Analogamente, se conclui que a função $x_2 \mapsto \mu_1(A_{x_2})$ é simples. Portanto,

$$\int_{X_1} \mu_2(A_{x_1})\, d\mu_1 = \mu_1(A_1) \times \mu_2(A_2) = \int_{X_2} \mu_1(A_{x_2})\, d\mu_2\,.$$

Logo, este cálculo para funções simples motiva a seguinte definição de μ

$$\mu(A) := \int_{X_1} \mu_2(A_{x_1})\, d\mu_1 = \int_{X_2} \mu_1(A_{x_2})\, d\mu_2\,.$$

Chamamos μ a **medida produto** e designamos por $\mu = \mu_1 \times \mu_2$. O seguinte teorema mostra que esta definição faz sentido para qualquer conjunto mensurável $F \in \mathcal{F}_1 \otimes \mathcal{F}_2$.

Teorema 2.43 (Medida produto)**.** *Sejam $(X_1, \mathcal{F}_1, \mu_1)$ e $(X_2, \mathcal{F}_2, \mu_2)$ dois espaços de medida e μ_1, μ_2 duas medidas σ-finitas. Considere-se a função $\mu : \mathcal{F}_1 \otimes \mathcal{F}_2 \to [0, \infty]$ definida por*

$$\mu(F) = \int_{X_1} \mu_2(F_{x_1})\, d\mu_1 = \int_{X_2} \mu_1(F_{x_2})\, d\mu_2\,.$$

Então μ é uma medida em $\mathcal{F}_1 \otimes \mathcal{F}_2$ e é a única medida tal que para qualquer rectângulo mensurável $A_1 \times A_2$ se tem

$$\mu(A_1 \times A_2) = \mu_1(A_1) \times \mu_2(A_2)\,.$$

Demonstração. Ver demonstração em [1]. □

Chegamos portanto ao seguinte teorema fundamental.

Teorema 2.44 (de Fubini)**.** *Nas condições do teorema anterior, se $f \in \mathcal{L}^1(\mu_1 \times \mu_2)$ então*

$$\int_X f\, d(\mu_1 \times \mu_2) = \int_{X_1} \int_{X_2} f\, d\mu_2\, d\mu_1 = \int_{X_2} \int_{X_1} f\, d\mu_1\, d\mu_2\,.$$

Demonstração. Ver demonstração em [1]. □

Parte II
Probabilidade

Na segunda parte deste curso, através da teoria da medida e integração desenvolvida anteriormente, iremos introduzir os principais conceitos da teoria da probabilidade e processos estocásticos.

3 Conceitos básicos

Considere-se uma experiência aleatória e seja Ω o conjunto formado por todos os resultados possíveis dessa experiência. Logo Ω é designado por **espaço dos resultados**. Um subconjunto $A \subset \Omega$ é designado por **acontecimento**. Uma colecção de acontecimentos $\mathcal{F}$ que forma uma σ-álgebra diz-se a **σ-álgebra de acontecimentos**. Portanto, $(\Omega, \mathcal{F})$ é um espaço mensurável.

Definição 3.1. Um **espaço de probabilidade** é um espaço de medida $(\Omega, \mathcal{F}, P)$ tal que $P(\Omega) = 1$.

A medida P designa-se por **medida de probabilidade**. Dado um acontecimento $A \in \mathcal{F}$ o número $P(A)$ é a **probabilidade do acontecimento** A.

Exemplo 3.1. Considere a experiência aleatória de lançar uma moeda "perfeita" ao ar e observar qual das suas faces fica voltada para cima, ou seja, se é cara $= \diamondsuit$ ou coroa $= \clubsuit$.

Formalmente, tem-se o espaço de resultados $\Omega = \{\diamondsuit, \clubsuit\}$, a σ-álgebra de acontecimentos $\mathcal{F} = \{\emptyset, \{\diamondsuit\}, \{\clubsuit\}, \{\diamondsuit, \clubsuit\}\}$ e a medida de probabilidade $P = \delta_{\diamondsuit}/2 + \delta_{\clubsuit}/2$ onde $\delta_{\diamondsuit}$ e $\delta_{\clubsuit}$ são medidas de Dirac. A probabilidade do acontecimento "sair cara" é $P(\{\diamondsuit\}) = 1/2$.

Exemplo 3.2. Considere a experiência aleatória de escolher um número "ao acaso" do intervalo $[0, 1]$.

Mais concretamente, considere-se o espaço de resultados $\Omega = [0, 1]$, a σ-álgebra de acontecimentos $\mathcal{F} = \mathcal{B}([0, 1])$ e a medida de probabilidade $P = m|_{[0,1]}$ (medida de Borel restrita ao intervalo $[0, 1]$). A probabilidade de o número que escolhemos "ao acaso" ser racional é $P(\mathbb{Q} \cap [0, 1]) = 0$.

Exemplo 3.3. Considere-se o espaço mensurável $(\Omega, \mathcal{P}(\Omega))$ com $\Omega = \{1, 2, \dots, n\}$. Seja $P : \mathcal{P}(\Omega) \to [0, \infty]$ a seguinte função,

$$P(A) = \sum_{k \in A} \binom{n}{k} p^k (1-p)^{n-k}, \quad A \subset \Omega,$$

onde $0 < p < 1$. É fácil demonstrar que P é uma medida de probabilidade. De facto, é claro que $P(\emptyset) = 0$. Por outro lado, se A_i é uma sucessão de conjuntos disjuntos então

$$\begin{aligned}
P\left(\bigcup_{i=1}^{\infty} A_i\right) &= \sum_{k\in\bigcup_{i=1}^{\infty} A_i} \binom{n}{k} p^k(1-p)^{n-k} \\
&= \sum_{i=1}^{\infty} \sum_{k\in A_i} \binom{n}{k} p^k(1-p)^{n-k} \\
&= \sum_{i=1}^{\infty} P(A_i)\,.
\end{aligned}$$

Finalmente,

$$P(\Omega) = \sum_{k=1}^{n} \binom{n}{k} p^k(1-p)^{n-k} = (p+1-p)^n = 1$$

Dada uma função mensurável $g : \Omega \to [0,\infty]$, é claro que g é simples. Logo,

$$\int_A g\, dP = \sum_{k\in A} g(k)P(\{k\}) = \sum_{k\in A} g(k)\binom{n}{k} p^k(1-p)^{n-k}\,.$$

3.1 Variáveis aleatórias

Por vezes, é conveniente associar ao resultado de uma experiência aleatória um número real. Considere-se um espaço de probabilidade $(\Omega, \mathcal{F}, P)$ e o espaço mensurável $(\mathbb{R}, \mathcal{B})$ onde $\mathcal{B}$ é a σ-álgebra de Borel de $\mathbb{R}$.

Definição 3.2. Uma função mensurável $X : \Omega \to \mathbb{R}$ diz-se uma **variável aleatória**.

Associada a uma variável aleatória X podemos definir uma medida na σ-álgebra de Borel que não é mais do que a medida imagem de P por X,

$$p_X(B) = P(X^{-1}(B)), \quad B \in \mathcal{B}\,.$$

Note que p_X é uma medida de probabilidade no espaço mensurável $(\mathbb{R}, \mathcal{B})$. Designamos esta medida por **distribuição de probabilidade da variável aleatória** X.

Observação 3.4. É também usual escrever-se

$$p_X(B) = P(X \in B)\,.$$

De acordo com a Proposição 1.7 podemos definir a **função de distribuição da variável aleatória** X, como a função $F_X : \mathbb{R} \to \mathbb{R}$ tal que

$$F_X(x) = p_X(]-\infty, x]) .$$

Veremos que F_X carateriza completamente a distribuição de probabilidade da variável aleatória X. De certa forma, a função de distribuição F_X constitui o modelo probabilístico que descreve a experiência aleatória, sendo também designada como a **lei de probabilidade** da variável aleatória X.

Observação 3.5. É usual escrever-se $F_X(x)$ de diversas maneiras, todas equivalentes entre si, isto é

$$F_X(x) = P(X^{-1}(]-\infty, x])) = P(\omega \in \Omega : X(\omega) \leq x) = P(X \leq x) .$$

Exercício 38. Considere o espaço de probabilidade $([0,1], \mathcal{B}, m)$ e a variável aleatória $X(\omega) = \min(\omega, 1-\omega)$. Calcule F_X.

