CAPÍTULO 8

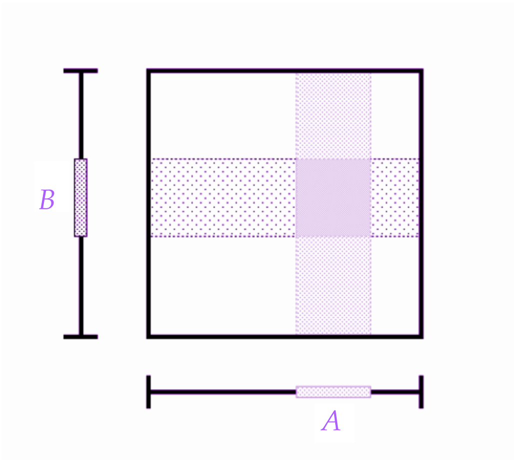
# Medida Produto

Nesse capítulo generalizamos o conceito de medidas para o produto de espaços. A integral que obtemos como subproduto dessas medidas são os análogos abstratos das integrais de várias variáveis e como tais podem ser calculadas como integrais iteradas.

**8.1 Definição** *Dados* $(\Omega_1,\mathcal{F}_1)$ *e* $(\Omega_2,\mathcal{F}_2,)$ *espaços mensuráveis, definimos a* $\sigma$*-álgebra* $\mathcal{F}_1 \otimes \mathcal{F}_2$ *como a* $\sigma$*-álgebra gerada pela coleção dos retângulos mensuráveis*

$$\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma\left\langle \{A \times B : A \in \mathcal{F}_1, B \in \mathcal{F}_2\}\right\rangle$$

Figura 8.1: Retângulos mensuráveis em $\mathcal{F}_1 \otimes \mathcal{F}_2$



Nesse caso também diremos que o produto dos espaços mensuráveis $(\Omega_1, \mathcal{F}_1)$ e $(\Omega_2, \mathcal{F}_2)$ é o espaço mensurável $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$.

Antes de prosseguirmos, observamos que podemos realizar a construção da $\sigma$-álgebra produto de um modo ligeiramente diferente. Para isso note que o produto cartesiano $\Omega_1 \times \Omega_2$ de dois conjuntos $\Omega_1$ e $\Omega_2$ é caracterizado pelas aplicações de projeção associadas

$$\pi_{\Omega_1} : \Omega_1 \times \Omega_2 \to \Omega_1 \quad \pi_{\Omega_2} : \Omega_1 \times \Omega_2 \to \Omega_2.$$

Podemos fazer uso dessas aplicações para construir a $\sigma$-álgebra produto. Dados dois espaços mensuráveis $(\Omega_1, \mathcal{F}_1)$ e $(\Omega_2, \mathcal{F}_2)$, podemos formar uma coleção de subconjuntos em $\Omega_1 \times \Omega_2$ puxando de $\mathcal{F}_1$ :

$$\pi^*_{\Omega_1}(\mathcal{F}_1) := \{\pi^{-1}_{\Omega_1}(E) : E \in \mathcal{F}_1\} = \{E \times \Omega_2 : E \in \mathcal{F}_1\}.$$

Agora, é fácil verificar que $\pi^*_{\Omega_1}(\mathcal{F}_1)$ é uma $\sigma$-álgebra.

Podemos fazer a mesma construção para a projeção no segundo espaço.

**8.2 Teorema (Produto Finito de $\sigma$-Álgebras)** *O produto das $\sigma$-álgebra $\mathcal{F}_1 \otimes \mathcal{F}_2$ é igual a $\sigma$-álgebra gerada pela união do pullback das $\sigma$-álgebras*

$$\mathcal{F}_1 \otimes \mathcal{F}_2 = \langle \pi^*_{\Omega_1}(\mathcal{F}_1) \cup \pi^*_{\Omega_2}(\mathcal{F}_2) \rangle.$$

**Demonstração.** *A demonstração é simples e será deixada como exercício ao leitor.* □

Suponha que $E \subset \Omega_1 \times \Omega_2$. Dado $x \in \Omega_1$ e $y \in \Omega_2$ , definimos a $x$-seção $E^x \subset \Omega_2$ e a $y$-seção $E^y \subset \Omega_1$ de $E$ como

$$E^y = \{x \in \Omega_1 : (x, y) \in E\}$$
$$E^x = \{y \in \Omega_2 : (x, y) \in E\}$$

Conforme indicado na próxima proposição, todas as seções de um conjunto mensurável são mensuráveis.

**8.3 Proposição** *Se $(\Omega_1, \mathcal{F}_1)$ e $(\Omega_2, \mathcal{F}_2)$ são espaços mensuráveis e $E \in \mathcal{F}_1 \otimes \mathcal{F}_2$, então $E^x \in \mathcal{F}_2$ para todo $x \in \Omega_1$ e $E^y \in \mathcal{F}_1$ para todo $y \in \Omega_2$.*

**Demonstração.** *Provaremos apenas para a seção $E^y$; a demonstração para a seção $E^x$ é análoga.*

*Considere a família $\mathcal{M} = \{E \in \mathcal{F}_1 \otimes \mathcal{F}_2 : E^y \in \mathcal{F}_1\}$. Mostraremos que $\mathcal{M}$ é uma $\sigma$-álgebra contendo os retângulos mensuráveis e logo deve ser $\mathcal{F}_1 \otimes \mathcal{F}_2$, e isso é suficiente para demonstrar a proposição.*

- *Suponha que $E = A \times B$ é um retângulo mensurável. Então $E^y = A$ quando $y \in \mathcal{F}_2$, caso contrário $E^y = \emptyset$. Em ambos os casos $E^y \in \mathcal{F}_1$, logo $E \in \mathcal{M}$.*

- *$\emptyset^y = \{x \in \Omega_1 : (x,y) \in \emptyset\} = \emptyset \in \mathcal{F}_1$, assim $\mathcal{M}$ contém $\emptyset$.*
- *Se $E_1, E_2, \ldots \in \mathcal{M}$, então $(\bigcup E_n)^y = \bigcup E_n^y \in \mathcal{F}_1$. Logo $\mathcal{M}$ é fechado em relação a união enumerável.*
- *Se $E \in \mathcal{M}$, e $F = E^{\complement}$, então $F^y = \{x \in \Omega_1 : (x,y) \notin E\} = \Omega_1 \setminus E^y \in \mathcal{F}_1$. Logo $\mathcal{M}$ é fechada sob complementação.* □

**8.4 Proposição** *Suponha que $\mathbb{R}^m$, $\mathbb{R}^n$ estejam equipados com suas $\sigma$-álgebras Borel $\mathcal{B}(\mathbb{R}^m), \mathcal{B}(\mathbb{R}^n)$ e seja $\mathbb{R}^{m+n} = \mathbb{R}^m \times \mathbb{R}^n$. Então*

$$\mathcal{B}(\mathbb{R}^{m+n}) = \mathcal{B}(\mathbb{R}^m) \otimes \mathcal{B}(\mathbb{R}^n).$$

Suponha que $(\Omega_1, \mathcal{F}_1)$ e $(\Omega_2, \mathcal{F}_2)$ sejam espaços mensuráveis. A interseção de retângulos mensuráveis é um retângulo mensurável

$$(A \times B) \cap (C \times D) = (A \cap C) \times (B \cap D),$$

e o complemento de um retângulo mensurável é uma união finita de retângulos

$$(A \times B)^c = (A^c \times B) \cup (A \times B^c) \cup (A^c \times B^c).$$

Assim, a família de uniões finitas de retângulos mensuráveis em $\Omega_1 \times \Omega_2$ forma uma álgebra, que denotamos por $\mathcal{F}_0$. Esta álgebra não é, em geral, uma $\sigma$-álgebra, mas obviamente gera a mesma $\sigma$-álgebra produto que os retângulos mensuráveis.

