

Projeto 1 - Geração de Números Pseudo-Aleatórios

Diogo Pedrosa - 94358

Rita Ferrolho - 88822

1. a)

- Pretende-se usar o método da rejeição para gerar valores do modelo contínuo X com a seguinte f.d.p., correspondente à função alvo:

$$f(x) = \frac{2}{2\pi} e^{(-\frac{x^2}{2})} I_{[0,+\infty[}(x)$$

- Para a variável proponente Y vamos usar o modelo com distribuição Exp(1) que vai ter a seguinte f.d.p.:

$$g(y) = \begin{cases} \lambda e^{-\lambda y}, & \text{se } y > 0 \\ 0, & \text{outros } y \end{cases}$$

- Agora, vai-se determinar a constante c para podermos implementar o método da rejeição:

$$h(y) = \frac{f(y)}{g(y)} = \frac{2}{2\pi} e^{(-\frac{y^2}{2}+y)} I_{[0,+\infty[}(y)$$

- Usando a derivada vai-se descobrir o valor máximo:

$$h'(y) = \frac{2}{2\pi} (-y + 1) e^{(-\frac{y^2}{2}+y)}, \text{ para } y > 0$$

$$h'(y) = 0 \Rightarrow y = 1$$

- Calcular o valor de c:

$$c = h(y=1) = \frac{2}{2\pi} e^{(\frac{1}{2})} = 1.3155$$

Para gerar u , NPA $u \sim U(0, 1)$, deve-se aceitar $x = y$ com a seguinte condição:

$$u < \frac{h(y)}{c} = e^{(-\frac{y^2}{2}+y-\frac{1}{2})} I_{[0,+\infty[}(y)$$

Deste modo, tem-se as condições para se implementar o algoritmo do método de rejeição (software utilizado foi o R):

- Definiu-se $n = 1000$, número de variáveis (NPA) a gerar de X;
- O número de tentativas previsto até serem gerados 1000 valores da distribuição X é $n.c = 1315$;

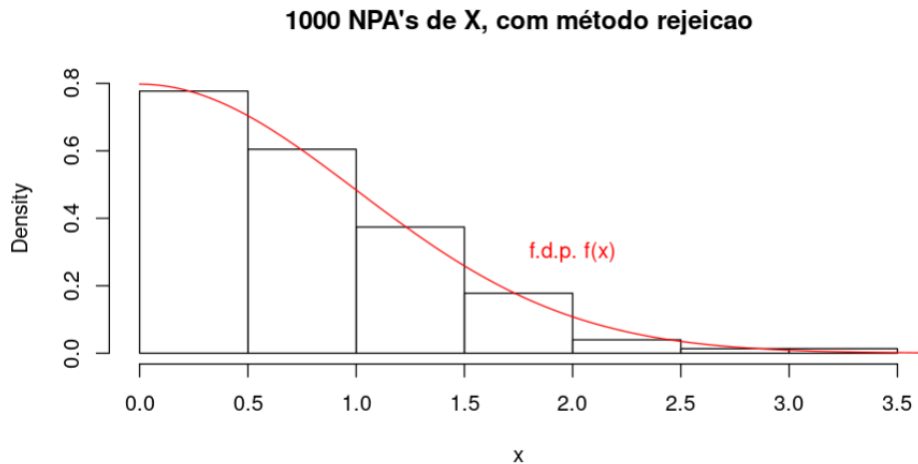


Figura 1 - Histograma que representa valores x gerados através do Método de Rejeição com f.d.p. (função alvo) sobreposta, $f(x)$.

O número esperado de observações para as $n=1000$ gerações é dado por $n.c \approx 1315$, sendo c o número de NPA's gerados, em média, para obter uma única observação de x .

Repetindo 500 vezes o algoritmo anterior obtivemos, em média, aproximadamente 759.39 observações para cada $n=1000$ gerações, com um erro de, aproximadamente, 0.584.

