimativa de I
est.I<-function(n,a,b){
  x<-runif(n,a,b)
  g<-exp(-x)
  est.I<-(b-a)*mean(g)
  return(est.I)
}
est.I(10000,3,9)

## [1] 0.05018693
```

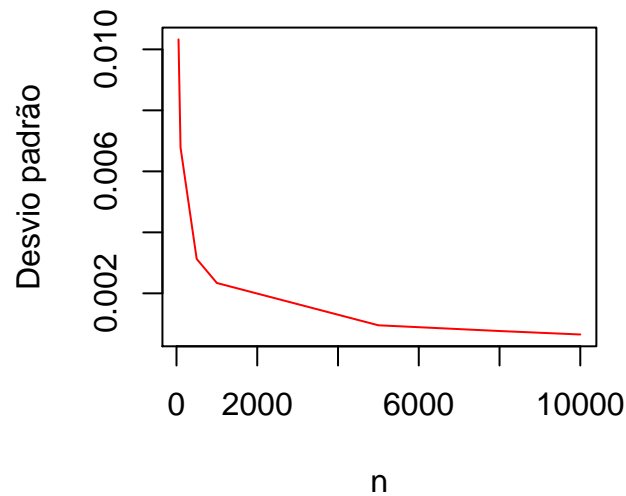
Para verificar a evolução do erro, considere os seguintes números de simulações: $n = 50, 100, 500, 1000, 5000, 8000, 10000$. Cada número de si-

mulações será repetido 100 vezes, encontrando a média e o desvio padrão.

```
n<-c(50,100,500,1000,5000,8000,10000)
#matriz para armazenar os valores da média e desvio padrão
resultado<-matrix(0,ncol=length(n),nrow=2)
for(j in 1:length(n)){
  nn<-c()
  for(i in 1:100){
    nn[i]<-est.I(n[j],3,9) #utilizando a função est.I
  }
  resultado[1,j]<-mean(nn)
  resultado[2,j]<-sd(nn)
}
round(resultado,4) #arredondando para 4 casas decimais

##          [,1]    [,2]    [,3]    [,4]    [,5]    [,6]    [,7]
## [1,] 0.0491 0.0507 0.0498 0.0495 0.0498 0.0497 0.0496
## [2,] 0.0103 0.0068 0.0031 0.0023 0.0010 0.0008 0.0007

#gráfico da evolução do erro
plot(n,resultado[2,],type="l",ylab="Desvio padrão", col="red")
```



EXEMPLO COMPUTACIONAL 4.8

Sabendo que $X \sim \text{Exponencial}(3)$, encontre uma estimativa Monte Carlo para $E(X^2)$. Calcule o erro de Monte Carlo.

Solução:

Sabendo que $X \sim \text{Exponencial}(3)$, tem-se que:

$$E(X^2) = \int_0^{\infty} x^2 f_X(x) dx = \int_0^{\infty} x^2 (3e^{-3x}) dx.$$

Neste caso, tem-se a esperança da função de variável aleatória $g(X) = X^2$.

Para encontrar uma estimativa Monte Carlo para $E(X^2)$, será implemen-

tado no *software* R o seguinte algoritmo:

1. Gere x_1, x_2, \dots, x_n da distribuição exponencial ($X \sim Exponencial(3)$);
2. Obtenha $g(x_1), g(x_2), \dots, g(x_n)$, em que $g(x_i) = x_i^2$;
3. Encontre $\frac{1}{n} \sum_{i=1}^n x_i^2$.

```
set.seed(123)
n<-10000
x<-rexp(n,3)
g<-x^2
estimativaMC<-mean(g)
print(estimativaMC)

## [1] 0.2230268
```

O erro de Monte Carlo (*EMC*) é calculado obtendo a raiz quadrada da variância do estimador de Monte Carlo, sendo:

$$EMC = \sqrt{\frac{\sum_{i=1}^n (g(x_i) - \bar{g})^2}{n^2}} = \sqrt{\frac{1}{n} \frac{\sum_{i=1}^n (g(x_i) - \bar{g})^2}{n}}.$$

Considerando as informações do exemplo, o *EMC* é obtido como a seguir.

```
soma2<-(g-estimativaMC)^2
msoma<-mean(soma2)
```

```
EMC<-sqrt(msoma/n)
print(EMC)

## [1] 0.004908889
```

EXEMPLO 4.3

Suponha que o tempo adicional de vida de um indivíduo x possua distribuição $T \sim Exponencial(\lambda)$ e que desejamos calcular o valor presente atuarial θ de um seguro de vida que paga 50 unidades monetárias no momento de morte com taxa de juros instantânea δ . Podemos encontrar θ fazendo

$$\theta = E[g(T)] = E[50e^{-\delta T}].$$

Como estimar por Monte Carlo o valor de θ ?

Solução:

Para encontrar uma estimativa Monte Carlo para θ , pode-se implementar o seguinte algoritmo, considerando valores fixos para λ e δ :

1. Gere t_1, t_2, \dots, t_n da distribuição exponencial ($T \sim Exponencial(\lambda)$);
 2. Obtenha $g(t_1), g(t_2), \dots, g(t_n)$, em que $g(t_i) = 50e^{-\delta t_i}$;
 3. Encontre $\frac{1}{n} \sum_{i=1}^n g(t_i)$.
-

4.2.5 Amostragem por importância

Suponha que o interesse seja calcular $I = \int g(\theta)d\theta$. Para tanto, considere Θ como uma variável aleatória com f.d.p $f_{\Theta}(\theta)$ e a variável aleatória $Y = \frac{g(\Theta)}{f_{\Theta}(\Theta)}$. Dessa maneira, tem-se que:

$$I = \int g(\theta)d\theta = \int \frac{g(\theta)}{f_{\Theta}(\theta)} f_{\Theta}(\theta)d\theta = \int y f_{\Theta}(\theta)d\theta = E(Y).$$

Com base numa amostra aleatória de Y , a estimativa de $E(Y)$ é encontrada por meio do estimador \bar{Y} :

$$\bar{Y} = \frac{1}{m} \sum_{i=1}^m Y_i = \frac{1}{m} \sum_{i=1}^m \frac{g(\theta_i)}{f_{\Theta}(\theta_i)},$$

em que as variáveis aleatórias $\theta_1, \dots, \theta_m$ são geradas de uma distribuição com f.d.p $f_{\Theta}(\theta)$, denominada de função de importância. A variância de \bar{Y} é igual a $\frac{Var(Y)}{m}$, sendo que a variância de Y deve ser pequena. Dessa maneira, a função de importância deverá ser aquela que resulta na menor variância de Y .

Suponha que agora o interesse seja calcular $I = \int_{-\infty}^{\infty} g(\theta)f_{\Theta}(\theta)d\theta$. Reescrevendo I , tem-se que:

$$I = \int_{-\infty}^{\infty} g(\theta)f_{\Theta}(\theta)d\theta = \int_{-\infty}^{\infty} \frac{g(\theta)f_{\Theta}(\theta)}{\phi_{\Theta}(\theta)}\phi_{\Theta}(\theta)d\theta = \int_{-\infty}^{\infty} w(\theta)\phi_{\Theta}(\theta)d\theta.$$

em que $\phi_{\Theta}(\theta)$ é chamada de função de importância, que deve ser de fácil

amostragem, e a razão $w(\theta)$ é denominada de peso. Dessa maneira, $I = E_{\phi_{\Theta}}[w(\Theta)]$, sendo que o seguinte algoritmo é utilizado para encontrar uma estimativa para I :

1. Gere $\theta_1, \theta_2, \dots, \theta_n$ de $\phi_{\Theta}(\theta)$;
2. Calcule os pesos $w(\theta_i) = \frac{g(\theta_i)f_{\Theta}(\theta_i)}{\phi_{\Theta}(\theta)_i}$;
3. Encontre $\hat{I} = \frac{1}{n} \sum_{i=1}^n w(\theta_i)$.

EXEMPLO COMPUTACIONAL 4.9

Encontre uma estimativa para $\int_0^8 2e^{-2\theta} d\theta$, utilizando amostragem por importância. Considere as funções de importância $f_1 = f_{\Theta}(\theta) = 5e^{-5\theta}$, para $0 < \theta < \infty$, e $f_2 = f_{\Theta}(\theta) = \frac{1}{8}$, para $0 < \theta < 8$.

Solução:

A seguir, são apresentados os comandos para encontrar estimativas para $\int_0^8 2e^{-2\theta} d\theta$, considerando m valores das funções de importância $f_1 = f_{\Theta}(\theta) = 5e^{-5\theta}$, sendo que $\Theta \sim \text{Exponencial}(5)$, e $f_2 = f_{\Theta}(\theta) = \frac{1}{8}$, sendo que $\Theta \sim \text{Uniforme}(0, 8)$. A função de importância escolhida foi $f_1 = f_{\Theta}(\theta) = 5e^{-5\theta}$, por apresentar a menor variância.

```
set.seed(123)
m<-10000

g<-function(theta) 2*exp(-2*theta) * (theta>0)*(theta<8)
```



```
#considerando f1
theta<-rexp(m,5)
f1<-5*exp(-5*theta)
r<-g(theta)/f1
estimativa<-mean(r)
variancia<-var(r)
estimativa
```

```
## [1] 0.9865448
```

```
variancia
```

```
## [1] 4.160063
```

```
#considerando f2
theta<-runif(m,0,8)
f2<-1/8
r<-g(theta)/f2
estimativa<-mean(r)
variancia<-var(r)
estimativa
```

