Notas de Teoria da Probabilidade e Processos Estocásticos

José Pedro Gaivão

Resumo

Estas notas destinam-se à disciplina de Teoria da Probabilidade e Processos Estocásticos do Mestrado de Matemática Financeira do ISEG do ano lectivo 2012/2013. Parte do material aqui apresentado foi retirado dos textos [1] e [2].

Conteúdo

Parte I
Medida e Integração

Nesta primeira parte do curso, faremos uma breve introdução à teoria da medida e integral de Lebesgue. Os resultados que iremos apresentar são fundamentais para o desenvolvimento da teoria da probabilidade.

Em teoria da medida, pretende-se construir funções que a cada subconjunto de $\mathbb{R}$ associam um valor positivo, uma medida que generalize o conceito usual de comprimento de um intervalo. Como motivação, relembremos brevemente o integral de Riemann. O integral de Riemann de uma função $f : [a, b] \to \mathbb{R}$ pode ser aproximada por somas do tipo

$$\sum_{i=1}^{n} f(x_i)\mathrm{comp}(E_i)\,,$$

onde E_i são intervalos disjuntos cuja união é $[a, b]$, $x_i \in E_i$ e $\mathrm{comp}(E_i)$ é o comprimento do intervalo E_i. Em teoria da probabilidade, o integral de Riemann é insatisfatório, uma vez que existem funções simples não integráveis, como é o caso da função (variável aleatória) definida no intervalo $[0, 1]$ que atribui 1 aos racionais e 0 aos irracionais. Com o objectivo de generalizar o integral de Riemann, o matemático francês Henri Lebesgue procurou uma medida m que tivesse as seguintes propriedades:

1. $m(\emptyset) = 0$.
2. $m(A) \geq 0$ para qualquer $A \subset \mathbb{R}$.
3. $m(I) = \mathrm{comp}(I)$ para qualquer intervalo $I \subset \mathbb{R}$.
4. $m(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} m(A_i)$ para quaisquer $A_i \subset \mathbb{R}$, $i = 1, 2, \ldots$, disjuntos dois a dois.
5. $m(A + x) = m(A)$ para todo $A \subset \mathbb{R}$ e $x \in \mathbb{R}$ onde $A + x = \{a + x : a \in A\}$.

No entanto, para que estas condições fossem compatíveis entre si, teve-se que restringir o cálculo da medida a certos subconjuntos de $\mathbb{R}$, denominados por *conjuntos mensuráveis*.

1 Espaços de medida

1.1 σ-álgebras

Definição 1.1. Seja X um conjunto não vazio e $\mathcal{P}(X)$ a colecção de todos os subconjuntos de X. Uma subcolecção $\mathcal{F} \subset \mathcal{P}(X)$ é uma **σ-álgebra** sse

1. $\emptyset \in \mathcal{F}$,
2. se $A \in \mathcal{F}$ então $A^{\mathrm{c}} \in \mathcal{F}$,
3. se $A_i \in \mathcal{F}$, $i = 1, 2, \ldots$ então $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

Ao par $(X, \mathcal{F})$ chamamos **espaço mensurável** e os elementos de $\mathcal{F}$ são os **conjuntos mensuráveis**.

Exemplo 1.1. Alguns exemplos de σ-álgebras:

1. $\mathcal{F} = \{\emptyset, X\}$
2. $\mathcal{F} = \{\emptyset, X, A, A^c\}$
3. $\mathcal{F} = \mathcal{P}(X)$

Exercício 1. Seja $(X, \mathcal{F})$ um espaço mensurável. Mostre que

1. $X \in \mathcal{F}$.
2. se $A_i \in \mathcal{F}$, $i = 1, 2, \ldots$ então $\bigcup_{i=1}^n A_i \in \mathcal{F}$.
3. se $A_i \in \mathcal{F}$, $i = 1, 2, \ldots$ então $\bigcap_{i=1}^\infty A_i \in \mathcal{F}$.
4. se $A_i \in \mathcal{F}$, $i = 1, 2, \ldots$ então $\bigcap_{i=1}^n A_i \in \mathcal{F}$
5. se $A, B \in \mathcal{F}$ então $A \setminus B \in \mathcal{F}$.

Proposição 1.2. *Sejam X um conjunto e $\mathcal{C}$ uma colecção de subconjuntos de X. Então existe uma única σ-álgebra $\mathcal{F}^*$ que contêm $\mathcal{C}$ tal que se $\mathcal{F}$ é uma σ-álgebra que contêm $\mathcal{C}$ então $\mathcal{F}^* \subset \mathcal{F}$.*

Demonstração. Seja $\mathfrak{F}$ o conjunto de todas as σ-álgebras que contêm $\mathcal{C}$. Note-se que $\mathfrak{F}$ é não vazio uma vez que $\mathcal{P}(X) \in \mathfrak{F}$. Seja $\mathcal{F}^*$ a intersecção de todas as σ-álgebras pertencentes a $\mathfrak{F}$, ou seja

$$\mathcal{F}^* = \{A \in \mathcal{P}(X) : A \in \mathcal{F}, \quad \forall \mathcal{F} \in \mathfrak{F}\} .$$

É claro que $\mathcal{C} \subset \mathcal{F}^*$. Resta apenas mostrar que $\mathcal{F}^*$ é também uma σ-álgebra. Iremos demonstrar que $\mathcal{F}^*$ é fechado para uniões numeráveis, deixando a verificação das restantes condições ao cuidado do leitor. Sejam $A_i \in \mathcal{F}^*$, $i = 1, 2, \ldots$. Por definição, temos que $A_i \in \mathcal{F}$ para todo o $\mathcal{F} \in \mathfrak{F}$ e $i = 1, 2, \ldots$. Segue do facto de $\mathcal{F}$ ser uma σ-álgebra que $\bigcup_i A_i \in \mathcal{F}$ para todo o $\mathcal{F} \in \mathfrak{F}$. Logo $\bigcup_i A_i \in \mathcal{F}^*$ como pretendíamos mostrar. $\square$

A σ-álgebra $\mathcal{F}^*$ obtida na Proposição 1.2 é designada por **σ-álgebra gerada por $\mathcal{C}$** e denota-se por $\sigma(\mathcal{C})$. Num certo sentido, $\sigma(\mathcal{C})$ é a "menor" σ-álgebra que contém a colecção $\mathcal{C}$. Uma aplicação relevante deste resultado é a seguinte. Seja X um espaço métrico ou mais geral um espaço topológico. A σ-álgebra $\mathcal{B}(X) = \sigma(\text{abertos de X})$ gerada pelos abertos de X é designada por **σ-álgebra de Borel** e os seus conjuntos mensuráveis designados por **Borelianos**. Note-se que $\mathcal{B}(X)$ contêm também os fechados de X, uniões numeráveis de abertos de X, intersecções numeráveis de abertos de X, etc.

Exercício 2. Mostre que a intersecção numerável de σ-álgebras é uma σ-álgebra. Será a união numerável de σ-álgebras também uma σ-álgebra?

Exercício 3. Seja $A \subset X$. Determine $\sigma(\{A\})$.

Exercício 4. Seja X um espaço topológico. Mostre que $\mathcal{B}(X) = \sigma$(fechados de X).

Exercício 5. Mostre que se A^c é numerável então A é Boreliano.

1.2 Medidas

Definição 1.2. Dado um espaço mensurável $(X, \mathcal{F})$, uma função $\mu : \mathcal{F} \to [0, \infty]$ é uma **medida** sse

1. $\mu(\emptyset) = 0$.
2. para qualquer sucessão de subconjuntos $A_i \in \mathcal{F}$, $i = 1, 2, \ldots$, disjuntos dois a dois $A_i \cap A_j = \emptyset$, $i \neq j$ têm-se

$$\mu\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} \mu(A_i)\,,$$

A segunda condição é conhecida por **σ-aditividade**. Ao tripleto $(X, \mathcal{F}, \mu)$ designamos por **espaço de medida**

Exemplo 1.3. 1. Seja X um conjunto qualquer, $\mathcal{F} = \mathcal{P}(X)$. Para qualquer $A \in \mathcal{P}(X)$ define-se a **medida de contagem** como

$$\mu(A) = \begin{cases} \#A & \text{se} \quad A \text{ é finito,} \\ \infty & \text{se} \quad A \text{ é infinito.} \end{cases}$$

2. Considere-se o espaço mensurável no exemplo anterior. Dado $x \in X$ e $A \in \mathcal{F}$ define-se

$$\delta_x(A) = \begin{cases} 1 & \text{se} \quad x \in A, \\ 0 & \text{se} \quad x \notin A. \end{cases}$$

Esta medida designa-se por **medida de Dirac** ou **massa unitária concentrada** em x.

Proposição 1.4 (Propriedades da medida)**.** *Seja $(X, \mathcal{F}, \mu)$ um espaço de medida. Então*

1. (aditividade finita) *Sejam $A_1, \ldots, A_n \in \mathcal{F}$, disjuntos dois a dois, então $\mu(A_1 \cup \cdots \cup A_n) = \mu(A_1) + \cdots + \mu(A_n)$.*

2. **(monotonia)** *Sejam* $A, B \in \mathcal{F}$ *tal que* $A \subset B$ *então* $\mu(A) \leq \mu(B)$.

3. **(continuidade inferior)** *Sejam* $A_i \in \mathcal{F}$, $i = 1, 2, \ldots$ *tal que*

$$A_1 \subset A_2 \subset A_3 \subset \cdots ,$$

então

$$\mu(\bigcup_{i=1}^{\infty} A_i) = \lim_{n\to\infty} \mu(A_n) .$$

4. **(continuidade superior)** *Sejam* $A_i \in \mathcal{F}$ *tal que* $\mu(A_1) < \infty$ *e*

$$A_1 \supset A_2 \supset A_3 \supset \cdots ,$$

então

$$\mu(\bigcap_{i=1}^{\infty} A_i) = \lim_{n\to\infty} \mu(A_n) .$$

Demonstração.

1. Sejam $B_i = A_i$ para $i = 1, \ldots, n$ e $B_i = \emptyset$ para $i > n$. É claro que $\mu(B_i) = \mu(A_i)$ para $i = 1, \ldots, n$ e $\mu(B_i) = 0$ para $i > n$. Pela σ-aditividade obtemos que

$$\mu(A_1 \cup \cdots \cup A_n) = \mu(\bigcup_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty} \mu(B_i) = \mu(A_1) + \cdots + \mu(A_n) .$$

2. Uma vez que $B = A \cup (B \setminus A)$ e $A \cap (B \setminus A) = \emptyset$, segue da propriedade anterior que $\mu(B) = \mu(A) + \mu(B \setminus A)$. Logo $\mu(A) \leq \mu(B)$.

3. Sejam $B_1 = A_1$ e $B_i = A_i \setminus A_{i-1}$ para $i = 2, 3, \ldots$. Note-se que os conjuntos B_i são disjuntos dois a dois. Mais, $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i$. Logo, pela σ-aditividade obtemos que

$$\mu(\bigcup_{i=1}^{\infty} A_i) = \mu(\bigcup_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty} \mu(B_i) .$$

Por outro lado, $\mu(B_i) = \mu(A_i) - \mu(A_{i-1})$ (mostre esta igualdade). Logo

$$\sum_{i=1}^{n} \mu(B_i) = \mu(A_n) .$$

Fazendo $n \to \infty$ têm-se $\sum_{i=1}^{\infty} \mu(B_i) = \lim_{n\to\infty} A_n$, de onde segue o resultado.

4. Deixa-se como exercício.

$\square$

Exercício 6. Seja $(X, \mathcal{F}, \mu)$ um espaço de medida. Mostre que para qualquer $A, B \in \mathcal{F}$ tal que $A \subset B$ tem-se $\mu(B \setminus A) = \mu(B) - \mu(A)$.

Exercício 7. (**σ-subaditividade**) Sejam $(X, \mathcal{F}, \mu)$ um espaço de medida e $(A_n)_{n\in\mathbb{N}}$ uma sucessão de elementos de $\mathcal{F}$. Mostre que

$$\mu(\bigcup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} \mu(A_n)\,.$$

Exercício 8. Sejam μ_1 e μ_2 duas medidas do espaço mensurável $(X, \mathcal{F})$. Mostre que $\mu_1 + \mu_2$ definida por $(\mu_1 + \mu_2)(A) = \mu_1(A) + \mu_2(A)$ para todo o $A \in \mathcal{F}$ é uma medida de $(X, \mathcal{F})$.

Exercício 9. Demonstre a propriedade (4) da proposição anterior, isto é, a continuidade superior da medida.

Exercício 10. Mostre que a condição $\mu(A_1) < \infty$ não pode ser retirada do enunciado da propriedade (4). (Dica: construa um contra-exemplo considerando um conjunto infinito X e a medida de contagem μ.)

1.3 Conjuntos de medida nula

Seja $(X, \mathcal{F}, \mu)$ um espaço de medida. Um subconjunto $A \subset X$ é um conjunto de **medida nula** se existir $B \in \mathcal{F}$ tal que $A \subset B$ e $\mu(B) = 0$. Do ponto de vista da teoria da medida, os conjuntos de medida nula são desprezáveis. No entanto, é conveniente que todos os conjuntos de medida nula sejam também conjuntos mensuráveis, isto é, elementos de $\mathcal{F}$. Um espaço de medida com esta propriedade diz-se **completo**.

Proposição 1.5. *A união numerável de conjuntos de medida nula é um conjunto de medida nula.*

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

A **diferença simétrica** entre dois conjuntos A e B é o conjunto

$$A \bigtriangleup B = (A \setminus B) \cup (B \setminus A) = (A \cup B) \setminus (A \cap B)\,.$$

Se $A \bigtriangleup B$ é um conjunto de medida nula então A e B são **μ-equivalentes**. Se adicionalmente, A e B são mensuráveis então $\mu(A) = \mu(B) = \mu(A \cap B) = \mu(A \cup B)$ (porquê?).

Exercício 11. Seja $(X, \mathcal{F}, \mu)$ um espaço de medida e $A, B \subset X$ dois conjuntos. Mostre que são equivalentes

1. A é μ-equivalente a B.
2. existe um conjunto $N \subset X$ de medida nula tal que $A \cap N^{\mathrm{c}} = B \cap N^{\mathrm{c}}$.
3. existe $C \in \mathcal{F}$, $\mu(C) = 0$ tal que $B \setminus C \subset A \subset B \cup C$.

Seja $\overline{\mathcal{F}}$ a seguinte colecção de subconjuntos de X,

$$\overline{\mathcal{F}} = \{A \in \mathcal{P}(X) \colon A \text{ é } \mu\text{-equivalente a algum } B \in \mathcal{F}\}$$

e $\overline{\mu} : \overline{\mathcal{F}} \to [0, \infty]$ a função definida por $\overline{\mu}(A) = \mu(B)$ para algum $B \in \mathcal{F}$ μ-equivalente a A. É claro que $\mathcal{F} \subset \overline{\mathcal{F}}$ (porquê?).

Proposição 1.6. $(X, \overline{\mathcal{F}}, \overline{\mu})$ *é um espaço de medida completo.*

Demonstração. Em primeiro lugar verifica-se que $\overline{\mathcal{F}}$ é uma σ-álgebra. Demonstramos o terceiro axioma, deixando como exercício a verificação dos restantes axiomas.

Sejam $A_i \in \overline{\mathcal{F}}$, $i = 1, 2, \ldots$. Por definição existem $B_i \in \mathcal{F}$ μ-equivalentes a A_i. Logo, pelo Exercício 11 existem $N_i \subset X$ conjuntos de medida nula tal que $A_i \cap N_i^{\mathrm{c}} = B_i \cap N_i^{\mathrm{c}}$. Segue que

$$\left(\bigcup_{i=1}^{\infty} A_i\right) \triangle \left(\bigcup_{i=1}^{\infty} B_i\right) \subset \bigcup_{i=1}^{\infty} N_i \,.$$

Como $\mathcal{F}$ é uma σ-álgebra temos que $\bigcup_{i=1}^{\infty} B_i \in \mathcal{F}$. Por outro lado, do Exercício 1.5 temos que $\bigcup_{i=1}^{\infty} N_i$ é um conjunto de medida nula. Logo $\bigcup_{i=1}^{\infty} A_i$ é μ-equivalente a $\bigcup_{i=1}^{\infty} B_i \in \mathcal{F}$, ou seja, $\bigcup_{i=1}^{\infty} A_i \in \overline{\mathcal{F}}$.

De seguida, mostra-se que $\overline{\mu}$ é uma medida. É claro que $\overline{\mu}(\emptyset) = \mu(\emptyset) = 0$. Para mostrar a σ-aditividade, sejam $A_i \in \overline{\mathcal{F}}$, $i = 1, 2, \ldots$ conjuntos disjuntos dois a dois. Como anteriormente, existem conjuntos $B_i \in \mathcal{F}$ μ-equivalentes a A_i tal que $\bigcup_{i=1}^{\infty} A_i$ é μ-equivalente a $\bigcup_{i=1}^{\infty} B_i$. Note-se que os conjuntos B_i são disjuntos dois a dois. Então

$$\overline{\mu}(\bigcup_{i=1}^{\infty} A_i) = \mu(\bigcup_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty} \mu(B_i) = \sum_{i=1}^{\infty} \overline{\mu}(A_i) \,.$$

Resta provar que $(X, \overline{\mathcal{F}}, \overline{\mu})$ é completo, ou seja, se A é um conjunto de medida nula então $A \in \overline{\mathcal{F}}$. Mas por definição de conjunto de medida nula, existe um $B \in \mathcal{F}$ tal que $A \subset B$ e $\mu(B) = 0$. Logo $A \triangle B$ é um conjunto de medida nula. Portanto A é μ-equivalente a B, isto é, $A \in \overline{\mathcal{F}}$. □

A σ-álgebra $\overline{\mathcal{F}}$ é designada por **μ-completação** da σ-álgebra $\mathcal{F}$.

1.4 Medidas de Lebesgue-Stieltjes

Nesta secção iremos estudar uma classe especial de medidas no espaço mensurável $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ onde $\mathcal{B}(\mathbb{R})$ é a σ-álgebra de Borel, ou seja, a menor σ-álgebra de partes de $\mathbb{R}$ que contém os abertos de $\mathbb{R}$. Estas medidas desempenham um papel essencial em teoria de probabilidades, como veremos mais à frente.

