Cálculo das Probabilidades II - Lista 2 (Gabarito) - 2020/PLE

Prof. Hugo Carvalho

29/10/2020

Questão 1:

a) A covariância entre W_1 e W_2 é dada por $\mathbb{E}[W_1W_2] - \mathbb{E}[W_1]\mathbb{E}[W_2]$. Para calcular essas esperanças, utilizaremos a lei da esperança iterada:

$$\begin{aligned}
\mathbb{E}[W_1W_2] &= \mathbb{E}_Y[\mathbb{E}[W_1W_2|Y]] \\
&\stackrel{\star}{=} \mathbb{E}_Y[\mathbb{E}[W_1|Y]\mathbb{E}[W_2|Y]] \\
&= \mathbb{E}_Y[Y^2] \\
&= \mathbb{V}_Y(Y) + \mathbb{E}_Y[Y]^2 \\
&= b^2 + m^2. \\
\mathbb{E}[W_i] &= \mathbb{E}_Y[\mathbb{E}[W_i|Y]] \\
&= \mathbb{E}_Y[Y] \\
&= m, \text{ para } i = 1, 2,
\end{aligned}$$

onde a igualdade marcada com $\star$ se justifica por W_1 e W_2 serem condicionalmente independentes dado Y. Dessa forma, temos que

$$\mathbb{E}[W_1W_2] - \mathbb{E}[W_1]\mathbb{E}[W_2] = b^2 + m^2 - m^2 = b^2.$$

b) Não se pode afirmar que W_1 e W_2 são independentes, pois sua covariância é dada por b^2, que sendo a variância de uma distribuição de probabilidade, é usualmente um valor diferente de zero. E mesmo tendo $b = 0$, somente poderíamos afirmar que W_1 e W_2 são descorrelacionadas, o que não necessariamente implica em sua independência, sem maiores informações sobre sua distribuição conjunta.

c) *(Resposta livre, sendo esse um modelo)* Podemos interpretar o modelo do enunciado da seguinte forma:

i) Sorteie um valor $Y = y$ segundo uma distribuição $N(m, b^2)$;

ii) Sorteie dois valores $W_i = w_i$, para $i = 1, 2$, segundo uma distribuição $N(y, \sigma^2)$.

Ao repetir esse procedimento n vezes, obtemos duas sequência de valores $w_1^{(1)}, \dots, w_1^{(n)}$ e $w_2^{(1)}, \dots, w_2^{(n)}$, sendo que cada $w_i^{(k)}$ tem média $y^{(k)}$, que foi sorteada de acordo com $N(m, b^2)$. Analisemos dois casos:

- Assumindo que b seja um valor relativamente "pequeno", haverá pouca variabilidade nos $y^{(k)}$ sorteados, e todos os $w_i^{(k)}$ sorteados terão uma média parecida, próxima de m. Como o único fator que introduz dependência entre W_1 e W_2 é a sua média em comum, é razoável interpretar que elas serão pouco correlacionados nesse cenário.
- Agora, se b é um valor relativamente "grande", haverá maior variabilidade nos $y^{(k)}$ sorteados, e cada par $(w_1^{(k)}, w_2^{(k)})$ estará centrado em torno de tais valores mais discrepantes. Dessa forma, conhecer, digamos, $w_1^{(k)}$ nos dá uma "dica" sobre o comportamento de $w_2^{(k)}$, indicando uma maior correlação entre essas variáveis.