Exercício 39. Suponha que um comboio parte aleatoriamente do Porto entre as 8h e as 10h da manhã com destino a Lisboa, que fica a 300km de distância. Suponha também que o comboio viaja a uma velocidade constante de 100km/h.

1. Determine a variável aleatória que descreve a distância entre o comboio e Lisboa às 12h.
2. Calcule a distribuição de probabilidade dessa variável aleatória e a respectiva função de distribuição.

No resultado que se segue resumimos as propriedades da função de distribuição F_X.

Proposição 3.6. *A função F_X tem as seguintes propriedades:*

1. *é uma função de distribuição no sentido da Definição 1.3.*
2. $p_X(]a,b]) = F_X(b) - F_X(a)$
3. $P(X = x) = F_X(x) - F_X(x^-)$
4. $\lim_{x\to+\infty} F_X(x) = 1$ *e* $\lim_{x\to-\infty} F_X(x) = 0$.

Demonstração. Deixa-se como exercício. □

Observação 3.7. Note-se que da segunda propriedade, podemos afirmar que p_X é uma medida de Borel-Stieltjes associada à função de distribuição F_X.

Exercício 40. Seja X uma variável aleatória e F_X a sua função de distribuição. Mostre que:

1. $P(a < X \leq b) = F_X(b) - F_X(a)$.
2. $P(a \leq X \leq b) = F_X(b) - F_X(a^-)$.
3. $P(a < X < b) = F_X(b^-) - F_X(a)$.
4. $P(a \leq X < b) = F_X(b^-) - F_X(a^-)$.

3.1.1 Classificação

Considere-se o espaço de probabilidade $(\mathbb{R}, \mathcal{B}, p_X)$ e denote-se por m a medida de Borel.

Definição 3.3. Uma variável aleatória X diz-se **discreta** sse existir um conjunto numerável $D \subset \mathbb{R}$ tal que p_X está concentrada em D.

Note-se que $p_X \perp m$ uma vez que D tem medida de Borel nula. Suponha que o conjunto de imagens de $X : \Omega \to \mathbb{R}$ é finito ou numerável. Isto é, $D = X(\Omega) = \{x_1, x_2, \ldots\}$. Então a variável aleatória X é discreta, uma vez que p_X se encontra concentrada em D. De facto, dado qualquer Boreliano B temos que

$$p_X(B) = p_X(B \cap D) + p_X(B \setminus D) = p_X(B \cap D)\,,$$

porque $p_X(B \setminus D) = P(X^{-1}(B \setminus D)) = P(\emptyset) = 0$. Por outro lado, podemos escrever

$$p_X = \sum_{n=1}^{\infty} P(X = x_n)\delta_{x_n}\,.$$

Note-se que D é o conjunto dos pontos de descontinuidade da função de distribuição. De facto F_X é uma função escada com pontos de descontinuidade D.

Definição 3.4. Uma variável aleatória X diz-se **contínua** sse $p_X(\{x\}) = 0$ para todo $x \in \mathbb{R}$.

Note-se que uma variável aleatória é contínua sse a sua função de distribuição F_X for contínua.

Um caso particular, é quando a distribuição de probabilidade p_X é absolutamente contínua relativamente à medida de Borel m.

Definição 3.5. Uma variável aleatória X diz-se **absolutamente contínua** sse $p_X \ll m$.

Uma variável absolutamente contínua é em particular contínua. De facto, pelo teorema de Radon-Nikodym existe uma função integrável $f \in \mathcal{L}^1(\mu)$ tal que

$$p_X(B) = \int_B f\,dm, \quad \forall\, B \in \mathcal{B}\,.$$

Como $m(\{x\}) = 0$ para todo $x \in \mathbb{R}$ segue que $p_X(\{x\}) = 0$.

Definição 3.6. A função integrável $f = \frac{dp_X}{dm}$ diz-se a **função densidade de probabilidade da variável aleatória** X.

Observação 3.8. Se p_X é absolutamente contínua relativamente à medida de Lebesgue m é possível demonstrar que F_X é diferenciável q.c. relativamente a m e tem-se $F_X' = f$ q.c. Por outro lado, se f é integrável segundo Riemann então

$$F_X(x) = \int_{-\infty}^{x} f(u)\, du\,.$$

Observação 3.9. A função de densidade de probabilidade é uma função não negativa que satisfaz

$$\int_{-\infty}^{\infty} f(u)\, du = 1\,.$$

Exemplo 3.10 (Distribuição Uniforme). Seja $B \subset \mathbb{R}$ um boreliano e considere a seguinte densidade de probabilidade

$$f(x) = \begin{cases} \frac{1}{m(B)} & \text{se } x \in B \\ 0 & \text{caso contrário} \end{cases}$$

Exemplo 3.11 (Distribuição Gaussiana ou Normal). Seja $\mu \in \mathbb{R}$ e $\sigma > 0$. A função de densidade de probabilidade de uma variável aleatória com distribuição normal é

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,.$$

Esta função é simétrica em torno de $x = \mu$ e tende para zero quando x tende para infinito.

Exemplo 3.12 (Distribuição Exponencial). O tempo de vida de uma componente electrónica pode ser modelado por uma variável aleatória absolutamente contínua $X : \Omega \to \mathbb{R}$ com a seguinte função de distribuição

$$F_X(x) = \begin{cases} 1 - e^{-\theta x} & \text{se } x \geq 0 \\ 0 & \text{se } x < 0 \end{cases}$$

com função densidade de probabilidade $f(x) = \theta e^{-\theta x} \chi_{[0,\infty[}(x)$.

Exemplo 3.13 (Distribuição Gama e **χ-quadrado).** Uma variável aleatória que segue uma distribuição gama tem como função de densidade de probabilidade,

$$f(x) = \begin{cases} \frac{1}{\Gamma(t)} \lambda^t x^{t-1} e^{-\lambda x} & \text{se } x \geq 0 \\ 0 & \text{caso contrário} \end{cases}$$

onde $t > 0$, $\lambda > 0$ e $\Gamma(t) = \int_0^\infty x^{t-1}e^{-x}\,dx$ é a função gama. A distribuição gama, indexada por dois parâmetros, contém uma grande classe de outras distribuições, tais como a distribuição exponencial ($t = 1$) e a distribuição chi-quadrado χ^2 com d graus de liberdade ($\lambda = 1/2$ e $t = d/2$).

Definição 3.7. Uma variável aleatória diz-se **singular** sse for contínua e $p_X \perp m$.

Exemplo 3.14. Um exemplo de uma variável aleatória singular é dada pela função de distribuição que passamos a descrever. Considere-se a representação ternária de um número $x \in [0, 1]$, isto é,

$$x = 0.a_1a_2a_3\cdots$$

onde $a_n \in \{0, 1, 2\}$. Seja $N = \min\{n \in \mathbb{N} : a_n = 1\}$. Note-se que para $x = 1/2$ temos que $N = 1$ enquanto que para $x = 1/6$ temos que $N = 2$. Definimos uma nova sucessão,

$$b_n = \begin{cases} \frac{a_n}{2} & \text{se } n < N \\ 1 & \text{se } n = N \\ 0 & \text{se } n > N \end{cases}.$$

Por fim, definimos a função de distribuição F_X da seguinte maneira,

$$F_X(x) = \begin{cases} 0 & \text{se } x < 0 \\ \displaystyle\sum_{n=1}^{\infty} \frac{b_n}{2^n} & \text{se } 0 \leq x \leq 1 \\ 1 & \text{se } x > 1 \end{cases}$$

A função F_X é contínua e constante em $\mathbb{R} \setminus C$ onde C é o conjunto ternário de Cantor. Portanto, p_X está concentrada em C. No entanto é possível mostrar que C é não numerável e tem medida de Lebesgue nula. Por outro lado, $F'_X = 0$ q.c. e

$$1 = F_X(1) - F_X(0) \neq \int_0^1 F'_X\,dm = 0\,.$$

3.1.2 Decomposição

Considere um espaço de probabilidade $(\Omega, \mathcal{F}, P)$ e $X : \Omega \to \mathbb{R}$ uma variável aleatória. O teorema da decomposição de Lebesgue estabelece que

$$p_X = \lambda_a + \lambda_s \quad \text{onde} \quad \lambda_a \ll m \quad \text{e} \quad \lambda_s \perp m\,,$$

onde m é a medida de Borel. No entanto é possível refinar esta decomposição, separando a parte singular λ_s numa soma de duas medidas singulares relativamente a m,

$$\lambda_s = \lambda_d + \lambda_{sc}, \quad \lambda_d, \lambda_{sc} \perp m\,,$$

tal que λ_d é uma **medida discreta**, isto é, existe um conjunto numerável $D \subset \mathbb{R}$ tal que λ_d se encontra concentrada em D e λ_{sc} é uma **medida singular contínua**, isto é, $\lambda_{sc}(\{x\}) = 0$ para todo $x \in \mathbb{R}$.