Definição 1.3. Uma função $F : \mathbb{R} \to \mathbb{R}$ é uma **função de distribuição** sse

1. F é não decrescente, isto é, se $x < y$ então $F(x) \leq F(y)$.
2. F é contínua à direita, $\lim_{x \to x_0^+} F(x) = F(x_0)$.

Uma medida μ definida em $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ diz-se **localmente finita** sse $\mu([a, b]) < \infty$ para quaisquer $a, b \in \mathbb{R}$.

Proposição 1.7. *Seja μ uma medida em $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ localmente finita e dado $a \in \mathbb{R}$ seja $F : \mathbb{R} \to \mathbb{R}$ a função definida por*

$$F(x) = \begin{cases} \mu(]a, x]) & \textit{se } x \geq a \\ -\mu(]x, a]) & \textit{se } x < a \end{cases}$$

Então F é uma função de distribuição.

Demonstração. De facto, F é não decrescente uma vez que para $x < a \leq y$ (os restantes casos são deixados como exercício) tem-se

$$F(y) - F(x) = \mu(]a, y]) + \mu(]x, a]) = \mu(]x, y]) \geq 0\,.$$

Quando à continuidade à direita para $x_0 \geq a$ (o caso $x_0 < a$ é deixado como exercício), tome-se uma qualquer sucessão decrescente x_n e convergente para x_0. Pela continuidade superior da medida μ (ver Proposição 1.4) e do facto de $\mu(]a, x_n]) < \infty$ segue que

$$F(x_0) = \mu(]a, x_0]) = \mu\Big(\bigcap_{n=1}^{\infty}]a, x_n]\Big) = \lim_{n \to \infty} \mu(]a, x_n]) = \lim_{n \to \infty} F(x_n)\,.$$

$\square$

Vimos na proposição anterior que qualquer medida localmente finita define uma função de distribuição. Agora consideremos o problema inverso, ou seja, dada uma função de distribuição F será possível associar uma medida m_F? O seguinte teorema responde positivamente a esta questão.

Teorema 1.8. *Dada uma função de distribuição F existe uma medida m_F em $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ tal que*

$$m_F(]a,b]) = F(b) - F(a), \quad \forall a, b \in \mathbb{R}.$$

A construção da medida m_F é reminiscente da construção da medida de Lesbegue em $\mathbb{R}$. A sua demonstração pode ser vista em [1].

Observação 1.9. Note-se que os intervalos considerados são abertos à esquerda e fechados à direita, garantindo assim que m_F é aditiva, isto é, $m_F(]a,b]) = m_F(]a,c]) + m_F(]c,b])$ para todo $a < c < b$ e por outro lado garante que m_F é contínua superiormente, ou seja, dada um sucessão decrescente $b_n \searrow b$ tem-se

$$\begin{aligned} m_F(\bigcap_{n=1}^{\infty}]a,b_n]) = m_F(]a,b]) &= F(b) - F(a) \\ &= \lim_{n\to\infty} F(b_n) - F(a) = \lim_{n\to\infty} m_F(]a,b_n]). \end{aligned}$$

Observação 1.10. Note-se que nem sempre o conjunto unitário $\{a\}$ tem medida nula. De facto, usando a continuidade superior da medida m_F temos que

$$m_F(\{a\}) = m_F(\bigcap_{n=1}^{\infty}]a-1/n,a]) = \lim_{n\to\infty} m_F(]a-1/n,a]) = F(a) - F(a^-).$$

Portanto, se a é um ponto de descontinuidade de F então $m_F(\{a\})$ é igual ao "salto" de F no ponto a.

A medida m_F designa-se por **medida de Borel-Stieltjes**. Denote-se por $\mathcal{M}_F$ a m_F-completação de $\mathcal{B}(\mathbb{R})$. Então $(\mathbb{R}, \mathcal{M}_F, m_F)$ é um espaço de medida completo e m_F designa-se por **medida de Lebesgue-Stieltjes**.

Exemplo 1.11. Sejam $a \in \mathbb{R}$ e $F : \mathbb{R} \to \mathbb{R}$ a função de distribuição definida por $F(x) = 1$ se $x \geq a$ e $F(x) = 0$ se $x < a$. A medida de Lebesgue-Stieltjes $m_F = \delta_a$ onde δ_a é a medida de Dirac, ou seja, para todo $A \subseteq \mathbb{R}$ tem-se

$$\delta_a(A) = \begin{cases} 1 & \text{se} \quad a \in A, \\ 0 & \text{se} \quad a \notin A. \end{cases}$$

Não é difícil verificar que $\mathcal{M}_F = \mathcal{P}(\mathbb{R})$. De facto, dado $A \in \mathcal{P}(\mathbb{R})$ tal que $a \in A$ (o caso contrário deixa-se como exercício) pode-se escrever $A = \{a\} \cup (A \setminus \{a\})$. Note-se que $\{a\} \in \mathcal{B}(\mathbb{R}) \subset \mathcal{M}_F$. Por outro lado,

$A \setminus \{a\} \subset]-\infty, a[\cup]a, \infty[\in \mathcal{B}(\mathbb{R})$ onde $\delta_a(]-\infty, a[\cup]a, \infty[) = 0$. Logo $A \setminus \{a\}$ é um conjunto de medida nula e portanto A é m_F-equivalente a $\{a\}$. Como $\mathcal{M}_F$ é a m_F-completação de $\mathcal{B}(\mathbb{R})$ tem-se $A \in \mathcal{M}_F$.

Em geral dada qualquer distribuição F constante por troços e com pontos de descontinuidade $\{a_1, \ldots, a_n\}$, a medida m_F é igual à soma

$$m_F = \sum_{i=1}^{n} \big[F(a_i) - F(a_i^-)\big]\big)\delta_{a_i}$$

e todo o subconjunto de $\mathbb{R}$ é Lebesgue-Stieltjes mensurável, ou seja, $\mathcal{M}_F = \mathcal{P}(\mathbb{R})$.

Exemplo 1.12. No caso em que a função de distribuição é $F(x) = x$, $\forall x \in \mathbb{R}$ então $(\mathbb{R}, \mathcal{M}_F, m_F)$ escreve-se apenas $(\mathbb{R}, \mathcal{M}, m)$ e designa-se m por **medida de Lebesgue**. Quando restringida aos Borelianos designa-se por **medida de Borel**. A medida de Lebesgue satisfaz algumas propriedades adicionais, em particular:

1. Dado um conjunto E Lebesgue mensurável, ou seja $E \in \mathcal{M}$, então para qualquer $x \in \mathbb{R}$ o conjunto $x + E = \{x + y : y \in E\}$ é Lebesgue mensurável e $m(x + E) = m(E)$.
2. Dado um intervalo $I \subset \mathbb{R}$ então $m(I) = \mathrm{Comp}(I)$.

Segundo o Teorema 1.8 tem-se que $\mathcal{B}(\mathbb{R}) \subseteq \mathcal{M} \subseteq \mathcal{P}(\mathbb{R})$. De facto é possível provar que

$$\mathcal{B}(\mathbb{R}) \subsetneq \mathcal{M} \subsetneq \mathcal{P}(\mathbb{R})$$

Na proposição que se segue, provamos que a segunda inclusão é de facto estrita. Ou seja, que existem subconjuntos de $\mathbb{R}$ não mensuráveis no sentido de Lebesgue. Para provar a primeira inclusão veja-se [1].

Proposição 1.13. *Existe um conjunto $V \subset [0, 1]$ denominado por* ***conjunto de Vitali*** *tal que V não é mensurável no sentido de Lebesgue, ou seja, $V \notin \mathcal{M}$.*

Demonstração. Define-se a seguinte relação de equivalência no intervalo $[0, 1]$. Dois números $x, y \in [0, 1]$ dizem-se equivalentes e escreve-se $x \sim y$ sse $y - x$ for um número racional. Esta relação de equivalência particiona $[0, 1]$ em classes de equivalência disjuntas A_α do tipo $a_\alpha + \mathbb{Q}$ para algum $a_\alpha \in [0, 1]$. É claro que existem um número não numerável de classes de equivalência. Seja $V \subset [0, 1]$ o conjunto formado por um e um só elemento de cada classe de equivalência A_α (porque se pode fazer esta escolha?). Seja q_n uma enumeração do racionais em $[-1, 1]$ e seja $V_n = q_n + V$ uma translação do conjunto V.

Os conjuntos V_n são disjuntos dois a dois. De facto, se $x \in V_n \cap V_m$ então $x = a_\alpha + q_n = a_\beta + q_m$ para algum $a_\alpha, a_\beta \in V$. Logo, $a_\alpha - a_\beta =$

$q_m - q_n \in \mathbb{Q}$ o que implica que $a_\alpha \sim a_\beta$. Como V contém um e um só elemento de cada classe A_α segue que $a_\alpha = a_\beta$, logo $n = m$.

Portanto, se V for mensurável então V_n também é mensurável e para além disso $m(V_n) = m(V)$ (porquê?). Por outro lado,

$$[0,1] \subset \bigcup_{n=1}^{\infty} V_n \subset [-1,2]\,.$$

Logo pela σ-aditividade e monotonia da medida m tem-se

$$1 = m([0,1]) \leq m(\bigcup_{n=1}^{\infty} V_n) = \sum_{n=1}^{\infty} m(V_n) = m(V) + m(V) + \cdots \leq 3\,.$$

A desigualdade anterior é claramente impossível visto que se $m(V) = 0$ então $1 \leq 0 \leq 3$. Por outro lado, se $m(V) > 0$ então $1 \leq \infty \leq 3$. Logo V não é mensurável. $\square$

Exercício 12. Mostre que se $A \subset \mathbb{R}$ é um conjunto finito ou numerável então A tem medida de Lebesgue nula.

Exercício 13. Seja F uma função de distribuição continua. Mostre que se $A \subset \mathbb{R}$ é um conjunto finito ou numerável então $m_F(A) = 0$.

Exercício 14. Dê um exemplo de uma função de distribuição F tal que $m_F(\{1\}) = 1$.

Exercício 15. Considere a função,

$$F(x) = \begin{cases} 0 & \text{se} \quad x < 0, \\ x^2 & \text{se} \quad x \geq 0. \end{cases}$$

1. Prove que F é uma função de distribuição.
2. Seja μ a medida de Borel-Stieltjes associada a F. Calcule $\mu(]3,9])$.

Exercício 16. Considere a medida $\mu = 3\delta_2 + 2\delta_3$ em $\mathbb{R}$. Determine a função de distribuição F tal que $\mu = m_F$.

Exercício 17. Mostre que uma função de distribuição tem no máximo um conjunto numerável de pontos de descontinuidade.

2 Integração

Na secção anterior definimos medidas que tomam valores na semi-recta estendida, ou seja, $[0,\infty]$. A introdução do infinito é necessária porque existem conjuntos de medida infinita, como é o caso da medida de Lebesgue de $\mathbb{R}$. Portanto, no desenvolvimento da teoria de integração temos de ter algum cuidado com a aritmética em $[0, \infty]$. Assim, definimos para a adição

$$a + \infty = \infty + a = \infty, \quad \forall\, a \in [0, \infty] \ ,$$

enquanto que para a multiplicação

$$a \times \infty = \infty \times a = \begin{cases} \infty & \text{se } a \in \,]0, \infty] \\ 0 & \text{se } a = 0 \end{cases}$$

Com estas convenções estendemos as leis usuais ao conjunto $[0, \infty]$, com excepção da lei do corte para o infinito, como demonstra o exemplo: $\infty = 0 + \infty = 1 + \infty$ e no entanto $0 \neq 1$.

2.1 Funções mensuráveis

Seja (X, τ) um **espaço topológico**, isto é, uma **topologia** $\tau \subset \mathcal{P}(X)$ cujos elementos, designados por **conjuntos abertos**, satisfazem as seguintes condições:

1. $\emptyset, X \in \tau$.
2. se $A_1, \ldots, A_n \in \tau$ então $V_1 \cap \cdots \cap V_n \in \tau$
3. se $\{A_\alpha\} \subset \tau$ então $\bigcup_\alpha A_\alpha \in \tau$.

Ao complementar dos conjuntos abertos designamos por **conjuntos fechados**, ou seja, se $A \subset X$ é aberto então A^c é fechado. Uma **vizinhança** de $x \in X$ é um qualquer conjunto $V \subset X$ tal que existe um aberto $A \subset V$ que contenha x.

Exemplo 2.1. Verifica-se facilmente que $\mathcal{P}(X)$ e $\{\emptyset, X\}$ são topologias de X, a topologia discreta e indiscreta, respectivamente.

Exemplo 2.2. Seja $d : X \times X \to [0, \infty[$ uma **distância** de X, ou seja, uma função que satisfaz

1. $d(x, y) = 0$ sse $x = y$.
2. $d(x, y) = d(y, x)$ para todo $x, y \in X$.
3. $d(x, y) \leq d(x, z) + d(z, y)$ para todo $x, y, z \in X$.

A terceira propriedade é conhecida por **desigualdade triangular**. O par (X, d) designa-se por **espaço métrico**. Dado $r > 0$, o conjunto $B(x, r) = \{y \in X : d(x, y) < r\}$ designa-se por **bola aberta** de centro x e raio r.

Um espaço métrico é um caso particular de um espaço topológico onde a topologia é formada por uniões arbitrárias (podem ser não numeráveis) de bolas abertas. Como exemplo, tome-se o espaço métrico $(\mathbb{R}^n, d)$ onde

$$d(x, y) = \|x - y\| = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2}\,.$$

A topologia em $\mathbb{R}^n$ induzida pela distância anterior é conhecida por **topologia usual** ou **Euclidiana**.

Observação 2.3. Regra geral, nestas notas a topologia subjacente a $\mathbb{R}^n$ será sempre a topologia usual.

Exemplo 2.4. A recta estendida $\overline{\mathbb{R}} = [-\infty, \infty]$ é um espaço topológico com a topologia formada por uniões arbitrárias de intervalos do tipo $[-\infty, a[$, $]a, \infty]$ e $]a, b[$.

Sejam X e Y dois espaços topológicos. Uma função $f : X \to Y$ é **contínua** sse $f^{-1}(A)$ é aberto de X para qualquer aberto A de Y.

Definição 2.1. Sejam $(X, \mathcal{F})$ um espaço mensurável e (Y, τ) um espaço topológico. Uma função $f : X \to Y$ é **mensurável** sse $f^{-1}(A) \in \mathcal{F}$ para todo o aberto $A \in \tau$ de Y.

Exemplo 2.5. Seja $E \subset X$ e tome-se a **função característica** de E, $\chi_E : X \to \mathbb{R}$ definida por

$$\chi_E(x) = \begin{cases} 1 & \text{se } x \in E \\ 0 & \text{se } x \notin E \end{cases}$$

A função χ_E é mensurável sse E é mensurável. De facto, dado um aberto de $A \subset \mathbb{R}$ tem-se: se $0 \notin A$ e $1 \notin A$ então $\chi_E^{-1}(A) = \emptyset$; se $0 \in A$ e $1 \notin A$ então $\chi_E^{-1}(A) = E^c$; se $0 \notin A$ e $1 \in A$ então $\chi_E^{-1}(A) = E$; se $0 \in A$ e $1 \in A$ então $\chi_E^{-1}(A) = X$.

De seguida enunciaremos duas proposições que permitem estabelecer a mensurabilidade de uma grande classe de funções. Antes de continuar convém relembrar as propriedades da imagem inversa, isto é, que se $f : X \to Y$ é uma função e $A, B \subset Y$ então

$$f^{-1}(A \cup B) = f^{-1}(A) \cup f^{-1}(B) \quad \text{e} \quad f^{-1}(Y \setminus A) = X \setminus f^{-1}(A)\,.$$

Note-se que a primeira propriedade por ser estendida a uniões numeráveis usando indução.

Proposição 2.6. *Sejam Y e Z espaços topológicos e X um espaço mensurável. Se $f : X \to Y$ é uma função mensurável e $g : Y \to Z$ é uma função contínua então $g \circ f : X \to Z$ é uma função mensurável.*

Exercício 18. Demonstre a proposição anterior. (Dica: use as definições de continuidade e mensurabilidade)

Proposição 2.7. *Sejam X um espaço mensurável, Y um espaço topológico, $u, v : X \to \mathbb{R}$ duas funções mensuráveis e $\phi : \mathbb{R}^2 \to Y$ uma função continua. Então $h : X \to Y$ definida por $h(x) = \phi(u(x), v(x))$ é uma função mensurável.*

Demonstração. É suficiente demonstrar que a função $f : X \to \mathbb{R}^2$ definida por $f(x) = (u(x), v(x))$ é mensurável. De facto, como $h = \phi \circ f$ e ϕ é contínua, basta aplicar a Proposição 2.6.

Resta então provar que f é mensurável, ou seja, para qualquer aberto $A \subseteq \mathbb{R}^2$ tem-se que $f^{-1}(A)$ é mensurável. Procedemos em dois passos:

1. Considere-se primeiro abertos do tipo $A = I \times J$ (rectângulos) onde $I, J \subseteq \mathbb{R}$ são dois intervalos abertos. Então, segue da definição de f que

$$f^{-1}(I \times J) = u^{-1}(I) \cap v^{-1}(J)\,.$$

   Como u e v são mensuráveis, logo $u^{-1}(I)$ e $v^{-1}(J)$ são conjuntos mensuráveis de X, tal como sua intersecção. Logo $f^{-1}(I \times J)$ é mensurável.

2. Como todo o aberto $A \subseteq \mathbb{R}^2$ pode ser escrito como uma união numerável de rectângulos abertos $I_n \times J_n$, $n = 1, 2, \ldots$ (porquê?) então

$$f^{-1}(A) = f^{-1}\Big(\bigcup_{n=1}^{\infty} I_n \times J_n\Big) = \bigcup_{n=1}^{\infty} f^{-1}(I_n \times J_n) = \bigcup_{n=1}^{\infty} u^{-1}(I_n) \cap v^{-1}(J_n)\,.$$

   Como $u^{-1}(I_n) \cap v^{-1}(J_n)$ é um conjunto mensurável (pelo argumento do ponto 1.) e a união numerável de conjuntos mensuráveis é mensurável, segue que $f^{-1}(A)$ é mensurável.

□

Exercício 19. Mostre, usando a proposição anterior, que se $f, g : X \to \mathbb{R}$ são duas funções mensuráveis então

1. $f + g$ é mensurável.
2. $f \cdot g$ é mensurável.

Observação 2.8. Do exercício anterior conclui-se que o conjunto das funções mensuráveis $f : X \to \mathbb{R}$ é um espaço linear.

