Medida e Probabilidade

Eduardo Horta

1 Espaços mensuráveis

Seja X um conjunto. Dizemos que uma coleção $\mathscr{X}$ de subconjuntos de X é uma σ–**álgebra em** X quando as seguintes propriedades são satisfeitas:

A1. $\varnothing \in \mathscr{X}$;

A2. se $A \in \mathscr{X}$, então $A^{\complement} \in \mathscr{X}$.

A3. se $A_n \in \mathscr{X}$ para todo $n \in \mathbb{N}$, então $\bigcup_{n=1}^{\infty} A_n \in \mathscr{X}$.

Note que, por definição, $\mathscr{X} \subset \mathscr{P}(X)$.

Se X é um conjunto e $\mathscr{X}$ é uma σ–álgebra em X, dizemos que o par $(X, \mathscr{X})$ é um **espaço mensurável**. Um subconjunto A de X pode ou não pertencer à σ–álgebra $\mathscr{X}$. Se $A \in \mathscr{X}$, dizemos que A é um conjunto $\mathscr{X}$–**mensurável**. Quando a σ–álgebra $\mathscr{X}$ estiver subentendida, dizemos simplesmente que A é um conjunto **mensurável**.

Exercício 1.1. Seja X um conjunto. Verifique que $\mathscr{X}$ é σ–álgebra em cada um dos seguintes casos:

(i) $\mathscr{X} = \mathscr{P}(X)$;

(ii) $\mathscr{X} = \{\varnothing, X\}$. //

Exercício 1.2. Seja $(X, \mathscr{X})$ um espaço mensurável. Verifique que, se $A_n \in \mathscr{X}$ para todo $n \in \mathbb{N}$, então $\bigcap_{n=1}^{\infty} A_n \in \mathscr{X}$. //

Exercício 1.3. Sejam $\mathscr{X}_1$ e $\mathscr{X}_2$ duas σ–álgebras em X. Verifique que o conjunto

$$\mathscr{X} := \mathscr{X}_1 \cap \mathscr{X}_2 = \{A \subset X : A \in \mathscr{X}_1 \wedge A \in \mathscr{X}_2\}$$

é uma σ–álgebra em X. //

Exercício 1.4. Sejam X e J conjuntos não-vazios. Verifique a seguinte afirmação: se para todo $\alpha \in J$ tem-se que $\mathscr{X}_\alpha$ é σ–álgebra em X, então $\bigcap_{\alpha\in J} \mathscr{X}_\alpha$ é também σ–álgebra em $\mathscr{X}$. //

Seja X um conjunto e considere uma classe $\mathscr{C}$ de subconjuntos de X. Defina $\sigma(\mathscr{C})$ como sendo a intersecção de todas as σ–álgebras que contém $\mathscr{C}$, isto é,

$$\sigma(\mathscr{C}) := \bigcap \{\mathscr{X} : \mathscr{X} \text{ é } \sigma\text{–álgebra em } X \wedge \mathscr{C} \subset \mathscr{X}\}$$

Assim, $\sigma(\mathscr{C})$ é unicamente determinada pelas seguintes propriedades:

(i) $\sigma(\mathscr{C})$ é uma σ–álgebra em X;

(ii) $\mathscr{C} \subset \sigma(\mathscr{C})$;

(iii) se $\mathscr{X}$ é qualquer outra σ–álgebra em X tal que $\mathscr{C} \subset \mathscr{X}$, então $\sigma(\mathscr{C}) \subset \mathscr{X}$.

Dizemos que $\sigma(\mathscr{C})$ é a *menor σ–álgebra em X contendo a classe $\mathscr{C}$*, ou então que $\sigma(\mathscr{C})$ é a σ–álgebra **gerada** pela classe $\mathscr{C}$.

Exemplo 1.1. Seja $X = \mathbb{R}^d$. Seja $\mathscr{U}$ a classe de subconjuntos de X formada pelos conjuntos do tipo $A_1 \times \cdots \times A_d$, onde cada A_j é um intervalo aberto da reta. A σ–álgebra $\sigma(\mathscr{U})$ é chamada a **σ–álgebra de Borel** em $\mathbb{R}^d$. Notação $\sigma(\mathscr{U}) =: \mathscr{B}(\mathbb{R}^d)$. Um conjunto $\mathscr{B}(\mathbb{R}^d)$–mensurável é dito um **Boreliano** de $\mathbb{R}^d$. //

Em probabilidade, conjuntos mensuráveis são chamadas de **eventos aleatórios**. Mais precisamente, se $(X, \mathscr{X})$ é um espaço mensurável e $A \in \mathscr{X}$, então A é dito **evento aleatório** em X. O evento aleatório X é dito o **espaço amostral**, ou **evento certo**.

2 Funções mensuráveis

Sejam $(X, \mathscr{X})$ e $(Y, \mathscr{Y})$ dois espaços mensuráveis. Uma função $f\colon X \to Y$ é dita $\mathscr{X}/\mathscr{Y}$–**mensurável** se satisfizer a seguinte condição:

(F1) $$f^{-1}(B) \in \mathscr{X}, \qquad \forall B \in \mathscr{Y}.$$

Quando as σ–álgebras estiverem subentendidas, diremos simplesmente que f é uma função **mensurável**. No caso em que $Y = \mathbb{R}^d$, sempre vamos considerar a σ–álgebra de Borel $\mathscr{B}(\mathbb{R}^d)$.

Uma palavra sobre notação: lembre que $f^{-1}(B)$ é, por definição, o conjunto $\{x \in X : f(x) \in B\}$. Ocasionalmente usaremos a notação $[f \in B]$ para denotar esse conjunto.

Exercício 2.1. Seja $(X, \mathscr{X})$ um espaço mensurável, e considere um subconjunto $A \subset X$ (não necessariamente mensurável). Defina a função $\mathbb{I}_A : X \to \mathbb{R}$ por

$$\mathbb{I}_A(x) := \begin{cases} 1, & x \in A \\ 0, & x \notin A. \end{cases}$$

A função $\mathbb{I}_A$ é dita a **função indicadora** do conjunto A. Utilize as definições para concluir que conjuntos unitários são Borelianos em $\mathbb{R}$; em particular tem-se $\{1\} \in \mathscr{B}(\mathbb{R})$ e $\{0\} \in \mathscr{B}(\mathbb{R})$. Utilize esse fato para mostrar que o conjunto A é mensurável se, e somente se, a função $\mathbb{I}_A$ é mensurável. //

Exercício 2.2. Uma função $f\colon X \to \mathbb{R}^d$ é dita **simples** se for mensurável e se a imagem de X por f for um subconjunto finito de $\mathbb{R}^d$, isto é, se existirem um número natural $k \in \mathbb{N}$, e elementos *distintos* $y_1, \dots y_k \in \mathbb{R}^d$ tais que $\operatorname{ran}(f) = \{y_1, \dots, y_k\}$. Note que cada y_j é um ponto em $\mathbb{R}^d$ e portanto $y_j = (y_{j1}, \dots, y_{jd})$. Considere, então, uma função simples $f\colon X \to \mathbb{R}^d$, e defina

$$A_j := f^{-1}(\{y_j\}) = \{x \in X : f(x) = y_j\}, \qquad j = 1, \dots k.$$

Mostre que a coleção formada pelos conjuntos A_j é uma partição de X. Conclua que podemos escrever

$$f(x) = \sum_{j=1}^{k} y_j \mathbb{I}_{A_j}(x). \tag{2.1}$$

Essa é chamada a **representação padrão** da função simples f. //

Exercício 2.3. Sejam f e g funções simples de X em $\mathbb{R}^d$, e seja $\alpha \in \mathbb{R}$. Mostre que as funções $x \mapsto f(x) + g(x)$ e $x \mapsto \alpha f(x)$ são simples. //

Exercício 2.4. Seja $f\colon X \to Y$ uma função $\mathscr{X}/\mathscr{Y}$–mensurável. Defina $\sigma(f) := \left\{f^{-1}(B) : B \in \mathscr{Y}\right\}$. Mostre que $\sigma(f)$ é uma σ–álgebra em X. $\sigma(f)$ é dita a **σ–álgebra gerada por** f, e é a menor σ–álgebra em X que torna f uma função mensurável. Note que, escrevendo $\mathscr{C} = \left\{f^{-1}(B) : B \in \mathscr{Y}\right\}$, temos que $\sigma(f) = \sigma(\mathscr{C}) = \mathscr{C}$. //

Exemplo 2.1. Seja J um conjunto não–vazio, e considere uma família indexada $(f_\alpha : \alpha \in J)$ de funções $\mathscr{X}/\mathscr{Y}$–mensuráveis $f_\alpha\colon X \to Y$. Definimos $\sigma(f_\alpha : \alpha \in J)$ como sendo a σ–álgebra gerada pela classe $\mathscr{C} := \left\{f_\alpha^{-1}(B) : \alpha \in J,\ B \in \mathscr{Y}\right\}$. Se J é um conjunto unitário, este exemplo se reduz ao Exercício 2.4, mas note que em geral a classe $\mathscr{C}$ *não é* uma σ–álgebra.

Em probabilidade, funções mensuráveis são chamadas de **variáveis aleatórias**. Mais precisamente, se $(X, \mathscr{X})$ e $(Y, \mathscr{Y})$ são dois espaços mensuráveis e $f\colon X \to Y$ é uma aplicação $\mathscr{X}/\mathscr{Y}$–mensurável, então dizemos que f é **variável aleatória** com valores em Y. Equivalentemente, f é dita um **elemento aleatório** em Y. No caso em que $(Y, \mathscr{Y}) = (\mathbb{R}, \mathscr{B}(\mathbb{R}))$, dizemos que f é **variável aleatória real**. Se $(Y, \mathscr{Y}) = \big(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d)\big)$, com $d \geq 2$, diz-se que f é **vetor aleatório**.

2.1 Caracterizações de funções mensuráveis

A definição de função mensurável a partir da condição (F1) nem sempre é conveniente, no sentido de que o elemento genérico $B \in \mathscr{Y}$, para o qual tal condição deve valer, pode não ter uma "forma conhecida". Por exemplo, não há uma caracterização simples de um conjunto Boreliano em $\mathbb{R}^d$, exceto que sabemos que tal conjunto é elemento da σ–álgebra gerada pela coleção de subconjuntos de $\mathbb{R}^d$ do tipo $A_1 \times \cdots \times A_d$, onde cada A_j é um intervalo aberto em $\mathbb{R}$. Os resultados seguintes tornam as coisas um pouco mais fáceis.

