CAPÍTULO 5
Exercícios Resolvidos

R5.1) Casais com no máximo 2 filhos

Consideremos o conjunto dos casais que têm no máximo dois filhos. Admitamos que dentro desse contexto, cada uma das possibilidades em termos do número de filhos, a saber, 0 filhos, 1 filho e 2 filhos têm a mesma probabilidade, ou seja, 1/3 para cada uma delas.
Admitamos também que as probabilidades de nascimento de homens e de mulheres são iguais.
Assim sendo, entre os que têm apenas 1 filho (o que ocorre com probabilidade 1/3), temos metade para cada sexo, isto é, 1/6 para 1 filho homem e 1/6 para uma filha mulher.
Analogamente, entre os que têm 2 filhos (o que também ocorre com probabilidade 1/3), de novo cada uma das 4 possibilidades de combinações dos sexos tem a mesma chance: 2 homens tem probabilidade 1/12, 2 mulheres tem probabilidade 1/12, 1 homem e 1 mulher tem probabilidade 1/6.
Sejam X e Y, respectivamente, o número de filhos homens e o número de filhas mulheres de um casal escolhido ao acaso.

(a) Qual a distribuição de probabilidade de X? E de Y?
(b) Calcule E(X), Var(X), E(Y) e Var(Y).
(c) X e Y são variáveis aleatórias independentes? Por que?
(d) Calcule E(X + Y) e Var(X + Y).
(e) Calcule Cov(X,Y).
(f) Verifique, neste caso, a validade das expressões E(X + Y) = E(X) + E(Y) e Var(X + Y) = Var(X) + Var(Y) + 2Cov(X,Y).

Solução:

(a) X e Y têm ambos a mesma distribuição de probabilidade. X (resp, Y) pode ser 0, 1 ou 2, com probabilidades 7/12, 4/12 e 1/12, respectivamente. Por que?

(b) E(X) = E(Y) = 0x7/12+1x4/12+2x1/12 = 1/2 e
$\text{Var}(X) = \text{Var}(Y) = 0^2x7/12+1^2x4/12+2^2x1/12 – (1/2)^2 = 5/12$.

(c) X e Y não são mais variáveis aleatórias independentes. Por que? Por exemplo, porque $P(X=0, Y=0) = 1/3 \neq 49/144 = (7/12)x(7/12) = P(X=0).P(Y=0)$.

(d) X + Y é o número total de filhos (de ambos os sexos) de um casal selecionado ao acaso. Já vimos que, por construção, X + Y pode assumir cada um dos valores 0, 1 ou 2 com probabilidade 1/3. Então
E(X + Y) = 0x1/3+1x1/3+2x1/3 = 1 e
$\text{Var}(X + Y) = 0^2x1/3+1^2x1/3+2^2x1/3 – 1^2 = 2/3$.

(e) A variável XY só pode assumir os valores 0 e 1, com probabilidades 5/6 e 1/6, respectivamente. (Por que?)
Daí, E(XY) = 0x5/6+1x1/6 = 1/6.
Pela propriedade (f), temos
Cov(X,Y) = E(XY) – E(X)E(Y) = 1/6 – (1/2).(1/2) = – 1/12.
Finalmente,

(f) E(X) + E(Y) = ½ + ½ = 1 = E(X+Y)
Var(X) + Var(Y) + 2Cov(X,Y) = 5/12 + 5/12 + 2 x (-1/12) = 2/3 = Var(X + Y)
conforme prevê a propriedade (e).

R5.2) Multiplicação de partículas

Um certo tipo de partícula se divide em 0, 1 ou 2 novas partículas (que serão chamadas suas descendentes) com probabilidades 30%, 40% e 30%, respectivamente, e depois se desintegra. As partículas individuais agem independentemente entre si. Dada uma partícula, seja X_1 o número dos seus descendentes e seja X_2 o número de descendentes dos seus descendentes.
Calcule:
(a) $P(X_2 = 0)$
(b) $P(X_1 = 1 \mid X_2 = 2)$

Solução:
Temos $P(X_1 = 0) = 0,3$ $P(X_1 = 1) = 0,4$ $P(X_1 = 2) = 0,3$

$P(X_2 = 0 \mid X_1 = 0) = 1$

$P(X_2 = 0 \mid X_1 = 1) = 0,3$
$P(X_2 = 1 \mid X_1 = 1) = 0,4$
$P(X_2 = 2 \mid X_1 = 1) = 0,3$

$P(X_2 = 0 \mid X_1 = 2) = 0,3 \times 0,3 = 0,09$
$P(X_2 = 1 \mid X_1 = 2) = 0,3 \times 0,4 + 0,4 \times 0,3 = 0,24$
$P(X_2 = 2 \mid X_1 = 2) = 0,3 \times 0,3 + 0,4 \times 0,4 + 0,3 \times 0,3 = 0,34$
$P(X_2 = 3 \mid X_1 = 2) = 0,3 \times 0,4 + 0,4 \times 0,3 = 0,24$
$P(X_2 = 4 \mid X_1 = 2) = 0,3 \times 0,3 = 0,09$

(a) $P(X_2 = 0) =$
$$= P(X_2 = 0 \mid X_1 = 0) \cdot P(X_1 = 0) + P(X_2 = 0 \mid X_1 = 1) \cdot P(X_1 = 1) + P(X_2 = 0 \mid X_1 = 2) \cdot P(X_1 = 2)$$
$$= 1 \times 0,3 + 0,3 \times 0,4 + 0,09 \times 0,3 = 0,447 .$$

(b) $P(X_1 = 1 \mid X_2 = 2) =$
$$= \frac{P(X_2 = 2 \mid X_1 = 1) \cdot P(X_1 = 1)}{P(X_2 = 2 \mid X_1 = 0) \cdot P(X_1 = 0) + P(X_2 = 2 \mid X_1 = 1) \cdot P(X_1 = 1) + P(X_2 = 2 \mid X_1 = 2) \cdot P(X_1 = 2)}$$

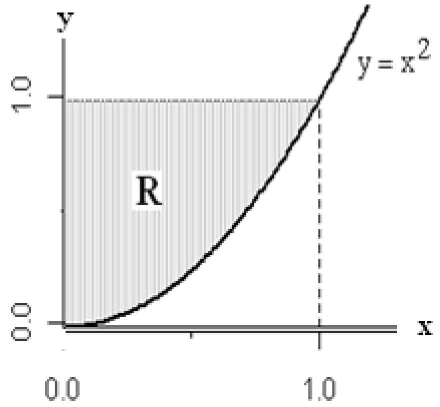
$$= \frac{0,3 \times 0,4}{0 \times 0,3 + 0,3 \times 0,4 + 0,34 \times 0,3} = \frac{0,12}{0,222} = 0,5405 .$$

R5.3) Distribuição uniforme em uma região do plano

Dizemos que uma v.a bidimensional (X,Y) tem distribuição uniforme em uma região R do plano real se sua função de densidade conjunta é

$$f(x,y) = \frac{1}{\text{Área de R}} \quad , \quad (x,y) \in R$$
$$= \quad 0 \quad , \text{caso contrário}$$

Seja R a região do plano limitada pela curva $y = x^2$, o eixo dos y e a reta y =1 (ver Figura a seguir).