Adicionalmente, é possível normalizar as medidas anteriores para obter uma decomposição de p_X em termos de medidas de probabilidades. Portanto é possível escrever,

$$p_X = \alpha_a p_a + \alpha_d p_d + \alpha_{sc} p_{sc}, \quad \text{com} \quad \alpha_a + \alpha_d + \alpha_{sc} = 1\,,$$

onde $\alpha_a, \alpha_d, \alpha_{sc} \geq 0$ e p_a é uma medida de probabilidade absolutamente contínua (relativamente e m), p_d é uma medida de probabilidade discreta e p_{sc} é uma medida de probabilidade singular contínua.

Observação 3.15. Segue da decomposição anterior que

$$F_X(x) = \alpha_a F_a(x) + \alpha_d F_d(x) + \alpha_{sc} F_{sc}(x)\,,$$

onde F_a é diferenciável q.c. e $F_a' = f_a$ para alguma função f_a integrável, F_d é uma função tipo escada com um conjunto numerável de "saltos" e F_{sc} é uma função contínua.

Observação 3.16. Note-se que se $\alpha_a = 1$ (implicando que $\alpha_d = \alpha_{sc} = 0$) então a variável aleatória X é absolutamente contínua. No entanto, se $\alpha_d = 1$ (implicando que $\alpha_a = \alpha_{sc} = 0$) então X é discreta. Finalmente, se $\alpha_{sc} = 1$ então X é singular.

Definição 3.8. No caso em que $\alpha_a \neq 0$ e $\alpha_d + \alpha_{sc} \neq 0$ então a variável aleatória X diz-se **mista**.

Exemplo 3.17. Um exemplo de uma variável aleatória mista é dada pela seguinte lei de probabilidade,

$$F_X(x) = \begin{cases} 0 & \text{se } x < 0 \\ 1/2 + x/2 & \text{se } 0 \leq x < 1 \\ 1 & \text{se } x \geq 1 \end{cases}$$

que corresponde à distribuição de probabilidade

$$p_X = \frac{\delta_0}{2} + \frac{m|_{[0,1]}}{2}\,.$$

As variáveis aleatórias mais usuais em probabilidade são as absolutamente contínuas, as discretas, ou então uma combinação destas, que designamos por mistas. As variáveis aleatórias singulares, como a do exemplo anterior, são menos usuais e servem na maior parte dos casos para construir contra-exemplos.

Acabamos esta secção com uma fórmula bastante útil no cálculo da densidade de uma função de uma variável aleatória.

Proposição 3.18. *Seja X uma variável aleatória com função de densidade de probabilidade f_X e $g : \mathbb{R} \to \mathbb{R}$ uma função estritamente crescente e diferenciável. Então,*

$$f_{g(X)}(y) = f_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y) .$$

Demonstração. Tomando as distribuições de probabilidade obtemos,

$$F_{g(X)}(y) = P(g(X) \leq y)) = P(X \leq g^{-1}(y)) = F_X(g^{-1}(y)) .$$

Derivando em ambos os lados da equação obtém-se o resultado. $\square$

3.2 Valor esperado

Definição 3.9. Seja $(\Omega, \mathcal{F}, P)$ um espaço de probabilidade. O **valor esperado**, **valor médio** ou **esperança matemática** da variável aleatória $X : \Omega \to \mathbb{R}$ é definido por

$$E(X) = \int_\Omega X \, dP .$$

Observação 3.19. Esta definição só faz sentido se a variável aleatória pertencer a $\mathcal{L}^1(P)$, uma vez que,

$$|E(X)| \leq \int_\Omega |X| \, dP < \infty .$$

Ou seja, $E(X)$ é um valor finito.

Observação 3.20. Dada uma função mensurável $g : \mathbb{R} \to \mathbb{R}$, segue do Teorema 2.42 que

$$\int_\mathbb{R} g \, dp_X = \int_\Omega g \circ X \, dP .$$

Logo, tomando $g(x) = x$ obtemos que

$$E(X) = \int_\Omega X \, dP = \int_\mathbb{R} x \, dp_X .$$

Uma vez que p_X é uma medida de Borel-Stieltjes com função de distribuição F_X então pode-se calcular o valor esperado de X usando qualquer uma das expressões,

$$E(X) = \int_\Omega X\, dP = \int_{\mathbb{R}} x\, dp_X = \int_{\mathbb{R}} x\, dF_X\,.$$

Observação 3.21. Se a variável aleatória for absolutamente contínua, isto é, $p_X \ll m$ então existe uma função de densidade de probabilidade f tal que

$$p_X(B) = \int_B f\, dm, \quad \forall\, B \in \mathcal{B}\,.$$

Segue então do Teorema 2.23 que

$$E(X) = \int_{\mathbb{R}} x\, dp_X = \int_{\mathbb{R}} x\, dF_X = \int_{\mathbb{R}} x f(x)\, dm\,.$$

Se adicionalmente f for integrável à Riemann então

$$E(X) = \int_{-\infty}^{+\infty} x f(x)\, dx\,.$$

Observação 3.22. Suponha que $X : \Omega \to \mathbb{R}$ é uma variável aleatória discreta, ou seja, que existe um conjunto numerável $D = \{x_1, x_2, \ldots\} \subset \mathbb{R}$ tal que p_X está concentrada em D. Então, por uma observação anterior temos que a distribuição de probabilidade de X satisfaz

$$p_X = \sum_{n=1}^{\infty} p_n \delta_{x_n}\,,$$

onde $p_n = P(X = x_n)$. Uma vez que $p_X(\mathbb{R} \setminus D) = 0$, então o valor esperado de X pode ser calculado da seguinte maneira,

$$E(X) = \int_{\mathbb{R}} x\, dp_X = \int_D x\, dp_X + \int_{\mathbb{R}\setminus D} x\, dp_X = \int_D x\, dp_X\,.$$

Portanto,

$$E(X) = \int_{\bigcup_{n=1}^{\infty}\{x_n\}} x\, dp_X = \sum_{n=1}^{\infty} \int_{\{x_n\}} x\, dp_X = \sum_{n=1}^{\infty} \int_{\mathbb{R}} \chi_{\{x_n\}} \cdot x\, dp_X\,.$$

Como $\chi_{\{x_n\}} \cdot x = x_n \chi_{\{x_n\}}$ é uma função simples temos que

$$E(X) = \sum_{n=1}^{\infty} x_n\, p_X(\{x_n\}) = \sum_{n=1}^{\infty} x_n p_n\,.$$

Definição 3.10. O **momento de ordem** n de uma variável aleatória X é o número $E(X^n)$, $n = 1, 2, \ldots$.

Seja $\mu = E(X)$, então o **momento central de ordem** n é definido por

$$E((X - \mu)^n), \quad n = 1, 2, \ldots$$

Definição 3.11. Ao momento central de segunda ordem de uma variável aleatória X designa-se por **variância** de X e escreve-se,

$$\mathrm{Var}(X) = E\left((X - E(X))^n\right).$$

3.3 Condicionamento e independência: primeira abordagem

Dado um espaço de probabilidade $(\Omega, \mathcal{F}, P)$ e um acontecimento $B \in \mathcal{F}$ considere a seguinte colecção de acontecimentos

$$\mathcal{F}_B = \{A \cap B : A \in \mathcal{F}\}\ .$$

A colecção $\mathcal{F}_B$ pode ser vista como a colecção de todos os acontecimentos de $\mathcal{F}$ restritos ao acontecimento B.

Exercício 41. Mostre que $\mathcal{F}_B$ é uma σ-álgebra de partes de B.

Definição 3.12. Dados dois acontecimentos $A, B \in \mathcal{F}$ tal que $P(B) > 0$, então o número

$$P(A|B) = \frac{P(A \cap B)}{P(B)},$$

é designado por **probabilidade condicionada de** A **dado** B.