Lema 2.1. *Sejam $(X, \mathscr{X})$ e $(Y, \mathscr{Y})$ dois espaços mensuráveis, onde $\mathscr{Y} = \sigma(\mathscr{C})$ para alguma classe $\mathscr{C}$ de subconjuntos de Y. Seja $f\colon X \to Y$. Se f satisfaz a condição*

$$\text{(F2)} \qquad f^{-1}(B) \in \mathscr{X}, \qquad \forall B \in \mathscr{C},$$

então f é $\mathscr{X}/\mathscr{Y}$–mensurável.

No caso em que $Y = \mathbb{R}$, temos o seguinte.

Corolário 2.1. *Sejam $(X, \mathscr{X})$ um espaço mensurável, e $f\colon X \to \mathbb{R}$. São equivalentes as seguintes propriedades:*

(i) f é mensurável;

(ii) para todo $\alpha \in \mathbb{R}$ tem-se $\{x \in X : f(x) > \alpha\} \in \mathscr{X}$.

(iii) para todo $\alpha \in \mathbb{R}$ tem-se $\{x \in X : f(x) \leq \alpha\} \in \mathscr{X}$.

(iv) para todo $\alpha \in \mathbb{R}$ tem-se $\{x \in X : f(x) \geq \alpha\} \in \mathscr{X}$.

(v) para todo $\alpha \in \mathbb{R}$ tem-se $\{x \in X : f(x) < \alpha\} \in \mathscr{X}$.

Lema 2.2. *Seja $(X, \mathscr{X})$ um espaço mensurável, e considere uma função mensurável não-negativa $f\colon X \to \mathbb{R}$. Então existe uma sequência de funções simples $f_n : X \to \mathbb{R}$ tais que*

(i) $0 \leq f_n(x) \leq f_{n+1}(x) \leq f(x)$ para todo $x \in X$ e $n \in \mathbb{N}$;

(ii) $f_n \to f$ *pontualmente.*

Corolário 2.2. *Seja* $(X, \mathscr{X})$ *um espaço mensurável, e considere uma função mensurável* $f\colon X \to \mathbb{R}$*. Então existe uma sequência de funções simples* $f_n : X \to \mathbb{R}$ *tais que*

(i) $|f_n(x)| \leq |f(x)|$ *para todo* $x \in X$ *e* $n \in \mathbb{N}$*;*

(ii) $f_n \to f$ *pontualmente.*

Seja X um conjunto e considere uma função $f\colon X \to \mathbb{R}^d$. Se $x \in X$, então $f(x)$ é um ponto em $\mathbb{R}^d$, isto é,

$$f(x) = (f_1(x), \ldots, f_d(x)).$$

Da expressão acima segue que f é unicamente determinada pelas funções $f_1, \ldots, f_d$, com $f_j : X \to \mathbb{R}$. O seguinte Lema relaciona mensurabilidade de f com a mensurabilidade de cada f_j.

Lema 2.3. *Seja* $(X, \mathscr{X})$ *um espaço mensurável. Uma função* $f\colon X \to \mathbb{R}^d$ *é mensurável se, e somente se, para cada* j *a função* $f_j : X \to \mathbb{R}$ *é mensurável.*

Exercício 2.5. Se $(X, \mathscr{X})$ é um espaço mensurável e $f\colon X \to \mathbb{C}$ é uma função complexa qualquer, de que maneira você abordaria a mensurabilidade de f? //

Dadas certas funções mensuráveis, é possível obter, a partir delas, novas funções mensuráveis. Este fato está enunciado nos Lemas seguintes.

Lema 2.4. *Sejam* $(X, \mathscr{X})$*,* $(Y, \mathscr{Y})$ *e* $(Z, \mathscr{Z})$ *espaços mensuráveis. Se* $f\colon X \to Y$ *e* $g\colon Y \to Z$ *são mensuráveis, então é também mensurável a aplicação* $g \circ f\colon X \to Z$*.*

Demonstração. Seja $C \in \mathscr{Z}$. Queremos verificar que $(g \circ f)^{-1}(C) \in \mathscr{X}$. Note que

$$(g \circ f)^{-1}(C) = f^{-1}\big(g^{-1}(C)\big).$$

Do fato de g ser mensurável, temos $g^{-1}(C) \in \mathscr{Y}$, quer dizer, $g^{-1}(C) = B$ para algum $B \in \mathscr{Y}$. Do fato de f ser mensurável, vale que $f^{-1}(B) \in \mathscr{X}$, isto é, $f^{-1}\big(g^{-1}(C)\big) \in \mathscr{X}$, como queríamos demonstrar. $\square$

Exemplo 2.2. Nas condições do Lema 2.4 é usual, principalmente em Teoria da Probabilidade, escrever $g(f)$ ao invés do mais correto $g \circ f$. Tal notação requer certa cautela, pois podem ocorrer ambiguidades. //

O resultado seguinte será extremamente importante quando introduzirmos o conceito de esperança condicional.

Lema 2.5 (Doob–Dynkin)**.** *Seja X um conjunto, $(Y, \mathscr{Y})$ um espaço mensurável e g uma função de X em Y. Então uma função $f\colon X \to \mathbb{R}^d$ é $\sigma(g)/\mathscr{B}(\mathbb{R}^d)$–mensurável se, e somente se, existe uma função $h\colon Y \to \mathbb{R}^d$ mensurável $\mathscr{Y}/\mathscr{B}(\mathbb{R}^d)$ tal que $f = h \circ g$.*

$$\begin{tikzcd} (X, \sigma(g)) \arrow[r, "g"] \arrow[dr, "f"'] & (Y, \mathscr{Y}) \arrow[d, "h"] \\ & \mathbb{R}^d \end{tikzcd}$$

Lema 2.6. *Seja $f\colon \mathbb{R}^d \to \mathbb{R}^k$. Se f é contínua, então f é $\mathscr{B}(\mathbb{R}^d)/\mathscr{B}(\mathbb{R}^k)$–mensurável.*

Corolário 2.3. *Sejam $f\colon X \to \mathbb{R}^d$ e $g\colon X \to \mathbb{R}^k$ aplicações mensuráveis, e seja $\alpha \in \mathbb{R}$. Então são também mensuráveis as seguintes aplicações:*

(i) $x \mapsto \alpha f(x)$;

(ii) $x \mapsto f(x) + g(x)$, *se* $d = k$;

(iii) $x \mapsto f(x)g(x)$, *se* $d = 1$ *e* $k \geq 1$;

(iv) $x \mapsto \|f(x)\|$;

(v) $x \mapsto \langle f(x), g(x)\rangle$, *se* $d = k$.

Se $f\colon X \to \mathbb{R}$ é uma função real qualquer, definimos a função $f^+ : X \to \mathbb{R}$, dita a **parte positiva** de f, por

$$f^+(x) := \max\{f(x), 0\}.$$

Semelhantemente, definimos a **parte negativa** de f por

$$f^-(x) := \max\{-f(x), 0\}.$$

Note que tanto f^+ quanto f^- assume somente valores reais maiores ou iguais a zero.

Exercício 2.6. Seja $f\colon X \to \mathbb{R}$ uma função. Mostre que:

(i) $f = f^+ - f^-$;

(ii) $|f| = f^+ + f^-$;

(iii) $f^+ = (|f| + f)/2$;

(iv) $f^- = (|f| - f)/2$.

Veja se você consegue justificar a seguinte afirmação: f é mensurável se, e somente se, f^+ e f^- são mensuráveis. //

Lema 2.7. *Seja $(X, \mathscr{X})$ um espaço mensurável. Seja (f_n) uma sequência de funções mensuráveis de X em $\mathbb{R}$. Então são mensuráveis as aplicações*

(i) $x \mapsto \inf_n f_n(x)$;

(ii) $x \mapsto \sup_n f_n(x)$.

Corolário 2.4. *Seja $(X, \mathscr{X})$ um espaço mensurável, e seja (f_n) uma sequência de funções mensuráveis de X em $\mathbb{R}^d$. Se para cada $x \in X$ existir o limite $\lim_n f_n(x)$, então é também mensurável a aplicação $x \mapsto \lim_n f_n(x)$.*

3 Medidas

Seja $(X, \mathscr{X})$ um espaço mensurável. Dizemos que uma função $\mu : \mathscr{X} \to \mathbb{R} \cup \{\infty\}$ é uma **medida** em $(X, \mathscr{X})$ se satisfizer as seguintes condições:

M1. $\mu(\varnothing) = 0$;

M2. $\mu(A) \geq 0$, $\forall A \in \mathscr{X}$;

M3. para toda sequência disjunta[1] (A_n) em $\mathscr{X}$, tem-se

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

A condição (M3) é dita a propriedade de σ–**aditividade** da medida μ. Se $\mu(X) < \infty$ diz-se que μ é **medida finita**. O exemplo mais importante de medida finita são as medidas de probabilidade: uma medida μ é dita uma **medida de probabilidade**, ou simplesmente uma **probabilidade**, em $(X, \mathscr{X})$ se $\mu(X) = 1$. Se existem conjuntos $A_n \in \mathscr{X}$, $n \in \mathbb{N}$, tais que $X = \bigcup_{n=1}^{\infty} A_n$ e $\mu(A_n) < \infty$ para todo n, então dizemos que μ é σ–**finita**. Toda medida finita é σ–finita: basta tomar $A_n = X$ para todo n.

Seja X um conjunto, $\mathscr{X}$ uma σ–álgebra em X, e μ uma medida definida em $\mathscr{X}$. Dizemos que a terna $(X, \mathscr{X}, \mu)$ é um **espaço de medida**. Quando μ é medida de probabilidade, $(X, \mathscr{X}, \mu)$ é dito **espaço de probabilidade**.

[1] Uma sequência (A_n) de elementos de uma σ–álgebra $\mathscr{X}$ é dita **disjunta** se $A_i \cap A_j = \varnothing$ sempre que $i \neq j$.