1. b)

A referida distribuição de X , $f(x)$, corresponde à distribuição $N(0,1)$ truncada, podendo-se verificar da seguinte forma:

- Seja $g(x)$ a f.d.p da distribuição $N(\mu, \sigma^2)$:

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}, \quad x \in \mathbb{R}, \mu \in \mathbb{R}, \sigma > 0$$

- Para $\mu = 0$ e $\sigma^2 = 1$:

$$g(x) = \frac{1}{\sqrt{2\pi}} e^{\left(-\frac{x^2}{2}\right)}, \quad x \in \mathbb{R}$$

- Ao truncar $g(x)$ com $P(x > 0)$, obtém-se $f(x)$, c.q.d:

$$f(x) = \frac{g(x)}{P(x \geq 0)} \Leftrightarrow f(x) = \frac{\frac{1}{\sqrt{2\pi}} e^{\left(-\frac{x^2}{2}\right)}}{1 - P(x < 0)}, \quad \text{para } x \geq 0$$

$$\Leftrightarrow f(x) = \frac{\frac{1}{\sqrt{2\pi}} e^{\left(-\frac{x^2}{2}\right)}}{1 - \Phi(0)}, \quad \text{para } x \geq 0$$

$$\Leftrightarrow f(x) = \frac{\frac{1}{\sqrt{2\pi}} e^{\left(-\frac{x^2}{2}\right)}}{\frac{1}{2}}, \quad \text{para } x \geq 0$$

$$\Leftrightarrow f(x) = \frac{2}{2\pi} e^{(-\frac{x^2}{2})} I_{[0,+\infty[}(x), \text{ c.q.d.}$$

O gráfico obtido neste método alternativo foi o seguinte:

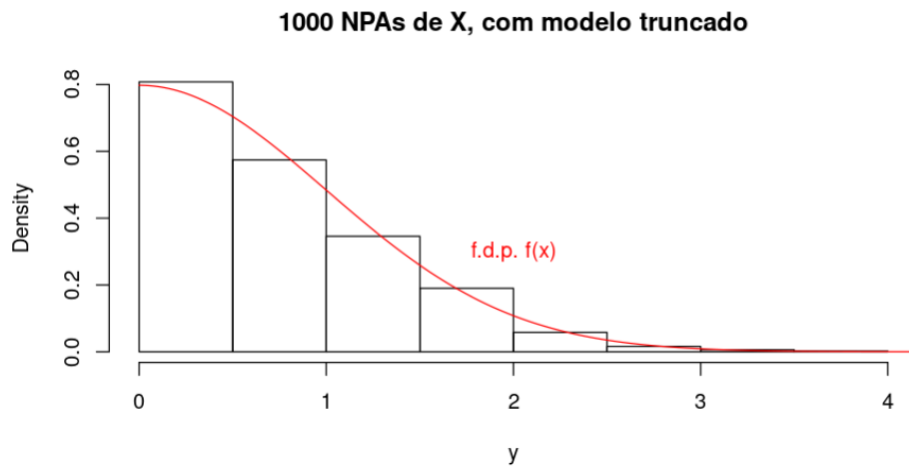


Figura 2 - Histograma que representa valores x gerados através do modelo truncado com f.d.p. (função alvo) sobreposta, $f(x)$.

O erro obtido neste exercício foi, aproximadamente, 0.019, ou seja, foi inferior ao erro obtido com o Método de Rejeição, usado no exercício anterior.

2. a)

Neste exercício pretende-se, começando por $f(x)$, chegar à distribuição T-Student, ou seja:

$$f(x) = \frac{\Gamma(\frac{v+1}{2})}{\sqrt{v\pi}\Gamma(\frac{v}{2})} \left(1 + \frac{x^2}{v}\right)^{-\frac{v+1}{2}}$$

Para cada distribuição temos as seguintes f.d.p.:

- $f_{N(0, \frac{1}{\tau})}(x) = \frac{\sqrt{\tau}}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2\tau}$
- $f_{G(\alpha, \alpha)}(\tau) = \frac{\alpha^\alpha}{\Gamma(\alpha)} \tau^{\alpha-1} e^{-\alpha\tau}$