Enunciamos mais dois resultados, que fornecem condições equivalentes de mensurabilidade de funções.

Proposição 2.9. *Sejam X um espaço mensurável, Y um espaço topológico e $f : X \to Y$. São equivalentes:*

1. *f é uma função mensurável.*
2. *$f^{-1}(A)$ é mensurável qualquer que seja o Boreliano A de Y.*

Demonstração. Uma vez que os abertos de Y são também Borelianos de Y então é claro que 2. implica 1. Mostremos então que 1. implica 2. Considere-se a colecção de conjuntos

$$\mathcal{C} = \left\{ A \subset Y : f^{-1}(A) \text{ é um conjunto mensurável} \right\} .$$

Quer-se provar que $\mathcal{B}(Y) \subseteq \mathcal{C}$. Uma vez que f é mensurável então $\mathcal{C}$ contém todos os abertos de Y. Por outro lado, usando as propriedades da imagem inversa mostra-se sem dificuldade que $\mathcal{C}$ é uma σ-álgebra de Y. Como a σ-álgebra de Borel $\mathcal{B}(Y)$ é a "menor" σ-álgebra que contém $\mathcal{C}$ segue que $\mathcal{B}(Y) \subseteq \mathcal{C}$. □

Proposição 2.10. *Seja X um espaço mensurável e $f : X \to \overline{\mathbb{R}}$. São equivalentes:*

1. *f é uma função mensurável.*
2. *$f^{-1}(]a, \infty])$ é mensurável qualquer que seja $a \in \mathbb{R}$.*

Demonstração. Como $]a, \infty]$ é aberto de $\overline{\mathbb{R}}$ segue que 1. implica 2. Para demonstrar a implicação inversa, considere-se (como na proposição anterior) o conjunto

$$\mathcal{C} = \left\{ A \subset \overline{\mathbb{R}} : f^{-1}(A) \text{ é um conjunto mensurável} \right\} .$$

Quer-se mostrar que $\mathcal{C}$ contém todos os abertos de $\overline{\mathbb{R}}$. Por hipótese sabe-se que $]a, \infty] \in \mathcal{C}$, ou seja, $f^{-1}(]a, \infty])$ é mensurável. Por outro lado

- $[-\infty, b] \in \mathcal{C}$, $\forall\, b \in \mathbb{R}$. De facto,

  $$f^{-1}([-\infty, b]) = X \setminus f^{-1}(]b, \infty]) .$$

  Como $f^{-1}(]b, \infty])$ é mensurável temos que $X \setminus f^{-1}(]b, \infty])$ é mensurável.

- $]a, b] \in \mathcal{C}$, $\forall\, a, b \in \mathbb{R}$. De facto,

  $$f^{-1}(]a, b]) = f^{-1}([-\infty, b] \cap]a, \infty]) = f^{-1}([-\infty, b]) \cap f^{-1}(]a, \infty]) .$$

  Como $f^{-1}([-\infty, b])$ e $f^{-1}(]a, \infty])$ são mensuráveis, segue que $f^{-1}(]a, b])$ é mensurável.

- $]a,b[\in \mathcal{C}$, $\forall\, a,b \in \mathbb{R}$. De facto, tome-se uma sucessão crescente b_n e convergente para b. Então

$$f^{-1}(]a,b[) = f^{-1}(\bigcup_{n=1}^{\infty}]a,b_n]) = \bigcup_{n=1}^{\infty} f^{-1}(]a,b_n])\,.$$

  Como $f^{-1}(]a,b_n])$ é mensurável para todo $n = 1,2,\ldots$ segue que $f^{-1}(]a,b[)$ é mensurável.

Finalmente, como todo o aberto $A \subset \overline{\mathbb{R}}$ pode ser escrito como uma união numerável de intervalos abertos do tipo $]a_n,b_n[$ com $a_n < b_n$ (porquê?) temos que

$$f^{-1}(A) = f^{-1}(\bigcup_{n=1}^{\infty}]a_n,b_n[) = \bigcup_{n=1}^{\infty} f^{-1}(]a_n,b_n[)\,,$$

de onde segue que A é mensurável. Logo $\mathcal{C}$ contém todos os abertos de $\overline{\mathbb{R}}$. $\square$

Exercício 20. Seja X um espaço topológico e $(X,\mathcal{B})$ o espaço mensurável onde $\mathcal{B}$ é a σ-álgebra de Borel. Mostre que se $f : X \to \mathbb{R}$ é uma função contínua então f é mensurável.

Exercício 21. Dado $c \in \mathbb{R}$, mostre que a função constante $f(x) = c$ para todo $x \in \mathbb{R}$ é uma função mensurável.

Exercício 22. Seja $f : X \to \mathbb{R}$ uma função e defina-se

$$f^+(x) = \begin{cases} f(x) & \text{se } f(x) > 0 \\ 0 & \text{se } f(x) \leq 0 \end{cases} \quad \text{e} \quad f^-(x) = \begin{cases} 0 & \text{se } f(x) > 0 \\ -f(x) & \text{se } f(x) \leq 0 \end{cases}$$

a **parte positiva** e **negativa** de f, respectivamente. Mostre que f é mensurável sse f^+ e f^- são mensuráveis. (Dica: $f^+ = f \cdot \chi_E$ onde $E = \{x \in X : f(x) > 0\}$)

Exercício 23. Mostre que se $f : X \to \mathbb{R}$ uma função mensurável então $|f|$ também é mensurável.

Exercício 24. Seja $f : X \to \overline{\mathbb{R}}$ uma função mensurável. Dado $a \in \mathbb{R}$, mostre que o **conjunto de nível** $\{x \in X : f(x) = a\}$ é mensurável.

Exercício 25. Considere um espaço mensurável $(X,\mathcal{F})$ e um espaço topológico Y. Seja $f : X \to Y$ uma função mensurável. Prove que a colecção de conjuntos $\mathcal{C} = \left\{A \subset Y : f^{-1}(A) \in \mathcal{F}\right\}$ é uma σ-álgebra.

De seguida veremos que limites de funções mensuráveis resultam em funções mensuráveis, ou seja, a passagem para o limite não destrói a mensurabilidade. Convém relembrar as seguintes definições. Dada uma sucessão $a_n \in \overline{\mathbb{R}}$, $n = 1, 2, \ldots$, o **limite superior** de a_n é o número

$$\limsup_{n\to\infty} a_n = \inf_{k\geq 1} \sup_{n\geq k} a_n \,.$$

Ou seja, o maior dos pontos de acumulação da sucessão a_n. Analogamente, o **limite inferior** de a_n é o número

$$\liminf_{n\to\infty} a_n = \sup_{k\geq 1} \inf_{n\geq k} a_n \,.$$

Não é difícil verificar que $\limsup_{n\to\infty} a_n = -\liminf_{n\to\infty}(-a_n)$. Note-se que uma sucessão a_n é convergente sse o seu limite superior e inferior coincidirem.

Considere-se agora uma sucessão de funções $f_n : X \to \overline{\mathbb{R}}$, $n = 1, 2, \ldots$. Para cada $x \in X$, tem-se que $f_n(x)$ é uma sucessão em $\overline{\mathbb{R}}$. Logo podemos definir as funções $\sup f_n$ e $\limsup f_n$ da seguinte maneira

$$(\sup f_n)(x) = \sup_{n\geq 1} f_n(x) \quad \text{e} \quad (\limsup f_n)(x) = \limsup_{n\to\infty} f_n(x)\,.$$

Define-se de forma análoga as funções $\inf f_n$ e $\liminf f_n$. Diz-se que f_n **converge pontualmente** sse $\liminf f_n = \limsup f_n$.

Proposição 2.11. *Se $f_n : X \to \overline{\mathbb{R}}$ é uma sucessão de funções mensuráveis então $\inf f_n$, $\sup f_n$, $\liminf f_n$ e $\limsup f_n$ são funções mensuráveis.*

Demonstração. Não é difícil verificar a seguinte igualdade

$$\left\{x \in X : \sup_{n\geq k} f_n(x) > a\right\} = \bigcup_{n=k}^{\infty} \{x \in X : f_n(x) > a\}\,.$$

Como f_n é uma função mensurável, segue da Proposição 2.10 que $f_n^{-1}((a,\infty])$ é um conjunto mensurável para todo $a \in \mathbb{R}$. Logo, como a união numerável de conjuntos mensuráveis é mensurável, temos que $(\sup f_n)^{-1}((a,\infty])$ é mensurável. Segue outra vez da Proposição 2.10 que $\sup f_n$ é uma função mensurável. A demonstração das restantes deixa-se como exercício. $\square$

2.2 Integração de funções não negativas

Considere-se um espaço de medida $(X, \mathcal{F}, \mu)$. Uma função $\varphi : X \to [0, \infty)$ é **simples** sse $\varphi(X) = \{a_1, a_2, \dots, a_n\}$ e $A_i = \varphi^{-1}(\{a_i\})$ é um conjunto mensurável para $i = 1, 2, \dots, n$. Ou seja, uma função é simples se tiver um número finito de imagens e suas imagens inversas foram mensuráveis.

Note-se que os conjuntos A_i são disjuntos e sua união é X. Por outro lado pode-se escrever,

$$\varphi(x) = \sum_{i=1}^{n} a_i \chi_{A_i}(x)\,,$$

onde χ_{A_i} são as funções características de A_i.

Exercício 26. Mostre toda a função simples é mensurável.

Definição 2.2. O **integral de Lebesgue de uma função simples** φ em $E \subseteq X$ relativo à medida μ é o número

$$\int_E \varphi \, d\mu = \sum_{i=1}^{n} a_i \mu(E \cap A_i)\,.$$

Observação 2.12. Uma vez que é possível $\mu(A_i) = \infty$, usa-se a convenção $0 \times \infty = 0$ descrita no início desta secção.

Observação 2.13. Note-se que

$$\int_E \varphi \, d\mu = \int_X \varphi \cdot \chi_E \, d\mu\,.$$

Exemplo 2.14. Considere o espaço de medida $(\mathbb{R}, \mathcal{B}, m)$ onde m é a medida de Borel. Segundo a definição anterior, o integral de Lebesgue da função simples $\chi_{\mathbb{Q}}$ em $\mathbb{R}$ relativo a m é

$$\int_{\mathbb{Q}} \chi_{\mathbb{Q}} \, dm = 0 \times m(\mathbb{R} \setminus \mathbb{Q}) + 1 \times m(\mathbb{Q}) = 0 \times \infty + 1 \times 0 = 0\,.$$

Note-se que $\chi_{\mathbb{Q}}$ não é integrável segundo Riemann.

Exercício 27. Sejam φ, ψ funções simples e $a > 0$. Mostre que $\varphi + \psi$ e $a\varphi$ são funções simples.

Exercício 28. Calcule o integral de Lebesgue em $A \subset [0, \infty)$ relativo a m das seguintes funções simples:

1. $\phi(x) = [x]$ e $A = [0, 10]$.
2. $\phi(x) = \left[x^2\right]$ e $A = [0, 2]$.

Nota: o símbolo $[x]$ denota a parte inteira do número x.

Exercício 29. Sejam φ, ψ duas funções simples em X. Mostre que se $\varphi \leq \psi$ então $\int_E \varphi \, d\mu \leq \int_E \psi \, d\mu$ para todo conjunto mensurável E.

Na proposição que se segue, reunimos algumas propriedades importantes do integral de Lebesgue de funções simples.

Proposição 2.15. *[Propriedades do integral de Lebesgue de funções simples] Seja $(X, \mathcal{F}, \mu)$ um espaço de medida e φ, ψ duas funções simples em X. Então:*

1. *Se $\varphi \leq \psi$ então $\int_E \varphi \, d\mu \leq \int_E \psi \, d\mu$.*
2. *Se $E_i \subset X$, $i = 1, 2, \ldots$ é uma sucessão de conjuntos disjuntos então*
$$\int_{\bigcup_{i=1}^{\infty} E_i} \varphi \, d\mu = \sum_{i=1}^{\infty} \int_{E_i} \varphi \, d\mu \, .$$
3. *Qualquer que seja $a > 0$, $a\varphi$ é uma função simples e*
$$\int_E a\varphi \, d\mu = a \int_E \varphi \, d\mu \, .$$
4. *A soma $\varphi + \psi$ é uma função simples e*
$$\int_E \varphi + \psi \, d\mu = \int_E \varphi \, d\mu + \int_E \psi \, d\mu \, .$$

Demonstração.

1. Ver Exercício 29.
2. Segue directo da definição de integral de Lebesgue e da σ-aditividade da medida μ que
$$\begin{aligned} \int_{\bigcup_{i=1}^{\infty} E_i} \varphi \, d\mu &= \sum_{j=1}^{n} a_j \mu\Big(\bigcup_{i=1}^{\infty} E_i \cap A_j\Big) = \sum_{j=1}^{n} \sum_{i=1}^{\infty} a_j \mu(E_i \cap A_j) \\ &= \sum_{i=1}^{\infty} \sum_{j=1}^{n} a_j \mu(E_i \cap A_j) = \sum_{i=1}^{\infty} \int_{E_i} \varphi \, d\mu \end{aligned}$$
3. Se $\varphi = \sum c_i \chi_{A_i}$ então $a\varphi = \sum a c_i \chi_{A_i}$. Logo
$$\int_E a\varphi \, d\mu = \sum a c_i \mu(E \cap A_i) = a \sum c_i \mu(E \cap A_i) = a \int_E \varphi \, d\mu \, .$$

4. Se $\varphi = \sum c_i \chi_{A_i}$ e $\psi = \sum b_j \chi_{B_j}$ então

$$\varphi + \psi = \sum_{i,j} (c_i + b_j) \chi_{A_i \cap B_j} .$$

Por outro lado,

$$\int_{A_i \cap B_j} \varphi + \psi \, d\mu = (c_i + b_j) \mu(A_i \cap B_j) = \int_{A_i \cap B_j} \varphi \, d\mu + \int_{A_i \cap B_j} \psi \, d\mu .$$

Logo, como $E = \bigcup_{i,j} A_i \cap B_j$ e $A_i \cap B_j$ são disjuntos, segue da segunda propriedade desta proposição que

$$\begin{aligned} \int_E \varphi + \psi \, d\mu &= \int_{\bigcup_{i,j} A_i \cap B_j} \varphi + \psi \, d\mu \\ &= \sum_{i,j} \int_{A_i \cap B_j} \varphi + \psi \, d\mu \\ &= \sum_{i,j} \int_{A_i \cap B_j} \varphi \, d\mu + \sum_{i,j} \int_{A_i \cap B_j} \psi \, d\mu \\ &= \int_E \varphi \, d\mu + \int_E \psi \, d\mu \end{aligned}$$

□

Seja $(X, \mathcal{F}, \mu)$ um espaço de medida. De seguida definimos o integral de Lebesgue para funções mensuráveis **não negativas** usando o já conhecido integral de Lebesgue para funções simples.

Definição 2.3. O **integral de Lebesgue de uma função mensurável não negativa** $f : X \to [0, \infty]$ em $E \subseteq X$ relativo à medida μ é o número

$$\int_E f \, d\mu = \sup \left\{ \int_E \varphi \, d\mu : \varphi \leq f \text{ é uma função simples} \right\}$$

Note-se que o integral é sempre não negativo e pode tomar o valor ∞.

Quando f é uma função simples, obtemos duas definições de integral de Lebesgue para funções simples. No entanto, segue da primeira propriedade da Proposição 2.15 que o supremo é atingido quando $\varphi = f$. Logo as duas definições são equivalentes quando f é simples.

Proposição 2.16. *[Propriedades do Integral de Lebesgue] Sejam $f, g : X \to [0, \infty]$ duas funções mensuráveis não negativas. Então*

1. *Se $f \leq g$ então $\int_E f \, d\mu \leq \int_E g \, d\mu$.*

2. *Se* $A \subset B$ *então* $\int_A f\,d\mu \leq \int_B f\,d\mu$.
3. *Se* $c \in [0,\infty)$ *então* $\int_E cf\,d\mu = c\int_E f\,d\mu$.
4. *Se* $f = 0$ *então* $\int_E f\,d\mu = 0$.
5. *Se* $\mu(E) = 0$ *então* $\int_E f\,d\mu = 0$.
6. $\int_E f\,d\mu = \int_X \chi_E f\,d\mu$.

Demonstração. Deixa-se como exercício. □

A proposição que se segue é bastante útil, porque permite generalizar para funções mensuráveis resultados que se conhecem para funções simples.

Proposição 2.17. *Seja* $f : X \to [0,\infty]$ *uma função mensurável. Então existe uma sucessão* φ_n, $n = 1, 2, \dots$ *de funções simples definidas em* X *tal que* $\varphi_n \nearrow f$. *Ou seja* $\varphi_1 \leq \varphi_2 \leq \varphi_3 \leq \cdots$ *e*

$$\lim_{n\to\infty} \varphi_n(x) = f(x),\ \forall x \in X\,.$$

Demonstração. Para cada $n \geq 1$ define-se

$$E_{n,k} = \left\{x \in X : \frac{k-1}{2^n} \leq f(x) \leq \frac{k}{2^n}\right\} \quad \text{e} \quad F_n = \{x \in X : f(x) \geq n\}\,.$$

Como f é mensurável segue que $E_{n,k}$ e F_n são conjuntos mensuráveis. Seja

$$\varphi_n = n\chi_{F_n} + \sum_{k=1}^{n2^n} \frac{k-1}{2^n}\chi_{E_{n,k}}\,.$$

É fácil verificar φ_n é uma sucessão crescente de funções simples e que converge pontualmente para f. □

Exercício 30. Sejam $f, g : X \to [0,\infty]$ duas funções mensuráveis. Mostre que $f + g$ e $f \cdot g$ são funções mensuráveis de $X \to [0,\infty]$. (Dica: Use as Proposições 2.11 e 2.17).

De seguida, apresentamos três resultados fundamentais que permitem efectuar as usuais trocas de limites com integrais para funções mensuráveis não negativas.