Exercício 3.1. Seja $x \in X$, onde $(X, \mathscr{X})$ é um espaço mensurável. Defina $\delta_x \colon \mathscr{X} \to \mathbb{R}$ por

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

Mostre que δ_x é medida de probabilidade em $(X, \mathscr{X})$. A medida δ_x é dita a **medida de Dirac em** x. Observe que $\delta_x(A) = \mathbb{I}_A(x)$. //

Exercício 3.2. Seja $X = \mathbb{N}$, e considere em X a σ–álgebra das partes $\mathscr{X} = \mathscr{P}(\mathbb{N})$. Defina $\mu : \mathscr{X} \to \mathbb{R} \cup \{\infty\}$ por.

$$\mu(A) := \begin{cases} \operatorname{card}(A), & \text{se } A \text{ for um subconjunto finito de } \mathbb{N}, \\ \infty, & \text{se } A \text{ for um subconjunto infinito de } \mathbb{N}. \end{cases}$$

Verifique que μ assim definida é uma medida σ-finita em $(\mathbb{N}, \mathscr{P}(\mathbb{N}))$. ($\mu$ é chamada a **medida da contagem** em $\mathbb{N}$). //

Exemplo 3.1. Seja $(X, \mathscr{X}) = (\mathbb{R}, \mathscr{B}(\mathbb{R}))$. Se $F \colon \mathbb{R} \to \mathbb{R}$ é uma função não decrescente e contínua à direita, então existe uma única medida σ–finita μ_F tal que $\mu_F(I) = F(b) - F(a)$, onde $I = (a, b]$ é um intervalo com extremos $a < b$. A medida μ_F é chamada **medida de Borel-Stieltjes** associada a F. O exemplo mais importante é o caso em que $F(x) = x$: tem-se $\mu_F(I) = b - a$, de forma que μ_F é a medida 'comprimento' em $\mathbb{R}$. Esta é a chamada **medida de Lebesgue**. O fato de que tais medidas existem e são únicas será tratado mais adiante. //

Exemplo 3.2. Seja $(X, \mathscr{X}, \mu)$ um espaço de medida, e $(Y, \mathscr{Y})$ um espaço mensurável. Se f é uma função $\mathscr{X}/\mathscr{Y}$–mensurável , então podemos induzir em $(Y, \mathscr{Y})$ uma medida μ_f, definida na σ–álgebra $\mathscr{Y}$, da seguinte maneira:

$$\mu_f(B) := \mu\big(f^{-1}(B)\big), \qquad B \in \mathscr{Y}.$$

No caso em que $(X, \mathscr{X}, \mu)$ é um espaço de probabilidade, dizemos que a medida μ_f é a **distribuição**, ou a **lei**, da variável aleatória f. Um dos casos mais importantes ocorre quando $Y = \mathbb{R}^d$. Nesse contexto, tem-se que $f = (f_1, \ldots, f_d)$, onde cada f_j é variável aleatória real, e então dizemos que μ_f é a **distribuição conjunta das variáveis aleatórias** $f_1, \ldots, f_d$. Outras notações para μ_f são $\mu \circ f^{-1}$ e $f_*\mu$. //

Exercício 3.3. Suponha dado um certo espaço de probabilidade $(Y, \mathscr{Y}, \lambda)$. Proponha um espaço de probabilidade $(X, \mathscr{X}, \mu)$ e uma variável aleatória f definida em X e tomando valores em Y tal que λ seja sua distribuição, isto é, tal que $\mu_f = \lambda$.//

Lema 3.1. *Seja μ uma medida definida em uma σ–álgebra $\mathscr{X}$. Se A e B são conjuntos mensuráveis e $A \subset B$, então $\mu(A) \leq \mu(B)$. Se $\mu(A) < \infty$, então $\mu(B \sim A) = \mu(B) - \mu(A)$.*

Uma sequência (A_n) de subconjuntos de um conjunto X é dita *crescente* se $A_n \subset A_{n+1}$, para todo n. A sequência (A_n) é dita *decrescente* se $A_n \supset A_{n+1}$, para todo n.

Lema 3.2. *Seja μ uma medida definida em uma σ–álgebra $\mathscr{X}$.*

(i) Se (A_n) é uma sequência crescente de conjuntos mensuráveis, então

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

(ii) Se (A_n) é uma sequência decrescente de conjuntos mensuráveis e $\mu(A_1) < \infty$, então

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

Seja $(X, \mathscr{X}, \mu)$ um espaço de medida, e seja Q uma certa propriedade (tipicamente, Q refere-se a elementos de X). Dizemos que Q vale para μ–**quase todo ponto** (μ–qtp) se $\mu\{x \in X : \neg Q(x)\} = 0$. Quando a medida μ está subentendida, é usual dizer simplesmente que a propriedade Q vale para **quase todo ponto** (qtp). Na linguagem de probabilidade, uma propriedade que valha para quase todo ponto é dita válida **quase certamente**.

Exemplo 3.3. Seja $(X, \mathscr{X}, \mu)$ um espaço de medida, e $(Y, \mathscr{Y})$ um espaço mensurável. Sejam f e g funções $\mathscr{X}/\mathscr{Y}$–mensuráveis. Dizemos que f **e** g **são iguais** μ**–qtp** se $\mu\{x \in X : f(x) \neq g(x)\} = 0$. Notação:

$$f = g, \qquad \mu\text{–qtp}.$$

Semelhantemente, escrevemos $f > 0$, μ–qtp (quando $Y = \mathbb{R}$), etc. //

Exemplo 3.4. Este exemplo complementa o anterior. Seja $(X, \mathscr{X}) = ([0,1], \mathscr{B}[0,1])$, e seja μ = medida de Lebesgue (comprimento) em $[0,1]$. Seja $(Y, \mathscr{Y}) = (\mathbb{R}, \mathscr{B}(\mathbb{R}))$, e considere as funções $f(x) := 1$, para todo $x \in [0,1]$, e

$$g(x) := \begin{cases} 1, & x \in [0,1) \\ 4, & x = 1. \end{cases}$$

Então $f \neq g$ mas $f = g$ μ–qtp. //

Exemplo 3.5. Seja $(X, \mathscr{X}, \mu)$ um espaço de medida. Sejam (f_n) uma sequência de funções mensuráveis de X em $\mathbb{R}^d$, e $f\colon X \to \mathbb{R}^d$ mensurável. Dizemos que f_n **converge para** f μ**–qtp**, se $\mu\{x \in X : \lim_n f_n(x) \neq f(x)\} = 0$. Notação

$$\lim_{n\to\infty} f_n(x) = f(x), \qquad \mu\text{–qtp}.$$

Equivalentemente, $f_n \to f$, μ–qtp. Se μ é medida de probabilidade, dizemos que a sequência de vetores aleatórios f_n **converge quase certamente** para o vetor aleatório f. //

3.1 Extensão de medidas

Esta seção busca revelar a importância da noção de uma σ–álgebra gerada por uma classe $\mathscr{C}$ de subconjuntos de X (conforme definido acima). A ideia básica é que, se μ é uma medida (finita ou σ–finita) definida em uma σ–álgebra gerada por uma classe $\mathscr{C}$ "suficientemente regular", então os valores $\mu(A)$, onde $A \in \mathscr{C}$, são suficientes para determinar μ completamente.

Lema 3.3. *Seja $\mathscr{C}$ uma classe de subconjuntos de X satisfazendo a seguinte propriedade:*

$$A \in \mathscr{C} \wedge B \in \mathscr{C} \implies A \cap B \in \mathscr{C}. \tag{3.1}$$

Se μ_1 e μ_2 são duas medidas definidas em $\sigma(\mathscr{C})$ tais que $\mu_1(X) = \mu_2(X) < \infty$ e

$$\mu_1(A) = \mu_2(A), \qquad \forall A \in \mathscr{C},$$

então

$$\mu_1(A) = \mu_2(A), \qquad \forall A \in \sigma(\mathscr{C}).$$

Uma classe de subconjuntos satisfazendo a condição (3.1) no Lema 3.3 é dita um π–**system**. Com essa definição podemos enunciar sucintamente o seguinte corolário.

Corolário 3.1. *Se duas medidas de probabilidade coincidem em um π–system, então elas coincidem na σ–álgebra gerada por esse π–system.*

Exemplo 3.6. A classe $\mathscr{U}$ do Exemplo 1.1 é um π–system. Logo, se μ é uma medida de probabilidade em $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$, e se conhecemos os valores $\mu(A_1 \times \cdots \times A_d)$, onde cada A_j é um intervalo aberto da reta, então conhecemos $\mu(B)$ para todo Boreliano B de $\mathbb{R}^d$. //

No Lema 3.3, as medidas μ_1 e μ_2 são *dadas*, e podemos afirmar que elas são idênticas, se coincidirem quando restritas a uma classe "suficientemente regular" (naquele caso, um π–system). Uma questão relacionada é a seguinte: dada uma classe "suficientemente regular" $\mathscr{C}$, e dada uma função $\mu_0 \colon \mathscr{C} \to \mathbb{R} \cup \{\infty\}$ que possua propriedades análogas à de uma medida, em que casos é possível estender μ_0 a uma medida μ definida na σ–álgebra gerada por $\mathscr{C}$? Nesse contexto, π–systems são classes restritivas demais; será necessário introduzir a seguinte noção. Dizemos que uma classe $\mathscr{A}$ de subconjuntos de X é uma **álgebra de subconjuntos de X** quando as seguintes propriedades são satisfeitas:

A1'. $\varnothing \in \mathscr{A}$;

A2'. se $A \in \mathscr{A}$, então $A^{\complement} \in \mathscr{A}$.

A3'. se $A_i \in \mathscr{A}$ para todo $i \in \{1, \dots, n\}$, então $\bigcup_{i=1}^{n} A_i \in \mathscr{A}$.

O que distingue σ–álgebras de álgebras, portanto, é o fato de que aquelas são estáveis sob uma quantidade enumerável de operações de união em seus elementos, enquanto estas o são sob uma quantidade *finita* de tais operações. Se $\mathscr{A}$ é uma álgebra de subconjuntos de X, dizemos que uma função $\mu_0 \colon \mathscr{A} \to \mathbb{R} \cup \{\infty\}$ é uma **pré–medida em $\mathscr{A}$** se (i) $\mu_0(\varnothing) = 0$; (ii) $\mu(A) \geq 0$ para todo $A \in \mathscr{A}$ e; (iii) $\mu_0(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu_0(A_n)$, sempre que (A_n) for uma sequência disjunta em $\mathscr{A}$ *cuja união está em $\mathscr{A}$*. Note que no item (iii) desta definição é necessária a ressalva de que a união enumerável da sequência disjunta esteja em $\mathscr{A}$ (isso nem sempre precisa ocorrer, já que $\mathscr{A}$ é apenas álgebra). Em outras palavras, a propriedade de σ–aditividade de uma pré–medida só tem que valer onde fizer sentido. Com estas definições em mãos, podemos enunciar o seguinte.