**8.5 Teorema (Existência da Medida Produto)** *Dados $(\Omega_1, \mathcal{F}_1, \mu_1)$ e $(\Omega_2, \mathcal{F}_2, \mu_2)$ espaços de medidas. Então existe uma única medida $\mu$ em $\mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2$ satisfazendo*

$$\mu(A \times B) = \mu_1(A) \times \mu_2(B).$$

*Diremos que $\mu$ é a medida produto de $\mu_1$ e $\mu_2$. E será denotada por $\mu_1 \otimes \mu_2$.*

**Demonstração.** *Seja $\mathcal{F}_0$ a álgebra dos retângulos, i.e.,*

$$\mathcal{F}_0 = f\left\langle \{A \times B : A \in \mathcal{F}_1, B \in \mathcal{F}_2\} \right\rangle$$

*Como $\mathcal{F}_0$ é uma álgebra, e $\mathcal{F} = \sigma\langle \mathcal{F}_0 \rangle$, basta mostrar que se $A \times B = \bigcup_{i=1}^{\infty} (A_i \times B_i)$ é uma união disjunta, então*

$$\mu(A \times B) = \sum_{i=1}^{\infty} \mu(A_i \times B_i)$$

*Observamos que a função indicadora tem a seguinte propriedade*

$$\mathbb{1}_{A\times B}(x,y) = \sum_{i=1}^{\infty} \mathbb{1}_{A_i\times B_i}(x,y) = \sum_{i=1}^{\infty} \mathbb{1}_{A_i}(x)\mathbb{1}_{B_i}(y)$$

*Integrando sobre $y$ para um $x \in \Omega_1$ fixo e usando o Teorema da Convergência Dominada, temos*

$$\mathbb{1}_A \mu_2(B) = \sum_{i=1}^{\infty} \mathbb{1}_{A_i}(x)\mu_2(B_i).$$

*Integrando com respeito a $x$, temos*

$$\mu_1(A)\mu_2(B) = \sum_{i=1}^{\infty} \mu_1(A_i)\mu_2(B_i).$$

*Pelo Teorema de Extensão de Carathéodory temos que existe uma única medida definida na $\sigma$-álgebra $\mathcal{F}_1 \otimes \mathcal{F}_2$.* □

**8.6 Teorema (Mensurabilidade de Funções com Variável Fixa)** *Sejam $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$ e $(Z, \mathcal{M})$ espaços mensuráveis.*

*Se $f: \Omega_1 \times \Omega_2 \to Z$ é mensurável, então as funções $f^y: \Omega_1 \to Z$, $f^x: \Omega_2 \to Z$ obtidas fixando uma variável também são mensuráveis.*

**Demonstração.** *Provaremos apenas para $f^y$. Seja $I^y: \Omega_1 \to \Omega_1 \times \Omega_2$ a aplicação de inclusão $I^y(x) = (x,y)$. Dado $E \in \mathcal{F}_1 \otimes \mathcal{F}_2$, por 8.3 $(I^y)^{-1}(E) = E^y \in \mathcal{F}_1$, logo $I^y$ é uma função mensurável. Mas $f^y = f \circ I^y$.* □

## 8.1 Áreas de Seções Transversais

**8.7 Teorema (Mensurabilidade da Função Área de Seção Transversal)** *Sejam $(\Omega_1, \mathcal{F}_1, \mu)$, $(\Omega_2, \mathcal{F}_2, \nu)$ espaços de medida. Se $E \in \mathcal{F}_1 \otimes \mathcal{F}_2$,*

- *$\nu(E^x)$ é uma função mensurável de $x \in \Omega_1$.*
- *$\mu(E^y)$ é uma função mensurável de $y \in \Omega_2$.*

**Demonstração.** *Provaremos para $\mu(E^y)$. Também assumiremos que $\mu(\Omega_1) < \infty$. Seja*

$$\mathcal{M} = \{E \in \mathcal{F}_1 \otimes \mathcal{F}_2 : \mu(E^y) \text{ é uma função mensurável de } y\}.$$

*$\mathcal{M}$ é igual à $\mathcal{F}_1 \otimes \mathcal{F}_2$, pois:*

- *Se $E = A \times B$ é um retângulo mensurável, então $\mu(E^y) = \mu(A)\mathbb{1}_{y\in B}$ que é uma função mensurável de $y$. Se $E$ é a união finita disjunta de retângulos mensuráveis $E_n$, então $\mu(E^y) = \sum_n \mu(E_n^y)$ que também mensurável.*

  *Então $\mathcal{M}$ contém a álgebra das uniões finita disjuntas de retângulos mensuráveis.*

- *Se $E_n$ são conjuntos crescentes em $\mathcal{M}$ então $\mu\big((\bigcup_n E_n)^y\big) = \mu(\bigcup_n E_n^y) = \lim_{n\to\infty} \mu(E_n^y)$ é mensurável, e logo $\bigcup_n E_n \in \mathcal{M}$.*

  *De modo análogo, se $E_n$ são conjuntos decrescentes in $\mathcal{M}$, então $\bigcap_n E_n \in \mathcal{M}$. (Aqui é crucial que $E_n^y$ tenham medida finita.)*

- *Assim $\mathcal{M}$ é a classe monótona, contendo a álgebra das uniões finita de retângulos mensuráveis. Pelo Teorema da Classe Monótona (1.29), $\mathcal{M} = \mathcal{F}_1 \otimes \mathcal{F}_2$.*

*Para demonstrar o caso de medida infinita seja $\Omega_n \uparrow \Omega$ com $\mu(\Omega_n) < \infty$, e reaplique o argumento trocando $\mu$ por uma medida finita $\mu_n(F) = \mu(F \cap \Omega_n)$. Então $\mu(E^y) = \lim_{n\to\infty} \mu_n(E^y)$* □

## 8.2 Integrais Iteradas

Até agora, construímos a medida produto que atribuí a um retângulo mensurável a medida que é o produto das medidas de cada um de seus lados. Essa medida foi obtida pelo processo de extensão de Carathéodory, agora apresentaremos uma fórmula integral explícita:

**8.8 Teorema (Medida produto como integral de seções transversais)** *Sejam $(\Omega_1, \mathcal{F}_1, \mu)$ e $(\Omega_2, \mathcal{F}_2, \nu)$ espaços de medida. Então existe uma única medida produto $\mu \otimes \nu \colon \mathcal{F}_1 \otimes \mathcal{F}_2 \to [0, \infty]$, tal que*

$$(\mu \otimes \nu)(E) = \int_{x\in\Omega_1} \underbrace{\int_{y\in\Omega_2} \mathbb{1}_E \, d\nu}_{\nu(E^x)} \; d\mu = \int_{y\in\Omega_2} \underbrace{\int_{x\in\Omega_1} \mathbb{1}_E \, d\mu}_{\mu(E^y)} \; d\nu \,.$$

**Demonstração.** *Denote por $\lambda_1(E)$ a integral dupla à esquerda, e por $\lambda_2(E)$ à integral a direita. Essa integrais existem pelo Teorema 8.7. As medidas $\lambda_1$ e $\lambda_2$ são $\sigma$-aditivas pelo Teorema da Convergência Monótona logo ambas são medidas em $\mathcal{F}_1 \otimes \mathcal{F}_2$. Mais ainda se $E = A \times B$, então expandindo as duas integrais temos $\lambda_1(E) = \mu(A)\nu(B) = \lambda_2(E)$.*

*E pelo Teorema de Unicidade das Medidas de Probabilidade 2.13 temos que $\lambda_1 = \lambda_2 = \mu \otimes \nu$ para todo $\mathcal{F}_1 \otimes \mathcal{F}_2$.* □

**8.9 Teorema (Fubini)** *Sejam $(\Omega_1, \mathcal{F}_1, \mu)$ e $(\Omega_2, \mathcal{F}_2, \nu)$ espaços de medida. Se $\mathrm{X} \colon \Omega_1 \times \Omega_2 \to \mathbb{R}$ é $\mu \otimes \nu$-integrável, então*

$$\int_{\Omega_1\times\Omega_2} \mathrm{X} \, d(\mu \otimes \nu) = \int_{x\in\Omega_1} \left[ \int_{y\in\Omega_2} \mathrm{X}(x,y) \, d\nu \right] d\mu = \int_{y\in\Omega_2} \left[ \int_{x\in\Omega_1} \mathrm{X}(x,y) \, d\mu \right] d\nu \,.$$

**Demonstração.** *Observamos primeiramente que se* $X = \mathbb{1}_E$ *então esse resultado é apenas 8.8.*

*Uma vez que as três integrais são aditivas, e como elas são iguais para as funções simples não negativas* X*, portanto, também o são para todas as funções* X *não negativas, por aproximação e convergência monótona.*

*Para uma função* X *não necessariamente não negativa, use a decomposição* $X = X^+ - X^-$ *como de costume e aplique linearidade. Agora pode acontecer que* $\int_{y\in\Omega_2} X^{\pm}(x,y)d\nu$ *possa ser* $\infty$ *para algum* $x \in \Omega_1$ *e, portanto,* $\int_{y\in\Omega_2} X(x,y)d\nu$ *não estaria definida. No entanto, se* X *é* $\mu\otimes\nu$*-integrável, isso pode acontecer apenas em um conjunto de medida nula em* $\Omega_1$ *(veja o próximo teorema). A integral terá sentido desde que ignoremos este conjunto de medida nula.* □

**8.10 Teorema (Tonelli)** *Sejam* $(\Omega_1, \mathcal{F}_1, \mu)$ *e* $(\Omega_2, \mathcal{F}_2, \nu)$ *espaços de medida, e* $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ *uma função* $\mu\otimes\nu$*-mensurável. Então* $f$ *é* $\mu\otimes\nu$*-integrável se e somente se*

$$\int_{x\in\Omega_1}\left[\int_{y\in\Omega_2} |f(x,y)|\ \mathrm{d}\nu\right] \mathrm{d}\mu < \infty$$

**Demonstração.** *Segue de imediato do Teorema de Fubini aplicado à função não negativa* $|f|$. □

Resumidamente, se uma integral dupla é absolutamente convergente, então é válido mudar a ordem de integração.