Teorema 2.18 (da convergência monótona)**.** *Seja* $f_n : X \to [0,\infty]$ *uma sucessão de funções mensuráveis não negativas tal que* $f_n \nearrow f$, *ou seja,* $f_1 \leq f_2 \leq \cdots$ *e* $\lim_{n\to\infty} f_n(x) = f(x)$ *para todo* $x \in X$. *Então* f *é mensurável e*

$$\lim_{n\to\infty} \int_X f_n\,d\mu = \int_X f\,d\mu\,.$$

Demonstração. Como $f_1 \leq f_2 \leq \cdots$ então $\int_X f_1\,d\mu \leq \int_X f_2\,d\mu \leq \cdots$ pela Proposição 2.16. Logo, existe um $\alpha \in [0,\infty]$ tal que

$$\lim_{n\to\infty} \int_X f_n\,d\mu = \alpha\,.$$

Pela Proposição 2.11, $f = \lim_{n\to\infty} f_n$ é mensurável. Portanto, resta provar que

$$\alpha = \int_X f\,d\mu\,.$$

- ($\alpha \leq \int_X f\,d\mu$): Como $f_n \leq f$ temos que $\int_X f_n\,d\mu \leq \int_X f\,d\mu$ para todo $n = 1,2,\ldots$. Logo $\alpha \leq \int_X f\,d\mu$.
- ($\alpha \geq \int_X f\,d\mu$): A prova desta desigualdade é um pouco mais elaborada. Vamos minorar o integral $\int_X f_n$ da seguinte maneira. Seja φ uma função simples tal que $\varphi \leq f$. Dado $0 < c < 1$ define-se

  $$E_n = \{x \in X : f_n(x) \geq c\varphi(x)\}\ .$$

  Note-se que $E_1 \subset E_2 \subset E_3 \subset \cdots$ (porque $f_1 \leq f_2 \leq \cdots$) e $\bigcup_n E_n = X$ (porque $f > c\varphi$ para todo $c \in]0,1[$ e $f_n \nearrow f$, logo para todo $x \in X$ existe um $n \geq 1$ tal que $f_n(x) \geq c\varphi(x)$). Então

  $$\int_X f_n\,d\mu \geq \int_{E_n} f_n\,d\mu \geq c\int_{E_n} \varphi\,d\mu\,, \quad n = 1,2,\ldots$$

  Tomando $n \to \infty$ na desigualdade anterior obtém-se

  $$\alpha \geq c\int_X \varphi\,d\mu\,.$$

  Como c e $\varphi \leq f$ são arbitrários, segue da definição do integral de Lebesgue que

  $$\alpha \geq \sup\left\{\int_X \varphi\,d\mu : \varphi \leq f\right\} = \int_X f\,d\mu\,.$$

$\square$

Exemplo 2.19. Tome-se $f_n = \chi_{[n,n+1]}$. É claro que $\int_{\mathbb{R}} f_n\,dm = 1$ para todo $n \geq 1$ e $\lim f_n = 0$. Logo $\int_{\mathbb{R}}(\lim f_n)\,dm \neq \lim \int_{\mathbb{R}} f_n\,dm$.

Teorema 2.20 (conhecido por Lema de Fatou)**.** *Seja $f_n : X \to [0,\infty]$ uma sucessão de funções mensuráveis. Então*

$$\int_X \liminf_{n\to\infty} f_n\,d\mu \leq \liminf_{n\to\infty} \int_X f_n\,d\mu\,.$$

Demonstração. Seja $g_n(x) = \inf_{k \geq n} f_k(x)$. Então $g_n \leq f_n$ e

$$\int_X g_n \, d\mu \leq \int_X f_n \, d\mu \,, \quad n = 1, 2, \ldots$$

Por outro lado $g_1 \leq g_2 \leq g_3 \leq \cdots$ onde cada g_n é mensurável e $\lim_{n\to\infty} g_n = \liminf f_n$ por definição de limite inferior. Logo, o Teorema da convergência monótona implica que

$$\lim_{n\to\infty} \int_X g_n \, d\mu = \int_X \liminf_{n\to\infty} f_n \, d\mu \,.$$

Tomando $n \to \infty$ na desigualdade anterior obtém-se

$$\int_X \liminf_{n\to\infty} f_n \, d\mu \leq \liminf_{n\to\infty} \int_X f_n \, d\mu \,.$$

□

Teorema 2.21. *Sejam $f_n : X \to [0, \infty]$, $n = 1, 2, \ldots$ funções mensuráveis e*

$$f(x) = \sum_{n=1}^{\infty} f_n(x), \quad \forall x \in X \,.$$

Então

$$\int_X f \, d\mu = \sum_{n=1}^{\infty} \int_X f_n \, d\mu \,.$$

Demonstração. Exercício. □

Exercício 31. Demonstre o Teorema 2.21. (Dica: Considere em primeiro lugar uma soma finita e use a Proposição 2.17 e o Teorema da convergência monótona.)

Teorema 2.22. *Considere um espaço de medida $(X, \mathcal{F}, \mu)$ e seja $f : X \to [0, \infty]$ uma função mensurável. Então a função $\nu : \mathcal{F} \to [0, \infty]$ definida por*

$$\nu(E) = \int_E f \, d\mu \,, \quad E \in \mathcal{F}$$

é uma medida em $(X, \mathcal{F})$. Adicionalmente, se $g : X \to [0, \infty]$ é uma função mensurável então

$$\int_X g \, d\nu = \int_X g \cdot f \, d\mu \,. \tag{2.1}$$

Demonstração. A demonstração de que a função ν é uma medida deixa-se como exercício. Façamos a demonstração da fórmula (2.1).

- A fórmula é válida quando $g = \chi_E$. De facto

$$\int_X \chi_E \, d\nu = \nu(E) = \int_E f \, d\mu = \int_X \chi_E \cdot f \, d\mu \, .$$

- A fórmula é válida para qualquer função simples, ou seja, $g = \sum_{i=1}^n c_i \chi_{E_i}$ com E_i disjuntos. De facto,

$$\begin{aligned}\int_X \sum_i c_i \chi_{E_i} \, d\nu &= \sum_i c_i \int_X \chi_{E_i} \, d\nu \\ &= \sum_i c_i \int_X \chi_{E_i} \cdot f \, d\mu \\ &= \int_X (\sum_i c_i \chi_{E_i}) \cdot f \, d\mu\end{aligned}$$

- A fórmula é válida para qualquer g mensurável. De facto, segundo a Proposição 2.17 existe uma sucessão de funções simples $\varphi_n \nearrow g$. Aplicando o teorema da convergência monótona tem-se

$$\int_X g \, d\nu = \lim_{n \to \infty} \int_X \varphi_n \, d\nu = \lim_{n \to \infty} \int_X \varphi_n \cdot f \, d\mu = \int_X g \cdot f \, d\mu$$

□

Exercício 32. Demonstre a primeira parte do Teorema 2.22, ou seja, que a função $\nu : \mathcal{F} \to [0, \infty]$ definida por

$$\nu(E) = \int_E f \, d\mu \, , \quad E \in \mathcal{F}$$

é uma medida. (Dica: Use a propriedade (6) da Proposição 2.16 e o Teorema 2.21 para mostrar a σ-aditividade.)

Observação 2.23. O teorema anterior estabelece que dada uma função mensurável, o integral de Lebesgue de uma união finita ou infinita numerável de conjuntos é igual à soma dos integrais relativos a cada conjunto.

Observação 2.24. No contexto do teorema anterior podemos escrever formalmente $d\nu = f \, d\mu$.

Exemplo 2.25. Seja $(X, \mathcal{P}(X), \delta_a)$ o espaço de medida onde δ_a é medida de massa unitária concentrada em $a \in X$. Dada uma função mensurável $f : X \to [0, \infty]$ iremos calcular o integral de Lebesgue de f. Atendendo ao Teorema 2.22 vem que

$$\int_X f \, d\delta_a = \int_{\{a\}} f \, d\delta_a + \int_{X \setminus \{a\}} f \, d\delta_a \, .$$

Uma vez que $\delta_a(X \setminus \{a\}) = 0$ tem-se que o segundo integral é zero. Logo

$$\int_X f\, d\delta_a = \int_{\{a\}} f\, d\delta_a = \int_X \chi_{\{a\}} f\, d\delta_a\,.$$

Mas $\chi_{\{a\}} f = f(a)\chi_{\{a\}}$ é uma função simples. Logo,

$$\int_X f\, d\delta_a = f(a) \times \delta_a(\{a\}) = f(a)\,.$$

Exercício 33. Considere o espaço de medida $(X, \mathcal{F}, \mu)$ onde μ é a medida de contagem. Seja $A = \{x_1, x_2, x_3\}$ um subconjunto de X e $h : X \to [0, \infty]$ uma função mensurável. Mostre que

1. $\chi_A \cdot h$ é uma função simples, onde χ_A é a função característica de A.
2. Calcule $\int_A h\, d\mu$.

2.3 Integração de funções complexas e espaço $\mathcal{L}^1(\mu)$

Seja μ uma medida no espaço mensurável $(X, \mathcal{F})$. A teoria de integração para funções complexas é uma generalização imediata da teoria de integração de funções não negativas.

Definição 2.4. Define-se $\mathcal{L}^1(\mu)$ o conjunto de todas as funções complexas mensuráveis $f : X \to \mathbb{C}$ tais que

$$\int_X |f| < \infty\,.$$

Observação 2.26. Segue da Proposição 2.6 que $|f| : X \to [0, \infty[$ é uma função mensurável. Logo, faz sentido $\int_X |f|$.

Uma função $f \in \mathcal{L}^1(\mu)$ diz-se uma **função integrável** em X no sentido de Lebesgue, relativamente à medida μ. Neste caso, tomando $f = u + iv$, onde $u, v : X \to \mathbb{R}$ são funções mensuráveis, define-se o **integral de Lebesgue da função complexa** f em E, relativo à medida μ como sendo o número

$$\int_E f\, d\mu = \int_E u^+\, d\mu - \int_E u^-\, d\mu + i \int_E v^+\, d\mu - i \int_E v^-\, d\mu$$

onde u^+, v^+ e u^-, v^- são a parte positiva e negativa de u, v respectivamente. Note-se que f é integrável em E sse $\int_E u^\pm\, d\mu$ e $\int_E v^\pm\, d\mu$ são finitos.

Teorema 2.27 (Propriedades)**.** *Sejam $f, g \in \mathcal{L}^1(\mu)$ e $\alpha, \beta \in \mathbb{C}$. Então*

1. $\alpha f + \beta g \in \mathcal{L}^1(\mu)$.

2. $\int_X (\alpha f + \beta g)\, d\mu = \alpha \int_X f\, d\mu + \beta \int_X g\, d\mu$.

3. $\left|\int_X f\, d\mu\right| \leq \int_X |f|\, d\mu$.

Demonstração. Deixa-se como exercício. □

Observação 2.28. $\mathcal{L}^1(\mu)$ é um espaço linear.

Exemplo 2.29. Tome-se o espaço de medida $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$ onde μ é a medida de contagem. Considere-se o Boreliano $E = \{-2, -1, 1, 2\}$ e $g : \mathbb{R} \to \mathbb{R}$ a função mensurável $g(x) = x$. Então, pela definição de integral de Lebesgue

$$\begin{aligned}\int_E g\, d\mu &= \int_E g^+\, d\mu - \int_E g^-\, d\mu \\ &= \int_X \chi_E \cdot g^+\, d\mu - \int_X \chi_E \cdot g^-\, d\mu\,.\end{aligned}$$

Note-se que $\chi_E \cdot g^+$ é uma função simples, de facto

$$\chi_E \cdot g^+ = g^+(-2)\chi_{\{-2\}} + g^+(-1)\chi_{\{-1\}} + g^+(1)\chi_{\{1\}} + g^+(2)\chi_{\{2\}}\,.$$

Logo,

$$\int_X \chi_E \cdot g^+\, d\mu = g^+(1)\mu(\{1\}) + g^+(2)\mu(\{2\}) = 1 \times 1 + 2 \times 1 = 3\,.$$

De maneira análoga tem-se $\int_X \chi_E \cdot g^-\, d\mu = 3$. Logo, $\int_E g\, d\mu = 0$.

Como já mencionado anteriormente, do ponto de vista da teoria da medida e integração, os conjuntos de medida nula são desprezáveis. De facto, se $f = g$ excepto num conjunto de medida nula N então

$$\int_X f\, d\mu = \int_X g\, d\mu\,.$$

Portanto, diz-se que uma função f satisfaz a propriedade (P) **quase certamente (q.c.)** em E se f satisfizer essa propriedade para todos os pontos em E à excepção de um conjunto de medida nula.

Teorema 2.30. *Seja $(X, \mathcal{F}, \mu)$ um espaço de medida. Então*

1. Se $f : X \to [0, \infty]$ é mensurável e $\int_E f\, d\mu = 0$ para algum $E \in \mathcal{F}$ então $f = 0$ q.c. em E.

2. Se $f \in \mathcal{L}^1(\mu)$ e $\int_E f\, d\mu = 0$ para todo $E \in \mathcal{F}$ então $f = 0$ q.c. em X.

Demonstração.

1. Tome-se o conjunto mensurável $E_n = \{x \in E : f(x) \geq 1/n\}$. Então $0 = \int_E f\, d\mu \geq \int_{E_n} f\, d\mu \geq \int_{E_n} 1/n\, d\mu = \mu(E_n)/n$. Logo $\mu(E_n) = 0$. Uma vez que $\mu(\bigcup_n E_n) = 0$, segue que $f > 0$ para um conjunto de medida nula, ou seja, $f = 0$ q.c.
2. Deixa-se como exercício.

□

2.4 Integral de Lebesgue-Stieltjes

Numa secção anterior, construiriam-se medidas de Lebesgue-Stieltjes em $\mathbb{R}$, isto é, dada uma função de distribuição F existe um espaço de medida completo $(\mathbb{R}, \mathcal{M}_F, m_F)$ onde m_F é designada por medida de Lebesgue-Stieltjes. Tome-se uma função $g : \mathbb{R} \to \mathbb{R}$ mensurável relativamente à σ-álgebra $\mathcal{M}_F$. Ao integral de g relativamente à medida m_F designa-se por **integral de Lebesgue-Stieltjes** e é usual escrever-se

$$\int g\, dm_F = \int g\, dF\,.$$

2.5 Teorema da convergência dominada

No contexto das funções mensuráveis não negativas o teorema da convergência monótona garante que para uma sucessão de funções que convergem monotonamente para uma função então o integral da função limite é igual ao limite dos integrais das respectivas funções.

Nesta secção enunciamos um teorema semelhante ao da convergência monótona. O teorema que se segue, estabelece um conjunto de condições suficientes para se proceder à troca de limites com integral quando as funções a integrar tomam valores complexos.

Teorema 2.31 (da convergência dominada)**.** *Seja $f_n : X \to \mathbb{C}$ uma sucessão de funções mensuráveis e $g : X \to [0, \infty)$ uma função integrável, isto é $g \in \mathcal{L}^1(\mu)$, tal que para todo $n \geq 1$ se tem $|f_n| \leq g$.*

Se $f = \lim_n f_n$ então $f \in \mathcal{L}^1(\mu)$ e

$$\lim_{n\to\infty} \int_X f_n\, d\mu = \int_X f\, d\mu\,.$$

Demonstração. Tomando parte real e imaginária, pode-se supor que f_n e f são funções reais. Logo $-g \leq f_n \leq g$. Como $f = \lim_n f_n$ temos que f é mensurável e $-g \leq f \leq g$. Aplicando o lema de Fatou à sucessão $f_n + g \geq 0$ obtém-se

$$\int_X f + g\, d\mu \leq \liminf_{n\to\infty} \int_X f_n + g\, d\mu\,.$$

Segue do facto de g ser integrável que

$$\int_X f\, d\mu \leq \liminf_{n\to\infty} \int_X f_n\, d\mu\,.$$

Aplicando mais uma vez o lema de Fatou, mas agora à sucessão $g - f_n \geq 0$, obtém-se

$$\int_X g - f\, d\mu \leq \liminf_{n\to\infty} \int_X g - f_n\, d\mu\,,$$

ou escrito de forma equivalente

$$\limsup_{n\to\infty} \int_X f_n\, d\mu \leq \int_X f\, d\mu\,.$$

□

Exercício 34. Considere-se o espaço mensurável $(X, \mathcal{P}(X))$. Seja $A = \{x_1, x_2, x_3, \ldots\}$ um subconjunto numerável de X e $\mu : \mathcal{P}(X) \to [0,\infty]$ a seguinte função

$$\mu(E) = \sum_{x_i \in E} \alpha_i\,,$$

onde α_i, $i = 1, 2, \ldots$ são números reais não negativos. Mostre que

1. μ é uma medida, designada por **medida discreta**.
2. $\mu = \sum_{i=1}^{\infty} \alpha_i \delta_{x_i}$ onde δ_{x_i} é a medida de Dirac.
3. Dada uma função mensurável $f : X \to [0,\infty]$,

$$\int_E f\, d\mu = \sum_{x_i \in E} f(x_i)\alpha_i\,.$$

2.6 Relação com integral de Riemann

Nesta secção relacionamos o integral de Riemann com o recém definido integral de Lebesgue. Relembremos o conceito de integral de Riemann. Seja $f : [a,b] \to \mathbb{R}$ uma função real limitada onde $a < b$. Uma **partição** do intervalo $[a,b]$ é um conjunto finito $P = \{a_0, a_1, \ldots, a_n\}$ onde

$$a = a_0 < a_1 < \ldots < a_n = b\,.$$

Dada uma partição P pode-se definir as **somas inferior e superior** de Riemann da função f relativas à partição P,

$$\underline{\Sigma}(f,P) = \sum_{i=1}^{n} m_i(a_i - a_{i-1}) \quad \text{e} \quad \overline{\Sigma}(f,P) = \sum_{i=1}^{n} M_i(a_i - a_{i-1})\,,$$

onde $m_i = \inf\{f(x) : a_{i-1} < x < a_i\}$ e $M_i = \sup\{f(x) : a_{i-1} < x < a_i\}$. Uma vez que f é limitada estes números existem. Por fim define-se

$$\overline{\int_a^b} f = \inf\left\{\overline{\Sigma}(f,P) : P \text{ é uma partição de } [a,b]\right\} ,$$
$$\underline{\int_a^b} f = \sup\left\{\underline{\Sigma}(f,P) : P \text{ é uma partição de } [a,b]\right\} .$$

Finalmente, diz-se que f é **integrável à Riemann** sse $\underline{\int_a^b} f = \overline{\int_a^b} f$ e é comum denotar-se este número por

$$\int_a^b f(x)\,dx\,.$$

O seguinte teorema permite relacionar o integral de Riemann com o integral de Lebesgue.