Teorema 3.1 (Teorema de Extensão de Hahn-Carathéodory)**.** *Se μ_0 é uma pré–medida definida em uma álgebra $\mathscr{A}$ de subconjuntos de X, então existe uma extensão de μ_0 a uma medida $\mu \colon \sigma(\mathscr{A}) \to \mathbb{R} \cup \{\infty\}$. Ademais, se μ é σ–finita, então tal extensão é única.*

Exemplo 3.7 (Continuação do Exemplo 3.1)**.** Este exemplo dá a importante caracterização de medidas de probabilidade em $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ em termos de *funções de distribuição*. Uma função $F \colon \mathbb{R}^d \to \mathbb{R}$ é dita uma **função de distribuição d–dimensional** (ou: função de distribuição em $\mathbb{R}^d$) se satisfaz as seguintes propriedades:

FD1. F é não decrescente em cada um de seus argumentos: para cada $i = 1, \dots, d$ tem-se

$$u \geq 0 \implies F(x_1, x_2, \dots, x_i + u, \dots, x_d) \geq F(x_1, x_2, \dots, x_i, \dots, x_d).$$

FD2. F é contínua à direita em cada um de seus argumentos: para cada $i = 1, \ldots, d$ tem-se

$$\lim_{u \downarrow 0} F(x_1, x_2, \ldots, x_i + u, \ldots, x_d) = F(x_1, x_2, \ldots, x_i, \ldots, x_d).$$

FD3. se *qualquer* dos argumentos de F vai para $-\infty$, então F converge para 0: para cada $i = 1, \ldots, d$,

$$\lim_{x_i \to -\infty} F(x_1, x_2, \ldots, x_i, \ldots, x_d) = 0.$$

FD4. para cada i, existe o limite $\lim_{x_i \to +\infty} F(x_1, \ldots, x_i, \ldots, x_d)$. Ademais, se *todos* os argumentos de F vão para $+\infty$, então F converge para 1: tem-se

$$\lim_{\min_i x_i \to +\infty} F(x_1, \ldots, x_d) = 1.$$

FD5. (positividade, condição técnica) Para quaisquer intervalos $J_k = (a_k, b_k] \subset \mathbb{R}$, $k = 1, \ldots, d$, tem-se

$$\Delta^1_{J_1} \Delta^2_{J_2} \cdots \Delta^d_{J_d} F \geq 0.$$

No item (FD5) acima, os operadores Δ^k_J são definidos da seguinte maneira: dados um inteiro $k \in \mathbb{N}$ e uma função $G \colon \mathbb{R}^k \to \mathbb{R}$, seja $J = (a, b] \subset \mathbb{R}$. Se $k \geq 2$, então $\Delta^k_J G$ é a função real definida em $\mathbb{R}^{k-1}$ tal que, para $(x_1, \ldots, x_{k-1}) \in \mathbb{R}^{k-1}$,

$$\Delta^k_J G(x_1, \ldots, x_{k-1}) = G(x_1, \ldots, x_{k-1}, b) - G(x_1, \ldots, x_{k-1}, a).$$

Se $k = 1$, então

$$\Delta^1_J G = G(b) - G(a).$$

Teorema 3.2. *Seja μ uma medida de probabilidade em $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$. Defina F_μ, para $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$, por*

$$F_\mu(x) = \mu\big((-\infty, x_1] \times (-\infty, x_2] \times \cdots \times (-\infty, x_d]\big).$$

Então $F_\mu \colon \mathbb{R}^d \to \mathbb{R}$ é função de distribuição em $\mathbb{R}^d$. Reciprocamente, se F é uma função de distribuição em $\mathbb{R}^d$, então existe uma única medida de probabilidade μ em $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ tal que F é sua função de distribuição, isto é, tal que $F_\mu = F$.

Quando $d = 1$, uma função de distribuição é às vezes chamada de **função de distribuição acumulada**. Note que, nesse caso, a condição (FD5) é implicada pela condição (FD1). O Teorema 3.2 dá uma resposta parcial à afirmação do Exemplo 3.1, de que a toda função $F \colon \mathbb{R} \to \mathbb{R}$ não decrescente e contínua à direita corresponde uma única medida σ–finita μ_F.

Para um exemplo de uma função F satisfazendo os itens (FD1), (FD2), (FD3) e (FD4), mas que não é uma função de distribuição de uma medida de probabilidade em $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$, ver Barry James (exemplo 2.4.14). //

Exercício 3.4. Seja $\mathscr{A}$ a coleção formada por uniões finitas disjuntas de intervalos do tipo

$$(a, b], \quad (-\infty, b], \quad (a, +\infty), \quad (-\infty, +\infty), \tag{3.2}$$

onde $a < b \in \mathbb{R}$. Isto é, os elementos de $\mathscr{A}$ são conjuntos da forma $A = J_1 \cup \cdots , J_m$, onde cada J_k é um intervalo do tipo (3.2) com extremos $a_k \leq b_k$ (possivelmente iguais a $\pm\infty$ onde permitido).

(i) Mostre que $\mathscr{A}$ é uma álgebra de subconjuntos de $\mathbb{R}$;

(ii) Dada uma função de distribuição acumulada $F\colon \mathbb{R} \to \mathbb{R}$, defina μ_0, para $A = J_1 \cup \cdots \cup J_m \in \mathscr{A}$, por

$$\mu_0(A) = \sum_{i=1}^{m} F(b_i) - F(a_i),$$

onde adotamos a convenção de escrever $F(+\infty) = 1$ e $F(-\infty) = 0$. Mostre que μ_0 é uma pré–medida em $\mathscr{A}$.

(iii) Conclua que μ_0 admite uma única extensão à uma medida μ em $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. Isso demonstra o Teorema 3.2 no caso $d = 1$. //

4 Integral de Lebesgue

Seja $(X, \mathscr{X})$ um espaço mensurável. Será conveniente introduzir uma notação para nos referirmos a funções mensuráveis, simples, etc., definidas em X. Nesse sentido, denotemos por $M\big(\mathscr{X}; \mathbb{R}^d\big)$ o conjunto formado pelas funções mensuráveis definidas em X e tomando valores em $\mathbb{R}^d$, isto é,

$$M\big(\mathscr{X}; \mathbb{R}^d\big) := \big\{f :\ f \text{ é } \mathscr{X}/\mathscr{B}(\mathbb{R}^d)\text{–mensurável}\big\}.$$

Semelhantemente, denotemos por $S\big(\mathscr{X}; \mathbb{R}^d\big)$ o subconjunto de $M\big(\mathscr{X}; \mathbb{R}^d\big)$ formado pelas funções simples:

$$S\big(\mathscr{X}; \mathbb{R}^d\big) := \big\{f :\ f \text{ é } \mathscr{X}/\mathscr{B}(\mathbb{R}^d)\text{–mensurável} \wedge \operatorname{ran}(f) \text{ é finito}\big\}.$$

Se $d = 1$, definimos adicionalmente $M^+(\mathscr{X}; \mathbb{R})$ e $S^+(\mathscr{X}; \mathbb{R})$ como sendo os subconjuntos respectivamente de $M(\mathscr{X}; \mathbb{R})$ e $S(\mathscr{X}; \mathbb{R})$ cujos elementos são funções não negativas.

No restante desta seção, $(X, \mathscr{X}, \mu)$ é um espaço de medida fixado. Considere uma função simples não negativa $f\colon X \to \mathbb{R}$, isto é, $f \in S^+(\mathscr{X}; \mathbb{R})$. Lembre que f admite a representação padrão

$$f(x) = \sum_{j=1}^{k} y_j \mathbb{I}_{A_j}(x), \tag{4.1}$$

em que $\operatorname{ran}(f) = \{y_1, \ldots, y_k\}$ e $A_j = f^{-1}(\{y_j\})$. Definimos a **integral de f em relação a μ** por

$$I_0^\mu(f) := \sum_{j=1}^{k} y_j \mu(A_j), \tag{4.2}$$

Note que não excluímos a possibilidade de que a quantidade acima seja igual a ∞. De fato, se para algum j tivermos $\mu(A_j) = \infty$ e $y_j \neq 0$, então $I_0^\mu(f) = \infty$. (Adota-se usualmente a convenção de que $0 \cdot \infty = 0$).

Exercício 4.1. Mostre que toda função simples não negativa f com representação padrão (4.1) admite infinitas representações como uma combinação linear de funções indicadoras, isto é, representações do tipo

$$f(x) = \sum_{j=1}^{\ell} z_j \mathbb{I}_{B_j}(x),$$

onde $z_j \geq 0$ e $B_j \in \mathscr{X}$. (Essa afirmação também é verdadeira para funções simples tomando valores em $\mathbb{R}^d$). Verifique que se f tem admite a representação (4.1), então $I_0^\mu(f) = \sum_{j=1}^{\ell} z_j \mu(B_j)$. //

Exercício 4.2. Sejam f e g funções simples não negativas, e $\alpha \geq 0$. Mostre que

(i) $\mathbb{I}_0^\mu(\alpha f) = \alpha I_0^\mu(f)$;

(ii) $I_0^\mu(f + g) = I_0^\mu(f) + I_0^\mu(g)$,

onde αf e $f + g$ denotam, respectivamente, as aplicações $x \mapsto \alpha f(x)$ e $x \mapsto f(x) + g(x)$. //

Seja agora $f\colon X \to \mathbb{R}$ mensurável, não necessariamente simples. Lembre que, pelo Lemma 2.2, existe uma sequência (f_n) de funções simples não negativas com $f_n \to f$ pontualmente. Definimos a **integral de f em relação a μ** por

$$I_1^\mu(f) := \lim_{n \to \infty} I_0^\mu(f_n) \tag{4.3}$$

Novamente, pode ocorrer de a quantidade acima ser igual a ∞.

