

# Ergodicidade, Mixing e Shifts em Espaços Simbólicos

L. Cioletti

Abril de 2026

## Resumo

Este texto faz parte de uma série de notas de aula de um curso de Teoria Ergódica ministrado pelo autor no Departamento de Matemática da Universidade de Brasília. O objetivo central é introduzir as propriedades de mistura fraca e forte (*weak-mixing* e *strong-mixing*) e aplicá-las ao estudo dos shifts no espaço simbólico. Iniciamos com uma consequência do Teorema Ergódico de Birkhoff que relaciona medidas ergódicas distintas, mostrando que ou coincidem ou são mutuamente singulares. Em seguida lembramos uma caracterização de ergodicidade por meio da convergência, no sentido de Cesàro, de  $\mu(T^{-n}A \cap B)$  para produto  $\mu(A)\mu(B)$ . Este é usado como motivação para introduzir os conceitos de dinâmicas misturadoras forte e fraca.

Em seguida, apresentamos uma versão do Teorema de Existência de Kolmogorov adequada à construção de medidas em espaços produto da forma  $Y^{\mathbb{Z}}$ , onde  $Y$  é um conjunto finito. Usamos este teorema para apresentar uma construção rigorosa dos shifts de Bernoulli e de Markov, duas das classes mais importantes de exemplos em Teoria Ergódica. No caso Markoviano, evidenciamos como a estocasticidade da matriz  $P$  corresponde à consistência das marginais à direita, enquanto a estacionariedade  $pP = p$  equivale à consistência das marginais à esquerda.

Encerramos estabelecendo, para shifts de Markov, a equivalência entre cinco formulações da ergodicidade: (i) ergodicidade no sentido dinâmico; (ii) igualdade das linhas da matriz limite de Cesàro  $Q$ ; (iii) positividade estrita das entradas de  $Q$ ; (iv) irreduzibilidade de  $P$ ; e (v) unidimensionalidade do autoespaço à esquerda de  $P$  associado ao autovalor 1. Este resultado conecta a teoria ergódica abstrata à teoria clássica das cadeias de Markov finitas estacionárias.

## 1. Introdução e Algumas Notações Básicas

Ao longo deste texto, utilizamos as seguintes notações e convenções. A tripla ordenada  $(X, \mathcal{B}, \mu)$  denota um espaço de **probabilidade**. Se  $T : X \rightarrow X$  denota uma função definida em  $X$  e tomando valores em  $X$  e  $n \in \mathbb{Z}_+ \equiv \mathbb{N} \cup \{0\}$  um inteiro não-negativo, denotamos por  $T^n$  a  $n$ -ésima iterada de  $T$  (com a convenção  $T^0 \equiv \text{Id}$ ). Como de costume, usamos a notação  $T^{-n}(E) \equiv \{x \in X : T^n(x) \in E\}$  para a imagem inversa, por  $T^n$ , do conjunto  $E \subseteq X$ .

**Definição 1** (Transformações que Preservam Medida). Seja  $(X, \mathcal{B}, \mu)$  um espaço de probabilidade. Dizemos que uma transformação  $T : X \rightarrow X$  é mensurável se, para todo  $B \in \mathcal{B}$ , temos que  $T^{-1}(B) \in \mathcal{B}$ . Se  $T : X \rightarrow X$  é uma transformação mensurável e satisfaz

$$\mu(T^{-1}(B)) = \mu(B), \quad \forall B \in \mathcal{B},$$

então dizemos que  $T$  é uma transformação que preserva a medida  $\mu$ .

**Definição 2 (Transformação Ergódica)**. Sejam  $(X, \mathcal{B}, \mu)$  um espaço de probabilidade e  $T : X \rightarrow X$  uma transformação mensurável. Dizemos que  $T$  é ergódica se para todo  $E \in \mathcal{B}$  satisfazendo  $T^{-1}(E) = E$ , temos  $\mu(E) = 0$  ou  $\mu(E) = 1$ .

A seguir enunciamos o Teorema Ergódico de Birkhoff para transformações que preservam medida definidas sobre um espaço de probabilidade. O enunciado e a prova deste teorema no caso em que  $(X, \mathcal{B}, \mu)$  é um espaço de medida  $\sigma$ -finito podem ser vistos na referência [2].

**Teorema 3** (Teorema Ergódico de Birkhoff). Sejam  $(X, \mathcal{B}, \mu)$  um espaço de probabilidade,  $T : (X, \mathcal{B}, \mu) \rightarrow (X, \mathcal{B}, \mu)$  uma transformação que preserva medida e  $f \in L^1(X, \mathcal{B}, \mu)$ . Então existe uma função mensurável  $f^* \in L^1(X, \mathcal{B}, \mu)$  tal que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) = f^*(x) \quad \mu - \text{qtp.} \quad (1)$$

A função  $f^* : X \rightarrow \mathbb{R} \cup \{\pm\infty\}$  é  $\mathcal{B}$ -mensurável e satisfaz a identidade  $f^* \circ T = f^*$   $\mu$ -qtp.

Além do mais, para todo  $W \in \mathcal{B}$ , satisfazendo

$$T^{-1}(W) = W \quad \text{temos que} \quad \int_W f^* d\mu = \int_W f d\mu.$$

**Teorema 4.** Seja  $T : X \rightarrow X$  uma aplicação mensurável definida sobre o espaço mensurável  $(X, \mathcal{B})$ . Sejam  $\mu, \nu$  medidas de probabilidade definidas sobre  $\mathcal{B}$ . Suponha que ambas  $\mu$  e  $\nu$  sejam medidas  $T$ -invariantes. Então:

- (1) se  $T$  é uma transformação ergódica em  $(X, \mathcal{B}, \mu)$ , e  $\nu \ll \mu$  é absolutamente contínua em relação a  $\mu$ , então  $\mu = \nu$ ;
- (2) se  $T$  é uma transformação ergódica em ambos espaços  $(X, \mathcal{B}, \mu)$  e  $(X, \mathcal{B}, \nu)$ , então ou  $\mu = \nu$  ou  $\mu \perp \nu$ .

**Prova.** Pelo Teorema Ergódico de Birkhoff temos para qualquer conjunto  $A \in \mathcal{B}$ , que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \chi_A(T^k x) = \mu(A), \quad \mu - \text{qtp.}$$

Como  $\nu$  é absolutamente contínua em relação a  $\mu$ , então a igualdade acima é válida para  $\nu$ -quase todo ponto. Pelo Teorema da Convergência Dominada de Lebesgue, temos

$$\lim_{n \rightarrow \infty} \int_X \frac{1}{n} \sum_{k=0}^{n-1} \chi_A(T^k x) d\nu = \int_X \mu(A) d\nu = \mu(A).$$

Por outro lado, para todo  $n \in \mathbb{N}$  temos

$$\int_X \frac{1}{n} \sum_{k=0}^{n-1} \chi_A(T^k x) d\nu = \nu(A).$$

Portanto  $\mu(A) = \nu(A)$ . Como  $A$  é arbitrário em  $\mathcal{B}$ , a afirmação (1) está provada.

(2) Suponha que  $\mu \neq \nu$ . Então existe algum subconjunto não-vazio  $A \in \mathcal{B}$ , tal que  $\mu(A) \neq \nu(A)$ . Seja

$$A(\mu) = \left\{ x \in X : \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \chi_A(T^k x) = \mu(A) \right\}$$

e

$$A(\nu) = \left\{ x \in X : \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \chi_A(T^k x) = \nu(A) \right\}.$$

Observe que  $A(\mu) = (\chi_A^*)^{-1}(\mu(A))$  e  $A(\nu) = (\chi_A^*)^{-1}(\nu(A))$  e portanto  $A(\mu) \cap A(\nu) = \emptyset$ . Além do mais, segue do Teorema Ergódico de Birkhoff que  $\mu(A(\mu)) = 1 = \nu(A(\nu))$ . Portanto, as medidas  $\mu$  e  $\nu$  são mutuamente singulares e o teorema está provado. ■

Outra consequência importante do Teorema Ergódico de Birkhoff que vimos em [2] foi o seguinte corolário.

**Corolário 5.** Sejam  $(X, \mathcal{B}, \mu)$  um espaço de probabilidade e  $T: X \rightarrow X$  uma transformação que preserva medida. Então  $T$  é ergódica se, e somente se, para todo  $A, B \in \mathcal{B}$ ,

$$\frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i} A \cap B) \xrightarrow{n \rightarrow \infty} \mu(A) \mu(B).$$

**Prova.** Suponha que  $T$  seja uma transformação ergódica. Já que  $(X, \mathcal{B}, \mu)$  é um espaço de probabilidade, então podemos aplicar o teorema ergódico à  $f = \chi_A$  para afirmar que

$$\frac{1}{n} \sum_{i=0}^{n-1} \chi_A(T^i(x)) \xrightarrow{n \rightarrow \infty} \mu(A) \quad \mu - \text{qtp.}$$

Multiplicando por  $\chi_B$ , em ambos lados da igualdade acima ficamos com

$$\frac{1}{n} \sum_{i=0}^{n-1} \chi_A(T^i(x)) \chi_B(x) \xrightarrow{n \rightarrow \infty} \mu(A) \chi_B(x) \quad \mu - \text{qtp.}$$

Aplicando o Teorema da Convergência Dominada, temos

$$\frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i} A \cap B) \xrightarrow{n \rightarrow \infty} \mu(A) \mu(B).$$

Reciprocamente, suponha que, fixados  $A, B \in \mathcal{B}$ , a sequência  $\mu(T^{-n} A \cap B)$  converge no sentido de Cesàro para  $\mu(A)\mu(B)$ . Seja  $E \in \mathcal{B}$  tal que  $T^{-1}E = E$ . Tomando  $A = B = E$ , então segue da hipótese que

$$\frac{1}{n} \sum_{i=0}^{n-1} \mu(E) \xrightarrow{n \rightarrow \infty} \mu(E)^2.$$

Logo  $\mu(E) = \mu(E)^2$  e, portanto,  $\mu(E) = 0$  ou  $1$ . ■

## 2. Transformações Misturadoras

O **Corolário 5** afirma que se a sequência  $\{\mu(T^{-n}A \cap B)\}_{n \in \mathbb{Z}_+}$  converge para  $\mu(A)\mu(B)$ , no sentido de Cesàro, então  $T$  é ergódica e vice-versa. As chamadas propriedades de “mixing” fraca e forte são obtidas alterando este modo de convergência da sequência. De forma mais precisa.

**Definição 6.** Seja  $T : X \rightarrow X$  uma aplicação que preserva medida sobre  $(X, \mathcal{B}, \mu)$ .

- (i) vamos dizer que  $T$  é uma transformação fracamente misturadora ou *weak-mixing* se  $\forall A, B \in \mathcal{B}$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} |\mu(T^{-i}A \cap B) - \mu(A)\mu(B)| = 0.$$

- (ii) vamos dizer que  $T$  é fortemente misturadora ou *strong-mixing* se  $\forall A, B \in \mathcal{B}$

$$\lim_{n \rightarrow \infty} \mu(T^{-n}A \cap B) = \mu(A)\mu(B).$$

### Observações

1. Toda transformação  $T$  fortemente misturadora é fracamente misturadora. Além do mais, toda transformação fracamente misturadora é ergódica. De fato, basta observar que para toda sequência de números reais  $\{a_n\}_{n \in \mathbb{Z}_+}$  satisfazendo:  $a_n \rightarrow 0$ , quando  $n \rightarrow \infty$ , temos

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} |a_i| = 0$$

e conseqüentemente pela desigualdade triangular e continuidade do valor absoluto

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} a_i = 0.$$

Fortemente Misturadora  $\implies$  Fracamente Misturadora  $\implies$  Ergódica.

2. Um exemplo de uma transformação ergódica que não é fracamente misturadora é dado por uma rotação  $T(z) = az$ , do círculo unitário que será denotado por  $\mathbb{K}$ . Para maiores detalhes veja [8]. Mas intuitivamente segue da seguinte observação. Se  $A$  e  $B$  forem intervalos pequenos em  $\mathbb{K}$ , então  $T^{-i}A$  será disjunto de  $B$  para pelo menos metade dos valores de  $i$ , de modo que

$$\frac{1}{n} \sum_{i=0}^{n-1} |\mu(T^{-i}A \cap B) - \mu(A)\mu(B)| \geq \frac{1}{2} \mu(A)\mu(B),$$

para  $n$  suficientemente grande. Este exemplo sugere que uma transformação fracamente misturadora de alguma forma acabará possuindo alguma propriedade de expansividade.

Ergódica  $\not\implies$  Fracamente Misturadora.

3. Existem exemplos de transformações  $T$  fracamente misturadoras que não são fortemente misturadoras. Kakutani [5] apresentou um exemplo construído por métodos combinatórios e Maruyama [6] construiu um exemplo utilizando processos Gaussianos. Chacon, Katok e Stepin também exibiram alguns exemplos. Além do mais, se  $(X, \mathcal{B}, \mu)$  é um espaço de probabilidade, seja  $\tau(X)$  a coleção de todas as transformações invertíveis que preservam a medida de  $(X, \mathcal{B}, \mu)$ . Se munimos  $\tau(X)$  de uma topologia “fraca” (veja Halmos [4]), a classe de transformações fracamente misturadoras é de segunda categoria, enquanto a classe de transformações fortemente misturadoras é de primeira categoria. Portanto, do ponto de vista desta topologia, a maioria das transformações são fracamente misturadoras, mas não fortemente misturadoras.
4. Descrições intuitivas de ergodicidade e de misturas fortes podem ser dadas da seguinte forma. Dizer que  $T$  é fortemente misturadora significa que, para qualquer conjunto  $A$ , a sequência de conjuntos  $T^{-n}A$  torna-se, assintoticamente, independente de qualquer outro conjunto  $B$ . A ergodicidade significa que  $T^{-n}A$  torna-se independente de  $B$  na média, para cada par de conjuntos  $A, B \in \mathcal{B}$ .

O teorema a seguir fornece uma maneira de verificar as propriedades de mistura para alguns exemplos, reduzindo os cálculos à uma classe de conjuntos para os quais o cálculo da condição de mistura é mais viável. Por exemplo, ele implica que precisamos considerar apenas retângulos mensuráveis ao lidar com as propriedades de mistura de deslocamentos (shifts) no espaço produto.

Antes de enunciarmos o teorema vamos lembrar algumas definições importantes. Uma coleção  $\mathcal{S}$  de subconjuntos de  $X$  é chamada de **semi-álgebra** se as seguintes três condições forem satisfeitas:

- (i)  $\emptyset \in \mathcal{S}$ ;
- (ii) se  $A, B \in \mathcal{S}$ , então  $A \cap B \in \mathcal{S}$ ;
- (iii) se  $A \in \mathcal{S}$ , então  $X \setminus A = \bigcup_{i=1}^n E_i$ , em que cada  $E_i \in \mathcal{S}$  e  $E_1, \dots, E_n$  são subconjuntos dois a dois disjuntos de  $X$ .

Por exemplo, a coleção de todos os subintervalos de  $[0, 1]$  é uma semi-álgebra. Além disso, a coleção de todos os subintervalos de  $[0, 1]$  das formas  $[0, b]$  e  $(a, b]$ , com  $0 \leq a < b \leq 1$ , mais o conjunto vazio, forma uma semi-álgebra.

