Um Primeiro Curso sobre Teoria Ergódica com Aplicações

Publicações Matemáticas

Um Primeiro Curso sobre Teoria Ergódica com Aplicações

Krerley Oliveira
UFAL

impa

25º Colóquio Brasileiro de Matemática

Estrada Dona Castorina, 110
22460-320 Rio de Janeiro, RJ

Impresso no Brasil / Printed in Brazil

Capa: Noni Geiger / Sérgio R. Vaz

25º Colóquio Brasileiro de Matemática

- A Short Introduction to Numerical Analysis of Stochastic Differential Equations - Luis José Roman
- An Introduction to Gauge Theory and its Applications - Marcos Jardim
- Aplicações da Análise Combinatória à Mecânica Estatística - Domingos H. U. Marchetti
- Dynamics of Infinite-dimensional Groups and Ramsey-type Phenomena - Vladimir Pestov
- Elementos de Estatística Computacional usando Plataformas de Software Livre/Gratuito - Alejandro C. Frery e Francisco Cribari-Neto
- Espaços de Hardy no Disco Unitário - Gustavo Hoepfner e Jorge Hounie
- Fotografia 3D - Paulo Cezar Carvalho, Luiz Velho, Anselmo Antunes Montenegro, Adelailson Peixoto, Asla Sá e Esdras Soares
- Introdução à Teoria da Escolha - Luciano I. de Castro e José Heleno Faro
- Introdução à Dinâmica de Aplicações do Tipo Twist - Clodoaldo G. Ragazzo, Mário J. Dias Carneiro e Salvador Addas-Zanata
- Schubert Calculus: an Algebraic Introduction - Letterio Gatto
- Surface Subgroups and Subgroup Separability in 3-manifold Topology - Darren Long and Alan W. Reid
- Tópicos em Processos Estocásticos: Eventos Raros, Tempos Exponenciais e Metaestabilidade - Adilson Simonis e Cláudia Peixoto
- Topics in Inverse Problems - Johann Baumeister and Antonio Leitão
- **Um Primeiro Curso sobre Teoria Ergódica com Aplicações - Krerley Oliveira**
- Uma Introdução à Simetrização em Análise e Geometria - Renato H. L. Pedrosa

Distribuição:
IMPA
Estrada Dona Castorina, 110
22460-320 Rio de Janeiro, RJ
E-mail: ddic@impa.br - http://www.impa.br
ISBN: 85-244-0223-7

Agradecimentos

Este texto é um subconjunto de um livro escrito em colaboração com o Prof. Marcelo Viana, o qual se torna imediatamente colaborador direto deste texto. Agradeço a ele por seu incentivo na escrita, pelas correções e as várias sugestões. Agradeço também aos colegas professores da UFAL, Adán Corcho e Hilário Alencar, que dividem comigo as dificuldades e alegrias do dia-a-dia, por seu apoio e pelo prazer de um ambiente científico agradável. Agradeço também aos estudantes Joao Gouveia, Vitor Saraiva, Ricardo Andrade, Maria João e Márcio Batista, que contibuiram com melhoria e clareza do texto com leituras e sugestões.

Agradeço à Fundação de Amparo à Pesquisa em Alagoas e ao CNPq pelo suporte financeiro e a Associação do Instituto Nacional de Matemática Pura e Aplicada e seus funcionários, onde parte deste trabalho foi escrito, pelo ambiente científico de excelente qualidade.

Finalmente agradeço à minha família, Eliséte, Batista, Kathy e Beto, que me dão o suporte necessário para o dia-a-dia e minha dileta esposa Marcela.

Krerley

Prefácio

Ao longo dos últimos anos diversas aplicações da Teoria Ergódica têm sido feitas para resolver problemas em áreas das mais diversas, desde problemas em Teoria dos Números e Topologia à aplicações em Geometria Diferencial e Probabilidade. Numa outra direção, a riqueza da motivação geralmente oriunda da Física, torna a área promissora e transportar resultados e heurística obtidos através de experimentos e simulações de problemas físicos para uma linguagem matemática precisa é uma tarefa árdua e que pode gerar excelentes resultados do ponto de vista matemático. Apesar da importancia e do apelo de tal disciplina, os estudantes a nível de graduação e mestrado não dispõem de muitos textos e cursos elementares sobre o assunto com intenção de promover o primeiro contato com a disciplina. Em geral, somente em cursos de doutorado o aluno pode tomar conhecimento dos mecanismos da área.

Este livro foi escrito com o objetivo de ser uma introdução à Teoria Ergódica para leitores com conhecimentos equivalentes aos de um estudante do último ano de graduação em matemática ou início de mestrado. Todo o texto foi focalizado em dois resultados principais: o Teorema de Recorrência e o Teorema Ergódico.

Para tornar todo o texto auto-suficiente até mesmo para leitores menos experientes, introduzimos um capítulo tratando dos principais resultados de Teoria da Medida utilizados. O leitor mais experiente pode dispensar a leitura deste capítulo, recorrendo a ele eventualmente quando lhe for necessário.

Ao longo do texto, obteremos várias aplicações do Teorema de Recorrência e do Teorema Ergódico a alguns exemplos concretos. Em particular, serão estudados os algoritmos de expansão decimal de um número e expansão em frações contínuas. Veremos também os importantes exemplos de fluxos que preservam volume e fluxos Hamiltonianos. O Capítulo 7 está especialmente dedicado ao estudo de duas aplicações interessantes e profundas em Teoria dos Números, que tratam da existência de progressões aritméticas de comprimento arbitrário num conjunto dado e da distribuição de valores de polinômios.

Krerley Oliveira [1]

[1]Departamento de Matemática, Universidade Federal de Alagoas, Campus A. C. Simões s/n, 57072-090 Maceió, Brasil. krerley@mat.ufal.br.

Capítulo 1

Elementos de Teoria da Medida

Neste capítulo inicial recordamos algumas noções e resultados básicos da Teoria da Medida que são úteis para o que segue. As demonstrações podem ser encontradas nos livros de Castro [Cas04], Fernandez [Fer02] ou Rudin [Rud87].

1.1 Espaços mensuráveis

Começamos por introduzir as noções de álgebra e σ-álgebra de subconjuntos. Em seguida definimos espaços mensuráveis e apresentamos uma técnica de construção de σ-álgebras. Seja M um conjunto.

Definição 1.1. Uma *álgebra* de subconjuntos de M é uma família $\mathcal{B}$ de subconjuntos que contém M e é fechada para as operações elementares de conjuntos:

- $A \in \mathcal{B}$ implica $A^c = M \setminus A \in \mathcal{B}$
- $A \in \mathcal{B}$ e $B \in \mathcal{B}$ implica $A \cup B \in \mathcal{B}$.

Então $A \cap B = (A^c \cup B^c)^c$ e $A \setminus B = A \cap B^c$ também estão em $\mathcal{B}$, quaisquer que sejam $A, B \in \mathcal{B}$. Além disso, por associatividade, a união e a intersecção de qualquer número finito de elementos de $\mathcal{B}$ também estão em $\mathcal{B}$.

Definição 1.2. Uma álgebra diz-se uma *σ-álgebra* de subconjuntos de M se também for fechada para uniões enumeráveis:

- $A_j \in \mathcal{B}$ para $j = 1, 2, \ldots, n, \ldots$ implica $\bigcup_{j=1}^{\infty} A_j \in \mathcal{B}$.

Observação. $\mathcal{B}$ também é fechada para intersecções enumeráveis: se $A_j \in \mathcal{B}$ para $j = 1, 2, \ldots, n, \ldots$ então $\bigcap_{j=1}^{\infty} A_j = \big(\bigcup_{j=1}^{\infty} A_j^c\big)^c \in \mathcal{B}$.

Definição 1.3. Um *espaço mensurável* é uma dupla $(M, \mathcal{B})$ onde M é um conjunto e $\mathcal{B}$ é uma σ-álgebra de subconjuntos de M. Os elementos de $\mathcal{B}$ são chamados *conjuntos mensuráveis*.

Em seguida apresentamos alguns exemplos de σ-álgebras remetendo para os exercícios o estudo de outros exemplos.

Exemplo 1.4. Seja M um conjunto qualquer.

1. Denotemos por 2^M a família de todos os subconjuntos de M. Então $\mathcal{B} = 2^M$ é claramente uma σ-álgebra.

2. $\mathcal{B} = \{\emptyset, M\}$ é também uma σ-álgebra.

Note que se $\mathcal{B}$ é uma álgebra de um conjunto M então $\{\emptyset, M\} \subset \mathcal{B} \subset 2^M$. Portanto $\{\emptyset, M\}$ é a menor álgebra e 2^M é a maior álgebra de um conjunto M. Considere uma família não-vazia $\{\mathcal{B}_i : i \in \mathcal{I}\}$ qualquer de σ-álgebras ($\mathcal{I}$ é um conjunto qualquer, que serve apenas para indexar os elementos da família). Então a intersecção

$$\mathcal{B} = \bigcap_{i \in \mathcal{I}} \mathcal{B}_i$$

é também uma σ-álgebra (veja o Exercício 1.1). Agora, dado um conjunto qualquer $\mathcal{E}$ de subconjuntos de M, podemos aplicar esta idéia à família de todas as σ-álgebras que contém $\mathcal{E}$. Note que esta família é não vazia, uma vez que contém a σ-álgebra 2^M, pelo menos. De acordo com a observação anterior, a intersecção de todas estas

σ-álgebras é também uma σ-álgebra, e é claro que contém $\mathcal{E}$. Além disso, do modo que é construída, ela está contida em todas as σ-álgebras que contêm $\mathcal{E}$. Portanto é a menor σ-álgebra que contém $\mathcal{E}$.

Definição 1.5. A *σ-álgebra gerada* por uma família $\mathcal{E}$ de subconjuntos de M é a menor σ-álgebra que contém a família $\mathcal{E}$.

No caso em que M vem munido da estrutura de espaço topológico, há uma escolha natural para $\mathcal{E}$, nomeadamente, o conjunto dos subconjuntos abertos. Isto nos conduz à noção de *σ-álgebra de Borel*.