Proposição 3.23. *$(B, \mathcal{F}_B, P(\cdot|B))$ é um espaço de probabilidade.*

Demonstração. Deixa-se como exercício. □

É natural dizer-se que o acontecimento $A \in \mathcal{F}$ é independente do acontecimento $B \in \mathcal{F}$ se a ocorrência do acontecimento B não influenciar a probabilidade do acontecimento A. Ou seja,

$$P(A|B) = P(A).$$

Da definição de $P(A|B)$ obtemos que

$$P(A \cap B) = P(A)P(B),$$

que é usada como a definição de independência de dois acontecimentos. Generalizando para um número finito de acontecimentos, dizemos que $A_1, \dots, A_n \in \mathcal{F}$ são **independentes** se para qualquer escolha de índices $i_1, \dots, i_k \in \{1, \dots, n\}$ se tem

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k}) .$$

Esta definição também pode ser generalizada para σ-álgebras de acontecimentos.

Definição 3.13. Seja $(\Omega, \mathcal{F}, P)$ um espaço de probabilidade e $\mathcal{F}_i \subset \mathcal{F}$, $i = 1, \dots, n$ um conjunto finito de σ-álgebras. Diz-se que as **σ-álgebras $\mathcal{F}_1, \dots, \mathcal{F}_n$ são independentes** se para qualquer escolha de índices $i_1, \dots, i_k \in \{1, \dots, n\}$ e acontecimentos $A_{i_m} \in \mathcal{F}_{i_m}$ se tem

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k}) .$$

Exercício 42. Mostre que dois acontecimentos são independentes sse as σ-álgebras geradas por esses acontecimentos são independentes.

3.3.1 σ-álgebra gerada por X

Considere um espaço de probabilidade $(\Omega, \mathcal{F}, P)$ e uma variável aleatória $X : \Omega \to \mathbb{R}$. Os valores de $X(\Omega)$ constituem todas as medições que podemos observar de uma experiência aleatória. Como tal, é importante entender o nível de complexidade, diga-se aleatoriedade, da informação produzida pela variável aleatória. Para esse efeito, defini-se a seguinte colecção de acontecimentos,

$$\sigma(X) = X^{-1}(\mathcal{B}) = \left\{X^{-1}(B) : B \in \mathcal{B}\right\} .$$

Note-se que $\sigma(X) \subset \mathcal{F}$. Além disso é fácil verificar que $\sigma(X)$ é uma σ-álgebra de acontecimentos de Ω, que designamos por **σ-álgebra gerada pela variável aleatória** X. Também é usual designar-se $\sigma(X)$ por σ-álgebra induzida por X.

Exemplo 3.24. Quando a variável aleatória é constante, isto é, $X(\omega) = a$ para algum $a \in \mathbb{R}$ então $\sigma(X) = \{\emptyset, \Omega\}$ é a σ-álgebra trivial.

Exercício 43. Mostre que $\sigma(X)$ é uma σ-álgebra e que a função X é mensurável no espaço mensurável $(\Omega, \sigma(X))$.

Exercício 44. Considere uma variável aleatória que toma dois valores distintos $a, b \in \mathbb{R}$. Calcule $\sigma(X)$.

Exercício 45. Mostre que a σ-álgebra $\sigma(X)$ é a menor das σ-álgebras de partes de Ω que tornam X uma função mensurável.

Definição 3.14. Dizemos que duas variáveis aleatórias X e Y são **independentes** sse $\sigma(X)$ e $\sigma(Y)$ são independentes.

A definição de independência anterior pode ser facilmente generalizada para um número finito de variáveis aleatórias.

Proposição 3.25. *Duas variáveis aleatórias X e Y são independentes sse para quaisquer Borelianos $B, C \in \mathcal{B}$ tem-se*

$$P(X^{-1}(B) \cap Y^{-1}(C)) = P(X^{-1}(B))P(Y^{-1}(C)).$$

Demonstração. Deixa-se como exercício. □

Exemplo 3.26. Considere o espaço de probabilidade $([0,1], \mathcal{B}, m)$ e as variáveis aleatórias $X = \chi_{[0,\frac{1}{2}]}$ e $Y = \chi_{[\frac{1}{4},\frac{3}{4}]}$. Então $\sigma(X) = \left\{\emptyset, [0,1], [0,\frac{1}{2}], (\frac{1}{2},1]\right\}$ e $\sigma(Y) = \left\{\emptyset, [0,1], [\frac{1}{4},\frac{3}{4}], [0,\frac{1}{4}) \cup (\frac{3}{4},1]\right\}$ são claramente independentes.

3.4 Vectores aleatórios

Considere-se o espaço de medida $(\mathbb{R}, \mathcal{B}, m)$ onde $\mathcal{B}$ é a σ-álgebra de Borel e m a medida de Borel. É possível definir um espaço de medida sobre $\mathbb{R}^n$ fazendo o produto de n cópias de $(\mathbb{R}, \mathcal{B}, m)$. Portanto, definimos a σ-álgebra produto e a medida produto da maneira usual,

$$\mathcal{B}_n = \bigotimes_{i=1}^{n} \mathcal{B} \quad \text{e} \quad m_n = \underbrace{m \times \cdots \times m}_{n \text{ vezes}}.$$

Obtemos assim um espaço de medida $(\mathbb{R}^n, \mathcal{B}_n, m_n)$ tal que para cada rectângulo mensurável $B_1 \times \cdots \times B_n$ é válido

$$m_n(B_1 \times \cdots \times B_n) = m(B_1) \cdots m(B_n).$$

Observação 3.27. Note-se que $\mathcal{B}_n$ coincide com a σ-álgebra de Borel de $\mathbb{R}^n$.

Definição 3.15. Seja $(\Omega, \mathcal{F}, P)$ um espaço de probabilidade. Um **vector aleatório** é uma aplicação $X : \Omega \to \mathbb{R}^n$ mensurável.

Observação 3.28. Seja $\mathrm{pr}_i : \mathbb{R}^n \to \mathbb{R}$ a i-ésima projecção canónica, ou seja, a função que retorna a i-ésima coordenada de um vector em $\mathbb{R}^n$. Formalmente $\mathrm{pr}_i(x_1, \ldots, x_i, \ldots, x_n) = x_i$. Então é fácil verificar que $\mathrm{pr}_i \circ X : \Omega \to \mathbb{R}$ é uma variável aleatória e escreve-se $X_i = \mathrm{pr}_i \circ X$. De facto, dado um Boreliano $B \in \mathcal{B}$ então

$$X_i^{-1}(B) = X^{-1} \circ \mathrm{pr}_i^{-1}(B) = X^{-1}(\mathbb{R} \times \cdots \times B \times \cdots \times \mathbb{R}) \in \mathcal{F}.$$

Por outro lado, se $X_i : \Omega \to \mathbb{R}$ são variáveis aleatórias então a função $X : \Omega \to \mathbb{R}^n$ definida por

$$X(\omega) = (X_1(\omega), \ldots, X_n(\omega))$$

é mensurável.

Portanto, pode-se dizer que um vector aleatório é um vector de variáveis aleatórias.

Por comodidade de notação considere-se um vector aleatório bidimensional $X = (X_1, X_2) : \Omega \to \mathbb{R}^2$. A **distribuição conjunta de probabilidade de** $X = (X_1, X_2)$ é dada por

$$p_X(B) = P(X^{-1}(B)), \quad B \in \mathcal{B}_2 .$$

Designa-se por **função de distribuição conjunta de** $X = (X_1, X_2)$, a função $F_X : \mathbb{R}^2 \to \mathbb{R}$ definida por

$$F_X(x_1, x_2) = p_X(]-\infty, x_1] \times]-\infty, x_2]) .$$

Observação 3.29. A função de distribuição pode representar-se usando as seguintes expressões equivalentes

$$\begin{aligned} F_X(x_1, x_2) &= P(\omega \in \Omega : X_1(\omega) \le x_1,\ X_2(\omega) \le x_2) \\ &= P(X_1 \le x_1,\ X_2 \le x_2) . \end{aligned}$$

A distribuição conjunta de probabilidade de (X_1, X_2) determina as distribuições unidimensionais de cada variável aleatória X_i,

$$p_{X_1}(A) = p_X(A \times \mathbb{R}) \quad \text{e} \quad p_{X_2}(A) = p_X(\mathbb{R} \times A),$$

onde $A \in \mathcal{B}$ é um Boreliano de $\mathbb{R}$. As distribuições de probabilidade p_{X_1} e p_{X_2} são designadas por **distribuições de probabilidade marginais** e as funções de distribuição de probabilidade

$$F_{X_1}(x_1) = p_{X_1}(]-\infty, x_1]) \quad \text{e} \quad F_{X_2}(x_2) = p_{X_2}(]-\infty, x_2])$$

são designadas por **funções de distribuição marginais**.