A região do plano onde (X,Y) está definido

Se (X,Y) é uniforme em R, determine:

(a) A densidade conjunta de (X,Y)
(b) As densidades marginais de X e de Y
(c) As densidades condicionais de X dado Y = y e de Y dado X = x
(d) As esperanças condicionais de X dado Y = y e de Y dado X = x
(e) As variâncias condicionais de X dado Y = y e de Y dado X = x

Solução:

(a) Área de R = $\int_0^1 \int_0^{\sqrt{y}} dxdy = 2/3$. Então a densidade conjunta de (X,Y) é

$$f(x,y) = 3/2 \quad , \text{ se } 0 \le x \le \sqrt{y} \le 1$$
$$= 0 \quad , \text{caso contrario}$$

(b) Temos, portanto:

$$f_X(x) = \int_{x^2}^1 \frac{3}{2} dy = \frac{3}{2}(1 - x^2) \quad , 0 \le x \le 1$$

$$f_Y(y) = \int_0^{\sqrt{y}} \frac{3}{2} dx = \frac{3}{2}\sqrt{y} \quad , 0 \le y \le 1$$

(c) $f(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{1}{\sqrt{y}} \quad , \quad 0 \le x \le \sqrt{y} \le 1$

$$f(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{1}{1-x^2} \quad , \quad 0 \le x \le \sqrt{y} \le 1 \quad \text{ou} \quad 0 \le x^2 \le y \le 1$$

(d) $E(X|y) = \int_0^{\sqrt{y}} x \frac{1}{\sqrt{y}} dx = \frac{\sqrt{y}}{2} \quad , \quad 0 \le y \le 1$

$$E(Y|x) = \int_{x^2}^1 y \frac{1}{1-x^2} dy = \frac{1}{2}(1+x^2), \quad 0 \le x \le 1$$

(e) $Var(X|y) = E(X^2|y) - \{E(X|y)\}^2$, onde

$$E(X^2|y) = \int_0^{\sqrt{y}} x^2 \frac{1}{\sqrt{y}} dx = \frac{y}{3} \quad , \quad \text{se } 0 \le y \le 1.$$

Portanto, $\mathrm{Var}(X|y) = \frac{y}{3} - (\frac{\sqrt{y}}{2})^2 = \frac{y}{12}$, se $0 \le y \le 1$

$\mathrm{Var}(Y|x) = E(Y^2|\, x) - \{E(Y|x)\}^2$, onde

$E(Y^2|x) = \int_{x^2}^{1} y^2 \frac{1}{1-x^2} dy = \frac{1-x^6}{3(1-x^2)} = \frac{1}{3}(1 + x^2 + x^4)$, se $0 \le x \le 1$

Portanto, $\mathrm{Var}(Y|x) = \frac{1}{3}(1 + x^2 + x^4) - \frac{1}{4}(1 + x^2)^2 = \frac{(1-x^2)^2}{12}$, se $0 \le x \le 1$

R5.4) Tempo gasto no caixa de uma loja

Para cada cliente que entra na fila do caixa de uma loja de roupas:

- O tempo de espera na fila segue uma distribuição de probabilidade exponencial com média de 5 minutos;
- O tempo de atendimento segue uma distribuição de probabilidade exponencial com média de 3 minutos;
- Esses dois tempos são v.a.'s independentes.

Para a variável "tempo total do cliente no caixa" – incluindo a espera na fila e o atendimento – , determine a FDA, a densidade, a esperança e o desvio padrão.

Solução:
Sejam X o tempo de espera na fila e Y o tempo de atendimento, ambos em minutos.
Então suas densidades são respectivamente:
$f(x) = (1/5)\exp\left(-\frac{x}{5}\right)$, para $x > 0$ e $g(y) = (1/3)\exp\left(-\frac{y}{3}\right)$, para $y > 0$.
Como X e Y são v.a.'s independentes, sua densidade conjunta é
$\varphi(x,y) = (1/5)\exp\left(-\frac{x}{5}\right).(1/3)\exp\left(-\frac{y}{3}\right)$, para $x > 0$ e $y > 0$.
Seja Z o tempo total do cliente em minutos. Sua FDA é então:
$H(z) = P[Z \le z] = P[X + Y \le z] = \int_0^z \int_0^{z-x} (1/5)\exp\left(-\frac{x}{5}\right).(1/3)\exp\left(-\frac{y}{3}\right) dy\, dx =$

$$= \int_0^z (1/5)\exp\left(-\frac{x}{5}\right)\left[\int_0^{z-x}(1/3)\exp\left(-\frac{y}{3}\right)dy\right]dx =$$

$$= \int_0^z (1/5)\exp\left(-\frac{x}{5}\right)\left[1 - \exp\left(-\frac{z-x}{3}\right)\right]dx =$$

$$= \int_0^z (1/5)\exp\left(-\frac{x}{5}\right)dx - (3/2)\exp\left(-\frac{z}{3}\right)\int_0^z \frac{2}{15}\exp\left(\frac{2x}{15}\right)dx =$$

$$= 1 - \exp\left(-\frac{z}{5}\right) - (3/2)\exp\left(-\frac{z}{3}\right)\left[\exp\left(\frac{2z}{15}\right) - 1\right] =$$

$= 1 + (1/2)\left[3\exp\left(-\frac{z}{3}\right) - 5\exp\left(-\frac{z}{5}\right)\right]$, para $z > 0$, sendo $H(z) = 0$, para $z \le 0$.

A densidade de Z é então
$h(z) = \frac{dH(z)}{dz} = (1/2)\left[\exp\left(-\frac{z}{5}\right) - \exp\left(-\frac{z}{3}\right)\right]$, para $z > 0$, sendo $h(z) = 0$, para $z \le 0$.

A esperança e o desvio padrão de Z podem ser ambos calculados diretamente via integração, a partir da sua densidade:
$E(Z) = \int_0^\infty z(1/2)\left[\exp\left(-\frac{z}{5}\right) - \exp\left(-\frac{z}{3}\right)\right]dz = \left(\frac{25}{2} - \frac{9}{2}\right)\Gamma(2) = 8$ min.
$\mathrm{Var}(Z) = \int_0^\infty z^2(1/2)\left[\exp\left(-\frac{z}{5}\right) - \exp\left(-\frac{z}{3}\right)\right]dz - 8^2 = \left(\frac{125}{2} - \frac{27}{2}\right)\Gamma(3) - 64 = 34$ min^2.

DP(Z) = $\sqrt{34} = 5{,}83$ min.
Observe que, para chegarmos à função Gama, foram feitas substituições de variáveis do tipo: z = 5t e z = 3t, nas integrais acima.
Uma outra forma de se obter a esperança e a variância de Z seria através do uso das propriedades:
E(Z) = E(X + Y) = E(X) + E(Y) = 5 + 3 = 8 min.
Var(Z) = Var(X + Y) = Var(X) + Var(Y) = $5^2 + 3^2 = 34$ min^2.   (usando a independência)
Então DP(Z) = $\sqrt{34} = 5{,}83$ minutos.