Definição 1.6. Seja (M, τ) um espaço topológico, isto é, M um conjunto e τ a família dos subconjuntos abertos de M. Então a *σ-álgebra de Borel* de M é a σ-álgebra gerada por τ, ou seja, a menor σ-álgebra que contém todos os subconjuntos abertos.

1.2 Espaços de medida

Agora introduzimos o conceito de medida e analisamos algumas das suas propriedades fundamentais. Em seguida apresentamos alguns resultados sobre construção de medidas. Finalmente, analisamos duas importantes classes de medidas: medidas de Lebesgue em espaços euclideanos e medidas produto em espaço de sequências.

Definição 1.7. Uma *medida* num espaço mensurável $(M, \mathcal{B})$ é uma função $\mu : \mathcal{B} \to [0, +\infty]$ que satisfaz:

1. $\mu(\emptyset) = 0$;
2. $\mu(\bigcup_{j=1}^{\infty} A_j) = \sum_{j=1}^{\infty} \mu(A_j)$ para quaisquer $A_j \in \mathcal{B}$ disjuntos dois-a-dois.

A tripla $(M, \mathcal{B}, \mu)$ é chamada *espaço de medida*. Quando $\mu(M) = 1$ dizemos que μ é uma medida de *probabilidade* e $(M, \mathcal{B}, \mu)$ é um espaço de probabilidade.

A segunda propriedade na definição de medida é chamada a *σ-aditividade*. Dizemos que uma função $\mu : \mathcal{B} \to [0, +\infty]$ é *finitamente aditiva* se:

$$\mu(\bigcup_{j=1}^{N} A_j) = \sum_{j=1}^{N} \mu(A_j)$$

para qualquer família finita $A_1, \dots, A_N \in \mathcal{B}$ de subconjuntos disjuntos dois-a-dois. Note que toda medida é, automaticamente, finitamente aditiva.

Exemplo 1.8. Seja M um conjunto e consideremos a σ-álgebra $\mathcal{B} = 2^M$. Dado qualquer $p \in M$, consideremos a função $\delta_p : 2^M \to [0, +\infty]$ definida por:

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

Temos que δ_p é uma medida, que é usualmente designada por *delta de Dirac* no ponto p.

Em seguida apresentamos um resultado muito útil na construção de medidas.

Teorema 1.9 (Extensão). *Seja $\mathcal{B}_0$ uma álgebra de subconjuntos de M e seja $\mu_0 : \mathcal{B}_0 \to [0, +\infty]$ uma função finita, finitamente aditiva. Então existe uma única função finita, finitamente aditiva $\mu : \mathcal{B} \to [0, +\infty]$ que é uma extensão de μ_0 (isto é, μ restrita a $\mathcal{B}_0$ coincide com μ_0) à σ-álgebra $\mathcal{B}$ gerada por $\mathcal{B}_0$. Se μ_0 é σ-aditiva então μ também o é.*

Em geral, ao aplicar este resultado o mais difícil é verificar a σ-aditividade. O critério mais usado para esse efeito é expresso no seguinte resultado. A sua demonstração é proposta como Exercício 1.7.

Teorema 1.10 (σ-aditividade). *Seja $\mathcal{B}_0$ uma álgebra e seja $\mu_0 : \mathcal{B}_0 \to [0, +\infty]$ uma função finitamente aditiva com $\mu_0(M) = 1$. Suponha que*

$$\lim_{n \to \infty} \mu_0\Big(\bigcap_{j=1}^{n} A_j\Big) = 0 \tag{1.1}$$

para toda a sequência $A_1 \supset \cdots \supset A_j \supset \cdots$ de conjuntos mensuráveis tal que $\bigcap_{j=1}^{\infty} A_j = \emptyset$. Então μ_0 é σ-aditiva.

O resultado seguinte nos diz que todo o elemento B da σ-álgebra gerada por uma álgebra é aproximado por algum elemento B_0 da álgebra, no sentido em que a medida da diferença simétrica $B \Delta B_0 = B \setminus B_0 \cup B_0 \setminus B$ é pequena.

Teorema 1.11 (Aproximação). *Seja $(M, \mathcal{B}, \mu)$ um espaço de probabilidade e seja $\mathcal{B}_0$ uma álgebra que gera a σ-álgebra $\mathcal{B}$. Então para todo o $\varepsilon > 0$ e todo $B \in \mathcal{B}$ existe $B_0 \in \mathcal{B}_0$ tal que $\mu(B \Delta B_0) < \varepsilon$.*

1.2.1 Medida de Lebesgue

A medida de Lebesgue corresponde ao que entendemos por volume de subconjuntos de $\mathbb{R}^d$. Para construí-la, recorremos ao Teorema de Extensão 1.9. Consideremos $M = [0,1]$ e seja $\mathcal{B}_0$ a família de todos os subconjuntos da forma $B = I_1 \cup \cdots \cup I_N$ onde $I_1, \ldots, I_N$ são intervalos disjuntos dois-a-dois. É fácil ver que $\mathcal{B}_0$ é uma álgebra de subconjuntos de M. Além disso, temos uma função $\mu_0 : \mathcal{B}_0 \to [0,1]$ definida nesta álgebra por

$$\mu_0\big(I_1 \cup \cdots \cup I_N\big) = |I_1| + \cdots + |I_N|,$$

onde $|I_j|$ representa o comprimento de I_j. Note que $\mu_0(M) = 1$. Além disso, a σ-álgebra gerada por $\mathcal{B}_0$ coincide com a σ-álgebra de Borel de M, já que todo aberto pode ser escrito como união enumerável de intervalos disjuntos dois-a-dois. Pelo Teorema 1.9, existe uma única probabilidade μ definida na σ-álgebra de $[0,1]$ que é uma extensão de μ_0 à σ-álgebra $\mathcal{B}$ gerada por $\mathcal{B}_0$. Chamamos μ de *medida de Lebesgue* em $[0,1]$. Mais geralmente, definimos medida de Lebesgue μ no cubo $M = [0,1]^d$ de qualquer dimensão $d \geq 1$ da seguinte maneira: chamamos retângulo em M qualquer subconjunto da forma $R = I_1 \times \cdots \times I_d$ onde os I_j são intervalos, e definimos

$$\mu_0(R) = |I_1| \times \cdots \times |I_d|.$$

Em seguida, consideramos a álgebra $\mathcal{B}_0$ dos subconjuntos de $[0,1]^d$ da forma $B = R_1 \cup \cdots \cup R_N$, onde $R_1, \ldots, R_N$ são retângulos disjuntos dois-a-dois, e definimos

$$\mu_0(B) = \mu_0(R_1) + \cdots \mu_0(R_N)$$

para todo B nessa álgebra. A medida de Lebesgue em $M = [0,1]^d$ é a extensão de μ_0 à σ-álgebra gerada por $\mathcal{B}_0$, que coincide com a σ-álgebra de Borel de M. Finalmente, definimos a medida de Lebesgue num espaço euclidiano $\mathbb{R}^d$ decompondo o espaço em cubos de lado unitário

$$\mathbb{R}^d = \bigcup_{m_1 \in \mathbb{Z}} \cdots \bigcup_{m_d \in \mathbb{Z}} [m_1, m_1 + 1) \times \cdots \times [m_d, m_d + 1)$$

e definindo, para cada subconjunto mensurável E,

$$\mu(E) = \sum_{m_1 \in \mathbb{Z}} \cdots \sum_{m_d \in \mathbb{Z}} \mu\big(E \cap [m_1, m_1 + 1) \times \cdots \times [m_d, m_d + 1)\big).$$

Exemplo 1.12 (Medida de Volume em S^1). Considere a aplicação sobrejetora $\gamma : [0, 1] \to S^1$ definida por:

$$\gamma(t) = e^{2\pi i t}.$$

A *medida de Lebesgue* em S^1 é a medida μ definida por $\mu(A) = m(\gamma^{-1}(A))$. Observe que com esta definição, a medida de A é igual a medida de $R_\alpha(A)$, onde $R_\alpha : S^1 \to S^1$ denota a rotação de ângulo α. Na verdade, módulo multiplicação por um número positivo, μ é a única medida que satisfaz essa condição para todo α.

Exemplo 1.13. Seja $\phi : [0, 1] \to \mathbb{R}$ uma função contínua e positiva. Defina a medida μ_ϕ num intervalo $[a, b]$ por:

$$\mu_\phi([a, b]) = \int_a^b \phi(x)\, dx.$$

Observe que μ_ϕ é aditiva e com o auxílio dos Teoremas 1.10 e 1.9 podemos estender μ_ϕ para toda σ-álgebra dos Borelianos de $[0, 1]$. A medida μ_ϕ tem a seguinte propriedade especial: se um conjunto $A \subset [0, 1]$ tem medida de Lebesgue 0 então $\mu_\phi(A) = 0$. Essa propriedade nos diz que μ_ϕ é *absolutamente contínua* com respeito à medida de Lebesgue. A densidade de μ_ϕ em relação a m é igual a ϕ. Estudaremos tais medidas com mais detalhes na Secção 1.3.2.

Exemplo 1.14. Vamos agora exibir uma medida que, apesar de ser positiva em qualquer aberto, não é absolutamente contínua com respeito a medida de Lebesgue. Para isso, considere uma enumeração $\{r_1, r_2, \dots\}$ do conjunto $\mathbb{Q}$ dos racionais. Defina μ por:

$$\mu(A) = \sum_{r_i \in A} \frac{1}{2^i}.$$

Observe que a medida de qualquer aberto da reta é positiva, pois necessariamente A contém algum i, e, apesar disso, $\mu(\mathbb{Q}) = 1$. Em particular, μ não é absolutamente contínua com respeito a medida de Lebesgue.

O exemplo anterior nos motiva a definir o *suporte* de uma medida:

Definição 1.15. Seja $(M, \mathcal{B}, \mu)$ um espaço de medida e M um espaço topológico. O suporte da medida μ é o fecho do conjunto de pontos $x \in M$ tais que para qualquer vizinhança aberta V_x contendo x, temos que $\mu(V_x) > 0$.