Uma coleção  $\mathcal{A}$  de subconjuntos de  $X$  é chamada de **álgebra** se as seguintes três condições forem satisfeitas:

- (i)  $\emptyset \in \mathcal{A}$ ;
- (ii) se  $A, B \in \mathcal{A}$ , então  $A \cap B \in \mathcal{A}$ ;
- (iii) se  $A \in \mathcal{A}$ , então  $X \setminus A \in \mathcal{A}$ .

**Teorema 7.** Seja  $(X, \mathcal{B}, \mu)$  um espaço de probabilidade e seja  $\mathcal{S}$  uma semi-álgebra que gera  $\mathcal{B}$ . Seja  $T : X \rightarrow X$  uma transformação que preserva a medida. Então:

- (i)  $T$  é ergódica se, e somente se,  $\forall A, B \in \mathcal{S}$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}A \cap B) = \mu(A)\mu(B),$$

(ii)  $T$  é fracamente misturadora se, e somente se,  $\forall A, B \in \mathcal{S}$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} |\mu(T^{-i}A \cap B) - \mu(A)\mu(B)| = 0, \quad \text{e}$$

(iii)  $T$  é fortemente misturadora se, e somente se,  $\forall A, B \in \mathcal{S}$

$$\lim_{n \rightarrow \infty} \mu(T^{-n}A \cap B) = \mu(A)\mu(B).$$

**Prova.** Já que para cada membro da álgebra,  $\mathcal{A}(\mathcal{S})$ , gerada por  $\mathcal{S}$ , pode ser escrito como uma união finita de membros de  $\mathcal{S}$ , segue das propriedades elementares de limites que se as igualdades acima valem para todos os membros de  $\mathcal{S}$ , então elas valem para todos os elementos de  $\mathcal{A}(\mathcal{S})$ .

Dado  $\varepsilon > 0$ , sejam  $A, B \in \mathcal{B}$ . Pela construção de medidas via o Teorema de Extensão de Carathéodory sabemos que existem  $A_0, B_0 \in \mathcal{A}(\mathcal{S})$  tais que  $\mu(A \Delta A_0) < \varepsilon$  e  $\mu(B \Delta B_0) < \varepsilon$ .

Para todo  $i \in \mathbb{N}$ ,

$$(T^{-i}A \cap B) \Delta (T^{-i}A_0 \cap B_0) \subseteq (T^{-i}A \Delta T^{-i}A_0) \cup (B \Delta B_0).$$

Já que  $\mu$  é uma medida  $T$ -invariante temos da relação acima que

$$\mu((T^{-i}A \cap B) \Delta (T^{-i}A_0 \cap B_0)) < 2\varepsilon.$$

Portanto, temos para todo  $i \in \mathbb{N}$  a seguinte desigualdade  $|\mu(T^{-i}A \cap B) - \mu(T^{-i}A_0 \cap B_0)| < 2\varepsilon$ .  
Portanto

$$\begin{aligned} |\mu(T^{-i}A \cap B) - \mu(A)\mu(B)| &\leq |\mu(T^{-i}A \cap B) - \mu(T^{-i}A_0 \cap B_0)| \\ &\quad + |\mu(T^{-i}A_0 \cap B_0) - \mu(A_0)\mu(B_0)| \\ &\quad + |\mu(A_0)\mu(B_0) - \mu(A)\mu(B)| \\ &\quad + |\mu(A)\mu(B) - \mu(A)\mu(B)| \\ &< 4\varepsilon + |\mu(T^{-i}A_0 \cap B_0) - \mu(A_0)\mu(B_0)|. \end{aligned}$$

Da desigualdade acima, segue que se assumimos (ii) e (iii) para  $A_0, B_0$ , então temos (ii) e (iii) para  $A$  e  $B$ .

Para obter (i), vamos provar que vale a seguinte desigualdade:

$$\left| \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}A \cap B) - \mu(A)\mu(B) \right| < 4\varepsilon + \left| \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}A_0 \cap B_0) - \mu(A_0)\mu(B_0) \right|. \quad (2)$$

Para facilitar a notação da prova, para cada  $n \in \mathbb{N}$  seja  $S_n(A, B) \equiv \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}A \cap B)$ . Desta forma nosso objetivo é limitar  $|S_n(A, B) - \mu(A)\mu(B)|$ . Usando desigualdade triangular e inserindo os seguintes termos  $S_n(A_0, B_0)$  e  $\mu(A_0)\mu(B_0)$  na diferença acima ficamos com

$$\begin{aligned} |S_n(A, B) - \mu(A)\mu(B)| &\leq |S_n(A, B) - S_n(A_0, B_0)| + |S_n(A_0, B_0) - \mu(A_0)\mu(B_0)| \\ &\quad + |\mu(A_0)\mu(B_0) - \mu(A)\mu(B)|. \end{aligned}$$

Desta maneira para obtermos a desigualdade desejada basta estimar a primeira e a terceira parcela do lado direito da desigualdade acima.

Para a primeira parcela temos:

$$\begin{aligned} |S_n(A, B) - S_n(A_0, B_0)| &= \left| \frac{1}{n} \sum_{i=0}^{n-1} \left[ \mu(T^{-i}A \cap B) - \mu(T^{-i}A_0 \cap B_0) \right] \right| \\ &\leq \frac{1}{n} \sum_{i=0}^{n-1} \left| \mu(T^{-i}A \cap B) - \mu(T^{-i}A_0 \cap B_0) \right|. \end{aligned} \quad (3)$$

Lembrando que para quaisquer  $E, F \in \mathcal{B}$  temos  $|\mu(E) - \mu(F)| \leq \mu(E \Delta F)$  e que vale a seguinte inclusão

$$(E_1 \cap F_1) \Delta (E_2 \cap F_2) \subset (E_1 \Delta E_2) \cup (F_1 \Delta F_2), \quad \forall E_1, E_2, F_1, F_2 \in \mathcal{B}$$

temos de (3)

$$\begin{aligned} |S_n(A, B) - S_n(A_0, B_0)| &\leq \frac{1}{n} \sum_{i=0}^{n-1} \mu((T^{-i}A \cap B) \Delta (T^{-i}A_0 \cap B_0)) \\ &\leq \frac{1}{n} \sum_{i=0}^{n-1} \left[ \mu(T^{-i}A \Delta T^{-i}A_0) + \mu(B \Delta B_0) \right]. \end{aligned}$$

Como  $T$  preserva a medida,  $\mu(T^{-i}A \Delta T^{-i}A_0) = \mu(A \Delta A_0) < \varepsilon$ . Além disso,  $\mu(B \Delta B_0) < \varepsilon$ . Logo, cada parcela da soma é menor que  $2\varepsilon$ , e portanto  $|S_n(A, B) - S_n(A_0, B_0)| < 2\varepsilon$ .

Para limitar o terceiro termo, somamos e subtraímos  $\mu(A)\mu(B_0)$ . Em seguida, usando a desigualdade triangular e novamente a desigualdade  $|\mu(E) - \mu(F)| \leq \mu(E \Delta F)$  obtemos

$$\begin{aligned} |\mu(A_0)\mu(B_0) - \mu(A)\mu(B)| &= |\mu(A_0)\mu(B_0) - \mu(A)\mu(B_0) + \mu(A)\mu(B_0) - \mu(A)\mu(B)| \\ &\leq |\mu(B_0)| \cdot |\mu(A_0) - \mu(A)| + |\mu(A)| \cdot |\mu(B_0) - \mu(B)| \\ &< |\mu(B_0)| \cdot \varepsilon + |\mu(A)| \cdot \varepsilon. \end{aligned}$$

Com esta estimativa encerramos a prova da desigualdade (2). Usando esta desigualdade e tomando o limite, quando  $n \rightarrow \infty$  concluímos a prova do item (i) do teorema. ■

### 3. O Teorema da Existência de Kolmogorov

Nesta seção vamos enunciar uma versão mais simples do famoso Teorema da Existência de Kolmogorov, adequada aos propósitos destas notas. Seus enunciados e prova, no caso geral podem ser encontrados em diversos livros de Processos Estocásticos ou de Teoria de Probabilidade. Veja, por exemplo, as referências [1, 3, 7].

O objetivo desta seção é dar suporte teórico a construção de duas classes de exemplos extremamente importantes em Teoria Ergódica que são os shifts de Bernoulli e Markov.

**Definição 8** (Família Consistente de Medidas). Suponha que para cada  $k \in \mathbb{N}$  fixado e para cada elemento  $(t_1, \dots, t_k) \in \mathbb{Z}^k$ , seja possível associar uma medida de probabilidade  $\mu_{t_1, \dots, t_k}$  definida sobre o espaço produto  $Y^k$ , munido de sua  $\sigma$ -álgebra produto. Denote a família de todas estas medidas por  $\{\mu_{t_1, \dots, t_k}\}$ . Vamos dizer que a família  $\{\mu_{t_1, \dots, t_k}\}$  é uma família consistente de medidas se as seguintes condições, chamadas de condições de consistência são satisfeitas:

$$\mu_{t_1, \dots, t_k}(H_1 \times \cdots \times H_k) = \mu_{t_{\pi(1)}, \dots, t_{\pi(k)}}(H_{\pi(1)} \times \cdots \times H_{\pi(k)}) \quad (\text{C1})$$

$$\mu_{t_1, \dots, t_{k-1}}(H_1 \times \cdots \times H_{k-1}) = \mu_{t_1, \dots, t_{k-1}, t_k}(H_1 \times \cdots \times H_{k-1} \times Y), \quad (\text{C2})$$

onde  $H_j \subseteq Y$ ,  $(t_{\pi(1)}, \dots, t_{\pi(k)})$  é uma permutação de  $(t_1, \dots, t_k)$ .

**Teorema 9** (Teorema de Existência de Kolmogorov). Seja  $\{\mu_{t_1, \dots, t_k}\}$  uma família consistente de medidas. Então existe uma única medida de probabilidade  $\mu$  definida sobre o espaço produto  $(Y^{\mathbb{Z}}, \mathcal{B})$  tal que para quaisquer  $k \geq 0$ ,  $m \in \mathbb{Z}$  e  $H_j \subseteq Y$  temos:

$$\mu(\{x \in X : x_m \in H_0, x_{m+1} \in H_1, \dots, x_k \in H_k\}) = \mu_{m, m+1, \dots, m+k}(H_0 \times H_1 \times \dots \times H_k).$$

### 3.1. Construindo Famílias de Medidas Consistentes

Para cada  $k \in \mathbb{N}$ , denotamos por  $\mathcal{M}(Y^k, \mathcal{B}_k)$  o conjunto de todas as medidas de probabilidade definidas sobre o espaço produto  $(Y^k, \mathcal{B}_k)$ , onde  $\mathcal{B}_k$  é a  $\sigma$ -álgebra produto  $\mathcal{P}(Y) \otimes \dots \otimes \mathcal{P}(Y)$ , onde  $\mathcal{P}(Y)$  denota a  $\sigma$ -álgebra das partes de  $Y$ .

O objetivo desta seção é mostrar que é sempre possível construir uma família consistente de medidas de probabilidade  $\{\mu_{t_1, \dots, t_k}\}$  à partir de uma família de medidas de probabilidade

$$\mathcal{M} \equiv \{\mu_{t_1, \dots, t_k} \in \mathcal{M}(Y^k, \mathcal{B}_k) : k \in \mathbb{N} \text{ e } t_1 < t_2 < \dots < t_k\}, \quad (4)$$

desde que esta família de medidas indexadas sobre listas  $(t_1, \dots, t_k)$  ordenadas, satisfaça as condições de consistência (5), (6) e (7) descritas abaixo:

$$\begin{aligned} & \mu_{t_1, \dots, t_{i-1}, t_{i+1}, \dots, t_k}(H_1 \times \cdots \times H_{i-1} \times H_{i+1} \times \cdots \times H_k) \\ & \quad \parallel \\ & \mu_{t_1, \dots, t_k}(H_1 \times \cdots \times H_{i-1} \times Y \times H_{i+1} \times \cdots \times H_k) \end{aligned} \quad (5)$$

para quaisquer  $H_j \subseteq Y$ , com  $j = 1, \dots, k$  e  $1 < i < k$ . Para  $i = 1$  e  $i = k$  exigimos que sejam satisfeitas as seguintes igualdades, respectivamente:

$$\mu_{t_2, \dots, t_k}(H_2 \times \cdots \times H_k) = \mu_{t_1, \dots, t_k}(Y \times H_2 \times \cdots \times H_k); \quad (6)$$

$$\mu_{t_1, \dots, t_{k-1}}(H_1 \times \cdots \times H_{k-1}) = \mu_{t_1, \dots, t_k}(H_1 \times \cdots \times H_{k-1} \times Y), \quad (7)$$

para quaisquer  $H_j \subseteq Y$ , com  $j = 1, \dots, k$ .

Observação. Embora possam parecer distintas, as condições (5), (6) (7) são simplesmente categorizações de todos os casos possíveis da condição (C2). Esta divisão é importante porque (C2) é formulada para listas arbitrárias de inteiros distintos, enquanto (5), (6) (7) lidam com todos os casos de consistência quando a lista de índices é composta por inteiros distintos que estão ordenados.

Continuando, a construção que nos referimos acima é feita da seguinte maneira. Suponha que  $\mathcal{M}$ , seja uma família de medidas de probabilidade indexada sobre  $k$ -úplas ordenadas de inteiros, como em (4), e que satisfaça as condições de consistência (5), (6) e (7).

Dada uma  $k$ -úpla arbitrária de inteiros dois-a-dois distintos  $(t_1, \dots, t_k) \in \mathbb{Z}^k$ , considere a única permutação  $\pi : \{1, \dots, k\} \rightarrow \{1, \dots, k\}$ , que age ordenando as coordenadas desta  $k$ -úpla, ou seja,  $t_{\pi(1)} < \dots < t_{\pi(k)}$ . Agora, para esta  $k$ -úpla arbitrária de inteiros defina  $\mu_{t_1, \dots, t_k} \in \mathcal{M}(Y^k, \mathcal{B}_k)$  por

$$\mu_{t_1, \dots, t_k} \equiv \mu_{t_{\pi(1)}, \dots, t_{\pi(k)}} \circ \varphi_{\pi}^{-1}, \quad (8)$$

onde  $\varphi_{\pi} : Y^k \rightarrow Y^k$  é a aplicação dada por  $\varphi_{\pi}(x_1, \dots, x_k) = (x_{\pi^{-1}(1)}, \dots, x_{\pi^{-1}(k)})$ .

**Teorema 10.** Seja  $\mathcal{M} \equiv \{\mu_{t_1, \dots, t_k} \in \mathcal{M}(Y^k, \mathcal{B}_k) : k \in \mathbb{N} \text{ e } t_1 < t_2 < \dots < t_k\}$ , uma família de medidas de probabilidade indexada sobre todas as  $k$ -úplas ordenadas de inteiros, com  $k$  variando sobre todos os elementos de  $\mathbb{N}$ . Suponha que esta família satisfaça as condições de consistência (5), (6) e (7). Então existe uma família consistente de medidas de probabilidade  $\{\mu_{t_1, \dots, t_k}\}$  (no sentido da Definição 8) contendo a família  $\mathcal{M}$ .

**Prova.** Primeiro, construção da família  $\{\mu_{t_1, \dots, t_k}\}$ , que aparece no enunciado do teorema, é feita por meio da equação (8). Construída esta família agora precisamos mostrar que ela satisfaz as condições de consistência (C1) e (C2).