**8.11 Exemplo** *Se* X *não é não-negativa, a hipótese de convergência absoluta é crucial para mudar a ordem de integração.*

*Um contra-exemplo elementar é a sequência duplamente indexada* $a_{m,n} = (-m)^n/n!$ *Integrado em relação a medida de contagem temos* $\sum_n a_{m,n} = e^{-m}$*, então* $\sum_m \sum_n a_{m,n} = (1-e^{-1})^{-1}$*, mas* $\sum_m a_{m,n}$ *diverge para todo* $n$.

## 8.3 Distribuição Conjunta

Lembramos a definição de distribuição (conjunta) para vetores aleatórios:

**8.12 Definição (Distribuição de Vetores Aleatórios)** *Se* $X = (X_1, \ldots, X_n)$ *for uma vetor aleatório, então* X *induz uma medida de probabilidade* **P** *em* $\mathbb{R}^n$ *definida tomando a pré-imagem para cada conjunto de Borel* $A$ :

$$\mathbf{P}(A) = \mathbf{P}(X^{-1}(A)) = \mathbf{P}(X \in A). \tag{8.1}$$

*Esta medida de probabilidade* **P** *é denominada* ***distribuição*** *de* X.

**8.13 Definição (Função de Distribuição Conjunta)** *Se* $X = (X_1, \dots, X_n)$ *for uma vetor aleatório, então a função de distribuição conjunta de* $(X_1, \dots, X_n)$

$$F_{(X_1,\dots,X_n)}(x_1, \dots, x_n) = \mathbf{P}(X \le x_i, \dots, X_n \le x_n). \tag{8.2}$$

A partir da definição da independência vemos que a função de distribuição conjunta fatora como o produto de funções de distribuição.

$$F_{(X_1,\dots,X_n)}(x_1, \dots, x_n) = F_{X_1}(x_1) \times \cdots \times F_{X_n}(x_n),$$

**8.14 Teorema** *As variáveis aleatórias* $X_1, \dots, X_n$ *são independentes se e somente se a distribuição conjunta* $\mathbf{P}$ *em* $\mathbb{R}^n$ *é o produto das distribuições* $\mathbf{P}_k$ *de* $X_k$ *em* $\mathbb{R}$.

**Demonstração.** *Na direção* $\Longrightarrow$*, pelo Teorema de unicidade de medidas, basta verificar que* $\mathbf{P}(A) = (\mathbf{P}_1 \otimes \cdots \otimes \mathbf{P}_n)(A)$ *para* $A \in \mathcal{A}$*. Como sabemos, cada* $A \in \mathcal{A}$ *tem a forma* $A = A_1 \times \cdots A_n$*, onde* $A_k$ *é Borel. Então, pela independência,*

$$\begin{aligned}
\mathbf{P}(A_1 \times \cdots \times A_n) &= \mathbf{P}\left((X_1, \dots, X_n) \in A_1 \times \cdots \times A_n\right) && (8.3)\\
&= \mathbf{P}(X_1 \in A_1, \dots, X_n \in A_n) && (8.4)\\
&= \mathbf{P}(X_1 \in A_1) \times \cdots \times \mathbf{P}(X_n \in A_n) && (8.5)\\
&= \mathbf{P}_1(A_1) \times \cdots \times \mathbf{P}_n(A_n). && (8.6)
\end{aligned}$$

*Na outra direção* ($\Longleftarrow$)*, para verificar a independência de* $X_1, \dots, X_n$*, é suficiente considerarmos os semi-intervalos. Ou seja, se* $\mathbf{P}(X_1 \le x_1, \dots, X_n \le x_n)$ *fatorar, o teorema estará demonstrado. Mas este último termo é (pela definição de distribuição conjunta)*

$$\begin{aligned}
\mathbf{P}((-\infty, x_1] \times \cdots \times (-\infty, x_n]) &= \mathbf{P}_1(-\infty, x_1] \times \cdots \times \mathbf{P}_n(-\infty, x_n] && (8.7)\\
&= \mathbf{P}(X_1 \le x_1) \times \cdots \times \mathbf{P}(X_n \le x_n). && (8.8)
\end{aligned}$$

*E assim terminamos.* □

Assim, o modo de variáveis aleatórias serem independentes é quando elas são definidas no espaço produto.

## 8.4 Produtos Infinitos: Teorema de Extensão de Kolmogorov I

Espaços produto são de importância central em diversas áreas da matemática, e na teoria da probabilidade não é diferente. Da construção de variáveis aleatórias independentes ao estudo formal de processos estocásticos, nos deparamos constantemente com necessidade de entender medidas de probabilidade em espaços produto.

Nessa seção nos concentraremos na construção de medidas no espaço $\mathbb{R}^{\mathbb{N}}$, onde todas as ideias centrais são apresentadas. Posteriormente, na Seção 8.8, o argumento que apresentaremos será generalizado para produtos mais gerais.

Para provar o Teorema de extensão de Kolmogorov, seguiremos os passos de Bochner[10, 2], e para isso, antes de entrarmos diretamente nos espaço produto, passaremos rapidamente por algumas definições e resultados que facilitarão nosso trabalho.

Considere então um conjunto $\Omega$ e uma sequência crescente $\{\mathcal{F}_n\}_{n=1}^{\infty}$ de $\sigma$-álgebras em $\Omega$. E para cada $n$, seja $\mu_n$ uma medida de probabilidade na $\sigma$-álgebra $\mathcal{F}_n$.

**8.15 Definição** *Diremos que a sequência $\{(\mathcal{F}_n, \mu_n)\}_{n=1}^{\infty}$ é* ***Kolmogorov consistente*** *se sempre que $m \leq n$ temos $\mathcal{F}_m \subset \mathcal{F}_n$ e $\mu_n|_{\mathcal{F}_m} = \mu_m$.*

Dado $\mathcal{A} = \bigcup_{n=1}^{\infty} \mathcal{F}_n$, uma **extensão** da sequência de probabilidades $\{\mu_n\}_{n=1}^{\infty}$ é uma medida $\mu$ em $\sigma\langle\mathcal{A}\rangle$ satisfazendo $\mu_n = \mu|_{\mathcal{F}_n}$.

**8.16 Teorema (Teorema de Extensão de Bochner)** *Dado $\{(\mathcal{F}_n, \mu_n)\}_{n=1}^{\infty}$ uma sequência Kolmogorov consistente. Suponha que existe uma classe compacta $\mathcal{K} \subseteq \bigcup_{n=1}^{\infty} \mathcal{F}_n$ de $\Omega$ tal que*

$$\mu_n(A) = \sup\{\mu_n(C) : C \in \mathcal{K} \cap \mathcal{F}_n \text{ e } C \subset A\}.$$

*Então existe uma única extensão de Kolmogorov para a $\sigma$-álgebra $\sigma\langle\bigcup_{n=1}^{\infty} \mathcal{F}_n\rangle$.*

**Demonstração.**

*Defina $\mu$ na álgebra $\mathcal{A} = \bigcup_{n=1}^{\infty} \mathcal{F}_n$ por $\mu(E) = \mu_n(E)$ para $E \in \mathcal{F}_n$. A condição de consistência de Kolmogorov garante que $\mu$ está bem definida. Também é simples ver que $\mu(\emptyset) = 0$ e $\mu(\Omega) = 1$.*

*A demonstração que $\mu$ é uma probabilidade finitamente aditiva é direta. Basta observar que se $A_i \in \mathcal{A}$ para $i = 1, \ldots, n$ então existe $m$ tal que $A_i \in \mathcal{F}_m$ para todo $i = 1, \ldots, n$.*

*Para concluir a demonstração, observamos que se $A \in \mathcal{A}$, então existe $n$ tal que $A \in \mathcal{F}_n$ e*

$$\begin{aligned}
\mu(A) &= \mu_n(A) \\
&= \sup\{\mu_n(C) : C \in \mathcal{K} \cap \mathcal{F}_n \text{ e } C \subset A\} \\
&\leq \sup\{\mu_n(C) : C \in \mathcal{K} \text{ e } C \subset A\} \\
&\leq \mu(A).
\end{aligned}$$

*Assim, pelo Teorema de Extensão Compacta (2.28) temos que $\mu$ é pré-medida e finalmente pelo Teorema de Extensão de Carathéodory podemos estender a uma probabilidade em $\sigma\langle\mathcal{A}\rangle$.* □

Postas estas considerações iniciais, podemos nos concentrar no objeto central dessa seção, ou seja, medidas de probabilidade em

$$\mathbb{R}^{\mathbb{N}} = \{(\omega_1, \omega_2, \dots) : \omega_i \in \mathbb{R}, i \in \mathbb{N}\}.$$

O primeiro passo é eleger uma $\sigma$-álgebra onde definiremos nossa medida, e para isso

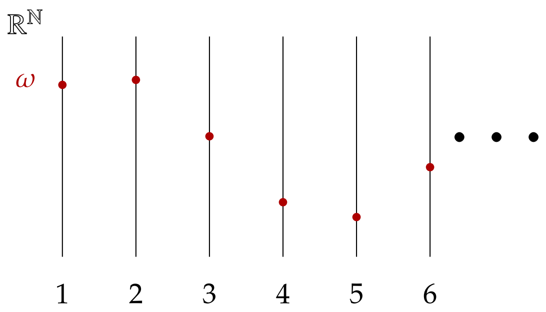



Figura 8.2: Representação visual de $\mathbb{R}^{\mathbb{N}}$.

temos um candidato natural. Nós equiparemos $\mathbb{R}^{\mathbb{N}}$ com a $\sigma$-álgebra produto gerada pelos retângulos finito dimensionais

**8.17 Definição** *A $\sigma$-álgebra produto gerada pelos retângulos finito dimensionais é dada por*

$$\mathcal{C} = \{A : A = \{\omega : \omega_i \in (a_i, b_i] \text{ para } i = 1, \dots, k \quad k \in \mathbb{N}\}\},$$

*onde* $-\infty \leq a_i < b_i \leq \infty$.