Proposição 3.30. *A função de distribuição conjunta de* (X_1, X_2) *verifica as seguintes propriedades:*

1. $F_X(x_1, x_2)$ *é contínua à direita relativamente a cada variável.*
2. *se* $x_1 \le y_1$ *e* $x_2 \le y_2$ *então* $F_X(x_1, x_2) \le F_X(y_1, y_2)$.
3. $\lim_{(x_1,x_2)\to(+\infty,+\infty)} F_X(x_1, x_2) = 1$
4. $\lim_{x_1\to-\infty} F_X(x_1, x_2) = \lim_{x_2\to-\infty} F_X(x_1, x_2) = 0.$
5. $\lim_{x_1\to+\infty} F_X(x_1, x_2) = F_{X_2}(x_2)$ *e* $\lim_{x_2\to+\infty} F_X(x_1, x_2) = F_{X_1}(x_1)$.

Demonstração. Deixa-se como exercício. □

Diz-se que o vector aleatório $X = (X_1, X_2)$ é **absolutamente contínuo** sse a distribuição conjunta de probabilidade p_X é absolutamente contínua relativamente à medida produto $m_2 = m \times m$. Segue do teorema de Radon-Nikodym que existe uma função integrável $f_X : \mathbb{R}^2 \to \mathbb{R}$, designada por **função de densidade conjunta de** X tal que

$$p_X(B) = \int_B f_X \, dm_2, \quad B \in \mathcal{B}_2 .$$

Proposição 3.31. *Seja $X = (X_1, X_2)$ um vector aleatório absolutamente contínuo. Então as variáveis aleatórias X_1 e X_2 são absolutamente contínuas com funções densidade de probabilidade,*

$$f_{X_1}(x_1) = \int_{\mathbb{R}} f_X(x_1, x_2) \, dm(x_2) \quad e \quad f_{X_2}(x_2) = \int_{\mathbb{R}} f_X(x_1, x_2) \, dm(x_1) .$$

Demonstração. Suponha que $m(B) = 0$ para algum Boreliano $B \in \mathcal{B}$. Então $m_2(B \times \mathbb{R}) = m(B) \times m(\mathbb{R}) = 0$. Logo $p_X(B \times \mathbb{R}) = 0$. Por outro lado, $p_{X_1}(B) = P_X(B \times \mathbb{R})$. Segue que, se $m(B) = 0$ então $p_{X_1}(B) = 0$, ou seja, $p_{X_1} \ll m$. Adicionalmente, temos pelo Teorema de Fubini que

$$\begin{aligned} p_{X_1}(B) = p_X(B \times \mathbb{R}) &= \int_{B \times \mathbb{R}} f_X(x_1, x_2) \, d(m \times m) \\ &= \int_B \left(\int_{\mathbb{R}} f_X(x_1, x_2) \, dm(x_2) \right) dm(x_1) . \end{aligned}$$

□

As funções de densidade de probabilidade da proposição anterior designam-se por **funções de densidade marginais**.

3.4.1 Independência

Suponha que as variáveis aleatórias X_1 e X_2 são independentes. Então para qualquer rectângulo Boreliano $B_1 \times B_2$ temos que

$$p_{(X_1, X_2)}(B_1 \times B_2) = P((X_1, X_2) \in B_1 \times B_2) = P(X_1^{-1}(B_1) \cap X_2^{-1}(B_2)) .$$

Logo, segue da Proposição 3.25 que

$$P(X_1^{-1}(B_1) \cap X_2^{-1}(B_2)) = P(X_1^{-1}(B_1)) P(X_2^{-1}(B_2)) .$$

Portanto,

$$p_{(X_1, X_2)}(B_1 \times B_2) = p_{X_1}(B_1) \times p_{X_2}(B_2) .$$

Ou seja, a distribuição conjunta de probabilidade $p_{(X_1,X_2)}$ coincide com a medida produto $p_{X_1} \times p_{X_2}$ se X_1 e X_2 forem independentes. De facto, a implicação no sentido inverso também é válida.

Proposição 3.32. *Duas variáveis aleatórias X_1 e X_2 são independentes sse*

$$p_{(X_1,X_2)} = p_{X_1} \times p_{X_2} \,.$$

Demonstração. Deixa-se como exercício. □

Segue da proposição anterior que se X_1 e X_2 forem independentes então

$$F_{(X_1,X_2)}(x_1,x_2) = F_{X_1}(x_1)F_{X_2}(x_2) \,.$$

Ou seja, a função de distribuição conjunta é o produto das funções de distribuição marginais.

Adicionalmente, se o vector aleatório (X_1, X_2) for absolutamente contínuo então temos o seguinte resultado.

Proposição 3.33. *Suponha que (X_1, X_2) é absolutamente contínuo. Então as variáveis aleatórias X_1 e X_2 são independentes sse*

$$f_{(X_1,X_2)}(x_1,x_2) = f_{X_1}(x_1)f_{X_2}(x_2) \,.$$

Demonstração. Deixa-se como exercício. □

A independência das variáveis aleatórias X_1 e X_2 permite estabelecer uma fórmula importante no cálculo de valores esperados.

Teorema 3.34. *Se X_1 e X_2 são duas variáveis aleatórias independentes então*

$$E(X_1X_2) = E(X_1)E(X_2) \,.$$

Demonstração. Seja $h : \mathbb{R}^2 \to \mathbb{R}$ a função mensurável definida por $h(x_1, x_2) = x_1x_2$ e $X : \Omega \to \mathbb{R}^2$ o vector aleatório $X = (X_1, X_2)$. Então $h \circ X(\omega) = X_1(\omega)X_2(\omega)$ é uma função integrável de $\Omega \to \mathbb{R}$. Logo, segue do teorema da mudança de variável que

$$E(X_1X_2) = \int_\Omega h \circ X \, dP = \int_{\mathbb{R}^2} h \, dp_X \,.$$

Por outro lado, segue do Teorema de Fubini e da Proposição 3.32 que

$$\begin{aligned}\int_{\mathbb{R}^2} h \, dp_X &= \int_{\mathbb{R}} \int_{\mathbb{R}} x_1x_2 \, d(p_{X_1} \times p_{X_2}) \\ &= \int_{\mathbb{R}} x_1 \, dp_{X_1} \int_{\mathbb{R}} x_2 \, dp_{X_2} \\ &= E(X_1)E(X_2) \,.\end{aligned}$$

□

Teorema 3.35 (Desigualdade de Cauchy-Schwarz). *Sejam X_1 e X_2 duas variáveis aleatórias com segundos momentos finitos, isto é,*

$$E(X_1^2) < \infty \quad e \quad E(X_2^2) < \infty \,.$$

Então

$$E^2(X_1X_2) \leq E(X_1^2)E(X_2^2)\,.$$

Demonstração. Seja $\lambda \in \mathbb{R}$ e considere-se a variável aleatória $Y = X_1 + \lambda X_2$. Então

$$\begin{aligned} 0 \leq E(Y^2) &= E(X_1^2 + 2\lambda X_1X_2 + \lambda^2 X_2^2) \\ &= E(X_1^2) + 2\lambda E(X_1X_2) + \lambda^2 E(X_2^2)\,. \end{aligned}$$

Um polinómio de segundo grau em λ é não negativo sse o seu discriminante é não positivo, ou seja,

$$(2E(X_1X_2))^2 - 4E(X_2^2)E(X_1^2) \leq 0\,.$$

□

Observação 3.36. Tomando uma das variáveis na proposição anterior igual à identidade obtém-se

$$E^2(X) \leq E(X^2)\,.$$

Ou seja, se o segundo momento da variável aleatória é finito então o seu valor esperado é finito.