Temos portanto:

$$\begin{aligned} f(x) &= \int_0^{+\infty} f_{N(0, \frac{1}{\tau})}(x) f_{G(\alpha, \alpha)}(\tau) d\tau \\ &= \int_0^{+\infty} \frac{\sqrt{\tau}}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2\tau} \frac{\alpha^\alpha}{\Gamma(\alpha)} \tau^{\alpha-1} e^{-\alpha\tau} d\tau \\ &= \frac{1}{\sqrt{2\pi}\Gamma(\alpha)} \int_0^{+\infty} \sqrt{\tau} e^{-\frac{1}{2}x^2\tau} \alpha^\alpha \tau^{\alpha-1} e^{-\alpha\tau} d\tau \end{aligned}$$

$$= \frac{1}{\sqrt{2\pi}\Gamma(\alpha)} \int_0^{+\infty} \alpha^\alpha \sqrt{\tau} \tau^{\alpha-1} e^{-\tau\left(\alpha + \frac{x^2}{2}\right)} d\tau$$

Fazendo $u = \tau\left(\alpha + \frac{x^2}{2}\right) \Leftrightarrow \tau = \frac{u}{\left(\alpha + \frac{x^2}{2}\right)}$

E alterando $d\tau = \frac{1}{\left(\alpha + \frac{x^2}{2}\right)} du$,

$$= \frac{1}{\sqrt{2\pi}\Gamma(\alpha)} \int_0^{+\infty} \alpha^\alpha \frac{\sqrt{u}}{\sqrt{\alpha + \frac{x^2}{2}}} \frac{u^{\alpha-1}}{\left(\alpha + \frac{x^2}{2}\right)^{\alpha-1}} e^{-u} \frac{1}{\left(\alpha + \frac{x^2}{2}\right)} du$$

$$= \frac{1}{\sqrt{2\pi}\Gamma(\alpha)} \int_0^{+\infty} \frac{1}{\sqrt{\alpha + \frac{x^2}{2}}} \frac{\alpha^\alpha}{\left(\alpha + \frac{x^2}{2}\right)^\alpha} u^{\left(\alpha + \frac{1}{2}\right)-1} e^{-u} du$$

$$= \frac{1}{\sqrt{2\pi}\Gamma(\alpha)} \int_0^{+\infty} \frac{\alpha^\alpha}{\left(\alpha + \frac{x^2}{2}\right)^{\left(\alpha + \frac{1}{2}\right)}} u^{\left(\alpha + \frac{1}{2}\right)-1} e^{-u} du$$

$$= \frac{\Gamma(\alpha + 1/2)}{\sqrt{2\pi}\Gamma(\alpha)\sqrt{\alpha}} \left(1 + \frac{x^2}{2\alpha}\right)^{-(\alpha + 1/2)}$$

$$= \frac{\Gamma(\alpha + 1/2)}{\sqrt{2\pi}\alpha\Gamma(\alpha)} \left(1 + \frac{x^2}{2\alpha}\right)^{-(\alpha + 1/2)}$$

Como $v=2\alpha$,

$$f(x) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\Gamma\left(\frac{v}{2}\right)} \left(1 + \frac{x^2}{2v}\right)^{-\left(\frac{v+1}{2}\right)} \text{ (T-Student com } v/2 \text{)}$$

Que representa uma distribuição T-Student, tal como queríamos demonstrar.

Curiosamente, a distribuição T-Student também pode ser escrita da seguinte forma, em que B corresponde à função Beta:

$$f(x) = \frac{1}{\sqrt{v}B\left(\frac{1}{2}, \frac{v}{2}\right)} \left(1 + \frac{x^2}{v}\right)^{-\left(\frac{v+1}{2}\right)}$$

2.b) Em todas as alíneas deste exercício vai-se assumir que $\alpha = 2$ e $T = 2$.

2.b.i)

```
# dimensao 100
n = 100
normal_gama100 = numeric(100)
for (i in 0:n) {
  taugamma = rgamma(n, shape = 2, rate = 2)
  xnorm = rnorm(n, 0, 1/taugamma)
  normal_gama100[i] = mean(xnorm)
}
mean(normal_gama100)
```

Figura 3 – Gerar valores de distribuição normal gama para dimensão $n = 100$.