Questão 2:

a) A densidade conjunta de Z_1 e Z_2 é dada por

$$f(z_1, z_2) = \frac{1}{2\pi} \exp\left\{-\frac{1}{2}(z_1^2 + z_2^2)\right\},$$

e a transformação de (Z_1, Z_2) para (X_1, X_2) é dada através da função

$$\begin{cases} x_1 = g_1(z_1, z_2) = \sigma_1 z_1 + \mu_1 \\ x_2 = g_2(z_1, z_2) = \sigma_2[\rho z_1 + (1-\rho^2)^{1/2} z_2] + \mu_2. \end{cases}$$

Isolando z_1 na primeira expressão e depois substituindo na segunda para podermos isolar z_2, obtemos a inversa de tal função, dada por

$$\begin{cases} z_1 = h_1(x_1, x_2) = \frac{x_1-\mu_1}{\sigma_1} \\ z_2 = h_2(x_1, x_2) = \frac{1}{(1-\rho^2)^{1/2}} \left[\left(\frac{x_2-\mu_2}{\sigma_2}\right) - \rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\right] \end{cases}.$$

O Jacobiano de tal transformação é dado por

$$J = \det \begin{bmatrix} \frac{\partial h_1}{\partial x_1} & \frac{\partial h_1}{\partial x_2} \\ \frac{\partial h_2}{\partial x_1} & \frac{\partial h_2}{\partial x_2} \end{bmatrix} = \det \begin{bmatrix} 1/\sigma_1 & 0 \\ - & 1/\sigma_2(1-\rho^2)^{1/2} \end{bmatrix} = \frac{1}{\sigma_1\sigma_2(1-\rho^2)^{1/2}}.$$

Note que não precisamos calcular $\frac{\partial h_2}{\partial x_1}$ pois tal termo multiplicará um zero na hora de computar o determinante. Finalmente, utilizando o método do Jacobiano e substituindo as expressões para z_1 e z_2 em função de x_1 e x_2 na densidade conjunta de Z_1 e Z_2, obtemos

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

que após abrir o quadrado no termo referente a z_2 e fazer os devidos cancelamentos, nos dá a densidade conjunta de X_1 e X_2.

b) Utilizando a definição da função geradora de momentos conjunta e a transformação do item a), temos que:

$$\begin{aligned} \psi(\mathbf{t}) &= \mathbb{E}[\exp\{t_1X_1 + t_2X_2\}] \\ &= \mathbb{E}[\exp\{t_1(\sigma_1 Z_1 + \mu_1) + t_2(\sigma_2[\rho Z_1 + (1-\rho^2)^{1/2}Z_2] + \mu_2)\}] \\ &= \mathbb{E}[\exp\{t_1\sigma_1 Z_1 + t_1\mu_1 + t_2\sigma_2\rho Z_1 + t_2\sigma_2(1-\rho^2)^{1/2}Z_2 + t_2\mu_2\}] \\ &= \mathbb{E}[\exp\{(t_1\sigma_1 + t_2\sigma_2\rho)Z_1 + t_2\sigma_2(1-\rho^2)^{1/2}Z_2 + t_1\mu_1 + t_2\mu_2\}] \\ &\overset{(1)}{=} \mathbb{E}[\exp\{(t_1\sigma_1 + t_2\sigma_2\rho)Z_1\}]\mathbb{E}[\exp\{t_2\sigma_2(1-\rho^2)^{1/2}Z_2\}]\exp\{t_1\mu_1 + t_2\mu_2\} \\ &\overset{(2)}{=} \exp\{(t_1\sigma_1 + t_2\sigma_2\rho)^2/2\}\exp\{(t_2\sigma_2(1-\rho^2)^{1/2})^2/2\}\exp\{t_1\mu_1 + t_2\mu_2\} \\ &= \exp\left\{\frac{1}{2}\left[t_1^2\sigma_1^2 + t_2^2\sigma_2^2\rho^2 + 2t_1t_2\sigma_1\sigma_2\rho + t_2^2\sigma_2^2(1-\rho^2)\right] + t_1\mu_1 + t_2\mu_2\right\} \\ &\overset{(3)}{=} \exp\left\{\mathbf{t}^T\boldsymbol{\mu} + \frac{1}{2}\mathbf{t}^T\boldsymbol{\Sigma}\mathbf{t}\right\}, \text{ para } t_1, t_2 \in \mathbb{R}, \end{aligned}$$

onde as igualdades (1), (2) e (3) se justificam respectivamente por: independência de Z_1 e Z_2; utilização da função geradora de momentos de Z_1 e Z_2 informada no enunciado; regras da multiplicação matricial.