Fica como exercício para o leitor mostrar que o suporte de uma medida é sempre um conjunto fechado (1.17).

1.2.2 Medida produto no espaço das sequências

Consideremos os espaços de probabilidade $(M_i, \mathcal{B}_i, \mu_i)$, com $i \in \mathbb{Z}$. Vamos construir uma probabilidade μ no conjunto

$$M = \prod_{i=-\infty}^{\infty} M_i$$

das sequências bilaterais $(x_i)_{i=-\infty}^{\infty}$ com $x_i \in M_i$ para cada i. Mais precisamente, a medida μ será definida na *σ-álgebra produto* $\mathcal{B}$ das σ-álgebras $\mathcal{B}_i$, que é caracterizada do seguinte modo: dados inteiros $m \leq n$ e conjuntos $A_j \in \mathcal{B}_j$ para $m \leq j \leq n$, consideremos

$$[m; A_m, \ldots, A_n] = \{(x_i)_{i \in \mathbb{Z}} : x_j \in A_j \text{ para } m \leq j \leq n\}.$$

Estes subconjuntos de M são chamados *cilindros*. A família $\mathcal{B}_0$ das uniões finitas de cilindros disjuntos dois-a-dois é uma álgebra. Por definição, a σ-álgebra produto $\mathcal{B}$ é a σ-álgebra gerada por $\mathcal{B}_0$. Para construir a medida μ procedemos do seguinte modo: primeiramente, consideramos a aplicação τ definida na família dos cilindros por

$$\tau([m; A_m, \ldots, A_n]) = \prod_{j=m}^{n} \mu_j(A_j).$$

Em seguida estendemos τ à álgebra $\mathcal{B}_0$, estipulando que a imagem de qualquer união finita de cilindros disjuntos dois-a-dois é igual à soma das imagens dos cilindros. Esta extensão está bem definida e é finitamente aditiva. Então, recorrendo aos Teoremas 1.10 e 1.9, obtemos uma medida de probabilidade μ em $(M, \mathcal{B})$ que estende τ.

Definição 1.16. O espaço de probabilidade $(M, \mathcal{B}, \mu)$ construído acima é designado *produto direto* dos espaços $(M_i, \mathcal{B}_i, \mu_i)$.

Existe um caso particular importante, que corresponde à situação onde os espaços $(M_i, \mathcal{B}_i, \mu_i)$ são todos iguais a um dado $(X, \mathcal{C}, \nu)$, em que $X = \{1, \ldots, d\}$ é um conjunto finito e $\mathcal{C} = 2^X$ é a σ-álgebra de todos os subconjuntos de X. Neste caso basta considerar apenas cilindros elementares, isto tais que cada A_j consiste de um único ponto de X. De fato, todo cilindro é uma união finita disjunta de tais cilindros elementares. Obtemos então subconjuntos de M da forma

$$[m; a_m, \ldots, a_n] = \{(x_i)_{i=-\infty}^{\infty} \in M : x_m = a_m, \ldots, x_n = a_n\}$$

onde $a_j \in \{1, \ldots, d\}$. A medida μ é designada *medida de Bernoulli* definida por ν e é caracterizada por $\mu([m; a_m, \ldots, a_n]) = \nu(\{a_m\}) \cdots \nu(\{a_n\})$.

1.3 Integração em espaços de medida

Nesta seção definimos a noção de integral de uma função em relação a uma medida e apresentamos teoremas fundamentais da Teoria da Medida. Para tanto, introduziremos algumas classes de funções. Ao longo desta seção $(M, \mathcal{B}, \mu)$ será sempre um espaço de medida.

Definição 1.17. Seja $\mathcal{B}(\mathbb{R})$ a σ-álgebra de Borel de $\mathbb{R}$. Uma função $f : M \to \mathbb{R}$ diz-se *mensurável* se $f^{-1}(D) \in \mathcal{B}$ para todo $D \in \mathcal{B}(\mathbb{R})$.

O espaço das funções mensuráveis possui diversas propriedades muito úteis. Vamos enunciá-las como proposição:

Proposição 1.18. *Sejam f_1, f_2 funções mensuráveis e $c_1, c_2 \in \mathbb{R}$. Então também são mensuráveis as seguintes funções:*

1. *então* $(c_1 f_1 + c_2 f_2)(x) = c_1 f_1(x) + c_2 f_2(x)$
2. $(f_1 \cdot f_2)(x) = f_1(x) \cdot f_2(x)$

3. $\max\{f_1, f_2\}(x) = \max\{f_1(x), f_2(x)\}$

Dizemos que uma função $s : M \to \mathbb{R}$ é *simples* se existem constantes $\alpha_1, \ldots, \alpha_k \in \mathbb{R}$ e conjuntos $A_1, \ldots, A_k \in \mathcal{B}$ disjuntos dois-a-dois tais que

$$s = \sum_{j=1}^{k} \alpha_k \mathcal{X}_{A_k},$$

onde $\mathcal{X}_A$ é a função característica do conjunto A, isto é, $\mathcal{X}_A(x)$ é igual a 1 se $x \in A$ e zero caso contrário. Introduzimos agora a noção de integral. Para tal começamos por definir integral de uma função simples.

Definição 1.19. Seja s uma função simples da forma acima. Então a *integral* de s em relação a μ é dado por:

$$\int s \; d\mu = \sum_{j=1}^{k} \alpha_k \mu(A_k).$$

É fácil verificar que esta definição é coerente: se duas combinações lineares de funções características definem uma mesma função simples, os valores das integrais obtidos a partir das duas combinações coincidem. O próximo passo é definir integral de uma função mensurável qualquer. Para isso, trataremos primeiro do caso da função ser não-negativa. Necessitamos do seguinte resultado, que nos diz que qualquer função mensurável é o limite de uma sequência de funções simples mensuráveis:

Teorema 1.20. *Seja $f : M \to [-\infty, \infty]$ uma função mensurável. Então existe uma sequência s_1, s_2, ... de funções simples mensuráveis tal que*

$$\lim_{k\to\infty} s_k(x) = f(x) \text{ para todo o } x \in M.$$

Se $f \geq 0$ então a sequência pode ser escolhida de modo que $0 \leq s_1 \leq s_2 \leq \cdots$.

A demonstração deste teorema é proposta como Exercício 1.16. Ele torna possível a seguinte

Definição 1.21. Seja $f : M \to [0, \infty]$ uma função mensurável não-negativa. Então

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

onde $s_1 \leq s_2 \leq \dots$ é uma sequência de funções simples crescentes para f, isto é, $\lim_{n\to\infty} s_n(x) = f(x)$ para todo $x \in M$.

É fácil verificar que o valor da integral não depende da escolha da sequência de funções simples, e portanto esta definição é coerente. Para estender a definição de integral a quaisquer funções mensuráveis, observemos que dada uma função $f : M \to \mathbb{R}$ sempre podemos escrever $f = f^+ - f^-$ onde $f^+(x) = \max\{f(x), 0\}$ e $f^-(x) = \max\{-f(x), 0\}$ são não-negativas. Mostra-se também que f^+ e f^- são mensuráveis se e só se, f é mensurável.

Definição 1.22. Seja $f : M \to [0, \infty]$ uma função mensurável. Então

$$\int f\, d\mu = \int f^+\, d\mu - \int f^-\, d\mu,$$

desde que alguma das integrais do lado direito seja finita.

Definição 1.23. Dizemos que uma função é *integrável* se for mensurável e tiver integral finita. Denotamos o conjunto das funções integráveis por $L^1(M, \mathcal{B}, \mu)$ ou, mais simplesmente, por $L^1(M, \mu)$.

Dada uma função mensurável $f : M \to \mathbb{R}$ e um conjunto mensurável E definimos a *integral de f sobre E* por

$$\int_E f d\mu = \int f \mathcal{X}_E\, d\mu,$$

onde $\mathcal{X}_E$ é a função característica do conjunto E.

Exemplo 1.24. Sejam $x_1, \dots, x_m \in M$ e $p_1, \dots, p_m > 0$ com $p_1 + \dots + p_m = 1$. Consideremos a medida de probabilidade $\mu : 2^M \to [0, 1]$ dada por:

$$\mu(A) = \sum_{x_i \in A} p_i.$$

Notemos que $\mu = \sum_{i=1}^{m} p_i \delta_{x_i}$, onde δ_{x_i} é a medida delta de Dirac em x_i. Neste caso temos que se f é uma função integrável então

$$\int f\, d\mu = \sum_{i=1}^{m} f(x_i) p_i.$$

1.3.1 Teorema de derivação de Lebesgue

Comecemos por introduzir a noção de "quase em toda a parte" em relação a uma medida. Dizemos que uma propriedade é válida *em μ-quase todo ponto* se é válida em todo o M exceto, possivelmente, num conjunto de medida nula. Por exemplo, dizemos que duas funções f, g são iguais em μ-quase todo ponto se existe um conjunto mensurável N com $\mu(N) = 0$ tal que $f(x) = g(x)$ para todo $x \in M \setminus N$.

Teorema 1.25 (Derivação de Lebesgue). *Seja $M = \mathbb{R}^d$, $\mathcal{B}$ a σ-álgebra de Borel e μ a medida de Lebesgue em $\mathbb{R}^d$. Seja $f : M \to \mathbb{R}$ uma função localmente integrável, isto é, tal que $f\mathcal{X}_K$ é integrável para todo compacto $K \subset \mathbb{R}^d$. Então*

$$\lim_{r\to 0} \frac{1}{\mu(B(x,r))} \int_{B(x,r)} |f(y) - f(x)| d\mu = 0.$$

em μ-quase todo ponto $x \in \mathbb{R}^d$. Em particular, em μ-quase todo o ponto $x \in \mathbb{R}^d$ tem-se

$$\lim_{r\to 0} \frac{1}{\mu(B(x,r))} \int_{B(x,r)} f(y) d\mu = f(x).$$

Dado um subconjunto mensurável A de $\mathbb{R}^d$, dizemos que um ponto $a \in A$ é um *ponto de densidade* de A se este conjunto preenche a maior parte de qualquer pequena vizinhança de a, i.e,

$$\lim_{\varepsilon\to 0} \frac{\mu(B(a,\varepsilon) \cap A)}{\mu(B(a,\varepsilon)} = 1. \tag{1.2}$$

O próximo resultado é uma consequência direta do teorema de derivação de Lebesgue. No Exercício 1.13 sugerimos uma demonstração.