Para gerar a mesma distribuição, mas com dimensão $n = 1000$ e $n = 10000$, repetiu-se o mesmo algoritmo, mas para esses valores de n .

Quanto maior o n , mais próximo é o valor da média de 0, ou seja, a média tende a ser 0 para $n \rightarrow +\infty$. O valor teórico da média de uma distribuição normal gama é zero, por isso os valores gerados experimentalmente não contradizem com o valor teórico.

2.b.ii) Distribuição T-Student, com $v = 2 * \alpha$ graus de liberdade:

```
## ex 2bii ##

xstudent100 = rt(100, 4)
mean(xstudent100)

xstudent1000 = rt(1000, 4)
mean(xstudent1000)

xstudent10000 = rt(10000, 4)
mean(xstudent10000)
```

Figura 4 - Gerar valores de distribuição T-Student para dimensões $n = 100$, $n = 1000$ e $n = 10000$, respetivamente.

Recorrendo à função `rt()` que representa a função T-Student, incluída no R, a resolução deste exercício torna-se trivial. Novamente, o valor da média aproxima-se de zero com o aumento da dimensão n , o que não contradiz o valor teórico da média da distribuição T-Student, que também é 0.

2.b.iii)

A média das distribuições das alíneas i e ii tendem a ser próximas do valor teórico da média dessas distribuições ($\mu = 0$) à medida que n aumenta, para $v > 0$.

Conclui-se assim que os valores gerados experimentalmente adequam-se aos valores teóricos.

2.b.iv)

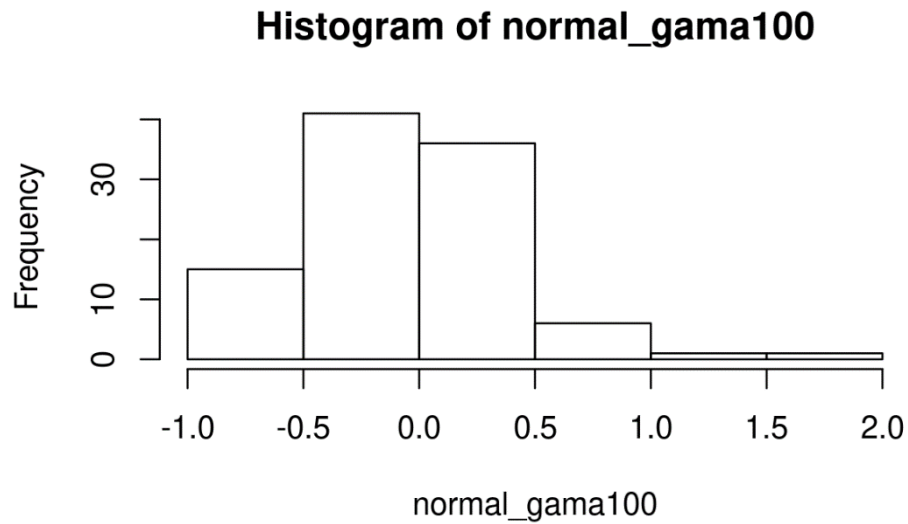


Figura 5 – Histograma da distribuição normal gama, para $n = 100$.

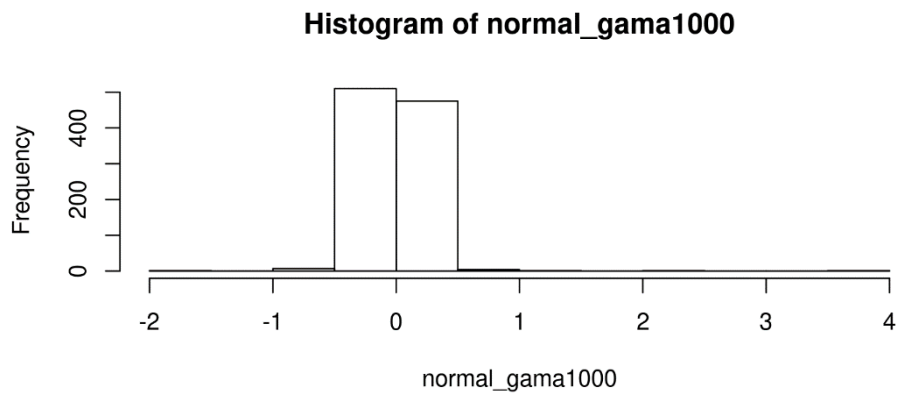


Figura 6 – Histograma da distribuição normal gama, para $n = 1000$.