Teorema 2.32. *Seja $f : [a,b] \to \mathbb{R}$ uma função limitada. Se f for integrável à Riemann então f é integrável relativamente à medida de Lebesgue e os integrais coincidem,*

$$\int_a^b f(x)\,dx = \int_{[a,b]} f\,dm\,.$$

Demonstração. Ver demonstração em [1]. □

Exemplo 2.33. 1. Se $f : [a,b] \to \mathbb{R}$ é contínua então é integrável à Riemann. Para além disso, se tiver uma **primitiva**, isto é, existir uma função $F : [a,b] \to \mathbb{R}$ tal que $F' = f$ então

$$F(b) - F(a) = \int_a^b f(x)\,dx\,.$$

2. Há funções simples que não são integráveis à Riemann, como é o caso da função característica $\chi_{\mathbb{Q}} : [0,1] \to \mathbb{R}$. De facto, $\underline{\Sigma}(\chi_{\mathbb{Q}}, P) = 0$ e $\overline{\Sigma}(\chi_{\mathbb{Q}}, P) = 1$ para toda a partição P. Logo $\overline{\int_0^1} \chi_Q q \neq \underline{\int_0^1} \chi_{\mathbb{Q}}$.

Exemplo 2.34. Considere-se a medida de Borel m no espaço mensurável $([0,1], \mathcal{B}([0,1]))$ e a seguinte sucessão de funções reais

$$f_n(x) = \frac{n \sin x}{1 + n^2 x^{1/2}} \quad n = 1, 2, \ldots$$

para $x \in [0,1]$. É claro que $\lim_n f_n = 0$. Para concluir que $\lim_n \int_{[0,1]} f_n \, dm = 0$ basta, usando o teorema da convergência dominada, encontrar uma função integrável $g \geq 0$ tal que $|f_n| \leq g$. Majorando f_n obtém-se

$$\left| \frac{n \sin x}{1 + n^2 x^{1/2}} \right| \leq \frac{n}{1 + n^2 x^{1/2}} \leq \frac{1}{n x^{1/2}} \leq \frac{1}{x^{1/2}} \, .$$

Por outro lado, $\frac{1}{x^{1/2}}$ é integrável à Riemann no intervalo $[0,1]$. Logo $|f_n|$ é majorada por uma função integrável (no sentido de Lebesgue). Segue do teorema da convergência dominada que $\lim_n \int_{[0,1]} f_n \, dm = 0$.

Exercício 35. Use o teorema da convergência dominada para calcular

$$\lim_{n \to \infty} \int_1^\infty \frac{\sqrt{x}}{1 + n x^3} \, dx \, .$$

2.7 Continuidade absoluta e Teorema de Radon-Nikodym

Seja $(X, \mathcal{F}, \mu)$ um espaço de medida e $f : X \to [0, \infty]$ uma função mensurável. Como foi visto, a função $\nu : \mathcal{F} \to [0, \infty]$ definida por

$$\nu(E) = \int_E f \, d\mu \, ,$$

é uma medida no espaço mensurável $(X, \mathcal{F})$. A medida ν tem a seguinte propriedade: se $\mu(E) = 0$ então $\nu(E) = 0$. Esta propriedade é de fundamental importância para teoria de probabilidades como veremos adiante.

Definição 2.5. Seja $(X, \mathcal{F})$ um espaço mensurável e μ, λ duas medidas definidas neste espaço. Diz-se que λ é **absolutamente contínua relativamente a** μ e escreve-se $\lambda \ll \mu$ sse para todo $E \in \mathcal{F}$ tal que $\mu(E) = 0$ então $\lambda(E) = 0$.

A definição anterior diz que $\lambda \ll \mu$ sse todos os conjuntos de medida nula de μ forem também conjuntos de medida nula para λ. No entanto, λ pode ter mais conjuntos de medida nula que μ.

Acabámos de ver que todas a medidas ν construídas através do integral $\int_E f \, d\mu$ são absolutamente contínuas relativamente a μ. A questão que se coloca é: será que todas a medidas absolutamente contínuas relativamente a μ podem ser obtidas dessa maneira? A resposta a esta questão é dada pelo teorema de Radon-Nikodym.

Teorema 2.35 (de Radon-Nikodym)**.** *Sejam λ e μ duas medidas finitas definidas em $(X, \mathcal{F})$ tal que $\lambda \ll \mu$. Então existe uma única função $h \in \mathcal{L}^1(\mu)$ tal que*

$$\lambda(E) = \int_E h \, d\mu, \quad \forall E \in \mathcal{F} \, .$$

Demonstração. A demonstração pode ser encontrada em [1]. $\square$

A função h designa-se por **derivada no sentido de Radon-Nikodym de** λ e escreve-se formalmente

$$h = \frac{d\lambda}{d\mu}\,.$$

Observação 2.36. A unicidade de h no teorema de Radon-Nikodym deve ser entendida no seguinte sentido: se f é outra função em $\mathcal{L}^1(\mu)$ tal que $\lambda(E) = \int_E f\, d\mu$ então $f = h$ q.c.

Observação 2.37. O teorema de Radon-Nikodym é válido para o caso mais geral de λ e μ serem duas medidas σ-finitas, como é o caso da medida de Lebesgue. Uma medida μ de $(X, \mathcal{F})$ diz-se **σ-finita** sse existirem conjuntos mensuráveis $A_n \in \mathcal{F}$, $n = 1, 2, \ldots$ tal que $\bigcup_{n=1}^{\infty} A_n = X$ e $\mu(A_n) < \infty$ para todo $n = 1, 2, \ldots$.

Definição 2.6. Seja $(X, \mathcal{F}, \mu)$ um espaço de medida e $A \in \mathcal{F}$. Diz-se que μ está **concentrada em** A sse

$$\mu(E) = \mu(E \cap A), \quad \forall E \in \mathcal{F}\,.$$

Definição 2.7. Duas medidas μ e λ definidas em $(X, \mathcal{F})$ dizem-se **mutuamente singulares** e escreve-se $\mu \perp \lambda$ sse existem dois conjuntos disjuntos $A, B \in \mathcal{F}$ tal que μ está concentrada em A e λ está concentrada em B.

Proposição 2.38 (Propriedades de $\ll$ e $\perp$.)**.** *Sejam $\mu, \lambda, \lambda_1, \lambda_2$ medidas no espaço mensurável $(X, \mathcal{F})$. Então*

1. *se $\lambda_1 \perp \mu$ e $\lambda_2 \perp \mu$ então $(\lambda_1 + \lambda_2) \perp \mu$.*
2. *se $\lambda_1 \ll \mu$ e $\lambda_2 \ll \mu$ então $(\lambda_1 + \lambda_2) \ll \mu$.*
3. *se $\lambda_1 \ll \mu$ e $\lambda_2 \perp \mu$ então $\lambda_1 \perp \lambda_2$.*
4. *se $\lambda \ll \mu$ e $\lambda \perp \mu$ então $\lambda = 0$.*

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

O resultado que se segue estabelece que qualquer medida finita pode ser decomposta numa parte absolutamente contínua e numa parte singular (com respeito à uma medida finita de referência).

Teorema 2.39 (da decomposição de Lebesgue)**.** *Sejam μ e λ duas medidas finitas no espaço mensurável $(X, \mathcal{F})$. Então existe um par de medidas (λ_a, λ_s) definidas em $(X, \mathcal{F})$ tal que*

$$\lambda = \lambda_a + \lambda_s, \quad \text{onde} \quad \lambda_a \ll \mu \quad \text{e} \quad \lambda_s \perp \mu\,.$$

Demonstração. Ver demonstração em [1]. □

Exercício 36. Considere-se as seguinte medidas em $([0,1],\mathcal{B})$: $\mu_1 = \delta_0$, $\mu_2 = m$ e $\mu_3 = \frac{\mu_1}{2} + \frac{\mu_2}{2}$. Para que $i \neq j$ se tem $\mu_i \ll \mu_j$? Determine a derivada no sentido de Radon-Nikodym em cada caso.

2.8 Medida imagem

Nesta secção introduzimos o conceito de medida imagem. Este conceito é fundamental no desenvolvimento da teoria da probabilidade.

Teorema 2.40 (Medida imagem)**.** *Sejam $(X,\mathcal{F},\mu)$ um espaço de medida e $(Y,\mathcal{B})$ um espaço mensurável onde $\mathcal{B}$ é a σ-álgebra de Borel do espaço topológico Y. Dada uma função mensurável $f : X \to Y$ então a função $f_*\mu : \mathcal{B} \to [0,\infty]$ definida por*

$$f_*\mu(B) = \mu(f^{-1}(B)), \quad B \in \mathcal{B}$$

é uma medida em $(Y,\mathcal{B})$.

Demonstração. Deixa-se como exercício. (Dica: use as propriedades da imagem inversa) □

A medida $f_*\mu$ do teorema anterior designa-se por **medida imagem de μ por f**.

Teorema 2.41 (Mudança de variável)**.** *Nas condições do teorema anterior, uma função mensurável $g : Y \to \mathbb{R}$ é integrável relativamente à medida imagem $f_*\mu$ sse a função composta $g \circ f : X \to \mathbb{R}$ é integrável relativamente a μ e*

$$\int_B g\, d(f_*\mu) = \int_{f^{-1}(B)} g \circ f \, d\mu\,.$$

Demonstração. Deixa-se como exercício. (Dica: prove o teorema para funções simples, de seguida use o teorema da convergência monótona e prove para funções mensuráveis não negativas e, por fim, use a definição de integral de Lebesgue para funções mensuráveis.) □

2.9 Produto de medidas e Teorema de Fubini

Sejam $(X_1,\mathcal{F}_1,\mu_1)$ e $(X_2,\mathcal{F}_2,\mu_2)$ dois espaços de medida. Designamos por X o produto cartesiano $X = X_1 \times X_2$ e ao conjunto $A = A_1 \times A_2$ onde $A_1 \in \mathcal{F}_1$ e $A_2 \in \mathcal{F}_2$ chamamos **rectângulo mensurável**. Seja $\mathcal{R} \subset \mathcal{P}(X)$ a colecção de todos os rectângulos mensuráveis.

Definição 2.8. A σ-álgebra gerada pelos rectângulos mensuráveis $\sigma(\mathcal{R})$ chama-se **σ-álgebra produto** e designa-se por $\mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2$.

Dado um conjunto mensurável $F \in \mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2$, **a secção de** F **em** x_1 é o conjunto

$$F_{x_1} = \{x_2 \in X_2 \colon (x_1, x_2) \in F\} .$$

Analogamente se define a secção de F em x_2 e designa-se por F_{x_2}.

Exercício 37. Mostre que se F é um conjunto mensurável de $\mathcal{F}$ então $F_{x_1} \in \mathcal{F}_2$ e $F_{x_2} \in \mathcal{F}_1$. (Dica: Mostre que a colecção $\mathcal{G} = \{F \in \mathcal{F} : F_{x_1} \in \mathcal{F}_2, \forall\, x_1 \in X_1\}$ é uma σ-álgebra que contém todos os rectângulos mensuráveis).

O nosso objectivo é encontrar uma medida μ definida em $\mathcal{F}_1 \otimes \mathcal{F}_2$ tal que para qualquer rectângulo mensurável $A = A_1 \times A_2$ se tem

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

Dado um rectângulo mensurável $A = A_1 \times A_2$, a função $x_1 \mapsto \mu_2(A_{x_1})$ é simples. De facto,

$$\mu_2(A_{x_1}) = \begin{cases} \mu_2(A_2) & \text{se } x_1 \in A_1 \\ 0 & \text{se } x_1 \notin A_1 \end{cases}$$

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 2.42 (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 2.43 (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 $\Omega = \{1, 2, \ldots, n\}$. 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,

$$x = 0.a_1 a_2 a_3 \cdots$$

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. □

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 2.41 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 2.22 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. $\square$

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 $A_1, \ldots, A_n \in \mathcal{F}$ são **independentes** se para qualquer escolha de índices $i_1, \ldots, i_k \in \{1, \ldots, n\}$ se tem

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

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

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

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

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 $\sigma(X) = \big\{\emptyset, [0,1], [0,\frac{1}{2}], (\frac{1}{2},1]\big\}$ e $\sigma(Y) = \big\{\emptyset, [0,1], [\frac{1}{4},\frac{3}{4}], [0,\frac{1}{4}) \cup (\frac{3}{4},1]\big\}$ 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

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

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* $x_1 \leq y_1$ *e* $x_2 \leq y_2$ *então* $F_X(x_1, x_2) \leq F_X(y_1, y_2)$.
3. $\displaystyle\lim_{(x_1,x_2)\to(+\infty,+\infty)} F_X(x_1, x_2) = 1$
4. $\displaystyle\lim_{x_1\to-\infty} F_X(x_1, x_2) = \lim_{x_2\to-\infty} F_X(x_1, x_2) = 0.$
5. $\displaystyle\lim_{x_1\to+\infty} F_X(x_1, x_2) = F_{X_2}(x_2)$ *e* $\displaystyle\lim_{x_2\to+\infty} F_X(x_1, x_2) = F_{X_1}(x_1).$

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 $h(x_1, x_2) = x_1 x_2$ 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

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

□

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

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

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

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*

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

Demonstração. Deixa-se como exercício. □

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

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

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,

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

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

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

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

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 2.30 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 2.30 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.

Definição 5.1. Para cada $\omega \in \Omega$, a sucessão de números reais

$$(X_1(\omega), X_2(\omega), X_3(\omega), \ldots)$$

é designada por **trajectória** ou **realização**.

Definição 5.2. As variáveis aleatórias $X_1, X_2, \ldots$ dizem-se **independentes** sse $X_1, \ldots, X_n$ são independentes para todo $n \geq 1$ e dizem-se **identicamente distribuídas** sse X_i, $i = 1, 2, \ldots$ tiverem a mesma distribuição de probabilidade.

Observação 5.1. Variáveis aleatórias independentes e identicamente distribuídas são designadas por **IID**.

Um problema fundamental relacionado com sucessões de variáveis aleatórias é o estudo do seu comportamento assimptótico. Como exemplo, considere-se o lançamento de uma moeda ao ar um número arbitrário de vezes. Seja $X_i = 1$ se a moeda fica de face voltada para cima no i-ésimo lançamento e $X_i = 0$ caso contrário. Assumindo que a moeda é "perfeita" temos que

$$P(X_i = 1) = P(X_i = 0) = \frac{1}{2}\,.$$

Supondo que os lançamentos são independentes entre si, temos uma sucessão de variáveis aleatórias $(X_1, X_2, \ldots)$ IID. Forme-se a soma,

$$S_n = X_1 + X_2 + \cdots + X_n\,,$$

que representa o número de faces em n lançamentos. Quando tomamos o limite $n \to \infty$ esperamos que a razão $\frac{S_n}{n}$ se aproxime da probabilidade real de a face ficar voltada para cima, ou seja, $\frac{1}{2}$. Este tipo de resultado é conhecido pela lei dos grandes números.

Proposição 5.2 (Lei fraca dos grandes números)**.** *Sejam* $X_1, X_2, \ldots$ *variáveis aleatórias independentes tal que*

$$E(X_i) = \mu < \infty \quad e \quad \mathrm{Var}(X_i) \leq K < \infty\,, \quad \forall\, i = 1, 2, \ldots\,.$$

Então $\frac{S_n}{n}$ *converge em probabilidade para* μ*, ou seja,*

$$\forall\, \epsilon > 0\,, \quad \lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - \mu\right| < \epsilon\right) = 1\,.$$

Demonstração. Em primeiro lugar note-se que $E(S_n) = n\mu$ pela linearidade do valor esperado e $\mathrm{Var}(S_n) \leq nK$ porque $X_1, X_2, \ldots$ são independentes. Logo, aplicando a desigualdade de Markov vem que,

$$P\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) = P\left((S_n - n\mu)^2 \geq n^2\epsilon^2\right) \leq \frac{\mathrm{Var}(S_n)}{n^2\epsilon^2} \leq \frac{K}{n\epsilon^2}\,.$$

Logo,

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) = 0\,.$$

Ou seja, dado $\epsilon > 0$ arbitrário tem-se

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - \mu\right| < \epsilon\right) = 1 - \lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) = 1\,.$$

□

Observação 5.3. Existem resultados de convergência mais fortes que o anterior, nomeadamente a lei forte dos grandes números que estabelece a convergência $S_n/n \to \mu$ à excepção de um conjunto de probabilidade nula: se $X_1, X_2, \ldots$ são variáveis aleatórias IID com valor esperado finito então

$$\lim_{n\to\infty} \frac{S_n}{n} = \mu\,, \quad P\text{-q.c.}\,.$$

5.1 Martingalas

O termo Marginala tem as suas origens nos jogos de apostas e descreve o conceito de jogo honesto. As martingalas encontram aplicações em diversas áreas, nomeadamente em matemática financeira, na modelação de activos financeiros. Antes de passarmos à definição de martingala temos de introduzir o conceito de filtração.

Definição 5.3. Diz-se que as σ-álgebras $\mathcal{F}_1, \mathcal{F}_2, \ldots$ de partes de Ω formam uma **filtração** sse $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots$.

Exemplo 5.4. Considere uma sucessão de variáveis aleatórias $(X_n)_{n\geq 1}$ e as σ-álgebras $\mathcal{F}_n \subset \mathcal{F}$ definidas da seguinte maneira,

$$\mathcal{F}_n = \sigma(X_1, \ldots, X_n)\,.$$

Então $\mathcal{F}_1, \mathcal{F}_2, \ldots$ formam uma filtração e representam toda a informação até ao instante n, ou seja, todos os acontecimentos possíveis de observar de $(X_1, \ldots, X_n)$. A filtração $(\mathcal{F}_n)_{n\geq 1}$ designa-se por **filtração canónica**.

Definição 5.4. Diz-se que uma sucessão de variáveis aleatórias $(X_n)_{n\geq 1}$ é **adaptada** à filtração $(\mathcal{F}_n)_{n\geq 1}$ sse X_n é $\mathcal{F}_n$-mensurável.

Exemplo 5.5. Qualquer sucessão de variáveis aleatórias é adaptada à filtração canónica.

Estamos em condições de definir o conceito de martingala.

Definição 5.5. Uma sucessão de variáveis aleatórias $(X_n)_{n\geq 1}$ é uma **martingala** relativamente à filtração $(\mathcal{F}_n)_{n\geq 1}$ sse

1. X_n é integrável para todo $n = 1, 2, \ldots$, ou seja, $E(|X_n|) < \infty$.
2. $(X_1, X_2, \ldots)$ é adaptada a $(\mathcal{F}_1, \mathcal{F}_2, \ldots)$.
3. $E(X_{n+1}|\mathcal{F}_n) = X_n$, P-q.c. para todo $n = 1, 2, \ldots$.