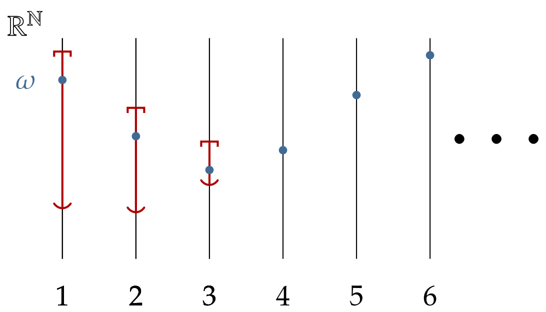



Figura 8.3: Representação visual de um cilindro em $\mathbb{R}^{\mathbb{N}}$. Em azul um ponto do cilindro.

**8.18 Observação** *Esta é, de fato, a $\sigma$-álgebra com a qual trabalhamos intuitivamente nos cursos mais básicos de probabilidade. Lembre, por exemplo, que quando queremos estudar o "lançamento de infinitas moedas", sempre descrevemos os eventos de interesse a partir de eventos que dependem do* *resultado de finitos lançamentos. Mesmo eventos do tipo $A = \{$observamos cara infinitas vezes$\}$ é escrito como*

$$A = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n,$$

*com $A_n = \{$observamos cara na $n$-ésima jogada$\}$.*

Dando sequência, vamos começar a preparar o caminho para usar o teorema 8.16 e introduziremos uma notação particularmente útil. Dado $B \subset \mathbb{R}^n$, denotaremos por $B \times \mathbb{R}^{\mathbb{N}-n}$ o conjunto:

$$B \times \mathbb{R}^{\mathbb{N}-n} := \{(\omega_1, \omega_2, \dots) \text{ tais que } (\omega_1, \dots, \omega_n) \in B\}$$

É fácil ver que se $A \in \mathcal{C}$ então existe $n \in \mathbb{N}$, e $B \in \mathcal{B}(\mathbb{R}^n)$ tal que $A = B \times \mathbb{R}^{\mathbb{N}-n}$.

Considere agora a sequência encaixante de $\sigma$-álgebras:

$$\mathcal{F}_n = \mathcal{B}(\mathbb{R}^n) \times \mathbb{R}^{\mathbb{N}-n} = \{A \times \mathbb{R}^{\mathbb{N}-n}, A \in \mathcal{B}(\mathbb{R}^n),\}$$

e note que dado $m < n$ e $B \in \mathcal{B}(\mathbb{R}^n)$, então

$$B \times \mathbb{R}^{\mathbb{N}-m} = (B \times \mathbb{R}^{n-m}) \times \mathbb{R}^{\mathbb{N}-n} \in \mathcal{F}_n,$$

de modo que $\mathcal{F}_m \subset \mathcal{F}_n$, e $\mathcal{A} = \bigcup_{n=1}^{\infty} \mathcal{F}_n$ é uma álgebra, com $\mathcal{C} \subseteq \mathcal{A}$.

**8.19 Proposição** $\sigma\langle\mathcal{A}\rangle = \sigma\langle\mathcal{C}\rangle$.

**Demonstração.** *Uma das inclusões é trivial. De fato, uma vez que $\mathcal{C} \subset \mathcal{A}$, segue trivialmente que $\sigma\langle\mathcal{C}\rangle \subset \sigma\langle\mathcal{A}\rangle$. Para ver a inclusão contrária, dado $n \geq 1$, faça $\mathcal{G}_n = \{A \in B(\mathbb{R}^n) : A \times \mathbb{R}^{\mathbb{N}-n} \in \sigma\langle\mathcal{C}\rangle\}$. Como $\mathcal{G}_n$ é uma $\sigma$-álgebra contendo os retângulos do tipo*

$$\Pi_{k=1}^{n}(a_k, b_k] \subset \mathbb{R}^n,$$

*então $\mathcal{G}_n = \mathcal{B}(\mathbb{R}^n)$. Segue que $\mathcal{F}_n \subseteq \sigma\langle\mathcal{C}\rangle$ para todo $n \geq 1$ e*

$$\sigma\langle\mathcal{A}\rangle \subseteq \sigma\langle\mathcal{C}\rangle.$$

**8.20 Teorema (Teorema de Extensão de Kolmogorov)** *Dadas medidas de probabilidade $\mathbf{P}_n$ em $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ tais que*

$$\mathbf{P}_{n+1}((a_1, b_1] \times \cdots \times (a_n, b_n] \times \mathbb{R}) = \mathbf{P}_n((a_1, b_1] \times \cdots \times (a_n, b_n]),$$

*para todo $n \geq 1$, então existe uma única medida de probabilidade $\mathbf{P}$ em $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^N))$ com*

$$\mathbf{P}(\omega : \omega_i \in (a_i, b_i], 1 \leq i \leq n) = \mathbf{P}_n((a_1, b_1] \times \cdots \times (a_n, b_n])$$

**Demonstração.** *Para começar, dado $n \geq 1$ defina a medida de probabilidade $Q_n$ em $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ por*

$$Q_n(A) = \mathbf{P}_{n+1}(A \times \mathbb{R}).$$

*Como $Q_n$ coincide com $\mathbf{P}_n$ no $\pi$-sistema dos retângulos do tipo $\Pi_{k=1}^n (a_k, b_k]$, então $Q_n$ e $\mathbf{P}_n$ coincidem em todo $\mathcal{B}(\mathbb{R}^n)$. Segue que para todo $n \geq 1$ e $A \in \mathcal{B}(\mathbb{R}^n)$,*

$$\mathbf{P}_{n+1}(A \times \mathbb{R}) = \mathbf{P}_n(A),$$

*e por indução*

$$\mathbf{P}_{n+m}(A \times \mathbb{R}^m) = \mathbf{P}_n(A).$$

*Com a notação introduzida anteriormente, defina a medida $\mu_n$ em $(\mathbb{R}^{\mathbb{N}}, \mathcal{F}_n)$ por*

$$\mu_n(B \times \mathbb{R}^{\mathbb{N}-n}) = \mathbf{P}_n(B).$$

*Assim se $m < n$ então para $B \in \mathcal{F}_m$,*

$$\mu_n(B \times \mathbb{R}^{\mathbb{N}-m}) = \mu_n(B \times \mathbb{R}^{n-m} \times \mathbb{R}^{\mathbb{N}-n}) = \mathbf{P}_n(B \times \mathbb{R}^{n-m}) = \mathbf{P}_m(B) = \mu_m(B \times \mathbb{R}^{\mathbb{N}-m}).$$

*Concluímos assim que a sequência $\{(\mathcal{F}_n, \mu_n)\}_{n\geq 1}$ é Kolmogorov consistente.*

*Considere agora as classes de cilindros*

$$\mathcal{K}_n = \{K \times \mathbb{R}^{\mathbb{N}-n}, K \in \mathcal{B}(\mathbb{R}^n) \text{ compacto }\} \quad \text{e} \quad \mathcal{K} = \bigcup_{n\geq 1} \mathcal{K}_n.$$

*Observe que se $K \in \mathcal{K}_n$ e $L \in \mathcal{K}_m$, com $m < n$ então $K \cap L \in \mathcal{K}_n$.*

*Afirmamos que classe $\mathcal{K}$ é compacta, o que provaremos no lema a seguir.*

*Dado agora um conjunto $A \in C$ então $A = D \times \mathbb{R}^{\mathbb{N}-n}$ com $D \in \mathcal{B}(\mathbb{R}^n)$, para algum $n \geq 1$, e logo existe um $K \subset D \in \mathcal{B}(\mathbb{R}^n)$ compacto, tal que $\mu_n(K \times \mathbb{R}^{\mathbb{N}-n}) = \mathbf{P}_n(K) < \mathbf{P}_n(D) + \epsilon = \mu_n(A) + \epsilon$. Consequentemente*

$$\mu_n(A) = \sup\{\mu_n(C) : C \in \mathcal{K} \cap \mathcal{F}_n \text{ e } C \subset A\}.$$

*e o teorema segue do Teorema 8.16.*

Dado $A \in \mathcal{A}$ não vazio, definimos $\dim(A) := \min\{n \geq 1 : A \in \mathcal{F}_n\}$, e $\dim(\mathbb{R}^{\mathbb{N}}) = 0$.

