FT043 - Fundamentos da Ciência de Dados

Atividade 2

Gustavo Nicolau Gonçalves

RA: 265556

Parte teórica

1.

$f_X(x_1,x_2)$		X_1 = 0	X_1 = 1	X_1 = 2	
X_2	0	0	0,25	0	0,25
	1	0,25	0	0,25	0,50
	2	0	0,25	0	0,25
		0,25	0,50	0,25	

a) Obtenha as distribuições marginais das variáveis X_1 e X_2

X_1 →

0	1	2
0,25	0,50	0,25

X_2 →

0	0,25
1	0,50
2	0,25

b-) Calcule $P(X_1 > X_2)$

$$P(x_1 > x_2) = \cancel{P(x} \; P(1,0) + P(2,0) + P(2,1)$$

$$P(x_1 > x_2) = 0{,}25 + 0 + 0{,}25 = \underline{0{,}50}$$

c-) Calcule $E\{x_1\}$ e $E\{x_2\}$

$$\underline{E\{x\} = \sum_{i=1}^{n} x_i \cdot p(x_i)}$$

$$\hookrightarrow E\{x_1\} = 0 \cdot \underbrace{\cancel{p_{x_1}(0)}}_{0,25} + 1 \cdot \underbrace{p_{x_1}(1)}_{0,50} + 2 \cdot \underbrace{p_{x_2}(2)}_{0,25} = 0{,}5 + 0{,}5 = \underline{1}$$

$$\hookrightarrow E\{x_2\} = 0 \cdot \cancel{p_{x_2}(0)} + 1 \cdot \overbrace{p_{x_2}(1)} + 2 \cdot \overbrace{p_{x_2}(2)} = 0{,}5 + 0{,}5 = \underline{1}$$

$$\underline{E\{x_1\} = E\{x_2\} = 1}$$

d) Calcule o desvio padrão das variáveis X_1 e X_2

$$\underbrace{\sigma}_{\text{desvio padrão}} = \sqrt{\underbrace{\sigma^2}_{\text{Variância}}}$$

→ Primeiro vamos calcular a Variância (σ^2) de X_1 e X_2

$$\sigma^2 = \underbrace{E\{x^2\}} - [E\{x\}]^2$$

$$\hookrightarrow E\{x_1^2\} = \cancel{0^2 \cdot p_{x_1}(0)} + 1^2 \cdot \underbrace{p_{x_1}(1)}_{0,5} + 2^2 \underbrace{p_{x_1}(2)}_{0,25} = \underline{\underline{1,5}}$$

$$E\{x_2^2\} = E\{x_1^2\} = 1,5$$

$$\hookrightarrow \sigma_{x_1}^2 = E\{x_1^2\} - [E\{x_1\}]^2 = 1,5 - 1 = 0,5$$

$$\hookrightarrow \underline{\sigma_{x_2}^2 = \sigma_{x_1}^2 = 0,5} /\!/$$

→ Agora calculamos o Desvio Padrão (σ) de X_1 e X_2

$$\sigma_{x_1} = \sqrt{0,5} \cong \underline{0,707106}$$

$$\underline{\sigma_{x_1} = \sigma_{x_2} \cong 0,707106} /\!/$$

e-) Calcule o coeficiente de correlação entre X_1 e X_2

$$\rho = \frac{E\{X_1 . X_2\} - \overbrace{E\{X_1\}}^{1} . \overbrace{E\{X_2\}}^{1}}{\underbrace{\sigma_{X_1} . \sigma_{X_2}}_{\sqrt{0,5} \cdot \sqrt{0,5} = \underline{0,5}}}$$

- $E\{X_1 . X_2\} = \sum_{i=0}^{2} x_i y_i . p(x_1, x_2)$

$$= 2.1.0,25 + 1.2.0,25 + \underbrace{0 \ldots + 0}_{\text{demais termos são multiplicados por zero}} = 1$$

$$\rho = \frac{1 - 1}{0,5} \Rightarrow \underline{\rho = 0} \longrightarrow$$ $\underline{X_1 \text{ e } X_2 \text{ são}}$ ~~independentes~~

f-) Verifique se as variáveis X_1 e X_2 são estatisticamente independentes

Se $P(X_1 | X_2) = P(X_1) \longrightarrow$ são estatisticamente independentes

$$f(1,1) = f_{X_1}(1) . f_{X_2}(1) \ ? \quad \begin{cases} f(1,1) = 0 \\ f_{X_1}(1) = 0,50 \\ f_{X_2}(1) = 0,50 \end{cases} \Rightarrow \underline{\underline{0 \neq 0,25}}$$