```
## [1] 1.010606
```

```
variancia
```

[1] 7.156242

4.2.6 Testes de hipóteses

Com base numa amostra, pode-se testar uma hipótese referente ao parâmetro θ de uma população, sendo necessário tomar a decisão de aceitar ou não tal hipótese. Dessa maneira, é verificado se a amostra traz evidência que apoie ou não a hipótese de interesse.

Considere H_0 como a hipótese de interesse, denominada de hipótese nula. Caso H_0 seja rejeitada, aceita-se a hipótese alternativa H_1 . A decisão de aceitar ou não H_0 é realizada por meio de uma região crítica (RC), que é a região de rejeição de H_0 com base numa estatística do teste.

Com base em H_0 e H_1 , têm-se os erros de tipo I e tipo II. O erro de tipo I ocorre quando rejeitamos H_0 , sendo H_0 verdadeira. O erro de tipo II ocorre quando aceitamos H_0 , sendo H_0 falsa.

A probabilidade de cometer o erro de tipo I, denotada por α , é definida como nível de significância do teste. Procura-se controlar o erro de tipo I, fixando α , e construindo a RC que resulte em $P(\text{RC} | H_0 \text{ verdadeira}) = \alpha$.

A probabilidade de cometer o erro de tipo II é denotada por β , que frequentemente não é calculada, pois H_1 é estabelecida sob um conjunto de valores para o parâmetro θ . Com isso, pode-se definir a função caracterísitca de operação do teste como $\beta(\theta) = P(\text{aceitar } H_0 | \theta)$, que representa a probabilidade de aceitar H_0 em função de θ . A probabilidade de rejeitar H_0 em função de θ é dada por $\pi(\theta) = 1 - \beta(\theta)$, chamada de função poder do teste.

A probabilidade de significância ou valor-p do teste é a probabilidade de ocorrer valores mais extremos da estatística do teste do que o valor observado pela amostra, dada que a hipótese H_0 seja verdadeira. Ou seja, o valor-p é o menor valor possível de α , tal que o valor observado da estatística do teste leve a rejeitar H_0 .

EXEMPLO 4.4

Considere que o valor de sinistro X segue uma distribuição normal com média μ e variância σ^2 . Uma seguradora deseja testar se o valor médio de sinistro é diferente de 1400 unidades monetárias ($\mu = 1400$). Suponha que o desvio padrão é igual a 200 ($\sigma = 200$). Foi selecionado aleatoriamente valores de 60 sinistros, obtendo a média amostral igual a 1520. Proponha um teste de hipóteses considerando $\alpha = 1\%$.

Solução:

Considere os seguintes passos para a construção do teste de hipóteses:

Passo 1 - Defina as hipóteses H_0 e H_1 .

De acordo com as informações do exemplo, tem-se um teste de hipóteses bilateral, com as seguinte hipóteses:

$$\begin{aligned}H_0: \mu &= \mu_0, \\H_1: \mu &\neq \mu_0,\end{aligned}$$

sendo $\mu_0 = 1400$.

Passo 2 - Use uma estatística para testar H_0 .

Se H_0 for verdadeira, tem-se pelo TCL que $\bar{X} \approx Normal(\mu_{\bar{X}}, \sigma_{\bar{X}}^2)$, sendo $\mu_{\bar{X}} = \mu = 1400$ e $\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n} = \frac{200^2}{60} \approx 666,67$.

Passo 3 - Fixe a probabilidade de cometer o erro de tipo I (α) e construa a região crítica (RC).

Pela hipótese H_1 , tem-se que H_0 deve ser rejeitada quando \bar{X} for muito pequena ou muito grande, caracterizando um teste de hipóteses bilateral. Utilizando a tabela da distribuição normal padrão para $\alpha = 1\%$, têm-se os valores tabelados $z_1 = -z_{\frac{\alpha}{2}} = -2,58$ e $z_2 = z_{\frac{\alpha}{2}} = 2,58$, que correspondem aos seguintes valores de \bar{x} :

$$z_1 = -2,58 = \frac{\bar{x}_1 - 1400}{25,82},$$

$$\bar{x}_1 = 1333,38.$$

$$z_2 = 2,58 = \frac{\bar{x}_2 - 1400}{25,82},$$

$$\bar{x}_2 = 1466,62.$$

Dessa maneira, a RC é dada por:

$$RC = \{\bar{x} \in \mathbb{R} | \bar{x} \leq 1333,38 \text{ ou } \bar{x} \geq 1466,62\}.$$

Passo 4 - Calcule o valor da estatística do teste com base na informação da amostra.

Considerando os valores de 60 sinistros, tem-se que $\bar{x} = 1520$.

Passo 5 - Decida se H_0 deve ser ou não rejeitada.

Como o valor observado da estatística do teste \bar{x} pertence a região crítica, a conclusão do teste é rejeitar H_0 .

EXEMPLO COMPUTACIONAL 4.10

Considere que o valor de sinistro X segue uma distribuição normal com média μ e variância σ^2 . Uma seguradora afirma que o valor médio de sinistro não é maior que 2100 unidades monetárias. Uma amostra de 40 sinistros apresentou valor médio de sinistro igual a 1860 e desvio padrão igual a 100.

a) Proponha um teste de hipóteses considerando $\alpha = 5\%$ para avaliar a afirmação da seguradora.

b) Simule amostras de tamanho 40, sendo que $X \sim Normal(2100, 100^2)$. Encontre a probabilidade de erro do tipo I e verifique se é aproximadamente igual a $\alpha = 5\%$.

Solução:

a)

Passo 1 - Defina as hipóteses H_0 e H_1 .

De acordo com as informações do exemplo, tem-se um teste de hipóteses unilateral à direita, com as seguinte hipóteses:

$$H_0: \mu = \mu_0,$$

$$H_1: \mu > \mu_0,$$

sendo $\mu_0 = 2100$.

Passo 2 - Use uma estatística para testar H_0 .

Sabendo que $X \sim Normal(\mu, \sigma^2)$, com variância amostral conhecida, pode-se utilizar a estatística

$$T = \frac{\sqrt{n}(\bar{X} - \mu_0)}{s},$$

sendo que s é o desvio padrão amostral e T segue uma distribuição t de Student com $n - 1$ graus de liberdade.

Passo 3 - Fixe a probabilidade de cometer o erro de tipo I (α) e construa a região crítica (RC).

Pela hipótese H_1 , tem-se que H_0 deve ser rejeitada quando T for muito grande, caracterizando um teste de hipóteses unilateral à direita. Utilizando a tabela t de Student, para $n - 1 = 39$ graus de liberdade e $\alpha = 5\%$, tem-se o valor tabelado t igual a 1,711. Ou seja, a RC é dada por:

$$RC = \{t \in \mathbb{R} | t \geq 1,711\}.$$

Passo 4 - Calcule o valor da estatística do teste com base na informação da amostra.

Considerando os valores de 40 sinistros, tem-se

$$t = \frac{\sqrt{40}(1860 - 2100)}{100} = -15,18.$$

Passo 5 - Decida se H_0 deve ser ou não rejeitada.

Como o valor observado da estatística do teste não pertence a região crítica, a conclusão do teste é aceitar H_0 .

b)

Os comandos a seguir apresentam a implementação do teste utilizando a estatística T . Para tanto, foi utilizada a função **t.test**, que permite realizar o teste unilateral à direita, obtendo os valores-p. Foram simuladas 10000 amostras de tamanho igual a 40, encontrando assim, 10000 valores-p. Dessa maneira, foi calculada a média de ocorrência de valores-p menores que α , encontrando a proporção de testes significativos, ou seja, que rejeitam H_0 a $\alpha\%$ de probabilidade.

```
n<-40 #tamanhos das amostras simuladas
alpha<-0.05 #nível de significância
mi0<-2100 #média
sigma<-100 #desvio padrão
m<-10000 #número de amostras simuladas
valoresp<-numeric(m)
for(i in 1:m){
  x<-rnorm(n,mi0,sigma)
  #comando t.test - teste unilateal à direita
```

```
testet<-t.test(x,alternative="greater",mu=mi0)
valoresp[i]<-testet$p.value
}
p.estimado<-mean(valoresp<alpha)
p.estimado

