Universidade de São Paulo
Instituto de Matemática e Estatística
Aluno: Guilherme Dias Vianna

Professor: Luis Fantozzi
Monitor: -
Disciplina: Verão de Estatística

Lista 2 (Integração e Expectativa)

(1) Seja $(\Omega, \Sigma, \mathbb{P})$ um espaço de probabilidade, e $Z_n \geq 0$, $n \in \mathbb{N}$, uma sequência de variáveis aleatórias não negativas.

(a) Mostre que $\mathbb{E}[\sum_{n=1}^{\infty} Z_n] = \sum_{n=1}^{\infty} \mathbb{E}[Z_n]$.

Resolução:

Defina as somas parciais $S_N := \sum_{n=1}^{N} Z_n$ e a soma (possivelmente estendida) $S := \sum_{n=1}^{\infty} Z_n$. Como $Z_n \geqslant 0$, temos $S_N(\omega) \uparrow S(\omega)$ ponto a ponto.
Pelo Teorema da Convergência Monótona,

$$\mathbb{E}[S] \;=\; \lim_{N\to\infty} \mathbb{E}[S_N].$$

Mostremos que, para quaisquer $X, Y \geqslant 0$ mensuráveis, $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.
Para funções simples não negativas $X = \sum_{i=1}^{k} a_i \mathbf{1}_{\{A_i\}}$ e $Y = \sum_{j=1}^{m} b_j \mathbf{1}_{\{B_j\}}$, refinamos para a partição disjunta $\{A_i \cap B_j\}_{i,j}$ e usamos a definição de integral de Lebesgue para funções simples:

$$\mathbb{E}[X+Y] = \sum_{i,j} (a_i + b_j)\, \mathbb{P}(A_i \cap B_j) = \sum_i a_i\, \mathbb{P}(A_i) + \sum_j b_j\, \mathbb{P}(B_j) = \mathbb{E}[X] + \mathbb{E}[Y].$$

Para X, Y gerais não negativas, tome as aproximações em escada $X_k := s_k \circ X$ e $Y_k := s_k \circ Y$. Então $X_k \uparrow X$, $Y_k \uparrow Y$ e, para cada k, $\mathbb{E}[X_k + Y_k] = \mathbb{E}[X_k] + \mathbb{E}[Y_k]$ (caso simples). Aplicando novamente a convergência monótona,

$$\mathbb{E}[X+Y] = \lim_{k\to\infty} \mathbb{E}[X_k + Y_k] = \lim_{k\to\infty} \mathbb{E}[X_k] + \lim_{k\to\infty} \mathbb{E}[Y_k] = \mathbb{E}[X] + \mathbb{E}[Y].$$

Por indução, a aditividade vale para qualquer soma finita de não negativas: $\mathbb{E}\left[\sum_{n=1}^{N} Z_n\right] = \sum_{n=1}^{N} \mathbb{E}[Z_n]$ (valores em $[0, \infty]$).
Combinando os dois passos:

$$\mathbb{E}\left[\sum_{n=1}^{\infty} Z_n\right] = \lim_{N\to\infty} \mathbb{E}\left[\sum_{n=1}^{N} Z_n\right] = \lim_{N\to\infty} \sum_{n=1}^{N} \mathbb{E}[Z_n] = \sum_{n=1}^{\infty} \mathbb{E}[Z_n].$$

A igualdade é válida sem assumir integrabilidade individual dos Z_n; ambos os lados são entendidos em $[0, \infty]$.

(b) Mostre que $\mathbb{E}[\sum_{n=1}^{\infty} Z_n] < \infty \implies \mathbb{P}[\{\omega : \sum_{n=1}^{\infty} Z_n(\omega) < \infty\}] = 1$.

Resolução:

Defina as somas parciais $S_N := \sum_{n=1}^{N} Z_n$ e $S := \sum_{n=1}^{\infty} Z_n = \lim_{N\to\infty} S_N$. Como $S_N \uparrow S$ ponto a ponto e cada S_N é mensurável, S é mensurável e $S \geq 0$.
Veja que $\mathbb{E}[S] < \infty$. Pelo lema de integração de Lebesgue para funções não negativas dos slides, "Se $f \geq 0$ é mensurável e $\mu(f) < \infty$, então $\mu(\{\omega : f(\omega) = \infty\}) = 0$", tomando $\mu = \mathbb{P}$ e $f = S$ obtemos $\mathbb{P}[S = \infty] = 0$.
Logo,

$$\mathbb{P}\left(\left\{\omega : \sum_{n=1}^{\infty} Z_n(\omega) < \infty\right\}\right) = 1,$$

isto é, a soma $\sum_{n=1}^{\infty} Z_n$ é finita $\mathbb{P}$-quase certamente.

(c) Mostre que $\mathbb{P}[\{\omega : \sum_{n=1}^{\infty} Z_n(\omega) < \infty\}] = 1 \implies \mathbb{P}[\{\omega : Z_n(\omega) \nrightarrow 0\}] = 0$.

Resolução:

Sejam $Z_n \geqslant 0$ e

$$E := \left\{\omega : \sum_{n=1}^{\infty} Z_n(\omega) < \infty\right\}, \qquad \mathbb{P}(E) = 1.$$

Para $\omega \in E$, as somas parciais $S_N(\omega) := \sum_{n=1}^{N} Z_n(\omega)$ são crescentes e $S_N(\omega) \uparrow S(\omega) := \sum_{n=1}^{\infty} Z_n(\omega) < \infty$.
Considere as caudas $T_n(\omega) := \sum_{k=n}^{\infty} Z_k(\omega) = S(\omega) - S_{n-1}(\omega)$. Então $T_{n+1}(\omega) \leqslant T_n(\omega)$ e $T_n(\omega) \downarrow 0$ quando $n \to \infty$. Além disso,

$$0 \leqslant Z_n(\omega) \leqslant T_n(\omega) \quad \text{para todo } n,$$

logo $Z_n(\omega) \to 0$ para todo $\omega \in E$.
Portanto $E \subseteq \{\omega : Z_n(\omega) \to 0\}$, e

$$\mathbb{P}\{\omega : Z_n(\omega) \nrightarrow 0\} = 1 - \mathbb{P}\{\omega : Z_n(\omega) \to 0\} \leqslant 1 - \mathbb{P}(E) = 0.$$

Conclui-se que $\mathbb{P}\{\omega : Z_n(\omega) \nrightarrow 0\} = 0$.