Observação 5.6. Da definição de esperança condicional tem-se

$$\int_A E(X_{n+1}|\mathcal{F}_n)\, dP = \int_A X_{n+1}\, dP\,, \quad \forall\, A \in \mathcal{F}_n\,.$$

Por outro lado, segue da terceira condição da definição de martingala que

$$\int_A X_n\, dP = \int_A X_{n+1}\, dP\,.$$

Portanto, tomando $A = \Omega$ temos que

$$E(X_{n+1}) = E(X_n)\,, \quad n = 1, 2, \ldots\,.$$

Usando indução prova-se que $E(X_n) = E(X_1)$ para todo $n = 1, 2, \ldots$. Ou seja, uma martingala $(X_n)_{n\geq 1}$ tem valor esperado constante.

Observação 5.7. Se $(X_n)_{n\geq 1}$ é uma martingala relativamente à filtração $(\mathcal{F}_n)_{n\geq 1}$ então

$$E(X_n|\mathcal{F}_k) = X_k\,, \quad \text{para } k \in \{1, 2, \ldots, n-1\}\,.$$

De facto, usando as propriedades da esperança condicional podemos escrever

$$E(X_n|\mathcal{F}_k) = E(E(X_n|\mathcal{F}_{n-1})|F_k) = E(X_{n-1}|\mathcal{F}_k)\,.$$

Usando indução temos que

$$E(X_n|\mathcal{F}_k) = E(X_k|\mathcal{F}_k) = X_k\,.$$

Exemplo 5.8. Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes com valor esperado nulo. O **passeio aleatório**,

$$S_n = X_1 + X_2 + \cdots + X_n\,,$$

é uma martingala relativamente à filtração canónica

$$\mathcal{F}_n = \sigma(X_1, \ldots, X_n)\,.$$

De facto, a primeira e segunda condição da definição de martingala verificam-se trivialmente. Relativamente à terceira condição temos que,

$$\begin{aligned}
E(S_{n+1}|X_1, \ldots, X_n) &= E(X_{n+1} + S_n|X_1, \ldots, X_n) \\
&= E(X_{n+1}|X_1, \ldots, X_n) + E(S_n|X_1, \ldots, X_n) \\
&= E(X_{n+1}) + S_n \\
&= S_n
\end{aligned}$$

Exemplo 5.9. Seja Y uma variável aleatória integrável, isto é $E(|Y|) < \infty$ e $(\mathcal{F}_n)_{n\geq 1}$ uma filtração. Então a sucessão de variáveis aleatórias

$$X_n = E(Y|\mathcal{F}_n)\,, \quad n = 1, 2, \ldots$$

é uma martingala relativamente à filtração $(\mathcal{F}_n)_{n\geq 1}$. De facto,

$$|X_n| = |E(Y|\mathcal{F}_n)| \leq E(|Y|\,|\mathcal{F}_n)\,.$$

Logo,

$$E(|X_n|) \leq E(E(|Y|\,|\mathcal{F}_n)) \leq E(|Y|) < \infty\,.$$

Por outro lado,

$$E(X_{n+1}|\mathcal{F}_n) = E(E(Y|\mathcal{F}_{n+1})|\mathcal{F}_n) = E(Y|\mathcal{F}_n) = X_n\,.$$

Esta martingala é conhecida como **processo de Doob**.

Exemplo 5.10. Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes tais que $E(X_n) = 1$ para todo $n \geq 1$. A sucessão de variáveis aleatórias

$$Z_n = X_1 \cdot X_2 \cdots X_n\,,$$

é uma martingala relativamente à filtração canónica $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$. De facto, uma vez que as variáveis X_n são independentes temos que

$$E(Z_n) = E(X_1 \cdots X_n) = \prod_{i=1}^{n} E(X_i) = 1\,.$$

Logo, $E(|Z_n|) < \infty$. Por outro lado,

$$E(Z_{n+1}|\mathcal{F}_n) = E(X_{n+1} \cdot Z_n|\mathcal{F}_n) = Z_n E(X_{n+1}|\mathcal{F}_n) = Z_n E(X_{n+1}) = Z_n\,.$$

Exercício 50. Seja $(X_n)_{n\geq 1}$ uma martingala relativamente à filtração $(\mathcal{F}_n)_{n\geq 1}$. Mostre que $(X_n)_{n\geq 1}$ é uma martingala relativamente à filtração canónica,

$$\mathcal{G}_n = \sigma(X_1, \ldots, X_n)\,.$$

Retomando o exemplo do lançamento da moeda, considere o seguinte jogo. Se no n-ésimo lançamento da moeda sair face então ganha 1 euro, caso contrário perde 1 euro. Formalizando, seja $\Omega = \{\mathrm{E}, \mathrm{F}\}^{\mathbb{N}}$ o conjunto de todas as sequências de letras possíveis de formar com $\{\mathrm{E}, \mathrm{F}\}$ e $X_n : \Omega \to \mathbb{R}$ a variável aleatória que toma os valores $X_n(\omega) = 1$ se a n-ésima letra da sequência ω for um F e $X_n(\omega) = -1$ caso contrário. Logo, concluímos do Exemplo 5.8 que o ganho total até ao n-ésimo lançamento,

$$S_n = X_1 + \cdots + X_n\,,$$

é uma martingala relativamente à filtração canónica. Portanto, da terceira condição da definição de martingala concluímos que o lançamento de uma moeda perfeita é um exemplo de um **jogo justo**, ou seja, que perante a informação acumulada até ao n-ésimo lançamento, espera-se que no lançamento seguinte o ganho total não se altere.

Nem todos os jogos são justos, no sentido em que podem existir jogos mais favoráveis que outros. Um exemplo de um jogo favorável seria o do lançamento de dois dados em que o apostador ganharia se a soma dos dados fosse maior ou igual a 6.

Definição 5.6. Uma sucessão de variáveis aleatórias $(X_n)_{n\geq 1}$ é uma **supermartingala** (**submartingala**) relativamente à filtração $(\mathcal{F}_n)_{n\geq 1}$ sse

1. X_n é integrável para todo $n = 1, 2, \ldots$.
2. $(X_1, X_2, \ldots)$ é adaptada a $(\mathcal{F}_1, \mathcal{F}_2, \ldots)$.
3. $E(X_{n+1}|\mathcal{F}_n) \leq X_n$ (respectivamente, $E(X_{n+1}|\mathcal{F}_n) \geq X_n$) para todo $n = 1, 2, \ldots$.

Portanto, uma supermartingala (submartingala) representa um jogo desfavorável (respectivamente, favorável).

5.1.1 Estratégias

Considere um jogo de apostas em que em cada jogada é possível apostar uma certa quantidade que pode ou não perder com certa probabilidade. Ou seja, considere uma sucessão de variáveis aleatórias $(X_n)_{n\geq 1}$ tais que $E(|X_n|) < \infty$ que representam o valor que pode ganhar ou perder em cada jogada. Se as suas apostas em cada jogada são igual à unidade então o ganho total no final de n jogadas é

$$S_n = X_1 + \cdots + X_n\,.$$

Considere a filtração canónica e por conveniência tome $S_0 = 0$ e $\mathcal{F}_0 = \{\emptyset, \Omega\}$.

Suponhamos que decide em cada jogada alterar a sua aposta com base nas jogadas anteriores. Ou seja, na jogada n faz uma aposta α_n que é uma função mensurável relativamente à informação das jogadas anteriores $\mathcal{F}_{n-1}$. Então o ganho total até à n-ésima jogada é dado por,

$$Z_n = \alpha_1 X_1 + \cdots + \alpha_n X_n\,.$$

Uma **estratégia** é uma sucessão de funções $(\alpha_n)_{n\geq 1}$ tais que α_n é mensurável relativamente a $\mathcal{F}_{n-1}$ para $n = 1, 2, \ldots$. Numa situação real, as apostas α_n terão de ser funções limitadas, uma vez que nenhum jogador pode dispôr de fundos ilimitados. Portanto, uma **estratégia limitada** é uma estratégia $(\alpha_n)_{n\geq 1}$ tal que α_n é uma função limitada para $n = 1, 2, \dots$.

A seguinte proposição demonstra que é impossível através de uma estratégia limitada alterar um jogo a nosso favor.

Proposição 5.11. *Se $(S_n)_{n\geq 1}$ é uma martingala relativamente a $(\mathcal{F}_n)_{n\geq 1}$ e $(\alpha_n)_{n\geq 1}$ é uma estratégia limitada. Então $(Z_n)_{n\geq 1}$ é uma martingala relativamente a $(\mathcal{F}_n)_{n\geq 1}$.*

Demonstração. Seja $C_n = \max_{1\leq k\leq n} |\alpha_n|$. Note-se que $C_n < \infty$ por hipótese. Então,

$$E(|Z_n|) = E(|\alpha_1 X_1 + \cdots + \alpha_n X_n|) \leq C_n E(|S_n|) < \infty\,.$$

Por outro lado,

$$\begin{aligned} E(Z_n|\mathcal{F}_{n-1}) &= E(Z_{n-1} + \alpha_n(S_n - S_{n-1})|\mathcal{F}_{n-1}) \\ &= E(Z_{n-1}|\mathcal{F}_{n-1}) + E(\alpha_n(S_n - S_{n-1})|\mathcal{F}_{n-1}) \\ &= Z_{n-1} + \alpha_n(E(S_n|\mathcal{F}_{n-1}) - S_{n-1})\,. \end{aligned}$$

Como α_n é limitada temos que $\alpha_n(E(S_n|\mathcal{F}_{n-1}) - S_{n-1}) = 0$ pela definição de martingala. Logo $E(Z_n|\mathcal{F}_{n-1}) = Z_{n-1}$. □

Observação 5.12. É válido um resultado análogo no contexto de supermartingalas (submartingalas). Mais concretamente, se α_n é uma estratégia limitada não negativa ($\alpha_n \geq 0$) e S_n uma supermartingala (submartingala) então Z_n é uma supermartingala (submartingala).

Exemplo 5.13 (A martingala)**.** Considere novamente o jogo do lançamento repetido de uma moeda perfeita. O ganho acumulado é dado por

$$S_n = X_1 + \cdots + X_n\,,$$

onde $X_1, X_2, \dots$ são variáveis aleatórias IID tais que $X_n = 1$ ou $X_n = -1$ com igual probabilidade.

A **martingala** é um tipo de estratégia usada nos jogos de apostas que teve origem em França no século XVIII. A estratégia consiste em cada jogada duplicar a aposta anterior caso o resultado anterior tenha sido desfavorável. Ou seja,

$$\alpha_n = \begin{cases} 2^n & \text{se } X_1 = X_2 = \cdots = X_{n-1} = -1 \\ 0 & \text{caso contrário} \end{cases}$$

Portanto, ao fim de n jogadas tem um ganho acumulado de

$$Z_n = \alpha_1 X_1 + \cdots + \alpha_n X_n\,,$$

Segue da proposição anterior que Z_n é uma martingala relativamente à filtração canónica,

$$\mathcal{F}_n = \sigma(X_1, \ldots, X_n)\,.$$

Note-se que $E(Z_n) = 0$ para $n = 1, 2, \ldots$ (uma das propriedades da martigala). Considere agora o menor n tal que o resultado da n-ésima jogada foi favorável, ou seja, a função

$$\tau(\omega) = \min\{n \in \mathbb{N} \colon X_n(\omega) = 1\}\ .$$

Note-se que o acontecimento $\{\tau = n\} \in \mathcal{F}_n$. De facto, como podemos escrever

$$\{\tau = n\} = \{X_1 = -1\} \cap \cdots \cap \{X_{n-1} = -1\} \cap \{X_n = 1\}\ .$$

segue que $\{\tau = n\} \in \mathcal{F}_n$. Por outras palavras, o jogador toma a decisão de parar o jogo na n-ésima jogada com base na informação de todas as jogadas que já fez. Designamos a função τ por **tempo de paragem**. Usando o tempo de paragem podemos escrever o ganho acumulado,

$$Z_n(\omega) = \begin{cases} -1 - 2 - \ldots - 2^{n-1} & \text{se } n < \tau(\omega) \\ 1 & \text{se } n \geq \tau(\omega) \end{cases}.$$

Sempre que $n = \tau(\omega) < \infty$ temos que $Z_{\tau(\omega)}(\omega) = 1$. De facto $Z_\tau = 1$, P-q.c. uma vez que $P(\tau < \infty) = 1$. Logo

$$E(Z_\tau) = 1\,.$$

Por outro lado,

$$E(Z_{\tau-1}) = \sum_{n=2}^{\infty} (-1 - \ldots - 2^{n-2}) P(\tau = n) = \sum_{n=2}^{\infty} \frac{1 - 2^{n-1}}{2^n} = -\infty\,.$$

Ou seja, um jogador pode ir acumulando pequenos ganhos ao longo do tempo dando a ilusão de que a sua estratégia é ganhadora. No entanto, a possibilidade remota de ter um prejuízo catastrófico contrabalança esses ganhos e pode levar o jogador à falência rapidamente. Portanto, a estratégia da martingala é apenas bem sucedida se o jogador dispôr de recursos e tempo ilimitados.

5.1.2 Tempos de paragem

Motivados pelo exemplo anterior, fazemos a seguinte definição.

Definição 5.7. Um **tempo de paragem relativamente à filtração** $\mathcal{F}_n$ é uma função $\tau : \Omega \to \mathbb{N} \cup \{\infty\}$ que satisfaz

$$\{\tau = n\} \in \mathcal{F}_n\,, \quad n = 1, 2, \ldots\,.$$

Quando é claro a que filtração se refere o tempo de paragem, dizemos apenas que τ é um tempo de paragem.

Exercício 51. Mostre que $\{\tau = n\} \in \mathcal{F}_n$ sse $\{\tau \leq n\} \in \mathcal{F}_n$.

Exercício 52. Seja $(X_n)_{n\geq 1}$ uma sucessão de variáveis aleatórias adaptada a uma filtração $\mathcal{F}_n$ e seja $B \subset \mathbb{R}$ um Boreliano. Mostre que o **tempo de primeira entrada** de X_n em B,

$$\tau(\omega) = \min\{n \in \mathbb{N}\colon X_n(\omega) \in B\} ,$$

é um tempo de paragem relativamente a $\mathcal{F}_n$. (Dica: generalizar o argumento do Exemplo 5.13 escrevendo $\{\tau = n\}$ como uma intersecção de acontecimentos em $\mathcal{F}_n$).

Se $(X_n)_{\geq 1}$ uma sucessão de variáveis aleatórias adaptada a uma filtração $(\mathcal{F}_n)_{n\geq 1}$. Dado um tempo de paragem τ podemos definir uma nova sucessão,

$$X_n^\tau(\omega) = \begin{cases} X_n(\omega) & \text{se } n \leq \tau(\omega) \\ X_{\tau(\omega)}(\omega) & \text{se } n > \tau(\omega) \end{cases} ,$$

que representa a **sucessão** $(X_n)_{n\geq 1}$ **parada no instante** τ. Por exemplo, se ω é tal que $\tau(\omega) = 3$ então a realização da sucessão original é

$$X_1(\omega), X_2(\omega), X_3(\omega), X_4(\omega), X_5(\omega), \ldots$$

enquanto que a realização da sucessão parada nesse instante é

$$X_1(\omega), X_2(\omega), X_3(\omega), X_3(\omega), X_3(\omega), \ldots .$$

Observação 5.14. É usual escrever-se,

$$X_n^\tau(\omega) = X_{\min\{n,\tau(\omega)\}}(\omega) = X_{\tau(\omega)\wedge n}(\omega) , \quad \forall\, \omega \in \Omega ,$$

onde $a \wedge b = \min\{a, b\}$.

A função,

$$X_\tau(\omega) = \begin{cases} X_{\tau(\omega)}(\omega) & \text{se } \tau(\omega) < \infty \\ 0 & \text{caso contrário} \end{cases} ,$$

é designada por X_n **no momento exacto do tempo de paragem**. Note-se que X_τ é $\mathcal{F}$-mensurável, uma vez que,

$$X_\tau = \sum_{n=1}^{\infty} X_n \cdot \chi_{\{\tau=n\}} .$$

Vimos na secção anterior que não é possível através de uma estratégia limitada alterar um jogo justo a nosso favor. A seguinte proposição mostra um resultado semelhante no contexto dos tempos de paragem.

Proposição 5.15. *Seja X_n uma martingala relativamente à filtração $\mathcal{F}_n$ e τ um tempo de paragem. Então $X_{\tau\wedge n}$ é uma martingala relativamente a $\mathcal{F}_n$.*

Demonstração. Seja

$$\alpha_n(\omega) = \begin{cases} 1 & \text{se } n \leq \tau(\omega) \\ 0 & \text{se } n > \tau(\omega) \end{cases} .$$

Então podemos escrever,

$$X_{\tau\wedge n} = \alpha_1 Y_1 + \alpha_2 Y_2 + \cdots + \alpha_n Y_n \,,$$

onde

$$Y_1 = X_1 \quad \text{e} \quad Y_k = X_k - X_{k-1} \,, \quad \forall\, k \geq 2 \,.$$

É fácil verificar que α_n é uma estratégia limitada. De facto, dado um Boreliano B de $\mathbb{R}$ temos que:

- se $0 \notin B$ e $1 \notin B$ então $\{\alpha_n \in B\} = \emptyset \in \mathcal{F}_{n-1}$;
- se $0 \in B$ e $1 \notin B$ então $\{\alpha_n \in B\} = \{\tau < n\} \in \mathcal{F}_{n-1}$;
- se $0 \notin B$ e $1 \in B$ então $\{\alpha_n \in B\} = \{\tau \geq n\} = \{\tau < n\}^{\mathrm{c}} \in \mathcal{F}_{n-1}$;
- se $0 \in B$ e $1 \in B$ então $\{\alpha_n \in B\} = \Omega \in \mathcal{F}_{n-1}$.

Logo α_n é $\mathcal{F}_{n-1}$-mensurável. Para provar que $X_{\tau\wedge n}$ é uma martingala basta mostrar, pela Proposição 5.11, que $S_n = Y_1 + \cdots + Y_n$ é uma martingala relativamente a $\mathcal{F}_n$. Mas

$$S_n = X_1 + (X_2 - X_1) + \cdots + (X_n - X_{n-1}) = X_n \,.$$

□

Observação 5.16. É válido um resultado análogo no contexto de supermartingalas (submartingalas).