Observe que se $A_1, A_2 \in \mathcal{A}$ são tais que $A_1 \cap A_2 \neq \emptyset$, então $\dim(A_1 \cap A_2) = \max\{\dim(A_1), \dim(A_2)\}$. De fato, se $m < n$

$$(B_1 \times \mathbb{R}^{\mathbb{N}-m}) \cap (B_2 \times \mathbb{R}^{\mathbb{N}-n}) = ((B_1 \times \mathbb{R}^{n-m}) \cap B_2) \times \mathbb{R}^{\mathbb{N}-n} \in \mathcal{F}_n.$$

**8.21 LEMA** *A classe $\mathcal{K}$ definida na demonstração anterior como*

$$\mathcal{K}_n = \{K \times \mathbb{R}^{\mathbb{N}-n}, K \in \mathcal{B}(\mathbb{R}^n) \text{ compacto }\} \quad \text{e} \quad \mathcal{K} = \bigcup_{n\geq 1} \mathcal{K}_n.$$

*é compacta.* ◁

**Demonstração.** *Para ver isso considere uma sequência de conjuntos* $\{C_i\}_{n=1}^{\infty} \in \mathcal{K}$ *e suponha que para todo* $m \geq 1$, $B_m = \bigcap_{i=1}^{m} C_i \neq \emptyset$. *Assim, para todo* $m \geq 1$ *e* $j \geq 1$, *temos que a projeção na* $j$*-ésima coordenada é não vazia, i.e.,* $\pi_j(B_m) \neq \emptyset$.

*Temos agora dois casos:* $\sup\{\dim(B_m); m \geq 1\} = d < \infty$ *ou* $\sup\{\dim(B_m); m \geq 1\} = \infty$.

*No primeiro caso, como a sequência* $(\dim(B_m))_{n\geq 1}$ *é não-decrescente, então existe* $m_0 \geq 1$ *tal que* $\dim(B_m) = d$ *para todo* $m \geq m_0$. *Assim, para todo* $m \geq m_0$ *existe* $K_m \subset \mathbb{R}^d$, *compacto de* $\mathbb{R}^d$ *tal que* $B_m = K_m \times R^{\mathbb{N}-d}$. *Vale assim que*

$$\bigcap_{m=m_0}^{\infty} K_m \neq \emptyset, \quad e \quad \bigcap_{i=1}^{\infty} C_i = \bigcap_{m=m_0}^{\infty} B_m = \left(\bigcap_{m=m_0}^{\infty} K_m\right) \times \mathbb{R}^{\mathbb{N}-d} \neq \emptyset.$$

*Agora, se* $\sup\{\dim(B_m); m \geq 1\} = \infty$, *então para todo* $j \geq 1$ *existe* $m_j \geq 1$ *tal que* $\dim(B_m) \geq j$, *para todo* $m \geq m_j$. *Em outras palavras, se* $m \geq m_j$, *então* $B_m = K_m \times \mathbb{R}^{\mathbb{N}-d}$ *para algum compacto* $K_m$ *de* $\mathbb{R}^d$ *e algum* $d \geq j$. *Mas isso significa que* $\pi_j(B_m)$ *é um compacto de* $\mathbb{R}$ *para todo* $m \geq m_j$. *Além disso, como* $B_m \neq \emptyset$, *então* $\pi_j(B_m) \neq \emptyset$ *e*

$$\bigcap_{m=m_j}^{\infty} \pi_j(B_m) \neq \emptyset.$$

*Como as projeções* $\pi_j$ *são contínuas e a sequência* $(B_m)_{m\geq 1}$ *é decrescente, então para todo* $j \geq 1$

$$\pi_j\left(\bigcap_{i=1}^{\infty} C_i\right) = \pi_j\left(\bigcap_{m=m_j}^{\infty} B_m\right) = \bigcap_{m=m_j}^{\infty} \pi_j(B_m) \neq \emptyset.$$

*Segue, portanto que* $\bigcap_{i=1}^{\infty} C_i \neq \emptyset$ *e* $\mathcal{K}$ *é uma classe compacta.*

## 8.5 Existência de Variáveis Aleatórias Independentes

**8.22 TEOREMA** *Dadas probabilidades* $\mathbf{P}_1, \ldots, \mathbf{P}_n$ *em* $\mathbb{R}$, *então existem variáveis aleatórias independentes* $X_1, \ldots, X_n$ *com distribuições* $\mathbf{P}_1, \ldots, \mathbf{P}_n$.

**Demonstração.** *Seja* $\mathbf{P} = \mathbf{P}_1 \otimes \cdots \otimes \mathbf{P}_n$ *a medida de probabilidade produto em* $\mathbb{R}^n$. *Como já provamos, toda medida de probabilidade em* $\mathbb{R}^n$ *é uma distribuição de alguma variável aleatória, veja Teorema 6.15.*

*Vamos definir uma variável aleatória (ou melhor, um vetor aleatório). Começamos definindo o espaço de probabilidade* $(\Omega, \mathcal{F}, \mathbf{P})$ *como* $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathbf{P})$.

*Então, agora, precisamos definir uma variável aleatória* $X$ *que é uma função desse conjunto. Seja* $X(x) = x$. *Como* $X$ *é a função identidade, então* $X \in B$ *se e somente se* $x \in B$. *Então* $\mathbf{P}(X \in B) = \mathbf{P}(B)$, *então a distribuição de* $X$ *é igual a* $\mathbf{P}$. *Considere* $X = (X_1, \ldots, X_n)$ *então as funções componentes* $X_1, \ldots, X_n$ *são independentes, pelo teorema anterior.* □

Podemos generalizar o resultado anterior para sequências de variáveis aleatórias i.i.d.

**8.23 Teorema ( Sequências de Variáveis i.i.d.)** *Seja $F$ uma distribuição em $\mathbb{R}$, então existe uma sequência de variáveis aleatórias independentes $X_1, \ldots, X_n, \ldots$ com distribuição $F$.*

**Demonstração.** *Exercício. Repita o argumento do teorema anterior usando o Teorema de Extensão de Kolmogorov.* □

Agora vamos demonstrar alguns fatos que mostram como a teoria desenvolvida até esse ponto reflete em densidades de variáveis aleatórias.

1. Uma função mensurável positiva $f \geq 0$ é a densidade de uma variável aleatória X se $\mathbf{P}(X \in A) = \int_A f(x) \quad \mathrm{d}x$, para cada conjunto de Borel $A \subset \mathbb{R}$. Claro, a densidade é definida apenas q.c.

2. A definição é a mesma para um vetor aleatório: se X for um vetor aleatório e $\mathbf{f}$ está definido em $\mathbb{R}^n$ e $A$ é um Boreliano de $\mathbb{R}^n$.

3. Densidade conjunta de $n$ variáveis aleatórias $f_1, \ldots, f_n$ é a densidade do vetor aleatório $(X_1, \ldots, X_n)$.

**8.24 Teorema** *Sejam $X_1, \ldots, X_n$ variáveis aleatórias com densidades $f_1, \ldots, f_n$ e densidade conjunta $\mathbf{f}$. Então $X_1, \ldots, X_n$ são independentes se e somente se $\mathbf{f}(x_1, \ldots, x_n) = f_1(x_1) \cdot \cdots \cdot f_n(x_n)$ para quase todo $(x_1, \ldots, x_n)$ em $\mathbb{R}^n$.*

Existem algumas dificuldades técnicas. Para obter uma densidade de uma distribuição, precisamos diferenciar. Se temos funções que não são contínuas, então a diferenciação é um problema. Precisamos do seguinte resultado da teoria das medidas:

**8.25 Teorema (Teorema de Diferenciação de Lebesgue)** *Seja $f : \mathbb{R} \to \mathbb{R}$ uma função Lebesgue integrável. Então*

$$\frac{1}{\epsilon} \int_x^{x+\epsilon} X(y)\, dy$$

*vai convergir para $X(x)$, para quase todo $x \in \mathbb{R}$.*

Esta é uma versão unidimensional do Teorema de diferenciação de Lebesgue. Você também possui a versão $n$-dimensional.

**8.26 Teorema** *Se $f : \mathbb{R}^n \to \mathbb{R}$ é Lebesgue integrável, então*

$$\frac{1}{\mathrm{vol}(Q)} \int_Q f(y)\, dy \to X(x)$$

*para quase todos os $x$, onde $Q$ é um cubo contendo $x$.*

**Demonstração (Prova do Teorema 8.24).** *Sem perda de generalidade, podemos assumir* $n = 2$.

X *tem densidade* $f$, Y *tem densidade* $g$ *e* (X, Y) *tem densidade* $\varphi$. *Isso significa que* $P(X \in A) = \int_A f(x)\ dx$, $P(Y \in B) = \int_B g(y)\,dy$ *e* $P((X, Y) \in A \times B) = \int_{A\times B} \varphi(x, y)\ dx\ dy$

*Então, vamos começar a demonstrar a implicação. Então supomos que* X *e* Y *são independentes. Então,* $\int_{A\times B} \varphi(x, y)\ dx\ dy$ *deve ser igual a, por definição, o produto das probabilidades individuais* $\int_A f(x)\ dx \times \int_B g(y)\,dy$. *Faça* $A = [x_0, x_0 + \epsilon]$ *e* $B = [y_0, y_0 + \epsilon]$. *Então* $A \times B$ *é exatamente um cubo* $Q$ *que contém* $(x_0, y_0)$.