É claro que a condição (C1) é satisfeita por construção, já que a família é definida satisfazendo (8). A condição (C2) também é automaticamente satisfeita, por construção, no caso em que  $(t_1, \dots, t_k)$  é uma  $k$ -úpla de inteiros ordenada, porque assumimos na hipótese do teorema que família de medidas  $\mathcal{M}$  satisfaz a condição de consistência (7). Portanto, resta verificar a validade da condição (C2) quando  $(t_1, \dots, t_k)$  é uma  $k$ -úpla arbitrária.

Sejam  $(t_1, \dots, t_k)$  é uma  $k$ -úpla arbitrária e  $\lambda$  a única permutação do conjunto de índices  $\{1, \dots, k-1\}$  que induz uma ordenação da  $(k-1)$ -úpla  $(t_1, \dots, t_{k-1})$ , ou seja,  $t_{\lambda(1)} < \dots < t_{\lambda(k-1)}$ .

Vamos considerar os seguintes três casos:

**Caso 1.** O índice  $t_k$  é o maior índice da  $k$ -úpla  $(t_1, \dots, t_k)$ . Neste caso a permutação  $\pi$  que induz a ordenação nesta  $k$ -úpla é dada por  $\pi(j) = \lambda(j)$ , para  $j = 1, \dots, k-1$  e  $\pi(k) = k$ . Assim temos diretamente das condições (8) e (7) que

$$\begin{aligned} \mu_{t_1, \dots, t_{k-1}}(H_1 \times \dots \times H_{k-1}) &= \mu_{t_{\lambda(1)}, \dots, t_{\lambda(k-1)}}(H_{\lambda(1)} \times \dots \times H_{\lambda(k-1)}) \\ &= \mu_{t_{\pi(1)}, \dots, t_{\pi(k-1)}, t_{\pi(k)}}(H_{\pi(1)} \times \dots \times H_{\pi(k-1)} \times Y) \\ &= \mu_{t_1, \dots, t_{k-1}, t_k}(H_1 \times \dots \times H_{k-1} \times Y) \end{aligned}$$

e assim a condição (C2) é satisfeita neste caso.

**Caso 2.** O índice  $t_k$  é o menor índice da  $k$ -úpla  $(t_1, \dots, t_k)$ . Neste caso a ordenação desta  $k$ -úpla é induzida pela permutação  $\pi$  dada por  $\pi(1) = k$  e  $\pi(j) = \lambda(j-1)$ , para todo  $j = 2, \dots, k$ . Analogamente, segue das condições (8) e (6) que

$$\begin{aligned} \mu_{t_1, \dots, t_{k-1}}(H_1 \times \dots \times H_{k-1}) &= \mu_{t_{\lambda(1)}, \dots, t_{\lambda(k-1)}}(H_{\lambda(1)} \times \dots \times H_{\lambda(k-1)}) \\ &= \mu_{t_{\pi(1)}, \dots, t_{\pi(k-1)}, t_{\pi(k)}}(Y \times H_{\pi(2)} \times \dots \times H_{\pi(k)}) \\ &= \mu_{t_1, \dots, t_{k-1}, t_k}(H_1 \times \dots \times H_{k-1} \times Y) \end{aligned}$$

o que novamente prova a validade da condição (C2).

**Caso 3.** O índice  $t_k$  é  $i$ -ésimo índice da  $k$ -úpla  $(t_1, \dots, t_k)$ , com  $1 < i < k$ . Neste caso a permutação de  $\{1, \dots, k\}$  que induz a ordenação desta  $k$ -úpla é dada em função de  $\lambda$  da seguinte maneira:

$$\pi(j) = \begin{cases} \lambda(j), & \text{se } j < i; \\ k, & \text{se } j = i; \\ \lambda(j-1), & \text{se } i < j \leq k. \end{cases}$$

Agora, usando as condições (8) e (5) temos

$$\begin{aligned} & \mu_{t_1, \dots, t_{k-1}}(H_1 \times \dots \times H_{k-1}) \\ & \quad \parallel \\ & \mu_{t_{\lambda(1)}, \dots, t_{\lambda(k-1)}}(H_{\lambda(1)} \times \dots \times H_{\lambda(k-1)}) \\ & \quad \parallel \\ & \mu_{t_{\lambda(1)}, \dots, t_{\lambda(i-1)}, t_k, t_{\lambda(i)}, t_{\lambda(i+1)}, \dots, t_{\lambda(k-1)}}(H_{\lambda(1)} \times \dots \times H_{\lambda(i-1)} \times Y \times H_{\lambda(i+1)} \times \dots \times H_{\lambda(k-1)}) \\ & \quad \parallel \\ & \mu_{t_{\pi(1)}, \dots, t_{\pi(i-1)}, t_{\pi(i)}, t_{\pi(i+1)}, \dots, t_{\pi(k)}}(H_{\pi(1)} \times \dots \times H_{\pi(i-1)} \times Y \times H_{\pi(i+1)} \times \dots \times H_{\pi(k)}) \\ & \quad \parallel \\ & \mu_{t_1, \dots, t_{k-1}, t_k}(H_1 \times \dots \times H_{k-1} \times Y). \end{aligned}$$

Esta igualdade encerra a verificação de que a condição (C2) é satisfeita pelos membros da família  $\{\mu_{t_1, \dots, t_k}\}$ . ■

Em resumo, se construímos uma família de medidas de probabilidade  $\mu_{t_1, \dots, t_k}$ , apenas para  $k$ -úplas ordenadas de forma que as condições (5), (6) e (7) sejam satisfeitas e em seguida, estendemos esta família, por meio de (8), para uma família indexada por  $k$ -úplas arbitrárias (**não necessariamente ordenada**), então obtemos uma família consistente de medidas de probabilidade que satisfaz as hipóteses do Teorema de Consistência de Kolmogorov.

Na prática, é frequentemente mais natural construir uma família de medidas de probabilidade apenas sobre  $(k+1)$ -úplas de *inteiros consecutivos*, da forma  $(m, m+1, \dots, m+k)$ , ao invés de  $(k+1)$ -úplas ordenadas de inteiros. Para destacar famílias indexada sobre  $(k+1)$ -úplas consecutivas vamos usar a notação

$$\mathcal{N} \equiv \{\mu_{m, m+1, \dots, m+k} \in \mathcal{M}(Y^{k+1}, \mathcal{B}_{k+1}) : k \geq 0 \text{ e } m \in \mathbb{Z}\}. \quad (9)$$

Observe que, nesta família, a operação de remover um índice central  $j$ , com  $m < j < m+k$ , de uma  $(k+1)$ -úpla consecutiva  $(m, m+1, \dots, m+k)$  produz uma lista que não é mais consecutiva. Logo, não faz sentido impor sobre  $\mathcal{N}$  uma condição de consistência análoga a (5). Já as condições (6) e (7) permanecem relevantes, pois a remoção do primeiro ou do último índice de uma  $(k+1)$ -úpla consecutiva produz uma  $k$ -úpla consecutiva (com ponto inicial deslocado, no caso de remover o primeiro índice). Desta forma é mais natural, exigir de tais tipos de famílias que satisfaçam as duas seguintes condições de consistência:

$$\mu_{m, m+1, \dots, m+k-1}(H_0 \times \dots \times H_{k-1}) = \mu_{m, m+1, \dots, m+k}(H_0 \times \dots \times H_{k-1} \times Y), \quad (10)$$

e

$$\mu_{m+1, m+2, \dots, m+k}(H_1 \times \dots \times H_k) = \mu_{m, m+1, \dots, m+k}(Y \times H_1 \times \dots \times H_k), \quad (11)$$

para todo  $m \in \mathbb{Z}$ ,  $k \geq 1$  e  $H_j \subseteq Y$ , com  $j = 0, 1, \dots, k$ .

**Observação 11.** Diferentemente do que ocorre para famílias indexadas sobre  $k$ -úplas ordenadas, as condições (10) e (11) são essencialmente independentes para famílias indexadas sobre  $k$ -úplas consecutivas. De fato, a condição (10) relaciona membros da família  $\mathcal{N}$  das formas:  $\mu_{m,\dots,m+k}$  e  $\mu_{m,\dots,m+k-1}$ . Enquanto a condição (11) relaciona  $\mu_{m,\dots,m+k}$  com  $\mu_{m+1,\dots,m+k}$ . Tratam-se, portanto, de relações envolvendo membros distintos da família.

A proposição a seguir mostra que, a partir de uma família  $\mathcal{N}$  satisfazendo (10) e (11), é sempre possível construir uma família do tipo  $\mathcal{M}$ , indexada sobre todas as  $k$ -úplas ordenadas de inteiros, satisfazendo as três condições de consistência (5), (6) e (7). Com esta proposição em mãos, a passagem de  $\mathcal{N}$  a uma família consistente, no sentido da Definição 8, segue imediatamente do Teorema 10.

**Proposição 12.** Seja  $\mathcal{N}$  uma família de medidas de probabilidade indexada sobre todas as  $k$ -úplas de inteiros consecutivos, como em (9), satisfazendo as condições de consistência (10) e (11). Então existe uma família  $\mathcal{M} \equiv \{\mu_{t_1,\dots,t_k} \in \mathcal{M}(Y^k, \mathcal{B}_k) : k \in \mathbb{N} \text{ e } t_1 < t_2 < \dots < t_k\}$  satisfazendo as condições de consistência (5), (6) e (7) e contendo  $\mathcal{N}$ , no sentido de que  $\mu_{m,m+1,\dots,m+k}$  coincide com o membro da família  $\mathcal{M}$  indexado pela mesma lista ordenada, para todo  $m \in \mathbb{Z}$  e  $k \geq 0$ .

**Prova.** Seja  $k \geq 1$ ,  $(t_1, \dots, t_k)$  uma  $k$ -úpla de inteiros ordenada, isto é,  $t_1 < t_2 < \dots < t_k$  e  $H_1, \dots, H_k \subseteq Y$ . Defina  $m \equiv t_1$  e  $n \equiv t_k - t_1$ . Considere os seguintes conjuntos de índices  $T \equiv \{t_1, t_2, \dots, t_k\}$  e  $M \equiv \{m, m+1, m+2, \dots, m+n\}$ . É claro que  $T \subseteq M$ . Assim, podemos definir para cada índice  $\ell \in \{0, 1, \dots, n\}$  o conjunto

$$G_\ell \equiv \begin{cases} H_j, & \text{se } m + \ell = t_j; \\ Y, & \text{caso contrário.} \end{cases}$$

Veja o exemplo abaixo para ver como funciona a construção dos conjuntos  $G$ 's à partir dos conjuntos  $H$ 's.

$G_0$	$G_1$	$G_2$	$G_3$	$G_4$	$G_5$	$G_6$	$G_7$	$G_8$
$H_1$	$Y$	$Y$	$H_2$	$Y$	$Y$	$Y$	$Y$	$H_3$
$t_1$			$t_2$					$t_3$
$m$	$m+1$	$m+2$	$m+3$	$\dots$	$\dots$	$\dots$	$\dots$	$m+8$

Agora defina

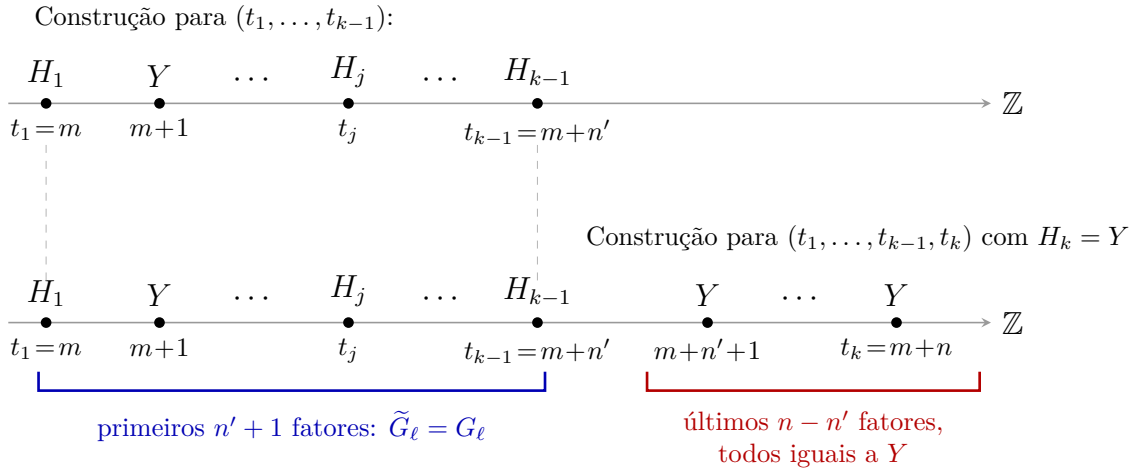
$$\mu_{t_1,\dots,t_k}(H_1 \times \dots \times H_k) \equiv \mu_{m,m+1,\dots,m+n}(G_0 \times G_1 \times \dots \times G_n). \quad (12)$$

É claro que, se  $(t_1, \dots, t_k)$  já é uma  $k$ -úpla consecutiva, então  $n = k - 1$  e  $G_\ell = H_{\ell+1}$  para todo  $\ell \in \{0, \dots, n\}$ , de modo que  $\mu_{t_1,\dots,t_k} = \mu_{m,m+1,\dots,m+(k-1)}$  e portanto a família  $\mathcal{M}$  assim definida contém  $\mathcal{N}$ .

Resta verificar que a família  $\mathcal{M}$  satisfaz as três condições de consistência.

**Condição de Consistência (7).** Sejam  $(t_1, \dots, t_k)$  uma  $k$ -úpla ordenada de inteiros e  $H_1, \dots, H_{k-1} \subseteq Y$ . Defina  $m \equiv t_1$ ,  $n \equiv t_k - t_1$  e  $n' \equiv t_{k-1} - t_1$ . Como  $t_{k-1} < t_k$ , temos  $0 \leq n' < n$ .

A ideia é comparar duas expressões fornecidas por (12): a primeira é obtida usando esta expressão para  $(k-1)$ -úpla  $(t_1, \dots, t_{k-1})$ , que produz a coleção  $(G_0, G_1, \dots, G_{n'})$ , e outra aplicada à  $k$ -úpla  $(t_1, \dots, t_{k-1}, t_k)$  com o último conjunto tomado igual a  $H_k \equiv Y$ , que produz uma segunda coleção, denotada por  $(\tilde{G}_0, \tilde{G}_1, \dots, \tilde{G}_n)$ , para distingui-la da primeira. A figura abaixo ilustra como os dois conjuntos de índices e os valores atribuídos aos  $G$ 's funcionam em cada caso.



A inspeção das duas construções nos dá três observações-chave:

- Para  $\ell \in \{0, 1, \dots, n'\}$ , vale  $\tilde{G}_\ell = G_\ell$ . De fato, se  $m + \ell = t_j$  para algum  $j$ , então  $t_j \leq t_{k-1}$ , ou seja,  $j \leq k - 1$ , e ambas as regras atribuem o mesmo valor  $H_j$ ; caso contrário, ambas atribuem  $Y$ .
- Para  $\ell \in \{n' + 1, \dots, n - 1\}$ , vale  $\tilde{G}_\ell = Y$ . Com efeito, uma eventual identidade  $m + \ell = t_j$  com  $j \in \{1, \dots, k\}$  implicaria  $t_{k-1} < t_j < t_k$ , o que é impossível, já que  $t_{k-1}$  e  $t_k$  são consecutivos na lista ordenada  $(t_1, \dots, t_k)$ .
- Para  $\ell = n$ , vale  $\tilde{G}_n = H_k = Y$ , pela escolha de  $H_k$ .