Encontrado um contra exemplo, dessa forma X_1 e X_2 são <u>dependentes estatisticamente</u>

Para verificar a correlação, podemos calcular a covariância entre X_1 e X_2:

$$\underset{X_1 X_2}{COV} = \underbrace{E\{X_1 . X_2\} - E\{X_1\} . E\{X_2\}}_{\text{podemos utilizar o resultado do item e.)}} = \underline{\underline{0}}$$

→ Com esse resultado concluímos que as V.A. X_1 e X_2 são <u>descorrelacionadas.</u>

2.

Os resultados numéricos anexos foram obtidos através da execução do codigo "atividade 2.py" anexo.

Para o item a) foi gerado gráfico das distribuições marginais de X_1 e X_2, os demais itens tiveram suas respostas diretamente no terminal.

É possível observar uma tendência de aproximação entre o numérico e teórico, principalmente ao rodar o código mais de uma vez, com o aumento do número de amostras.

É notável, entretanto, que para alguns casos, e dado a natureza das operações como o item (e), para poucas o resultado ~~se aprox~~ ficou mais próximo ao teorico que em quantidades mais elevadas de amostras. Em outras execuções o mesmo não ocorreu, e o que se percebeu foi que para uma menor quantidade de amostras houve uma maior variação de respostas a cada execução, enquanto que para maiores quantidades de amostras menor variação de resposta a cada execução.

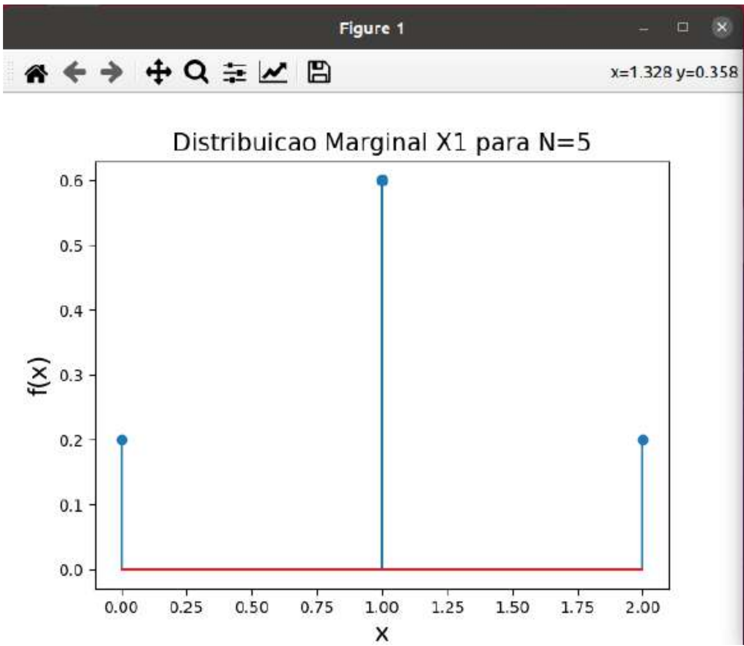
Figure 1
x=1.328 y=0.358
Distribuicao Marginal X1 para N=5
f(x)
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
x

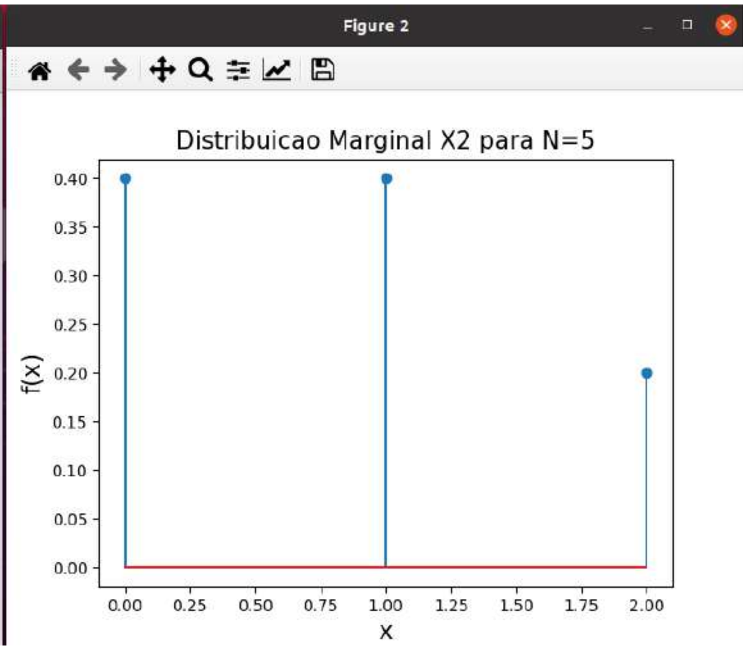
Figure 2
Distribuicao Marginal X2 para N=5
f(x)
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
x