5.1.3 Teorema de paragem opcional

Vimos na secção anterior que se X_n é uma martingala e τ um tempo de paragem então $X_{\tau\wedge n}$ é também uma martingala. Adicionalmente,

$$E(X_{\tau\wedge n}) = E(X_n) = E(X_1) \,.$$

No entanto, se considerarmos X_n no momento exacto do tempo de paragem,

$$X_\tau(\omega) = \begin{cases} X_{\tau(\omega)}(\omega) & \text{se } \tau(\omega) < \infty \\ 0 & \text{caso contrário} \end{cases} ,$$

então nem sempre $E(X_\tau) = E(X_1)$. Como exemplo veja-se o caso da estratégia de martingala. O teorema de paragem opcional estabelece condições suficientes para que a propriedade da martingala se estenda a tempos de paragem, isto é,

$$E(X_\tau) = E(X_1)\,.$$

Antes de enunciar o teorema, necessitamos de uma definição técnica.

Definição 5.8. Uma sucessão de variáveis aleatórias $(Y_n)_{n\geq 1}$ diz-se **uniformemente integrável** sse existir uma função integrável $g : \Omega \to [0, \infty]$ tal que $|Y_n| \leq g$ para todo $n = 1, 2, \ldots$.

Estamos em condições de enunciar o teorema.

Teorema 5.17 (da paragem opcional de Doob)**.** *Seja $(X_n)_{n\geq 1}$ uma martingala relativamente a uma filtração $(\mathcal{F}_n)_{n\geq 1}$ e τ um tempo de paragem. Se $P(\tau < \infty) = 1$ e $(X_{\tau\wedge n})_{n\geq 1}$ é uniformemente integrável então $E(X_\tau) = E(X_1)$.*

Ou seja, não é possível alterar um jogo a nosso favor usando uma estratégia de paragem que satisfaça as condições do teorema. Antes de comentarmos acerca das condições do teorema vejamos a sua demonstração.

Demonstração. Uma vez que $P(\tau < \infty) = 1$ segue que

$$\lim_{n\to\infty} X_{\tau\wedge n} = X_\tau\,, \quad P\text{-q.c.}$$

De facto, dado $\omega \in \Omega$ tal que $\tau(\omega) < \infty$, existe $k \geq 1$ tal que $X_{\tau(\omega)\wedge n}(\omega) = X_\tau(\omega)$ para todo $n \geq k$. Logo, segue do Teorema da convergência dominada que

$$E(X_\tau) = \lim_{n\to\infty} E(X_{\tau\wedge n}) = \lim_{n\to\infty} E(X_1) = E(X_1)\,.$$

$\square$

Das duas condições do teorema da paragem opcional a mais difícil de verificar é $X_{\tau\wedge n}$ uniformemente integrável. Contudo existem alguns casos em que podemos verificar essa condição com facilidade.

Proposição 5.18. *A martingala $(X_{\tau\wedge n})_{n\geq 1}$ é uniformemente integrável se alguma das seguintes condições se verificar:*

1. *A sucessão $(X_{\tau\wedge n})_{n\geq 1}$ é limitada, ou seja, existe um $M \geq 0$ tal que $|X_{\tau\wedge n}| \leq M$ para todo $n \geq 1$.*

2. O tempo de paragem é limitado: $P(\tau \leq k) = 1$ para algum $k \geq 1$.

3. $E(\tau) < \infty$ e existe um $M > 0$ tal que

$$\forall\, k \geq 1 \quad E(|X_{k+1} - X_k|\,|\mathcal{F}_k) \leq M\,.$$

Demonstração.

1. Segue directamente da Definição 5.8.
2. Se $\tau \leq k$ P-q.c. então

$$|X_{\tau\wedge n}| \leq \max\left\{|X_1|\,, \ldots, |X_k|\right\}\,.$$

Logo, $X_{\tau\wedge n}$ é uniformemente integrável.

3. Seja,

$$W = |X_1 - X_0| + |X_2 - X_1| + \ldots + |X_\tau - X_{\tau-1}|\,,$$

onde por conveniência se define $X_0 = 0$ e $\mathcal{F}_0 = \{\emptyset, \Omega\}$ a σ-álgebra trivial. Segue da definição de $X_{\tau\wedge n}$ que

$$|X_{\tau\wedge n}| \leq W\,, \quad \forall\, n = 1, 2, \ldots$$

Basta mostrar que W é integrável, ou seja, $E(W) < \infty$. Em primeiro lugar note-se que,

$$W = \sum_{k=0}^{\infty} |X_{k+1} - X_k|\, \chi_{\{\tau \geq k+1\}}\,.$$

Por outro lado, como $E(|X_{k+1} - X_k|\, \mathcal{F}_k) \leq M$ temos que

$$\begin{aligned} E(|X_{k+1} - X_k|\, \chi_{\{\tau\geq k+1\}} |\mathcal{F}_k) &= \chi_{\{\tau\geq k+1\}} E(|X_{k+1} - X_k|\,|\mathcal{F}_k) \\ &\leq M \chi_{\{\tau\geq k+1\}}\,. \end{aligned}$$

Logo, $E(|X_{k+1} - X_k|\, \chi_{\tau\geq k+1}) \leq MP(\{\tau \geq k+1\})$. Assim temos que,

$$\begin{aligned} E(W) = \sum_{i=0}^{\infty} E(|X_{k+1} - X_k|\, \chi_{\{\tau\geq k+1\}}) &\leq \sum_{i=0}^{\infty} MP(\{\tau \geq k+1\}) \\ &= ME(\tau) < \infty\,. \end{aligned}$$

□

Exemplo 5.19 (O jogo da ruína)**.** Retomemos o jogo do lançamento repetitivo de uma moeda perfeita. Seja $X_1, X_2, \ldots$ variáveis aleatórias IID tais que $X_n \in \{-1, 1\}$ com igual probabilidade. Então o passeio aleatório,

$$S_n = X_1 + \ldots + X_n \,,$$

é uma martingala relativamente a $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$. Considere-se o tempo de paragem,

$$\tau = \min \{n \in \mathbb{N} \colon S_n \in \{-a, b\}\} \,,$$

onde $0 < a, b$. Ou seja, o jogo termina quando o jogador obtiver um prejuízo de a euros ou um lucro de b euros. Desejamos calcular a probabilidade $p_a = P(S_\tau = -a)$, isto é, a probabilidade de o jogador acabar o jogo com um prejuízo de a euros em vez de um lucro de b euros.

Suponhamos que as condições de aplicabilidade do teorema da paragem opcional se verificam, isto é,

- $P(\tau < \infty) = 1$
- $S_{\tau \wedge n}$ é uniformemente integrável

Então, do teorema da paragem opcional segue que $E(S_\tau) = E(S_1) = 0$. Por outro lado,

$$0 = E(S_\tau) = -a p_a + b(1 - p_a) \,.$$

Logo,

$$p_a = \frac{b}{b + a} \,.$$

Portanto, se o jogador tiver a euros e estiver disposto jogar enquanto não os perder, então a probabilidade de ruína é,

$$\lim_{b \to \infty} p_a = \lim_{b \to \infty} \frac{b}{b + a} = 1 \,.$$

Verifiquemos agora as condições do teorema.

- Uma vez que $X_{\tau \wedge n} \in [-a, b]$ logo $X_{\tau \wedge n}$ é uma sucessão limitada. Portanto uniformemente integrável.
- Demonstremos $P(\tau < \infty) = 1$. Seja $l = a + b$ o comprimento do intervalo $[-a, b]$ e A_k o acontecimento,

  $$A_k = \left\{X_{(k-1)l+1} = X_{(k-1)l+2} = \cdots = X_{kl} = 1\right\} \,, \quad k = 1, 2, \ldots$$

  Note-se que $P(A_k) = \frac{1}{2^l}$ para todo $k = 1, 2, \ldots$. Se A_1 ocorrer então o jogo terminou antes de se chegarem às l jogadas, uma vez que $S_l > b$. Logo, $A_1 \subset \{\tau \leq l\}$ e segue que

  $$P(\tau > l) \leq P(A_1^c) = 1 - \frac{1}{2^l} \,.$$

Se no entanto $A_1 \cup A_2$ ocorrer então $S_m = b$ para algum $m \in \{1, \ldots, 2l\}$. Logo o jogo terminou antes de se chegar à $2l$-jogada. Logo, $A_1 \cup A_2 \subset \{\tau \leq 2l\}$. Portanto,

$$P(\tau > 2l) \leq P((A_1 \cup A_2)^c) = P(A_1^c)P(A_2^c) = \left(1 - \frac{1}{2^l}\right)^2 .$$

Generalizando este argumento obtemos que,

$$P(\tau > kl) \leq P\left(\bigcap_{i=1}^{k} A_i^c\right) = \left(1 - \frac{1}{2^l}\right)^k , \quad k = 1, 2, \ldots$$

Logo,

$$P(\tau = \infty) = \lim_{k \to \infty} P(\tau > kl) = 0 .$$

Proposição 5.20 (Equação de Wald)**.** *Seja X_n uma sucessão de variáveis aleatórias IID com $E(|X_n|) \leq M$ para algum $M \geq 0$ e τ um tempo de paragem relativamente à filtração $\sigma(X_1, \ldots, X_n)$ tal que $P(\tau < \infty) = 1$ e $E(\tau) < \infty$. Então*

$$E(\sum_{i=1}^{\tau} X_i) = \mu E(\tau) ,$$

onde $\mu = E(X_n)$ para todo $n = 1, 2, \ldots$.

Demonstração. Seja $Z_n = \sum_{i=1}^{n} X_i - n\mu$. Uma vez que Z_n é uma martingala então se o teorema da paragem opcional for aplicável temos que

$$0 = E(Z_\tau) = E(\sum_{i=1}^{\tau} X_i - \tau\mu) = E(\sum_{i=1}^{\tau} X_i) - \mu E(\tau) .$$

Logo, resta verificar as condições do teorema. De facto, usando a Proposição 5.18 basta provar que existe um $M \geq 0$ tal que,

$$E(|Z_{n+1} - Z_n| \, |\mathcal{F}_n) \leq M , \quad \forall n = 1, 2, \ldots .$$

Mas,

$$\begin{aligned} E(|Z_{n+1} - Z_n| \, |\mathcal{F}_n) &= E(|X_{n+1} - \mu| \, |\mathcal{F}_n) \\ &= E(|X_{n+1} - \mu|) \\ &\leq E(|X_{n+1}|) + \mu \\ &\leq M + \mu < \infty . \end{aligned}$$

□

Exemplo 5.21. Seja $X_1, X_2, \dots$ variáveis aleatórias IID tais que $X_n \in \{0,1\}$ e $\alpha = P(X_n = 1)$. Forme-se a soma,

$$S_n = X_1 + \cdots + X_n\,,$$

e considere-se o tempo de paragem,

$$\tau_k = \min\{n \in \mathbb{N}\colon S_n = k\}\ .$$

Uma vez que $P(\tau_k < \infty)$ temos pela equação de Wald que

$$k = E(S_{\tau_k}) = \alpha E(\tau_k)\,.$$

Logo, $E(\tau_k) = \frac{k}{\alpha}$.

Exemplo 5.22. Retomando o Exemplo 5.19, desejamos calcular $E(\tau)$. A equação de Wald não ajuda uma vez que $E(S_\tau) = 0$ e $\mu = 0$. No entanto se considerarmos uma nova sucessão $Z_n = S_n^2 - n$ é possível calcular $E(\tau)$. Note-se que Z_n é uma martingala, logo $Z_{\tau\wedge n}$ é também uma martingala. Logo,

$$0 = E(Z_1) = E(Z_{\tau\wedge n}) = E(S^2_{\tau\wedge n} - \tau\wedge n)\,.$$

Portanto,

$$E(S^2_{\tau\wedge n}) = E(\tau\wedge n)\,.$$

Como $\tau\wedge n \nearrow \tau$ e $S_{\tau\wedge n} \to S_\tau$ onde $S_{\tau\wedge n}$ é uma sucessão limitada, então pelo teorema da convergência monótona e da convergência limitada temos que

$$E(S_\tau^2) = E(\tau)\,.$$

Mas como, $E(S_\tau^2) = a^2\frac{b}{b+a} + b^2(1 - \frac{b}{b+a}) = ab$, logo,

$$E(\tau) = ab\,.$$

Teorema 5.23. *Seja $(X_n)_{n\geq 1}$ uma supermartingala (submartingala) relativamente a uma filtração $(\mathcal{F}_n)_{n\geq 1}$ e τ um tempo de paragem. Se $P(\tau < \infty) = 1$ e $(X_{\tau\wedge n})_{n\geq 1}$ é uniformemente integrável então $E(X_\tau) \leq E(X_1)$ $(E(X_\tau) \geq E(X_1))$.*

Demonstração. Deixa-se como exercício. □

5.1.4 Convergência de martingalas

O resultado que se segue estabelece a convergência de uma martingala com probabilidade 1 desde que a sucessão $E(|X_n|)$ seja limitada. Sua demonstração pode ser vista em [1].

Teorema 5.24. *Seja X_n uma martingala tal que $E(|X_n|) \leq M$ para todo $n = 1, 2, \ldots$ e algum $M \geq 0$. Então existe uma variável aleatória X_∞ integrável tal que*

$$X_\infty = \lim_{n\to\infty} X_n\,, \quad P - q.c.$$

Como corolário tem-se o seguinte resultado.

Corolário 5.25. *Se X_n é uma martingala não negativa então existe uma variável aleatória X_∞ tal que*

$$X_\infty = \lim_{n\to\infty} X_n\,, \quad P - q.c.$$

Demonstração. Uma vez que X_n é não negativa tem-se

$$E(|X_n|) = E(X_n) = E(X_1)\,, \quad \forall n = 1, 2, \ldots$$

O resultado segue do teorema anterior. □

Exemplo 5.26. Considere novamento o jogo honesto,

$$S_n = X_1 + \ldots + X_n\,,$$

onde X_n é uma sucessão de variáveis aleatórias IID tais que $X_n \in \{-1, 1\}$ com igual probabilidade. Suponha que o jogador dispõe de crédito limitado, a euros. Portanto, é obrigado a abandonar o jogo quando o seu prejuízo atingir esse limite,

$$\tau = \min\{n \geq 1 : S_n = -a\}\ .$$

Logo, a martingala $S_{\tau\wedge n}+a$ é não negativa e segue do corolário anterior que

$$S_\tau = \lim_{n\to\infty} S_{\tau\wedge n}\,, \quad P - \text{q.c.}$$

Uma vez que

$$|S_{n+1} - S_n| = 1\,, \quad \forall n = 1, 2, 3, \ldots$$

então τ é finito P-q.c., caso contrário a sucessão $S_{\tau\wedge n}$ não seria convergente P-q.c.

Conclui-se assim que com probabilidade 1, o jogador será obrigado a abandonar o jogo ao fim de um número finito de jogadas.

Exemplo 5.27. Sejam $X_1, X_2, \ldots$ variáveis aleatórias IID tais que $X_n \in \{0, 2\}$ com igual probabilidade. Uma vez que $E(X_n) = 1$ então

$$Z_n = \prod_{i=1}^{n} X_i$$

é uma martingala. Como $Z_n \geq 0$ então aplicando o corolário anterior tem-se

$$Z_\infty = \lim_{n \to \infty} Z_n \,, \quad P - \text{q.c.}$$

De facto, $Z_\infty = 0$, P-q.c.

6 Processos Estocásticos

6.1 Definições gerais

Definição 6.1. Seja $(\Omega, \mathcal{F}, P)$ um espaço de probabilidade. Um **processo estocástico** X é uma colecção $X = \{X_t : t \in T\}$ de variáveis aleatórias $X_t : \Omega \to \mathbb{R}$ indexadas por um parâmetro $t \in T \subset \mathbb{R}$.

Quando o **conjunto dos parâmetros** T é um conjunto numerável, tipicamente $T = \{0, 1, 2, \ldots\}$ ou $T = \mathbb{Z}$, então o processo estocástico é de **parâmetro ou tempo discreto**.

Observação 6.1. Neste contexto, uma sucessão de variáveis aleatórias é um processo estocástico de tempo discreto.

Quando T é um intervalo, tipicamente $T = [0, \infty[$ ou $T = \mathbb{R}$, então o processo estocástico é de **parâmetro ou tempo contínuo**. Ao conjunto dos valores que as variáveis aleatórias X_t podem tomar designa-se por E, o **conjunto dos estados do processo estocástico**. Quando o conjunto dos estados E é finito ou numerável então o processo estocástico diz-se **discreto**, caso contrário diz-se **contínuo**.

Observação 6.2. É usual escrever-se $X(t)$ quando o processo é de tempo contínuo.

Para cada $\omega \in \Omega$ a função

$$t \mapsto X_t(\omega) \,, \quad t \in T$$

é designada por **trajectória** ou **realização** do processo. Uma trajectória de um processo referente a um período limitado de tempo é designada por **série temporal**.

A **lei de probabilidade** do processo estocástico é dada por todas as distribuições de probabilidade conjuntas de um número finito de variáveis aleatórias $(X_{t_1}, \ldots, X_{t_n})$. Portanto, a lei de probabilidade do processo consiste na família de funções de distribuição de probabilidade conjuntas,

$$F_{(X_{t_1},\ldots,X_{t_n})}(x_1, \ldots, x_n) = P(X_{t_1} \leq x_1, \ldots, X_{t_n} \leq x_n)\,,$$

onde $t_1, \ldots, t_n$ são quaisquer conjunto finito de índices pertencentes a T e $n \geq 1$.

Definição 6.2. Dois processos estocásticos $X = \{X_t : t \in T\}$ e $Y = \{Y_t : t \in T\}$ dizem-se **identicamente distribuídos** sse tiverem a mesma família de funções de distribuição de probabilidade conjuntas, ou seja,

$$F_{(X_{t_1},\ldots,X_{t_n})} = F_{(Y_{t_1},\ldots,Y_{t_n})}$$

para todo o conjunto finito de índices $t_1, \ldots, t_n \in T$ e $n \in \mathbb{N}$.

Definição 6.3. Dois processos estocásticos $X = \{X_t : t \in T\}$ e $Y = \{Y_t : t \in T\}$ dizem-se **independentes** sse os vectores aleatórios

$$(X_{t_1}, \ldots, X_{t_n}) \quad \text{e} \quad (Y_{t_1}, \ldots, Y_{t_n})$$

são independentes para quaisquer $t_1, \ldots, t_n \in T$ e $n \in \mathbb{N}$.

Observação 6.3. Note-se que dois vectores aleatórios são independentes sse as suas σ-álgebras induzidas são independentes.