R5.5) Vôos domésticos e vôos internacionais

Seja X o número de aeronaves que chegam a um determinado aeroporto, no intervalo de 4 horas, provenientes de vôos domésticos. Seja Y o número de aeronaves que chegam a esse mesmo aeroporto, ao longo do mesmo intervalo de 4 horas, porém provenientes de vôos internacionais. Sabe-se que:

- A distribuição marginal de X+Y (nº total de chegadas em 4 horas) é Poisson com parâmetro λ, isto é, $P[X+Y=j] = \frac{e^{-\lambda}\lambda^j}{j!}$, para todo j = 0,1,2,...
- A distribuição condicional de X dado que X+Y = j é Binomial(j;p), isto é, $P[X=k|X+Y=j] = \binom{j}{k} p^k (1-p)^{j-k}$, para todo k = 0,1,2,...,j.

Mostre que, nessas condições:

(a) a distribuição marginal de X (chegadas de vôos domésticos em 4 horas) é Poisson com parâmetro λp, isto é, $P[X=k] = \frac{e^{-\lambda p}(\lambda p)^k}{k!}$, para todo k = 0,1,2,...

(b) a distribuição condicional de X+Y (total de chegadas em 4 horas), dado que X = k (chegam k vôos domésticos em 4 horas) é uma "Poisson truncada", isto é, $P[X+Y=j|X=k] = \frac{e^{-\lambda(1-p)}(\lambda(1-p))^{j-k}}{(j-k)!}$, para todo j = k, k+1, k+2,...

(c) Determine o número esperado total de chegadas (entre as provenientes de vôos domésticos e internacionais) ao longo de 4 horas, dado que durante esse período chegaram k vôos nacionais, ou seja, E[X+Y| X=k].

Solução:

(a) $P(X=k) = \sum_{j=k}^{\infty} P(X=k|X+Y=j)P(X+Y=j) =$

$= \sum_{j=k}^{\infty} \binom{j}{k} p^k (1-p)^{j-k} \frac{\exp(-\lambda)\lambda^j}{j!}$.

Fazendo i = j – k , temos i = 0,1,2,...   e   j = i + k. Então,

$P(X=k) = \sum_{i=0}^{\infty} \frac{(i+k)!}{i!k!} p^k (1-p)^i \frac{\exp(-\lambda)\lambda^{i+k}}{(i+k)!} =$

$= \frac{(\lambda p)^k \exp(-\lambda)}{k!} \sum_{i=0}^{\infty} \frac{(1-p)^i \lambda^i}{i!} = \frac{(\lambda p)^k}{k!} \exp(-\lambda)\exp(\lambda(1-p)) = \frac{(\lambda p)^k}{k!} \exp(-\lambda p)$,

para todo k = 0,1,2,...
Isso mostra que X, o número de chegadas de vôos domésticos em 4 horas, segue uma distribuição de Poisson com parâmetro λp.

(b) $P(X+Y=j|X=k) = \frac{\binom{j}{k}p^k(1-p)^{j-k}\exp(-\lambda)\lambda^j}{j!}\frac{k!}{(\lambda p)^k\exp(-\lambda p)} =$

$= \frac{j!k!}{k!(j-k)!j!}\frac{p^k\lambda^{j-k}(1-p)^{j-k}\exp(-\lambda(1-p))}{\lambda^k p^k} = \frac{\exp(-\lambda(1-p))(\lambda(1-p))^{j-k}}{(j-k)!}$

para j = k, k+1, k+2,...
Isso mostra que a distribuição condicional de X + Y, dado que X = k, é uma "Poisson truncada" com parâmetro $\lambda(1-p)$.

(c) Decorre do item (b) que a v.a. X + Y – k é uma Poisson com parâmetro $\lambda(1-p)$. Logo, dado que em 4 horas houve k chegadas de vôos domésticos, o número esperado de chegadas nesse mesmo intervalo de tempo, sejam elas provenientes de vôos domésticos ou internacionais, é

$$E(X+Y|X=k) = E(k + (X+Y-k)) = k + \lambda(1-p).$$

R5.6) Mais uma vez o recadastramento

Consideremos, mais uma vez, a situação do Exemplo 5.9. Isto é, o processo de recadastramento vai evoluindo progressiva e uniformemente ao longo do ano, de modo que, se x é a proporção de indivíduos já recadastrados, então x = 0 no início do ano e x = 1 no fim do ano. Porém, agora:

- X é o tempo (em fração de ano) a contar do início do ano até o momento em que se realiza um experimento no qual são sorteados sucessivamente tantos membros da população quantos forem necessários até que apareça o primeiro já recadastrado;
- Y é o número de sorteios realizados até aparecer o primeiro recadastrado.

Determine:

(a) A distribuição condicional de Y dado que X = x.
(b) A esperança condicional de Y dado que X = x.
(c) A distribuição marginal de Y.
(d) O valor esperado de Y.
(e) A distribuição condicional de X dado Y.
(f) A esperança condicional de X dado Y.

Solução:

(a) Aqui se trata de uma distribuição geométrica com parâmetro x, isto é,

$$P(Y=y|X=x) = x(1-x)^{y-1}, \qquad \text{para todo } y = 1,2,3,...$$

(b) Por isso, E(Y|X=x) = 1/x.

Então, se, por exemplo, esse experimento for realizado no final de fevereiro, ou seja, depois de passados 2 meses – o que corresponde a x = 2/12 = 1/6 do ano – espera-se que sejam necessários 1/x = 6 sorteios para que apareça o primeiro indivíduo já recadastrado.

(c) $P(Y{=}y) = \int_0^1 x(1-x)^{y-1}dx = \beta(2;y) = \frac{\Gamma(2)\Gamma(y)}{\Gamma(2+y)} = \frac{(y-1)!}{(y+1)!} = \frac{1}{y(y+1)}$.

(d) $E(Y) = \sum_{y=1}^{\infty} y\frac{1}{y(y+1)} = \sum_{y=1}^{\infty}\frac{1}{y+1} = +\infty$.

Isso significa que se esses sorteios forem realizados em um momento escolhido aleatoriamente ao longo do ano, espera-se que seja necessário um número infinito de sorteios para que apareça o primeiro indivíduo já recadastrado.

(e) $f(x|Y=y) = \frac{P(Y=y|X=x)f_X(x)}{P(Y=y)} = \frac{x(1-x)^{y-1}}{1/_{y(y+1)}} = y(y+1)x(1-x)^{y-1}$, $0 < x < 1$

Esta é a distribuição Beta(a,b), com a = 2 e b = y.

(f) Por isso, $E(X|Y=y) = \frac{a}{a+b} = \frac{2}{2+y}$.

Logo, se em um determinado momento ao longo do ano realizou-se o experimento e, por exemplo, foram necessários 4 sorteios para que aparecesse o primeiro indivíduo já recadastrado, isso deve ter ocorrido em torno do final de abril, isto é, depois de passados y = 4 meses, ou 2/(2+4) = 1/3 do ano.