Teorema 1.26. *Seja A um subconjunto mensurável de $\mathbb{R}^d$ com medida de Lebesgue $\mu(A)$ maior que zero. Então μ-quase todo ponto $a \in A$ é ponto de densidade de A.*

Muitos dos resultados envolvendo funções vão se apoiar no chamado "Teorema da Convergência Dominada", que garante que se uma sequência de funções convergente é menor que uma função integrável, então o limite das suas integrais converge e podemos tomar o limite sob o sinal da integral. Mais precisamente:

Teorema 1.27 (Teorema da Convergência Dominada). *Consideremos $f_n : M \to \mathbb{R}$ uma sequência de funções mensuráveis e g uma função integrável tal que $|f_n(x)| \leq |g(x)|$ para μ-quase todo x em M. Se para μ-quase todo $x \in M$ a sequência $f_n(x)$ converge para o valor $f(x)$, então a função f é integrável e vale:*

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

1.3.2 Teorema de Radon-Nikodym

Sejam μ e ν duas medidas num espaço mensurável $(M, \mathcal{B})$. Dizemos que ν é *absolutamente contínua* em relação a μ se $\mu(E) = 0$ implica $\nu(E) = 0$, qualquer que seja o conjunto mensurável. Nesse caso escrevemos $\nu \ll \mu$. O Teorema de Radon–Nikodym afirma que nesse caso a medida ν pode ser vista como o produto de μ por alguma função mensurável, que é chamada *densidade* ou *derivada de Radon-Nikodym* de ν relativamente a μ.

Teorema 1.28 (Radon-Nikodym). *Se μ e ν são medidas finitas tais que $\nu \ll \mu$ então existe uma função mensurável $\rho : M \to [0, +\infty]$ tal que $\nu = \rho\mu$, ou seja, tal que*

$$\nu(E) = \int_E \rho \, d\mu \quad \textit{para todo o conjunto mensurável } E \subset M.$$

Além disso, essa função é essencialmente única: duas quaisquer coincidem μ quase em toda a parte.

1.4 Exercícios

1.1. Seja M um conjunto e, para cada i pertencente a um conjunto de índices $\mathcal{I}$, seja $\mathcal{B}_i$ uma σ-álgebra de subconjuntos de M. Mostre que

$$\mathcal{B} = \bigcap_{i \in \mathcal{I}} \mathcal{B}_i$$

é uma σ-álgebra.

1.2. Seja M um conjunto e considere a família de conjuntos

$$\mathcal{B}_0 = \{A \subset M : A \text{ é finito ou } A^c \text{ é finito}\}.$$

Mostre que $\mathcal{B}_0$ é uma álgebra. Além disso, $\mathcal{B}_0$ é uma σ-álgebra se e somente se o conjunto M é finito.

1.3. Seja M um conjunto e considere a seguinte família de conjuntos

$$\mathcal{B}_1 = \{A \subset M : A \text{ é finito ou enumerável ou } A^c \text{ é finito ou enumerável}\}.$$

Mostre que $\mathcal{B}_1$ é uma σ-álgebra. De fato, $\mathcal{B}_1$ é a σ-álgebra gerada pela álgebra $\mathcal{B}_0$ do Exercício 1.2.

1.4. Seja $\mathcal{E}$ uma família de subconjuntos de um conjunto M. Mostre que existe a menor *álgebra* $\mathcal{B}_0$ que contém $\mathcal{E}$. Que relação existe entre $\mathcal{B}_0$ e a σ-álgebra $\mathcal{B}$ gerada por $\mathcal{E}$?

1.5. Seja $(M, \mathcal{B}, \mu)$ um espaço de medida. Mostre que se A_1,A_2, ... estão em $\mathcal{B}$ então

$$\mu(\bigcup_{j=1}^{\infty} A_j) \leq \sum_{j=1}^{\infty} \mu(A_j).$$

1.6. Seja $\mathcal{B} = 2^M$ e considere $\mu : 2^M \to [0, +\infty]$ definido por:

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

Mostre que μ é uma medida. Esta medida é designada *medida de contagem.*

1.7. Demonstre o Teorema 1.10. *Dica*: Dados quaisquer conjuntos disjuntos dois a dois $B_1, \ldots, B_n, \ldots$ em $\mathcal{B}_0$ tais $B = \bigcup_{j=1}^{\infty} B_j$ também está em $\mathcal{B}_0$, defina $C_j = B_1 \cup \cdots \cup B_j$ para cada $j \geq 1$. Verifique que os conjuntos $A_j = B \backslash C_j$ satisfazem a hipótese (1.1) no Teorema 1.10.

1.8. Seja $(M, \mathcal{B})$ um espaço mensurável.

1. Mostre que se $\mu : \mathcal{B} \to [0, +\infty]$ é uma medida então

$$\mu(\bigcup_{j=1}^{\infty} A_j) = \lim_{j \to \infty} \mu(A_j).$$

   para qualquer sequência crescente $A_1 \subset A_2 \subset \cdots A_n \subset \cdots$ de elementos de $\mathcal{B}$.

2. Reciprocamente, mostre que se $\mu_0 : \mathcal{B} \to [0, +\infty]$ é uma função finitamente aditiva que satisfaz a condição do item anterior então μ_0 é σ-aditiva.

1.9. Seja $(M, \mathcal{B})$ um espaço mensurável, onde o conjunto M é não-enumerável e a σ-álgebra $\mathcal{B}$ é definida como no Exercício 1.3. Mostre que $\mu : \mathcal{B} \to [0, +\infty]$ definida por:

$$\mu(A) = \begin{cases} 0 & \text{se } A \text{ é finito ou enumerável} \\ 1 & \text{se } A^c \text{ é finito ou enumerável} \end{cases}$$

é uma medida de probabilidade.

1.10. Sejam f e g funções mensuráveis. Mostre que f é integrável se e somente se $|f|$ é integrável e, nesse caso,

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

Além disso, se f é integrável e $|f| \geq |g|$ então g é integrável.

1.11. Seja E um conjunto mensurável com $\mu(E) = 0$. Mostre que $\int_E f \, d\mu = 0$ para qualquer função mensurável f.

1.12. Mostre que a é um ponto de densidade do conjunto A se e só se

$$\lim_{\varepsilon\to 0} \inf \left\{ \frac{\mu(B\cap A)}{\mu(B)} : B \text{ bola contida em } B(a,\varepsilon) \text{ e contendo } a \right\} = 1$$

1.13. Demonstre o Teorema 1.26.

1.14. Seja x_1, $x_2 \in M$ e p_1, p_2, q_1, $q_2 > 0$ com $p_1 + p_2 = q_1 + q_2 = 1$. Considere as medidas de probabilidade μ e ν dadas por

$$\mu(A) = \sum_{x_i \in A} p_i, \qquad \nu(A) = \sum_{x_i \in A} q_i,$$

ou seja, $\mu = p_1\delta_{x_1} + p_2\delta_{x_2}$ e $\nu = q_1\delta_{x_1} + q_2\delta_{x_2}$. Mostre que $\nu \ll \mu$ e $\mu \ll \nu$ e calcule as respectivas derivadas de Radon-Nikodym.

1.15. Seja $f : M \to \mathbb{R}$. Mostre que se $f^{-1}((-\infty, c)) \in \mathcal{B}$ para todo $c \in \mathbb{R}$ então f é mensurável. *Dica*: Mostre que a família $\mathcal{C} = \{A \subset \mathbb{R} : f^{-1}(A) \in \mathcal{B}\}$ é uma σ-álgebra e contém todos os subconjuntos abertos.

1.16. Prove o Teorema 1.20. *Dica*: Trate primeiro o caso onde f é não-negativa.

1.17. Mostre que o suporte de uma medida é sempre um conjunto fechado. Conclua que se M é compacto, o suporte de qualquer medida também é compacto.

1.18. Mostre que toda função $f : \mathbb{R}^m \to \mathbb{R}^n$ contínua é mensurável. Dê exemplo de uma função mensurável que não é contínua em nenhum ponto.

1.19. Seja $T : M \to M$ uma função mensurável e ν uma medida. Defina $T_*\nu(A) = \nu(T^{-1}(A))$. Mostre que $T_*\nu$ é uma medida.

Capítulo 2

Teorema de Recorrência de Poincaré

Entenderemos por *sistemas dinâmicos* as transformações $f : M \to M$ em algum espaço métrico ou topológico M. Heuristicamente, pensamos em f como associando a cada estado $x \in M$ do sistema o estado $f(x) \in M$ em que o sistema se encontrará uma unidade de tempo depois. Um dos principais objetivos do estudo dos sistemas dinâmicos é descrever o comportamento assintótico de um ponto x segundo ação de f. Entretanto, na maioria dos casos, esse trabalho feito com total generalidade é tarefa impossível. É aí que entra a Teoria Ergódica, disciplina que a grosso modo estuda o comportamento assintótico para *quase* todo ponto x, num sentido que iremos precisar.

Um ponto $x \in M$ diz-se recorrente se a sua trajetória pelo sistema dinâmico $f : M \to M$ volta arbitrariamente perto de x quando o tempo vai para infinito. A dinâmica no conjunto dos pontos não-recorrentes é, em certo sentido, sempre a mesma, independentemente do sistema dinâmico. Por isso, é fundamental compreender o conjunto dos pontos recorrentes, já que ele contém toda a dinâmica interessante do sistema.