Comentário: o item (c) é exatamente o análogo probabilístico do "teste do termo geral" para séries: para cada ω, $\sum_{n\geqslant 1} Z_n(\omega)$ é uma série de números reais, se converge, então $Z_n(\omega) \to 0$.

2 Seja $(\Omega, \Sigma, \mathbb{P})$ um espaço de probabilidade, e X uma variável aleatória. Mostre que, para todo $f : \mathbb{R} \mapsto \mathbb{R}$ mensurável tal que $\mathbb{E}[|f(X)|] < \infty$.

$$\mathbb{E}[f(X)] = \int f(x)\mathbb{P}_X(dx)\,,$$

onde $\mathbb{P}_X$ é a medida de probabilidade induzida por X sobre $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. *Dica:* mostre para funções f simples, depois aproxime para funções não-negativas pelo teorema da covnergência monótona, depois use a definição de integral para funções gerais.

Solução:

Caso simples.
Se f é simples, $f = \sum_{i=1}^{k} a_i \mathbf{1}_{A_i}$ com $A_i \in \mathcal{B}(\mathbb{R})$ disjuntos,

$$\mathbb{E}[f(X)] = \sum_{i=1}^{k} a_i\, \mathbb{E}[\mathbf{1}_{A_i}(X)] = \sum_{i=1}^{k} a_i\, \mathbb{P}(X \in A_i) = \sum_{i=1}^{k} a_i\, \mathbb{P}_X(A_i) = \int f\, \mathbb{P}_X(\mathrm{d}x),$$

onde usamos a definição de integral de Lebesgue para funções simples e a definição de $\mathbb{P}_X$.
Caso $f \geqslant 0$.
Tome $(f_n)_n$ funções simples não negativas tais que $f_n \uparrow f$ em $\mathbb{R}$ (por exemplo, aproximações por funções-escada). Então $f_n(X) \uparrow f(X)$ em Ω. Pelo Teorema da Convergência Monótona (TCM),

$$\mathbb{E}[f(X)] = \lim_{n\to\infty} \mathbb{E}[f_n(X)] = \lim_{n\to\infty} \int f_n\, \mathbb{P}_X(\mathrm{d}x) = \int f\, \mathbb{P}_X(\mathrm{d}x),$$

onde a segunda igualdade vem do caso simples e a última do TCM aplicado no espaço $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_X)$.
Caso geral integrável.
Escreva $f = f^+ - f^-$, com $f^+, f^- \geqslant 0$. Como $\mathbb{E}[|f(X)|] < \infty$, temos $\mathbb{E}[f^\pm(X)] < \infty$. Aplicando o caso $f \geqslant 0$ a f^+ e f^-,

$$\mathbb{E}[f(X)] = \mathbb{E}[f^+(X)] - \mathbb{E}[f^-(X)] = \int f^+\, \mathbb{P}_X(\mathrm{d}x) - \int f^-\, \mathbb{P}_X(\mathrm{d}x) = \int f\, \mathbb{P}_X(\mathrm{d}x).$$

Comentário: Veja que esse problema indica que não é necessário, ao lidarmos com funções de variáveis aleatórias, saber muito mais do que a distribuição da variável aleatória original. A esperança em questão, pela definição, seria calculada como:

$$\mathbb{E}[f(X)] = \int_\Omega f(X(\omega))\, \mathbb{P}(\mathrm{d}\omega) = \int_\mathbb{R} y\, \mathbb{P}_{f(X)}(\mathrm{d}y),$$

mas **nada sabemos (pelo exposto em aula) sobre** $\mathbb{P}_{f(X)}$, esse teorema mostra que isso é uma complicação desnecessária e pode ser facilmente ignorada (tanto que na prática muitas pessoas esquecem que isso é um resultado que precisa ser provado e tomam isso como parte da definição), por isso coloquialmente é chamado de **Teorema do Estatístico Inconsciente**.

3 Demonstre a desigualdade de Jensen *estrita*: se X é uma variável aleatória integrável sobre um espaço de probabilidade $(\Omega, \Sigma, \mathbb{P})$, $f : A \mapsto \mathbb{R}$ uma função **estritamente** convexa sobre $A \subseteq \mathbb{R}$ convexo e aberto, com $\mathbb{P}[X \in A] = 1$ e $\mathbb{E}[|X|] < \infty$, $\mathbb{E}[|f(X)|] < \infty$. Se $\mathbb{P}[X \neq \mathbb{E}[X]] > 0$, então:

$$f(\mathbb{E}[X]) < \mathbb{E}[f(X)]$$

Resolução:

Denote $m := \mathbb{E}[X]$ (note que, como A é um intervalo aberto e $X \in A$ q.c., necessariamente $m \in A$). Suponha $\mathbb{P}[X \neq m] > 0$. Mostraremos que

$$f(m) \;<\; \mathbb{E}[f(X)].$$

Considere os eventos $B := \{X \leq m\}$ e $C := \{X > m\}$. Da hipótese $\mathbb{P}[X \neq m] > 0$ e do fato de $\mathbb{E}[X] = m$, segue que $\mathbb{P}(B) \in (0,1)$ e $\mathbb{P}(C) \in (0,1)$ (se, por exemplo, $\mathbb{P}(C) = 0$, então $X \leq m$ q.c. e, com $\mathbb{P}(X < m) > 0$, teríamos $\mathbb{E}[X] < m$, contradição). Defina

$$m_1 := \mathbb{E}[X \,|\, B], \qquad m_2 := \mathbb{E}[X \,|\, C].$$

Como $X \leq m$ em B e $\mathbb{P}(X < m) > 0$, temos $m_1 < m$; analogamente, $m_2 > m$. Além disso,

$$m = \mathbb{E}[X] = \mathbb{E}[X\mathbf{1}_B] + \mathbb{E}[X\mathbf{1}_C] = \mathbb{P}(B)\, m_1 + \mathbb{P}(C)\, m_2.$$