Exercício 4.3. Note que cada f_n na equação (4.3) é simples e não negativa, e portanto podemos calcular sua integral $I_0^\mu(f_n)$ pela definição dada na equação (4.2). Lembre ainda que, pelo Lema 2.2, tem-se $0 \leq f_n(x) \leq f_{n+1}(x)$, $\forall x \in X$, $\forall n \in \mathbb{N}$. Mostre que $I_0^\mu(f_n) \leq I_0^\mu(f_{n+1})$. Conclua que a sequência de números reais $(I_0^\mu(f_n))_{n\in\mathbb{N}}$ é não decrescente e, portanto, ou converge para um número real não negativo, ou diverge para ∞. //

Iniciamos nossa construção da integral considerando apenas funções não negativas, e admitimos que suas integrais, dadas pelas equações (4.2) e (4.3), assumissem o valor ∞. No caso de uma função mensurável que assuma tanto valores positivos quanto negativos, será necessário nos restringirmos a um subconjunto de $M(\mathscr{X};\mathbb{R})$, no qual faça sentido definir a integral de uma função. Nesse sentido, a definição de integral dada a seguir não é, estritamente falando, uma extensão das definições anteriores.

Considere uma função mensurável $f\colon X \to \mathbb{R}$. Como tanto sua parte positiva quanto sua parte negativa são funções mensuráveis não negativas (quer dizer, f^+ e f^- são elementos de $M^+(\mathscr{X};\mathbb{R})$), podemos usar a definição dada em (4.3) para calcular suas integrais. No caso de $I_1^\mu(f^+)$ e $I_1^\mu(f^-)$ serem ambos finitos, dizemos que f é **μ–integrável**, e definimos a **integral de f em relação a μ** por

$$\int f\, d\mu := I_1^\mu(f^+) - I_1^\mu(f^-). \tag{4.4}$$

A hipótese de que $I_1^\mu(f^+)$ e $I_1^\mu(f^-)$ são ambos finitos garante que não ocorram indeterminações do tipo $\infty - \infty$ em (4.4). Notação:

$$\int f\, d\mu =: \int f(x)\, d\mu(x) =: \int f(x)\mu(dx) =: \mu(f).$$

Quando μ é medida de probabilidade, a integral de f em relação a μ é também chamada a **esperança de f**, denotada por $\mathbb{E}(f)$. Outros nomes para a esperança são: esperança matemática, valor esperado, média, etc. Se a medida μ estiver subentendida, dizemos simplesmente que f é **integrável**.

Exercício 4.4. Seja $f \in M(\mathscr{X};\mathbb{R})$ uma função integrável. Mostre que, para todo $A \in \mathscr{X}$, é também integrável a função $x \mapsto f(x)\mathbb{I}_A(x)$. Nesse caso, definimos a **integral de f sobre A, em relação a μ** por

$$\int_A f\, d\mu := \int f(x)\mathbb{I}_A(x)\, d\mu(x)$$

As notações $\int_A f(x)\, d\mu(x)$, $\mu(f\mathbb{I}_A)$, etc., também são utilizadas. //

É comum na Teoria da Integração utilizar o mesmo símbolo $\int$ para designar integrais de funções simples não negativas, de funções mensuráveis não negativas e de funções integráveis. Em geral está subentendida a classe à qual pertence a função sendo integrada, e portanto é possível saber qual das definições dentre (4.2), (4.3) ou (4.4) deve ser utilizada. No que segue, adotaremos essa convenção. É importante observar que as definições não geram ambiguidades.

Lema 4.1. *Sejam $f, g \in M^+(\mathscr{X}; \mathbb{R})$, e $A, B \in \mathscr{X}$.*

(i) se $f \leq g$, então $\int f\, d\mu \leq \int g\, d\mu$;

(ii) se $A \subset B$, então $\int_A f\, d\mu \leq \int_B f\, d\mu$.

O seguinte Teorema é um dos resultados mais importantes da Teoria da Integração; ele implica, em particular, que a sequência de funções simples usada na definição da integral I_1^μ não precisa ser tomada, necessariamente, como sendo aquela dada no Lema 2.2.

Teorema 4.1 (Teorema da Convergência Monótona)**.** *Seja (f_n) uma sequência de funções mensuráveis reais tais que $0 \leq f_n(x) \leq f_{n+1}(x)$, para todo $x \in X$ e todo $n \in \mathbb{N}$. Então*

$$\lim \int f_n\, d\mu = \int \lim f_n\, d\mu.$$

Corolário 4.1. *Se $f \in M^+(\mathscr{X}; \mathbb{R})$, seja $\lambda\colon \mathscr{X} \to \mathbb{R} \cup \{\infty\}$ definida por*

$$\lambda(A) := \int_A f\, d\mu, \qquad A \in \mathscr{X}. \tag{4.5}$$

Então λ é uma medida em $(X, \mathscr{X})$.

Corolário 4.2. *Seja $f \in M^+(\mathscr{X}; \mathbb{R})$. Então tem-se que $f = 0$ μ–qtp se, e somente se, $\int f\, d\mu = 0$.*

Teorema 4.2 (Desigualdade de Hölder)**.** *Sejam f e g funções $\mathscr{X}/\mathscr{B}(\mathbb{R})$–mensuráveis. Sejam $p > 1$ e q tais que $(1/p) + (1/q) = 1$. Assuma que*

(i) $\int |f(x)|^p\, d\mu(x) < \infty$;

(ii) $\int |g(x)|^q\, d\mu(x) < \infty$.

Então

$$\int |f(x)g(x)|\, d\mu(x) \leq \Big(\int |f(x)|^p\, d\mu(x) \Big)^{1/p} \Big(\int |g(x)|^q\, d\mu(x) \Big)^{1/q}. \tag{4.6}$$

Exercício 4.5. Em alguns casos, é conveniente considerar funções $f\colon X \to [0,+\infty]$, isto é, funções não negativas que assumam o valor $+\infty$. A mensurabilidade de tais funções pode ser abordada através de pequenas modificações nas definições dadas acima. Se adaptarmos nossa definição de função simples não negativa de forma a permitir que o valor $+\infty$ seja assumido, então as integrais I_0^μ e I_1^μ são definidas de maneira idêntica ao acima exposto. Mostre que, se $I_1^\mu(f) < \infty$, então $\mu\{x \in X : f(x) = +\infty\} = 0$. //

Exemplo 4.1 (Continuação dos Exemplos 3.1 e 3.7)**.** Seja $F\colon \mathbb{R} \to \mathbb{R}$ uma função não decrescente e contínua à direita. Seja μ a medida de Borel–Stieltjes associada a F. Nesse caso, para uma função μ–integrável $g\colon \mathbb{R} \to \mathbb{R}$, escrevemos

$$\int g(x)\,dF(x) := \int g(x)\,d\mu(x).$$

Semelhantemente, escrevemos $\int_B g(x)\,dF(x)$, $\int_a^b g(x)\,dF(x)$, etc, cujo sentido deve ser evidente. //

Teorema 4.3 (Mudança de Variáveis)**.** *Seja $f\colon X \to \mathbb{R}$ uma função $\mathscr{X}/\mathscr{B}(\mathbb{R})$–mensurável. Uma aplicação $g\colon \mathbb{R} \to \mathbb{R}$ é μ_f–integrável se, e somente se, $g \circ f$ é μ–integrável. Nesse caso,*

$$\int_B g(t)\,d\mu_f(t) = \int_{f^{-1}(B)} g \circ f(x)\,d\mu(x), \qquad B \in \mathscr{B}(\mathbb{R}).$$

4.1 Medidas absolutamente contínuas

Se λ e μ são medidas em $(X, \mathscr{X})$, então dizemos que λ **é absolutamente contínua em relação a** μ se valer a seguinte condição: para cada $A \in \mathscr{X}$, se $\mu(A) = 0$, então $\lambda(A) = 0$. Notação: $\lambda \ll \mu$.

Corolário 4.3. *Suponha que $f \in M^+(\mathscr{X};\mathbb{R})$, e defina λ pela equação (4.5). Então λ é absolutamente contínua em relação a μ.*

O resultado abaixo, conhecido como Teorema de Radon-Nikodým, dá uma recíproca para o Corolário 4.3. A demonstração desse fato não decorre diretamente dos conteúdos abordados neste texto, e de fato foge ao nosso escopo. Enunciamos o Teorema para que o leitor tenha conhecimento dessa importante caracterização.

Teorema 4.4 (Teorema de Radon–Nikodým)**.** *Sejam λ e μ medidas σ–finitas definidas em $\mathscr{X}$, e suponha que λ é absolutamente contínua em relação μ. Então existe uma função $f \in M^+(\mathscr{X};\mathbb{R})$ tal que vale a identidade (4.5). Ademais, f é única μ–qtp.*

Se $\lambda \ll \mu$ são medidas σ–finitas, a função f dada no Teorema 4.4 é chamada a **derivada de Radon–Nikodým** de λ com respeito a μ.