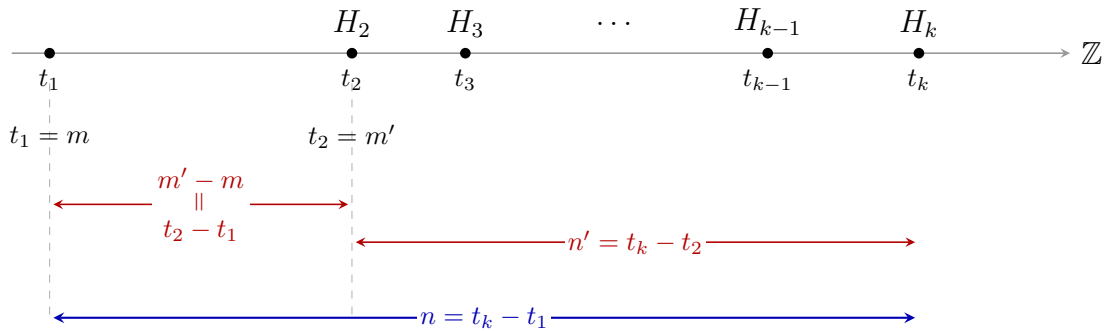
Em suma: os primeiros  $n' + 1$  fatores das duas coleções coincidem, e os  $n - n'$  fatores finais de  $(\tilde{G}_0, \dots, \tilde{G}_n)$  são todos iguais a  $Y$  — o que corresponde, geometricamente, ao colchete azul e ao colchete vermelho da figura, respectivamente. Com estas observações obtemos a cadeia

$$\begin{aligned}
\mu_{t_1, \dots, t_{k-1}}(H_1 \times \dots \times H_{k-1}) &= \mu_{m, \dots, m+n'}(G_0 \times \dots \times G_{n'}) \\
&= \mu_{m, \dots, m+n', \dots, m+n} \left( G_0 \times \dots \times G_{n'} \times \underbrace{Y \times \dots \times Y}_{n-n' \text{ fatores}} \right) \\
&= \mu_{m, \dots, m+n}(\tilde{G}_0 \times \tilde{G}_1 \times \dots \times \tilde{G}_n) \\
&= \mu_{t_1, \dots, t_k}(H_1 \times \dots \times H_{k-1} \times Y),
\end{aligned}$$

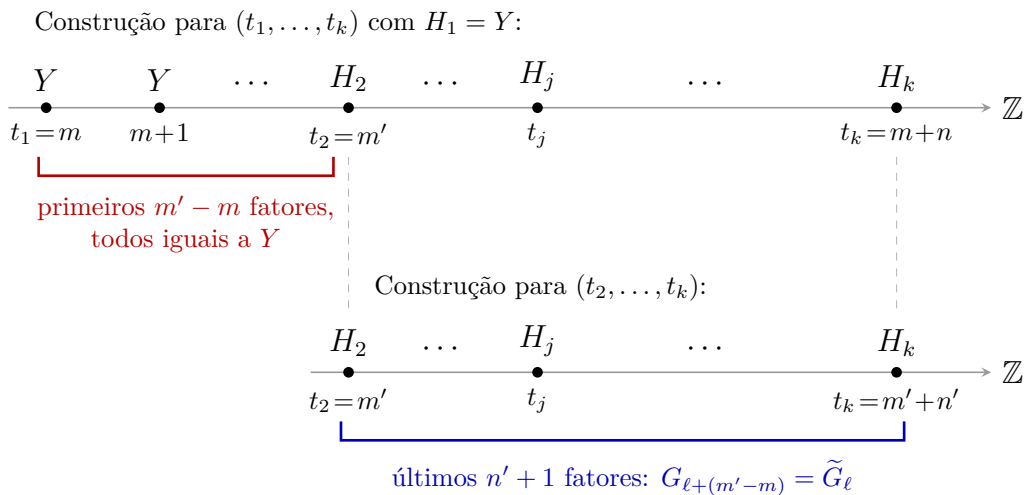
onde as igualdades acima são justificadas da seguinte maneira:

- na primeira, aplicamos a definição (12) à  $(k - 1)$ -úpla  $(t_1, \dots, t_{k-1})$ ;
- na segunda, aplicamos  $(n - n')$  vezes, sucessivamente, a condição de consistência (10), cada aplicação anexando um novo fator  $Y$  à direita;
- na terceira, usamos as três observações acima, que identificam termo a termo o integrando com a coleção  $(\tilde{G}_0, \tilde{G}_1, \dots, \tilde{G}_n)$ ;
- na quarta, aplicamos novamente a definição (12), desta vez à  $k$ -úpla  $(t_1, \dots, t_{k-1}, t_k)$  com  $H_k = Y$ .

**Condição de Consistência (6).** Sejam  $(t_1, \dots, t_k)$  uma  $k$ -úpla ordenada de inteiros e  $H_2, \dots, H_k \subseteq Y$ . Defina  $m \equiv t_1$ ,  $m' \equiv t_2$ ,  $n \equiv t_k - t_1$  e  $n' \equiv t_k - t_2$ . Como  $t_1 < t_2$ , temos  $m < m'$ , e note que  $n = n' + (m' - m)$ .



Novamente, vamos comparar as duas expressões obtidas via (12): a primeira vem da aplicação desta definição à  $k$ -úpla  $(t_1, \dots, t_k)$  com a escolha  $H_1 \equiv Y$ , produzindo a coleção  $(G_0, G_1, \dots, G_n)$ ; a segunda é a aplicação à  $(k - 1)$ -úpla  $(t_2, \dots, t_k)$ , produzindo a coleção que denotaremos por  $(\tilde{G}_0, \tilde{G}_1, \dots, \tilde{G}_{n'})$ . A figura abaixo ilustra o alinhamento das duas construções.



A inspeção geométrica das construções nos revela que:

- Para  $\ell \in \{0, 1, \dots, m' - m - 1\}$ , vale  $G_\ell = Y$ . Isto ocorre porque nenhum dos índices  $m, m + 1, \dots, m' - 1$  pertence ao conjunto  $\{t_2, t_3, \dots, t_k\}$ , e o conjunto associado ao próprio  $t_1$  foi definido como  $H_1 = Y$ .

- Para  $\ell \in \{0, 1, \dots, n'\}$ , vale  $\tilde{G}_\ell = G_{\ell+(m'-m)}$ . De fato, a posição  $\ell$  a partir de  $m'$  na segunda construção refere-se ao índice  $m' + \ell$ , que corresponde exatamente à posição  $\ell + (m' - m)$  a partir de  $m$  na primeira construção. As regras de atribuição de  $H_j$  ou  $Y$  são idênticas a partir de  $t_2$ .

Assim, a primeira coleção possui um bloco inicial de  $m' - m$  conjuntos  $Y$ , seguido exatamente pela segunda coleção. Com isso, obtemos:

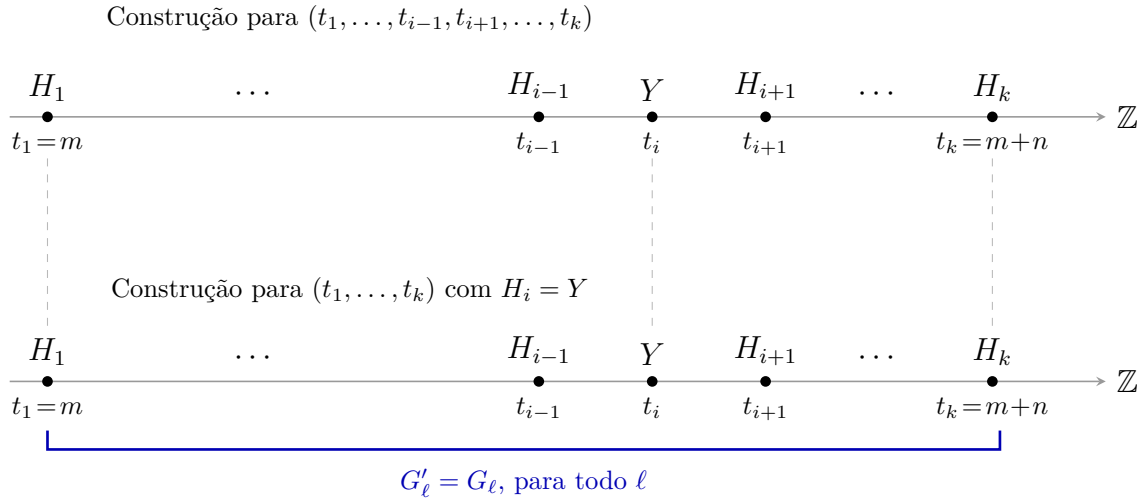
$$\begin{aligned}
\mu_{t_1, \dots, t_k}(Y \times H_2 \times \dots \times H_k) &= \mu_{m, \dots, m+n}(G_0 \times \dots \times G_n) \\
&= \mu_{m, \dots, m', \dots, m+n} \left( \underbrace{Y \times \dots \times Y}_{m'-m \text{ fatores}} \times G_{m'-m} \times \dots \times G_n \right) \\
&= \mu_{m', \dots, m+n}(G_{m'-m} \times \dots \times G_n) \\
&= \mu_{m', \dots, m'+n'}(\tilde{G}_0 \times \dots \times \tilde{G}_{n'}) \\
&= \mu_{t_2, \dots, t_k}(H_2 \times \dots \times H_k),
\end{aligned}$$

onde cada uma das igualdades acima é justificada da seguinte maneira

- na primeira igualdade, usamos a definição (12) para a  $k$ -úpla  $(t_1, \dots, t_k)$  com  $H_1 = Y$ ;
- na segunda, substituímos os primeiros fatores por  $Y$ , conforme a primeira observação;
- na terceira, aplicamos  $(m' - m)$  vezes, sucessivamente, a condição de consistência à esquerda (11), a qual tem o efeito de tomar a marginal (remove) nas primeiras coordenadas  $Y$ 's e desloca o índice inicial até  $m'$ ;
- na quarta, reescrevemos o integrando usando a identidade termo a termo da segunda observação, notando que  $m + n = m' + n'$ ;
- na quinta, usamos novamente a definição (12), agora de trás para frente, para a  $(k - 1)$ -úpla  $(t_2, \dots, t_k)$ .

**Condição de Consistência (5).** Sejam  $(t_1, \dots, t_k)$  uma  $k$ -úpla ordenada,  $1 < i < k$  e  $H_j \subseteq Y$  para  $j \in \{1, \dots, k\} \setminus \{i\}$ . Defina  $m \equiv t_1$  e  $n \equiv t_k - t_1$ .

Vamos mais uma vez confrontar duas aplicações da definição (12). O primeiro cenário será para a  $(k - 1)$ -úpla ordenada  $(t_1, \dots, t_{i-1}, t_{i+1}, \dots, t_k)$ , que omite  $t_i$  e produz uma coleção  $(G'_0, G'_1, \dots, G'_n)$ . O segundo cenário será para a  $k$ -úpla original  $(t_1, \dots, t_k)$ , porém impondo  $H_i \equiv Y$ , o que produz a coleção  $(G_0, G_1, \dots, G_n)$ . Como os extremos  $t_1$  e  $t_k$  são idênticos em ambos os casos, o comprimento  $n$  das duas coleções é rigorosamente o mesmo.



A análise das coleções geradas torna a prova direta, baseada na seguinte observação:

- Para todo  $\ell \in \{0, 1, \dots, n\}$ , temos  $G'_\ell = G_\ell$ . A única posição onde as construções poderiam diferir seria em  $\ell^* \equiv t_i - m$ . Na primeira construção (que omite  $t_i$ ), o índice  $m + \ell^* = t_i$  não figura entre os conjuntos especificados, logo a regra dita que  $G'_{\ell^*} = Y$ . Na segunda construção, o índice  $t_i$  está presente, mas especificamos que  $H_i = Y$ , de modo que a regra dita  $G_{\ell^*} = H_i = Y$ . Nos demais índices, ambas as regras coincidem perfeitamente.

Sendo as sequências de conjuntos idênticas, basta reescrever os termos:

$$\begin{aligned}
& \mu_{t_1, \dots, t_{i-1}, t_{i+1}, \dots, t_k}(H_1 \times \dots \times H_{i-1} \times H_{i+1} \times \dots \times H_k) \\
&= \mu_{m, \dots, m+n}(G'_0 \times \dots \times G'_n) \\
&= \mu_{m, \dots, m+n}(G_0 \times \dots \times G_n) \\
&= \mu_{t_1, \dots, t_k}(H_1 \times \dots \times H_{i-1} \times Y \times H_{i+1} \times \dots \times H_k).
\end{aligned}$$

A cadeia acima é validada pelos seguintes passos:

- na primeira igualdade, avaliamos a definição (12) para a coleção com a ausência de  $t_i$ ;
- na segunda igualdade, utilizamos a observação feita acima de que  $(G'_0, \dots, G'_n) = (G_0, \dots, G_n)$ ;
- na terceira igualdade, aplicamos a definição (12) de trás para frente para a  $k$ -úpla completa, alocando  $H_i = Y$ .

Isto encerra a prova da condição (5). ■

Com a **Proposição 12** estabelecida, o teorema de construção a partir de famílias indexadas sobre  $k$ -úplas consecutivas torna-se um corolário imediato.

**Teorema 13.** Seja  $\mathcal{N}$  uma família de medidas de probabilidade indexada sobre todas as  $k$ -úplas de inteiros consecutivos, como em (9). Suponha que esta família satisfaça as condições de consistência (10) e (11). Então existe uma única medida de probabilidade  $\mu$  definida sobre o espaço produto  $(Y^{\mathbb{Z}}, \mathcal{B})$  tal que para quaisquer  $k \in \mathbb{N}$ ,  $m \in \mathbb{Z}$  e  $H_j \subseteq Y$  temos:

$$\mu(\{x \in X : x_m \in H_0, x_{m+1} \in H_1, \dots, x_{m+k} \in H_k\}) = \mu_{m, m+1, \dots, m+k}(H_0 \times \dots \times H_k).$$

**Prova.** Pela **Proposição 12**, existe uma família  $\mathcal{M} \equiv \{\mu_{t_1, \dots, t_k} : k \in \mathbb{N}, t_1 < \dots < t_k\}$  satisfazendo as condições de consistência (5), (6) e (7), e contendo  $\mathcal{N}$ . Pelo **Teorema 10**, existe uma família consistente  $\{\mu_{t_1, \dots, t_k}\}$  (no sentido da **Definição 8**) contendo  $\mathcal{M}$ . Aplicando o Teorema da Existência de Kolmogorov a esta família consistente, obtemos uma medida de probabilidade  $\mu$  sobre  $(Y^{\mathbb{Z}}, \mathcal{B})$  satisfazendo a identidade desejada nos cilindros finitos-dimensionais com índices consecutivos.