Pela estrita convexidade de f em A e pelo fato de $m_1 \neq m_2$ e $\mathbb{P}(B) \in (0,1)$,

$$f(m) = f(\mathbb{P}(B)m_1 + \mathbb{P}(C)m_2) \;<\; \mathbb{P}(B)f(m_1) + \mathbb{P}(C)f(m_2).$$

Por outro lado, aplicando Jensen condicional à σ-álgebra gerada pela partição $\{B, C\}$,

$$\mathbb{E}[f(X) \,|\, \sigma(B)] \;\geq\; f(\mathbb{E}[X \,|\, \sigma(B)]) = f(m_1\mathbf{1}_B + m_2\mathbf{1}_C) \quad \text{q.c.}$$

Tomando esperanças,

$$\mathbb{E}[f(X)] \;\geq\; \mathbb{P}(B)\, f(m_1) + \mathbb{P}(C)\, f(m_2).$$

Combinando, obtemos

$$f(\mathbb{E}[X]) \;<\; \mathbb{E}[f(X)],$$

como desejado.

4 Seja $(\Omega, \Sigma, \mathbb{P})$ um espaço de probabilidade. Para $A, B \in \mathcal{L}_2(\Omega, \Sigma, \mathbb{P})$ a covariância entre A e B é definida como:

$$\mathbb{C}(A, B) := \mathbb{E}[AB] - \mathbb{E}[A]\mathbb{E}[B]\,.$$

Para duas variáveis aleatórias X, Y definidas em $(\Omega, \Sigma, \mathbb{P})$, mostre que X é independente de Y se, e somente se, para quaisquer funções $h : \mathbb{R} \mapsto \mathbb{R}$ e $g : \mathbb{R} \mapsto \mathbb{R}$ limitadas e mensuráveis:

$$\mathrm{C}(h(X), g(Y)) = 0\,.$$

Dica: para a direção "se", observe que, para todo $B \in \mathcal{B}(\mathbb{R})$, $\mathbb{P}[X^{-1}(B)] = \mathbb{E}[\mathbf{1}_{X^{-1}(B)}] = \mathbb{E}[\mathbf{1}_B(X)]$. Para a outra direção, mostre para funções simples e depois aproxime para funções limitadas por funções-escada.

Resolução:

(⇐) Suponha que para quaisquer h, g limitadas e mensuráveis temos $\mathbb{E}[h(X)g(Y)] = \mathbb{E}[h(X)]\mathbb{E}[g(Y)]$. Tome $A, B \in \mathcal{B}(\mathbb{R})$ e escolha $h = \mathbf{1}_A$, $g = \mathbf{1}_B$. Então

$$\mathbb{E}\left[\mathbf{1}_A(X)\mathbf{1}_B(Y)\right] = \mathbb{E}\left[\mathbf{1}_A(X)\right]\mathbb{E}\left[\mathbf{1}_B(Y)\right].$$

Usando que $\mathbf{1}_A(X) = \mathbf{1}_{X^{-1}(A)}$ e $\mathbb{E}[\mathbf{1}_E] = \mathbb{P}(E)$ para qualquer evento $E \in \Sigma$, obtemos

$$\mathbb{P}\left(X \in A,\ Y \in B\right) = \mathbb{P}\left(X \in A\right)\mathbb{P}\left(Y \in B\right).$$

Como A, B são arbitrários em $\mathcal{B}(\mathbb{R})$, segue que X e Y são independentes.

(⇒) Suponha agora que X e Y são independentes, isto é, para todo $A, B \in \mathcal{B}(\mathbb{R})$,

$$\mathbb{P}\left(X \in A,\ Y \in B\right) = \mathbb{P}\left(X \in A\right)\mathbb{P}\left(Y \in B\right).$$

Primeiro verificamos a fatoração de esperanças para funções simples e depois passamos ao caso geral por aproximação por funções-escada e pelo teorema da convergência limitada.

Indicadoras.

Para $A, B \in \mathcal{B}(\mathbb{R})$,

$$\mathbb{E}\left[\mathbf{1}_A(X)\mathbf{1}_B(Y)\right] = \mathbb{P}\left(X \in A,\ Y \in B\right) = \mathbb{P}\left(X \in A\right)\mathbb{P}\left(Y \in B\right) = \mathbb{E}\left[\mathbf{1}_A(X)\right]\mathbb{E}\left[\mathbf{1}_B(Y)\right].$$

Funções simples.

Sejam $h = \sum_{i=1}^{k} a_i \mathbf{1}_{A_i}$ e $g = \sum_{j=1}^{m} b_j \mathbf{1}_{B_j}$ funções simples com $A_i, B_j \in \mathcal{B}(\mathbb{R})$ (podemos supor A_i e B_j disjuntos). Então

$$h(X)g(Y) = \sum_{i=1}^{k}\sum_{j=1}^{m} a_i b_j\, \mathbf{1}_{A_i}(X)\mathbf{1}_{B_j}(Y),$$

e, pela linearidade da esperança,

$$\mathbb{E}\left[h(X)g(Y)\right] = \sum_{i=1}^{k}\sum_{j=1}^{m} a_i b_j\, \mathbb{E}\left[\mathbf{1}_{A_i}(X)\mathbf{1}_{B_j}(Y)\right].$$