## [1] 0.0504
```

A média de ocorrência de valores-p menores que α foi igual a 0,0504, que é próximo ao valor teórico de 5%.

4.2.6.1 Testes de aderência

Pode-se estar interessado em testar a hipótese de que um conjunto de dados segue uma distribuição específica. Por exemplo, pode-se testar as seguintes hipóteses nula e alternativa:

H_0 : Não existe diferença entre as frequências obtidas na amostra e as frequências esperadas, dado que a população tem distribuição de probabilidade conhecida,

H_1 : Existe diferença entre as frequências obtidas na amostra e as frequências esperadas, dado que a população tem distribuição de probabilidade conhecida.

Os testes de aderência avaliam a adequabilidade de uma distribuição específica a um conjunto de dados observados. Existem vários testes de aderência, como os testes de Kolmogorov-Smirnov, Shapiro-Wilk, Anderson-Darling e Qui-quadrado. A seguir, será apresentado o teste Qui-quadrado,

que pode ser aplicado a distribuições discretas ou contínuas, exigindo uma categorização dos dados. Também será apresentado o teste de Kolmogorov-Smirnov, utilizado para distribuições contínuas.

Teste Qui-quadrado

A estatística de teste de aderência Q , para dados agrupados em intervalos, testa se as frequências observadas diferem muito das frequências esperadas, sendo definida como:

$$Q = \sum_{j=1}^k \frac{(O_j - E_j)^2}{E_j},$$

em que O_j representa o número de observações no grupo ou categoria j (frequência observada) e E_j representa o número esperado de observações no grupo ou categoria j (frequência esperada), sob a hipótese H_0 de que os dados seguem uma distribuição específica. No caso, tem-se $E_j = np_j$, sendo p_j a probabilidade dada pelo modelo sob a hipótese H_0 referente a categoria j e n é o número total de observações. A estatística Q segue uma distribuição qui-quadrado com $\phi = k - 1$ graus de liberdade (χ_ϕ^2), sendo que se $Q > \chi_\phi^2$, rejeita-se a hipótese H_0 para o nível de significância α . No teste Qui-quadrado, é necessário especificar o número de grupos ou categorias, onde as contagens serão realizadas. Na literatura, é mencionado que deve-se escolher grupos de maneira que o número esperado de observações em cada grupo seja de no mínimo 5.

EXEMPLO 4.5 (CAS/SOA)

Durante um ano, o número de acidentes N por dia foi distribuído como

a seguir. Teste a hipótese de que os dados seguem uma distribuição de Poisson com média 0,6 usando o número máximo de grupos, de modo que cada grupo tenha no mínimo 5 observações esperadas. Use um nível de significância de 5%.

Número de acidentes	Dias
0	209
1	111
2	33
3	7
4	3
5	2

Solução:

Sob a hipótese H_0 de que os dados seguem uma distribuição de Poisson, os valores esperados E_j para t_j acidentes foram obtidos, com $n = 365$ observações. Dessa maneira, tem-se que:

$$E_j = np_j = nP(N = t_j) = n \frac{\lambda^{t_j} e^{-\lambda}}{t_j!} = 365 \frac{0,6^{t_j} e^{-0,6}}{t_j!}.$$

A seguir, é apresentada uma tabela de frequências com os valores observados e esperados.

Número de acidentes (n_j)	Observado (O_j)	Esperado (E_j)
0	209	200,32
1	111	120,19
2	33	36,06
3	7	7,21
4	3	1,08
5	2	0,14

Considerando o número máximo de grupos, de modo que cada grupo tenha no mínimo 5 observações esperadas, tem-se os seguintes grupos, considerando o agrupamento das últimas três linhas da tabela anterior, obtendo um número observado de observações igual a 12 e um número esperado de observações igual a 8,43.

Grupo (j)	Observado (O_j)	Esperado (E_j)
1	209	200,32
2	111	120,19
3	33	36,06
4	12	8,43

Assim, a estatística Q é obtida como a seguir:

$$Q = \sum_{j=1}^4 \frac{(O_j - E_j)^2}{E_j} = \frac{(209 - 200,32)^2}{200,32} + \dots + \frac{(12 - 8,43)^2}{8,43} = 2,85.$$

No *software* R, obtem-se χ_ϕ^2 por meio do comando **qchisq(1- α , ϕ)**, que calcula o quantil correspondente a $(1 - \alpha)\%$ de probabilidade. Para $\phi = 3$ e $\alpha = 0,05$, tem-se que $\chi_3^2 = 7,81$. Como $Q < \chi_3^2$, não rejeita a hipótese H_0 à 5% de significância, ou seja, o número de acidentes por dia possui, estatisticamente, uma distribuição de Poisson com média 0,6.

Teste de Kolmogorov-Smirnov

O teste de Kolmogorov-Smirnov se baseia na distância entre a distribuição acumulada observada (distribuição empírica), $F_n(x)$, e a distribuição acumulada teórica, $F_X(x)$, que é a distribuição acumulada esperada. O objetivo é testar se uma amostra é proveniente de uma população com distribuição

acumulada esperada $F_X(x)$. Dessa maneira, as hipóteses a serem testadas são:

$$\begin{aligned} H_0: F_n(x) &= F_X(x), \\ H_1: F_n(x) &\neq F_X(x). \end{aligned}$$

A estatística de teste de Kolmogorov-Smirnov é dada seguinte maneira:

$$D_n = \sup_x |F_n(x) - F_X(x)|,$$

em que D_n é o menor limite superior de todas as diferenças pontuais $|F_n(x) - F_X(x)|$. Na literatura, é mencionado que os valores críticos para o teste de Kolmogorov-Smirnov podem ser: $1,22/\sqrt{n}$ para $\alpha = 0,1$; $1,36/\sqrt{n}$ para $\alpha = 0,05$ e; $1,63/\sqrt{n}$ para $\alpha = 0,01$

Observação: Para um conjunto de valores reais x_1, \dots, x_n ($i = 1, \dots, n$), a função de distribuição empírica $F_n(x)$ é uma função $F_n(x) : \mathbb{R} \rightarrow [0, 1]$, tal que, para qualquer $x \in \mathbb{R}$, tem-se que $F_n(x) = \frac{\#\{x_i \leq x\}}{n}$, que representa a proporção dos valores x_i que são menores ou iguais a x , sendo n o número total de observações. No *software* R, a função de distribuição empírica dos dados pode ser obtida por meio do comando **ecdf**.

EXEMPLO 4.6 (CAS/SOA)

Considere uma amostra aleatória com as seguintes observações: 0,1; 0,2; 0,5; 0,7 e 1,3. Você testa a hipótese de que a f.d.p é $f_X(x) = \frac{4}{(1+x)^5}$. Calcule a estatística de teste de Kolmogorov-Smirnov.

Solução:

Como a f.d.p teórica é $f_X(x) = \frac{4}{(1+x)^5}$, obtém-se a função de distribuição teórica $F_X(x) = 1 - \frac{1}{(1+x)^4}$. Considerando a amostra aleatória com 5 observações, tem-se a seguinte função de distribuição empírica $F_n(x) = F_5(x)$:

$$F_5(x) = \begin{cases} 0, & x < 0,1, \\ 0,2, & 0,1 \leq x < 0,2, \\ 0,4, & 0,2 \leq x < 0,5, \\ 0,6, & 0,5 \leq x < 0,7, \\ 0,8, & 0,7 \leq x < 1,3, \\ 1,0, & x > 1,3. \end{cases}$$

Dessa maneira, comparamos a função de distribuição teórica $F_X(x)$ com a função de distribuição empírica $F_5(x_i^-)$, obtendo as diferenças $|F_5(x_i^-) - F_X(x_i)|$.

x_i	$F_5(x_i^-)$	$F_5(x_i)$	$F_X(x_i)$	$ F_5(x_i^-) - F_X(x_i) $
0,1	0	0,2	0,3170	0,3170
0,2	0,2	0,4	0,5177	0,3177
0,5	0,4	0,6	0,8025	0,4025
0,7	0,6	0,8	0,8803	0,2803
1,3	0,8	1,0	0,9643	0,1643

Com isso, a estatística de teste de Kolmogorov-Smirnov é igual ao maior valor de $|F_5(x_i^-) - F_X(x_i)|$, ou seja, $D_5 = 0,4025$. O valor crítico pode ser calculado conforme o critério $1,36/\sqrt{n}$, para $\alpha = 0,05$. Como $1,36/\sqrt{n} = 0,6082 > 0,4025$, rejeita-se a hipótese nula, ou seja, a função de distribuição teórica $F_X(x) = 1 - \frac{1}{(1+x)^4}$ não é geradora dos dados.

4.3 Estatística bayesiana

Na estatística clássica, dada uma variável aleatória X , com f.d.p ou f.p, $f_{X|\Theta}(x|\theta)$, o parâmetro θ é fixo e desconhecido. Na estatística bayesiana, θ é considerado como uma variável aleatória, com f.d.p ou f.p, $\pi_{\Theta}(\theta)$, denominada de distribuição *a priori* de θ , que representa uma informação subjetiva sobre o modelo, antes de ser coletada uma amostra aleatória de n observações de X . Coletada uma amostra de X , a informação de Θ será atualizada com base no teorema de Bayes para variáveis aleatórias, obtendo a distribuição *a posteriori* de Θ dado X . O teorema de Bayes é dado da seguinte maneira:

$$\pi_{\Theta|X}(\theta|x) = \frac{f_{X,\Theta}(x, \theta)}{f_X(x)} = \frac{f_{X|\Theta}(x|\theta)\pi_{\Theta}(\theta)}{f_X(x)},$$

em que:

$f_{X,\Theta}(x, \theta)$ é a distribuição conjunta de X e Θ ;

$f_{X|\Theta}(x|\theta)$ é a função de verossimilhança de X , sendo $f_{X|\Theta}(x|\theta) = \prod_{i=1}^n f_{X_i|\Theta}(x_i|\theta)$;

$f_X(x)$ é a distribuição marginal de X , sendo que $f_X(x) = \int_{\theta \in \Omega_{\Theta}} f_{X,\Theta}(x, \theta) d\theta$;

$\pi_{\Theta}(\theta)$ é a distribuição *a priori* de Θ ;

$\pi_{\Theta|X}(\theta|x)$ é a distribuição *a posteriori* de Θ dado X .

Considerando que o interesse seja somente nos termos que envolvem Θ , pode-se reescrever o teorema de Bayes, considerando $f_X(x)$ como uma cons-

tante de normalização. Dessa maneira, tem-se o teorema de Bayes de forma simplificada:

$$\pi_{\Theta|X}(\theta|x) \propto f_{X|\Theta}(x|\theta)\pi_{\Theta}(\theta).$$

Com relação a classificação das distribuições *a priori*, uma *priori* é dita ser não informativa quando reflete uma ignorância total ou um conhecimento muito limitado sobre o parâmetro de interesse. Uma distribuição *a priori* é dita ser conjugada quando possui a mesma distribuição *a posteriori*, envolvendo uma mudança nos parâmetros da distribuição *a priori*, que são denominados de hiperparâmetros.

Em muitas situações, por facilitar manipulações algébricas, são utilizadas distribuições *a priori* conjugadas, sendo que nem sempre, tais distribuições refletem a realidade. Com o não uso de distribuições *a priori* conjugadas, pode-se obter distribuições *a posteriori* de alta complexidade, não possuindo solução fechada. Com isso, métodos de simulação estocástica podem ser utilizados, obtendo amostras simuladas da distribuição *a posteriori*. Um dos métodos utilizado é o Monte Carlo via Cadeias de Markov (*Markov Chain Monte Carlo - MCMC*).

EXEMPLO COMPUTACIONAL 4.11

Suponha que X tem distribuição exponencial com parâmetro Λ e que Λ tem distribuição gama com parâmetros α e β . Encontre a distribuição *a posteriori* de Λ dado X .

Considere:

$$\pi_{\Lambda}(\lambda) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda},$$

$$f_{X|\Lambda}(x|\lambda) = \prod_{i=1}^n \lambda e^{-\lambda x_i}.$$

Solução:

Sabendo que $X \sim \text{Exponencial}(\Lambda)$ e $\Lambda \sim \text{Gama}(\alpha, \beta)$, a distribuição *a posteriori* de Λ dado X pode ser encontrada via teorema de Bayes de forma simplificada. Para tanto, serão considerados os núcleos das distribuições, ou seja, os termos que envolvem Λ .

$$\pi_{\Lambda|X}(\lambda|x) \propto f_{X|\Lambda}(x|\lambda) \pi_{\Lambda}(\lambda),$$

$$\pi_{\Lambda|X}(\lambda|x) \propto \left[\prod_{i=1}^n \lambda e^{-\lambda x_i} \right] \left[\lambda^{\alpha-1} e^{-\beta\lambda} \right],$$

$$\pi_{\Lambda|X}(\lambda|x) \propto \left[\lambda^n e^{-\lambda \sum_{i=1}^n x_i} \right] \left[\lambda^{\alpha-1} e^{-\beta\lambda} \right],$$

$$\pi_{\Lambda|X}(\lambda|x) \propto \lambda^{\alpha+n-1} e^{-\lambda(\sum_{i=1}^n x_i + \beta)}.$$

A distribuição *a posteriori* é proporcional ao núcleo de uma distribuição gama com parâmetros $\alpha^* = \alpha + n - 1$ e $\beta^* = \sum_{i=1}^n x_i + \beta$. Dessa maneira, $\Theta|X \sim \text{Gama}(\alpha^*, \beta^*)$. Pode-se observar que a distribuição *a priori*

é uma distribuição conjugada, pois a distribuição *a posteriori* é da mesma família da distribuição *a priori*, ocorrendo apenas uma mudança nos hiperparâmetros. Vale ressaltar que a média da distribuição *a posteriori*, $E(\Theta|X)$, é um estimador bayesiano, sendo que:

$$\hat{\lambda} = E(\Theta|X) = \frac{\alpha^*}{\beta^*},$$

$$\hat{\lambda} = \frac{\alpha + n - 1}{\sum_{i=1}^n x_i + \beta}.$$

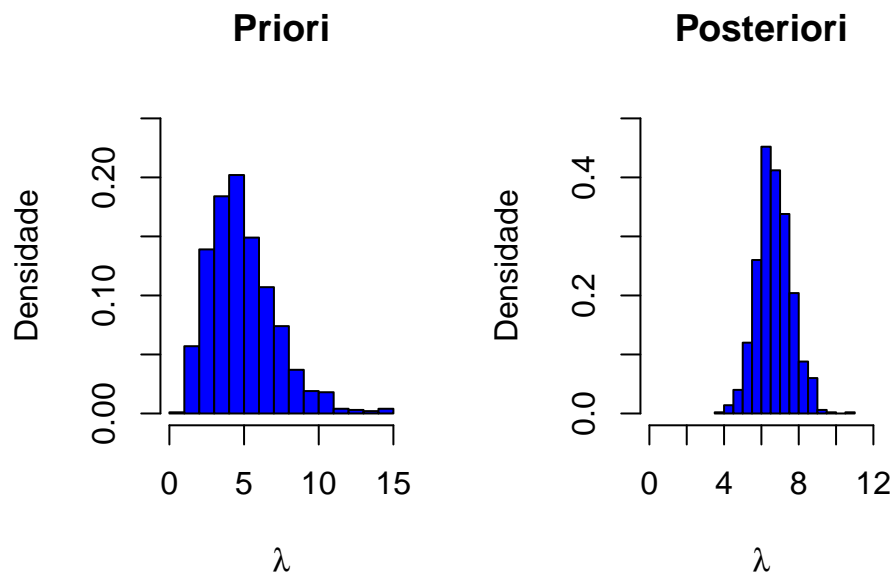
Os comandos a seguir apresentam os resultados para a distribuição *a posteriori* de Λ dado X , considerando que $X \sim \text{Exponencial}(8)$ e $\Lambda \sim \text{Gama}(5, 1)$. Foram consideradas uma amostra de tamanho 50 de X e uma amostra de tamanho 1000 da distribuição *a posteriori*.