Para a unicidade, suponha que  $\mu^{(1)}$  e  $\mu^{(2)}$  sejam duas medidas de probabilidade sobre  $(Y^{\mathbb{Z}}, \mathcal{B})$  satisfazendo a identidade do enunciado. Então ambas coincidem sobre todos os cilindros da forma  $\{x \in X : x_m \in H_0, x_{m+1} \in H_1, \dots, x_{m+k} \in H_k\}$ , com  $m \in \mathbb{Z}$ ,  $k \geq 0$  e  $H_j \subseteq Y$ . A família destes cilindros gera a  $\sigma$ -álgebra produto  $\mathcal{B}$  e é fechada por interseções finitas (a interseção de dois cilindros com índice consecutivos pode ser vista como um cilindro cujo o conjunto de índices é o subconjunto de pontos consecutivos de  $\mathbb{Z}$  delimitado pelo mínimo do conjuntos de índices da união dos dois e pelo máximo. Note que para ver esta interseção como um cilindro alguns dos fatores devem ser tomados iguais a  $Y$ ). Logo, pelo Teorema  $\pi$ - $\lambda$  de Dynkin, segue que  $\mu^{(1)} = \mu^{(2)}$  em  $\mathcal{B}$ , o que encerra a prova da unicidade. ■

## 4. Shifts de Bernoulli e Shifts de Markov

Nesta seção vamos construir em detalhe as duas classes de sistemas dinâmicos simbólicos que serão estudadas ao longo deste texto: os *shifts de Bernoulli* e os *shifts de Markov*. Em ambos os casos, o espaço subjacente é o espaço produto  $X \equiv Y^{\mathbb{Z}}$ , onde  $Y$  é um conjunto finito, e a transformação  $T : X \rightarrow X$  é o shift para a esquerda. O que distingue os dois modelos é a medida de probabilidade considerada em  $X$ : no primeiro caso, temos uma medida produto (coordenadas independentes); no segundo, temos a medida  $(p, P)$ -Markoviana (dependência de primeira ordem).

Nosso objetivo é construir rigorosamente estas medidas, verificando em cada etapa que as distribuições finito-dimensionais associadas satisfazem as condições de consistência do Teorema de Extensão de Kolmogorov.

### 4.1. O Espaço Produto e o Shift para a Esquerda

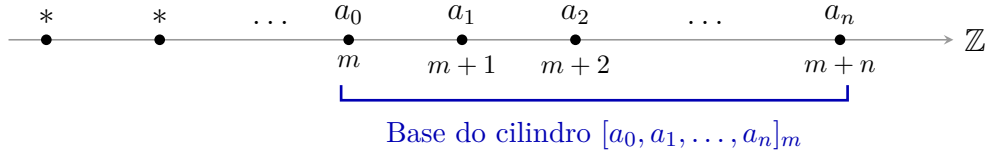
Seja  $Y \equiv \{0, 1, \dots, k-1\}$  um conjunto finito com  $k \geq 2$  elementos, equipado com a  $\sigma$ -álgebra das partes. Vamos considerar o espaço produto

$$X \equiv Y^{\mathbb{Z}} = \{x = (x_n)_{n \in \mathbb{Z}} : x_n \in Y, \forall n \in \mathbb{Z}\}$$

equipado com a  $\sigma$ -álgebra produto. Como estamos assumindo que  $Y$  é um conjunto finito então a  $\sigma$ -álgebra produto é gerada pelos conjuntos cilíndricos da forma

$$[a_0, a_1, \dots, a_n]_m \equiv \{x \in X : x_m = a_0, x_{m+1} = a_1, \dots, x_{m+n} = a_n\}, \quad (13)$$

onde  $m \in \mathbb{Z}$ ,  $n \in \mathbb{Z}_+$  e  $a_0, \dots, a_n \in Y$  são fixados.



A coleção de todos os cilindros, definidos em (13), juntamente com o conjunto vazio e o próprio  $X$ , forma uma semi-álgebra que denotaremos por  $\mathcal{S}$ . É um resultado clássico de Teoria da Medida que  $\sigma(\mathcal{S})$  coincide com a  $\sigma$ -álgebra produto em  $X$ , denotada por  $\mathcal{B}$ .

O shift para a esquerda  $T : X \rightarrow X$  é a aplicação definida por

$$(Tx)_n \equiv x_{n+1}, \quad \forall n \in \mathbb{Z}, \forall x \in X.$$

Não é difícil verificar que  $T$  é uma bijeção mensurável com inversa mensurável, e portanto os conceitos de medida  $T$ -invariante e ergodicidade estão bem definidos em  $(X, \mathcal{B})$ .

A ação de  $T$  sobre cilindros da forma (13) é especialmente simples e será usada repetidamente nas construções que seguem:

$$T^{-1}([a_0, a_1, \dots, a_n]_m) = [a_0, a_1, \dots, a_n]_{m+1}. \quad (14)$$

De fato,  $x \in T^{-1}([a_0, \dots, a_n]_m)$  se, e somente se,  $Tx \in [a_0, \dots, a_n]_m$ , o que por sua vez equivale a  $(Tx)_{m+j} = a_j$  para todo  $j \in \{0, 1, \dots, n\}$ . Como, por definição,  $(Tx)_{m+j} = x_{m+j+1}$ , isso implica que  $x_{m+j+1} = a_j$ . Reescrevendo, isto é exatamente a afirmação de que  $x_{(m+1)+j} = a_j$  para todo  $j \in \{0, \dots, n\}$ , e portanto  $x \in [a_0, \dots, a_n]_{m+1}$ .

## 4.2. Shifts de Bernoulli

Nesta subseção vamos construir a primeira das duas medidas  $T$ -invariantes naturais sobre  $X = Y^{\mathbb{Z}}$ . Essa medida é chamada de medida produto ou medida de Bernoulli.

**Definição 14.** Um vetor  $p = (p_0, p_1, \dots, p_{k-1}) \in \mathbb{R}^k$  é dito um *vetor de probabilidade* se  $p_i \geq 0$  para todo  $i \in Y$  e  $\sum_{i \in Y} p_i = 1$ .

Fixado um vetor de probabilidade  $p$ , defina, para cada cilindro  $[a_0, a_1, \dots, a_n]_m$ , a quantidade

$$\mu([a_0, a_1, \dots, a_n]_m) \equiv p_{a_0} p_{a_1} \cdots p_{a_n}. \quad (15)$$

Observe que o lado direito da expressão acima não depende de  $m$ , isto é, não depende da posição em que começa o conjunto de índices que determina o cilindro. Esta é a observação será a chave para mostrar que a função de conjuntos definida em (15) pode ser estendida à uma medida de probabilidade que é invariante pelo shift.

Nosso próximo objetivo é mostrar que a função de conjunto definida em (15) se estende unicamente a uma medida de probabilidade em  $(X, \mathcal{B})$ . Para isto, vamos construir a partir (15) uma família de consistente de medidas de probabilidade e invocar o Teorema da Existência de Kolmogorov.

Para cada  $m \in \mathbb{Z}$  e  $n \in \mathbb{Z}_+$ , definimos uma medida de probabilidade  $\mu_{m,n}$ , associada aos índices de inteiros consecutivos  $m, m+1, \dots, m+n$ , como sendo a única medida de probabilidade em  $Y^{n+1}$  determinada pelos seguintes valores:

$$\mu_{m,n}(\{a_0\} \times \{a_1\} \times \dots \times \{a_n\}) \equiv \mu_{m,n}(a_0, a_1, \dots, a_n) \equiv p_{a_0} p_{a_1} \cdots p_{a_n}, \quad (16)$$

para todo  $a_0, a_1, \dots, a_n \in Y$ .

Vamos mostrar abaixo que é possível usar o **Teorema 13** e concluir que existe uma única medida de probabilidade  $\mu$  sobre  $X$ , tal que as probabilidades dos conjuntos cilíndricos são completamente determinadas pela família  $\{\mu_{m,n}\}_{m \in \mathbb{Z}, n \in \mathbb{N}}$ .

**Proposição 15.** A família  $\{\mu_{m,n}\}_{m \in \mathbb{Z}, n \in \mathbb{Z}_+}$ , com  $\mu_{m,n}$  definida como (16), é uma família de medidas de probabilidade satisfazendo as condições de consistência (10) e (11).

**Prova.** Primeiro, observamos que usando a notação do **Teorema 13**, temos a seguinte igualdade entre famílias de medidas de probabilidade

$$\{\mu_{m,k}\}_{m \in \mathbb{Z}, k \in \mathbb{Z}_+} = \{\mu_{m,m+1,\dots,m+k} \in \mathcal{M}(Y^{k+1}, \mathcal{B}_{k+1}) : k \geq 0 \text{ e } m \in \mathbb{Z}\}.$$

Por questão de simplicidade, vamos começar verificando a validade das condições de consistência (10) e (11) para conjuntos  $H_0 \times \dots \times H_n$  da forma  $\{a_0\} \times \dots \times \{a_n\}$ .

Nosso primeiro passo será verificar a validade da condição (10) - marginal à direita. Neste caso, temos que mostrar que

$$\sum_{a_n \in Y} \mu_{m,n}(a_0, \dots, a_{n-1}, a_n) = \mu_{m,n-1}(a_0, \dots, a_{n-1}).$$

Usando a definição de  $\mu_{m,n}$  dada em (16), temos

$$\begin{aligned} \sum_{a_n \in Y} \mu_{m,n}(a_0, \dots, a_{n-1}, a_n) &= \sum_{a_n \in Y} p_{a_0} p_{a_1} \cdots p_{a_{n-1}} p_{a_n} \\ &= p_{a_0} p_{a_1} \cdots p_{a_{n-1}} \sum_{a_n \in Y} p_{a_n} \\ &= p_{a_0} p_{a_1} \cdots p_{a_{n-1}} \cdot 1 \\ &= \mu_{m,n-1}(a_0, \dots, a_{n-1}), \end{aligned}$$

onde usamos  $\sum_{a_n \in Y} p_{a_n} = 1$ , que segue de  $p$  ser um vetor de probabilidade.

Vamos verificar, agora, que a validade da condição (11) - marginal à esquerda,

$$\sum_{a_0 \in Y} \mu_{m,n}(a_0, a_1, \dots, a_n) = \mu_{m+1,n-1}(a_1, \dots, a_n),$$

pode ser provada de forma completamente análoga à anterior. De fato, segue diretamente da definição (16) que

$$\begin{aligned} \sum_{a_0 \in Y} \mu_{m,n}(a_0, a_1, \dots, a_n) &= \sum_{a_0 \in Y} p_{a_0} p_{a_1} \cdots p_{a_n} \\ &= \left( \sum_{a_0 \in Y} p_{a_0} \right) p_{a_1} \cdots p_{a_n} \\ &= p_{a_1} \cdots p_{a_n} \\ &= \mu_{m+1,n-1}(a_1, \dots, a_n), \end{aligned}$$

onde usamos novamente apenas  $\sum_{a_0 \in Y} p_{a_0} = 1$ .

Para o caso geral, consideremos conjuntos arbitrários  $H_j \subseteq Y$ . Por definição, a medida de um retângulo da forma  $H_0 \times \cdots \times H_n$  é dada pela soma das medidas de seus átomos, ou seja,

$$\mu_{m,n}(H_0 \times \cdots \times H_n) = \sum_{a_0 \in H_0} \cdots \sum_{a_n \in H_n} \mu_{m,n}(a_0, \dots, a_n).$$

Utilizando a forma de produto obtida para os átomos, temos

$$\mu_{m,n}(H_0 \times \cdots \times H_n) = \left( \sum_{a_0 \in H_0} p_{a_0} \right) \cdots \left( \sum_{a_n \in H_n} p_{a_n} \right).$$

Assim, para verificar a validade de (10), basta notar que ao tomarmos  $H_n = Y$ , o último fator na expressão acima torna-se a soma  $\sum_{a_n \in Y} p_{a_n} = 1$ . Portanto,

$$\begin{aligned} \mu_{m,n}(H_0 \times \cdots \times H_{n-1} \times Y) &= \prod_{j=0}^{n-1} \left( \sum_{a_j \in H_j} p_{a_j} \right) \cdot 1 \\ &= \mu_{m,n-1}(H_0 \times \cdots \times H_{n-1}). \end{aligned}$$

De forma análoga, para a condição (11), ao tomarmos  $H_0 = Y$ , o primeiro fator da decomposição resulta em 1, restando apenas os termos que definem  $\mu_{m+1,n-1}(H_1 \times \cdots \times H_n)$ . ■

**Observação 16.** É instrutivo observar que, no caso Bernoulli, as duas condições de consistência seguem da *mesma* propriedade elementar de  $p$ : a de ser um vetor de probabilidade, isto é,  $\sum_{i \in Y} p_i = 1$ . Esta uniformidade reflete a independência das coordenadas: como nenhuma coordenada “conversa” com as outras, tomar qualquer uma das marginais (à esquerda ou à direita) tem o mesmo efeito algébrico. Veremos a seguir que, no caso Markoviano, esta simetria se quebra: a consistência para cada tipo de marginal corresponderá a uma propriedade estrutural distinta do par  $(p, P)$ .

**Teorema 17** (Existência e Unicidade da Medida de Bernoulli). Seja  $Y \equiv \{0, 1, \dots, k-1\}$  um conjunto finito e considere o espaço produto  $X \equiv Y^{\mathbb{Z}}$  equipado com a  $\sigma$ -álgebra produto  $\mathcal{B}$ . Dado um vetor de probabilidade  $p = (p_0, p_1, \dots, p_{k-1})$ , existe uma única medida de probabilidade  $\mu$  em  $(X, \mathcal{B})$  tal que, para qualquer cilindro da forma  $[a_0, a_1, \dots, a_n]_m \in \mathcal{S}$ , temos

$$\mu([a_0, a_1, \dots, a_n]_m) = p_{a_0} p_{a_1} \cdots p_{a_n}. \quad (17)$$

Esta medida  $\mu$  é denominada *medida de Bernoulli* ou *medida produto* determinada pelo vetor de probabilidade  $p$ .

**Prova.** Como vimos na **Proposição 15**, sabemos que a família  $\{\mu_{m,n}\}_{m \in \mathbb{Z}, n \in \mathbb{Z}_+}$ , definida a partir de (16), satisfaz as condições de consistência (10) e (11). Portanto, podemos aplicar o **Teorema 13** a esta família, mostrando assim a existência e a unicidade da medida de probabilidade  $\mu$  em  $(X, \mathcal{B})$  satisfazendo (17). ■

**Proposição 18.** A medida  $\mu$  de Bernoulli é  $T$ -invariante.

**Prova.** Como a coleção  $\mathcal{S}$  de todos os cilindros em  $X$  da forma  $[a_0, a_1, \dots, a_n]_m$  é uma semi-álgebra e  $\sigma(\mathcal{S}) = \mathcal{B}$ , basta verificar que  $\mu(T^{-1}C) = \mu(C)$  para todo elemento  $C \in \mathcal{S}$ . Usando (14) e (15), temos

$$\begin{aligned} \mu(T^{-1}[a_0, a_1, \dots, a_n]_m) &= \mu([a_0, a_1, \dots, a_n]_{m+1}) \\ &= p_{a_0} p_{a_1} \cdots p_{a_n} \\ &= \mu([a_0, a_1, \dots, a_n]_m). \end{aligned}$$

O que encerra a prova da proposição. ■

O sistema  $(X, \mathcal{B}, \mu, T)$  assim construído é chamado *shift de Bernoulli* com vetor de probabilidade  $p$ .

### 4.3. Shifts de Markov

Passamos agora à construção da medida  $(p, P)$ -Markoviana, que generaliza a medida de Bernoulli permitindo dependência de primeira ordem entre coordenadas consecutivas. Esta medida será de suma importância na prova do [Lema 27](#) e do [Teorema 28](#), que são resultados centrais na teoria de cadeias de Markov estacionárias. Além do mais, estas medidas fornecem exemplos muito importantes em Teoria Ergódica.

**Definição 19.** Uma matriz  $P \equiv (P_{ij})_{i,j \in Y} \in M_{k \times k}(\mathbb{R})$  é dita uma *matriz estocástica* se  $P_{ij} \geq 0$  para todos  $i, j \in Y$  e  $\sum_{j \in Y} P_{ij} = 1$  para todo  $i \in Y$ .