Pela independência (no caso de indicadoras), $\mathbb{E}\left[\mathbf{1}_{A_i}(X)\mathbf{1}_{B_j}(Y)\right] = \mathbb{E}\left[\mathbf{1}_{A_i}(X)\right]\mathbb{E}\left[\mathbf{1}_{B_j}(Y)\right]$, logo

$$\begin{aligned}
\mathbb{E}\left[h(X)g(Y)\right] &= \sum_{i=1}^{k}\sum_{j=1}^{m} a_i b_j\, \mathbb{E}\left[\mathbf{1}_{A_i}(X)\right]\mathbb{E}\left[\mathbf{1}_{B_j}(Y)\right] \\
&= \left(\sum_{i=1}^{k} a_i\, \mathbb{E}\left[\mathbf{1}_{A_i}(X)\right]\right)\left(\sum_{j=1}^{m} b_j\, \mathbb{E}\left[\mathbf{1}_{B_j}(Y)\right]\right) \\
&= \mathbb{E}\left[h(X)\right]\mathbb{E}\left[g(Y)\right].
\end{aligned}$$

Funções limitadas e mensuráveis. Sejam agora $h, g : \mathbb{R} \to \mathbb{R}$ limitadas e mensuráveis. Escolha $M_h = \|h\|_\infty$ e defina $u = h + M_h$, de modo que $u \geqslant 0$ e $u \leqslant 2M_h$. Considere as funções-escada s_n (definidas em $[0, \infty)$) e ponha

$$u_n = s_n \circ u, \qquad h_n = u_n - M_h.$$

Então u_n é simples, $0 \leqslant u_n \leqslant u$ e $u_n \uparrow u$ ponto a ponto; portanto h_n é simples, $|h_n| \leqslant M_h$ e $h_n \to h$ ponto a ponto. Fazendo o mesmo para g, obtemos funções simples g_n com $|g_n| \leqslant M_g = \|g\|_\infty$ e $g_n \to g$ ponto a ponto.
Como h_n e g_n são simples, já mostramos que para todo n,

$$\mathbb{E}\left[h_n(X)g_n(Y)\right] = \mathbb{E}\left[h_n(X)\right]\mathbb{E}\left[g_n(Y)\right].$$

Além disso, temos convergência quase certa $h_n(X) \to h(X)$, $g_n(Y) \to g(Y)$ e $h_n(X)g_n(Y) \to h(X)g(Y)$, e as cotas uniformes

$$|h_n(X)| \leqslant M_h, \qquad |g_n(Y)| \leqslant M_g, \qquad |h_n(X)g_n(Y)| \leqslant M_h M_g \quad \text{q.c.}$$

Pelo Teorema da Convergência Limitada, concluímos que

$$\mathbb{E}\left[h_n(X)\right] \to \mathbb{E}\left[h(X)\right], \qquad \mathbb{E}\left[g_n(Y)\right] \to \mathbb{E}\left[g(Y)\right], \qquad \mathbb{E}\left[h_n(X)g_n(Y)\right] \to \mathbb{E}\left[h(X)g(Y)\right].$$

Tomando limite na identidade para cada n, obtemos

$$\mathbb{E}\left[h(X)g(Y)\right] = \mathbb{E}\left[h(X)\right]\mathbb{E}\left[g(Y)\right],$$

isto é, $C\left(h(X), g(Y)\right) = 0$.
Combinando as duas implicações, provamos o que foi pedido.

5 Seja $(\Omega, \Sigma, \mathbb{P})$ e $Y \in L^2(\Omega, \Sigma, \mathbb{P})$. Seja $\mathcal{G}$ uma sub-σ-álgebra de Σ.

(a) Mostre que $\mathbb{E}[Y|\mathcal{G}] \in L^2(\Omega, \Sigma, \mathbb{P})$.

Resolução:

Como $Y \in L^2(\Omega, \Sigma, \mathbb{P})$, pela monotonicidade das normas L^p (com $p = 1$, $q = 2$) temos $Y \in L^1$. Logo $\mathbb{E}\left[Y \mid \mathcal{G}\right]$ está bem definida.
Pela desigualdade de Jensen condicional, para a função convexa $\phi(t) = t^2$ e usando que $\mathbb{E}\left[|Y|^2\right] < \infty$,

$$\mathbb{E}\left[Y^2 \mid \mathcal{G}\right] \geqslant \left(\mathbb{E}\left[Y \mid \mathcal{G}\right]\right)^2 \quad \text{quase certamente.}$$

Tomando esperança em ambos os lados e usando preservação da esperança,

$$\mathbb{E}\left[\left(\mathbb{E}\left[Y \mid \mathcal{G}\right]\right)^2\right] \leqslant \mathbb{E}\left[\mathbb{E}\left[Y^2 \mid \mathcal{G}\right]\right] = \mathbb{E}\left[Y^2\right] < \infty.$$

Portanto, $\mathbb{E}\left[Y \mid \mathcal{G}\right] \in L^2(\Omega, \Sigma, \mathbb{P})$.

(b) Mostre que $\mathbb{E}[Y|\mathcal{G}]$ é a única (a não ser num evento de probabilidade zero) solução ao problema de minimização:

$$\min_{S \in L^2(\Omega, \mathcal{G}, \mathbb{P})} \mathbb{E}[(Y - S)^2]$$

Resolução:

Fixe $S \in L^2(\Omega, \mathcal{G}, \mathbb{P})$ e defina:

$$M := \mathbb{E}\left[Y \mid \mathcal{G}\right].$$

Então $M \in L^2$ pelo item (a), M é $\mathcal{G}$-mensurável, e $S - M \in L^2(\Omega, \mathcal{G}, \mathbb{P})$. Expandindo,

$$(Y - S)^2 = (Y - M + M - S)^2 = (Y - M)^2 + (S - M)^2 + 2(Y - M)(M - S).$$

Tomando esperança,

$$\mathbb{E}\left[(Y-S)^2\right] = \mathbb{E}\left[(Y-M)^2\right] + \mathbb{E}\left[(S-M)^2\right] + 2\,\mathbb{E}\left[(Y-M)(M-S)\right].$$

Mostremos que o termo cruzado é nulo. Primeiro, pela linearidade e pela propriedade de mensurabilidade,

$$\mathbb{E}\left[Y-M \mid \mathcal{G}\right] = \mathbb{E}\left[Y \mid \mathcal{G}\right] - \mathbb{E}\left[M \mid \mathcal{G}\right] = M - M = 0 \quad \text{quase certamente.}$$