R5.7) Carteira de aplicações financeiras

Uma pessoa investe um total de C = 10000 reais em duas aplicações cujas taxas de retorno são variáveis aleatórias independentes X_1 e X_2, com médias 5% e 14% e desvios padrão 1% e 8%, respectivamente. O desvio padrão $\sigma(R)$ do seu retorno total $R = C_1X_1 + C_2X_2$ será usado aqui como uma medida do risco envolvido em selecionar essa dada carteira de aplicações.

(a) Caso se deseje manter o risco no mínimo possível, que quantias C_1 e C_2 devem ser investidas nas respectivas aplicações? Quais são a média do retorno e o risco correspondentes a essa carteira?

(b) Qual é o tamanho do risco a ser corrido para se atingir uma carteira cujo retorno médio seja de 770 reais?

(c) Através da Desigualdade de *Chebyshev*, obtenha um intervalo simétrico em torno de 770 reais que, com probabilidade superior a 80%, conterá o retorno R da carteira obtida no item (b).

Obs.: A Desigualdade de *Chebyshev* afirma que se Y é uma variável aleatória com esperança e variância finitas e ε é uma constante positiva, então

$$P\left[|Y - E(Y)| \geq \varepsilon\right] < \frac{Var(Y)}{\varepsilon^2}.$$

Solução:

$$R = C_1X_1 + C_2X_2 = C_1X_1 + (C - C_1)X_2$$

$$E(R) = C_1E(X_1) + (C - C_1)E(X_2)$$

$Var(R) = C_1^2 Var(X_1) + (C - C_1)^2 Var(X_2)$, devido à independência entre X_1 e X_2.

(a) Minimizar o desvio padrão é o mesmo que minimizar a variância. Então, para minimizar o risco (desvio padrão de R), devemos igualar a zero a derivada de Var(R) com relação a C_1.

$$\frac{dVar(R)}{dC_1} = 2C_1Var(X_1) - 2(C - C_1)Var(X_2)$$

Então, $\dfrac{dVar(R)}{dC_1} = 0$ implica que

$$C_1 = \frac{var(X_2)}{var(X_1)+var(X_2)}C = \frac{0{,}08^2}{0{,}01^2+0{,}08^2}\times 10000 = 9846{,}15 \text{ reais}$$

$$C_2 = C - C_1 = 153{,}85 \text{ reais}$$

$$E(R) = 9846{,}15 \times 0{,}05 + 153{,}85 \times 0{,}14 = 513{,}85 \text{ reais}$$

$$Var(R) = 9846{,}15^2 \times 0{,}01^2 + 153{,}85^2 \times 0{,}08^2 = 9846{,}15 \text{ reais}^2$$

Então, $\sigma(R) = \sqrt{9846{,}15} = 99{,}23 \text{ reais}$

(b) E(R) = 770 implica que $\sigma(R) = ?$

$$E(R) = C_1E(X_1) + (C - C_1)E(X_2) = 770 .$$

Então $C_1 = \dfrac{C \cdot E(X_2) - 770}{E(X_2) - E(X_1)} = \dfrac{10000 \times 0{,}14 - 770}{0{,}14 - 0{,}08} = 7000 \text{ reais}$

e $C_2 = 10000 - 7000 = 3000 \text{ reais}$.

$$\sigma(R) = \sqrt{7000^2 \times 0{,}01^2 + 3000^2 \times 0{,}08^2} = 250 \text{ reais}$$

(c) Aplicando a Desigualdade de *Chebyshev* à variável aleatória R, temos

$$P[|R - E(R)| \geq \varepsilon] < \frac{Var(R)}{\varepsilon^2}$$

Então

$$P[|R - E(R)| < \varepsilon] = P[E(R) - \varepsilon < R < E(R) + \varepsilon] \geq 1 - \frac{Var(R)}{\varepsilon^2}$$

Por outro lado, do item (b) sabemos que se E(R) = 770 então $\sigma(R) = 250$.
Então, para que o intervalo centrado em E(R) = 770 tenha probabilidade > 0,80, devemos igualar $1 - \dfrac{Var(R)}{\varepsilon^2}$ a 0,80.

$1 - \dfrac{Var(R)}{\varepsilon^2} = 0{,}80$ implica que $\varepsilon = \sqrt{\dfrac{250^2}{0{,}20}} = 559{,}02 \text{ reais}$.

O intervalo desejado é então $(770 - 559{,}02\,;\, 770 + 559{,}02)$,
ou seja, (210,98 ; 1329,02) em reais.

R5.8) Mais sobre o Movimento Browniano

Sob as mesmas condições do Exercício R3.6, mostre que se $X = |(1 - W)S + WT|$, onde:

- S é Normal($-x_0$; 2Dt);
- T é Normal(x_0; 2Dt);
- W é Bernoulli com p = ½;
- As variáveis aleatórias W, S e T são independentes entre si;

então X tem densidade f.

Solução:
Sejam S ~ N(–x_0; 2Dt), T ~ N(x_0; 2Dt) e W Bernoulli(p) com p = ½, onde as variáveis W, S e T são independentes entre si.

Suponhamos agora que $X = |(1-W)S + WT|$.

Desejamos mostrar que a densidade f(.) da v.a. X assim obtida é dada pela expressão

$$f(x) = \begin{cases} \dfrac{1}{2\sqrt{\pi Dt}}\left\{\exp\left[-\dfrac{(x+x_0)^2}{4Dt}\right] + \exp\left[-\dfrac{(x-x_0)^2}{4Dt}\right]\right\}, & \text{para } x \geq 0 \\ 0, & \text{para } x < 0 \end{cases}$$

A FDA de X é, por definição,

$$F(x) = P[X \leq x] = P[|(1-W)S + WT| \leq x] = P[-x \leq (1-W)S + WT \leq x] =$$
$$= \tfrac{1}{2}\{P[-x \leq (1-W)S + WT \leq x|W=0] + P[-x \leq (1-W)S + WT \leq x|W=1]\}=$$
$$= \frac{1}{2}\{P[-x \leq S \leq x] + P[-x \leq T \leq x]\} =$$
$$= \frac{1}{2}\left(\Phi\left(\frac{x+x_0}{\sqrt{2Dt}}\right) - \Phi\left(\frac{-x+x_0}{\sqrt{2Dt}}\right) + \Phi\left(\frac{x-x_0}{\sqrt{2Dt}}\right) - \Phi\left(\frac{-x-x_0}{\sqrt{2Dt}}\right)\right)$$

Já que:

- $\Phi(-a) = 1 - \Phi(a)$ implica em $\Phi(a) - \Phi(-a) = 2\,\Phi(a) - 1$, para todo a real;

  e
- $\frac{-x-x_0}{\sqrt{2Dt}} = -\frac{x+x_0}{\sqrt{2Dt}}$ e $\frac{-x+x_0}{\sqrt{2Dt}} = -\frac{x-x_0}{\sqrt{2Dt}}$;

concluímos que:

$$F(x) = \tfrac{1}{2}\left(2\Phi\left(\tfrac{x+x_0}{\sqrt{2Dt}}\right) + 2\Phi\left(\tfrac{x-x_0}{\sqrt{2Dt}}\right) - 2\right) = \Phi\left(\tfrac{x+x_0}{\sqrt{2Dt}}\right) + \Phi\left(\tfrac{x-x_0}{\sqrt{2Dt}}\right) - 1.$$