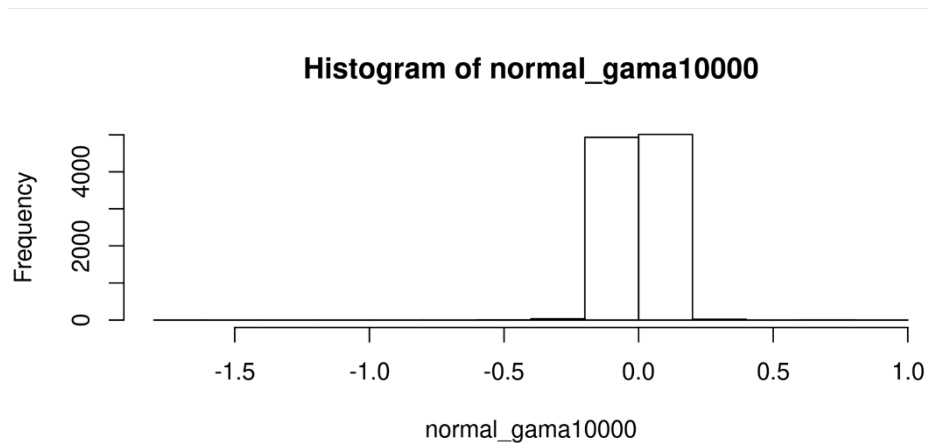


Figura 7 - Histograma da distribuição normal gama, para $n = 10000$.

Pelos histogramas das figuras 5, 6 e 7 é possível reparar que, com o aumento de n , menor é a variedade dos valores dos NPA's gerados (variância). Com estes histogramas induz-se que, para $n \rightarrow +\infty$, o valor de todos os NPA's gerados é nulo.

Se observarmos os gráficos das misturas para estes histogramas, obtemos os seguintes resultados:

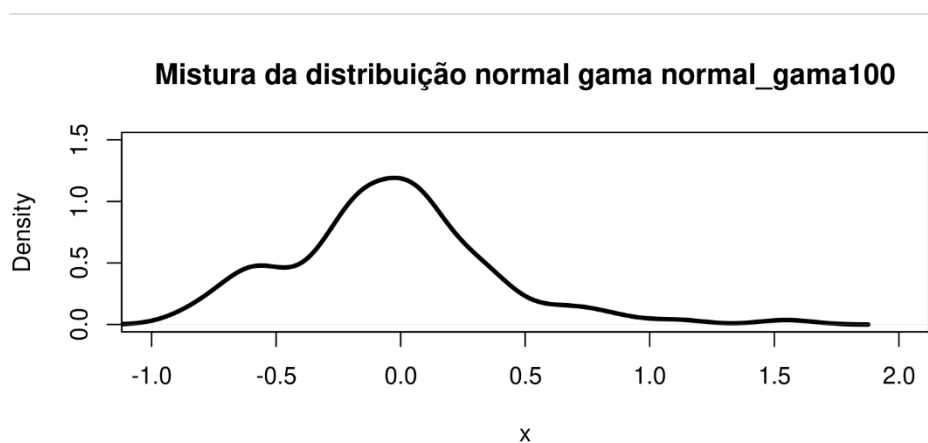


Figura 8 - Gráfico da mistura da distribuição normal gama para $n = 100$.