Como $M-S$ é $\mathcal{G}$-mensurável e $Y-M \in L^2$, $M-S \in L^2$, podemos usar a propriedade "extraindo o que é conhecido" na forma L^p–L^q (com $p=q=2$) para obter

$$\mathbb{E}\left[(Y-M)(M-S) \mid \mathcal{G}\right] = (M-S)\,\mathbb{E}\left[Y-M \mid \mathcal{G}\right] = (M-S)\cdot 0 = 0 \quad \text{q.c.}$$

Logo,

$$\mathbb{E}\left[(Y-M)(M-S)\right] = \mathbb{E}\left[\mathbb{E}\left[(Y-M)(M-S) \mid \mathcal{G}\right]\right] = 0.$$

Substituindo na identidade anterior,

$$\mathbb{E}\left[(Y-S)^2\right] = \mathbb{E}\left[(Y-M)^2\right] + \mathbb{E}\left[(S-M)^2\right] \geqslant \mathbb{E}\left[(Y-M)^2\right].$$

Assim, $M = \mathbb{E}\left[Y \mid \mathcal{G}\right]$ é o minimizador de $f(s) = \mathbb{E}[(Y-S)^2]$
Para a unicidade (a menos de um evento de probabilidade zero), suponha que $S^\star \in L^2(\Omega, \mathcal{G}, \mathbb{P})$ também atinja o mínimo. Então

$$\mathbb{E}\left[(Y-S^\star)^2\right] = \mathbb{E}\left[(Y-M)^2\right] \quad \Rightarrow \quad \mathbb{E}\left[(S^\star-M)^2\right] = 0.$$

Como $(S^\star - M)^2 \geqslant 0$, do lema básico sobre integrais de funções não negativas conclui-se que

$$\mathbb{P}\left[(S^\star-M)^2 > 0\right] = 0,$$

isto é, $S^\star = M$ quase certamente.

(c) Mostre que, quando $\mathcal{G} = \{\emptyset, \Omega\}$, $\mathbb{E}[Y|\mathcal{G}] = \mathbb{E}[Y]$, e que, quando $\mathcal{G} = \Sigma$, $\mathbb{E}[Y|\mathcal{G}] = Y$. À luz do item anterior, qual a interpretação desses resultados?

Resolução:

Se $\mathcal{G} = \{\varnothing, \Omega\}$, toda variável aleatória $\mathcal{G}$-mensurável é constante quase certamente. Seja $Z = \mathbb{E}\left[Y \mid \mathcal{G}\right]$. Então existe $c \in \mathbb{R}$ tal que $Z = c$ q.c. Pela definição da esperança condicional,

$$\mathbb{E}\left[Z\right] = \mathbb{E}\left[Y\right] \quad \Rightarrow \quad c = \mathbb{E}\left[Y\right],$$

logo $\mathbb{E}\left[Y \mid \mathcal{G}\right] = \mathbb{E}\left[Y\right]$ q.c.
Se $\mathcal{G} = \Sigma$, então Y é $\mathcal{G}$-mensurável e, pela propriedade de mensurabilidade,

$$\mathbb{E}\left[Y \mid \Sigma\right] = Y \quad \text{q.c.}$$

Interpretação pelo item (b): quando $\mathcal{G} = \{\varnothing, \Omega\}$, o agente não tem informação após o sorteio (só pode escolher aproximações constantes), e a melhor aproximação quadrática é a constante $\mathbb{E}[Y]$. Quando $\mathcal{G} = \Sigma$, o agente observa toda a informação possível (pode escolher $S = Y$), e o erro quadrático mínimo é zero.

(d) Seja $E \in \Sigma$, com $1 > \mathbb{P}[E] > 0$. Tome $\mathcal{G} = \{\emptyset, E, E^\complement, \Omega\}$. Mostre que a função:

$$g(\omega) = \begin{cases} \frac{\mathbb{E}[Y\mathbf{1}_E]}{\mathbb{P}[E]}, & \text{se } \omega \in E \\ \frac{\mathbb{E}[Y\mathbf{1}_{E^\complement}]}{\mathbb{P}[E^\complement]}, & \text{se } \omega \in E^\complement \end{cases},$$

é uma versão de $\mathbb{E}[Y|\mathcal{G}]$. Qual a interpretação desse resultado, à luz do item (b)?

Resolução:

Defina $Z : \Omega \to \mathbb{R}$ por

$$Z(\omega) = \frac{\mathbb{E}[Y\mathbf{1}_E]}{\mathbb{P}[E]}\mathbf{1}_E(\omega) + \frac{\mathbb{E}[Y\mathbf{1}_{E^c}]}{\mathbb{P}[E^c]}\mathbf{1}_{E^c}(\omega).$$

Como Z é constante em E e em E^c, temos que Z é $\mathcal{G}$-mensurável, onde $\mathcal{G} = \{\varnothing, E, E^c, \Omega\}$. Além disso, $Z \in L^2$ (em particular, é integrável): por Cauchy–Schwarz,

$$\left(\mathbb{E}[Y\mathbf{1}_E]\right)^2 \leqslant \mathbb{E}\left[Y^2\mathbf{1}_E\right]\mathbb{P}[E] \leqslant \mathbb{E}\left[Y^2\right]\mathbb{P}[E],$$

logo

$$\mathbb{E}\left[Z^2\right] = \frac{\left(\mathbb{E}[Y\mathbf{1}_E]\right)^2}{\mathbb{P}[E]} + \frac{\left(\mathbb{E}[Y\mathbf{1}_{E^c}]\right)^2}{\mathbb{P}[E^c]} \leqslant \mathbb{E}\left[Y^2\right] + \mathbb{E}\left[Y^2\right] < \infty.$$

Para verificar que Z é uma versão de $\mathbb{E}[Y \mid \mathcal{G}]$, basta checar a identidade de definição nos geradores E e E^c. Para $A = E$,

$$\mathbb{E}[Z\mathbf{1}_E] = \mathbb{E}\left[\frac{\mathbb{E}[Y\mathbf{1}_E]}{\mathbb{P}[E]}\mathbf{1}_E\right] = \frac{\mathbb{E}[Y\mathbf{1}_E]}{\mathbb{P}[E]}\mathbb{P}[E] = \mathbb{E}[Y\mathbf{1}_E].$$

Analogamente, para $A = E^c$,

$$\mathbb{E}[Z\mathbf{1}_{E^c}] = \mathbb{E}[Y\mathbf{1}_{E^c}].$$

Por linearidade, a identidade vale também para $A = \Omega$ e $A = \varnothing$. Logo Z satisfaz a caracterização e é uma versão de $\mathbb{E}[Y \mid \mathcal{G}]$.