*Fazendo* $\epsilon \longrightarrow 0$, *obtemos que o lado esquerdo será* $\phi(x_0, y_0)$ *e o lado direito será* $f(x_0)g(y_0)$ *para quase todos* $(x_0y_0)$.

*Para demonstrar a recíproca. Assumiremos que a densidade conjunta é o produto das duas densidades para quase todo* $(x, y) \in \mathbb{R}^2$. *Agora, usaremos o Teorema de Fubini.*

$$
\begin{aligned}
P(X \in A, Y \in B) &= \int_{A\times B} \varphi(x, y)\ dx\ dy && (8.9)\\
&= \int_{A\times B} f(x)g(y)\, dx\ dy && (8.10)\\
&= \int_A f(x)\, dx \int_B g(y)\, dy, \text{ por Fubini} && (8.11)\\
&= P(X \in A) \cdot P(Y \in B). && (8.12)
\end{aligned}
$$

*Então, por definição,* X *e* Y *são independentes.* □

Vejamos alguns exemplos.

**8.27 Exemplo (Distribuição Uniforme)** *Sejam* X *e* Y *variáveis aleatórias independentes, distribuídas uniformemente em* [0,1]. *Assim, a densidade de* X *e* Y *é dada pela função de densidade*

$$f(x) = \begin{cases} 1, & x \in [0,1] \\ 0, & \textit{caso contrário.} \end{cases}$$

*Portanto, a densidade conjunta de* X *e* Y *é*

$$\varphi(x, y) = \begin{cases} 1, & \textit{se } (x, y) \in [0,1] \times [0,1] \\ 0, & \textit{caso contrário.} \end{cases}$$

**8.28 Exemplo (Distribuição Gaussiana em $\mathbb{R}^n$)** *Também denominada de distribuição normal multivariada. A densidade de cada coordenada* $X_k$ *é dada pela gaussiana padrão unidimensional*

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, x \in \mathbb{R}$$

*e as coordenadas são independentes, então a densidade de* $(X_1, \dots, X_n)$ *é*

$$\varphi(x_1, \dots, x_n) = \frac{1}{(2\pi)^{1/2}} e^{-\frac{1}{2}|x|^2},$$

*onde* $|x| = x_1^2 + x_2^2 + \cdots + x_n^2$.

Faremos agora uma demonstração alternativa da afirmação de que para variáveis independentes a esperança do produto é o produto das esperanças. Compare essa demonstração com a demonstração do Teorema 7.21.

**8.29 Teorema** *Sejam* $X, Y \in \mathcal{L}^1(\mathbf{P})$ *independentes. Então* $(XY) \in \mathcal{L}^1(\mathbf{P})$ *e*

$$\mathbf{E}[XY] = \mathbf{E}[X] \times \mathbf{E}[Y].$$

**Demonstração.** *Sejam* $\mathbf{P}_X$ *e* $\mathbf{P}_Y$ *as distribuições de* $X$ *e* $Y$ *respectivamente. Então* $\mathbf{P}_X$ *e* $\mathbf{P}_Y$ *são medidas em* $\mathbb{R}$. *Então*

$$\mathbf{E}[X] = \int x \, d\mathbf{P}_X$$

*e*

$$\mathbf{E}[Y] = \int y \, d\mathbf{P}_Y,$$

*e, por outro lado*

$$\mathbf{E}[XY] = \int xy \, \mathrm{d}(\mathbf{P}_X \times \mathbf{P}_Y)$$

*uma vez que* $X$ *e* $Y$ *são variáveis aleatórias independentes. Por Fubini, podemos escrever a última integral como uma integral iterada, que é exatamente* $\mathbf{E}[X] \times \mathbf{E}[Y]$.

*Porquê podemos usar Fubini?* □

## 8.6 Convolução e Soma de Variáveis Aleatórias Independentes

Vamos começar com um exemplo. Se X e Y são variáveis aleatórias com distribuições conhecidas, qual é a distribuição de $X+Y$ é determinada? Certamente a expectativa é determinada (pela propriedade da esperança da soma). Mas a distribuição não é completamente determinada. Considere o seguinte exemplo: tome qualquer variável aleatória X, tal que $X \overset{d.}{=} -X$. Um exemplo de tal variável é o lançamento de uma moeda, onde estaremos denotando cara por 1 e coroa por $-1$.

Considere agora $X + X$ e $X + (-X)$. Apesar dos quatro somandos anteriores possuírem a mesma distribuição a primeira soma é igual a $2X$ e a segunda é igual a 0.

Ou seja, conhecer a distribuição dos somandos não é suficiente. Você realmente precisa da distribuição conjunta. Se X e Y forem independentes, então podemos calcular a distribuição da soma.

Se X e Y forem independentes podemos então calcular a distribuição de X + Y.

**8.30 Teorema** *Se X e Y são variáveis aleatórias independentes com funções de distribuição $F$ e $G$, respectivamente. Então, a função de distribuição da soma X + Y é dada por*

$$H(z) = \int_{\mathbb{R}} F(z-y)dG(y) = \int_{\mathbb{R}} G(z-y)dF(y).$$

**Demonstração.** *A distribuição conjunta de X e Y é $\mathbf{P}_X \otimes \mathbf{P}_X$, uma vez que as variáveis aleatórias são independentes. Então $H(z) = \mathbf{P}(X+Y \leq z)$ é igual a $\mathbf{P}((X,Y) \in \{(x,y) \in \mathbb{R}^2 : x+y \leq z\})$.*

*Então*

$$H(z) = \int_{\{(x,y)\in\mathbb{R}^2 : x+y\leq z\}} d(\mathbf{P}_X \times \mathbf{P}_Y) \tag{8.13}$$

$$= \int \int_{\mathbb{R}} \mathbf{1}_{\{(x,y)\in\mathbb{R}^2 : x+y\leq z\}} d(\mathbf{P}_X \times \mathbf{P}_Y) \tag{8.14}$$

$$= \int_{\mathbb{R}} d\mathbf{P}_Y \cdot \int_{\mathbb{R}} \mathbf{1}_{\{(x,y)\in\mathbb{R}^2 : x+y\leq z\}} d\mathbf{P}_X \quad (\textit{por Fubini}) \tag{8.15}$$

$$= \int_{\mathbb{R}} d\mathbf{P}_Y \cdot \int_{\mathbb{R}} \mathbf{1}_{\{(x,y)\in\mathbb{R}^2 : x\leq z-y\}} d\mathbf{P}_X \tag{8.16}$$

$$= \int_{\mathbb{R}} d\mathbf{P}_Y \cdot \int_{\mathbb{R}} \mathbf{P}_X(X \in (-\infty, z-y]) d\mathbf{P}_X \tag{8.17}$$

$$= \int_{\mathbb{R}} F(z-y) d\mathbf{P}_Y(y) \tag{8.18}$$

*Agora observamos que a integral em relação a uma medida é a mesma que a integral Stieltjes em relação a uma função de distribuição. e assim* $= \int_{\mathbb{R}} F(z-y)\,\mathrm{d}G(y)$. □

**8.31 Corolário** *Se X e Y tiverem densidades $f$ e $g$, respectivamente, então X + Y tem densidade*

$$Z(x) = \int_{\mathbb{R}} f(x-y)g(y)\,dy.$$

A integral anterior é denotada por $h = f \star g$ e é conhecida como **convolução** de $f$ e $g$.

**Demonstração (Prova do Corolário).** *fazer.* □

**8.32 Exemplo** *Sejam X e Y variáveis aleatórias uniformes independentes em* [0,1]. *Calcule a soma das duas variáveis. A densidade de X + Y é dada por*

$$Z(x) = \int_{\mathbb{R}} \mathbf{1}_{[0,1]}(x-y)\mathbf{1}_{[0,1]}(y)$$

*Então $0 \leq y \leq 1$, mas também $0 \leq x-y \leq 1$, então $y \leq x$ e $y \geq x-1$.*

*A soma terá valores entre 0 e 2. Portanto, $x \in [0,2]$.*

*Se $x \in [0,1]$, então a única condição a ser satisfeita é $y \leq x$, então temos* $\displaystyle\int_0^x dy = x$.

*Se $x \in [1,2]$, então temos* $\displaystyle\int_{x-1}^1 dy = 2 - x$. ◁

Nesta seção discutimos uma técnica para aproximar funções integráveis arbitrárias por funções suaves.

**8.33 Definição (Funções Teste)** *Uma função é dita **função teste** ou suave de suporte compacto, denotada por $f \in C_c^\infty(\mathbb{R})$ se $f$ é de classe $C^\infty$ e seu suporte*

$$\operatorname{supp}(f) := \overline{\{x \in X \mid f(x) \neq 0\}} = \overline{f^{-1}\left(\{0\}^c\right)}.$$

*é um conjunto compacto.*

Para começar, estabelecemos a existência de uma função teste em $[-1,1]$.