Definição 3.16. A **covariância** de duas variáveis aleatórias X_1 e X_2 é o número,

$$\operatorname{Cov}(X_1, X_2) = E\left[(X_1 - E(X_1))(X_2 - E(X_2))\right]\,.$$

Exercício 46. Mostre que $\operatorname{Cov}(X_1, X_2) = E(X_1X_2) - E(X_1)E(X_2)\,.$

Note-se que se X_1 e X_2 forem independentes então pelo teorema anterior tem-se

$$\operatorname{Cov}(X_1, X_2) = 0\,.$$

A desigualdade de Cauchy-Schwarz permite definir uma quantidade entre -1 e 1 que mede o grau de dependência de duas variáveis aleatórias X_1 e X_2.

Definição 3.17. O **coeficiente de correlação** de duas variáveis aleatórias X_1 e X_2 é o número,

$$\rho_{X_1,X_2} = \frac{\operatorname{Cov}(X_1, X_2)}{\sqrt{\operatorname{Var}(X_1)\operatorname{Var}(X_2)}}\,.$$

Observação 3.37. Segue da desigualdade de Cauchy-Schwarz que $|\rho_{X_1,X_2}| \leq 1$. Adicionalmente, se X_1 e X_2 forem independentes então $\rho_{X_1,X_2} = 0$. No entanto, se $X_1 = X_2$ então $\rho_{X_1,X_2} = 1$.

4 Esperança Condicional

A esperança condicional é um conceito central em teoria da probabilidade, em particular no estudo de processos estocásticos. Considere uma variável aleatória $X : \Omega \to \mathbb{R}$ definida num espaço de probabilidade $(\Omega, \mathcal{F}, P)$. Suponha que desejamos obter a melhor aproximação de X dado que conhecemos alguma informação de $\mathcal{F}$. A melhor aproximação é num certo sentido dada pela esperança condicional. Faremos a sua definição progressivamente, começando por definir a esperança condicional dado um acontecimento $A \in \mathcal{F}$.

4.1 Esperança condicional dado um acontecimento

Definição 4.1. Dada uma variável aleatória X e um acontecimento $A \in \mathcal{F}$ tal que $P(A) > 0$, a **esperança condicional de** X **dado** A é definida por

$$E(X|A) = \frac{1}{P(A)} \int_A X \, dP \, .$$

Exemplo 4.1. Considere duas moedas de 20 e 50 cêntimos. Lançam-se as moedas ao ar e somam-se os montantes das moedas que ficaram com a face voltada para cima. Esse montante é o valor da variável aleatória X. Seja A o acontecimento de uma moeda ficar com a face voltada para cima. Vamos calcular $E(X|A)$. Note-se que $A = \{EF, FE\}$ onde E simboliza escudo e F face. Então

$$X(EF) = 50 \quad \text{e} \quad X(FE) = 20 \, .$$

Logo,

$$E(X|A) = \frac{1}{\frac{1}{2}} \left(50 \times \frac{1}{4} + 20 \times \frac{1}{4} \right) = 35 \, .$$

Exercício 47. Mostre que $E(X|\Omega) = E(X)$.

Proposição 4.2. *Sejam A e B dois acontecimentos de $\mathcal{F}$ tais que $P(B) > 0$. Então*

$$E(\chi_A|B) = P(A|B) \, .$$

Demonstração. Deixa-se como exercício. □

Através do resultado anterior, podemos definir a probabilidade condicionada de A dado B usando da esperança condicional da variável aleatória χ_A dado B.

4.2 Esperança condicional dado uma variável aleatória discreta

Seja $Y : \Omega \to \mathbb{R}$ uma variável aleatória discreta. Suponha que $Y(\Omega) = \{y_1, y_2, \ldots\}$ e que $P(Y = y_n) > 0$. Desejamos condicionar uma variável aleatória X dado a variável aleatória Y. Como não sabemos *a priori* qual dos acontecimentos $A_n = Y^{-1}(y_n)$ pode ocorrer, é necessário considerar todas as esperanças condicionais,

$$E(X|A_1), E(X|A_2), \ldots .$$

Definição 4.2. Seja X uma variável aleatória e Y uma variável aleatória discreta. A **esperança condicional de** X **dado** Y é a função $E(X|Y) : \Omega \to \mathbb{R}$ definida por

$$E(X|Y)(\omega) = E(X|A_n) \quad \text{se } \omega \in A_n .$$

Observação 4.3. É possível escrever a esperança condicional de X dado Y como,

$$E(X|Y) = \sum_{n=1}^{\infty} E(X|A_n)\chi_{A_n} ,$$

onde $A_n = Y^{-1}(y_n)$.

Exemplo 4.4. Continuando com o exemplo anterior, seja Y a variável aleatória que retorna o montante da primeira moeda, de 20 cêntimos, caso esta se encontre de face voltada para cima. Então

$$E(X|Y^{-1}(0)) = 25 \quad \text{e} \quad E(X|Y^{-1}(20)) = 45 .$$

Logo,

$$E(X|Y)(\omega) = \begin{cases} 25 & \text{se } Y(\omega) = 0 \\ 45 & \text{se } Y(\omega) = 20 \end{cases}$$

Exercício 48. Mostre que $E(X|Y) = E(X)$ caso $Y(\omega) = c$ para todo $\omega \in \Omega$.

Proposição 4.5. *Seja X uma variável aleatória e Y uma variável aleatória discreta. Então*

1. *$E(X|Y)$ é uma função mensurável relativamente à σ-álgebra $\sigma(Y)$.*
2. *Para todo $A \in \sigma(Y)$,*

$$\int_A E(X|Y)\, dP = \int_A X\, dP .$$

Demonstração. Que $E(X|Y)$ é uma função mensurável relativamente a $\sigma(Y)$ segue da Observação 4.3.

Por outro lado, para cada $n \geq 1$ temos que

$$\int_{A_n} E(X|Y)\, dP = \int_{\Omega} E(X|A_n)\chi_{A_n}\, dP = E(X|A_n)P(A_n) = \int_{A_n} X\, dP\,.$$

Como os conjuntos A_n são disjuntos tem-se o resultado. □

4.3 Esperança condicional dado uma variável aleatória arbitrária

Usando a igualdade da proposição anterior podemos definir esperança condicional dado qualquer variável aleatória.

Definição 4.3. Seja X e Y duas variáveis aleatórias. A **esperança condicional de** X **dado** Y é uma função $E(X|Y) : \Omega \to \mathbb{R}$ que satisfaz

1. $E(X|Y)$ é mensurável relativamente a $\sigma(Y)$.
2. Para todo $A \in \sigma(Y)$,

$$\int_A E(X|Y)\, dP = \int_A X\, dP\,.$$

Observação 4.6. A existência de uma função que satisfaça os requisitos da definição anterior é garantida pelo Teorema de Radon-Nikodym. De facto, supondo que X é integrável, a função $\nu^{\pm} : \sigma(Y) \to [0, \infty]$ definida por

$$\nu^{\pm}(A) = \int_A X^{\pm}\, dP\,,$$

é uma medida finita no espaço mensurável $(\Omega, \sigma(Y))$. Por outro lado, se $P(A) = 0$ então $\nu^{\pm}(A) = 0$, ou seja, $\nu^{\pm} \ll P$. Logo, existem funções integráveis $h, g : \Omega \to \mathbb{R}$ (mensuráveis relativamente a $\sigma(Y)$) tal que para todo $A \in \sigma(Y)$,

$$\nu^{+}(A) = \int_A h\, dP \quad \text{e} \quad \nu^{-}(A) = \int_A g\, dP\,.$$

Portanto, a função integrável $f = h - g$ satisfaz

$$\int_A f\, dP = \int_A X\, dP\,, \quad \forall\, A \in \sigma(Y)\,.$$

Adicionalmente, a função f é única P-q.c. Logo, f é a esperança condicional de X dado Y (ou se quiser, uma versão da esperança condicional a menos de um conjunto de medida nula).