Interpretação pelo item (b): ao restringir S a ser $\mathcal{G}$-mensurável, estamos restringindo S a ser constante em E e em E^c, isto é, uma previsão baseada apenas na informação binária de ocorrer E ou não. O minimizador do erro quadrático médio escolhe, em cada átomo da partição $\{E, E^c\}$, o melhor valor constante, que é a média de Y condicionada ao átomo:

$$Z(\omega) = \mathbb{E}[Y \mid E] \ \text{em } E, \qquad Z(\omega) = \mathbb{E}[Y \mid E^c] \ \text{em } E^c.$$

6. Seja $(\Omega, \Sigma, \mathbb{P})$, e X e Y duas variáveis aleatórias reais. Considere $\mathbb{P}_{X,Y} := \mathbb{P}[\{\omega : (X(\omega), Y(\omega)) \in B]$, $B \in \mathcal{B}(\mathbb{R}^2)$, a medida de probabilidade induzida por (X, Y) em $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$. Dizemos que $\mathbb{P}_{X,Y}$ admite densidade f com respeito a medida de Lebesgue λ^2 em $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ [1]se, para todo $B \in \mathcal{B}(\mathbb{R}^2)$:

$$\mathbb{P}_{X,Y} = \int f(\omega_1, \omega_2)\mathbf{1}_B(\omega_1, \omega_2)\lambda(d\omega),$$

onde, por uma extensão do resultado visto em aula, $\lambda(d\omega)$ pode ser substituída pela integral dupla (de Riemann) quando $(\omega_1, \omega_2) \mapsto f(\omega_1, \omega_2)\mathbf{1}_B(\omega_1, \omega_2)$ for Riemann-integrável.

[1]A medida de Lebesgue em $\mathbb{R}^2$ é igual à medida produto $\lambda \otimes \lambda$ sobre $\mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}) = \mathcal{B}(\mathbb{R}^2)$, onde λ é a medida de Lebesgue em $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$

(a) Mostre que $f_1(\omega) := \int f(\omega, y)\lambda(dy)$ define uma densidade para a probabilidade $\mathbb{P}_X$ induzida sobre $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. *Dica:* tome $A \in \mathcal{B}(\mathbb{R})$ e use o teorema de Fubini para mostrar que $\mathbb{P}_X = \mathbb{P}_{X,Y}[A \times \mathbb{R}] = \int \mathbf{1}_{A\times\mathbb{R}} f d\lambda^2 = \int \mathbf{1}_A f_1 d\lambda_1$.

Resolução:

Como $\mathbb{P}_{X,Y}$ admite densidade f com respeito a λ^2, temos $f \geqslant 0$ mensurável e

$$\int_{\mathbb{R}^2} f(x,y)\, \lambda^2(\mathrm{d}(x,y)) = P_{X,Y}\left(\mathbb{R}^2\right) = 1,$$

logo $f \in L^1(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2), \lambda^2)$. Pelo Teorema de Fubini, a função

$$f_1(x) := \int_{\mathbb{R}} f(x,y)\, \lambda(\mathrm{d}y)$$

é bem definida para λ-q.t.p. x, é mensurável e não negativa.
Para $A \in \mathcal{B}(\mathbb{R})$,

$$\begin{aligned}
\mathbb{P}_X(A) = \mathbb{P}(X \in A) &= \mathbb{P}\left((X,Y) \in A \times \mathbb{R}\right) = \mathbb{P}_{X,Y}\left(A \times \mathbb{R}\right) \\
&= \int_{\mathbb{R}^2} \mathbf{1}_{A\times\mathbb{R}}(x,y)\, f(x,y)\, \lambda^2(\mathrm{d}(x,y)) \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbf{1}_A(x)\, f(x,y)\, \lambda(\mathrm{d}y) \right) \lambda(\mathrm{d}x) \\
&= \int_{\mathbb{R}} \mathbf{1}_A(x) \left(\int_{\mathbb{R}} f(x,y)\, \lambda(\mathrm{d}y) \right) \lambda(\mathrm{d}x) = \int_A f_1(x)\, \lambda(\mathrm{d}x).
\end{aligned}$$

Em particular, tomando $A = \mathbb{R}$ obtemos $\int_{\mathbb{R}} f_1\, d\lambda = 1$. Portanto, f_1 é uma densidade de P_X com respeito a λ.

(b) Suponha que $\mathbb{E}[|Y|] < \infty$. Mostre que a função $g : \mathbb{R} \mapsto \mathbb{R}$:

$$g(x) = \begin{cases} \frac{1}{f_1(x)} \int y f(y,x)\lambda(dy), & \text{se } f_1(x) > 0 \\ 0, & \text{se } f_1(x) = 0 \end{cases},$$

define uma função de expectativa condicional de Y em X, i.e. que $g(X)$ é uma versão de $\mathbb{E}[Y|X]$. A escolha, na definição de g, do valor 0 quando $f_1(x) = 0$, faz alguma diferença? Por quê?