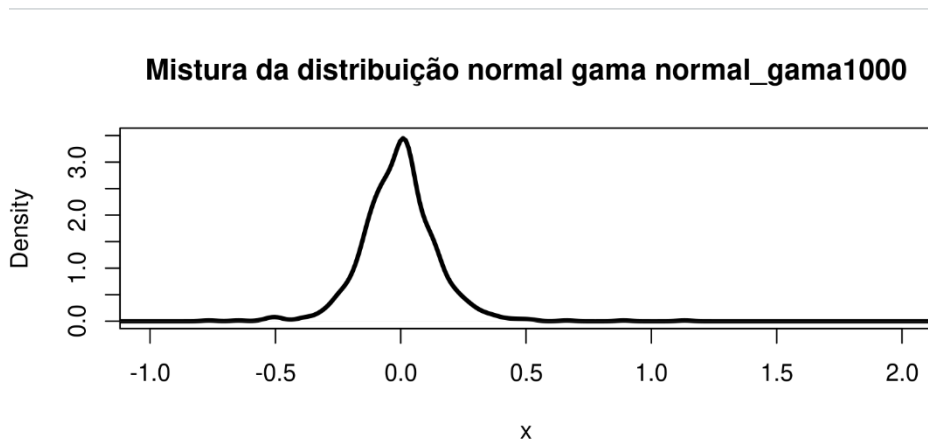


Figura 9 – Gráfico da mistura da distribuição normal gama para $n = 1000$.

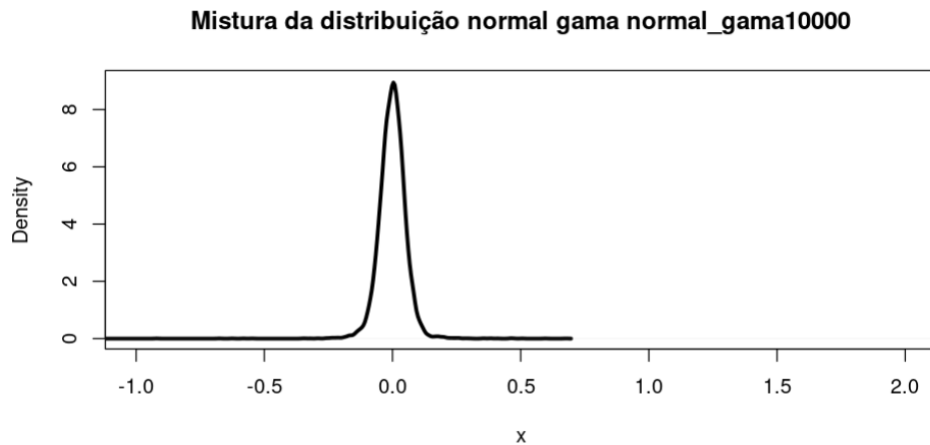


Figura 10 - Gráfico da mistura da distribuição normal gama para $n = 10000$.

A análise das misturas leva-nos a uma conclusão idêntica à análise dos histogramas. A variedade dos valores dos NPA's gerados (variância) reduz com o aumento da dimensão n . Para $n \rightarrow +\infty$, esta função teria função nula em todos os valores de x , com exceção da média, $x = 0$, cujo valor tende para $+\infty$. Com o aumento do n , é cada vez menos evidente que o gráfico representa uma mistura.

Para os histogramas das distribuições de T-Student, com sobreposição da f.d.p.:

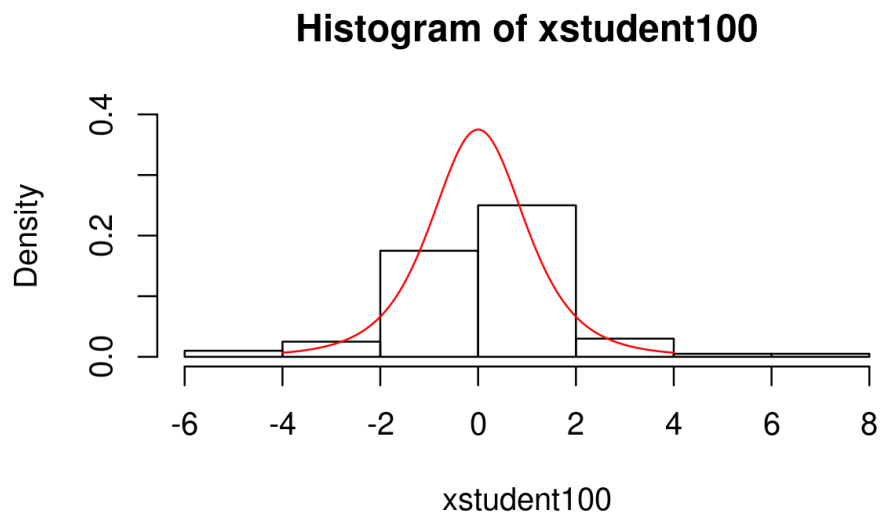


Figura 11 – Histograma da distribuição T-Student, para $n = 100$.