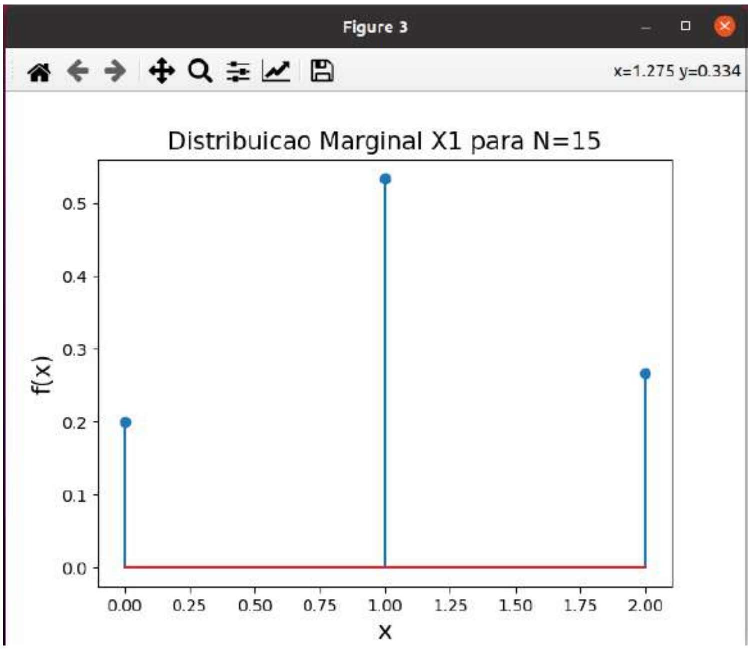
Figure 3
x=1.275 y=0.334
Distribuicao Marginal X1 para N=15
f(x)
x

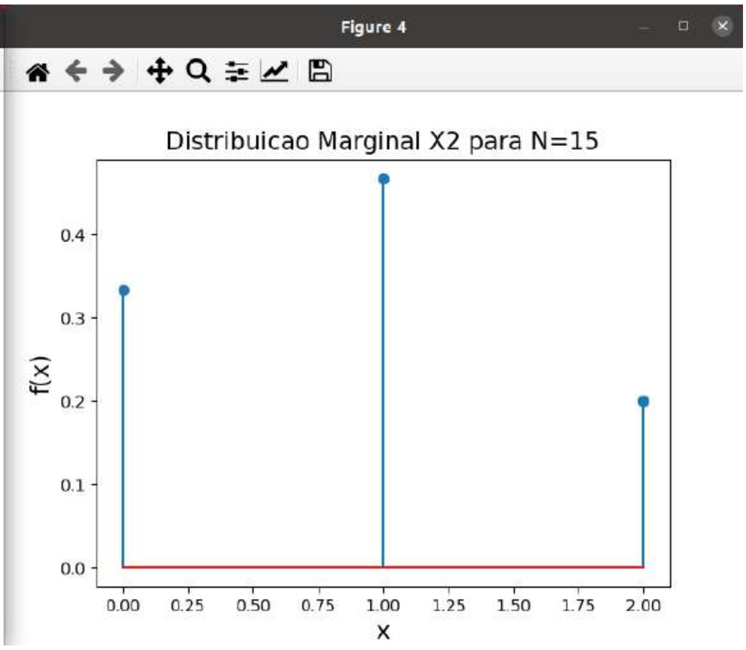
Figure 4
Distribuicao Marginal X2 para N=15
f(x)
x