Conseqüentemente,

$$f(x) = \frac{dF(x)}{dx} = \frac{d}{dx}\left[\Phi\left(\frac{x+x_0}{\sqrt{2Dt}}\right) + \Phi\left(\frac{x-x_0}{\sqrt{2Dt}}\right) - 1\right] = \varphi\left(\frac{x+x_0}{\sqrt{2Dt}}\right)\frac{1}{\sqrt{2Dt}} + \varphi\left(\frac{x-x_0}{\sqrt{2Dt}}\right)\frac{1}{\sqrt{2Dt}}$$
$$= \tfrac{1}{2\sqrt{\pi Dt}}\left\{\exp\left[-\tfrac{(x+x_0)^2}{4Dt}\right] + \exp\left[-\tfrac{(x-x_0)^2}{4Dt}\right]\right\}, \text{ se } x \geq 0.$$

Se x < 0, é claro que f(x) = 0, já que $X = |(1-W)S + WT|$ e o módulo é obrigatoriamente não-negativo.

R5.9) Tempo de deslocamento da residência até o local de trabalho

Considere uma pessoa que, toda manhã, faz uma viagem de carro desde sua residência no subúrbio até a estação ferroviária e, dali, toma um trem rumo ao seu local de trabalho no centro da cidade. Ela costuma sair de casa entre 7:00 e 7:30. O percurso de carro até a estação ferroviária leva entre 10 e 20 minutos. Admita que tanto o instante de partida quanto a duração do percurso de carro são variáveis aleatórias independentes, cada uma delas com distribuição uniforme no seu respectivo intervalo. Há três trens que ela pode tomar, sendo que todos eles são absolutamente pontuais em seus horários de partida e de chegada. O primeiro trem parte às 7:30 e chega às 8:20. O segundo trem parte às 7:45 e chega às 8:25. O terceiro trem parte às 8:00 e chega às 8:45.

(a) Considerando que o tempo é contado em minutos a partir de 7:00, mostre que a função de distribuição acumulada da variável aleatória Y, instante de chegada dessa pessoa à estação ferroviária, é dada por:

$$F_Y(y) = \begin{cases} 0, & \text{se } y < 10 \\ \frac{1}{600}(y-10)^2, & \text{se } 10 \le y \le 20 \\ \frac{1}{30}(y-15), & \text{se } 20 \le y \le 40 \\ \frac{1}{300}\left(300 - \frac{1}{2}(50-y)^2\right), & \text{se } 40 \le y \le 50 \\ 1, & \text{se } y > 50 \end{cases}$$

e que a sua correspondente função de densidade é

$$f_Y(y) = \begin{cases} \frac{1}{300}(y-10), & \text{se } 10 \le y < 20 \\ \frac{1}{30}, & \text{se } 20 \le y < 40 \\ \frac{1}{300}(50-y), & \text{se } 40 \le y \le 50 \\ 0, & \text{nos demais casos} \end{cases}.$$

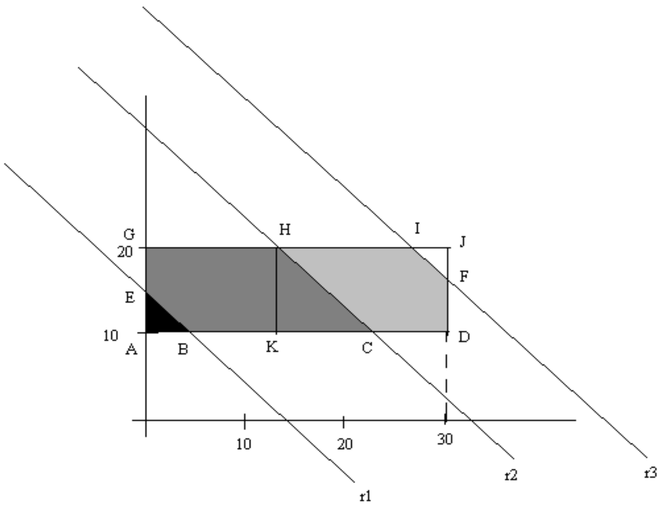
(b) Determine a média e o desvio padrão do horário de chegada dessa pessoa ao centro da cidade.

Solução:

(a) Podemos escrever que $Y = X_1 + X_2$, sendo X_1 e X_2 v.a.'s independentes e tais que X_1 é $U[0; 30]$ e X_2 é $U[10;20]$.

$F_Y(y) = P(Y \le y) = P(X_1 + X_2 \le y)$.

Para calcular essa probabilidade temos que considerar as 3 retas r1, r2 e r3 da figura a seguir:



Todas as 3 retas têm como equação $x_1 + x_2 = y$, porém:
No caso da reta r1, temos $10 \le y \le 20$, e

$$F_Y(y) = \frac{\text{Area do triângulo AEB}}{\text{Area do retangulo ADJG}} = \frac{(y-10)^2/2}{10 \times 30} = \frac{(y-10)^2}{600}$$

No caso da reta r2, temos $20 \le y \le 40$, e

$$F_Y(y) = \frac{\text{Area do retângulo AKHG} + \text{Area do triângulo KCH}}{\text{Area do retangulo ADJG}} =$$

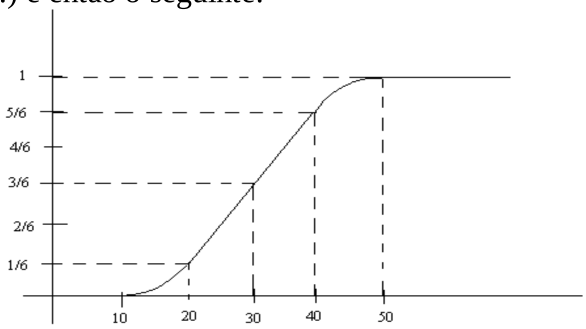
$$= \frac{10(y-20) + 10^2/2}{10 \times 30} = \frac{1}{30}(y-15)$$

No caso da reta r3, temos $40 \le y \le 50$, e

$$F_Y(y) = \frac{\text{Area do retângulo ADJG - Area do triângulo IJF}}{\text{Area do retangulo ADJG}} =$$

$$= \frac{10 \times 30 - (50-y)^2/2}{10 \times 30} = \frac{1}{300}\left(300 - \frac{(50-y)^2}{2}\right)$$

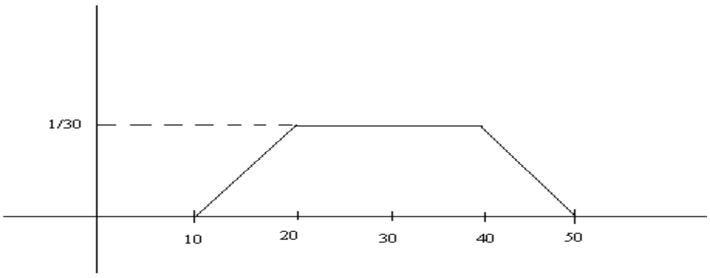
O gráfico da FDA $F_Y(.)$ é então o seguinte:



Calculando a derivada, obtemos:

$$f_Y(y) = \begin{cases} \frac{1}{300}(y-10), & \text{se } 10 \le y < 20 \\ \frac{1}{30}, & \text{se } 20 \le y < 40 \\ \frac{1}{300}(50-y), & \text{se } 40 \le y \le 50 \end{cases}$$

cujo gráfico é o seguinte:



(b) Consideremos agora a viagem de trem:

- Para que ela consiga tomar o primeiro trem é necessário que chegue à estação ferroviária no máximo até 7:30, o que corresponde a 30 minutos, a partir das 7:00. A probabilidade de que isso aconteça pode ser calculada como o valor da função de distribuição acumulada de Y no ponto y = 30:

$$F_Y(30) = \frac{1}{30}(30-15) = \frac{1}{2}.$$

Neste caso ela chegaria ao centro da cidade às 8:20.