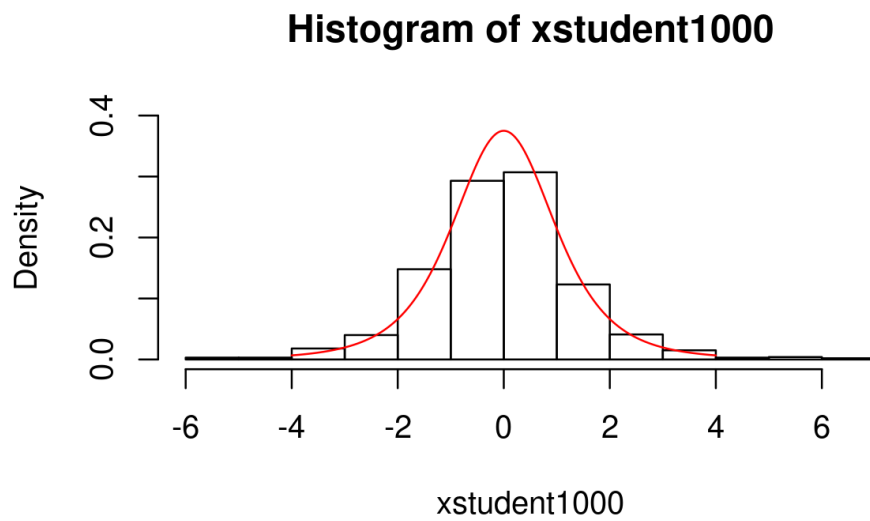


Figura 12– Histograma da distribuição T-Student, para $n = 1000$.

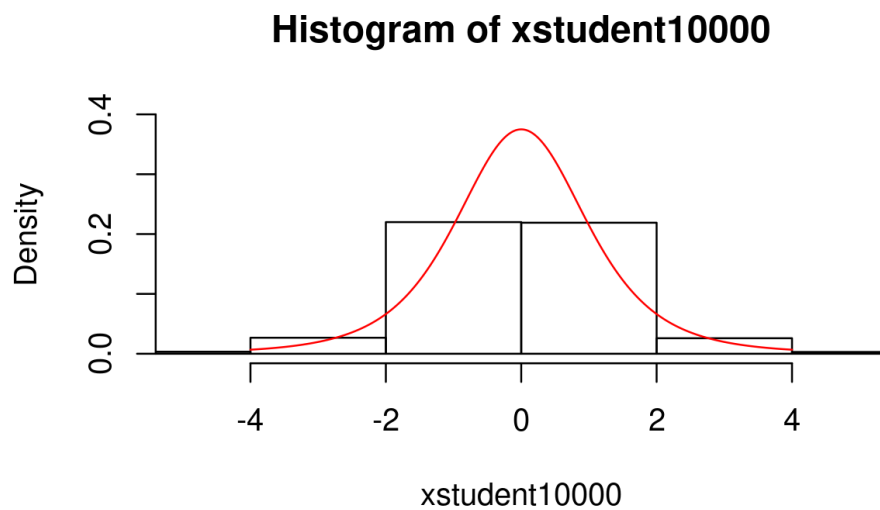


Figura 13 – Histograma da distribuição T-Student, para $n = 10000$.

Por observação destes gráficos, o valor da média dos NPA's gerados tende a aproximar-se do valor da média teórico ($\mu = 0$) quanto maior for o valor de n .