Resolução:

Como $\mathbb{E}\left[|Y|\right] < \infty$ e λ^2 admite densidade, generalizando o **Exercício 2**,

$$\mathbb{E}\left[|Y|\right] = \int_{\mathbb{R}^2} |y|\, f(x,y)\, \lambda^2(\mathrm{d}(x,y)) < \infty,$$

logo a função $(x,y) \mapsto y f(x,y)$ é integrável em valor absoluto. Pelo Teorema de Fubini,

$$m(x) := \int_{\mathbb{R}} y f(x,y)\, \lambda(\mathrm{d}y)$$

é (finita para λ-q.t.p. x e) mensurável.
Defina

$$g(x) := \begin{cases} \dfrac{m(x)}{f_1(x)}, & \text{se } f_1(x) > 0, \\ 0, & \text{se } f_1(x) = 0. \end{cases}$$

Então g é mensurável e $g(X)$ é $\sigma(X)$-mensurável.
Além disso, $g(X)$ é integrável (pelo mesmo argumento do **Exercício 2**):

$$\begin{aligned}\mathbb{E}\left[|g(X)|\right] = \int_{\mathbb{R}} |g(x)|\, f_1(x)\,\lambda(\mathrm{d}x) &= \int_{\{f_1>0\}} \left|\frac{m(x)}{f_1(x)}\right| f_1(x)\,\lambda(\mathrm{d}x)\\ &= \int_{\{f_1>0\}} |m(x)|\,\lambda(\mathrm{d}x) \leqslant \int_{\mathbb{R}} \left(\int_{\mathbb{R}} |y| f(x,y)\,\lambda(\mathrm{d}y)\right)\lambda(\mathrm{d}x)\\ &= \int_{\mathbb{R}^2} |y| f(x,y)\,\lambda^2(\mathrm{d}(x,y)) = \mathbb{E}\left[|Y|\right] < \infty.\end{aligned}$$

Para verificar a caracterização de esperança condicional com respeito a $\sigma(X)$, basta checar nos eventos $\{X \in A\}$, $A \in \mathcal{B}(\mathbb{R})$. Note que, se $f_1(x) = 0$, então

$$\int_{\mathbb{R}} f(x,y)\,\lambda(\mathrm{d}y) = 0.$$

Como $y \mapsto f(x,y)$ é não negativa, pelo lema básico para integrais de funções não negativas,

$$\lambda\left(\{y : f(x,y) > 0\}\right) = 0,$$

isto é, $f(x,y) = 0$ para λ-q.t.p. y, e portanto $m(x) = \int y f(x,y)\,\mathrm{d}y = 0$. Logo, para todo x,

$$g(x) f_1(x) = m(x).$$

Assim, para $A \in \mathcal{B}(\mathbb{R})$,

$$\begin{aligned}\mathbb{E}\left[g(X)\mathbf{1}_{\{X\in A\}}\right] = \int_{\mathbb{R}} g(x)\mathbf{1}_A(x)\, f_1(x)\,\lambda(\mathrm{d}x) &= \int_A g(x) f_1(x)\,\lambda(\mathrm{d}x) = \int_A m(x)\,\lambda(\mathrm{d}x)\\ &= \int_A \left(\int_{\mathbb{R}} y f(x,y)\,\lambda(\mathrm{d}y)\right)\lambda(\mathrm{d}x) = \int_{\mathbb{R}^2} y\,\mathbf{1}_A(x)\, f(x,y)\,\lambda^2(\mathrm{d}(x,y)).\end{aligned}$$

Usando de novo o **Exercício 2** generalizado em duas dimensões, o último termo é

$$\int_{\mathbb{R}^2} y\,\mathbf{1}_A(x)\, f(x,y)\,\lambda^2(\mathrm{d}(x,y)) = \mathbb{E}\left[Y\,\mathbf{1}_{\{X\in A\}}\right].$$

Portanto, para todo $A \in \mathcal{B}(\mathbb{R})$,

$$\mathbb{E}\left[g(X)\mathbf{1}_{\{X\in A\}}\right] = \mathbb{E}\left[Y\mathbf{1}_{\{X\in A\}}\right],$$

e como $g(X)$ é $\sigma(X)$-mensurável e integrável, concluímos que $g(X)$ é uma versão de $\mathbb{E}\left[Y \mid \sigma(X)\right]$, isto é, de $\mathbb{E}[Y \mid X]$.
Quanto à escolha do valor 0 quando $f_1(x) = 0$: ela não faz diferença para a esperança condicional, pois

$$\mathbb{P}\left(f_1(X) = 0\right) = \mathbb{P}_X\left(\{x : f_1(x) = 0\}\right) = \int_{\{f_1=0\}} f_1(x)\,\lambda(\mathrm{d}x) = 0.$$

Assim, qualquer outra definição de g em $\{f_1 = 0\}$ produziria uma função que coincide com $g(X)$ quase certamente, e versões de esperança condicional são únicas a menos de eventos de probabilidade zero.

(7) Seja $(\Omega, 2^\Omega, \mathbb{P})$ um espaço de probabilidade em que Ω é finito. Defina a *medida de contagem* c sobre $(\Omega, 2^\Omega)$ como, para todo $B \subseteq \Omega$:

$$c(B) = |B| = \sum_{\omega \in \Omega} \mathbf{1}_B(\omega) .$$

Mostre que $\mathbb{P}$ admite densidade com respeito a c, e que essa densidade é dada por $g(\omega) = \mathbb{P}[\{\omega\}]$, $\omega \in \Omega$.

Resolução:

Como Ω é finito, $\Sigma = 2^\Omega$ e toda função $g : \Omega \to \mathbb{R}$ é Σ-mensurável. Além disso, para todo $B \subseteq \Omega$ podemos escrever a decomposição disjunta

$$B = \bigcup_{\omega \in B} \{\omega\}.$$

Pela aditividade finita de $\mathbb{P}$ em uniões disjuntas (aqui a soma é finita), segue que

$$\mathbb{P}(B) = \sum_{\omega \in B} \mathbb{P}\left(\{\omega\}\right) .$$

Defina $g : \Omega \to [0, 1]$ por

$$g(\omega) := \mathbb{P}\left(\{\omega\}\right), \qquad \omega \in \Omega.$$

A integral com respeito à medida de contagem c coincide com a soma, para qualquer $h : \Omega \to \mathbb{R}_+$,

$$\int_\Omega h(\omega)\, c(\mathrm{d}\omega) = \sum_{\omega \in \Omega} h(\omega).$$