O resultado que estudaremos nesta capítulo, enunciado por Poincaré perto do final do século XIX, afirma que *quase todo ponto é recor-*

rente, relativamente a qualquer medida invariante *finita* do sistema dinâmico. Daremos duas versões deste resultado, a primeira numa linguagem mensurável e a segunda de natureza mais topológica. Também comentaremos que a hipótese de finitude da medida não pode ser omitida.

2.1 Versão mensurável

Sempre consideraremos medidas μ definida na σ-álgebra de Borel do espaço M. Dizemos que μ é uma probabilidade se $\mu(M) = 1$. Na maior parte dos casos trataremos com medidas finitas, isto é, tais que $\mu(M) < \infty$. Neste caso sempre podemos transformar μ numa probabilidade ν: para isso basta definir

$$\nu(E) = \frac{\mu(E)}{\mu(M)} \quad \text{para cada conjunto mensurável } E \subset M.$$

Em geral, uma medida μ diz-se invariante pela transformação f se

$$\mu(E) = \mu(f^{-1}(E)) \quad \text{para todo conjunto mensurável } E \subset M. \tag{2.1}$$

Heuristicamente, isto significa que a probabilidade de um ponto estar num dado conjunto e a probabilidade de que a sua imagem esteja nesse conjunto são iguais. Note que a definição (2.1) faz sentido, uma vez que a pré-imagem de um conjunto mensurável por uma transformação mensurável ainda é um conjunto mensurável.

No caso de fluxos, substituímos (2.1) por

$$\mu(E) = \mu(f^{-t}(E)) \quad \text{para todo mensurável } E \subset M \text{ e todo } t \in \mathbb{R}. \tag{2.2}$$

Teorema 2.1. *Seja $f : M \to M$ uma transformação mensurável e μ uma medida invariante finita. Seja $E \subset M$ qualquer conjunto mensurável com $\mu(E) > 0$. Então, μ-quase todo ponto $x \in E$ tem algum iterado $f^n(x)$, $n \geq 1$, que também está em E.*

Em outras palavras, o teorema afirma que quase todo ponto de E regressa a E no futuro. Antes mesmo de demonstrar este fato, podemos mostrar que ele implica outro aparentemente mais forte: quase todo ponto de E regressa a E infinitas vezes:

Corolário 2.2. *Nas condições do Teorema 2.1, para μ-quase todo ponto $x \in E$ existem infinitos valores de $n \geq 1$ tais que $f^n(x)$ está em E.*

Demonstração. Para cada $k \geq 1$ vamos representar por E_k o conjunto dos pontos $x \in E$ que regressam a E exatamente k vezes: existem exatamente k valores de $n \geq 1$ tais que $f^n(x) \in E$. Observe que o conjunto dos pontos que regressam a E apenas um número finito de vezes é precisamente

$$\bigcup_{k=1}^{\infty} E_k \,.$$

Portanto, para provar o corolário, basta mostrar que $\mu(E_k) = 0$ para todo $k \geq 1$. A demonstração será por contradição.

Suponhamos que $\mu(E_k) > 0$ para algum $k \geq 1$. Então, aplicando o Teorema 2.1 com este E_k no lugar de E, obtemos que quase todo ponto $x \in E_k$ tem algum iterado $f^n(x)$ que está em E_k. Fixemos um tal x e denotemos $y = f^n(x)$. Por definição, y tem exatamente k iterados futuros que estão em E. Como y é um iterado de x, isso implica que x tem $k+1$ iterados futuros em E. Mas isso contradiz o fato de que $x \in E_k$. Esta contradição prova que E_k tem medida nula, relativamente a μ, e portanto o corolário está demonstrado. □

Vamos agora dar a

Demonstração do Teorema 2.1. Representemos por E^0 o conjunto dos pontos $x \in E$ que nunca regressam a E. O nosso objetivo é provar que E^0 tem medida nula. Para isso, começamos por afirmar que as suas pré-imagens $f^{-n}(E^0)$ são disjuntas duas-a-duas. De fato, suponhamos que existem $m > n \geq 1$ tais que $f^{-m}(E^0)$ intersecta $f^{-n}(E^0)$. Seja x um ponto na intersecção e seja $y = f^n(x)$. Então $y \in E^0$ e $f^{m-n}(y) = f^m(x) \in E^0$, que está contido em E. Isto quer dizer que y volta pelo menos uma vez a E, o que contradiz a definição de E^0. Esta contradição, prova que as pré-imagens são disjuntas duas-a-duas, como afirmamos.

Isto implica que

$$\mu\Big(\bigcup_{n=0}^{\infty} f^{-n}(E^0)\Big) = \sum_{n=0}^{\infty} \mu(f^{-n}(E^0)) = \sum_{n=0}^{\infty} \mu(E^0).$$

Na última igualdade usamos a hipótese de que μ é invariante, que implica que $\mu(f^{-n}(E^0)) = \mu(E^0)$ para todo $n \geq 1$. Como supomos que a medida é finita, a expressão do lado esquerdo é finita. Por outro lado, à direita temos uma soma de infinitos termos, todos iguais. O único jeito desta soma ser finita é que as parcelas sejam nulas. Portanto, devemos ter $\mu(E^0) = 0$, tal como foi afirmado. $\square$

2.2 Versão topológica

Dizemos que um ponto $x \in M$ é *recorrente* para uma transformação $f : M \to M$ se, para toda vizinhança U de x, existe algum iterado $f^n(x)$ que está em U. A definição para fluxos é análoga, apenas nesse caso o tempo n é um número real.

Na formulação topológica do teorema de recorrência supomos que o espaço M admite uma base enumerável de abertos, ou seja, um família enumerável $\{U_k : k \in \mathbb{N}\}$ de abertos tal que todo aberto de M pode ser escrito como união de elementos U_k dessa família. Esta hipótese é satisfeita na maioria dos exemplos interessantes.

Teorema 2.3. *Suponhamos que M admite uma base enumerável de abertos. Seja $f : M \to M$ uma transformação mensurável e μ uma medida invariante finita. Então, μ-quase todo ponto $x \in M$ é recorrente para f.*

Demonstração. Para cada k representamos por U_k^0 o conjunto dos pontos $x \in U_k$ que nunca regressam a U_k. De acordo com o Teorema 2.1, todo U_k^0 tem medida nula. Consequentemente, a união enumerável

$$\tilde{U} = \bigcup_{k \in \mathbb{N}} U_k^0$$

tem medida nula. Portanto, para demonstrar o teorema será suficiente que mostremos que todo ponto x que não está em $\tilde{U}$ é recorrente. Isso é fácil, como vamos ver.

Seja $x \in M \setminus \tilde{U}$ e seja U uma vizinhança qualquer de x. A definição de base de abertos implica que existe algum $k \in \mathbb{N}$ tal que $x \in U_k$ e $U_k \subset U$. Como x não está em $\tilde{U}$, também $x \notin U_k^0$. Em outras palavras, x tem algum iterado $f^n(x), n \geq 1$ que está em U_k. Em particular, $f^n(x)$ também está em U. Como a vizinhança U é arbitrária, isto prova que x é um ponto recorrente, como havíamos afirmado. □

2.3 Recorrência para medidas infinitas

As conclusões dos Teoremas 2.1 e 2.3 não são verdadeiras, em geral, se omitirmos a hipótese de que a medida μ é finita. O exemplo mais simples é o seguinte:

Exemplo 2.4. Seja $f : \mathbb{R} \to \mathbb{R}$ a translação de 1 unidade, isto é, $f(x) = x+1$ para todo $x \in \mathbb{R}$. É fácil verificar que f deixa invariante a medida de Lebesgue em $\mathbb{R}$ (que é infinita). Por outro lado nenhum ponto é recorrente para f.

No entanto, é possível estender estes enunciados para certos casos de medidas infinitas como, por exemplo, no Exercício 2.2.

Uma transformação $f : M \to M$ diz-se invertível se é uma bijeção e a sua inversa é também uma transformação mensurável. Uma medida μ diz-se σ-finita se existe uma sequência crescente de subconjuntos M_k cuja união é o espaço M inteiro e tal que cada $\mu(M_k)$ é finito. Neste caso, diremos que um ponto x "vai para infinito"se, para qualquer k, existe apenas um número finito de iterados de x que estão em M_k.

2.4 Exercícios

2.1. Mostre que o seguinte enunciado é equivalente ao Teorema 2.1, isto é, qualquer um dos dois pode ser deduzido a partir do outro: Seja $f : M \to M$ uma transformação mensurável e μ uma medida invariante finita. Seja $E \subset M$ qualquer conjunto mensurável com $\mu(E) > 0$. Então existe $N \geq 1$ e um conjunto $D \subset E$ com medida positiva, tal que $f^N(x) \in E$ para todo ponto $x \in D$.

2.2. Suponha que $f : M \to M$ é invertível e que μ é uma medida σ-finita invariante por f. Mostre que, dado qualquer conjunto mensurável $E \subset M$ com $\mu(E) > 0$, quase todo ponto $x \in E$ ou regressa a E ou "vai para infinito".

Dica: Considere o conjunto $E^{0,k}$ dos pontos $x \in E$ que nunca regressam a E e têm um número infinito de iterados em M_k. Comece por mostrar que os seus iterados $f^n(E^{0,k})$ são dois-a-dois disjuntos. Usando que $\mu(M_k)$ é finito, deduza que $\mu(E^{k,0}) = 0$ para todo k.

2.3. Dadas funções $f : M \to M$ e $\tau : M \to \mathbb{N}$, defina o *mapa induzido* F por $F(x) = f^{\tau(x)}(x)$. Mostre que:

1. Se f é invertível e μ é uma medida f-invariante, então μ é F-invariante. Dê exemplo de alguma função τ onde μ é f-invariante mas não é F-invariante.

2. Se μ_F é uma medida F invariante, então a medida

$$\mu_f = \sum_{j=0}^{\infty} f_*^j \left(\mu_F \mid \{\tau_f > j\}\right)$$

é medida f-invariante.