Exemplo 6.4. Considere uma sucessão $(X_n)_{n\geq 1}$ de variáveis aleatórias independentes. O processo estocástico de tempo discreto

$$S = \{S_n = X_1 + \cdots + X_n \colon n \in \mathbb{N}\}\ ,$$

é designado por passeio aleatório. A variável aleatória S_n por ser interpretada como o deslocamento até ao instante n, podendo ser escrita

$$S_n = S_{n-1} + X_n\,.$$

6.2 Estacionariedade

Em termos gerais, um processo estacionário é um processo cujas características de aleatoriedade não se alteram ao longo do tempo. Existem diversas definições de estacionariedade.

Definição 6.4. Um processo X_t diz-se **estacionário em média** sse

$$E(X_t) = \mu\,, \quad \forall\, t \in T\,.$$

Num processo estacionário em média o seu valor esperado não evolui ao longo do tempo.

Definição 6.5. Um processo X_t diz-se de **covariâncias estacionárias** sse

1. Os momentos de segunda ordem são finitos, isto é
$$E(X_t^2) < \infty\,, \quad \forall t \in T\,,$$
2. Existe uma função $\gamma : T \to \mathbb{R}$ tal que
$$\mathrm{Cov}(X_s, X_t) = \gamma(t-s)\,, \quad \forall s \le t\,.$$

Observação 6.5. Num processo X_t com covariâncias estacionárias tem-se
$$\mathrm{Var}(X_t) = \gamma(0)\,.$$
Ou seja, a variância do processo não se altera ao longo do tempo.

Definição 6.6. Um processo X_t diz-se **estacionário até à segunda ordem** sse for estacionário em média e tiver covariâncias estacionárias.

Uma definição de estacionariedade mais forte que as anteriores é a seguinte.

Definição 6.7. Um processo X_t diz-se **fortemente ou estritamente estacionário de ordem** $k \in \mathbb{N}$ sse para quaisquer instantes $t_1, \ldots, t_k \in T$ e qualquer h tal que $t_i + h \in T$, $i = 1, \ldots, k$, os vectores aleatórios $(X_{t_1}, \ldots, X_{t_k})$ e $(X_{t_1+h}, \ldots, X_{t_k+h})$ são identicamente distribuídos.

Chegamos assim à definição de estacionariedade mais forte.

Definição 6.8. Um processo é **fortemente ou estritamente estacionário** sse for fortemente estacionário de ordem k para qualquer $k \in \mathbb{N}$.

Esta definição de estacionariedade implica todas as outras. Adicionalmente, num processo fortemente estacionário todas as variáveis aleatórias X_t tem a mesma distribuição. Analogamente, todos os pares (X_t, X_s) são identicamente distribuídos desde os instantes estejam igualmente espaçados no tempo.

Exemplo 6.6. Qualquer sucessão de variáveis aleatórias IID é um processo fortemente estacionário.

Exemplo 6.7. Considere um processo estocástico $\{X_n : n = 1, 2, \ldots\}$ tal que $E(X_n) = 0$ e $\mathrm{Var}(X_n) = \sigma^2$ para todo $n \ge 1$ e $\mathrm{Cov}(X_n, X_m) = 0$ para $n \ne m$. Este processo é estacionário até à segunda ordem, sendo usualmente designado por **ruído branco**.

Exemplo 6.8. Seja $(X_n)_{n\geq 1}$ uma sucessão de variáveis aleatórias IID, $\theta \in \mathbb{R}$ e $Y_n = X_n - \theta X_{n-1}$. O processo estocástico

$$Y = \{Y_n \colon n \in \mathbb{N}\}$$

é estacionário até à segunda ordem. É usualmente designado por **processo de médias móveis de primeira ordem**, MA(1).

6.3 Processo com incrementos estacionários e independentes

Definição 6.9. Um processo estocástico X_t diz-se de:

- **incrementos independentes** sse para quaisquer instantes $t_1 < t_2 < \cdots < t_n$ as variáveis aleatórias

  $$X_{t_2} - X_{t_1}, X_{t_3} - X_{t_2}, \ldots, X_{t_n} - X_{t_{n-1}}$$

  são independentes.
- **incrementos estacionários** se para quaisquer instantes $t, s \in T$ e $h > 0$ tal que $t+h, s+h \in T$, as variáveis aleatórias

  $$X_{t+h} - X_{s+h} \quad \text{e} \quad X_t - X_s$$

  são identicamente distribuídas.

Exercício 53. Mostre que um processo estocástico $\{X_t : t \in T\}$ tal que $X_0 = 0$ tem incrementos estacionários sse para todo $t \geq s$ tal que $t - s \in T$ as variáveis aleatórias $X_t - X_s$ e X_{t-s} são identicamente distribuídas.

Como o seguinte resultado demonstra é fácil caraterizar os processos estocásticos em tempo discreto com incrementos independentes e estacionários.

Proposição 6.9. *Uma sucessão de variáveis aleatórias* $(X_0, X_1, X_2, \ldots)$ *tem incrementos independentes e estacionários sse existir uma sucessão de variáveis aleatórias* $(U_1, U_2, \ldots)$ *IID tal que*

$$X_n = \sum_{i=1}^{n} U_i \quad n \in \mathbb{N}\,.$$

Demonstração. Se $(X_0, X_1, X_2, \ldots)$ tem incrementos independentes e estacionários então seja $U_n = X_n - X_{n-1}$ para $n = 1, 2, \ldots$. Segue da construção que a sucessão $(U_1, U_2, \ldots)$ é IID e

$$X_n = U_1 + \cdots + U_n\,.$$

Se por outro lado $(U_1, U_2, \ldots)$ é IID então mostremos que $X_n = U_1 + \cdots + U_n$ tem incrementos independentes e estacionários. De facto, dados $n > m$ temos que

$$X_n - X_m = U_{m+1} + \cdots + U_n \,.$$

Uma vez que $U_1, U_2, \ldots$ são IID segue que

$$X_{n+k} - X_{m+k} \quad \text{e} \quad X_n - X_m$$

são independentes. A estacionariedade prova-se de maneira semelhante. $\square$

Num processo estacionário $\{X_t : t \in T\}$ com incrementos independentes e estacionários, se sabemos a função de densidade de probabilidade de X_t para todo $t \in T$ então podemos determinar a função de densidade conjunta $f_{t_1,\ldots,t_n}$ de qualquer vector aleatório $(X_{t_1}, \ldots, X_{t_n})$.

Proposição 6.10. *Seja $X = \{X_t : t \in T\}$ um processo estacionário com incrementos independentes e estacionários tal que $X_0 = 0$ e X_t tem função densidade de probabilidade f_t para todo $t \in T$. Se $t_i \in T$, $i = 1, \ldots, n$, tal que*

$$0 < t_1 < t_2 < \cdots < t_n$$

então

$$f_{t_1,t_2,\ldots,t_n}(x_1, x_2, \ldots, x_n) = f_{t_1}(x_1) f_{t_2-t_1}(x_2-x_1) \cdots f_{t_n-t_{n-1}}(x_n-x_{n-1}) \,.$$

Demonstração. Seja $U_1 = X_{t_1}$ e $U_i = X_{t_i} - X_{t_{i-1}}$ para $i = 2, \ldots, n$. Então

$$(X_{t_1}, X_{t_2}, \ldots, X_{t_n}) = (U_1, U_1 + U_2, \ldots, U_1 + U_2 + \cdots + U_n) \,.$$

Segue de o processo ter incrementos estacionários e independentes que as variáveis aleatórias $U_1, U_2, \ldots, U_n$ são independentes e têm funções de densidade marginais $f_{t_1}, f_{t_2-t_1} \ldots, f_{t_n-t_{n-1}}$. Logo,

$$f_{U_1,\ldots,U_n}(x_1, \ldots, x_n) = f_{t_1}(x_1) f_{t_2-t_1}(x_2) \cdots f_{t_n-t_{n-1}}(x_n) \,.$$

Seja $X = (X_{t_1}, X_{t_2}, \ldots, X_{t_n})$, $U = (U_1, \ldots, U_2)$ e considere a função $h : \mathbb{R}^n \to \mathbb{R}^n$ definida por

$$h(x_1, \ldots, x_n) = (x_1, x_1 + x_2, \ldots, x_1 + x_2 + \cdots + x_n) \,.$$

Segue do teorema da mudança de variável que,

$$\begin{aligned}
\int_{\mathbb{R}^n} g\, dp_X &= \int_{\mathbb{R}^n} g \circ h\, dp_U \\
&= \int_{\mathbb{R}^n} g \circ h \cdot f_U\, dm_n \\
&= \int_{\mathbb{R}^n} g(x_1, x_1 + x_2, \ldots, x_1 + \cdots + x_n) f_U(x_1, \ldots, x_n)\, dm_n \\
&= \int_{\mathbb{R}^n} g(x_1, \ldots, x_n) f_U(x_1, x_2 - x_1, \ldots, x_n - x_{n-1})\, dm_n
\end{aligned}$$

Como a função g é arbitrária concluímos que

$$\frac{dp_X}{dm_n}(x_1, \ldots, x_n) = f_U(x_1, x_2 - x_1, \ldots, x_n - x_{n-1})\,.$$

□

Proposição 6.11. *Seja $X = \{X_t : t \in T\}$ um processo estocástico com incrementos independentes e estacionários tal que $X_0 = 0$ e $E(X_t^2) < \infty$ para todo $t \in T$. Então existem constantes $\mu \in \mathbb{R}$ e $\sigma \geq 0$ tal que*

1. *$E(X_t) = \mu t$ para $t \in T$.*
2. *$\operatorname{Cov}(X_s, X_t) = \sigma^2 \min\{s, t\}$ para todo $t, s \in T^2$.*

Demonstração.

1. Define-se a função $g(t) = E(X_t)$. Note-se que dados $t, s \in T$ tal que $t + s \in T$, as variáveis aleatórias $X_{t+s} - X_s$ e $X_t - X_0$ são identicamente distribuídas. Logo,

$$\begin{aligned}
g(t+s) = E(X_{t+s}) &= E(X_{t+s} - X_s + X_s) \\
&= E(X_{t+s} - X_s) + E(X_s) \\
&= E(X_t - X_0) + E(X_s) \\
&= E(X_t) + E(X_s) \\
&= g(t) + g(s)\,.
\end{aligned}$$

   Portanto, g é uma função aditiva. É possível provar que as únicas funções aditivas e limitadas em intervalos limitados são as funções lineares $g(t) = \mu t$ para algum $\mu \in \mathbb{R}$ (ver [1]). Logo, $E(X_t) = \mu t$.

2. Demonstração análoga à anterior.

□

Exercício 54. Seja $X = \{X_t : t \in T\}$ um processo estocástico com incrementos independentes e estacionários tal que $X_0 = 0$ e $E(X_t^2) < \infty$ para todo $t \in T$. Mostre que existe uma constante positiva σ tal que

$$\operatorname{Var}(X_t - X_s) = \sigma^2 |t - s|\,.$$

6.4 Processo de Markov

Um processo de Markov é um processo estocástico em que o futuro é independente do passado quando se conhece o presente.

Definição 6.10. Um processo estocástico $X = \{X_t \colon t \in T\}$ é um **processo de Markov** se satisfizer a **propriedade de Markov**,

$$P(X_t \in B | X_{t_1}, \dots, X_{t_n}) = P(X_t \in B | X_{t_n}) \, ,$$

sempre que $t_1 < t_2 < \cdots < t_n < t \in T$ e para todo o Boreliano B de $\mathbb{R}$.

Observação 6.12. Note-se que $P(X \in B|Y) = P(X^{-1}(B)|\sigma(Y))$ é a probabilidade condicionada do acontecimento $\{X \in B\}$ dado a informação de Y, como foi definida anteriormente.

Exemplo 6.13 (Passeio aleatório é um processo de Markov)**.** Considere-se o passeio aleatório

$$S_n = X_1 + \cdots + X_n \, ,$$

onde $X_1, X_2, \dots$ são variáveis aleatórias IID tais que $X_i \in \{-1, 1\}$. Uma vez que

$$S_{n+1} = S_n + X_{n+1} \, ,$$

onde X_{n+1} é independente de S_i, $i = 1, \dots, n$ temos que

$$\begin{aligned} P(S_{n+1} = s | S_1 = s_1, \dots, S_n = s_n) &= P(S_n + X_{n+1} = s | S_1 = s_1, \dots, S_n = s_n) \\ &= P(X_{n+1} = s - s_n | S_1 = s_1, \dots, S_n = s_n) \\ &= P(X_{n+1} = s - s_n) \, . \end{aligned}$$

Por outro lado,

$$\begin{aligned} P(S_{n+1} = s | S_n = s_n) &= P(S_n + X_{n+1} = s | S_n = s_n) \\ &= P(X_{n+1} = s - s_n | S_n = s_n) \\ &= P(X_{n+1} = s - s_n) \, . \end{aligned}$$

Logo,

$$P(S_{n+1} = s | S_1 = s_1, \dots, S_n = s_n) = P(S_{n+1} = s | S_n = s_n)$$

de onde se pode verificar com facilidade a propriedade de Markov.

Generalizando o exemplo anterior obtem-se o seguinte resultado.

Proposição 6.14. *Um processo estocástico $X = \{X_t : t \in T\}$ de incrementos independentes com $T = [0, \infty[$ ou $T = \{0, 1, 2, \dots\}$ é um processo de Markov.*

Demonstração. Deixa-se como exercício. □

Exemplo 6.15. No entanto existem processos não Markovianos, como é o caso do **processo autoregressivo de ordem** 2, designado por AR(2),

$$Y_n = a_1 Y_{n-1} + a_2 Y_{n-2} + X_n \,,$$

onde X_n é uma sucessão de variáveis IID.

Quando o conjunto de estados do processo de Markov X é finito ou numerável então diz-se que X é uma **cadeia de Markov**. Note-se que dependendo da natureza do conjunto dos parâmetros, uma cadeia de Markov pode ser a **tempo discreto ou contínuo**. Quando o conjunto de estados e dos parâmetros é um intervalo diz-se que X é um **processo de difusão**.

6.5 Processo de Wiener

O processo de Wiener é um processo de difusão usado para modelar o **movimento Browniano**, isto é, o movimento descrito por uma partícula imersa num líquido quando esta colide com as moléculas do líquido. O processo de Wiener é também amplamente utilizado em matemática financeira.

Definição 6.11. Um processo estocástico a tempo contínuo $W = \{W_t : t \in [0, \infty[\}$ diz-se um **processo de Wiener** sse:

1. $W_0 = 0$, P-q.c.
2. as trajectórias $t \mapsto W_t$ são contínuas P-q.c.
3. para quaisquer $0 < t_1 < \cdots < t_n$ o vector aleatório $(W_{t_1}, \ldots, W_{t_n})$ tem função de densidade conjunta,

   $$f_{t_1,\ldots,t_n}(x_1, \ldots, x_n) = p_{t_1}(x_1) p_{t_2-t_1}(x_2 - x_1) \cdots p_{t_n - t_{n-1}}(x_n - x_{n-1}) \,,$$

   onde

   $$p_t(x) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}} \,,$$

   é a função de densidade da distribuição Gaussiana.

Exercício 55. Mostre que a função de densidade de W_t é

$$f_t(x) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}} \,,$$

e calcule $E(W_t)$ e $\mathrm{Var}(W_t)$.

Proposição 6.16. *Para quaisquer* $0 \leq s < t$ *o incremento* $W_t - W_s$ *tem distribuição normal com valor esperado* 0 *e variância* $t - s$.

Demonstração. Segue da terceira condição da definição do processo de Wiener que a função de densidade conjunta de (W_s, W_t) é

$$f_{s,t}(x,y) = p_s(x) p_{t-s}(y-x)\,.$$

Logo,

$$\begin{aligned} P(W_t - W_s \leq z) &= \int_{\{y-x\leq z\}} p_s(x)p_{t-s}(y-x)\,dxdy \\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{z+x} p_s(x)p_{t-s}(y-x)\,dydx \\ &= \int_{-\infty}^{z}\int_{-\infty}^{+\infty} p_s(x)p_{t-s}(y)\,dxdy \\ &= \int_{-\infty}^{z} p_{t-s}(y)\,dy \end{aligned}$$

Logo,

$$f_{t-s}(y) = \frac{1}{\sqrt{2\pi(t-s)}} e^{-\frac{y^2}{2(t-s)}}$$

□

Proposição 6.17. *Um processo de Wiener tem incrementos independentes.*

Demonstração. Para quaisquer $0 < t_1 < t_2 < \cdots < t_n$ queremos mostrar que os incrementos,

$$W_{t_1} - W_0, W_{t_2} - W_{t_1}, \ldots, W_{t_n} - W_{t_{n-1}}$$

são independentes. É possível demonstrar que se $X_1, \ldots, X_n$ são variáveis aleatórias que seguem uma distribuição normal então são independentes sse não são correlacionadas, ou seja, $\mathrm{Cov}(X_i, X_j) = 0$ para todo $i \neq j$. Portanto, como o valor esperado dos incrementos é nulo, é suficiente mostrar que,

$$E\left[(W_u - W_t)(W_s - W_r)\right] = 0\,, \quad t \leq u \leq r \leq s\,.$$

Mas para $s \leq t$ temos que,

$$\begin{aligned} E(W_s W_t) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyp_s(x)p_{t-s}(y-x)\,dxdy \\ &= \int_{-\infty}^{+\infty} xp_s(x)\left(\int_{-\infty}^{+\infty} yp_{t-s}(y-x)\,dy\right)dx \\ &= \int_{-\infty}^{+\infty} x^2 p_s(x)\,dx = s\,. \end{aligned}$$

Logo,

$$\begin{aligned} E\left[(W_u - W_t)(W_s - W_r)\right] &= E(W_u W_s) - E(W_u W_r) - E(W_t W_s) + E(W_t W_r) \\ &= u - u - t + t = 0 \end{aligned}$$

$\square$

Segue das proposições anteriores que,

Corolário 6.18. *Um processo de Wiener tem incrementos independentes e estacionários.*

Exercício 56. Mostre que um processo de Wiener é estacionário em média, mas não tem covariâncias estacionárias.

Referências

[1] Robert B. Ash, Catherine A. Doléans-Dade, *Probability and Measure Theory*. Academic Press 2nd Edition, 1999.

[2] Daniel Müller, *Probabilidade e processos estocásticos: uma abordagem rigorosa com vista aos modelos em finanças*. Coimbra, Almedina, 2011.