Exercício 4.6. Sejam λ e μ medidas σ–finitas em $(X, \mathscr{X})$. Assuma que $\lambda \ll \mu$, e seja f a derivada de Radon–Nikodým de λ com respeito a μ. Mostre que, se $g\colon X \to \mathbb{R}$ é uma função mensurável não negativa, então

$$\int g(x)\, d\lambda(x) = \int g(x) f(x)\, d\mu(x).$$

Adapte para o caso em que g é λ–integrável. //

Exercício 4.7. Seja $F\colon \mathbb{R} \to \mathbb{R}$ uma função de distribuição acumulada, e seja μ a medida de Borel–Stieltjes correspondente. Seja λ a medida de Lebesgue na reta, e assuma que $\mu \ll \lambda$. Mostre que, se a derivada de Radon–Nikodým de μ em relação a λ for uma função contínua (vamos denotá-la por f), então $F' = f$. (Dica: utilize o Teorema Fundamental do Cálculo). Na verdade, a igualdade $f = F'$ ainda é válida λ–qtp mesmo quando f é descontínua, desde que $\mu \ll \lambda$. //

Exemplo 4.2. O conceito de medida absolutamente contínua é extremamente importante em probabilidade. O leitor certamente já se deparou, em cursos introdutórios, com as definições de variável aleatória discreta e contínua. Com um pequeno desvio em relação à notação adotada até aqui, seja $(\Omega, \mathscr{F}, \mathbb{P})$ um espaço de probabilidade, e seja $X\colon \Omega \to \mathbb{R}$ uma variável aleatória. Lembre que a **distribuição** da variável aleatória X é a medida de probabilidade $\mathbb{P}_X$ no espaço mensurável $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$, definida por

$$\mathbb{P}_X(B) := \mathbb{P}\big(X^{-1}(B)\big), \qquad B \in \mathscr{B}(\mathbb{R}).$$

Em textos introdutórios, a variável aleatória X é dita *contínua* se as probabilidades de eventos associados a X são calculadas utilizando-se uma função densidade de probabilidade f_X, através da fórmula

$$\mathbb{P}[a \le X \le b] = \int_a^b f_X(x)\, dx.$$

A integral no lado direito da igualdade acima é, usualmente, interpretada como uma integral de Riemann. Será conveniente interpretá-la aqui como a integral da função mensurável $f_X\colon \mathbb{R} \to \mathbb{R}$ em relação à medida de Lebesgue (comprimento), a qual denotaremos por λ. Logo, $\lambda(I) = b - a$ se $I \subset \mathbb{R}$ é um intervalo com extremos a e b. Assim, das identidades acima obtemos

$$\mathbb{P}_X(B) = \int_B f_X(x)\, d\lambda(x), \qquad B \in \mathscr{B}(\mathbb{R}).$$

Claramente $\mathbb{P}_X \ll \lambda$. Isso mostra que variáveis aleatórias contínuas são aquelas cuja distribuição é uma medida absolutamente contínua em relação à medida de Lebesgue. //

Exercício 4.8. Na notação do Exemplo acima, justifique a seguinte afirmação: se $\mathbb{P}[a \leq X \leq b] = \int_a^b f_X(x)\, d\lambda(x)$ para quaisquer $a < b$ reais, então $\mathbb{P}_X(B) = \int_B f_X(x)\, d\lambda(x)$ para qualquer Boreliano B. //

Exercício 4.9. Ainda na notação do Exemplo 4.2, seja F_X a função de distribuição acumulada correspondente à distribuição $\mathbb{P}_X$. Temos

$$\mathbb{E}(g(X)) = \mathbb{E}(g \circ X) = \int g \circ X(\omega)\, d\mathbb{P}(\omega) \stackrel{*}{=} \int g(x)\, d\mathbb{P}_X(x) = \int g(x)\, dF_X(x),$$

para toda função $\mathbb{P}_X$–integrável $g \colon \mathbb{R} \to \mathbb{R}$. (O termo mais à esquerda na igualdade acima segue a notação usual a que nos referimos anteriormente. Perceba a potencial ambiguidade nessa notação). A passagem $*$ se justifica pelo Teorema 4.3.

1. Seja λ a medida de Lebesgue em $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. No que segue, usaremos a notação $dx = d\lambda(x)$. Lembre que a variável aleatória X é dita **(absolutamente) contínua** se, e somente se, sua distribuição é uma medida absolutamente contínua em relação à medida de Lebesgue (isto é, $\mathbb{P}_X \ll \lambda$). Assumindo que X é contínua, seja f_X a derivada de Radon–Nikodým de $\mathbb{P}_X$ em relação a λ. Nesse contexto, f_X é dita a **função densidade de probabilidade de X**. Mostre que
$$\mathbb{E}(g(X)) = \int g(x) f_X(x)\, dx.$$
Em particular, se X é integrável,
$$\mathbb{E}(X) = \int x f_X(x)\, dx.$$

2. A variável aleatória X é dita **discreta** se existir um subconjunto enumerável $B_0 \subset \mathbb{R}$ tal que $\mathbb{P}_X(B_0) = 1$. *Defina*, para cada $x \in B_0$,
$$p_X(x) := \mathbb{P}_X(\{x\}) = \mathbb{P}\{\omega \in \Omega : X(\omega) = x\}.$$
Estenda p_X a toda a reta pondo $p_X(x) = 0$ se $x \notin B_0$. A função $p_X \colon \mathbb{R} \to \mathbb{R}$ é dita a **função massa de probabilidade de X**. Mostre que
$$\mathbb{E}(g(X)) = \sum_{x \in B_0} g(x) p_X(x),$$
e, em particular, $\mathbb{E}(X) = \sum_{x \in B_0} x p_X(x)$ se X é integrável. //

Finalizamos esta seção com um importante resultado sobre decomposição de medidas. Essencialmente, dadas quaisquer duas medidas σ–finitas λ e μ, é sempre verdade que ao menos *parte* de λ é absolutamente contínua em relação a μ. Duas medidas λ e μ definidas em $\mathscr{X}$ são ditas **mutuamente singulares** se existirem subconjuntos mensuráveis A e B de X tais que $X = A \cup B$ e $\lambda(A) = \mu(B) = 0$. Nesse caso, escrevemos $\lambda \perp \mu$. Embora a relação de singularidade seja simétrica em λ e μ, também diz-se que λ é **singular com respeito a** μ.

Teorema 4.5 (Teorema de Decomposição de Lebesgue)**.** *Sejam λ e μ medidas σ–aditivas definidas em uma σ–álgebra $\mathscr{X}$. Então existem medidas $\lambda_a \ll \mu$ e $\lambda_s \perp \mu$ tais que $\lambda = \lambda_a + \lambda_s$. Ademais, as medidas λ_a e λ_s são únicas.*

No caso em que $\lambda \perp \mu$, tem-se $\lambda_a(A) = 0$ para todo $A \in X$.

4.2 Integrais vetoriais

Seja $(X, \mathscr{X}, \mu)$ um espaço de medida, e seja E um espaço de Banach. Estaremos interessados principalmente nos casos em que $E = \mathbb{R}^d$, e em que E é o espaço $M_{\mathbb{R}}(d \times k)$ formado pelas matrizes $d \times k$ com entradas reais. Se o leitor não estiver familiarizado com a definição abstrata de um espaço de Banach, não haverá perda de generalidade em considerar, no que segue, que E é um dos dois espaços acima mencionados. Em qualquer caso, sempre consideramos em E a σ–**álgebra de Borel** $\mathscr{B}(E)$, isto é, a σ–álgebra $\sigma(\mathscr{U})$ gerada pela classe $\mathscr{U}$ dos conjuntos abertos de E. (O leitor deve convencer-se de que, no caso em que $E = \mathbb{R}^d$, essa σ–álgebra coincide com aquela dada no Exemplo 1.1.).

No que segue, denotaremos por $M(\mathscr{X}; E)$ o conjunto das funções $\mathscr{X}/\mathscr{B}(E)$–mensuráveis. Dadas f e g em $M(\mathscr{X}; E)$, e $\alpha \in \mathbb{R}$, adotaremos a notação usual em que $f+g$ e αf denotam respectivamente as aplicações $x \mapsto f(x)+g(x)$ e $x \mapsto \alpha f(x)$.

Exemplo 4.3. Lembre que todas as normas em um espaço vetorial E de dimensão finita são *equivalentes*, isto é, se $\|\cdot\|_1$ e $\|\cdot\|_2$ são duas normas em E, então existem constantes positivas c_1 e c_2 tais que, para todo $v \in E$,

$$c_1\|v\|_1 \leq \|v\|_2 \leq c_2\|v\|_1.$$

Lembre também que se E é um espaço vetorial de dimensão finita, digamos $\dim(E) = d$, então E é *isomorfo* a $\mathbb{R}^d$, isto é, existe uma bijeção linear entre esses dois espaços. Lembre ainda que o espaço $M_{\mathbb{R}}(d \times k)$ tem dimensão finita e igual a $d \times k$. Basta tomar, por exemplo, a base formada pelas matrizes E_{ij} tendo entrada (i, j) igual a 1 e demais entradas iguais a zero. Aqui o índice i varia de 1 a d e o índice j varia de 1 a k, mostrando que de fato a base tem cardinalidade $d \times k$. //

Dizemos que uma função mensurável $f\colon X \to E$ é μ–**simples** se $\operatorname{ran}(f)$ é um subconjunto finito[2] de E e $\mu\{x : f(x) = y\} < \infty$ para cada $y \neq 0$ em $\operatorname{ran}(f)$. Logo, se f é μ–simples, então vale a representação

$$f(x) = \sum_{j=1}^{k} y_j \mathbb{I}_{A_j}, \tag{4.7}$$

onde $\operatorname{ran}(f) = \{y_1, \ldots, y_k\}$ e $A_j = f^{-1}(\{y_j\})$, com $\mu(A_j) < \infty$ se $y_j \neq 0$. A diferença em relação à definição anterior de *função simples* está no requerimento adicional de que os conjuntos A_j tenham medida finita (e, claro, no fato de que somente havíamos introduzido a noção de função simples para o caso em que $E = \mathbb{R}^d$ – mas evidentemente aquela noção se aplica mais geralmente a espaços de Banach quaisquer). Definimos a integral de uma função μ–simples $f\colon X \to E$ com representação (4.7) como sendo o elemento $I_2^\mu(f) \in E$ dado por

$$I_2^\mu(f) = \sum_{j=1}^{k} y_j \mu(A_j).$$

Uma função $f\colon X \to E$ é dita μ–**integrável** se existe uma sequência (f_n) de funções μ–simples $f_n\colon X \to E$ tal que

(i) $f_n \to f$, μ–qtp;

(ii) $\lim_{n\to\infty} \int \|f_n(x) - f(x)\|\, d\mu(x) = 0$.

Se $f\colon X \to E$ é uma função μ–integrável, definimos a **integral de f em relação a μ** como sendo o elemento $\int f\, d\mu \in E$ dado por

$$\int f\, d\mu := \lim_{n\to\infty} I_2^\mu(f_n), \tag{4.8}$$

onde (f_n) é a sequência aproximante dada na definição. Notação:

$$\int f\, d\mu =: \int f(x)\, d\mu(x) =: \int f(x)\mu(dx) =: \mu(f).$$

Convém reiterar que $\int f\, d\mu$ é aqui um elemento do espaço E; em particular, se $E = \mathbb{R}^d$, então $\int f\, d\mu$ é um vetor; se $E = M_{\mathbb{R}}(d \times k)$, então $\int f\, d\mu$ é uma matriz. O leitor deve convencer-se de que, no caso em que $E = \mathbb{R}$, a nova definição de integral coincide com a anterior. Evidentemente, se a medida μ estiver subentendida, dizemos simplesmente que f é **integrável**.