```
set.seed(123)
x<-rexp(50,8)

par(mfrow=c(1,2))

priori<-rgamma(1000,5,1)
hist(priori,freq=F,xlab=expression(lambda),ylab="Densidade",
xlim=c(0,15),ylim=c(0,0.25),col="blue",main="Priori")

posteriori<-rgamma(1000,54,1+sum(x))
hist(posteriori,freq=F,xlab=expression(lambda),
ylab="Densidade",xlim=c(0,12),ylim=c(0,0.5),col="blue",
main="Posteriori")
```



```
summary(posteriori)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  3.618   6.102   6.624   6.684   7.278  10.587
```

```
sd(posteriori)
```

```
## [1] 0.9104898
```

```
#computando os percentis 2,5% e 97,5%.
```

```
quantile(posteriori,c(0.025,0.975))
```

```
##      2.5%      97.5%
## 4.971265 8.585219
```

Os percentis 2,5% e 97,5% da distribuição *a posteriori* representam os limites do intervalo de credibilidade bayesiano de 95% para λ ($ICB_{95\%}(\lambda)$). Assim, $ICB_{95\%}(\lambda)$ é o intervalo de valores mais prováveis de λ que soma probabilidade de 0,95.

4.3.1 Inferência preditiva

A inferência preditiva, baseada na distribuição *a posteriori*, tem como finalidade obter previsões, realizando inferências sobre dados futuros (x_{n+1}) com base em dados passados (\mathbf{x}). Considerando que o parâmetro Θ é uma variável aleatória contínua, a distribuição preditiva, $f_{X_{n+1}|\mathbf{X}}(x_{n+1}|\mathbf{x})$, é calculada da seguinte maneira:

$$f_{X_{n+1}|\mathbf{X}}(x_{n+1}|\mathbf{x}) = \frac{f_{X_{n+1},\mathbf{X}}(x_{n+1},\mathbf{x})}{f_{\mathbf{X}}(\mathbf{x})},$$

$$f_{X_{n+1}|\mathbf{X}}(x_{n+1}|\mathbf{x}) = \frac{\int_{\theta \in \Omega_{\Theta}} [\Pi_{i=1}^{n+1} f_{X_i|\Theta}(x_i|\theta)] \pi_{\Theta}(\theta) d\theta}{f_{\mathbf{X}}(\mathbf{x})},$$

$$f_{X_{n+1}|\mathbf{X}}(x_{n+1}|\mathbf{x}) = \frac{\int_{\theta \in \Omega_{\Theta}} f_{X_{n+1}|\Theta}(x_{n+1}|\theta) [\Pi_{i=1}^n f_{X_i|\Theta}(x_i|\theta)] \pi_{\Theta}(\theta) d\theta}{f_{\mathbf{X}}(\mathbf{x})},$$

$$f_{X_{n+1}|\mathbf{X}}(x_{n+1}|\mathbf{x}) = \frac{\int_{\theta \in \Omega_{\Theta}} f_{X_{n+1}|\Theta}(x_{n+1}|\theta) \pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\theta}{f_{\mathbf{X}}(\mathbf{x})},$$

$$f_{X_{n+1}|\mathbf{X}}(x_{n+1}|\mathbf{x}) = \int_{\theta \in \Omega_{\Theta}} f_{X_{n+1}|\Theta}(x_{n+1}|\theta) \pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) d\theta.$$

Não encontrando uma solução fechada para $f_{X_{n+1}|\mathbf{X}}(x_{n+1}|\mathbf{x})$, pode-se obter uma solução aproximada via integração Monte Carlo, sendo:

$$f_{X_{n+1}|\mathbf{X}}(x_{n+1}|\mathbf{x}) \approx \frac{1}{m} \sum_{i=1}^m f_{X_{n+1}|\Theta}(x_{n+1}|\theta_i),$$

para uma amostra grande de θ .

Em atuária, a média da distribuição preditiva é chamada de Prêmio Bayesiano, que pode ser determinada da seguinte maneira:

$$E(X_{n+1}|\mathbf{X} = \mathbf{x}) = \int_0^{\infty} x_{n+1} f_{X_{n+1}|\mathbf{X}}(x_{n+1}|\mathbf{x}) dx_{n+1},$$

$$E(X_{n+1}|\mathbf{X} = \mathbf{x}) = \int_0^{\infty} x_{n+1} \left[\int_{\theta \in \Omega_{\Theta}} f_{X_{n+1}|\Theta}(x_{n+1}|\theta) \pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) d\theta \right] dx_{n+1},$$

$$E(X_{n+1}|\mathbf{X} = \mathbf{x}) = \int_{\theta \in \Omega_{\Theta}} \left[\int_0^{\infty} x_{n+1} f_{X_{n+1}|\Theta}(x_{n+1}|\theta) dx_{n+1} \right] \pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) d\theta,$$

$$E(X_{n+1}|\mathbf{X} = \mathbf{x}) = \int_{\theta \in \Omega_{\Theta}} E(X_{n+1}|\Theta = \theta) \pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) d\theta,$$

$$E(X_{n+1}|\mathbf{X}) = E[E(X_{n+1}|\Theta)|\mathbf{x}].$$

EXEMPLO 4.7 (CAS/SOA)

Considere as seguintes informações de uma seguradora.

Classe de risco	Número de segurados	Probabilidade do número de sinistros				
		0	1	2	3	4
1	3000	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	0	0
2	2000	0	$\frac{1}{6}$	$\frac{2}{3}$	$\frac{1}{6}$	0
3	1000	0	0	$\frac{1}{6}$	$\frac{2}{3}$	$\frac{1}{6}$

Um segurado é selecionado aleatoriamente e apresentou um sinistro no ano

1. Determine o número esperado de sinistros no ano 2.

Solução:

O número esperado de sinistros no ano 2, que é a média da distribuição preditiva apresentada anteriormente, será obtido como a seguir:

$$E(X_2|X_1 = x_1) = \sum_{j=1}^3 E(X_2|c_j) \pi_{C|X}(c_j|x_1),$$

em que $\pi_{C|X}(c_j|x_1) = \frac{f_{X|C}(x_1|c_j)\pi_C(c_j)}{f_X(x_1)}$, sendo que $x_1 = 1$ e a j -ésima classe de risco é representada por c_j , para $j = 1, 2, 3$.

Com base nas informações da seguradora, tem-se a seguinte distribuição *a priori* $\pi_C(c)$:

c	$\pi_C(c)$
1	$\frac{3000}{6000} = \frac{1}{2}$
2	$\frac{2000}{6000} = \frac{1}{3}$
3	$\frac{1000}{6000} = \frac{1}{6}$

A distribuição marginal $f_X(x_1)$ é obtida da seguinte maneira:

$$f_X(x_1) = \sum_{j=1}^3 f_{X|C}(x_1|c_j) \pi_C(c_j),$$

$$f_X(x_1) = f_{X|C}(1|1) \pi_C(1) + f_{X|C}(1|2) \pi_C(2) + f_{X|C}(1|3) \pi_C(3),$$

$$f_X(x_1) = \frac{1}{3} \cdot \frac{1}{2} + \frac{1}{6} \cdot \frac{1}{3} + 0 \cdot \frac{1}{6} = \frac{2}{9}.$$

Com base na distribuição *a priori* $\pi_C(c)$ e na distribuição marginal $f_X(x_1)$, obtém-se a distribuição *a posteriori* $\pi_{C|X}(c_j|x_1)$:

$$\pi_{C|X}(1|1) = \frac{f_{X|C}(1|1) \pi_C(1)}{f_X(1)} = \frac{3}{4}.$$

$$\pi_{C|X}(2|1) = \frac{f_{X|C}(1|2) \pi_C(2)}{f_X(1)} = \frac{1}{4}.$$

$$\pi_{C|X}(3|1) = \frac{f_{X|C}(1|3) \pi_C(3)}{f_X(1)} = 0.$$

Dessa maneira, o número esperado de sinistros no ano 2 é encontrado como a seguir:

$$E(X_2|1) = \sum_{j=1}^3 E(X_2|c_j) \pi_{C|X}(c_j|1),$$

$$E(X_2|1) = \left[0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3}\right] \left(\frac{3}{4}\right) + \left[1 \cdot \frac{1}{6} + 2 \cdot \frac{2}{3} + 3 \cdot \frac{1}{6}\right] \left(\frac{1}{4}\right) + 0,$$

$$E(X_2|1) = 1,25.$$

4.4 Método *bootstrap*

O método *bootstrap* consiste em realizar reamostragens com reposição da amostra original, podendo ser utilizado na estimação de parâmetros. A amostra original é considerada como uma pseudo-população, com características semelhantes a população original. Com a geração de amostras aleatórias a partir da amostra original, pode-se encontrar a distribuição de um estimador, no caso, distribuição *bootstrap*.

Pode-se trabalhar com os métodos *bootstrap* não-paramétrico e paramétrico. O método *bootstrap* não-paramétrico consiste em substituir a distribuição desconhecida da população pela distribuição empírica. Já o método *bootstrap* paramétrico consiste em assumir uma distribuição para os dados amostrais. Neste material, será apresentado somente o método *bootstrap* não-paramétrico.

Suponha que X_1, X_2, \dots, X_n é uma amostra aleatória de tamanho n de uma população com distribuição desconhecida. Considere a distribuição empírica, em que $P(X_i^* = x_i) = \frac{1}{n}$, para $i = 1, \dots, n$, sendo que $X_1^*, X_2^*, \dots, X_n^*$ é uma amostra aleatória (amostra *bootstrap*) obtida por reamostragem com reposição da amostra aleatória original. A partir de várias amostras *bootstrap*, pode-se estimar um parâmetro de interesse.

A seguir, é apresentado um algoritmo para gerar amostras *bootstrap* para estimar um parâmetro θ desconhecido de uma população.

1. Encontrar a distribuição empírica, atribuindo a probabilidade $\frac{1}{n}$ a cada observação X_i , $i = 1, \dots, n$;
2. Gerar uma amostra *bootstrap* com reposição da distribuição empírica, dada por $x^* = (x_1^*, x_2^*, \dots, x_n^*)$;
3. Calcular uma estimativa do estimador $\hat{\theta}$, com base na amostra *bootstrap*;
4. Repetir os passos 2 e 3 B vezes.

Com base no algoritmo, serão obtidas B amostras *bootstrap*, denotadas por $x^{*(b)} = (x_1^*, x_2^*, \dots, x_n^*)$ e B estimativas de $\hat{\theta}$, denotadas por $\hat{\theta}^{(b)}$, para $b = 1, \dots, B$. Dessa maneira, é obtida uma distribuição *bootstrap* do estimador $\hat{\theta}$. Com base em tal distribuição, pode-se encontrar uma aproximação *bootstrap* para o erro padrão de $\hat{\theta}$, representada por $\hat{s}(\hat{\theta}^*)$, sendo:

$$\hat{s}(\hat{\theta}^*) = \sqrt{\frac{1}{B-1} \sum_{b=1}^B \left(\hat{\theta}^{(b)} - \bar{\hat{\theta}}^* \right)^2},$$

em que $\bar{\hat{\theta}}^* = \frac{\sum_{b=1}^B \hat{\theta}^{(b)}}{B}$.

EXEMPLO 4.8

Considere a amostra aleatória $\{7, 9, 1\}$ de uma variável aleatória X . Dado o estimador $\hat{\theta} = g(x_1, x_2, x_3) = x_3$. Determine a aproximação *bootstrap* para o erro padrão de $\hat{\theta}$.

Solução:

A seguir, são apresentadas as amostras *bootstrap* e as estimativas para $\hat{\theta}$.

Amostra <i>bootstrap</i> $(x^{*(b)})$	$\hat{\theta}^{(b)}$	Amostra <i>bootstrap</i> $(x^{*(b)})$	$\hat{\theta}^{(b)}$	Amostra <i>bootstrap</i> $(x^{*(b)})$	$\hat{\theta}^{(b)}$
$\{1, 1, 1\}$	1	$\{7, 7, 1\}$	1	$\{9, 9, 7\}$	7
$\{1, 1, 7\}$	7	$\{1, 9, 9\}$	9	$\{9, 7, 9\}$	9
$\{1, 1, 9\}$	9	$\{9, 1, 9\}$	9	$\{7, 9, 9\}$	9
$\{1, 7, 1\}$	1	$\{9, 9, 1\}$	1	$\{1, 7, 9\}$	9
$\{1, 9, 1\}$	1	$\{7, 7, 7\}$	7	$\{1, 9, 7\}$	7
$\{7, 1, 1\}$	1	$\{7, 7, 9\}$	9	$\{7, 1, 9\}$	9
$\{9, 1, 1\}$	1	$\{7, 9, 7\}$	7	$\{7, 9, 1\}$	1
$\{1, 7, 7\}$	7	$\{9, 7, 7\}$	7	$\{9, 1, 7\}$	7
$\{7, 1, 7\}$	7	$\{9, 9, 9\}$	9	$\{9, 7, 1\}$	1

Dessa maneira, a aproximação *bootstrap* para o erro padrão de $\hat{\theta}$ é:

$$\hat{s}(\hat{\theta}^*) = \sqrt{\frac{1}{B-1} \sum_{b=1}^B \left(\hat{\theta}^{(b)} - \bar{\hat{\theta}}^* \right)^2},$$

$$\hat{s}(\hat{\theta}^*) = \sqrt{\frac{1}{26} [(1-5,67)^2 + (7-5,67)^2] + \dots + (1-5,67)^2} = 3,46.$$

EXEMPLO 4.9 (CAS/SOA)

Considere a amostra aleatória $\{5, 9\}$ de uma variável aleatória X . Você estima $Var(X)$ utilizando o estimador $\hat{\theta} = g(x_1, x_2) = \frac{1}{2} \sum_{i=1}^2 (x_i - \bar{x})^2$. Determine a aproximação *bootstrap* para o erro quadrático médio (*EQM*) de g .

Solução:

A seguir, são apresentadas as amostras *bootstrap* e as estimativas para $\hat{\theta}$.

Amostra <i>bootstrap</i> ($x^{*(b)}$)	$\hat{\theta}^{(b)}$
$\{5, 5\}$	0
$\{5, 9\}$	4
$\{9, 5\}$	4
$\{9, 9\}$	0

Considerando que a variância da amostra original (variância empírica) é igual a 4, obtém-se a aproximação *bootstrap* para o *EQM* de $\hat{\theta}$, fazendo:

$$\widehat{EQM}(\hat{\theta}^*) = \frac{1}{B} \sum_{b=1}^B \left(\hat{\theta}^{(b)} - \hat{\theta}^* \right)^2,$$

$$\widehat{EQM}(\hat{\theta}^*) = \frac{1}{4} [(0-4)^2 + (4-4)^2 + (4-4)^2 + (0-4)^2] = 8.$$

EXEMPLO COMPUTACIONAL 4.12

Considere a amostra aleatória $\{1020, 785, 903, 815, 522, 1007\}$ da variável aleatória X , que representa o valor de sinistro ocorrido. Encontre a distribuição *bootstrap* para a média amostral \bar{x} .

Solução:

No *software* R, para gerar amostras *bootstrap* com ou sem reposição da amostra original, pode-se utilizar o comando **sample**, como a seguir.