- Para que ela perca o primeiro trem, mas consiga tomar o segundo trem é necessário que chegue à estação ferroviária entre 7:30 e 7:45, o que corresponde ao intervalo que vai desde 30 minutos até 45 minutos, a partir das 7:00. A probabilidade de que isso aconteça pode ser calculada como

$$F_Y(45) - F_Y(30) = \frac{1}{300}\left(300 - \frac{1}{2}(50-45)^2\right) - \frac{1}{30}(30-15) = \frac{23}{24} - \frac{1}{2} = \frac{11}{24}.$$

Neste caso ela chegaria ao centro da cidade às 8:25.

- Para que ela perca os dois primeiros trens, sendo portanto obrigada a tomar o terceiro trem é necessário que chegue à estação ferroviária após as 7:45, o que corresponde a 45 minutos, a partir das 7:00. A probabilidade de que isso aconteça pode ser calculada como

$$1 - F_Y(45) = 1 - \frac{1}{300}\left(300 - \frac{1}{2}(50-45)^2\right) = 1 - \frac{23}{24} = \frac{1}{24}.$$

Neste caso ela chegaria ao centro da cidade às 8:45.

Assim sendo, se W é a variável aleatória que corresponde ao momento em que ela chegará ao centro da cidade, contado em minutos a partir de 7:00, temos:

$$W = \begin{cases} 80, & \text{com probabilidade } \frac{1}{2} \\ 85, & \text{com probabilidade } \frac{11}{24} \\ 105, & \text{com probabilidade } \frac{1}{24} \end{cases}$$

Ou, de outra forma, calculando as probabilidades como áreas sob a curva de f_Y:

$$P[W=80] = P[10 < Y < 30] = \frac{1}{2} \times \frac{1}{30} \times (20-10) + \frac{1}{30} \times (30-20) = \frac{1}{2}$$

$$P[W=85] = P[30 < Y < 45] = \frac{1}{30} \times (40-30) + \left(\frac{1}{2} \times \frac{1}{30} \times (50-40) - \frac{1}{2} \times \frac{1}{60} \times (50-45)\right) = \frac{11}{24}$$

$$P[W=105] = P[45 < Y < 50] = \frac{1}{2} \times \frac{1}{60} \times (50-45) = \frac{1}{24}$$

Daí, $E(W) = 80 \times \frac{1}{2} + 85 \times \frac{11}{24} + 105 \times \frac{1}{24} = 83{,}33$ minutos a partir de 7:00 e

$$DP(W) = \sqrt{80^2 \times \frac{1}{2} + 85^2 \times \frac{11}{24} + 105^2 \times \frac{1}{24} - 83{,}33^2} = 5{,}14 \text{ minutos.}$$

Isso significa que o horário esperado da chegada ao centro é 8 horas, 23 minutos e 20 segundos, com um desvio padrão de 5 minutos e 8 segundos.

Exercícios Propostos

P5.1) Vendas semanais de carros importados e carros nacionais

Uma concessionária de automóveis vem mantendo semanalmente em estoque 2 carros importados e 3 de fabricação nacional, para atender aos seus clientes. Sejam X e Y as variáveis aleatórias que representam respectivamente o número de carros importados e o número de carros nacionais que ela vende ao longo de uma semana. Assim sendo, X pode assumir os valores 0, 1, 2 e Y os valores 0, 1, 2, 3. A função de probabilidade conjunta de X e Y é dada pela tabela abaixo:

Função de probabilidade conjunta de X e Y

x	**y**			
	0	1	2	3
0	0,01	0,05	0,05	0,04
1	0,05	0,20	0,15	0,10
2	0,04	0,15	0,10	0,06

Qual a probabilidade de que, em uma determinada semana:

(a) Não seja vendido nenhum carro importado?
(b) Todos os carros nacionais sejam vendidos?
(c) Sejam vendidos no máximo um carro importado e um carro nacional?
(d) Sejam vendidos mais carros importados do que nacionais?
(e) Sejam vendidos ao todo pelo menos 4 carros?

P5.2) Novamente as vendas semanais de carros importados e nacionais

Considerando novamente a concessionária do exercício anterior, obtenha:

(a) as distribuições marginais de X e de Y.
(b) as distribuições condicionais de X dado Y, e de Y dado X.
(c) Cov(X,Y) e ρ(X,Y).

P5.3) Erro grave

Na resolução do exercício abaixo foi cometido um erro grave.
Pergunta:
Sejam X e Y duas v.a.'s independentes e tais que $X \sim N(80; 9)$ e $Y \sim N(50; 16)$. Qual a distribuição de probabilidade da v.a. $Z = X - Y$?
Resposta:
$E(Z) = E(X - Y) = E(X) - E(Y) = 80 - 50 = 30$
$Var(Z) = Var(X - Y) = Var(X) - Var(Y) = 9 - 16 = -7$.
Conclusão: $Z \sim N(30; -7)$.

(a) Qual foi o erro cometido aqui?
(b) Qual a solução correta?

P5.4) Casais com exatamente 2 filhos

Admitamos que as probabilidades de nascimento de homens e de mulheres são iguais, ou seja, 50% para cada sexo. Consideremos apenas casais que tenham dois filhos. Então cada uma das 4 possibilidades de combinações quanto aos sexos dos filhos (MM, MF, FM, FF) tem 25% de chance de acontecer. Seja X igual a 0 ou 1 conforme o primeiro filho seja homem ou mulher. Seja Y igual a 0 ou 1 conforme o segundo filho seja homem ou mulher. Mostre que:

(a) As variáveis aleatórias X e Y são independentes e cada uma delas tem distribuição de Bernoulli com p =1/2.
(b) Qual a distribuição de probabilidade de X + Y, o número de crianças do sexo feminino entre as duas?
(c) Verifique, neste caso particular, a validade das propriedades:
$$E(X + Y) = E(X) + E(Y) \text{ e } Var(X + Y) = Var(X) + Var(Y).$$
(d) Qual a distribuição de probabilidade da v.a. XY?
(e) Verifique que, neste caso particular, Cov(X,Y) = 0. Por que?