[2] O mais correto seria afirmar que existe um subconjunto finito E_0 de E com $\mu(f \notin E_0) = 0$.

No caso em que $(X, \mathscr{X}, \mu)$ é espaço de probabilidade, a integral de f em relação a μ é também chamada a **esperança de** f, denotada por $\mathbb{E}(f)$. Outros nomes para a esperança são: esperança matemática, valor esperado, média, etc. Notação: $\mathbb{E}(f)$, $\mathbb{E}f$.

Exercício 4.10. Veja se você consegue justificar as desigualdades

$$\begin{aligned} \|I_2^\mu(f_n) - I_2^\mu(f_m)\| &\leq \int \|f_n(x) - f_m(x)\| \, d\mu(x) \\ &\leq \int \|f_n(x) - f(x)\| + \|f(x) - f_m(x)\| \, d\mu(x). \end{aligned}$$

Conclua que a sequência $(I_2^\mu(f_n))_{n\in\mathbb{N}}$ é uma sequência de Cauchy em E e que, portanto, o limite em (4.8) está bem definido. //

Exercício 4.11. Seja $f\colon X \to E$ uma função μ–integrável. Mostre que, para todo $A \in \mathscr{X}$, é também integrável a função $x \mapsto f(x)\mathbb{I}_A(x)$. Nesse caso, definimos a **integral de f sobre A, em relação a μ** por

$$\int_A f \, d\mu := \int f(x)\mathbb{I}_A(x) \, d\mu(x)$$

Assim como no caso escalar, as notações $\int_A f(x) \, d\mu(x)$, $\mu(f\mathbb{I}_A)$, etc., também são utilizadas. //

É possível dar uma segunda caracterização de funções μ–integráveis.

Proposição 4.1. *Seja $(X, \mathscr{X}, \mu)$ um espaço de medida, e seja E um espaço de Banach. Uma função $f\colon X \to E$ é μ–integrável se, e somente se,*

$$\int \|f(x)\| \, d\mu(x) < \infty.$$

Nesse caso, tem-se

$$\left\| \int f(x) \, d\mu(x) \right\| \leq \int \|f(x)\| \, d\mu(x).$$

O Teorema abaixo trata da linearidade da integral. Esse resultado é de suma importância.

Proposição 4.2. *Sejam $f, g\colon X \to E$ funções μ–integráveis, e $\alpha \in \mathbb{R}$. Então $f + g$ e αf são μ–integráveis, e*

$$\int \alpha f + g \, d\mu = \alpha \int f \, d\mu + \int g \, d\mu.$$

Exercício 4.12. Sejam A e B conjuntos mensuráveis disjuntos, e seja $f\colon X \to E$ tal que a aplicação $x \mapsto f(x)\mathbb{I}_{A\cup B}(x)$ é μ–integrável. Mostre que $\int_{A\cup B} f\,d\mu = \int_A f\,d\mu + \int_B f\,d\mu$. //

O resultado seguinte, conhecido como o Teorema da Convergência Dominada de Lebesgue, é um dos mais importantes em toda a Matemática.

Teorema 4.6 (Teorema da Convergência Dominada)**.** *Seja $f_n\colon X \to E$ uma sequência de funções μ–integráveis. Assuma que existam funções $f\colon X \to E$ e $g\colon X \to \mathbb{R}$ tais que*

(i) $\lim_{n\to\infty} f_n(x) = f(x)$, *$\mu$–qtp;*

(ii) $\|f_n(x)\| \leq |g(x)|$, *μ–qtp, para todo n;*

(iii) $\int |g(x)|\,d\mu(x) < \infty$.

Então f é μ–integrável e

$$\lim_{n\to\infty} \int \|f_n(x) - f(x)\|\,d\mu(x) = 0.$$

Em particular[3],

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

Teorema 4.7 (Hille)**.** *Sejam E e E' espaços de Banach, e seja $T\colon E \to E'$ uma transformação linear contínua. Se $f\colon X \to E$ é μ–integrável, então $T\circ f\colon X \to E'$ é μ–integrável e*

$$T\Big(\int f\,d\mu\Big) = \int T\circ f\,d\mu.$$

Nos exemplos abaixo usaremos dois resultados conhecidos da Álgebra Linear. Primeiramente, lembre que se E e E' são espaços vetoriais reais de dimensão finita, digamos $\dim(E) = k$ e $\dim(E') = d$, então toda transformação linear $T\colon E \to E'$ pode ser representada por uma matriz $A \in M_{\mathbb{R}}(d\times k)$. O segundo fato que usaremos é o seguinte: toda transformação linear entre espaços normados cujo domínio tem dimensão finita é contínua. Nesses exemplos, μ é medida de probabilidade.

Exemplo 4.4. Sejam $f\colon X \to \mathbb{R}^d$ vetor aleatório μ–integrável, e A uma matriz $k\times d$. Então

$$A\cdot \mathbb{E}f = \mathbb{E}A\cdot f, \tag{4.9}$$

[3] Evidentemente, o limite em questão significa que a sequência de números reais $\left(\left\|\int f_n\,d\mu - \int f\,d\mu\right\|\right)_{n\in\mathbb{N}}$ converge para zero.

onde $\cdot$ denota multiplicação de uma matriz por um vetor, e $A \cdot f$ é a aplicação $x \mapsto A \cdot f(x)$. De fato, defina $T_A : \mathbb{R}^d \to \mathbb{R}^k$ como sendo a transformação linear associada a A, isto é,

$$T_A(y) := A \cdot y, \qquad y \in \mathbb{R}^d.$$

Claramente T_A é linear, logo contínua. Pelo Teorema 4.7, temos $T_A(\mathbb{E}f) = \mathbb{E}(T_A \circ f)$, isto é, vale a identidade (4.9). //

Exemplo 4.5. Seja $f\colon X \to M_{\mathbb{R}}(d \times k)$ uma função mensurável. Dizemos que f é **matriz aleatória**. Se f é μ–integrável, e A é uma matriz $m \times d$ dada, então

$$A \cdot \mathbb{E}f = \mathbb{E}A \cdot f. \tag{4.10}$$

onde $\cdot$ denota multiplicação de matrizes e $A \cdot f$ é a aplicação $x \mapsto A \cdot f(x)$. De fato, defina $T_A\colon M_{\mathbb{R}}(d \times k) \to M_{\mathbb{R}}(m \times k)$ por

$$T_A(B) = A \cdot B, \qquad B \in M_{\mathbb{R}}(d \times k).$$

Claramente T_A é linear, logo contínua. Pelo Teorema 4.7, temos $T_A(\mathbb{E}f) = \mathbb{E}(T_A \circ f)$, isto é, vale a identidade (4.10). //

Exercício 4.13. Seja $f\colon X \to \mathbb{R}^d$ uma aplicação mensurável. Claramente, $f = (f_1, \ldots, f_d)$, onde $f_j\colon X \to \mathbb{R}$ são funções mensuráveis reais. Mostre que f é μ–integrável se, e somente se, f_j é μ–integrável para todo j. Use o Teorema 4.7 para concluir que $\left(\int f \, d\mu\right)_j = \int f_j \, d\mu$. Generalize para o caso $f\colon X \to M_{\mathbb{R}}(d \times k)$. //

4.3 Os espaços de Bochner–Lebesgue

Seja $1 \leq p < \infty$. Definimos o conjunto $L^p_\mu(E)$ como sendo a coleção formada por aquelas funções mensuráveis $f\colon X \to E$ tais que a aplicação $x \mapsto \|f(x)\|^p$ é μ–integrável, isto é

$$L^p_\mu(E) := \left\{ f \in M(\mathscr{X}; E) : \int \|f(x)\|^p \, d\mu(x) < \infty \right\}.$$

Se $f \in L^p_\mu(E)$, definimos a **norma–p de** f como sendo o número real

$$\|f\|_p := \left(\int \|f(x)\|^p \, d\mu(x) \right)^{1/p}.$$

Note que a norma dentro da integral no lado direito da equação acima é a norma em E. A primeira parte Proposição 4.1 pode ser re-enunciada da seguinte forma: f é um elemento de $L^1_\mu(E)$ se, e somente se, f é μ–integrável.

O caso $p = \infty$ requer um tratamento distinto. Definimos, para $f \in M_{\mathbb{R}}(\mathscr{X}; E)$,

$$\|f\|_\infty := \inf\{r \geq 0 : \mu\{x : \|f(x)\| > r\} = 0\}$$

e

$$L^\infty_\mu(E) := \{f \in M(\mathscr{X}; E) : \|f\|_\infty < \infty\}.$$

$L^\infty_\mu(E)$ é dito o espaço das funções μ–**essencialmente limitadas**.

Apesar da terminologia, ainda não sabemos se $\|\cdot\|_p$ de fato satisfaz as propriedades de uma norma. Evidentemente $\|\alpha f\|_p = |\alpha| \|f\|_p$. Isso já demonstra o primeiro item do Teorema abaixo; o segundo item é a desigualdade triangular.

Exemplo 4.6. Com a notação introduzida, a desigualdade de Hölder (4.6) pode ser re-enunciada da seguinte maneira: se $f \in L^p_\mu(\mathbb{R})$ e $g \in L^q_\mu(\mathbb{R})$, onde $p > 1$ e q são tais que $(1/p) + (1/q) = 1$, então

$$\|fg\|_1 \leq \|f\|_p \|g\|_q.$$

onde fg denota a aplicação $x \mapsto f(x)g(x)$. //

Teorema 4.8. *Sejam f e g funções em $L^p_\mu(E)$, e seja α um número real qualquer. Então*

(i) $\alpha f \in L^p_\mu(E)$*;*

(ii) $f + g \in L^p_\mu(E)$ *e*

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p. \tag{4.11}$$

A desigualdade (4.11) é conhecida como a **desigualdade de Minkowski**

Note que *não* verificamos que $\|\cdot\|_p$ é uma norma. Faltaria mostrar que

$$\|f\|_p = 0 \iff f = 0.$$

Vejamos que isso não ocorre. Evidentemente, a expressão $f = 0$ significa $f(x) = 0$ para todo x. Todavia, é suficiente que $f = 0$ μ–qtp para que valha $\|f\|_p = 0$. Isto é, $\|\cdot\|_p$ *não é* uma norma em $L^p_\mu(E)$.