```
set.seed(123)

#amostra original
amostra.original<-c(1020,785,903,815,522,1007)

#gerando uma amostra bootstrap de tamanho 6 com reposição
amostra.bootstrap1<-sample(amostra.original,6,replace=T)
amostra.bootstrap1

## [1] 785 522 903 1007 1007 1020
```

```
mean(amostra.bootstrap1)

## [1] 874

#gerando uma amostra bootstrap de tamanho 6 sem reposição
amostra.bootstrap2<-sample(amostra.original,6,replace=F)
amostra.bootstrap2

## [1] 815 522 903 785 1007 1020

mean(amostra.bootstrap2)

## [1] 842
```

Para encontrar a distribuição *bootstrap*, pode-se utilizar o pacote *gtools* para obter todas as amostras *bootstrap* com reposição da amostra original. Pelo comando **dim**, é verificado que no total são obtidas 46656 amostras *bootstrap*. Pode-se observar que a média das médias *bootstrap* é igual a média da amostra original.

```
library(gtools)

#gerando todas as amostras com reposição da amostra original
amostras.bootstrap3<-permutations(n=6,r=6,v=amostra.original,
repeats.allowed=T)
```

```
dim(amostras.bootstrap3)

## [1] 46656      6

#calculando a média de cada amostra bootstrap
medias.bootstrap<-apply(amostras.bootstrap3,1,mean)

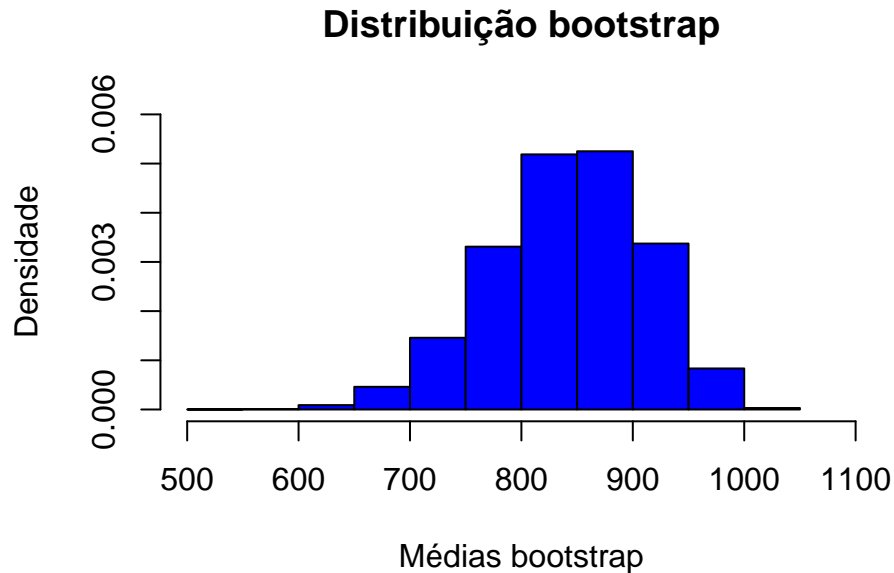
#média e erro padrão das médias bootstrap
med.medias.bootstrap<-mean(medias.bootstrap)
ep.medias.bootstrap<-sd(medias.bootstrap)
med.medias.bootstrap

## [1] 842

ep.medias.bootstrap

## [1] 68.51837

#distribuição bootstrap da média amostral
hist(medias.bootstrap,freq=F,xlab="Médias bootstrap",
ylab="Densidade",xlim=c(500,1100),ylim=c(0,0.006),
col="blue",main="Distribuição bootstrap")
```



Dependendo do tamanho da amostra original, é inviável obter todas as amostras *bootstrap* com reposição. Dessa maneira, pode-se utilizar simulação Monte Carlo para aproximar uma distribuição *bootstrap*. Utilizando a função *replicate*, são simuladas 10000 amostras *bootstrap*, como a seguir.

```
set.seed(123)

#simulando 10000 amostras bootstrap
medias.montecarlo<-replicate(10000,
mean(sample(amostra.original,8,replace=T)))

#média e erro padrão das médias bootstrap
m.medias.montecarlo<-mean(medias.montecarlo)
ep.medias.montecarlo<-sd(medias.montecarlo)
```

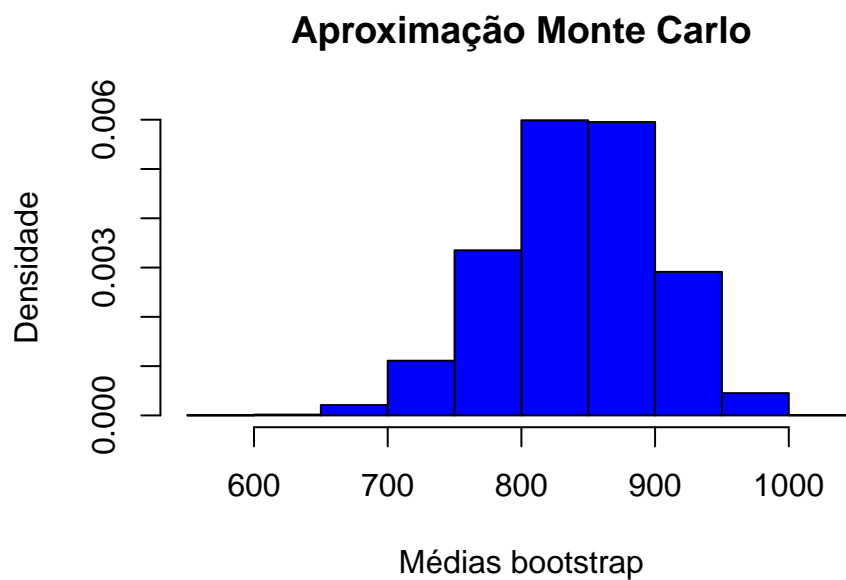
```
m.medias.montecarlo