c) Fazendo $t_2 = 0$, obtemos a função geradora de momentos de X_1, dada por:

$$\psi(t_1, 0) = \exp\left\{\frac{1}{2}t_1^2\sigma_1^2 + t_1\mu_1\right\},$$

que reconhecemos como sendo de uma distribuição Normal de média μ_1 e variância σ_1^2. Analogamente, fazendo $t_1 = 0$, temos

$$\begin{aligned} \psi(0, t_2) &= \exp\left\{\frac{1}{2}[t_2^2\sigma_2^2\rho^2 + t_2^2\sigma_2^2(1-\rho^2)] + t_2\mu_2\right\} \\ &= \exp\left\{\frac{1}{2}t_2^2\sigma_2^2 + t_2\mu_2\right\}, \end{aligned}$$

reconhecida como sendo de uma distribuição Normal de média μ_2 e variância σ_2^2.

d) Escrever!

e) Se X_1 e X_2 são independentes, então são descorrelacionadas, como visto em aula. Agora, se X_1 e X_2 são descorrelacionadas, temos $\rho = 0$, de modo que a sua função densidade de probabilidade conjunta pode ser reescrita como

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

que pode ser fatorada como o produto das densidades marginais.

f) Note que, se $X_1 = x_1$, então temos que $Z_1 = z_1 = \dfrac{x_1 - \mu_1}{\sigma_1}$, de modo que a distribuição de $X_2|(X_1 = x_1)$ pode ser escrita como $X_2|(X_1 = x_1) = \sigma_2 \left[\rho\left(\dfrac{x_1-\mu_1}{\sigma_1}\right) + (1-\rho^2)^{1/2} Z_2\right] + \mu_2$, e daí temos que:

$$\begin{aligned}
\mathbb{E}[X_2|X_1 = x_1] &= \mathbb{E}\left[\sigma_2\left[\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right) + (1-\rho^2)^{1/2} Z_2\right] + \mu_2\right] \\
&= \sigma_2\left[\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right) + (1-\rho^2)^{1/2}\mathbb{E}[Z_2]\right] + \mu_2 \\
&= \mu_2 + \sigma_2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right); \\
\mathbb{V}(X_2|X_1 = x_1) &= \mathbb{V}\left(\sigma_2\left[\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right) + (1-\rho^2)^{1/2} Z_2\right] + \mu_2\right) \\
&= (1-\rho^2)\mathbb{V}(Z_2) \\
&= (1-\rho^2)\sigma_2^2.
\end{aligned}$$

g) Note que repetir o argumento do item f) e substituir $X_2 = x_2$ na segunda equação da transformação do item a) levará a uma relação mais complicada entre Z_1 e Z_2, com a qual será potencialmente tediosa de se trabalhar. Porém note que a distribuição conjunta de X_2 e X_1 também será normal bivariada, porém com os índices 1 e 2 trocados. Dessa forma, pelo resultado concluído no item f), temos que

$$\begin{aligned}
\mathbb{E}[X_1|X_2 = x_2] &= \mu_1 + \sigma_1\rho\left(\frac{x_2-\mu_2}{\sigma_2}\right) \\
\mathbb{V}(X_1|X_2 = x_2) &= (1-\rho^2)\sigma_1^2.
\end{aligned}$$

h) *(Resposta livre, sendo esse um modelo)* Caso não conheçamos o peso do indivíduo, a melhor previsão que podemos fazer para sua altura é a média, ou seja, $\mathbb{E}[X_2] = \mu_2$. Se temos conhecimento sobre seu peso, sabemos que o melhor preditor para a altura é a esperança condicional, ou seja, a melhor previsão para sua altura é dada por $\mathbb{E}[X_2|X_1 = x_1] = \mu_2 + \sigma_2\rho\left(\dfrac{x_1-\mu_1}{\sigma_1}\right)$.
Obs.: Aqui estamos usando resultado visto em aula onde consideramos otimalidade de um preditor com respeito ao erro quadrático médio.

