Projeto 1

Estatística Computacional e Simulação – 41165

2022/2023

Docente: Professora Doutora Isabel Pereira

Gonçalo Freitas - 98012

Tiago Alvim - 95584

Vasco Costa – 97746

Exercício 1

a) Querendo-se gerar valores do modelo contínuo X com uma função de densidade de probabilidade f tal que:

$$f(x) = \frac{2}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) I_{[0.+\infty[}(x)$$

Para usar o método da rejeição é necessário calcular primeiramente o valor de c, tal que:

$$h(x) = \frac{f(x)}{g(x)} \leq c$$

Em que g é a função proponente da variável Y, neste caso $Y \sim Exp(1)$, então $g(x) = e^{-x}$. Assim, calculando o valor de c obtém-se:

$$\frac{f(x)}{g(x)} \leq c \iff \frac{\frac{2}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}}{e^{-x}} \leq c \iff \frac{2}{\sqrt{2\pi}} e^{\frac{-x^2+2x}{2}} \leq c$$

Pretendendo-se verificar quais os extremos da função h, neste caso o seu máximo, basta descobrir o zero da sua derivada.

$$h'(x) = 0 \iff \frac{2}{\sqrt{2\pi}} (x - \frac{x^2}{2})' e^{x-\frac{x^2}{2}} = 0 \iff \frac{\sqrt{4}}{\sqrt{2\pi}} (1-x) e^{x-\frac{x^2}{2}} = 0$$

$$\iff \sqrt{\frac{2}{\pi}}\, e^{x-\frac{x^2}{2}} (x-1) = 0 \iff x = 1$$

De forma a verificar se este é um máximo ou um mínimo um método possível de o concluir é com o cálculo do sinal da segunda derivada, indicativo da concavidade da função.

$$h''(x) = \sqrt{\frac{2}{\pi}} \left(2x e^{x-\frac{x^2}{2}} - e^{x-\frac{x^2}{2}} x^2\right)$$

$$h''(1) = \sqrt{\frac{2}{\pi}} \left(2 e^{1-\frac{1^2}{2}} - e^{1-\frac{1^2}{2}} 1^2\right) = \sqrt{\frac{2e}{\pi}} > 0$$

Pelo que será um máximo uma vez que a segunda derivada é positiva.

Calculando então $h(1)$, obtém-se:

$$c = h(1) = \frac{2}{\sqrt{2\pi}} \sqrt{e}$$

Na Figura 1 é possível verificar o histograma dos NPA's gerados através do método da rejeição com a linha vermelha para comparar com a função de distribuição de probabilidade alvo. Daqui consegue-se verificar que o histograma aproxima em certa forma a linha vermelha, mas não de uma forma perfeita.

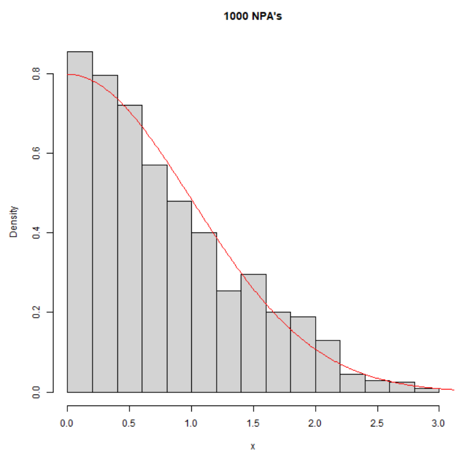

Figura 1 - Histograma dos 1000 NPA's gerados com a função de distribuição de probabilidade alvo a vermelho.

Em média seriam esperadas 1315 iterações necessárias para obter 1000 valores. No entanto, foram necessárias 1123 assim pode-se afirmar que o método aplicado foi eficiente.

b) Primeiramente, tem de se demonstrar que $f(x)$ é uma função de probabilidade de uma $N(0,1)$ truncada, para isso tem de se verificar que existe a função de distribuição para:

$f_{X|\ a\leq x\leq b}\ (x) = \frac{f_X(x)}{P(x\in(a,b))}$ e truncando o intervalo para $a = -\infty$ e $b = 0$ pois são partes da distribuição normal que não fazem parte do domínio de $f_X(x)$.

$$f_{X|\ a\leq x\leq b}\ (x) = \frac{f_X(x)}{F(0)-F(-\infty)} = \frac{\frac{2}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right) I_{[0.+\infty[}(x)}{\frac{1}{2}}$$

Neste caso, foram precisas 2001 iterações para gerar 1000 NPA's, mais iterações do que aquelas para a alínea anterior, sendo assim menos eficiente.