## [1] 842.5469

ep.medias.montecarlo

## [1] 59.18166

#distribuição bootstrap da média amostral
hist(medias.montecarlo,freq=F,xlab="Médias bootstrap",
ylab="Densidade",col="blue",main="Aproximação Monte Carlo")
```



Capítulo 5

Introdução aos Métodos Monte Carlo via Cadeias de Markov

5.1 Cadeias de Markov

Uma cadeia de Markov é um processo estocástico $\{X_t\}$ indexado no tempo $t \geq 0$. Considerando que $X_t = i$ represente o processo no estado i no tempo t , a sequência de variáveis aleatórias $\{X_0, X_1, \dots\}$ é uma cadeia de Markov se a probabilidade de transição do estado atual i para o estado futuro j é definida como

$$\begin{aligned} P(X_{t+1} = j | X_0 = i_0, X_1 = i_1, \dots, X_{t-1} = i_{t-1}, X_t = i) = \\ = P(X_{t+1} = j | X_t = i) = p_{ij}, \end{aligned}$$

para todos os pares de estados (i, j) . Pode-se observar que a probabilidade de que a cadeia de Markov assuma um determinado valor futuro, quando o

seu estado atual é conhecido, não se altera se o seu comportamento passado for conhecido.

Se o espaço de estado é finito, as probabilidades de transição p_{ij} podem ser representadas por uma matriz de transição P . Os elementos de P satisfazem as seguintes condições:

i) $p_{ij} \geq 0, i, j \geq 0$;

ii) $\sum_{j=0}^{\infty} p_{ij} = 1, i = 0, 1, \dots$

EXEMPLO 4.10

Uma seguradora de automóveis oferece três categorias de desconto no prêmio, sendo elas:

- Nenhum desconto - Estado 1;
- 20% de desconto - Estado 2;
- 40% de desconto - Estado 3.

Considere as seguintes informações:

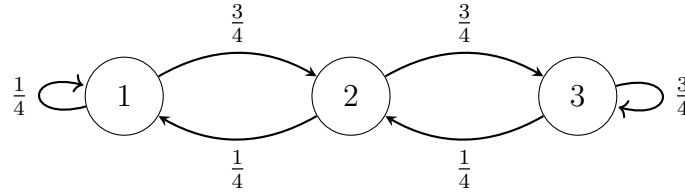
- Um novo segurado inicia no Estado 1;
- Após um ano sem ocorrência de sinistro, o segurado sobe um nível de desconto;

- Após um ano com ocorrência de um sinistro, o segurado desce um nível de desconto;
- Se o segurado está no Estado 1 e ocorre sinistro, ele permanece no Estado 1;
- Se o segurado está no Estado 3 e não ocorre sinistro, ele permanece no Estado 3;
- A probabilidade de que não ocorra sinistro em um ano é igual a $\frac{3}{4}$;
- A probabilidade de que a cadeia assuma um certo valor futuro, quando o seu estado atual é conhecido, não se altera com o conhecimento do seu comportamento passado (propriedade de Markov);
- As probabilidades de transição não dependem do tempo (cadeia de Markov estacionária).

A seguir, é apresentada a matriz de transição P .

$$P = \begin{bmatrix} \frac{1}{4} & \frac{3}{4} & 0 \\ \frac{1}{4} & 0 & \frac{3}{4} \\ 0 & \frac{1}{4} & \frac{3}{4} \end{bmatrix}$$

Pode-se apresentar a matriz de transição P através de um diagrama de transição, também denominado de grafo, como a seguir.



Com base em tais informações, calcule a probabilidade de obter um desconto de 40% no prêmio daqui a três anos, dado que atualmente não tem nenhum desconto no prêmio.

Solução:

O objetivo é encontrar a probabilidade da cadeia de Markov estacionária ir do estado atual i para o estado futuro j em n passos, ou seja, $p_{ij}^{(n)} = P(X_{n+k} = j | X_k = i)$, $n \geq 0$, $i, j \geq 0$. No caso, têm-se estado atual $i = 1$, estado futuro $j = 3$ e $n = 3$ (três anos). Pelo grafo, pode-se observar que existem duas possibilidades para ir do estado 1 ao estado 3, sendo elas: $1 \rightarrow 1 \rightarrow 2 \rightarrow 3$ e $1 \rightarrow 2 \rightarrow 3 \rightarrow 3$. Dessa maneira, obtém-se:

$$p_{13}^{(3)} = p_{11}p_{12}p_{23} + p_{12}p_{23}p_{33} = \frac{9}{16}.$$

Observação: As equações de Chapman-Kolmogorov fornecem um método para encontrar as probabilidades de transição em n passos. Tais equações são definidas como

$$p_{ij}^{(n+m)} = \sum_{k=0}^{\infty} p_{ik}^n p_{kj}^m$$

para todo $n, m \geq 0$ e i, j . Considerando que $P^{(n)}$ representa a matriz das probabilidades de transição a n passos, as equações de Chapman-Kolmogorov podem ser obtidas fazendo $P^{(n+m)} = P^{(n)}P^{(m)}$, encontrando $P^{(n)} = P^n$, ou seja, $P^{(n)}$ é obtida multiplicando n vezes a matriz de transição P .

Considerando a matriz de transição P do exemplo, obtém-se:

$$P^{(3)} = P^3 = PPP = \begin{bmatrix} p_{11}^{(3)} & p_{12}^{(3)} & p_{13}^{(3)} \\ p_{21}^{(3)} & p_{22}^{(3)} & p_{23}^{(3)} \\ p_{31}^{(3)} & p_{32}^{(3)} & p_{33}^{(3)} \end{bmatrix} = \begin{bmatrix} \frac{7}{64} & \frac{21}{64} & \frac{9}{16} \\ \frac{7}{64} & \frac{3}{16} & \frac{45}{64} \\ \frac{1}{16} & \frac{15}{64} & \frac{45}{64} \end{bmatrix},$$

encontrando $p_{13}^{(3)} = \frac{9}{16}$.

A distribuição de uma cadeia de Markov no tempo 0 (distribuição inicial) pode ser representada da seguinte maneira:

$$\alpha_i^{(0)} = P(X_0 = i),$$

em que $\sum_{i=0}^{\infty} \alpha_i^{(0)} = 1$ e $i \geq 0$. Considerando a matriz de transição P e a distribuição inicial representada pelo vetor linha $\alpha^{(0)} = (\alpha_1^{(0)}, \alpha_2^{(0)}, \dots)$, a distribuição no tempo n satisfaz:

$$\alpha^{(n)} = \alpha^{(0)} P^n.$$

Pode-se dizer que o estado i se comunica com o estado j quando a cadeia de Markov tem probabilidade positiva de atingir j , quando inicia a partir do estado i , ou seja, para um determinado valor de n , tem-se que

$$P(X_{m+n} = j | X_m = i) > 0.$$

Uma cadeia de Markov é dita ser irredutível se para todos os pares de estados (i, j) , tem-se que i e j se intercomunicam-se. Caso contrário, a cadeia é dita ser redutível.

Pode-se definir o período de um estado i como o maior divisor comum do conjunto de vezes que a cadeia pode retornar a i , iniciando em $X_0 = i$. Se o maior divisor comum for igual a 1, o estado i é dito ser aperiódico. Uma cadeia de Markov é dita ser periódica quando todos os estados são periódicos. Caso contrário, a cadeia é chamada de aperiódica.

Um estado i é dito ser recorrente se a cadeia retorna a i com probabilidade 1. Caso contrário, o estado i é dito ser transiente. Se o tempo esperado até a cadeia retornar a i é finito, então i é recorrente positivo. Uma cadeia de Markov que apresenta todos os estados positivos recorrentes e aperiódicos é denominada de cadeia ergódica. Uma cadeia ergódica apresenta probabilidades de transição que convergem para uma distribuição estacionária π , independente do estado inicial. Para todos os estados j , tem-se a distribuição estacionária $\pi = \{\pi_j\}$, independente do estado inicial i , sendo que:

$$(i) \quad \pi_j = \lim_{n \rightarrow \infty} p_{ij}^{(n)};$$

- (ii) $\pi_j \geq 0$;
- (iii) $\sum_{j=0}^{\infty} \pi_j = 1$;
- (iv) $\pi_j = \sum_{i=0}^{\infty} \pi_i p_{ij}$.

5.2 Integração Monte Carlo via Cadeias de Markov

Em estatística bayesiana, o interesse pode estar em encontrar a seguinte esperança

$$I = E[g(\Theta|X)] = \int g(\theta) \pi_{\Theta|X}(\theta|x) d\theta,$$

em que X e Θ são variáveis aleatórias, sendo que X é representada pela função de verossimilhança $f_{X|\Theta}(x|\theta)$ e Θ pela distribuição *a priori* $\pi_{\Theta}(\theta)$. Com base no teorema de Bayes, obtém-se a distribuição *a posteriori* $\pi_{\Theta|X}(\theta|x)$. Se m observações independentes de Θ forem obtidos partir de $\pi_{\Theta|X}(\theta|x)$, tem-se que $\hat{I} = \frac{1}{m} \sum_{i=1}^m g(\theta_i)$ converge para $E[g(\Theta|X)]$ pela lei forte dos grandes números. Uma cadeia de Markov pode ser utilizada para gerar observações dependentes da distribuição $\Theta|X$.

Se $\{\theta_0, \theta_1, \dots\}$ é uma realização de uma cadeia de Markov irreduzível e ergódica, com distribuição estacionária π , tem-se que

$$\hat{I}_m = \frac{1}{m} \sum_{t=0}^m g(\theta_t)$$

converge para $E[g(\Theta|X)]$ quando $m \rightarrow \infty$, em que $\Theta|X$ tem distribuição estacionária π .

5.3 Algoritmo de Metropolis-Hastings

O algoritmo de Metropolis-Hastings gera uma cadeia de Markov $\{X_t | t = 0, 1, \dots\}$, utilizando uma distribuição proposta q que depende do estado atual da cadeia (X_t) . Um novo valor da cadeia (X_{t+1}) é gerado a partir de $q(\cdot|X_t)$. Para isso, utiliza-se um ponto candidato Y gerado de $q(\cdot|X_t)$, sendo que se Y for aceito, então $X_{t+1} = Y$. O novo valor Y é aceito com probabilidade

$$\alpha(X_t, Y) = \min \left(1, \frac{\pi(Y)q(X_t|Y)}{\pi(X_t)q(Y|X_t)} \right),$$

em que π é a distribuição de interesse. Em estatística bayesiana, a distribuição de interesse é a distribuição *a posteriori* $\pi_{\Theta|X}(\theta|x)$.

O algoritmo de Metropolis-Hastings pode ser simplificado da seguinte maneira:

1. Escolha uma distribuição proposta $q(\cdot|X_t)$;
2. Inicie o contador de iterações $t = 0$;
3. Gere x_t da distribuição q ;
4. Gere um novo valor y da distribuição q ;

5. Calcule a probabilidade de aceitação $\alpha(x_t, y)$ e gere u da distribuição Uniforme $(0, 1)$;
6. Se $u \leq \alpha$, então aceite y e faça $x_{t+1} = y$, caso contrário, rejeite e faça $x_{t+1} = x_t$;
7. Incremente o contador t para $t + 1$ e retorne ao passo 4.

Em alguns casos, as cadeias geradas não convergem para a distribuição de interesse, sendo desconhecido o número de iterações necessárias para atingirem a convergência. O monitoramento da convergência da cadeia pode ser realizado por meio de gráficos que apresentam o comportamento sequencial da cadeia. Alguns critérios também podem ser utilizados, como os critérios de Geweke, Raftery-Lewis e Gelman-Rubin.

Observação: um caso particular do algoritmo de Metropolis-Hastings é o algoritmo de Gibbs, onde $q(X_t|Y)$ é obtida a partir de distribuições condicionais completas.

Capítulo 6

Aplicações

6.1 Modelo de risco coletivo

Considere a variável aleatória sinistro agregado S , sendo:

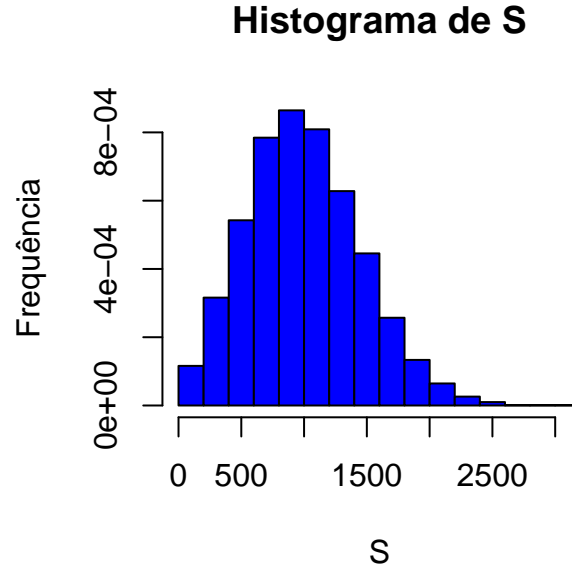
$$S = \sum_{i=1}^N X_i,$$

em que X_i representa o valor do i -ésimo sinistro individual e N representa o número de sinistros ocorridos (frequência de sinistros). Considerando a teoria do risco coletivo, os valores dos sinistros individuais serão assumidos independentes entre si e identicamente distribuídos, sendo X independente de N . Observe que a variável aleatória S está condicionada as variáveis aleatórias N e X . Existem vários métodos para aproximar a distribuição de S , como por exemplo, aproximação por uma distribuição normal e aproximação por uma distribuição gama translada. A seguir, será apresentada a aproximação da distribuição de S por meio de simulação de variáveis

aleatórias.

Considere que $N \sim \text{Poisson}(\lambda)$, com $\lambda = 5$ e, $X \sim \text{Gama}(\alpha, \beta)$, com $\alpha = 4000$ e $\beta = 20$. A seguir, são apresentados os comandos para gerar $m = 10000$ valores de S .