**8.34 Lema** *A função*

$$f(x) = \begin{cases} e^{\frac{-1}{1-x^2}} & |x| < 1 \\ 0 & |x| \geq 1 \end{cases}$$

*é de suporte compacto em $[-1,1]$ e possui derivada contínua em todos os pontos.* ◁

**8.35 Lema** *Seja $\rho(x)$ uma função positiva em $C_c^\infty(\mathbb{R})$ tal que $\rho(x)$ é tem suporte em $[-1,1]$ e*

$$\int_{-\infty}^{\infty} \rho(x)\, dx = \int_{-1}^{1} \rho(x)\, dx = 1.$$

*Seja $f : \mathbb{R} \longrightarrow \mathbb{R}$ uma função contínua. Defina*

$$f_n(x) = n \int_{-n}^{n} \rho(n(x-y)) f(y) dy$$

*Então $f_n \in C_c^\infty(\mathbb{R})$, $f_n^{(m)}(x) = n \displaystyle\int_{-n}^{n} \rho^{(m)}(n(x-y)) f(y) dy$ e $f_n$ convergem para $f$ uniformemente em conjuntos compactos. Além disso, se $f$ for limitada, então $\|f_n\|_\infty \leq \|f\|_\infty$.* ◁

**Demonstração.** *Primeiro, note que, como $\rho(x)$ e todas as suas derivadas são compactas, elas também são limitadas. Em particular, existe um $M > 0$ tal que $|\rho'(x)| \leq M$. Para limpar a notação um pouco, defina $\rho_n(y) = n\rho(ny)$ e assim temos*

$$f_n(x) = \int_{-n}^{n} \rho_n(x-y) f(y) dy$$

*Como o suporte de $\rho_n(x)$ está contido em $[-\frac{1}{n}, \frac{1}{n}]$, logo se consideramos $\rho_n(x-y)$ como uma função de $y$, seu suporte é contido em $[x-\frac{1}{n}, x+\frac{1}{n}]$. Assim, o suporte de $f_n(x)$ está contido em $[-n-\frac{1}{n}, n+\frac{1}{n}]$.*

*Para examinar a derivada de $f_n(x)$, escolha $h > 0$ e seja*

$$\frac{f_n(x+h) - f_n(x)}{h} = \frac{1}{h}\int_{-n}^{n} (\rho_n(x+h-y) - \rho_n(x-y)) f(y)dy$$

*O Teorema de Taylor nos diz que $\frac{1}{h}(\rho_n(x+h-y) - \rho_n(x-y)) = \rho_n'(c)$ para algum $c \in [x+h-y, x-y]$. Assim sendo, $\left|\frac{1}{h}(\rho_n(x+h-y) - \rho_n(x-y))f(y)\right| \leq M\,|f(y)|$ e pela integrabilidade de $f(y)$ no intervalo $[-n, n]$ podemos usar a Convergência Dominada para concluir que*

$$\begin{aligned} f_n'(x) &= \lim_{h\to 0} \frac{f_n(x+h) - f_n(x)}{h} \\ &= \lim_{h\to 0} \frac{1}{h}\int_{-n}^{n}(\rho_n(x+h-y) - \rho_n(x-y))f(y)dy \\ &= \int_{-n}^{n} \lim_{h\to 0} \frac{1}{h}(\rho_n(x+h-y) - \rho_n(x-y))f(y)dy \\ &= \int_{-n}^{n} \rho_n'(x-y)f(y)dy \end{aligned}$$

*A continuidade de $f_n'(x)$ segue da continuidade de $f(y)$ e $\rho_n'(x-y)$ e o Teorema da Convergência Dominada como acima. Uma indução simples estende o resultado para derivadas de ordem arbitrária.*

*Em seguida, mostraremos a convergência. Escolha um conjunto compacto $K \subset \mathbb{R}$ e $\varepsilon > 0$, Como $f$ é uniformemente contínua em $K$, existe um $\delta > 0$ tal que para qualqueis $x,y \in K$ temos que se $|x-y| \leq \delta$ então $|f(x) - f(y)| \leq \varepsilon$. Escolha $N_1 > 0$ tal que $\frac{1}{n} < \delta$ para todo $n \geq N_1$. A hipótese $\displaystyle\int_{-\infty}^{\infty} \rho(y)\,dy = \int_{-1}^{1} \rho(y)\,dy = 1$ e por uma mudança de variáveis temos $\displaystyle\int_{-\infty}^{\infty} \rho_n(x-y)\,dy = \int_{x-\frac{1}{n}}^{x+\frac{1}{n}} \rho_n(x-y)\,dy = 1$ para todos $x \in \mathbb{R}$ e $n > 0$. Escolha $N_2 > 0$ de modo que para todos $n > N_2$, temos $K \subset [-n+\frac{1}{n}, n-\frac{1}{n}]$. Portanto, podemos escrever $f(x) = \displaystyle\int_{-n}^{n} \rho_n(x-y)f(x)\,dy = 1$ para* *qualquer $x \in K$ e $n > N_2$. Temos então que para qualquer $n \geq \max(N_1, N_2)$*

$$
\begin{aligned}
|f_n(x) - f(x)| &= \left| \int_{-n}^{n} (\rho_n(x-y)f(y) - \rho_n(x-y)f(x))\, dy \right| \\
&= \left| \int_{x-\frac{1}{n}}^{x+\frac{1}{n}} (\rho_n(x-y)f(y) - \rho_n(x-y)f(x))\, dy \right| \qquad \text{pois } n > N_2 \\
&\leq \int_{x-\frac{1}{n}}^{x+\frac{1}{n}} \rho_n(x-y)\, |f(y) - f(x)|\, dy \\
&\leq \varepsilon \int_{x-\frac{1}{n}}^{x+\frac{1}{n}} \rho_n(x-y)\, dy \qquad \text{pois } \frac{1}{n} < \delta \\
&\leq \varepsilon \qquad \text{pois } \rho_n \text{ é positivo e } \int_{-\infty}^{\infty} \rho_n(x)\, dx = 1
\end{aligned}
$$

*A última coisa a provar é a desigualdade da norma caso $f$ seja limitada.*

$$
\begin{aligned}
|f_n(x)| &\leq n \int_{-n}^{n} \rho(n(x-y))\, |f(y)|\, dy \qquad \text{pois } \rho \text{ é positivo} \\
&\leq n \|f\|_\infty \int_{-\infty}^{\infty} \rho(n(x-y)) dy = \|f\|_\infty
\end{aligned}
$$

### Testando Convergência com Funções Suaves

**8.36 LEMA** *Seja $\{X_i\}_{i=1}^{\infty}$ uma sequência de variáveis aleatórias então $X_n \xrightarrow{d} X$ se e somente se $\lim_{n\to\infty} \mathbf{E}[f(X_n)] = \mathbf{E}[f(X)]$ para todos as funções $f \in C_c^\infty(\mathbb{R}, \mathbb{R})$.* ◁

**Demonstração.** *Como qualquer $f \in C_c^\infty(\mathbb{R}, \mathbb{R})$ é temos que $X_n \xrightarrow{d} X$ implica $\lim_{n\to\infty} \mathbf{E}[f(X_n)] = \mathbf{E}[f(X)]$.*

*Na outra direção, considere uma função limitada arbitrária $f$ e escolha $\varepsilon > 0$. Então podemos encontrar $f_n \in C_c^\infty(\mathbb{R}, \mathbb{R})$ tal que $f_n$ converge uniformemente em conjuntos compactos e $\|f_n\|_\infty \leq \|f\|_\infty$.*

*A ideia da prova é observar que, para qualquer $n, k \geq 0$, temos*

$$
|\mathbf{E}[f(X_n) - f(X)]| \leq |\mathbf{E}[f(X_n) - f_k(X_n)]| + |\mathbf{E}[f_k(X_n) - f_k(X)]| + |\mathbf{E}[f_k(X) - f(X)]|
$$

*e depois limitar cada termo no lado direito. O segundo termo será fácil de lidar devido à nossa hipótese e à suavidade de $f_k$. O primeiro e terceiros termos exigirão que examinemos a aproximação fornecida pela a convergência uniforme do $f_k$ em todos os conjuntos compactos.*

*A primeira tarefa que temos é escolher o conjunto compacto. Para tanto basta considerar os intervalos fechados centrados na origem. Para qualquer $R \in \mathbb{R}$ com $R > 0$, existe um $\psi_R \in C_c^\infty(\mathbb{R}, \mathbb{R})$*

*com* $\mathbb{1}|x| \leq \frac{R}{2} \leq \psi_R(x) \leq \mathbb{1}|x| \leq R$*, assim sendo*

$$\begin{aligned}
\lim_{n\to\infty} \mathbf{P}\left[|X_n| > R\right] &= 1 - \lim_{n\to\infty} \mathbf{E}\left[\mathbb{1}|X_n| \leq R\right] \\
&\leq 1 - \lim_{n\to\infty} \mathbf{E}\left[\psi_R(X_n)\right] \\
&= 1 - \mathbf{E}\left[\psi_R(X)\right] \\
&\leq 1 - \mathbf{E}\left[\mathbb{1}|X| \leq \frac{R}{2}\right] \\
&= \mathbf{P}\left[|X| > \frac{R}{2}\right]
\end{aligned}$$