Na teoria de Processos Estocásticos costumamos interpretar  $P_{ij}$  como a probabilidade de um sistema fazer uma transição de seu estado  $i$  para outro estado  $j$ , em um passo.

**Definição 20.** Dada uma matriz estocástica  $P$ , um vetor de probabilidade  $p \in \mathbb{R}^k$  é dito uma *distribuição estacionária* de  $P$  se  $pP = p$ , isto é,

$$\sum_{i \in Y} p_i P_{ij} = p_j, \quad \forall j \in Y.$$

Fixado um par  $(p, P)$  onde  $p$  é uma distribuição estacionária de  $P$ , definimos, para cada cilindro  $[a_0, a_1, \dots, a_n]_m$ , a quantidade

$$\mu([a_0, a_1, \dots, a_n]_m) \equiv p_{a_0} P_{a_0 a_1} P_{a_1 a_2} \cdots P_{a_{n-1} a_n}. \quad (18)$$

Assim como no caso Bernoulli, o lado direito *não depende de  $m$* . Esta propriedade será, ao final, responsável pela  $T$ -invariância da medida resultante. Quando olhamos para estes objetos do ponto de vista probabilístico a diferença crucial, com respeito a medida de Bernoulli é que, agora, os fatores  $P_{a_{j-1} a_j}$  representam probabilidades condicionais. Por exemplo,  $P_{a_{j-1} a_j}$  é interpretado como probabilidade de observar  $a_j$  na posição  $m + j$ , dado que observamos  $a_{j-1}$  na posição  $m + j - 1$ . Embora, as interpretações probabilísticas sejam muito úteis, as demonstrações e o restante da discussão é desenvolvido de modo independente da Teoria de Processos Estocásticos.

Como no caso das medidas de Bernoulli, vamos mostrar como é possível estender a função de conjuntos definida em (18) a uma única medida  $\mu$ , sobre  $X$ , usando o [Teorema 13](#) que é na verdade, uma versão do Teorema de Extensão de Kolmogorov.

Nesta seção vamos trabalhar com a seguinte família de medidas de probabilidade

$$\mu_{m,n}(a_0, a_1, \dots, a_n) \equiv p_{a_0} P_{a_0 a_1} P_{a_1 a_2} \cdots P_{a_{n-1} a_n}, \quad \forall a_0, \dots, a_n \in Y, \quad (19)$$

que, como observado, não depende de  $m$ . Passemos à verificação das três condições de consistência.

**Proposição 21.** Seja  $P$  uma matriz estocástica e  $p$  um vetor de probabilidade. A família  $\{\mu_{m,n}\}_{m \in \mathbb{Z}, n \in \mathbb{Z}_+}$  definida em (19) é uma família de medidas de probabilidade satisfazendo as condições de consistência (10) e (11) se, e somente se,  $p$  é uma distribuição estacionária de  $P$ , isto é,  $pP = p$ .

**Prova.** Vamos examinar cada uma das duas condições de consistência separadamente, mostrando que cada uma corresponde a uma propriedade estrutural distinta do par  $(p, P)$ .

Vamos começar mostrando que a condição de consistência (10) segue de  $P$  ser uma matriz estocástica. Como na prova da [Proposição 15](#), vamos começar mostrando a validade da condição de consistência (10) para retângulos do espaço produto que são produtos cartesianos de conjuntos unitários. Ou seja, vamos mostrar inicialmente que

$$\sum_{a_n \in Y} \mu_{m,n}(a_0, \dots, a_{n-1}, a_n) = \mu_{m,n-1}(a_0, \dots, a_{n-1}), \quad \text{para cada } (a_0, \dots, a_{n-1}) \in Y^n.$$

De fato, usando (19),

$$\begin{aligned} \sum_{a_n \in Y} \mu_{m,n}(a_0, \dots, a_{n-1}, a_n) &= \sum_{a_n \in Y} p_{a_0} P_{a_0 a_1} \cdots P_{a_{n-2} a_{n-1}} P_{a_{n-1} a_n} \\ &= p_{a_0} P_{a_0 a_1} \cdots P_{a_{n-2} a_{n-1}} \underbrace{\sum_{a_n \in Y} P_{a_{n-1} a_n}}_{=1} \\ &= p_{a_0} P_{a_0 a_1} \cdots P_{a_{n-2} a_{n-1}} \\ &= \mu_{m,n-1}(a_0, \dots, a_{n-1}), \end{aligned}$$

onde usamos que  $\sum_{a_n \in Y} P_{a_{n-1} a_n} = 1$ , que é precisamente a condição de  $P$  ser estocástica. Observe que esta condição é precisamente uma das condições que define uma matriz estocástica. Portanto, para tomar a marginal à direita não é necessário exigir nenhuma hipótese sobre existência ou não do autovetor à esquerda  $p$ .

Agora, vamos verificar a validade da condição de consistência (11) para retângulos que são da mesma forma que no argumento anterior. Para fazer isto veremos que será crucial assumir a validade da hipótese  $pP = p$ . Na verdade, para esta família de medidas de probabilidade a condição (11) é equivalente à  $pP = p$ .

Primeiro, observe que para os tipos de retângulos que estamos considerando a condição de consistência (11) é simplesmente

$$\sum_{a_0 \in Y} \mu_{m,n}(a_0, a_1, \dots, a_n) = \mu_{m+1,n-1}(a_1, \dots, a_n).$$

Veja que segue de (19) e da definição de produto de vetor por matriz que

$$\begin{aligned} \sum_{a_0 \in Y} \mu_{m,n}(a_0, a_1, \dots, a_n) &= \sum_{a_0 \in Y} p_{a_0} P_{a_0 a_1} P_{a_1 a_2} \cdots P_{a_{n-1} a_n} \\ &= \left( \sum_{a_0 \in Y} p_{a_0} P_{a_0 a_1} \right) P_{a_1 a_2} \cdots P_{a_{n-1} a_n} \\ &= (pP)_{a_1} \cdot P_{a_1 a_2} \cdots P_{a_{n-1} a_n}. \end{aligned}$$

Por outro lado, temos novamente de (19) que

$$\mu_{m+1,n-1}(a_1, \dots, a_n) = p_{a_1} P_{a_1 a_2} \cdots P_{a_{n-1} a_n}.$$

Igualando as duas expressões obtidas acima ficamos com

$$(pP)_{a_1} \cdot P_{a_1 a_2} \cdots P_{a_{n-1} a_n} = p_{a_1} P_{a_1 a_2} \cdots P_{a_{n-1} a_n}$$

Como a igualdade acima é válida para todos  $(a_2, \dots, a_n) \in Y^{n-1}$ , podemos somar ambos lados da igualdade acima sobre todas estas  $(n-1)$ -úplas e usar que  $P$  uma matriz estocástica para obter a seguinte igualdade

$$(pP)_{a_1} = (pP)_{a_1} \cdot \sum_{(a_2, \dots, a_n)} P_{a_1 a_2} \cdots P_{a_{n-1} a_n} = p_{a_1} \sum_{(a_2, \dots, a_n)} P_{a_1 a_2} \cdots P_{a_{n-1} a_n} = p_{a_1},$$

para todo  $a_1 \in Y$ . Deste fato, concluímos que a condição de consistência (11) para esta família é consequência de

$$(pP)_{a_1} = p_{a_1}, \quad \forall a_1 \in Y.$$

Isto é,  $pP = p$ .

Reciprocamente, suponha que a família  $\{\mu_{m,n}\}$  satisfaça a condição de consistência (11). Tomando  $n = 1$  e fixando  $a_1 \in Y$ , segue da condição de consistência (11) que

$$(pP)_{a_1} = \sum_{a_0 \in Y} p_{a_0} P_{a_0 a_1} = \sum_{a_0 \in Y} \mu_{m,1}(a_0, a_1) = \mu_{m+1,0}(a_1) = p_{a_1},$$

o que mostra que  $pP = p$ .

Resta estender a validade das duas condições de consistência aos retângulos arbitrários da forma  $H_0 \times \cdots \times H_n$ , com  $H_j \subseteq Y$ . Por definição, a medida de um tal retângulo é dada por

$$\begin{aligned} \mu_{m,n}(H_0 \times \cdots \times H_n) &= \sum_{a_0 \in H_0} \cdots \sum_{a_n \in H_n} \mu_{m,n}(a_0, \dots, a_n) \\ &= \sum_{a_0 \in H_0} \cdots \sum_{a_n \in H_n} p_{a_0} P_{a_0 a_1} P_{a_1 a_2} \cdots P_{a_{n-1} a_n}. \end{aligned}$$

Para verificar a condição (10), tomamos  $H_n = Y$  na expressão acima e efetuamos primeiro a soma sobre  $a_n$ . Como o único fator que depende de  $a_n$  é  $P_{a_{n-1} a_n}$ , e  $\sum_{a_n \in Y} P_{a_{n-1} a_n} = 1$  pela

estocasticidade de  $P$ , obtemos

$$\begin{aligned}
\mu_{m,n}(H_0 \times \cdots \times H_{n-1} \times Y) &= \sum_{a_0 \in H_0} \cdots \sum_{a_{n-1} \in H_{n-1}} p_{a_0} P_{a_0 a_1} \cdots P_{a_{n-2} a_{n-1}} \underbrace{\sum_{a_n \in Y} P_{a_{n-1} a_n}}_{=1} \\
&= \sum_{a_0 \in H_0} \cdots \sum_{a_{n-1} \in H_{n-1}} p_{a_0} P_{a_0 a_1} \cdots P_{a_{n-2} a_{n-1}} \\
&= \mu_{m,n-1}(H_0 \times \cdots \times H_{n-1}).
\end{aligned}$$

Analogamente, para verificar a condição (11), tomamos  $H_0 = Y$  e efetuamos primeiro a soma sobre  $a_0$ . Os únicos fatores que dependem de  $a_0$  são  $p_{a_0}$  e  $P_{a_0 a_1}$ , de modo que

$$\begin{aligned}
\mu_{m,n}(Y \times H_1 \times \cdots \times H_n) &= \sum_{a_1 \in H_1} \cdots \sum_{a_n \in H_n} \left( \sum_{a_0 \in Y} p_{a_0} P_{a_0 a_1} \right) P_{a_1 a_2} \cdots P_{a_{n-1} a_n} \\
&= \sum_{a_1 \in H_1} \cdots \sum_{a_n \in H_n} (pP)_{a_1} P_{a_1 a_2} \cdots P_{a_{n-1} a_n}.
\end{aligned}$$

Usando agora a hipótese  $pP = p$ , temos  $(pP)_{a_1} = p_{a_1}$  para todo  $a_1 \in Y$ , e portanto

$$\begin{aligned}
\mu_{m,n}(Y \times H_1 \times \cdots \times H_n) &= \sum_{a_1 \in H_1} \cdots \sum_{a_n \in H_n} p_{a_1} P_{a_1 a_2} \cdots P_{a_{n-1} a_n} \\
&= \mu_{m+1,n-1}(H_1 \times \cdots \times H_n).
\end{aligned}$$

Observe que a recíproca já está contemplada, pois, tomando  $H_j = \{a_j\}$  para cada  $j$ , recuperamos o caso dos retângulos formados por conjuntos unitários tratado anteriormente. ■

**Observação 22.** Em contraste com o **Observação 16** do caso Bernoulli, no caso Markoviano as duas condições de consistência correspondem a duas propriedades estruturais distintas:

- (a) Consistência das marginais à direita  $\longleftrightarrow$  estocasticidade de  $P$  (linhas somam 1);
- (b) Consistência das marginais à esquerda  $\longleftrightarrow$  estacionariedade  $pP = p$ .

A propriedade (a) é trivialmente satisfeita pela definição de matriz estocástica, mas a propriedade (b) é uma condição genuinamente não-trivial que restringe a escolha do vetor  $p$ .

**Teorema 23** (Medida  $(p, P)$ -Markoviana). Seja  $Y \equiv \{0, 1, \dots, k-1\}$  um conjunto finito e considere o espaço produto  $X \equiv Y^{\mathbb{Z}}$  equipado com a  $\sigma$ -álgebra produto  $\mathcal{B}$ . Dada uma matriz estocástica  $P = (P_{ij})_{i,j \in Y}$  e um vetor de probabilidade  $p = (p_0, p_1, \dots, p_{k-1})$ , satisfazendo  $pP = p$ , então existe uma única medida de probabilidade  $\mu$  em  $(X, \mathcal{B})$  tal que, para qualquer cilindro da forma  $[a_0, a_1, \dots, a_n]_m \in \mathcal{S}$ , temos

$$\mu([a_0, a_1, \dots, a_n]_m) = p_{a_0} P_{a_0 a_1} P_{a_1 a_2} \cdots P_{a_{n-1} a_n}, \quad (20)$$

Esta medida  $\mu$  é chamada de medida  $(p, P)$ -Markoviana.

**Prova.** Como estamos assumindo que  $pP = p$ , segue da **Proposição 21** que a família  $\{\mu_{m,n}\}_{m \in \mathbb{Z}, n \in \mathbb{Z}_+}$ , definida a partir de (19), satisfaz as condições de consistência (10) e (11). Portanto, podemos aplicar o **Teorema 13** a esta família, mostrando assim a existência e a unicidade da medida de probabilidade  $\mu$  em  $(X, \mathcal{B})$  satisfazendo (20). ■

**Proposição 24.** Se  $\mu$  é uma medida  $(p, P)$ -Markoviana, então  $\mu$  é  $T$ -invariante.

**Prova.** Como no caso Bernoulli, basta verificar a invariância sobre a semi-álgebra  $\mathcal{S}$  dos cilindros. Usando (14) e (18), temos

$$\begin{aligned}\mu(T^{-1}[a_0, a_1, \dots, a_n]_m) &= \mu([a_0, a_1, \dots, a_n]_{m+1}) \\ &= p_{a_0} P_{a_0 a_1} P_{a_1 a_2} \cdots P_{a_{n-1} a_n} \\ &= \mu([a_0, a_1, \dots, a_n]_m),\end{aligned}$$

onde a segunda igualdade usa a fórmula (18) aplicada ao cilindro  $[a_0, \dots, a_n]_{m-1}$ , e o fato de que esta fórmula não depende do índice inicial  $m - 1$ . ■

**Observação 25.** É conceitualmente importante observar que a  $T$ -invariância da medida  $\mu$  e a estacionariedade  $pP = p$  são manifestações equivalentes do mesmo fenômeno. De fato,

( $\Rightarrow$ ) Se  $\mu$  é  $T$ -invariante, então para todo  $j \in Y$ ,

$$\begin{aligned}p_j &= \mu(\{x_0 = j\}) = \mu(T^{-1}\{x_0 = j\}) \\ &= \mu(\{x_1 = j\}) \\ &= \sum_{i \in Y} \mu(\{x_0 = i, x_1 = j\}) \\ &= \sum_{i \in Y} p_i P_{ij} = (pP)_j,\end{aligned}$$

o que mostra  $pP = p$ .

( $\Leftarrow$ ) Se  $pP = p$ , a  $T$ -invariância de  $\mu$  sobre cilindros foi demonstrada na **Proposição 24**.

Em resumo: para a família de fórmulas  $(p, P)$ -Markovianas (18), ser uma medida de probabilidade bem definida, ser  $T$ -invariante, e ter  $p$  como autovetor à esquerda de  $P$  com autovalor 1, são três formulações da mesma propriedade estrutural.