P5.5) Aposentadoria

Todos os servidores aposentados de um certo país estão pleiteando que seja revisto o valor de sua aposentadoria. Por outro lado, eles estão sendo recadastrados ao longo de um ano. Admita que o percentual p de aposentados já recadastrados cresce uniformemente desde p = 0 no início do ano até p = 1 no final do ano. Em determinado momento ao longo do ano serão sorteados 50 entre esses servidores, para que seus pleitos sejam analisados. Somente serão considerados os pleitos daqueles que já estiverem recadastrados.

Calcule a probabilidade de que:

a) Pelo menos 25 dos servidores selecionados tenham seus pleitos analisados, se essa seleção for feita no final de maio;
b) Entre 20 e 30 dos servidores selecionados tenham seus pleitos analisados, se essa seleção for feita no final de junho;
c) No máximo 25 dos servidores selecionados tenham seus pleitos analisados, se essa seleção for feita no final de agosto.

P5.6) Detector de Mentiras

Um detector de mentiras será usado pela polícia para investigar 10 suspeitos de envolvimento em um determinado crime. Admita que entre eles 5 são culpados (mas alegarão inocência) e os outros 5 são realmente inocentes. Sabe-se também que:

- mesmo quando uma pessoa diz a verdade, o detector tem uma chance de 5% de falhar, indicando que ela mentiu;
- mesmo quando ela mente, o detector tem uma chance de 30% de não conseguir detectar a mentira.

Qual a probabilidade de que:

(a) todos os 10 diagnósticos obtidos através do detector estejam corretos?
(b) o detector libere todos os 10 suspeitos?
(c) ao mesmo tempo, pelo menos 3 dos culpados sejam pegos e pelo menos 4 dos inocentes sejam liberados?

P5.7) Pesquisa de mercado

Está sendo realizada uma pesquisa de mercado para se investigar a demanda potencial por um novo produto a ser lançado proximamente. Dois entrevistadores, A e B, estão abordando aleatoriamente os consumidores que circulam por determinado local dentro de um Shoppíng Center – o público alvo da pesquisa – para que estes respondam às perguntas de um questionário. Admita que aqueles que estariam propensos a comprar o novo produto correspondem a uma determinada proporção p do público alvo. Seja X (respectivamente Y) o número de entrevistas a serem feitas por A (respectivamente B), até que ele encontre o primeiro consumidor disposto a comprar o novo produto.

(a) Quais os valores possíveis e a distribuição de probabilidade da variável X + Y, o número total de entrevistas feitas por A e por B até que cada um deles encontre pela primeira vez um consumidor potencial do novo produto? Você identifica a distribuição de probabilidade obtida como pertencente a alguma família conhecida de modelos probabilísticos. Qual?

(b) Qual a distribuição de probabilidade condicional de X dado que X+Y = j.

(c) Determine $E[Y|X+Y=j]$. Ou seja, se A e B juntos tiveram que abordar j pessoas até que cada um deles encontrasse o primeiro consumidor potencial do produto, em média quantas entrevistas B terá feito até esse ponto?

P5.8) Produção de milho

Na safra de 2000/2001, a produtividade do solo, em toneladas por hectare, das plantações de milho no Brasil teve uma média de 3,3 t/ha e um desvio padrão de 0,5 t/ha. Por outro lado, a área, em hectares, das propriedades rurais dedicadas ao plantio do milho tinha nessa ocasião uma de média de 3,6 ha e um desvio padrão de 1,1 ha.

Com base nessas informações, calcule, para a safra de 2000/2001:

(a) A média e o desvio padrão da produção de milho, em toneladas, de uma propriedade rural.

(b) A produção média de milho, em toneladas, correspondente às propriedades rurais onde a produtividade do solo era de exatamente 4 t/ha.

(c) O desvio padrão da produção de milho, em toneladas, correspondente às propriedades rurais cuja área era exatamente 4,2 ha.

(d) O coeficiente de correlação entre produtividade do solo e produção de milho.

(e) O coeficiente de correlação entre área da propriedade rural e produção de milho.

Obs.:

1. Os valores dos parâmetros deste problema são aproximações obtidas a partir de dados reais.
2. Admita que a produtividade do solo e a área da propriedade rural são variáveis aleatórias independentes.

P5.9) Consumo de combustível

Sabe-se que, em uma certa localidade:

- 60% dos carros são pequenos;
- 30% dos carros são médios;
- 10% dos carros são grandes;
- O desempenho de um carro grande em km/litro é uma variável aleatória com distribuição uniforme no intervalo [6; 10];
- O desempenho de um carro médio em km/litro é uma variável aleatória com distribuição uniforme no intervalo [8; 12];
- O desempenho de um carro pequeno em km/litro é uma variável aleatória com distribuição uniforme no intervalo [10; 14];

- A rodagem mensal dos carros em km/mês é uma variável aleatória com distribuição exponencial de média 1000;
- A rodagem e o desempenho são variáveis aleatórias independentes.

Qual o consumo médio de combustível dos carros do local em litros/mês?
Sugestão:
Se C é o consumo, R é a rodagem e D é o desempenho, então C = R $\frac{1}{D}$ e, pela independência, E(C) = E(R) E $\left(\frac{1}{D}\right)$.
Além disso E $\left(\frac{1}{D}\right)$ = E $\left(\frac{1}{D}|\text{Grd}\right)$. 0,1+ E $\left(\frac{1}{D}|\text{Med}\right)$.0,3+ E $\left(\frac{1}{D}|\text{Peq}\right)$.0,6. Por que?
Como para carros grandes, D ~ U[6; 10], E $\left(\frac{1}{D}|\text{Grd}\right) = \int_6^{10} \frac{1}{x} \cdot \frac{1}{10-6} dx$. Analogamente para carros médios e pequenos...

P5.10) Soma e Produto de uniformes independentes

Sejam X e Y duas v.a.'s iid ambas Uniformes no intervalo [0,1].
Sejam U = X + Y e V = XY.

(a) Obtenha a função de densidade conjunta de U e V.
(b) Obtenha as funções de densidade marginais de U e de V.

P5.11) Marcando um encontro

Dois amigos combinaram de se encontrar em determinado local entre 14:00 e 16:00, sendo que cada um deles esperaria pelo outro no máximo até 15 minutos. Qual a probabilidade de que eles realmente se encontrem?
Sugestão:
Contando o tempo em minutos a partir das 14:00, o instante de chegada de cada um pode ser visto como uma v.a. Uniforme no intervalo [0; 120]. Além disso, essas v.a.'s podem ser consideradas independentes. Desenhe uma figura em que cada eixo do plano bidimensional representa o instante de chegada de uma pessoa. Verifique qual é o subconjunto do quadrado [0; 120]x[0; 120] que corresponde a um encontro entre eles.

P5.12) Ainda o problema do encontro

Considere novamente o problema anterior. Sejam S o tempo em minutos desde as 14:00 até o momento em que chega o primeiro dos dois e T o tempo em minutos desde as 14:00 até o momento em que ambos já chegaram. Obtenha:

(a) a expressão algébrica das densidade marginais de S e de T.
(b) E(S), Var(S), E(T), Var(T).

Sugestão: Note que $P(T \leq t) = P(X \leq t, Y \leq t)$ e $P(S \geq s) = P(X \geq s, Y \geq s)$, sendo X e Y os instantes de chegada dessas duas pessoas.