Exercício 4.14. Dadas f e g em $M(\mathscr{X}; E)$, defina a relação de equivalência $\sim$ por

$$f \sim g \iff f = g \ \mu\text{–qtp}.$$

Seja $\mathscr{L}^p_\mu(E)$ o espaço quociente $L^p_\mu(E)/\sim$, isto é, $\mathscr{L}^p_\mu(E)$ é o conjunto cujos elementos são as classes de equivalência $[f]$, onde $g \in [f]$ se, e somente se, $g \sim f$. Mostre que $\|\cdot\|_p$ define uma norma em $\mathscr{L}^p_\mu(E)$. //

Exercício 4.15. Verifique a seguinte afirmação: se μ é medida finita, então $L^r_\mu(E) \subset L^p_\mu(E)$, para quaisquer $1 \le r \le p \le \infty$. //

Exemplo 4.7. Seja μ medida de probabilidade. Lembre que a variância da variável aleatória $f\colon X \to \mathbb{R}$ é definida por $\mathrm{Var}(f) = \mathbb{E}(f - \mathbb{E}f)^2$. Segue que $\mathscr{L}^2_\mu(\mathbb{R})$ é o espaço das (classes de equivalência de) variáveis aleatórias com variância finita (isso segue da identidade $\mathrm{Var}(f) = \mathbb{E}f^2 - \mathbb{E}^2 f$). O espaço $\mathscr{L}^2_\mu(\mathbb{R})$ é um espaço de Hilbert com o produto-interno $\langle f, g\rangle = \mathbb{E}fg$. //

5 Medidas em espaços produto

Sejam $(X, \mathscr{X})$ e $(Y, \mathscr{Y})$ espaços mensuráveis. Queremos introduzir no conjunto $X \times Y$ uma σ–álgebra a partir de 'blocos fundamentais' do tipo $A \times B$, onde $A \in \mathscr{X}$ e $B \in \mathscr{Y}$. Chamaremos um tal conjunto de **retângulo mensurável** em $X \times Y$. Defina $\mathscr{Z}_0 \subset \mathscr{P}(X \times Y)$ como sendo a coleção cujos elementos são uniões finitas de retângulos mensuráveis. Isto é, o elemento típico de $\mathscr{Z}_0$ é um conjunto do tipo

$$(A_1 \times B_1) \cup (A_2 \times B_2) \cup \cdots \cup (A_m \times B_m),$$

onde $m \in \mathbb{N}$, cada A_j é $\mathscr{X}$–mensurável e cada B_j é $\mathscr{Y}$–mensurável. Denotamos por $\mathscr{X} \otimes \mathscr{Y}$ a σ–álgebra em $X \times Y$ gerada pela classe $\mathscr{Z}_0$. Note que $\mathscr{X} \otimes \mathscr{Y}$ *não é* o produto cartesiano das σ–álgebras $\mathscr{X}$ e $\mathscr{Y}$.

Exercício 5.1. Mostre que

(i) $\mathscr{Z}_0$ coincide com a coleção cujos elementos são uniões finitas de retângulos mensuráveis *disjuntos*;

(ii) $\mathscr{Z}_0$ é uma álgebra de subconjuntos de $X \times Y$. //

Quando $(X, \mathscr{X})$ e $(Y, \mathscr{Y})$ são munidos de medidas μ e ν, respectivamente, é possível introduzir no espaço mensurável $(X \times Y, \mathscr{X} \otimes \mathscr{Y})$ uma medida $\mu \otimes \nu$ que possui a seguinte propriedade:

$$\mu \otimes \nu(A \times B) = \mu(A)\nu(B), \qquad A \in \mathscr{X}, B \in \mathscr{Y}. \tag{5.1}$$

Ademais, se μ e ν são medidas σ–finitas, então existe uma única medida possuindo essa propriedade. Esse é o conteúdo do Teorema 5.1. A medida $\mu \otimes \nu$ é dita a **medida produto** de μ e ν. O espaço $(X \times Y, \mathscr{X} \otimes \mathscr{Y})$ é dito o **espaço produto** de $(X, \mathscr{X})$ e $(Y, \mathscr{Y})$. Veremos adiante que a noção de medida produto está intrinsicamente ligada ao importante conceito de independência em teoria da probabilidade.

Teorema 5.1 (Medida Produto). *Se $(X, \mathscr{X}, \mu)$ e $(Y, \mathscr{Y}, \nu)$ são espaços de medida, então existe uma medida $\mu \otimes \nu$ definida em $\mathscr{X} \otimes \mathscr{Y}$ tal que a condição (5.1) é satisfeita. Se μ e ν são medidas σ–finitas, então $\mu \otimes \nu$ é única.*

Lema 5.1. *Seja $(X \times Y, \mathscr{X} \otimes \mathscr{Y})$ o espaço produto de $(X, \mathscr{X})$ e $(Y, \mathscr{Y})$. Valem as seguintes propriedades:*

(i) Se E é um conjunto $\mathscr{X} \otimes \mathscr{Y}$–mensurável então, para cada $x \in X$ fixo, o conjunto

$$\{y \in Y : (x, y) \in E\}.$$

é $\mathscr{Y}$–mensurável;

(ii) Se $f : X \times Y \to \mathbb{R}$ é uma função $\mathscr{X} \otimes \mathscr{Y}$–mensurável então, para cada $x \in X$ fixo, a aplicação

$$y \mapsto f(x, y)$$

é $\mathscr{Y}$–mensurável.

Evidentemente, o Lema acima permanece válido para valores fixados de $y \in Y$, com as devidas modificações.

Será conveniente, em vista do Lema 5.1, introduzirmos a notação

$$\mu(y, E) := \mu\{x \in X : (x, y) \in E\}, \qquad y \in Y, E \in \mathscr{X} \otimes \mathscr{Y}$$

e, semelhantemente,

$$\nu(x, E) := \nu\{y \in Y : (x, y) \in E\}, \qquad x \in X, E \in \mathscr{X} \otimes \mathscr{Y}.$$

Essa notação nos permite interpretar a medida produto de um conjunto $E \in \mathscr{X} \otimes \mathscr{Y}$ como uma decomposição em medidas de 'secções' de E. Veremos mais adiante que o Lema abaixo é um exemplo de *desintegração* da medida $\mu \otimes \nu$.

Lema 5.2. *Sejam $(X, \mathscr{X}, \mu)$ e $(Y, \mathscr{Y}, \nu)$ espaços de medida σ–finitos. Então*

(i) para cada $x \in X$ fixado, a aplicação $E \mapsto \nu(x, E)$ é medida em $\mathscr{X} \otimes \mathscr{Y}$; para cada $E \in \mathscr{X} \otimes \mathscr{Y}$ fixado, a aplicação $x \mapsto \nu(x, E)$ é $\mathscr{X}$–mensurável.

(ii) para cada $y \in Y$ fixado, a aplicação $E \mapsto \mu(y, E)$ é medida em $\mathscr{X} \otimes \mathscr{Y}$; para cada $E \in \mathscr{X} \otimes \mathscr{Y}$ fixado, a aplicação $y \mapsto \mu(y, E)$ é $\mathscr{Y}$–mensurável.

Ademais, vale a identidade

$$\int \nu(x, E)\, d\mu(x) = \mu \otimes \nu(E) = \int \mu(y, E)\, d\nu(y).$$

Teorema 5.2 (Teorema de Tonelli)**.** *Sejam* $(X, \mathscr{X}, \mu)$ *e* $(Y, \mathscr{Y}, \nu)$ *espaços de medida* σ*–finitos. Se* $f\colon X \times Y \to \mathbb{R}$ *é uma aplicação mensurável não negativa, então são também mensuráveis as aplicações* $x \mapsto \int f(x,y)\,d\nu(y)$ *e* $y \mapsto \int f(x,y)\,d\mu(x)$*. Ademais, vale a identidade*

$$\int_X \int_Y f(x,y)\,d\nu(y)d\mu(x) = \int_{X\times Y} f(z)\,d\mu\otimes\nu(z) = \int_Y \int_X f(x,y)\,d\mu(x)d\nu(y).$$

Teorema 5.3 (Teorema de Fubini)**.** *Sejam* $(X, \mathscr{X}, \mu)$ *e* $(Y, \mathscr{Y}, \nu)$ *espaços de medida* σ*–finitos. Se* $f\colon X \times Y \to \mathbb{R}$ *é uma aplicação* $\mu\otimes\nu$*–integrável, então as aplicações* $x \mapsto \int f(x,y)\,d\nu(y)$ *e* $y \mapsto \int f(x,y)\,d\mu(x)$ *(definidas* μ*–qtp e* ν*–qtp, respectivamente) são integráveis, e vale a identidade*

$$\int_X \int_Y f(x,y)\,d\nu(y)d\mu(x) = \int_{X\times Y} f(z)\,d\mu\otimes\nu(z) = \int_Y \int_X f(x,y)\,d\mu(x)d\nu(y).$$

Exercício 5.2. Sejam f e g variáveis aleatórias, ambas definidas em um mesmo espaço de probabilidade $(\Omega, \mathscr{F}, \mathbb{P})$ e tomando valores nos espaços mensuráveis $(X, \mathscr{X})$ e $(Y, \mathscr{Y})$, respectivamente. Dizemos que as variáveis aleatórias f e g são (mutuamente) **independentes** se, e somente se, sua distribuição conjunta é a medida produto das suas marginais, isto é, $\mathbb{P}_{(f,g)} = \mathbb{P}_f \otimes \mathbb{P}_g$. Mostre que, se f e g são variáveis aleatórias reais independentes e integráveis, então fg é integrável e $\mathbb{E}fg = \mathbb{E}f\,\mathbb{E}g$. //

Índice Remissivo

Referências

P. Billingsley. *Convergence of probability measures*, volume 493. John Wiley & Sons, 2009.

B. R. James. *Probabilidade: um curso em nível intermediário*. Instituto de Matemática Pura e Aplicada, CNPq, 1996.

O. Kallenberg. *Foundations of modern probability*. Springer Science & Business Media, 2006.

L. C. G. Rogers and D. Williams. *Diffusions, Markov processes and Martingales*, volume 1. Cambridge University Press, 2000.

J. van Neerven. Stochastic evolution equations. *OpenCourseWare, TU Delft*, 2008.