O sistema  $(X, \mathcal{B}, \mu, T)$  assim construído é chamado *shift de Markov* com dados  $(p, P)$ .

**Observação 26.** Um caso particular importante da construção Markoviana é obtido quando tomamos  $P$  como uma matriz em que todas suas linhas são iguais e determinadas por vetor de probabilidade  $p$ , ou seja,  $P_{ij} = p_j$  para todos  $i, j \in Y$ . Neste caso:

- (a)  $P$  é claramente estocástica ( $\sum_j P_{ij} = \sum_j p_j = 1$ );
- (b)  $p$  é estacionária, pois  $(pP)_j = \sum_i p_i P_{ij} = p_j \sum_i p_i = p_j$ ;
- (c) A fórmula Markoviana (18) reduz-se a

$$p_{a_0} P_{a_0 a_1} \cdots P_{a_{n-1} a_n} = p_{a_0} p_{a_1} \cdots p_{a_n},$$

que é exatamente a expressão que determina a medida Bernoulli em (15).

Portanto, os shifts de Bernoulli são um caso particular dos shifts de Markov, correspondendo à situação de completa ausência de correlação entre coordenadas.

## 5. Condições de Ergodicidade para Shifts de Markov

**Lema 27.** Sejam  $Y \equiv \{0, 1, \dots, k-1\}$  e  $P \equiv (P_{ij})_{i,j \in Y}$  uma matriz estocástica possuindo um autovetor à esquerda  $p \in \mathbb{R}^k$ , satisfazendo  $pP = p$  e também  $p_i > 0$  e  $p_0 + \dots + p_{k-1} = 1$ . Então existe o seguinte limite, com respeito à topologia em  $M_{k \times k}(\mathbb{R})$  induzida pela norma  $A \mapsto \|A\|_{\text{op}} \equiv \sup\{\|Av\|_{\infty} : \|v\|_{\infty} = \max\{|v_0|, \dots, |v_{k-1}| \} = 1\}$ ,

$$Q \equiv \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} P^n. \quad (21)$$

Além do mais, a matriz  $Q$  também é uma matriz estocástica e satisfaz  $QP = PQ = Q$ . Se  $v \in \mathbb{R}^k$  é um autovetor de  $P$  associado ao autovalor 1, à direita ou à esquerda, então  $v$  também é autovetor de  $Q$  associado ao autovalor 1, do mesmo lado. Adicionalmente, a matriz  $Q$  é idempotente, ou seja,  $Q^2 = Q$ .

**Prova.** Vamos mostrar inicialmente que, individualmente, cada entrada da sequência de matrizes que aparece em (21) converge. Seja  $\mu$  a medida  $(p, P)$ -Markoviana fornecida pelo Teorema 23 e definida por meio da família  $\{\mu_{m,n}\}_{m \in \mathbb{Z}, n \in \mathbb{Z}_+}$  dada por

$$\mu_{m,n}(a_0, \dots, a_n) \equiv p_{a_0} P_{a_0 a_1} P_{a_1 a_2} \dots P_{a_{n-1} a_n}, \quad \forall a_0, \dots, a_n \in Y. \quad (22)$$

Denote por  $X \equiv Y^{\mathbb{Z}}$  e  $T : X \rightarrow X$  o shift para a esquerda. Para cada  $i \in Y$  fixado seja  $\chi_i : X \rightarrow \mathbb{R}$  a função indicadora do cilindro  $\{x \in X : x_0 = i\}$ . Já que  $\mu$  é  $T$ -invariante segue do Teorema Ergódico de Birkhoff que para cada  $j \in Y$  fixado temos

$$\frac{1}{N} \sum_{n=0}^{N-1} \chi_j(T^n(x)) \xrightarrow{N \rightarrow \infty} \chi_j^*(x) \quad \mu\text{-qtp.} \quad (23)$$

Multiplicando ambos lados da expressão acima por  $\chi_i(x)$  concluímos que

$$\frac{1}{N} \sum_{n=0}^{N-1} \chi_j(T^n(x)) \chi_i(x) \xrightarrow{N \rightarrow \infty} \chi_j^*(x) \chi_i(x) \quad \mu\text{-qtp.} \quad (24)$$

Já que  $T^{-n}(\{x \in X : x_0 = j\}) = \{x \in X : x_n = j\}$ , para cada  $n \in \mathbb{Z}_+$  temos

$$\chi_j(T^n(x)) = \chi_{T^{-n}(\{x \in X : x_0 = j\})}(x)$$

e conseqüentemente

$$\begin{aligned} \int_X \frac{1}{N} \sum_{n=0}^{N-1} \chi_j(T^n(x)) \chi_i(x) d\mu(x) &= \frac{1}{N} \sum_{n=0}^{N-1} \int_X \chi_j(T^n(x)) \chi_i(x) d\mu(x) \\ &= \frac{1}{N} \sum_{n=0}^{N-1} \mu(T^{-n}(\{x \in X : x_0 = j\}) \cap \{x \in X : x_0 = i\}) \\ &= \frac{1}{N} \sum_{n=0}^{N-1} \mu(\{x \in X : x_0 = i, x_n = j\}). \end{aligned} \quad (25)$$

O próximo passo é escrever a expressão que aparece no lado direito da igualdade acima em termos das potências da matriz estocástica  $P$ . A ideia é usar a identidade (22) da seguinte maneira:

$$\begin{aligned}
\mu(\{x \in X : x_0 = i, x_n = j\}) &= \sum_{a_1, \dots, a_{n-1} \in Y} \mu(\{x_0 = i, x_1 = a_1, \dots, x_{n-1} = a_{n-1}, x_n = j\}) \\
&= \sum_{a_1, \dots, a_{n-1} \in Y} p_i P_{ia_1} P_{a_1 a_2} \dots P_{a_{n-1} j} \\
&= p_i \sum_{a_1, \dots, a_{n-1} \in Y} P_{ia_1} P_{a_1 a_2} \dots P_{a_{n-1} j} \\
&= p_i P_{ij}^{(n)},
\end{aligned}$$

onde na última igualdade usamos a notação  $P_{ij}^{(n)} \equiv (P^n)_{ij}$ . Substituindo a expressão obtida acima em (25) ficamos com a seguinte igualdade

$$\int_X \frac{1}{N} \sum_{n=0}^{N-1} \chi_j(T^n(x)) \chi_i(x) d\mu(x) = \frac{1}{N} \sum_{n=0}^{N-1} p_i P_{ij}^{(n)}.$$

Já que a expressão do integrando que aparece acima é dominada pela função constante igual a um e que por (24) converge  $\mu$ -qtp, podemos aplicar o Teorema da Convergência Dominada para afirmar que valem as seguintes igualdades

$$\begin{aligned}
\int_X \chi_j^* \chi_i d\mu &= \lim_{N \rightarrow \infty} \int_X \frac{1}{N} \sum_{n=0}^{N-1} \chi_j(T^n(x)) \chi_i(x) d\mu(x) \\
&= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} p_i P_{ij}^{(n)} \\
&= p_i \cdot \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} P_{ij}^{(n)}.
\end{aligned}$$

Como estamos assumindo que  $p_i > 0$ , para todo  $i \in Y$ , segue da igualdade acima que para cada  $i \in Y$  fixado temos a seguinte identidade:

$$Q_{ij} \equiv \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} P_{ij}^{(n)} = \frac{1}{p_i} \int_X \chi_j^* \chi_i d\mu, \quad \forall i, j \in Y. \quad (26)$$

Vamos mostrar agora que a matriz  $Q$ , definida acima é uma matriz estocástica. Primeiro, observe que segue da igualdade (26) que cada entrada dessa matriz, isto é,  $Q_{ij}$  é um produto de um número positivo por uma integral de uma função não-negativa, e conseqüentemente  $Q_{ij} \geq 0$ , para todo  $i, j \in Y$ . Observando que  $\chi_0 + \chi_1 + \dots + \chi_{k-1} = 1$  e usando a igualdade (23), podemos concluir das propriedades elementares de limites que  $\chi_0^* + \chi_1^* + \dots + \chi_{k-1}^* = 1$

$\mu$ -qtp. Desta observação e da definição da medida  $\mu$ , segue

$$\begin{aligned} \sum_{j \in Y} Q_{ij} &= \sum_{j \in Y} \frac{1}{p_i} \int_X \chi_j^* \chi_i \, d\mu \\ &= \frac{1}{p_i} \int_X \left( \sum_{j \in Y} \chi_j^* \right) \chi_i \, d\mu \\ &= \frac{1}{p_i} \int_X \chi_i \, d\mu = \frac{1}{p_i} \mu(\{x_0 = i\}) = 1, \quad \forall i \in Y, \end{aligned}$$

o que encerra a prova de que  $Q$  é uma matriz estocástica.

Como estamos trabalhando em dimensão finita, podemos mostrar que a convergência da sequência acima também ocorre no sentido da norma de operadores em  $M_{k \times k}(\mathbb{R})$ . De fato,

$$\begin{aligned} \left\| Q - \frac{1}{N} \sum_{n=0}^{N-1} P^n \right\|_{\text{op}} &= \sup_{\substack{v \in \mathbb{R}^k \\ \|v\|_\infty = 1}} \left\| \left( Q - \frac{1}{N} \sum_{n=0}^{N-1} P^n \right) v \right\|_\infty \\ &= \sup_{\substack{v \in \mathbb{R}^k \\ \|v\|_\infty = 1}} \max_{i \in Y} \left| \sum_{j \in Y} \left( Q_{ij} - \frac{1}{N} \sum_{n=0}^{N-1} P_{ij}^{(n)} \right) v_j \right| \\ &\leq \sup_{\substack{v \in \mathbb{R}^k \\ \|v\|_\infty = 1}} \sum_{i \in Y} \left| \sum_{j \in Y} \left( Q_{ij} - \frac{1}{N} \sum_{n=0}^{N-1} P_{ij}^{(n)} \right) v_j \right| \\ &\leq \sum_{i \in Y} \sum_{j \in Y} \left| Q_{ij} - \frac{1}{N} \sum_{n=0}^{N-1} P_{ij}^{(n)} \right| \xrightarrow{N \rightarrow \infty} 0. \end{aligned} \tag{27}$$

Para cada  $N \in \mathbb{N}$ , defina a matriz

$$A_N = \frac{1}{N} \sum_{n=0}^{N-1} P^n.$$

Observe que vale a seguinte identidade algébrica

$$\begin{aligned} P A_N &= \frac{1}{N} \sum_{n=0}^{N-1} P^{n+1} = \frac{1}{N} \left( \sum_{n=0}^{N-1} P^n + P^N - I \right) \\ &= A_N + \frac{1}{N} (P^N - I). \end{aligned} \tag{28}$$

Como  $P$  é uma matriz estocástica, temos que

$$\|P\|_{\text{op}} \equiv \sup_{\substack{v \in \mathbb{R}^k \\ \|v\|_\infty = 1}} \|Pv\|_\infty = \sup_{\substack{v \in \mathbb{R}^k \\ \|v\|_\infty = 1}} \max_{i \in Y} \left| \sum_{j=0}^{k-1} P_{ij} v_j \right| \leq \sup_{\substack{v \in \mathbb{R}^k \\ \|v\|_\infty = 1}} \left( \|v\|_\infty \cdot \sup_{i \in Y} \left| \sum_{j=0}^{k-1} P_{ij} \right| \right) = 1.$$

Note que segue da desigualdade acima e das propriedades elementares de norma de operadores que

$$\|P^N\|_{\text{op}} \leq 1, \quad \forall N \in \mathbb{N}.$$

Usando esta estimativa e a desigualdade triangular para a norma de operadores, podemos verificar que o limite abaixo existe e é nulo

$$\lim_{N \rightarrow \infty} \left\| \frac{1}{N} (P^N - I) \right\|_{\text{op}} \leq \lim_{N \rightarrow \infty} \frac{1}{N} \left( \|P^N\|_{\text{op}} + \|I\|_{\text{op}} \right) \leq \lim_{N \rightarrow \infty} \frac{2}{N} = 0.$$

Usando a desigualdade acima em (28) juntamente com o resultado de convergência obtido em (27) concluímos pela continuidade do produto matricial que

$$PQ = P \left( \lim_{N \rightarrow \infty} A_N \right) = \lim_{N \rightarrow \infty} PA_N = \lim_{N \rightarrow \infty} \left( A_N + \frac{1}{N} (P^N - I) \right) = Q.$$

De maneira análoga, podemos verificar que  $QP = Q$ .

Para mostrar que a matriz  $Q$  é idempotente, basta observar que

$$\begin{aligned} Q^2 &= Q \cdot Q = Q \left( \lim_{N \rightarrow \infty} A_N \right) \\ &= \lim_{N \rightarrow \infty} (QA_N) && \text{(continuidade do produto matricial)} \\ &= \lim_{N \rightarrow \infty} \left( Q \cdot \frac{1}{N} \sum_{n=0}^{N-1} P^n \right) \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} QP^n. \end{aligned}$$

Usando o fato demonstrado acima  $QP = Q$ , segue, por indução em  $n$ , que  $QP^n = Q$ , para todo  $n \geq 0$ . Como consequência desta observação e da igualdade acima concluímos que

$$Q^2 = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} QP^n = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} Q = Q.$$

Para encerrar a prova do lema, resta mostrar que se  $v \in \mathbb{R}^k$  é um autovetor de  $P$  associado ao autovalor 1, à direita (respectivamente, à esquerda), então  $v$  também é um autovetor de  $Q$  associado ao autovalor 1, à direita (respectivamente, à esquerda).

Suponha primeiramente que  $Pv = v$ . Por indução em  $n$ , segue que  $P^n v = v$  para todo  $n \in \mathbb{Z}_+$ , e conseqüentemente

$$A_N v = \frac{1}{N} \sum_{n=0}^{N-1} P^n v = v, \quad \forall N \geq 1.$$

Passando ao limite quando  $N \rightarrow \infty$  e usando (21), concluímos que  $Qv = v$ . O caso em que  $v$  é autovetor à esquerda de  $P$  é inteiramente análogo: de  $vP = v$  segue  $vP^n = v$  para todo  $n \in \mathbb{Z}_+$ , logo  $vA_N = v$  para todo  $N \geq 1$ , e portanto  $vQ = v$ . ■

**Teorema 28.** Seja  $Y \equiv \{0, 1, \dots, k-1\}$  e considere o espaço produto  $X \equiv Y^{\mathbb{Z}}$  equipado com a topologia produto e a  $\sigma$ -álgebra de Borel  $\mathcal{B}$  associada. Sejam  $T : X \rightarrow X$  a aplicação de deslocamento para a esquerda (shift para a esquerda),  $p \in \mathbb{R}^k$  um vetor de probabilidade com todas as entradas estritamente positivas,  $p_j > 0$ , para todo  $j \in Y$  e  $P \equiv (P_{ij})_{i,j \in Y}$  uma matriz estocástica, isto é,  $P_{ij} \geq 0$  e  $P_{i0} + P_{i1} + \dots + P_{i(k-1)} = 1$ , para todo  $i \in Y$ . Seja  $Q$  a matriz obtida no **Lema 27**. Sob estas condições, as seguintes afirmações são equivalentes:

- (i)  $T$  é ergódica.
- (ii) Todas as linhas da matriz  $Q$  são iguais.
- (iii) Todas as entradas de  $Q$  são estritamente positivas.
- (iv)  $P$  é irredutível.
- (v)  $\dim(\{v \in \mathbb{R}^k : vP = v\}) = 1$ .