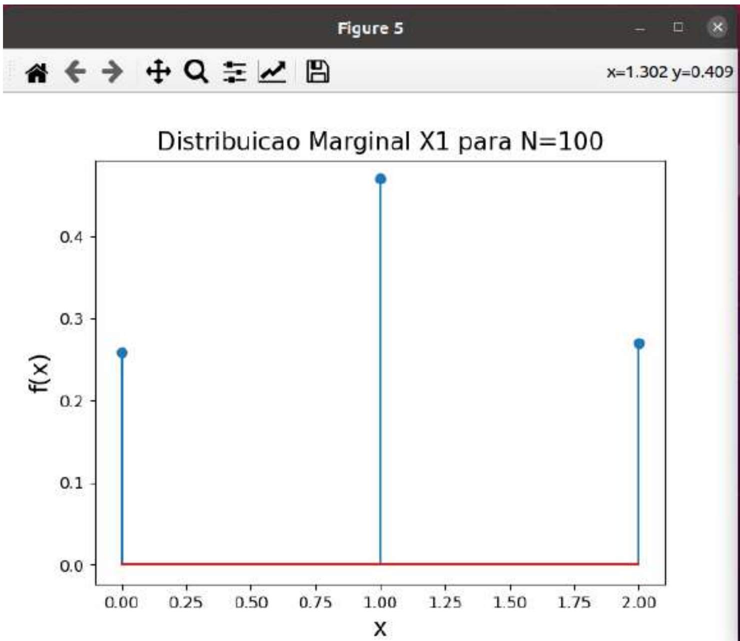
Figure 5
x=1.302 y=0.409
Distribuicao Marginal X1 para N=100
f(x)
x
0.0
0.1
0.2
0.3
0.4
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00

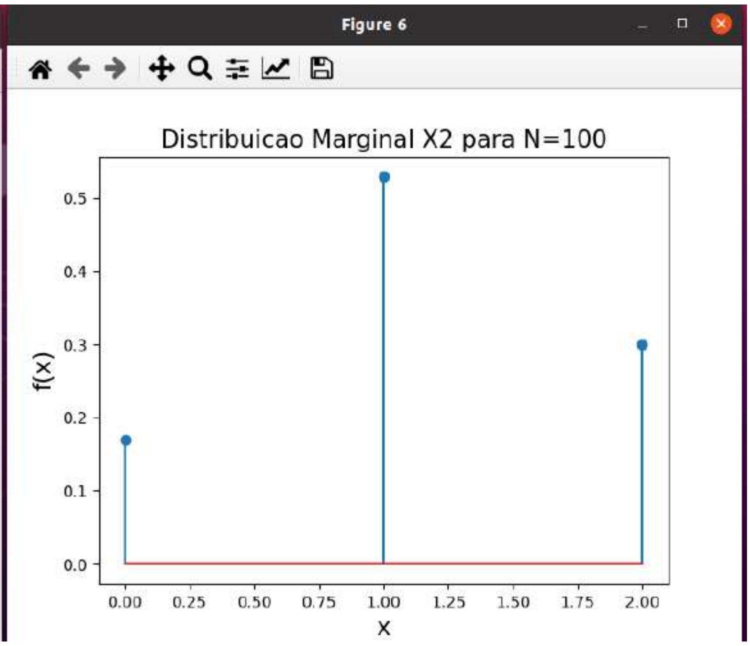
Figure 6
Distribuicao Marginal X2 para N=100
f(x)
x
0.0
0.1
0.2
0.3
0.4
0.5
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00

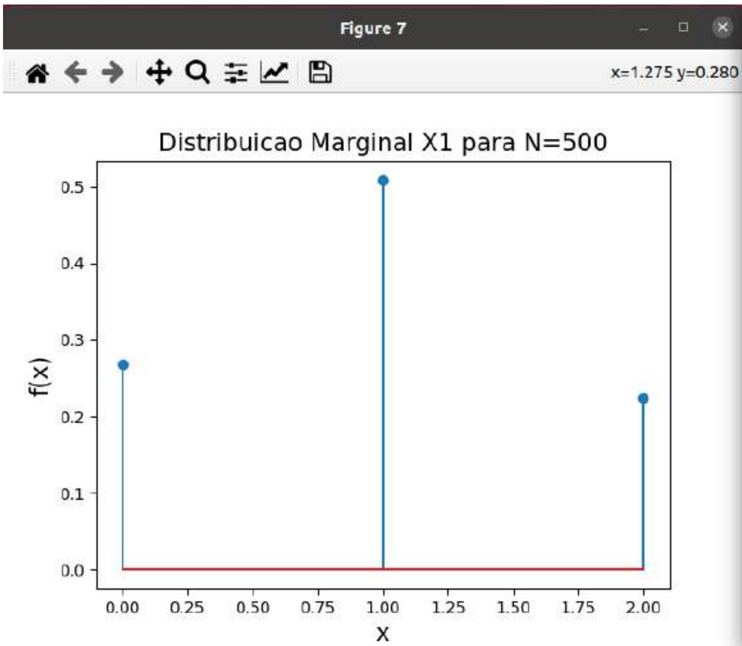
Figure 7
x=1.275 y=0.280
Distribuicao Marginal X1 para N=500
0.5
0.4
0.3
0.2
0.1
0.0
f(x)
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
X