4.3.1 Probabilidade condicionada

Definição 4.4 (Probabilidade condicionada)**.** Seja X uma variável aleatória. A função $P(\cdot|X) : \mathcal{F} \times \Omega \to [0,1]$ definida por

$$P(A|X)(\omega) = E(\chi_A|X)(\omega)$$

é designada por **probabilidade condicionada de um acontecimento A dado a variável aleatória X.**

Segue do ponto 2. da definição da esperança condicional que dado um acontecimento $B \in \sigma(X)$ então

$$\int_B P(A|X)\, dP = P(A \cap B)\,.$$

4.3.2 Esperança e probabilidade condicionada como funções de $\mathbb{R}$

Sejam X e Y duas variáveis aleatórias. A esperança condicional $E(X|Y) : \Omega \to \mathbb{R}$ é por definição uma função $\sigma(Y)$-mensurável. Logo, qualquer que seja $y \in \mathbb{R}$ temos que

$$E(X|Y)(\omega_1) = E(X|Y)(\omega_2)\,, \quad \forall\, \omega_1, \omega_2 \in Y^{-1}(y)\,.$$

Isto é, a função $E(X|Y)$ é constante nos conjuntos $Y^{-1}(y)$. Portanto, existe uma única função $g : \mathbb{R} \to \mathbb{R}$ mensurável tal que $E(X|Y)(\omega) = g \circ Y(\omega)$, ou seja, é válido o seguinte diagrama,

$$\begin{tikzcd} \Omega \arrow[r, "Y"] \arrow[d, "E(X|Y)"'] & \mathbb{R} \arrow[dl, "g"] \\ \mathbb{R} & \end{tikzcd}$$

A função g obtida é a **esperança condicional de X dado Y vista como uma função de** $\mathbb{R}$ e é usual escrever-se

$$g(y) = E(X|Y = y)\,.$$

Analogamente, a **probabilidade condicionada de um acontecimento A dado Y vista como uma função de** $\mathbb{R}$ é

$$P(A|Y = y) = E(\chi_A|Y = y)\,.$$

Quando o vector (X, Y) é absolutamente contínuo, a seguinte proposição permite calcular $E(X|Y = y)$ de forma explícita.

Proposição 4.7. *Seja (X,Y) um vector aleatório absolutamente contínuo. Então*

$$E(X|Y=y) = \int_{\mathbb{R}} x\, f_{X|Y}(x,y)\, dm(x)\,,$$

onde

$$f_{X|Y}(x,y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}\,.$$

Demonstração. Aplicando o teorema da mudança de variável temos que

$$\int_{Y^{-1}(B)} E(X|Y)\, dP = \int_B E(X|Y=y)\, dp_Y(y) \quad \forall B \in \mathcal{B}\,.$$

Por outro lado, aplicando novamente o teorema da mudança de variável vem que

$$\int_{Y^{-1}(B)} X\, dP = \int_{\mathbb{R}\times B} x\, dp_{X,Y}(x,y) \quad \forall B \in \mathcal{B}\,.$$

Logo, segue da definição de esperança condicional que

$$\int_B E(X|Y=y)\, dp_Y(y) = \int_{\mathbb{R}\times B} x\, dp_{X,Y}(x,y)\,.$$

Como X e Y são absolutamente contínuas segue do Teorema de Fubini que

$$\begin{aligned}\int_B E(X|Y=y) f_Y(y)\, dm(y) &= \int_{\mathbb{R}\times B} x f_{X,Y}(x,y)\, dm_2(x,y)\\ &= \int_B \left(\int_{\mathbb{R}} x f_{X,Y}(x,y)\, dm(x)\right) dm(y)\end{aligned}$$

Como a igualdade anterior é valida para todo o Boreliano B de $\mathbb{R}$ então

$$E(X|Y=y) = \int_{\mathbb{R}} x \frac{f_{X,Y}(x,y)}{f_Y(y)}\, dm(x)\,.$$

□

Definição 4.5. A função $f_{X|Y}$ obtida na proposição anterior é designada por **função de densidade condicional de X dado Y**.

Proposição 4.8. *Seja (X,Y) um vector aleatório absolutamente contínuo. Então*

$$P(X^{-1}(B)|Y=y) = \int_B f_{X|Y}(x,y)\, dm(x)\,, \quad \forall\, B \in \mathcal{B}\,.$$

Figure 1: Exemplo 4.10: gráficos de X, Y e $E(X|Y)$

Demonstração. Deixa-se como exercício. $\square$

Observação 4.9. É usual escrever-se a probabilidade condicionada de X dado Y como,

$$P(X \in B|Y = y) = P(X^{-1}(B)|Y = y)\,.$$

Exemplo 4.10. Considere-se o espaço de probabilidade $([0,1], \mathcal{B}, m)$ onde $\mathcal{B}$ e m são a σ-álgebra de Borel e medida de Lebesgue restritas ao intervalo $[0,1]$. Sejam $X, Y : [0,1] \to \mathbb{R}$ variáveis aleatórias definidas por

$$X(\omega) = \omega \quad \text{e} \quad Y(\omega) = \begin{cases} 2\omega & \text{se } 0 \leq \omega \leq \frac{1}{2} \\ 2\omega - 1 & \text{se } \frac{1}{2} < \omega \leq 1 \end{cases}.$$

Note-se que X e Y são variáveis aleatórias contínuas. Iremos calcular $E(X|Y)$.

Em primeiro lugar determinamos $\sigma(Y)$. Tome-se um Boreliano $B \subset [0,1]$, então $Y^{-1}(B) = \frac{B}{2} \cup (\frac{B}{2} + \frac{1}{2})$. Portanto,

$$\sigma(Y) = \left\{ A \cup \left(A + \frac{1}{2} \right) : A \subset [0, \frac{1}{2}] \text{ é Boreliano} \right\}.$$

Uma vez que $E(X|Y) : [0,1] \to \mathbb{R}$ tem de ser $\sigma(Y)$-mensurável então $E(X|Y)^{-1}(\{y\}) = A \cup \left(A + \frac{1}{2}\right)$ para algum Boreliano $A \subset [0, 1/2]$. Logo

$$E(X|Y)(\omega) = E(X|Y)(\omega + \frac{1}{2})\,, \quad \forall\, \omega \in [0, \frac{1}{2}]\,. \tag{4.1}$$

Por outro lado, pela definição $E(X|Y)$ tem que satisfazer a igualdade

$$\int_C E(X|Y)\, dm = \int_C X\, dm\,, \quad \forall\, C \in \sigma(Y)\,. \tag{4.2}$$

Como $C = A \cup (A + \frac{1}{2})$ para algum Boreliano $A \subset [0, 1/2]$ podemos reescrever o primeiro integral,

$$\begin{aligned}\int_{A\cup(A+\frac{1}{2})} E(X|Y)\, dm &= \int_A E(X|Y)\, dm + \int_{A+\frac{1}{2}} E(X|Y)\, dm \\ &= \int_A E(X|Y)\, dm + \int_A E(X|Y)\, dm \\ &= 2\int_A E(X|Y)\, dm\,,\end{aligned}$$

onde a segunda igualdade é obtida através de mudança de variável $\omega \mapsto \omega + \frac{1}{2}$ e tendo em conta (4.1). Analogamente, reescrevendo o segundo integral de (4.2) obtemos,

$$\begin{aligned}\int_{A\cup(A+\frac{1}{2})} \omega\, dm &= \int_A \omega\, dm + \int_{A+\frac{1}{2}} \omega\, dm \\ &= \int_A \omega\, dm + \int_A (\omega + \frac{1}{2})\, dm \\ &= \int_A (2\omega + \frac{1}{2})\, dm\,.\end{aligned}$$

Ou seja, para todo o Boreliano $A \subset [0, \frac{1}{2}]$ temos que

$$\int_A E(X|Y)\, dm = \int_A (\omega + \frac{1}{4})\, dm\,.$$

Logo $E(X|Y) = \omega + \frac{1}{4}$ para $\omega \in [0, 1/2]$. Segue de (4.1) que

$$E(X|Y)(\omega) = \begin{cases} \omega + \frac{1}{4} & \text{se } 0 \le \omega \le \frac{1}{2} \\ \omega - \frac{1}{4} & \text{se } \frac{1}{2} < \omega \le 1 \end{cases}$$

4.4 Esperança condicional dado uma σ-álgebra

É fácil verificar que a esperança condicional de X dado uma variável aleatória arbitrária Y não depende dos valores que Y pode tomar e que depende apenas da σ-álgebra gerada por Y.

Proposição 4.11. *Sejam Y e Z duas variáveis aleatórias tal que $\sigma(Y) = \sigma(Z)$. Então $E(X|Y) = E(X|Z)$ P-q.c.*

Demonstração. Seja $\mathcal{G} = \sigma(Y) = \sigma(Z)$. Segue da definição de esperança condicional que

$$\int_A E(X|Y)\, dP = \int_A E(X|Z)\, dP\,, \quad \forall\, A \in \mathcal{G}\,.$$

Logo, segue do Teorema 2.31 que $E(X|Y) = E(X|Z)$ P-q.c. $\square$

Portanto, é natural definir-se a esperança condicional de X dado uma σ-álgebra de partes de Ω contida em $\mathcal{F}$. A definição que se segue generaliza todas as anteriores.