**Prova.** Ao longo de toda esta prova  $\mu$  denota a medida  $(p, P)$ -Markoviana definida em (22) por meio da expressão que aparece em (24).

(i)  $\Rightarrow$  (ii). Argumentando como na demonstração do **Lema 27**, segue do Teorema Ergódico de Birkhoff que, para cada  $i, j \in Y$ ,

$$\frac{1}{N} \sum_{n=0}^{N-1} \mu(\{x \in X : x_0 = i, x_n = j\}) \xrightarrow{N \rightarrow \infty} p_i Q_{ij}.$$

Já que  $T$  é ergódica, segue do **Corolário 5** do Teorema Ergódico de Birkhoff que temos, por outro lado, a seguinte igualdade

$$\frac{1}{N} \sum_{n=0}^{N-1} \mu(\{x \in X : x_0 = i, x_n = j\}) \xrightarrow{N \rightarrow \infty} \mu(\{x_0 = i\}) \mu(\{x_0 = j\}) = p_i p_j.$$

Pela unicidade do limite, segue das duas expressões acima que  $p_i Q_{ij} = p_i p_j$ . Como  $p_i > 0$  para todo  $i \in Y$ , concluímos que  $Q_{ij} = p_j$ , para todo  $i, j \in Y$ , e portanto todas as linhas de  $Q$  são iguais.

(ii)  $\Rightarrow$  (iii). Suponha que todas as linhas de  $Q$  sejam iguais, digamos  $Q_{ij} = q_j$  para todo  $i, j \in Y$ . Como estamos assumindo que  $pP = p$  segue do **Lema 27** que o autovetor, à esquerda,  $p$  é também um autovetor, à esquerda, associado ao autovalor 1 de  $Q$ , ou seja,  $pQ = p$ . Deste fato decorre que

$$p_j = \sum_{i \in Y} p_i Q_{ij} = \sum_{i \in Y} p_i q_j = q_j, \quad \forall j \in Y.$$

Assim,  $Q_{ij} = p_j > 0$  para todo  $i, j \in Y$ .

(iii)  $\Rightarrow$  (iv). Fixe  $i, j \in Y$ . Como

$$\frac{1}{N} \sum_{n=0}^{N-1} P_{ij}^{(n)} \xrightarrow{N \rightarrow \infty} Q_{ij} > 0,$$

existe  $n \in \mathbb{Z}_+$  tal que  $P_{ij}^{(n)} > 0$ . Logo,  $P$  é irredutível.

(iv)  $\Rightarrow$  (iii). Fixe  $i \in Y$  e defina

$$S_i \equiv \{j \in Y : Q_{ij} > 0\}.$$

Do **Lema 27** temos que  $Q$  é uma matriz estocástica e portanto podemos concluir  $S_i \neq \emptyset$ , pois caso contrário a soma da  $i$ -ésima linha de  $Q$  seria nula. Já que  $QP = Q$ , temos, para qualquer  $\ell \in Y$ , a seguinte desigualdade:

$$Q_{ij} = \sum_{m \in Y} Q_{im} P_{mj} \geq Q_{i\ell} P_{\ell j}, \quad \forall j \in Y.$$

Segue da desigualdade acima que se  $P_{\ell j} > 0$ , para algum  $\ell \in S_i$ , então  $j \in S_i$ , ou seja,  $Q_{ij} > 0$ . Desta observação, podemos concluir que

$$1 = \sum_{j \in Y} P_{\ell j} = \sum_{j \in S_i} P_{\ell j} + \sum_{j \notin S_i} P_{\ell j} = \sum_{j \in S_i} P_{\ell j}, \quad \forall \ell \in S_i$$

De fato, se existisse  $j_0 \notin S_i$  com  $P_{\ell j_0} > 0$ , para algum  $\ell \in S_i$ , então teríamos  $j_0 \in S_i$  pela observação anterior, uma contradição; logo  $P_{\ell j} = 0$  para todo  $j \notin S_i$  e a soma das entradas de  $P$  restrita a  $S_i$  é igual a 1.

Suponha, por absurdo, que  $S_i \neq Y$ . Como  $S_i \neq \emptyset$ , podemos afirmar que existe pelo menos um par de elementos  $\ell, r \in Y$  com  $\ell \in S_i$  e  $r \notin S_i$ . Como  $P$  é irredutível, existe  $n \in \mathbb{Z}_+$  tal que  $P_{\ell r}^{(n)} > 0$ . Contudo segue, por indução sobre  $n$ , que podemos generalizar a observação feita acima (se  $\ell \in S_i$  então  $P_{\ell j} = 0$ , para todo  $j \notin S_i$ ) e que esta generalização implica que  $P_{\ell r}^{(n)} = 0$ , o que é absurdo. Portanto  $S_i = Y$  e  $Q_{ij} > 0$  para todo  $j \in Y$ . Como  $i \in Y$  escolhido no início da prova é um ponto arbitrário em  $Y$ , concluimos que todas as entradas de  $Q$  são estritamente positivas.

**(iii)  $\Rightarrow$  (ii).** Fixe  $j \in Y$  e defina  $\bar{Q}_j \equiv \max_{i \in Y} Q_{ij}$ . Pelo **Lema 27** sabemos que  $Q^2 = Q$ . Suponha, por absurdo, que exista  $i_0 \in Y$  tal que  $Q_{i_0 j} < \bar{Q}_j$ . Então, para cada  $\ell \in Y$ ,

$$Q_{\ell j} = \sum_{i \in Y} Q_{\ell i} Q_{ij} < \sum_{i \in Y} Q_{\ell i} \bar{Q}_j = \bar{Q}_j,$$

onde a desigualdade estrita usa que  $Q_{\ell i_0} > 0$  (hipótese (iii)) e  $Q_{i_0 j} < \bar{Q}_j$ , e a última igualdade usa que  $Q$  é estocástica (também pelo **Lema 27**). Isto vale para todo  $\ell \in Y$ , em particular para aquele que realiza o máximo  $\bar{Q}_j$ , o que contradiz a definição de  $\bar{Q}_j$ . Logo,  $Q_{ij} = \bar{Q}_j$  para todo  $i \in Y$  e as linhas de  $Q$  são iguais.

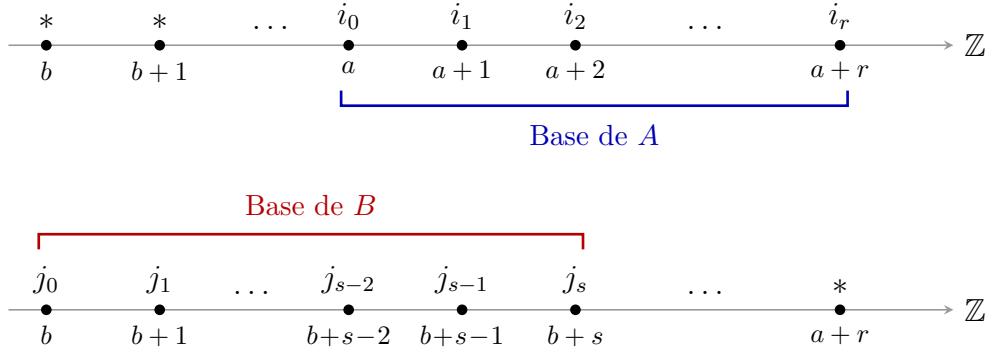
**(ii)  $\Rightarrow$  (i).** Pelo **Teorema 7**, para mostrar que  $T$  é ergódica basta verificar que, para quaisquer dois cilindros da forma

$$A \equiv \{x \in X : x_a = i_0, x_{a+1} = i_1, \dots, x_{a+r} = i_r\},$$

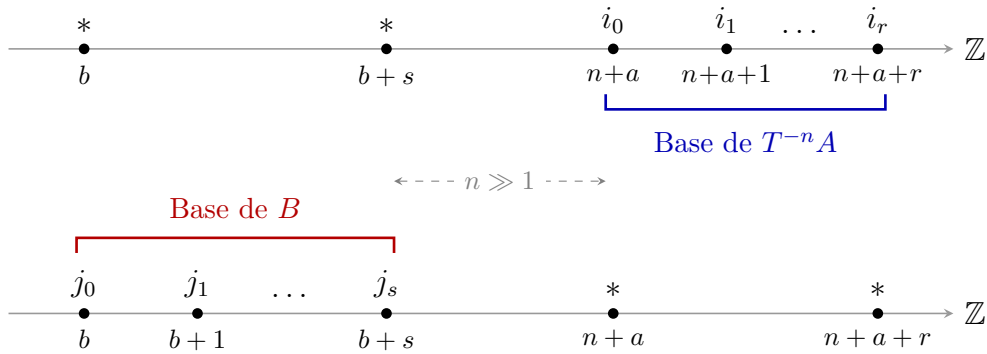
$$B \equiv \{x \in X : x_b = j_0, x_{b+1} = j_1, \dots, x_{b+s} = j_s\},$$

vale

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu(T^{-n} A \cap B) = \mu(A) \mu(B).$$



Para  $n \in \mathbb{N}$  tal que  $n + a > b + s$ , os conjuntos de índices das bases dos cilindros  $T^{-n}A$  e  $B$  não se interceptam.



Desta forma, já que  $\mu$  é uma medida  $(p, P)$ -Markoviana, temos

$$\mu(T^{-n}A \cap B) = p_{j_0} P_{j_0 j_1} \cdots P_{j_{s-1} j_s} P_{j_s i_0}^{(n+a-(b+s))} P_{i_0 i_1} \cdots P_{i_{r-1} i_r}. \quad (29)$$

Como estamos assumindo que todas as linhas da matriz  $Q$  são iguais, sabemos do argumento usado na prova de que (ii) $\Rightarrow$ (iii) que  $Q_{ij} = p_j$  para todo  $i, j \in Y$ . Além do mais,

$$\frac{1}{N} \sum_{n=0}^{N-1} P_{j_s i_0}^{(n)} \xrightarrow{N \rightarrow \infty} Q_{j_s i_0} = p_{i_0}.$$

Usando a igualdade acima, que a sequência  $P_{j_s i_0}^{(n)}$  converge no sentido de Cesàro para  $p_{i_0}$ , e que os demais fatores que aparecem em (29) não dependem de  $n$ , concluímos que

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu(T^{-n}A \cap B) &= p_{j_0} P_{j_0 j_1} \cdots P_{j_{s-1} j_s} p_{i_0} P_{i_0 i_1} \cdots P_{i_{r-1} i_r} \\ &= \mu(A) \mu(B). \end{aligned}$$

(ii)  $\Rightarrow$  (v). Assumindo a validade de (ii), já verificamos na prova de que (ii) $\Rightarrow$ (iii) que  $Q_{ij} = p_j$ , para todo  $i, j \in Y$ . Logo, os autovetores à esquerda de  $Q$  associados ao autovalor 1 são precisamente os múltiplos escalares de  $p \in \mathbb{R}^k$ . De fato, suponha que  $v \in \mathbb{R}^k$  seja um autovetor à esquerda de  $Q$ , associado ao autovalor 1, então

$$(v_0, \dots, v_{k-1}) \begin{pmatrix} p_0 & \cdots & p_{k-1} \\ \vdots & & \end{pmatrix} = (v_0, \dots, v_{k-1}) \implies p_j(v_0 + v_1 + \cdots + v_{k-1}) = v_j, \quad \forall j \in Y.$$

Pelo [Lema 27](#), todo autovetor à esquerda de  $P$  associado ao autovalor 1 também é autovetor à esquerda de  $Q$  associado ao mesmo autovalor. Neste caso, segue do argumento acima que este autovetor é necessariamente múltiplo de  $p$ . Portanto, podemos concluir que a multiplicidade geométrica de 1 como autovalor de  $P$  é exatamente 1, ou seja, 1 é autovalor simples de  $P$ .

(v)  $\Rightarrow$  (ii). Suponha que o subespaço  $\mathcal{V} \equiv \{v \in \mathbb{R}^k : vP = v\}$  seja unidimensional. Já que  $QP = Q$  ([Lema 27](#)),

$$\begin{pmatrix} Q_{00} & \cdots & Q_{0(k-1)} \\ \vdots & & \vdots \\ Q_{j0} & \cdots & Q_{j(k-1)} \\ \vdots & & \vdots \\ Q_{(k-1)0} & \cdots & Q_{(k-1)(k-1)} \end{pmatrix} \begin{pmatrix} P_{00} & \cdots & P_{0(k-1)} \\ \vdots & & \vdots \\ P_{j0} & \cdots & P_{j(k-1)} \\ \vdots & & \vdots \\ P_{(k-1)0} & \cdots & P_{(k-1)(k-1)} \end{pmatrix} = \begin{pmatrix} Q_{00} & \cdots & Q_{0(k-1)} \\ \vdots & & \vdots \\ Q_{j0} & \cdots & Q_{j(k-1)} \\ \vdots & & \vdots \\ Q_{(k-1)0} & \cdots & Q_{(k-1)(k-1)} \end{pmatrix}$$

cada linha de  $Q$  é um autovetor à esquerda de  $P$  associado ao autovalor 1.

Como estamos assumindo que o autoespaço associado ao autovalor 1 é unidimensional todas essas linhas são múltiplos escalares do autovetor  $p \in \mathbb{R}^k$  mencionado no enunciado do teorema. Portanto para cada  $j \in Y$  temos

$$(Q_{j0}, \dots, Q_{j(k-1)}) = \lambda_j p.$$

Como  $Q$  é uma matriz estocástica e  $p \in \mathbb{R}^k$  é um vetor de probabilidade, obtemos somando as coordenadas dos vetores que aparecem na igualdade acima

$$1 = \sum_{\ell \in Y} Q_{j\ell} = \lambda_j \sum_{\ell} p_{\ell} = \lambda_j, \quad \forall j \in Y.$$

O que determina univocamente as constantes de proporcionalidade. Mostrando que cada linha de  $Q$  é formada pelas entradas do vetor  $p \in \mathbb{R}^k$  e portanto todas as linhas de  $Q$  são iguais. ■

## Referências

- [1] P. Billingsley. *Probability and measure*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, anniversary edition, 2012. With a foreword by Steve Lalley and a brief biography of Billingsley by Steve Koppes.
- [2] L. Cioletti. O teorema ergódico de birkhoff e aplicações. 2026. MAT-UnB. Lecture Notes in Ergodic Theory.
- [3] L. Cioletti and L. Lucinger. O Teorema da Existência de Kolmogorov e o Movimento Browniano. <https://mat.unb.br/cioletti/ensino/kolmogorov-browniano.pdf>, 2025.
- [4] P. Halmos. *Lectures on ergodic theory*. Chelsea Publishing Co., New York, 1960.
- [5] S. Kakutani. Examples of ergodic measure preserving transformations which are weakly mixing but not strongly mixing. In *Recent advances in topological dynamics (Proc. Conf. Topological Dynamics, Yale Univ., New Haven, Conn., 1972; in honor of Gustav Arnold Hedlund)*, Lecture Notes in Math., Vol. 318, pages 143–149. Springer, Berlin-New York, 1973.

- [6] G. Maruyama. The harmonic analysis of stationary stochastic processes. *Mem. Fac. Sci. Kyūsyū Univ. A*, 4:45–106, 1949.
- [7] D. Stroock and S. Varadhan. *Multidimensional diffusion processes*, volume 233 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, Berlin-New York, 1979.
- [8] P. Walters. *An introduction to ergodic theory.*, volume 7 of *Graduate Texts in Mathematics*. Springer-Verlag, New York-Berlin., 1982.