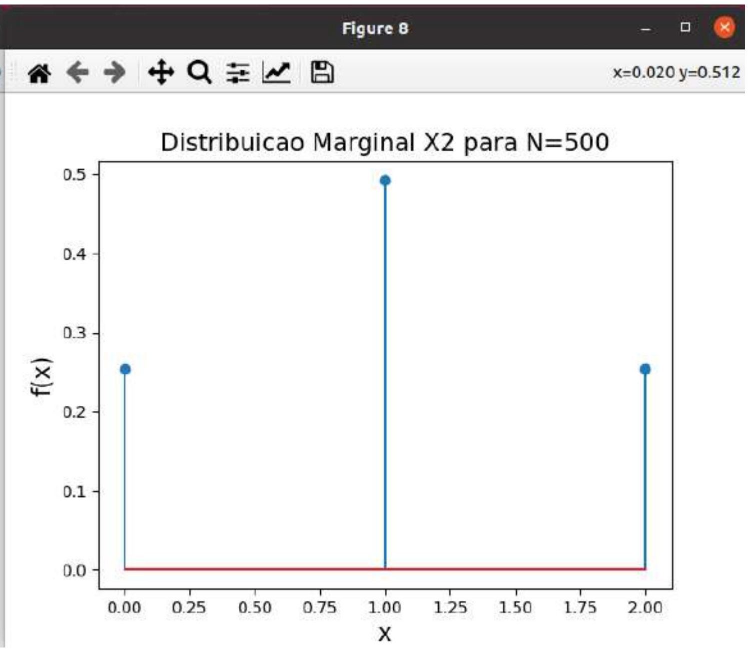
Figure 8
x=0.020 y=0.512
Distribuicao Marginal X2 para N=500
0.5
0.4
0.3
0.2
0.1
0.0
f(x)
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
X

```
Questão 2:
Item b
Frequência em que X1 > X2 (N=5)=  0.6
Frequência em que X1 > X2 (N=15)=  0.6
Frequência em que X1 > X2 (N=100)=  0.44
Frequência em que X1 > X2 (N=500)=  0.478


Item c
Média X1 e X2 (N=5) =  [1.  0.8]
Média X1 e X2 (N=15) =  [1.06666667 0.86666667]
Média X1 e X2 (N=100) =  [1.01 1.13]
Média X1 e X2 (N=500) =  [0.956 1.   ]


Item d
Desvio Padrão X1 e X2 (N=5) =  [0.63245553 0.74833148]
Desvio Padrão X1 e X2 (N=15) =  [0.67986927 0.71802197]
Desvio Padrão X1 e X2 (N=100) =  [0.72794231 0.67312703]
Desvio Padrão X1 e X2 (N=500) =  [0.70004571 0.71274119]


Item e
Coeficiente de Correlaçao X1 e X2 (N=5) =  0.0
Coeficiente de Correlação X1 e X2 (N=15) =  0.018208926018230737
Coeficiente de Correlação X1 e X2 (N=100) =  -0.0026530739871213293
Coeficiente de Correlação X1 e X2 (N=500) =  0.0


Item f
Será verificado se F(1,1) = Fx1(1).Fx2(1), assim como calculado na parte teórica

(N=5) F(1,1) = 0.0 ; Fx1(1).Fx2(1) =  0.24
(N=15) F(1,1) = 0.0 ; Fx1(1).Fx2(1) =  0.2488888888888888
(N=100) F(1,1) = 0.0 ; Fx1(1).Fx2(1) =  0.2491
(N=500) F(1,1) = 0.0 ; Fx1(1).Fx2(1) =  0.24993600000000002
```