```
calc_sin_agreg=function(m,lambda,alpha,beta){  
  sin_agreg=numeric()  
  for(i in 1:m){  
    n=rpois(1,lambda)  
    x=rgamma(n,alpha,beta)  
    sin_agreg[i]=sum(x)  
  }  
  
  #histograma dos valores simulados de S.  
  hist(sin_agreg,main="Histograma de S",xlab="S",  
    ylab="Frequência",col="blue",freq=FALSE)  
  
  summary(sin_agreg)  
}  
  
calc_sin_agreg(10000,5,4000,20)
```



##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	0.0	607.5	997.2	999.9	1211.1	3190.4

Em teoria do risco coletivo, sabe-se que $E(S) = E(N)E(X)$. No exemplo, $E(S) = \lambda \frac{\alpha}{\beta} = 1000$, sendo que pela aproximação da distribuição de S por meio de simulação de variáveis aleatórias, obteve-se uma média de S aproximadamente igual a 1000.

6.2 Modelo clássico de ruína

O modelo clássico de Cramér-Lundberg (modelo clássico de ruína) é definido da seguinte forma:

$$U(t) = u + P(t) - S(t), t \geq 0,$$

em que:

$U(t)$ é a reserva da seguradora até o tempo t ;

u é o capital inicial da seguradora no tempo $t = 0$;

$P(t)$ representa a taxa constante de prêmios recebidos até o tempo t ;

$S(t) = \sum_{i=1}^{N(t)} X_i$ representa o sinistro agregado, sendo que X_i é o valor do i -ésimo sinistro e $N(t)$ é o número de sinistros até o tempo t .

Considere que $N \sim Poisson(\lambda)$, com $\lambda = 100$, $X \sim Gama(\alpha, \beta)$, com $\alpha = 500$ e $\beta = 0,5$, $P(t) = 100000$, $u = 100000$ e o período de simulação igual a 120 meses. A seguir, são apresentados os comandos para encontrar a frequência de ruína observada.

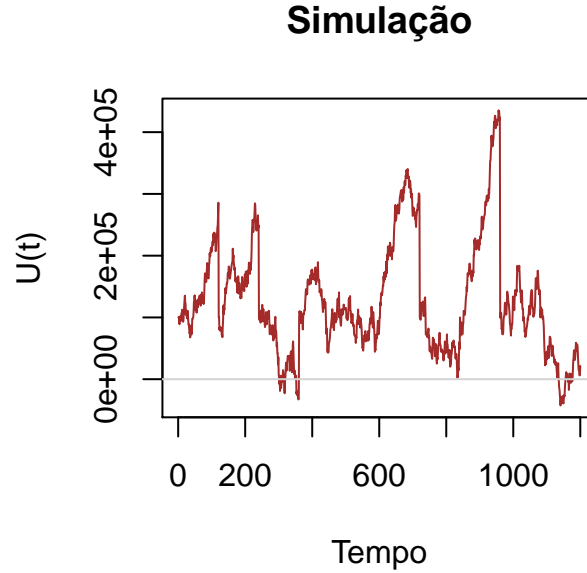
```
mc<-c()  
mct<-c()  
P<-100000  
a<-1  
  
for(i in 1:1000)  
{  
  R<-(1:120)*0  
  Uo<-100000  
  for(t in 1:120)
```

```
{
  mct[a]<-Uo
  a<-a+1
  N<-rpois(1,100)
  X<-rgamma(N,500,0.5)
  S<-sum(X)
  Ut<-Uo+P-S
  if(Ut<0){R[t]<-1}
  Uo<-Ut
}
mc[i]<-sum(R)
}

#frequência de ruína observada.
fequRuina<-length(which(mct<0))/length(mct)
fequRuina

## [1] 0.08074167

#gráfico referente ao nível de reserva ao longo do tempo.
plot(mct[1:1200],type='l',xlab="Tempo", ylab="U(t)",
main="Simulação", col="brown")
abline(h=0,col='lightgray')
```



Por meio de simulação de variáveis aleatórias, a frequência de ruína observada foi aproximadamente igual a 9%.

6.3 Resseguro não proporcional

Num contrato de resseguro não proporcional é fixado o nível de prioridade ou nível de retenção da seguradora (M), sendo que a resseguradora se compromete a assumir os valores dos sinistros que excederem M até o limite de cobertura acordado. Para uma indenização X , a seguradora cede à resseguradora o valor $X^{ced} = \max\{X - M, 0\}$, sendo que a seguradora retém $X^{ret} = \min\{X, M\}$. Considerando um modelo de risco coletivo, o prêmio de resseguro é dado por $\Pi^{RES} = E(S^{ced}) = E(N)E(X^{ced})$, em que S^{ced} é o sinistro agregado cedido à resseguradora.

Considere que $X \sim \text{Pareto}(\alpha, \beta)$ e que $N \sim \text{Poisson}(\lambda)$. Encontre o prêmio

de resseguro para $M = 4$ quando $\lambda = 10$, $\alpha = 3$ e $\beta = 2$, usando 100000 simulações.

Para gerar valores de X , vamos utilizar o método da transformação integral. Sabendo que $F_X(x) = 1 - \left(\frac{\beta}{x}\right)^\alpha$, para $x \geq b$, $\alpha > 0$ e $\beta > 0$, pelo método da transformação integral obtém-se $x = \beta(1 - u)^{-\frac{1}{\alpha}}$, sendo que u vem da distribuição $U \sim Uniforme(0, 1)$.

A seguir, são apresentados os comandos para encontrar o prêmio de resseguro Π^{RES} .

```
premioress<-function(n,lambda,alfa,beta,M){  
  set.seed(123)  
  u<-runif(n)  
  x<-beta*((1-u)^(-1/alfa))  
  xced<-pmax(x-M,0) #cálculo dos máximos  
  premio<-lambda*mean(xced)  
  return(premio)  
}  
  
premioress(100000,10,3,2,4)  
  