3. Mostre que se τ é μ_F integrável, então a medida μ_f acima é finita.

Definição 2.5. Chama-se $L^2(\mu)$ o espaço das funções [1] mensuráveis $\psi : S^1 \to \mathbb{C}$ cujo quadrado é integrável:

$$\int |\psi|^2 \, d\mu < \infty.$$

É claro que este espaço contém todas as funções mensuráveis limitadas e, em particular, todas as funções características de conjuntos mensuráveis.

2.4. Definindo o produto interno $< \phi, \psi >= \int \phi\psi \, d\mu$, mostre que L^2 é um espaço vetorial normado completo (i.e., toda sequência de Cauchy converge).

2.5. Mostre que f preserva uma medida μ se, e somente se, $U_f : L^2(\mu) \to L^2(\mu)$ definida por $U_f(\phi)(x) = \phi(f(x))$ é uma isometria com respeito a norma que provêm do produto interno definido no exercício anterior.

[1] Quando lidamos com $L^2(\mu)$ sempre identificamos funções que diferem apenas num conjunto de medida nula.

2.6. Dê exemplo de transformação f preservando uma medida μ tal que $U_f : L^2(\mu) \to L^2(\mu)$ não seja sobrejetora.

Capítulo 3

Exemplos de Medidas Invariantes

Neste capítulo vamos descrever alguns exemplos simples de medidas invariantes por transformações ou por fluxos. Antes porém, vamos mostrar uma proposição caracterizando quando uma medida é invariante:

Proposição 3.1. *Seja $f : M \to M$ uma transformação e μ uma medida. Então f preserva μ se, e somente se, para toda função integrável $\phi : M \to \mathbb{R}$ vale:*

$$\int \phi \, d\mu = \int \phi \circ f \, d\mu.$$

Demonstração. Assuma que f preserva a medida μ. Se ϕ é função característica de algum conjunto, digamos $\phi = \chi_A$, é imediato verificar que $\mu(f^{-1}(A)) = \int \phi \circ f \, d\mu$, já que $\chi_{f^{-1}(A)} = \phi \circ f$. Assim, fica provado que $\int \phi \, d\mu = \int \phi \circ \ f \, d\mu$, quando ϕ é uma função característica. Observe que segue diretamente da linearidade da integral que se ϕ é uma função simples, então a igualdade ainda vale. Finalmente, se ϕ é uma função integrável qualquer, pela definição de integral

$$\int \phi \, d\mu = \lim_{n \to \infty} \int \phi_n \, d\mu,$$

onde ϕ_n é uma sequência de funções simples crescendo para ϕ. Por outro lado, $\phi_n \circ f$ é uma sequência de funções simples crescendo para $\phi \circ f$. Logo,

$$\int \phi \circ f \, d\mu = \lim_{n\to\infty} \int \phi_n \circ f \, d\mu.$$

Como $\int \phi_n \, d\mu = \int \phi_n \circ f \, d\mu$, tomando o limite em ambos os lados, vem que

$$\int \phi \, d\mu = \int \phi \circ f \, d\mu.$$

A recíproca é imediata, desde que dado um boreliano A, tomando $\phi = \chi_A$, então

$$\mu(A) = \mu(f^{-1}(A)) \Leftrightarrow \int \phi \, d\mu = \int \phi \circ f \, d\mu.$$

□

3.1 Expansão decimal

O nosso primeiro exemplo é

$$f : [0,1] \to [0,1], \quad f(x) = 10x - [10x]$$

onde $[10x]$ representa o maior inteiro menor ou igual a $10x$. Em outras palavras, f associa a cada $x \in [0,1]$ a parte fracionária de $10x$. O gráfico da transformação f está descrito na Figura 3.1.

Afirmamos que a medida de Lebesgue μ no intervalo é invariante pela transformação f, isto é, satisfaz a condição (2.1). Comecemos por supor que E é um intervalo. Então, como ilustrado na Figura 3.1, a pré-imagem $f^{-1}(E)$ consiste de dez intervalos, cada um deles dez vezes mais curto do que E. Logo, a medida de Lebesgue de $f^{-1}(E)$ é igual à medida de Lebesgue de E. Isto mostra que (2.1) é satisfeita no caso de intervalos. Por outro lado, a família dos intervalos gera a σ-álgebra de Borel de $[0,1]$. Portanto, para concluir a demonstração basta usar o seguinte fato geral (veja o Exercício 3.1):

Lema 3.2. *Seja $f : M \to M$ uma transformação mensurável e μ uma medida finita em M. Suponha que existe uma sub-álgebra geradora $\mathcal{I}$ da σ-álgebra de M tal que $\mu(E) = \mu(f^{-1}(E))$ para todo*

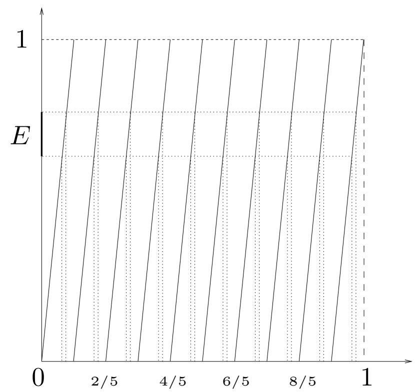

Figura 3.1: Transformação parte fracionária de $10x$

$E \in \mathcal{I}$. Então o mesmo vale para todo conjunto mensurável E, isto é, a medida μ é invariante por f.

Agora vamos explicar como, a partir do fato de que a medida de Lebesgue é invariante pela transformação f, podemos obter conclusões interessantes e não-triviais usando o teorema de recorrência de Poincaré.

Comecemos por observar que f tem uma expressão muito simples em termos de expansões decimais: se x é dado por

$$x = 0, a_0 a_1 a_2 a_3 \cdots$$

com $a_i \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$, então a sua imagem é dada por

$$f(x) = 0, a_1 a_2 a_3 \cdots .$$

Com isso, fica muito fácil escrever a expressão do iterado n-ésimo, para qualquer $n \geq 1$:

$$f^n(x) = 0, a_n a_{n+1} a_{n+2} \cdots \tag{3.1}$$

Agora, seja E o subconjunto dos $x \in [0, 1]$ cuja expansão decimal começa com o dígito 7, ou seja, tais que $a_0 = 7$. De acordo com o

Corolário 2.2, quase todo elemento de E tem infinitos iterados que também estão em E. Levando em conta a expressão (3.1), isto quer dizer que existem infinitos valores de n tais que $a_n = 7$. Portanto, provamos que *quase todo número x cuja expansão decimal começa por 7 tem infinitos dígitos iguais a 7!*

Claro que no lugar de 7 podemos considerar qualquer outro dígito. Além disso, podemos considerar blocos de dígitos mais complicados. Veja os Exercícios 3.2–3.3.

Mais tarde iremos provar resultados mais fortes: para quase todo número $x \in [0,1]$, todo dígito aparece com frequência $1/10$ na sua expansão decimal. O enunciado preciso aparecerá na Proposição 6.2, que será provada a partir do teorema ergódico de Birkhoff.

3.2 Sistemas conservativos

Seja U um aberto em algum espaço euclidiano $\mathbb{R}^d$, $d \geq 1$ e seja $f : U \to U$ um difeomorfismo de classe C^1. Isto quer dizer que f é uma bijeção e tanto ele quanto a sua inversa são deriváveis com derivada contínua.

Representaremos por vol a medida de Lebesgue, ou volume, em $\mathbb{R}^k$. Em outras palavras,

$$\operatorname{vol}(B) = \int_B dx_1 \dots dx_d \quad \text{e} \quad \int_B \varphi \, d\operatorname{vol} = \int_B \varphi(x_1, \dots, x_d)\, dx_1 \dots dx_d$$

para qualquer conjunto mensurável B e qualquer função integrável φ.

A fórmula de mudança de variáveis afirma que, para qualquer conjunto mensurável $B \subset U$,

$$\operatorname{vol}(f(B)) = \int_B |\det Df| \, d\operatorname{vol} \tag{3.2}$$

Daqui se deduz facilmente

Lema 3.3. *Um difeomorfismo $f : M \to M$ de classe C^1 deixa invariante o volume se e somente se o valor absoluto $|\det Df|$ do seu jacobiano é constante igual a 1.*

Demonstração. Suponha primeiro que o valor absoluto do jacobiano é igual 1 em todo ponto. Considere um conjunto mensurável E e seja $B = f^{-1}(E)$. A fórmula (3.2) dá que

$$\operatorname{vol}(E) = \int_B 1\, d\operatorname{vol} = \operatorname{vol}(B) = \operatorname{vol}(f^{-1}(E)).$$

Isto significa que f deixa invariante o volume e, portanto, provamos a parte "se"do enunciado.

Para provar a parte "somente se", suponha que $|\det Df|$ fosse maior que 1 em algum ponto x. Então, como o jacobiano é contínuo, existiria uma vizinhança U de x e algum número $\sigma > 1$ tais que

$$|\det Df(y)| \geq \sigma \quad \text{para todo } y \in U.$$

Então a fórmula (3.2) aplicada a $B = U$ daria

$$\operatorname{vol}(f(U)) \geq \int_U \sigma\, d\operatorname{vol} \geq \sigma \operatorname{vol}(U).$$

Denotando $E = f(U)$, isto implica que $\operatorname{vol}(E) > \operatorname{vol}(f^{-1}(E))$ e, portanto, f não deixa invariante o volume. Do mesmo modo se mostra que se o valor absoluto do jacobiano é menor que 1 em algum ponto então f não deixa invariante o volume. □

Os Exercícios 3.4–3.5 estendem este lema para transformações não necessariamente invertíveis e também para uma classe mais ampla de medidas. As suas conclusões nos serão úteis mais tarde.

Agora vamos considerar o caso de fluxos $f^t : U \to U$, $t \in \mathbb{R}$. Suporemos que o fluxo é de classe C^1. Claro que o Lema 3.3 se aplica neste contexto: o fluxo deixa invariante o volume se e somente se

$$\det Df^t(x) = 1 \quad \text{para todo } x \in U \text{ e todo } t \in \mathbb{R}. \tag{3.3}$$

Façamos duas observações simples antes de prosseguirmos. A primeira é que segue da definição de fluxo que todo f^t é invertível (um difeomorfismo, neste caso): a sua inversa é f^{-t}. A segunda observação é que o jacobiano de f^t é sempre positivo. Isso é claro quando $t = 0$ porque, outra vez por definição de fluxo, f^0 é a identidade. Segue que o mesmo é verdade para todo $t \in \mathbb{R}$, porque o jacobiano varia continuamente com t e, como acabamos de ver, nunca se anula.