Definição 4.6. Seja $(\Omega, \mathcal{F}, P)$ um espaço de probabilidade, $X : \Omega \to \mathbb{R}$ uma variável aleatória e $\mathcal{G} \subset \mathcal{F}$ uma σ-álgebra de partes de Ω. A **esperança condicional de X dado $\mathcal{G}$** é uma função $E(X|\mathcal{G}) : \Omega \to \mathbb{R}$ que satisfaz

1. $E(X|\mathcal{G})$ é mensurável relativamente a $\mathcal{G}$.
2. Para todo $A \in \mathcal{G}$,

$$\int_A E(X|\mathcal{G})\, dP = \int_A X\, dP\,.$$

Observação 4.12. Analogamente ao que foi feito nas secções anteriores, a **probabilidade condicionada de um acontecimento dado uma σ-álgebra** $\mathcal{G} \subset \mathcal{F}$ é a função $P(\cdot|\mathcal{G}) : \mathcal{F} \times \Omega \to \mathbb{R}$ definida por

$$P(A|\mathcal{G})(\omega) = E(\chi_A|\mathcal{G})(\omega)\,.$$

Exercício 49. Mostre que se $\mathcal{G} = \{\emptyset, \Omega\}$ então $E(X|\mathcal{G}) = E(X)$ P-q.c.

Observação 4.13. Sejam X e Y duas variáveis aleatórias. Note-se que a esperança condicional de X dado Y é igual à esperança condicional de X dado a σ-álgebra gerada por Y (a informação de Y), ou seja,

$$E(X|Y) = E(X|\sigma(Y))\,.$$

A Definição 4.6 permite também definir a **esperança condicional da variável aleatória X dado um vector aleatório** $Y = (Y_1, \ldots, Y_n)$. De facto, denote-se por $\sigma(Y_1, \ldots, Y_n)$ a σ-álgebra induzida pelo vector aleatório Y, ou seja, $\sigma(Y) = Y^{-1}(\mathcal{B}_n)$ onde $\mathcal{B}_n$ são os Borelianos de $\mathbb{R}^n$. Então definimos,

$$E(X|Y_1, \ldots, Y_n) = E(X|\sigma(Y_1, \ldots, Y_n))\,.$$

Resumimos na proposição seguinte as propriedades da esperança condicional. Todas as igualdades que se seguem devem ser entendidas P-q.c.

Proposição 4.14. *Sejam X e Y variáveis aleatórias, $\mathcal{G}, \mathcal{H} \subset \mathcal{F}$ σ-álgebras de partes de Ω e $\alpha, \beta \in \mathbb{R}$. Então*

1. $E(\alpha X + \beta Y|\mathcal{G}) = \alpha E(X|\mathcal{G}) + \beta E(Y|\mathcal{G})$ *(linearidade).*

2. *$E(E(X|\mathcal{G})) = E(X)$ (esperança condicional de X tem o mesmo valor esperado de X).*
3. *Se X é $\mathcal{G}$-mensurável então $E(X|\mathcal{G}) = X$ (se conhecemos a informação de X então a melhor aproximação de X é X.)*
4. *Se Y é $\mathcal{G}$-mensurável e XY é integrável então $E(XY|\mathcal{G}) = YE(X|\mathcal{G})$ (pode-se retirar para fora da esperança condicional a informação que se conhece).*
5. *Se $\sigma(X)$ é independente de $\mathcal{G}$ então $E(X|\mathcal{G}) = E(X)$ (se $\mathcal{G}$ não fornece nenhuma informação sobre X então o seu valor esperado é a melhor aproximação).*
6. *Se $\mathcal{H} \subset \mathcal{G}$ então $E(E(X|\mathcal{G})|\mathcal{H}) = E(X|\mathcal{H})$.*
7. *Se $X \geq 0$ então $E(X|\mathcal{G}) \geq 0$.*

Demonstração.

1. Para todo $A \in \mathcal{G}$, segue da linearidade do integral de Lebesgue e da definição de esperança condicional que

$$\begin{aligned}\int_A E(\alpha X + \beta Y|\mathcal{G})\, dP &= \int_A \alpha X + \beta Y\, dP \\ &= \alpha \int_A X\, dP + \beta \int_A \\ &= \alpha \int_A E(X|\mathcal{G})\, dP + \beta \int_A E(Y|\mathcal{G})\, dP \\ &= \int_A \left(\alpha E(X|\mathcal{G}) + \beta E(Y|\mathcal{G})\right)\, dP\,.\end{aligned}$$

   Logo, segue do Teorema 2.31 que $E(\alpha X + \beta Y|\mathcal{G}) = \alpha E(X|\mathcal{G}) + \beta E(Y|\mathcal{G})$ P-q.c.

2. Basta tomar $A = \Omega$ na definição de esperança condicional.
3. Se X é $\mathcal{G}$-mensurável então, pela definição, é uma versão da esperança condicional de X dado $\mathcal{G}$.
4. Suponha que $Y = \chi_B$ para algum $B \in \mathcal{G}$. Então, para todo $A \in \mathcal{G}$ temos que

$$\begin{aligned}\int_A E(\chi_B X|\mathcal{G})\, dP &= \int_A \chi_B X\, dP \\ &= \int_{A\cap B} X\, dP \\ &= \int_{A\cap B} E(X|\mathcal{G})\, dP \\ &= \int_A \chi_B E(X|\mathcal{G})\, dP\,.\end{aligned}$$

Logo $E(\chi_B X|\mathcal{G}) = \chi_B E(X|\mathcal{G})$. Aproximando Y por funções simples e usando o teorema da convergência dominada obtém-se o resultado.

5. Uma vez que $\sigma(X)$ é independente de $\mathcal{G}$ então as variáveis aleatórias χ_A e X são independentes para todo $A \in \mathcal{G}$. Aplicando o Teorema 3.34 segue que
$$\int_A E(X|\mathcal{G})\, dP = E(\chi_A \cdot X) = E(\chi_A)E(X) = \int_A E(X)\, dP\,.$$
Como a igualdade anterior é válida para todo $A \in \mathcal{G}$ temos que $E(X|\mathcal{G}) = E(X)$.

6. Como $\mathcal{H} \subset \mathcal{G}$ temos que
$$\int_B E(X|\mathcal{H})\, dP = \int_B X\, dP = \int_B E(X|\mathcal{G})\, dP\,, \quad \forall\, B \in \mathcal{H}\,.$$
Por outro lado, como $E(X|\mathcal{G})$ é uma função $\mathcal{G}$-mensurável, a sua esperança condicional dado $\mathcal{H}$ satisfaz,
$$\int_B E(E(X|\mathcal{G})|\mathcal{H})\, dP = \int_B E(X|\mathcal{G})\, dP\,, \quad \forall\, B \in \mathcal{H}\,.$$

7. Seja $A_n = \left\{\omega \in \Omega : E(X|\mathcal{G})(\omega) \leq -\frac{1}{n}\right\}$. Como $A_n \in \mathcal{G}$ temos que
$$0 \leq \int_{A_n} X\, dP = \int_{A_n} E(X|\mathcal{G})\, dP \leq -\frac{P(A_n)}{n}\,.$$
Logo, $P(A_n) = 0$ para todo $n = 1, 2, \ldots$. Então
$$P(\{\omega \in \Omega \colon E(X|\mathcal{G})(\omega) < 0\}) = P(\bigcup_{n=1}^{\infty} A_n) = 0\,.$$

□

5 Sucessões de variáveis aleatórias

Nesta secção estudaremos **sucessões de variáveis aleatórias**, ou seja, sucessões $(X_1, X_2, \ldots)$ onde $X_i : \Omega \to \mathbb{R}$, $i = 1, 2, \ldots$ são variáveis aleatórias.

Sucessões de variáveis aleatórias são processos estocásticos em tempo discreto, cuja definição veremos na secção seguinte. São geralmente usados para modelar fenómenos aleatórios que variam ao longo do tempo e onde se podem realizar observações em momentos no tempo espaçados entre si. Como exemplo, considere-se o lançamento de uma moeda um número arbitrário de vezes ou a cotação de um índice em bolsa observado no final de cada dia.