Logo, para todo $B \subseteq \Omega$,

$$\begin{aligned} \int_\Omega \mathbf{1}_B(\omega)\, g(\omega)\, c(\mathrm{d}\omega) &= \sum_{\omega \in \Omega} \mathbf{1}_B(\omega)\, g(\omega) = \sum_{\omega \in B} g(\omega) \\ &= \sum_{\omega \in B} \mathbb{P}\left(\{\omega\}\right) = \mathbb{P}(B). \end{aligned}$$

Portanto, $\mathbb{P}$ admite densidade com respeito a c, e uma densidade é exatamente $g(\omega) = \mathbb{P}(\{\omega\})$. Como verificação adicional, note que

$$\int_\Omega g(\omega)\, c(\mathrm{d}\omega) = \sum_{\omega \in \Omega} \mathbb{P}\left(\{\omega\}\right) = \mathbb{P}(\Omega) = 1,$$

como deve ocorrer para uma densidade de uma medida de probabilidade.

8. Se $(\Omega, \Sigma, \mathbb{P})$ um espaço de probabilidade, e X uma variável aleatória real cuja imagem está contida num intervalo $[a, b]$ de $\mathbb{R}$. Suponha que a função de distribuição de X, F_X é contínua em $[a, b]$, e diferenciável em (a, b). Mostre que a função $g : \mathbb{R} \mapsto \mathbb{R}$, dada por:

$$g(x) = \begin{cases} F'(x), & \text{se } x \in (a, b) \\ 0, & \text{se } x \notin (a, b) \end{cases}$$

define uma densidade de $\mathbb{P}_X$ com respeito à medida de Lebesgue em $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. *Dica:* use o teorema fundamental do cálculo e o lema do π-sistema.

Resolução:

Defina μ em $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ por

$$\mu(B) := \int_{\mathbb{R}} \mathbf{1}_B(x)\, g(x)\, \lambda(\mathrm{d}x), \qquad B \in \mathcal{B}(\mathbb{R}),$$

onde

$$g(x) = \begin{cases} F_X'(x), & x \in (a, b), \\ 0, & x \notin (a, b). \end{cases}$$

Primeiro, g é $\mathcal{B}(\mathbb{R})$-mensurável: como F_X é contínua, para cada $n \geqslant 1$ a função

$$x \longmapsto n\left(F_X\left(x + \frac{1}{n}\right) - F_X(x)\right)$$

é contínua (logo Borel-mensurável), e em (a, b) vale

$$F_X'(x) = \lim_{n\to\infty} n\left(F_X\left(x + \frac{1}{n}\right) - F_X(x)\right),$$

portanto F_X' é mensurável em (a, b) como limite pontual de funções mensuráveis; estendendo por zero fora de (a, b), segue que g é mensurável em $\mathbb{R}$. Além disso, $g \geqslant 0$ em (a, b), pois F_X é não-decrescente e, para $h > 0$,

$$\frac{F_X(x + h) - F_X(x)}{h} \geqslant 0 \quad \Rightarrow \quad F_X'(x) \geqslant 0.$$

Como $X(\Omega) \subseteq [a, b]$, temos

$$F_X(t) = \mathbb{P}(X \leqslant t) = 0 \text{ se } t < a, \qquad F_X(t) = 1 \text{ se } t \geqslant b.$$

Pela continuidade de F_X em $[a, b]$, em particular em a e b, obtemos

$$F_X(a) = 0, \qquad F_X(b) = 1.$$

Pelo teorema fundamental do cálculo,

$$\int_a^b F_X'(x)\, \lambda(dx) = F_X(b) - F_X(a) = 1,$$

e como $g = F_X'$ em (a, b) e $g = 0$ fora, segue que

$$\int_{\mathbb{R}} g(x)\, \lambda(\mathrm{d}x) = 1,$$

logo μ é uma medida de probabilidade.
Agora considere o π-sistema

$$\mathcal{P} := \{(-\infty, t] : t \in \mathbb{R}\}.$$

Temos $\mathcal{P}$ é um π-sistema pois

$$(-\infty, t_1] \cap (-\infty, t_2] = (-\infty, \min\{t_1, t_2\}],$$

e $\sigma(\mathcal{P}) = \mathcal{B}(\mathbb{R})$.
Mostremos que μ e $\mathbb{P}_X$ coincidem em $\mathcal{P}$. Fixe $t \in \mathbb{R}$. Então

$$\mu\left((-\infty, t]\right) = \int_{(-\infty,t]} g(x)\,\lambda(\mathrm{d}x).$$

Analisando casos:
- Se $t \leqslant a$, então $g = 0$ em $(-\infty, t]$, logo $\mu((-\infty, t]) = 0 = F_X(t) = \mathbb{P}_X((-\infty, t])$.
- Se $a < t < b$, então

$$\mu\left((-\infty, t]\right) = \int_a^t F_X'(x)\,\lambda(\mathrm{d}x) = F_X(t) - F_X(a) = F_X(t),$$

isto é, $\mu((-\infty, t]) = \mathbb{P}_X((-\infty, t])$.
- Se $t \geqslant b$, então

$$\mu\left((-\infty, t]\right) = \int_a^b F_X'(x)\,\lambda(\mathrm{d}x) = 1 = F_X(t),$$

isto é, $\mu((-\infty, t]) = \mathbb{P}_X((-\infty, t])$.
Logo, μ e $\mathbb{P}_X$ coincidem em $\mathcal{P}$, então pelo **lema do π-sistema**, conclui-se que $\mu = \mathbb{P}_X$ em $\mathcal{B}(\mathbb{R})$, isto é, para todo $B \in \mathcal{B}(\mathbb{R})$,

$$\mathbb{P}_X(B) = \mu(B) = \int_{\mathbb{R}} \mathbf{1}_B(x)\,g(x)\,\lambda(\mathrm{d}x).$$

Portanto, $\mathbb{P}_X$ admite densidade com respeito à medida de Lebesgue, e essa densidade é g.