Embora a resposta que acabamos de dar esteja inteiramente correta, ela não é muito útil na prática porque em geral não temos uma expressão explícita para f^t, e portanto não é claro como verificar a condição (3.3). Felizmente, existe uma expressão razoavelmente explícita para o jacobiano, de que iremos falar em seguida, que pode ser usada em muitas situações interessantes.

Suponhamos que o fluxo f^t corresponde às trajetórias de um campo de vetores $F : U \to U$ de classe C^1, quer dizer $f^t(x)$ é o valor no tempo t da solução da equação diferencial

$$\frac{dx}{dt} = F(x) \tag{3.4}$$

(quando tratando de equações diferenciáveis sempre suporemos que as suas soluções estão definidas para todo tempo). A fórmula de Liouville exprime o jacobiano de f^t em termos do divergente $\operatorname{div} F$ do campo de vetores F:

$$\det Df^t(x) = \exp\big(\int_0^t \operatorname{div} F(f^s(x))\, ds\big).$$

Lembre que o divergente de um campo de vetores F é o traço da sua matriz jacobiana, isto é

$$\operatorname{div} F = \frac{\partial F}{\partial x_1} + \cdots + \frac{\partial F}{\partial x_d}. \tag{3.5}$$

Combinando esta fórmula com (3.3) obtemos

Lema 3.4. *O fluxo f^t associado a um campo de vetores F de classe C^1 deixa invariante o volume se e somente se o divergente de F é identicamente nulo.*

O Exercício 3.6 é uma aplicação deste fato no caso, muito importante, de fluxos hamiltonianos.

3.3 Deslocamentos ("shifts") de Bernoulli

Estes sistemas modelam sequências de experimentos aleatórios em que o resultado de cada experimento é independente dos demais.

Supõe-se que em cada experimento há um número finito de resultados possíveis, designados por $1, 2, \ldots, d$, com probabilidades $p(1), p(2), \ldots, p(d)$ de ocorrerem, sendo

$$p(1) + p(2) + \cdots + p(d) = 1\,.$$

O conjunto M das sequências $\underline{\alpha} = (\alpha_n)_{n\in\mathbb{Z}}$ com cada $\alpha_n \in \{1, 2, \ldots, d\}$ contém os possíveis resultados da sequência de experimentos. Chamam-se *cilindros* os subconjuntos da forma

$$[k, l; a_k\,, \ldots, a_l] = \{\underline{\alpha} \in M : \alpha_k = a_k\,, \ldots, \alpha_l = a_l\}$$

onde $k, l \in \mathbb{Z}$, com $k \le l$, e cada $a_j \in \{1, 2, \ldots, d\}$. Definimos

$$\mu([k, l; a_k\,, \ldots, a_l]) = p(a_k) \cdots p(a_l) \tag{3.6}$$

Heuristicamente, isto significa que a probabilidade do evento composto

$$\alpha_k = a_k \quad \text{e} \quad \alpha_{k+1} = a_{k+1} \quad \text{e} \quad \cdots \quad \text{e} \quad \alpha_l = a_l$$

é o produto das probabilidades de cada um deles. Isto traduz, precisamente, que os resultados sucessivos são independentes entre si.

Consideramos em M a σ-álgebra $\mathcal{B}$ gerada pelos cilindros. A família $\mathcal{B}_0$ das uniões disjuntas finitas dos cilindros é uma álgebra (por convenção, M é um cilindro e $\mu(M) = 1$). Estendemos μ de modo a que seja finitamente aditiva: se $E \in \mathcal{B}_0$ é a união disjunta de cilindros $C_1\,, \ldots, C_N$, definimos

$$\mu(E) = \mu(C_1) + \cdots + \mu(C_N).$$

Verifica-se que esta função μ é, de fato, σ-aditiva em $\mathcal{B}_0$; por exemplo, isso pode ser feito usando o Teorema 1.10. Portanto existe uma única probabilidade na σ-álgebra $\mathcal{B}$ gerada por $\mathcal{B}_0$ que é uma extensão de μ, isto é, que coincide com ela restrita a $\mathcal{B}_0$. Chamamos essa probabilidade *medida de Bernoulli* definida por $p(1), p(2), \ldots, p(d)$ e, para não complicar desnecessariamente a notação, a representamos também por μ.

No espaço M consideramos a transformação deslocamento ("shift") à esquerda

$$f : M \to M \quad f\big((\alpha_n)_{n\in\mathbb{Z}}\big) = (\alpha_{n+1})_{n\in\mathbb{Z}}$$

que corresponde a fazer uma translação no tempo. Observe que a medida de Bernoulli é invariante por essa transformação. De fato, se $E = [k, l; a_k, \ldots, a_l]$ então $f^{-1}(E) = [k+1, l+1; a_k, \ldots, a_l]$ e a definição (3.6) dá que

$$\mu(E) = \mu(f^{-1}(E))$$

neste caso. Como a família dos cilindros gera a σ-álgebra $\mathcal{B}$, isto juntamente com o Lema 3.2, prova que a medida μ é invariante para f.

3.4 Transformação de Gauss

A transformação de Gauss $G : (0, 1] \to [0, 1]$ é definida por $G(x) =$ parte fracionária de $1/x$, ou seja,

$$G(x) = \frac{1}{x} - \left[\frac{1}{x}\right].$$

O gráfico de G pode ser esboçado facilmente, a partir da seguinte observação.

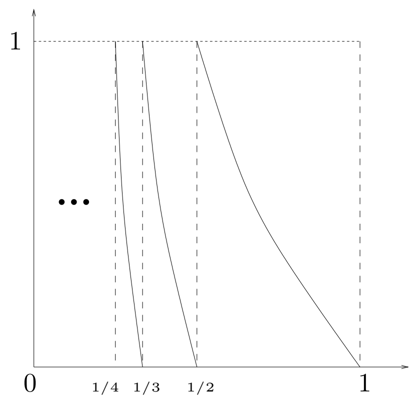

Figura 3.2: Transformação de Gauss

- Se $x \in (1/2, 1]$ então $1/x \in [1, 2)$ e portanto a sua parte inteira $[1/x]$ é igual a 1. Isto quer dizer que neste intervalo a transformação é dada por $G(x) = (1/x) - 1$.

- Mais geralmente, se $x \in (1/(k+1), 1/k)$ para algum $k \in \mathbb{N}$ então a parte inteira de $1/x$ é igual a k, e tem-se $G(x) = 1/x - k$. Veja também a Figura 3.2.

Note que G não está definida no ponto $x = 0$. Além disso, $G(1/k) = 0$ para todo $k \in \mathbb{N}$ e portanto o segundo iterado $G^2(1/k)$ não está definido nestes pontos (e o terceiro iterado não está definido nas suas pré-imagens, etc). Isto quer dizer, a rigor, que G não é um sistema dinâmico segundo a definição que demos antes. No entanto, isto não coloca nenhum problema para o que pretendemos fazer. De fato, todos os iterados estão bem definidos no conjunto dos números irracionais: basta observar que a imagem de um irracional também é irracional. Isto é suficiente para os nossos objetivos porque sempre tratamos de propriedade que valem para quase todo ponto, e o conjunto dos números irracionais tem medida de Lebesgue total no intervalo.

O que torna esta transformação interessante do ponto de vista ergódico é que G admite uma probabilidade invariante que é equivalente à medida de Lebesgue no intervalo. De fato, considere a medida definida por

$$\mu(E) = \int_E \frac{c}{1+x}\, dx \quad \text{para cada mensurável } E \subset [0, 1]$$

onde c é uma constante positiva. Note que a integral está bem definida, já que a função integranda é contínua no intervalo $[0, 1]$. Note também que

$$\frac{c}{2}\, m(E) \le \mu(E) \le c m(E) \quad \text{para todo mensurável } E \subset [0, 1].$$

Em particular, μ é de fato equivalente à medida de Lebesgue m: as duas medidas têm os mesmos conjuntos com medida nula.

Proposição 3.5. *A medida μ é invariante por G. Além disso, se escolhermos $c = 1/log 2$ então μ é uma probabilidade.*

Demonstração. Vamos usar o critério dado pelo exercício 3.5: a medida μ é invariante por G se tivermos

$$\sum_{x\in f^{-1}(y)} \frac{\rho(x)}{|G'(x)|} = \rho(y) \qquad \text{onde } \rho(x) = \frac{c}{1+x} \tag{3.7}$$

para todo y. Comece por observar que cada y tem exatamente uma pré-imagem x_k em cada intervalo $(1/(k+1), 1/k]$, dada por

$$G(x_k) = \frac{1}{x_k} - k = y \quad \Leftrightarrow \quad x_k = \frac{1}{y+k}\,.$$

Note também que $G'(x) = (1/x)' = -1/x^2$. Portanto, (3.7) se reescreve como

$$\sum_{k=1}^{\infty} \frac{cx_k^2}{1+x_k} = \frac{c}{1+y} \quad \Leftrightarrow \quad \sum_{k=1}^{\infty} \frac{1}{(y+k)(y+k+1)} = \frac{c}{1+y} \tag{3.8}$$

Para verificar que esta igualdade é realmente satisfeita, observe que

$$\frac{1}{(y+k)(y+k+1)} = \frac{1}{y+k} - \frac{1}{y+k+1}\,.$$

Isto quer dizer que a última soma em (3.8) pode ser escrita na forma telescópica: todos os termos, exceto o primeiro, aparecem duas vezes, com sinais contrários, e portanto se cancelam. Logo a soma é igual ao primeiro termo, que é precisamente o que se afirma em (3.8). Isto prova a invariância.