Questão 3:

a) Usando a definição de função geradora de momentos, temos que:

$$\begin{aligned}
\psi(t) &= \mathbb{E}[e^{tX_i}] \\
&= e^{-t} \times \mathbb{P}(X_i = -1) + e^t \times \mathbb{P}(X_i = 1) \\
&= \frac{e^t + e^{-t}}{2}.
\end{aligned}$$

b) Por definição $(X_n)_{n\in\mathbb{N}}$ satisfaz à Lei Forte dos Grandes Números se $\overline{X}_n - \mathbb{E}[\overline{X}_n] \overset{qc}{\to} 0$. Sendo tal sequência identicamente distribuída e de média zero, temos que $\mathbb{E}[\overline{X}_n] = 0$, e portanto, a condição anterior é equivalente a termos $\overline{X}_n \overset{qc}{\to} 0$.

c) Por definição, temos que $\overline{X}_n \overset{qc}{\to} 0$ se e somente se $\mathbb{P}\left(\left\{\omega \in \Omega \;\middle|\; \lim_{n\to\infty} \overline{X}_n(\omega) = 0\right\}\right) = 1$, que é equivalente a termos $\mathbb{P}\left(\left\{\omega \in \Omega \;\middle|\; \lim_{n\to\infty} \overline{X}_n(\omega) \neq 0 \text{ ou não existe}\right\}\right) = 0$. Lembremos que a sequência numérica $(\overline{X}_n(\omega))_{n\in\mathbb{N}}$ converge para zero se e somente se para todo $\varepsilon > 0$ temos que $|\overline{X}_n| < \varepsilon$, para todo n suficientemente grande. Portanto, negar essa afirmação é dizer que existe $\varepsilon > 0$ tal que não temos $|\overline{X}_n| < \varepsilon$ para todo n suficientemente grande, ou seja, para tal ε temos que $|\overline{X}_n| \geq \varepsilon$ para infinitos valores de n. Dessa forma, conclui-se a equivalência entre $\overline{X}_n \overset{qc}{\to} 0$ e $\mathbb{P}(|\overline{X}_n| \geq \varepsilon \text{ infinitas vezes}) = 0$.

d) Como cada variável aleatória X_i é simétrica em torno do zero (ou seja, X_i e $-X_i$ têm a mesma distribuição), temos que $\overline{X}_n$ também o é (ou seja, $\overline{X}_n$ e $-\overline{X}_n$ também têm a mesma distribuição), e portanto:

$$\begin{aligned}\mathbb{P}(|\overline{X}_n| \geq \varepsilon) &= \mathbb{P}(\overline{X}_n \geq \varepsilon) + \mathbb{P}(\overline{X}_n \leq -\varepsilon)\\ &= \mathbb{P}(\overline{X}_n \geq \varepsilon) + \mathbb{P}(-\overline{X}_n \geq \varepsilon)\\ &= 2\mathbb{P}(\overline{X}_n \geq \varepsilon),\end{aligned}$$

onde na primeira igualdade usamos a definição da função módulo e na terceira igualdade usamos a simetria de $\overline{X}_n$ em torno de zero.

e) Argumentemos igualdade por igualdade:

(1): Simplesmente multiplica-se ambos os lados da desigualdade por n.

(2): Note que função $z \mapsto e^{tz}$ é bijetiva, e sendo $t > 0$, é também crescente. Ambos os fatos justificam a igualdade (2) e a preservação do sinal da desigualdade dentro da probabilidade.

(3): Sendo a variável aleatória e^{tS_n} positiva, podemos aplicar a desigualdade de Markov, justificando assim essa igualdade.