Exercício 2

Considerando a distribuição variada com a função densidade de probabilidade conjunta $f_{(X,Y)}(x,y) = \begin{cases} 2 & ,0 < x < y < 1 \\ 0, outros(x,y) \in \mathbb{R}^2 \end{cases}$ pode-se utilizar o método multivariado da transformada inversa.

$$f_Y(y) = \int_0^y 2\,dx = 2y \quad ,0 < y < 1$$

$$F_Y(y) = \int_0^y 2t\,dt = y^2 \quad ,0 < y < 1$$

$$f_{X|Y}(x|Y=y) = \frac{f(x,y)}{f_Y(y)} = \frac{2}{2y} = \frac{1}{y} \quad ,0 < x < y\ e\ 0 < y < 1$$

$$F_{X|Y}(x|Y=y) = \int_0^x \frac{1}{y}dt = \frac{x}{y} \quad ,0 < x < y\ e\ 0 < y < 1$$

Deste modo é possível através de duas distribuições normais gerar os pontos como se pode verificar na Figura 2.

$$\begin{cases} U_1 = F_Y(X) \\ U_2 = F_{X|Y}(Y|X) \end{cases} \Leftrightarrow \begin{cases} U_1 = Y^2 \\ U_2 = \frac{X}{Y} \end{cases} \Leftrightarrow \begin{cases} Y = \sqrt{U_1} \\ X = U_2 Y \end{cases}$$

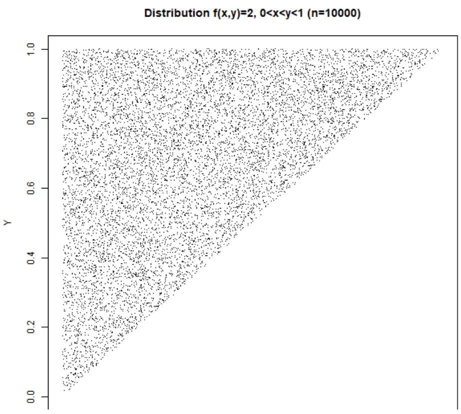

Figura 2 - Representação da distribuição bivariada f(x,y)

Exercício 3

a) Pretende-se mostrar que,

$$f(x) = \int_0^{+\infty} f_{N(0,1/\tau)}(x) f_{Ga(\alpha,\alpha)}(\tau)\,d\tau,\ \ x \in R$$

representa a função densidade de uma t-Student com $v = 2\alpha$ graus de liberdade.

Sabendo que $f(x) = \frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ é a função de densidade de probabilidade, f.d.p, da distribuição Gaussiana[1] de valor médio esperado μ e de desvio padrão σ. Assim como a f.d.p da distribuição Gamma[2] é $f(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}$ de parâmetro de forma $\alpha = k$ e de escala inversa de $\beta = \frac{1}{\theta}$. Chega-se às funções de densidade:

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

$$f_{Ga(\alpha,\alpha)}(\Gamma) = \frac{\alpha^\alpha}{\Gamma(\alpha)}\tau^{\alpha-1}e^{-\alpha\tau}$$

Verificando-se também que a função de densidade de probabilidade da distribuição t-Student é da forma $f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{x^2}{\nu}\right)^{-\left(\frac{\nu+1}{2}\right)}$, com ν graus de liberdade. Esta, ao ter $\nu = 2\alpha$ leva a $f(x) = \frac{\Gamma\left(\frac{2\alpha+1}{2}\right)}{\sqrt{2\alpha\pi}\Gamma(\alpha)}\left(1+\frac{x}{2\alpha}\right)^{-\left(\frac{2\alpha+1}{2}\right)}$

Assim, o integral apresentado é:

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

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

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

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

Efetuando a transformação $z = \tau\left(\alpha+\frac{x^2}{2}\right)$ o que leva $d\tau = \frac{1}{\left(\alpha+\frac{x^2}{2}\right)}dz$, substituindo no integral:

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

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

[1] Formula obtida em: https://pt.wikipedia.org/wiki/Distribui%C3%A7%C3%A3o_normal
[2] Formula obtida em: https://pt.wikipedia.org/wiki/Distribui%C3%A7%C3%A3o_gama

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

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

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

Verificando-se que $\Gamma(z)=\int_0^{+\infty}e^{-t}t^{z-1}\,dt$

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

O que representa a f.d.p de uma t-Student com $v = 2\alpha$, tal como se queria demonstrar.

a) Considerando $\alpha = 3$
   i. No código
   ii. No código
   iii. Segue em baixo

Tabela 1 - Comparativo das estatísticas sumárias amostrais para os valores obtidos e para os valores teóricos

N	Média			Mediana			Variância		
	Normal-Gama	R	Teoria	Normal-Gama	R	Teoria	Normal-Gama	R	Teoria
100	0.002995	0.1012	0	0.016585	0.1956	0	0.03626396	1.640768	$\frac{v}{v-2}=\frac{3}{3-2}=3$
1000	0.0006962	-0.03545		-0.0002404	-0.01409		0.004761161	1.36386	
10000	6.959e-05	0.03001		-2.809e-05	0.01860		0.0004539935	1.482451	

Da Tabela 1[3] é possível verificar os resultados obtidos com os dois métodos, com a decomposição Normal-Gama e através da função do R e com os valores teóricos de uma distribuição t-Student. Verifica-se que os valores obtidos com a Normal-Gama apresentam tanto uma média como uma mediana mais próxima dos valores teóricos que a função do R. No que toca à variância esta não apresenta os valores certos em nenhuma das situações.