P5.13) Mistura de Normais

Sejam X, Y e W v.a.'s independentes e tais que

$X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$ e $W \sim$ Bernoulli(p).

Definamos agora uma outra v.a.: $Z = (1 - W)X + WY$. Neste caso, dizemos que Z é uma mistura de duas Normais.

Prove que:

(a) A densidade de Z é $h(z) = (1-p)\frac{1}{\sqrt{2\pi}\sigma_X}\exp\left(-\frac{1}{2}\left(\frac{z-\mu_X}{\sigma_X}\right)^2\right) + p\frac{1}{\sqrt{2\pi}\sigma_Y}\exp\left(-\frac{1}{2}\left(\frac{z-\mu_Y}{\sigma_Y}\right)^2\right)$, $\forall z$.

(b) $E(Z) = (1-p)\,\mu_X + p\mu_Y$.

(c) $Var(Z) = (1-p)\,\sigma_X^2 + p\,\sigma_Y^2 + p(1-p)\left(\mu_X - \mu_Y\right)^2$.

Obs.: Note que no Exercício Resolvido 3.? temos uma mistura de Normais.

P5.14) Pureza do Minério de Ferro

Sejam X e Y duas v.a.'s tais que:

- X = teor de pureza de um minério de ferro
- Y = 0, se é tomada a decisão de não aproveitar esse minério na produção de aço
- Y = 1, se é tomada a decisão de aproveitar esse minério na produção de aço
- A distribuição marginal de X é Uniforme entre 0 e 1;
- A distribuição condicional de Y dado que X = x é Bernoulli(x).

(a) Determine $E(X|Y = 0)$ e $Var(X|Y = 0)$, ou seja, a média e a variância do teor de pureza do minério, dado que foi tomada a decisão de não aproveitar esse minério na produção de aço.

(b) Determine $E(X|Y = 1)$ e $Var(X|Y = 1)$, ou seja, a média e a variância do teor de pureza do minério, dado que foi tomada a decisão de aproveitar esse minério na produção de aço.

P5.15) Pedidos de informação em um aeroporto

Seja Z o número de pessoas em geral (nacionais ou estrangeiros) que recorrem ao balcão de informações de um aeroporto ao longo de uma hora. Sabe-se que:

- o número X de usuários nacionais que recorrem a esse balcão em uma hora é uma variável aleatória cuja lei de probabilidade é Poisson com freqüência média de chegada $\lambda_1 = 10$ pessoas por hora;
- o número Y de usuários estrangeiros que recorrem a esse balcão em uma hora segue também uma lei de probabilidade de Poisson com freqüência média de chegada $\lambda_2 = 4$ pessoas por hora;
- X e Y são variáveis aleatórias independentes.

(a) Calcule a probabilidade $P(11 \le Z \le 17)$ de que entre 11 e 17 pessoas, nacionais ou estrangeiras, recorrerão ao balcão de informações em uma hora.

(b) Mostre que a lei de probabilidade condicional de X dado que Z = z é uma binomial com parâmetros z e p = 10/14.

P5.16) Número de atendentes e tamanho da fila em um cartório

O número de atendentes que, em determinado momento, estão à disposição do público em um cartório pode variar desde 1 até 4, com 25% de chance para cada valor possível. Dado que há k atendentes naquele momento, o tamanho da fila única de pessoas que estão aguardando para serem atendidas segue uma distribuição de Poisson com parâmetro $\lambda = 12/k$. Alguém acaba de chegar ao cartório nesse momento. Calcule:

(a) A probabilidade de que haja 4 pessoas na fila.
(b) A probabilidade de que haja 2 atendentes, dado que há 4 pessoas na fila.
(c) A média e a variância do número de atendentes.
(d) A média e a variância do tamanho da fila.

P5.17) Distribuição Normal Bivariada *

Se $\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ é um vetor aleatório com distribuição normal bi-variada sendo: $E(X_1) = \mu_1$, $E(X_2) = \mu_2$, $Var(X_1) = \sigma_1^2$, $Var(X_2) = \sigma_2^2$ e $Cov(X_1, X_2) = \rho\, \sigma_1\, \sigma_2$, então a densidade conjunta de (X_1, X_2) é dada por

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\right],$$

para todo par (x_1, x_2) de R^2.

Mostre que:

(a) a densidade marginal de X_i é uma Normal (μ_i, σ_i^2), i = 1,2
(b) X_1 e X_2 são independentes se e só se X_1 e X_2 são não correlacionadas
(c) a densidade condicional de X_2 dado que $X_1 = x_1$ é uma Normal cuja média é $\mu_2 + \frac{\sigma_2}{\sigma_1}(X_1 - \mu_1)$ e cuja variância é $\sigma_2^2(1-\rho^2)$.

Obs.: Nas condições do enunciado acima, dizemos que X tem vetor de médias $\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$ e matriz de covariâncias $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$

Sugestão:

Pode ser provado que se $\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ é um vetor aleatório cuja distribuição é Normal bi-variada com os parâmetros listados acima, então podemos escrever:

$$\begin{cases} X_1 = \mu_1 + \sigma_1\sqrt{1-\rho^2}U_1 + \sigma_1\rho U_2 \\ X_2 = \mu_2 + \sigma_2 U_2 \end{cases}$$, ou seja, $$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} \sigma_1\sqrt{1-\rho^2} & \sigma_1\rho \\ 0 & \sigma_2 \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \end{bmatrix},$$

onde U_1 e U_2 são v.a.'s iid, ambas com distribuição Normal padrão.

Neste caso também é possível expressar U_1 e U_2 em função de X_1 e X_2:

$$\begin{bmatrix} U_1 \\ U_2 \end{bmatrix} = \frac{1}{\sigma_1\sigma_2\sqrt{1-\rho^2}} \begin{bmatrix} \sigma_2 & -\sigma_1\rho \\ 0 & \sigma_1\sqrt{1-\rho^2} \end{bmatrix} \begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \end{bmatrix}.$$

P5.18) Exportações e Importações

A distribuição conjunta das variáveis x = ln(exportações) e y = ln(importações) – onde exportações e importações, definidas para qualquer país do mundo, estão ambas expressas em bilhões de dólares relativos ao ano de 2007 – pode ser modelada como uma Normal bivariada com

Vetor de médias = $\begin{bmatrix} 2{,}17 \\ 2{,}38 \end{bmatrix}$ e Matriz de covariâncias = $\begin{bmatrix} 6{,}29 & 5{,}18 \\ 5{,}18 & 4{,}65 \end{bmatrix}$.

(a) Dado que um país exportou 150 bilhões de dólares em 2007, quanto em média ele deve ter importado nesse ano? Qual o desvio padrão?

(b) Dado que um país importou 100 bilhões de dólares em 2007, quanto em média ele deve ter exportado nesse ano? Qual o desvio padrão?

Obs.: Se Y = ln(X) tem distribuição Normal(μ; σ^2), então

E(X) = exp $(\mu + \frac{1}{2}\sigma^2)$ e Var(X) = $\exp(2\mu + 2\sigma^2) - \exp(2\mu + \sigma^2)$.