3.a)

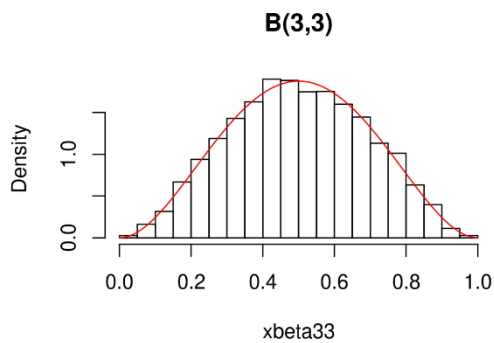


Figura 11: Histograma 5000 NPA's gerados de uma B(3,3)

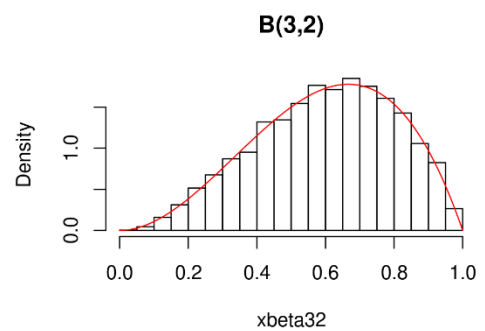


Figura 12: Histograma 5000 NPA's gerados de uma B(3,2)

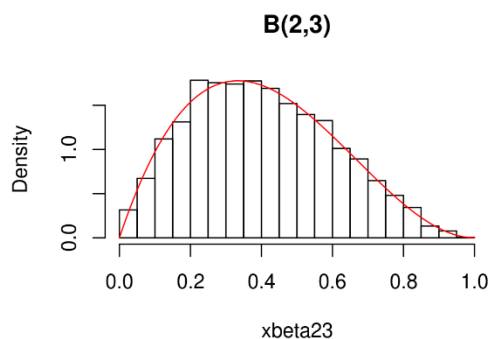


Figura 13: Histograma 5000 NPA's gerados de uma B(2,3)

3.b)

Calculamos o valor de b_1 no script em R do nosso trabalho.

Depois fomos calcular o coeficiente de assimetria ($\sqrt{b_1}$) para cada um dos Betas. De notar que os valores obtidos podem variar conforme os NPA's gerados.

De referir, que de acordo com os histogramas gerados no exercício 3a) que se espera que o coeficiente de assimetria de B(3,3) tenha um valor próximo de zero e que o coeficiente de assimetria de B(2,3) e B(3,2) deverão ter valores que não estão próximos de zero.

Summary dos coeficientes de assimetria gerados 50 vezes:

B(3,3)

Mínimo	1º quartil	Mediana	Média	3º quartil	Máximo
-0.044562	-0.011138	0.005955	0.001922	0.018076	0.041237

B(3,2)

Mínimo	1º quartil	Mediana	Média	3º quartil	Máximo
-0.3215	-0.2991	-0.2852	-0.2848	-0.2706	-0.2422

B(2,3)

Mínimo	1º quartil	Mediana	Média	3º quartil	Máximo
0.2435	0.2695	0.2858	0.2834	0.2983	0.3324

Média dos coeficientes de assimetria:

Para B(3,3): $\sqrt{b_1} = 0.001922 \approx 0$

Para B(3,2): $\sqrt{b_1} = -0.2848 \approx -0.3$

Para B(2,3): $\sqrt{b_1} = 0.2834 \approx 0.3$

De modo geral, os valores obtidos experimentalmente não contradizem os valores teóricos.

- B(3,3) é uma distribuição simétrica, o que significa que o valor do seu coeficiente de assimetria deve ser 0. O valor desse coeficiente obtido experimentalmente é muito próximo de zero, conforme as expectativas.
- B(3,2) é uma distribuição assimétrica negativa, o que implica que o valor teórico do seu coeficiente de assimetria é negativo. Experimentalmente, o valor desse coeficiente obtido é negativo (coeficiente de assimetria = -0.2848), conforme as expectativas.
- B(2,3) é uma distribuição assimétrica positiva, o que implica que o valor teórico do seu coeficiente de assimetria é positivo. Além disso, esse coeficiente deve ser simétrico ao da distribuição B(3,2). Experimentalmente, isso verificou-se, pois o coeficiente obtido foi 0.2834, que é muito próximo do valor simétrico do coeficiente obtido experimentalmente para B(3,2).