## [1] 2.479637
```

6.4 Cálculo do prêmio utilizando estatística bayesiana

Considere o modelo de risco coletivo, em que a variável aleatória sinistro agregado S é dada como a seguir:

$$S = \sum_{i=1}^N X_i,$$

em que X_i representa o valor do i -ésimo sinistro individual e N representa o número de sinistros ocorridos (frequência de sinistros). Considerando que os valores dos sinistros individuais são independentes entre si e identicamente distribuídos, sendo X independente de N , tem-se que $E(S) = E(N)E(X)$.

Com relação ao número de sinistros ocorridos, considere que $N \sim Poisson(\Lambda)$ e $\Lambda \sim Gama(\alpha_1, \beta_1)$. Com relação ao valor do sinistro individual, considere que $X \sim Gama(\alpha_2, \beta_2)$. Encontre o valor do prêmio Π com base no princípio do valor esperado, sendo:

$$\Pi = (1 + \theta) E(S),$$

em que θ é o carregamento de segurança.

Para o cálculo de Π , é necessário encontrar $E(S)$, sendo que $E(S) = E(N)E(X) = \hat{\lambda} \left(\frac{\alpha_2}{\beta_2} \right)$. Para tanto, $\hat{\lambda} \approx E(\Lambda|N)$, que é a média da distribuição *a posteriori* de Λ dado X .

Pelo teorema de Bayes, obtém-se:

$$\pi_{\Lambda|N}(\lambda|n) \propto f_{N|\Lambda}(n|\lambda)\pi_{\Lambda}(\lambda),$$

$$\pi_{\Lambda|N}(\lambda|n) \propto \left[\prod_{i=1}^t \lambda^{n_i} e^{-\lambda} \right] \left[\lambda^{\alpha_1} e^{-\lambda\beta_1} \right],$$

$$\pi_{\Lambda|N}(\lambda|n) \propto \left[\lambda^{t\bar{n}} e^{-t\lambda} \right] \left[\lambda^{\alpha_1} e^{-\lambda\beta_1} \right],$$

$$\pi_{\Lambda|N}(\lambda|n) \propto \lambda^{\alpha_1+t\bar{n}-1} e^{-\lambda(t+\beta_1)}.$$

Dessa maneira, tem-se que $\Lambda|N \sim Gama(\alpha_1 + t\bar{n} - 1, t + \beta_1)$, com $E(\Lambda|N) = \frac{\alpha_1+t\bar{n}-1}{t+\beta_1}$.

Com isso, $E(S) = \left(\frac{\alpha_1+t\bar{n}-1}{t+\beta_1} \right) \left(\frac{\alpha_2}{\beta_2} \right)$.

Considere que o conhecimento prévio sobre o número médio de sinistros é descrito pela distribuição gama com $\alpha_1 = 10$ e $\beta_1 = 2$, ou seja, $\Lambda \sim Gama(10, 2)$. O conhecimento obtido por uma amostra de tamanho igual a 45 ($t = 45$) e $\bar{n} = 3,5$ é descrito pela função de verossimilhança. Com base em tal amostra, tem-se que $X \sim Gama(4000, 5)$.

Dessa maneira, $\Lambda|N \sim Gama(166, 5; 47)$, com $E(\Lambda|N) = 3,54$. Com isso, $E(S) = 2832$. Para um carregamento de seguranaça igual 0,3 ($\theta = 0,3$), o prêmio Π é igual a 3681,6.

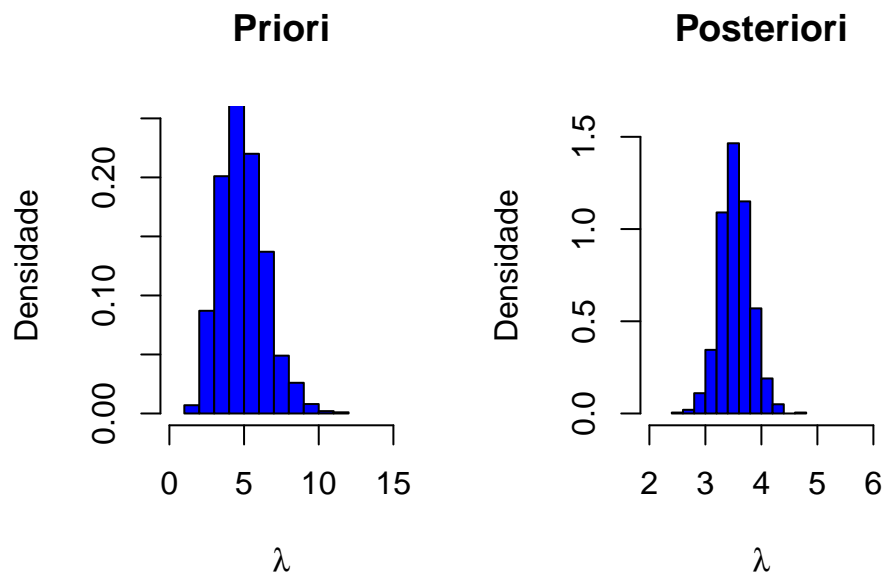
A seguir, são apresentados os comandos para encontrar o prêmio Π , utilizando simulação de variáveis aleatórias. Como resultado, o valor do prêmio encontrado via simulação de variáveis aleatórias é próximo ao valor encontrado sem simulação.

```
set.seed(123)
par(mfrow=c(1,2))

#distribuição a priori
priori<-rgamma(1000,10,2)
hist(priori,freq=F,xlab=expression(lambda),ylab="Densidade",
      xlim=c(0,15),ylim=c(0,0.25),col="blue",main="Priori")

#amostra de N, com tamanho igual a 45
n<-rpois(45,5)
#amostra de X, com tamanho igual ao número total de sinistros
x<-rgamma(sum(n),4000,5)

#distribuição a posteriori
posteriori<-rgamma(1000,166.5,47)
hist(posteriori,freq=F,xlab=expression(lambda),
      ylab="Densidade",xlim=c(2,6),ylim=c(0,1.6),col="blue",
      main="Posteriori")
```



```
#Esperança do sinistro agregado S
ES<-mean(posteriori)*mean(x)
ES

## [1] 2830

#Prêmio - Princípio do valor esperado
Premio<-(1+0.3)*ES
Premio

## [1] 3679
```

6.5 Cálculo do Valor em Risco (VaR)

Na análise de modelos de perdas, pode-se estar interessado em quantificar perdas extremas, utilizando quantis da distribuição de perdas. Uma função quantil é a inversa da função de distribuição acumulada. Considerando que

$$F_X(x_\delta) = P(X \leq x_\delta) = \delta,$$

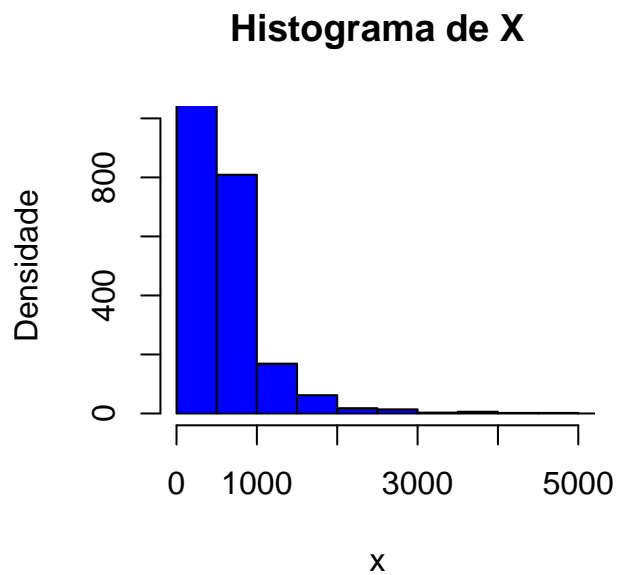
obtem-se $x_\delta = F_X^{-1}(\delta)$, sendo $F_X^{-1}(\cdot)$ definida como função quantil e x_δ como o δ -quantil. O Valor em Risco (VaR) a um nível de probabilidade δ , denotado por $\text{VaR}_\delta(X)$, é o δ -quantil da variável aleatória X .

Considerando que $X \sim \text{LogNormal}(\mu, \sigma^2)$, a seguir são apresentados os comandos para encontrar $\text{VaR}_{0,5}(X)$, $\text{VaR}_{0,95}(X)$ e $\text{VaR}_{0,975}(X)$. Para tanto, foi utilizada a função *quantile()*, podendo escolher um determinado tipo de algoritmo para encontrar os quantis. Por exemplo, o algoritmo tipo 4 é baseado na interpolação dos dados. Foram simulados 10000 valores de X , com $\mu = 5$ e $\sigma^2 = 1$.

```
set.seed(123)
x<-exp(rnorm(10000,5,1))
summary(x)

##      Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
##   3.173   76.099  146.776  244.344  291.007 6958.837

hist(x,freq=T,xlab="x", ylab="Densidade",xlim=c(0,5000),
      ylim=c(0,1000),col="blue",main="Histograma de X")
```



```
niveis<-c(0.5,0.95,0.975)
quantile(x,niveis,type=4)
```

```
##      50%      95%      97.5%
## 146.7701 766.4615 1044.3156
```

Capítulo 7

Referências

BOLAND, P. J. *Statistical and probabilistic methods in actuarial science*. Boca Raton: Chapman & Hall/CRC, 2007. 368 p.

BOLFARINE, H.; SANDOVAL, M. C. *Introdução à inferência estatística*. 2. ed. Rio de Janeiro: SBM, 2010. 172 p.

BOLVIKEN, E. *Computation and modelling in insurance and finance*. Cambridge: International Series on Actuarial Science, 2014. 685 p.

BRAUN, J. *A first course in statistical programming with R*. New York: Cambridge University Press, 2007. 163 p.

BUENO, L. P. *Métodos estatísticos básicos em seguros gerais*. Rio de Janeiro: ENS-CPES, 2017. 244 p.

CASELLA, G.; ROBERT, C. P. *Introduction Monte Carlo Methods with R*. New York: Springer, 2010. 284 p.

CHARPENTIER, A (ed.). *Computacional actuarial science with R*. Boca Raton: Chapman & Hall/CRC, 2014. 650 p.

PIRES, M. D.; FERREIRA, L.; COSTA, L. H.; MARQUES, R. *Teoria do risco atuarial: fundamentos e conceitos*. Curitiba: CRV, 2020. 260 p.

PIRES, M. D.; MARQUES, R.; FERREIRA, L.; COSTA, L. H. *Fundamentos da matemática atuarial: vida e pensões*. Curitiba: CRV, 2021. 304 p.

DICKSON, D. C. M. *Insurance risk and ruin*. Cambridge: Cambridge University Press, 2005. 229 p.

EHLERS, R. S. *Introdução a inferência bayesiana*. Disponível em: <http://www.leg.ufpr.br/~paulojus/CE227/ce227/>. Acesso em: 07 mai. 2023.

FERREIRA, D. F. *Estatística básica*. 2. ed. Lavras: UFLA, 2009. 664 p.

FERREIRA, D. F. *Estatística computacional em Java*. Lavras: UFLA, 2013. 695 p.

FERREIRA, D. F. *Fundamentos de probabilidade*. Lavras: UFLA, 2020. 707 p.

FINAN, M. B. *An introductory guide in the construction of actuarial models: A preparation for the actuarial Exam C/4*. Arkansas Tech University, 2017. 802 p.

KASS, R. et al. *Modern actuarial risk theory: using R*. New York:

Springer Verlag, 2008. 381 p.

KLUGMAN, S. A. *Loss models: from data to decisions*. 4. ed. New Jersey: J. Wiley, 2012. 514 p.

KORN, R.; KORN, E.; KROISANDT, G. *Monte Carlo methods and models in finance and insurance*. Boca Raton: CRC Press/Taylor Francis, 2010. 470 p.

MORALES, J. C. C.; CAUSIL, C. J. B. *Introducción a la estadística bayesiana*. Instituto Tecnológico Metropolitano, 2018. 222 p.

MORETTIN, P. A. *Estatística básica*. 7. ed. São Paulo: Saraiva, 2012. 540 p.

PACHECO, R. *Matemática atuarial de seguros de danos*. São Paulo: Atlas, 2014. 282 p.

PETERNELLI, L. A.; MELLO, M. P. *Conhecendo o R: uma visão mais que estatística*. Viçosa: UFV, 2013. 222 p.

RIBEIRO JUNIOR, P. J. *Introdução ao ambiente estatístico R*. Disponível em: <http://leg.ufpr.br/~paulojus/embrapa/Rembrapa/>. Acesso em: 10 jun. 2023.

RIZZO, M. L. *Statistical computing with R*. 2. ed. Boca Raton: Chapman & Hall/CRC, 2019. 490 p.

ROSS, S. M. *Simulation*. 6. ed. Cambridge: Academic Press, 2022. 336 p.

RUBINSTEIN, R. Y.; KROESE, D. P. *Simulation and the Monte Carlo Method*. 2. ed. New Jersey: J. Wiley, 2008. 384 p.

TSE, Y. K. *Nonlife actuarial models: Theory, Methods and Evaluation*. Cambridge: International Series on Actuarial Science, 2009. 542 p.

WEISHAUS, A. *Study manual for Exam C/Exam 4: Construction and evaluation of actuarial models*. 17. ed. Actuarial Study Materials, 2014. 1664 p.