[3] Valores teóricos obtidos em https://en.wikipedia.org/wiki/Student%27s_t-distribution

Exercício 4

a) Nas Figura 3 é possível verificar os histogramas para as distribuições Beta, Beta (3,3), Beta (2,3) e Beta (3,2), respetivamente.

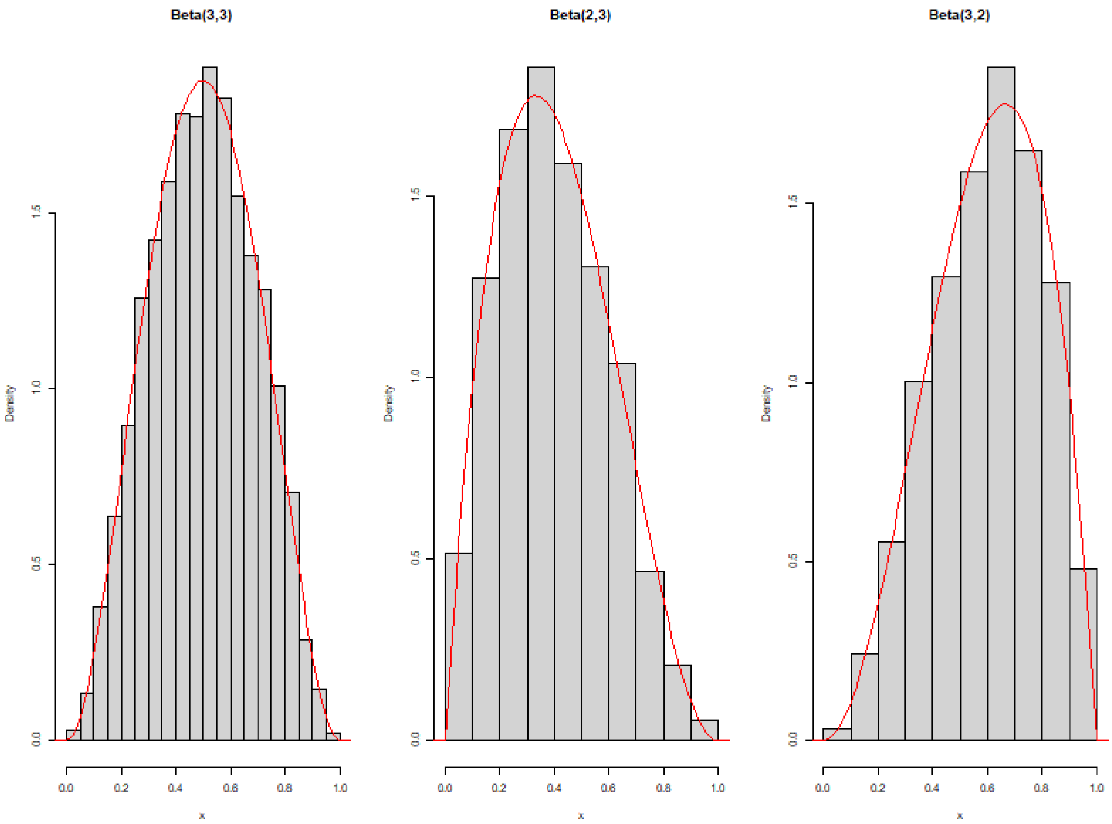

Figura 3 - Histogramas das distribuições de Beta (3,3), Beta (2,3) e Beta (3,2), respetivamente da esquerda para a direita, com a sua f.d.p respetiva sobreposta.

Verifica-se da Figura 3 que cada distribuição é única e que à exceção da Beta (3,3) as distribuições são assimétricas, verificando-se que a linha a vermelho aproxima as contagens no histograma.

b) Repetindo o processo 50 vezes de gerar as 3 distribuições Beta com 5000 valores calculando o seu coeficiente de assimetria e analisando os valores sumários dos valores registados para os valores dos coeficientes obtém-se os valores da Tabela 2.

Tabela 2 - Valores sumários dos 50 coeficientes de assimetria obtidos para cada uma das distribuições Beta, Beta (3,3), Beta (2,3) e Beta (3,2).

	B - Beta (3,3)	B - Beta (2,3)	B - Beta (3,2)
1º Quartil	0.35	0.24	0.45
Mediana	0.50	0.38	0.61
Média	0.50	0.39	0.59
3º Quartil	0.64	0.54	0.75

Verificando os valores da Tabela 2, apura-se, como esperado, que para a distribuição Beta (3,3) que esta é simétrica, mediana e média de 0.5. Para a distribuição Beta (2,3) que esta tem o máximo mais à esquerda, mediana de 0.38 e média de 0.39, e que a distribuição Beta (3,2) tem o máximo mais à direita, mediana de 0.61 e média de 0.59. Confirmando-se assim a simetria e a assimetria esperada em cada caso.