(4): Como as variáveis aleatórias $X_1, \ldots, X_n$ são independentes, a função geradora de momentos de S_n é o produto das respectivas funções geradoras de momentos das X_i. Como elas são identicamente distribuídas, justifica-se a potência n.

f) Novamente, argumentemos a cada passo:

(1): Do item e), temos que $\mathbb{P}(\overline{X}_n \geq \varepsilon) \leq e^{-tn\varepsilon}\left[\frac{e^t + e^{-t}}{2}\right]^n$, que utilizando a desigualdade dada no enunciado, nos dá $\mathbb{P}(\overline{X}_n \geq \varepsilon) \leq e^{-tn\varepsilon}e^{nt^2/2}$. Como tal desigualdade vale para todo $t > 0$, podemos tomar o mínimo em $t > 0$ e concluir que $\mathbb{P}(\overline{X}_n \geq \varepsilon) \leq \min_{t>0} e^{-tn\varepsilon}e^{nt^2/2} = \min_{t>0} e^{-n[t\varepsilon - t^2/2]}$.

(2): Pelo fato da exponencial ser uma função contínua, convexa, basta tomar o mínimo do seu argumento. Derivando e igualando a zero, temos que: $\frac{d}{dt}[-n(t\varepsilon - t^2/2)] = -n[\varepsilon - t] = 0$, de modo que o mínimo de tal função é atingido para $t = \varepsilon$. Note que, por ser um polinômio de grau dois cujo coeficiente associado ao termo quadrático é positivo, de fato esse é o mínimo da função. Assim, ao substituir o valor de $t = \varepsilon$ na função, obtemos a igualdade de tal passo.

(3): Somente reescrevemos o termo $e^{-n\varepsilon^2/2}$ de modo a ressaltar a potência n, ou seja, $e^{-n\varepsilon^2/2} = (e^{-\varepsilon^2/2})^n$.

g) Relembrando os itens b) e c), $(X_n)_{n\in\mathbb{N}}$ satisfaz à Lei Forte dos Grandes Números se e somente se $\overline{X}_n \xrightarrow{qc} 0$, e isso é equivalente a $\mathbb{P}(|\overline{Y}_n| \geq \varepsilon$ infinitas vezes$) = 0$. Nos itens d), e) e f) concluímos que $\mathbb{P}(|\overline{X}_n| \geq \varepsilon) \leq 2[e^{-(\varepsilon^2/2+\ln(2))}]^n$. Note que $\sum_{n=1}^{\infty}[e^{-(\varepsilon^2/2+\ln(2))}]^n < \infty$, pois é a soma de uma PG de razão menor que 1, em módulo. Portanto, pelos Lemas de Borel-Cantelli, temos que $\mathbb{P}(|\overline{X}_n| \geq \varepsilon$ infinitas vezes$) = 0$, e portanto, $\overline{X}_n \xrightarrow{qc} 0$, de modo que a Lei Forte dos Grandes Números é provada nesse caso.

h) *(Resposta livre, sendo esse um modelo)* A utilização da desigualdade de Chebyshev no item e) nos daria:

$$\mathbb{P}(\overline{X}_n \geq \varepsilon) \leq \frac{\mathbb{V}(\overline{X}_n)}{\varepsilon^2} = \frac{\frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}(X_i)}{\varepsilon^2} = \frac{\frac{1}{n^2}\sum_{i=1}^{n}1}{\varepsilon^2} = \frac{1}{n\varepsilon}.$$

Essa estimativa nos levaria a ter que usar a segunda parte dos Lemas de Borel-Cantelli, já que $\sum_{n=1}^{\infty}\frac{1}{n\varepsilon} = \infty$, porém esse resultado requer que os eventos $\{\overline{X}_n \geq \epsilon\}$ sejam independentes, o que não vale nesse cenário. Portanto, a utilização da desigualdade de Chebyshev não leva a conclusão alguma.