*Por outro lado, sabemos que* $\lim_{R\to\infty} \mathbb{1}|X| \leq \frac{R}{2} = 0$ *q.c. e, portanto, pelo Teorema da Convergência Monótona,* $\lim_{R\to\infty} \mathbf{P}\left[|X| > \frac{R}{2}\right] = 0$*. Escolha* $R > 0$ *tal que*

$$\mathbf{P}\left[|X| > R\right] \leq \mathbf{P}\left[|X| > \frac{R}{2}\right] \leq \frac{\varepsilon}{4\|f\|_\infty}$$

*Então podemos escolher* $N_1 > 0$ *tal que* $\mathbf{P}\left[|X_n| > R\right] \leq \frac{\varepsilon}{2\|f\|_\infty}$ *para todo* $n > N_1$*.*

*Tendo escolhido* $R > 0$*, sabemos que* $f_n$ *converge uniformemente para* $f$ *em* $|x| \leq R$ *e, portanto, podemos encontrar um* $K > 0$ *tal que se* $k > K$ *e* $|x| \leq R$ *temos* $|f_k(x) - f(x)| < \varepsilon$*. Assim sendo,*

$$\begin{aligned}
|\mathbf{E}\left[f_k(X) - f(X)\right]| &\leq \mathbf{E}\left[|f_k(X) - f(X)|\,;|X| \leq R\right] + \mathbf{E}\left[|f_k(X) - f(X)|\,;|X| > R\right] \\
&\leq \varepsilon \mathbf{P}\left[|X| \leq R\right] + 2\|X\|_\infty \mathbf{P}\left[|X| > R\right] \\
&\leq \varepsilon + \frac{\varepsilon}{2} < 2\varepsilon
\end{aligned}$$

*e pelo mesmo cálculo, por* $n > N_1$

$$|\mathbf{E}\left[f_k(X_n) - f(X_n)\right]| \leq \varepsilon + 2\|X_n\|_\infty \mathbf{P}\left[|X_n| > R\right] \leq 2\varepsilon$$

*Para terminar a prova, escolha* $k > K$ *e então podemos encontrar* $N_2 > 0$ *tal que para todo* $n > N_2$*, temos* $|\mathbf{E}\left[f_k(X_n) - f_k(X)\right]| < \varepsilon$*. Juntando essas três estimativas, temos* $n > \max(N_1, N_2)$*,*

$$|\mathbf{E}\left[f(X_n) - f(X)\right]| \leq 5\varepsilon$$

## 8.7 Aplicações

### Lei $0 - 1$ de Hewitt-Savage

**8.37 Definição** *Uma aplicação bijetiva* $\pi = (\pi_1, \pi_2, \ldots)$ *do conjunto dos números naturais* $\mathbb{N}$ *em si mesmo é dita* ***permutação finita*** *se* $\pi_n = n$ *para todos os* $n > N_0$*.*

Se $X = (X_1, X_2, \dots) \in (\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}})$ é uma sequência de variáveis aleatórias, $\pi(X)$ denota a sequência $(X_{\pi_1}, X_{\pi_2}, \dots)$. Se $A$ é um evento $\{X \in B\}$ com $B \in \mathcal{B}^{\mathbb{N}}$ então $\pi(A)$ denotará o evento $\{\pi(X) \in B\}$.

Dado uma sequência $(X_n)_{n\in\mathbb{N}}$ de variáveis aleatórias definimos a $\sigma$-álgebra permutável ou $\sigma$-álgebra de eventos simétricos $\mathcal{E}$ como o conjunto de eventos na $\sigma$-álgebra das variáveis $\{X_n\}_{n=1}^{\infty}$ que são invariantes sob permutações finitas dos índices na sequência $\{X_n\}_{n=1}^{\infty}$. Isto é, o evento $A = \{X \in B\}$ é simétrico se $A = \pi(A)$.

8.38 **Teorema (Lei 0 – 1 de Hewitt-Savage)** *Dado uma sequência $(X_n)$ de variáveis aleatórias independentes e identicamente distribuídas. Então a sigma álgebra dos eventos permutáveis $\mathcal{F}$ satisfaz*

$$\mathbf{P}(A) \in \{0,1\}, \forall A \in \mathcal{F}$$

**Demonstração.** *Seja $A = \{X \in B\}$ um evento simétrico, isto é*

$$A = \{X \in B\} = \pi_n(A) \equiv \{\pi_n(X) \in B\}.$$

*Aproximaremos $A$ por eventos cilíndricos. Para isso escolha conjuntos $B_n \in \mathcal{B}^{\mathbb{N}}$ tais que, para $A_n = \{\omega : (X_1, \dots X_n) \in B_n\}$,*

$$\mathbf{P}(A \triangle A_n) \to 0, \ n \to \infty. \tag{8.19}$$

*Como as variáveis aleatórias $\{X_i\}_{n\in\mathbb{N}}$ são independentes e identicamente distribuídas, temos que as probabilidades $\mathbf{P}(X \in B)$ e $\mathbf{P}(\pi_n(X) \in B)$ coincidem. Assim sendo*

$$\mathbf{P}(A \triangle A_n) = \mathbf{P}(X \in B \triangle B_n) = \mathbf{P}(\pi_n(X) \in B \triangle B_n). \tag{8.20}$$

*Consequentemente*

$$\begin{aligned}\mathbf{P}(\pi_n(X) \in B \triangle B_n) &= \mathbf{P}\{(\pi_n(X) \in B) \triangle (\pi_n(X) \in B_n)\} \\ &= \mathbf{P}\{(X \in B) \triangle (\pi_n(X) \in B_n)\} = \mathbf{P}\{A \triangle \pi_n(A_n)\}.\end{aligned} \tag{8.21}$$

*Por 8.20 e 8.21 temos que*

$$\mathbf{P}(A \triangle A_n) = \mathbf{P}(A \triangle \pi_n(A_n)). \tag{8.22}$$

*Logo por 8.19 temos que*

$$\mathbf{P}(A \triangle (A_n \cap \pi_n(A_n))) \to 0, \ n \to \infty. \tag{8.23}$$

*Assim, por 8.19, 8.22 e 8.23, obtemos*

$$\mathbf{P}(A_n) \to \mathbf{P}(A) \ \ e \ \ \mathbf{P}(\pi_n(A)) \to \mathbf{P}(A)$$

$$\mathbf{P}(A_n \cap \pi_n(A_n)) \to \mathbf{P}(A).$$

*Além disso, como* $\{X_i\}$ *são independentes, escolhendo* $\pi_n$ *como a permutação que leva* $(X_1, \dots X_n)$ *para* $(X_{n+1}, \dots X_{2n})$

*temos*

$$\begin{aligned}\mathbf{P}(A_n \cap \pi_n(A_n)) &= \mathbf{P}\{(X_1, \dots X_n) \in B_n,\ (X_{n+1}, \dots X_{2n}) \in B_n\} \\ &= \mathbf{P}\{\ (X_1, \dots X_n) \in B_n\}.\ \mathbf{P}\{(X_{n+1}, \dots X_{2n}) \in B_n\} \\ &= \mathbf{P}(A_n)\mathbf{P}(\pi_n(A_n)\end{aligned}$$

*Logo* $\mathbf{P}(A) = \mathbf{P}^2(A)$ *e, portanto,* $\mathbf{P}(A) = 0$ *ou 1.* □

## Passeios Aleatórios

Dado uma sequência $X_i$ de vetores aleatórios i.i.d. em $\mathbb{R}^n$ então

$$S_n = \sum_{i=1}^{n} X_i$$

é um passeio aleatório em $\mathbb{R}^n$

8.39 **Teorema** *Se* $S_n$ *é um passeio aleatório em* $\mathbb{R}$ *então uma das seguintes alternativas ocorre*

- $S_n = 0$ *para todo* $n$
- $S_n \to \infty$
- $S_n \to -\infty$
- $-\infty = \liminf S_n < \limsup S_n = \infty$

**Demonstração.** *Começamos observando que* $\{\omega : \limsup S_n(\omega) = c\}$ *é um evento permutável para todo* $c \in \mathbb{R}$. *Assim pela lei 0-1 de Hewitt-Savage,*

$$\mathbf{P}(\{\omega : \limsup S_n(\omega) = c\}) \in \{0,1\}$$

*e logo* $\limsup Sn$ *é uma constante* $c \in [-\infty, \infty]$ *q.c.*

*Agora considere* $S'_n = S_n - X_1$, *como* $S'_n$ *tem a mesma distribuição que* $S_n$ *tomando limsup em ambos os lados temos*

$$c = c - X_1.$$

*Assim, ou* $X_1 = 0$ *ou* $c \in \{-\infty, \infty\}$. *De modo análogo,* $\liminf S_n \in \{-\infty, \infty\}$. □