Finalmente, usando a primitiva $c\log(1+x)$ da função $\rho(x)$ vemos que

$$\mu([0,1]) = \int_0^1 \frac{c}{1+x}\,dx = c\log 2.$$

Logo, escolhendo $c = 1/\log 2$ obtemos que μ é uma probabilidade. $\square$

A transformação de Gauss tem um papel muito importante em teoria dos números, devido à sua relação com o processo de expansão dos números em fração contínua. Recordemos do que se trata.

Dado um número $x_0 \in (0,1)$, seja

$$a_1 = \left[\frac{1}{x_0}\right] \qquad \text{e} \qquad x_1 = \frac{1}{x_0} - a_1 = G(x_0).$$

Note que a_1 é um número natural, $x_1 \in [0,1)$ e tem-se

$$x_0 = \frac{1}{a_1 + x_1}.$$

Agora, supondo que x_1 seja diferente de zero, podemos repetir o processo, definindo

$$a_2 = \left[\frac{1}{x_1}\right] \qquad \text{e} \qquad x_2 = \frac{1}{x_1} - a_2 = G(x_1).$$

Então

$$x_1 = \frac{1}{a_1 + x_2} \quad \text{portanto} \quad x_0 = \cfrac{1}{a_1 + \cfrac{1}{a_2 + x_2}}.$$

Por recorrência, para cada $n \geq 1$ tal que $x_{n-1} \in (0,1)$ se define

$$a_n = \left[\frac{1}{x_{n-1}}\right] \qquad \text{e} \qquad x_n = \frac{1}{x_{n-1}} - a_n = G(x_{n-1})$$

e tem-se

$$x_0 = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\cdots + \cfrac{1}{a_n + x_n}}}}. \tag{3.9}$$

Não é difícil mostrar (verifique!) que a sequência

$$z_n = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\cdots + \cfrac{1}{a_n}}}}.$$

converge para x_0 quando $n \to \infty$, e é usual traduzir este fato escrevendo

$$x_0 = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\cdots + \cfrac{1}{a_n + \cfrac{1}{\cdots}}}}}, \tag{3.10}$$

que é chamada *expansão em fração contínua* de x_0.

Note que a sequência z_n consiste de números racionais. De fato se mostra que estes são os números racionais que melhor aproximam o número x_0, no sentido de que z_n está mais próximo de x_0 do que qualquer outro número racional com denominador menor ou igual que o denominador de z_n (escrito em forma irredutível). Observe também que para obter (3.10) supusemos que $x_n \in (0,1)$ para todo $n \in \mathbb{N}$. Se encontramos algum $x_n = 0$, o processo para nesse momento e consideramos (3.9) a expansão em fração contínua de x_0. Claro que este último caso ocorre somente se x_0 é um número racional.

Estas idéias de Teoria Ergódica podem ser usadas para obter conclusões não triviais em Teoria dos Números. Por exemplo (veja o Exercício 3.7), para quase todo número $x_0 \in (1/8, 1/7)$ o número 7 aparece infinitas vezes na sua expansão em fração contínua, isto é, tem-se $a_n = 7$ para infinitos valores de $n \in \mathbb{N}$.

De fato, mais tarde provaremos um fato muito mais preciso: para quase todo $x_0 \in (0,1)$ o número 7 aparece com frequência

$$\frac{1}{\log 2} \log \frac{64}{63}$$

na sua expansão em fração contínua. Tente intuir desde já de onde vem este número!

3.5 Exercícios

3.1. Demonstre o Lema 3.2. Dica: mostre que a família de todos os conjuntos E tais que $\mu(E) = \mu(f^{-1}(E))$ é uma σ-álgebra.

3.2. Prove que, para quase todo número $x \in [0,1]$ cuja expansão decimal contém o bloco 617 (por exemplo $x = 0,3375\mathbf{617}264\cdots$), esse bloco aparece infinitas vezes na expansão.

3.3. Prove que o dígito 7 aparece infinitas vezes na expansão decimal de quase todo número $x \in [0,1]$. Dica: Comece por mostrar que o conjunto dos números cuja expansão decimal *nunca* exibe o dígito 7 tem medida nula.

3.4. Suponha que $f : U \to U$ é um difeomorfismo local (isto é: o seu jacobiano é não nulo em todo ponto) de classe C^1. Mostre que f deixa invariante o volume se e somente se

$$\sum_{x \in f^{-1}(y)} \frac{1}{|\det Df(x)|} = 1 \quad \text{para todo } y \in U.$$

3.5. Dada uma função $\rho : U \to [0, \infty)$, denotamos por $\mu = \rho \,\mathrm{vol}$ a medida definida por $\mu(E) = \int_E \rho \, d\,\mathrm{vol}$. Suponha que $f : U \to U$ é um difeomorfismo local de classe C^1 e que ρ é uma função contínua. Mostre que f deixa invariante a medida $\mu = \rho \,\mathrm{vol}$ se e somente se

$$\sum_{x \in f^{-1}(y)} \frac{\rho(x)}{|\det Df(x)|} = \rho(y) \quad \text{para todo } y \in U.$$

Em particular, no caso em que f é invertível, f deixa invariante a medida μ se e somente se $\rho(x) = \rho(f(x))|\det Df|(x)$ para todo $x \in U$.

3.6. Seja U um aberto de $\mathbb{R}^{2d}$ e $H : U \to \mathbb{R}$ uma função de classe C^2. Denotamos as variáveis em $\mathbb{R}^{2d}$ por $(p_1, q_1, \ldots, \ldots, p_d, q_d)$. O campo de vetores hamiltoniano associado a H é definido por

$$F(p_1, \ldots, p_d, q_1, \ldots, q_d) = \left(\frac{\partial H}{\partial q_1}, \ldots, \frac{\partial H}{\partial q_d}, -\frac{\partial H}{\partial p_1}, \ldots, -\frac{\partial H}{\partial p_d} \right).$$

Verifique que o fluxo definido por F preserva o volume.

3.7. Para (Lebesgue) quase todo número $x_0 \in (1/8, 1/7)$ o número 7 aparece infinitas vezes na sua expansão em fração contínua, isto é, tem-se $a_n = 7$ para infinitos valores de $n \in \mathbb{N}$.

3.8. Considere a sequência $1, 2, 4, 8, \ldots, a_n = 2^n, \ldots$. Mostre que dado um dígito $i \in 0, \ldots, 9$, existe uma quantidade infinita de valores n tal que a_n começa com este dígito.

3.9. Mostre que se A é uma matriz $n \times n$ com coeficientes inteiros, então a transformação induzida $[A] : \Pi^n \to Pi^n$ definida por $[A](\bar{x}) = A\overline{(x)}$ preserva a medida de Lebesgue de Π^n.

3.10. Mostre que o deslocamento σ definido na Secção 3.3 é transitivo e que o conjunto de suas órbitas periódicas é denso.

Capítulo 4

Existência de Medidas Invariantes

Neste capítulo provaremos o seguinte resultado, que garante a existência de medidas invariantes em grande generalidade:

Teorema 4.1. *Seja $f : M \to M$ uma transformação contínua num espaço métrico compacto. Então existe pelo menos uma probabilidade invariante por f. O mesmo resultado vale para fluxos.*

Antes de demonstrarmos este resultado, mencionemos alguns exemplos que mostram que nenhuma das duas hipóteses, continuidade e compacidade, podem ser omitidas.

4.1 Alguns exemplos simples

Considere $f : (0,1] \to (0,1]$ dada por $f(x) = x/2$. Suponha que f admite alguma probabilidade invariante (o objetivo é mostrar que isso não acontece). Pelo Teorema de Recorrência 2.3, relativamente a essa probabilidade quase todo ponto de $(0,1]$ é recorrente. Mas é imediato que não existe nenhum ponto recorrente: a órbita de qualquer $x \in (0,1]$ converge para zero e, em particular, não acumula no ponto inicial x. Isto mostra que f é um exemplo de transformação contínua num espaço não compacto que não admite nenhuma medida probabilidade invariante.

Modificando um pouco o exemplo, podemos mostrar que o mesmo fenômeno pode ocorrer em espaços compactos, se a transformação não é contínua. Considere $f : [0,1] \to [0,1]$ dada por $f(x) = x/2$ se $x \neq 0$ e $f(0) = 1$. Pela mesma razão que antes, nenhum ponto $x \in (0,1]$ é recorrente. Portanto, se existe alguma probabilidade invariante μ ela tem dar peso total ao único ponto recorrente que é $x = 0$. Em outras palavras, μ precisa ser a medida de Dirac δ_0 suportada em zero, que é definida por

$$\delta_0(E) = 1 \text{ se } 0 \in E \quad \text{e} \quad \delta_0(E) = 0 \text{ se } 0 \notin E.$$

Mas a medida δ_0 *não é* invariante por f: tomando $E = \{0\}$ temos que E tem medida 1 mas a sua pré-imagem $f^{-1}(E)$ é o conjunto vazio, que tem medida nula. Portanto, esta transformação também não tem nenhuma probabilidade invariante.

O nosso terceiro exemplo é de natureza um pouco diferente. Consideremos $f : [0,1] \to [0,1]$ dada por $f(x) = x/2$. Trata-se de uma transformação contínua num espaço compacto. Logo, pelo teorema que iremos demonstrar, admite alguma probabilidade invariante. Pelos mesmos argumentos que usamos no caso anterior, se conclui que de fato há uma única probabilidade invariante, que é a medida de Dirac δ_0 suportada no ponto zero. Note que neste caso δ_0 é de fato invariante.

Mencionamos este último caso para enfatizar as limitações do Teorema de Existência (que são inerentes à sua grande generalidade): as medidas que ele garante existirem podem ser bastante triviais; por exemplo, neste caso quando falamos de "quase todo ponto" estamos nos referindo apenas ao ponto $x = 0$. Por isso, um objetivo importante é obter resultados mais sofisticados de existência de medidas com propriedades adicionais que as tornem mais interessantes, por exemplo serem equivalentes à medida de Lebesgue.

4.2 A topologia fraca* no espaço das medidas

Nesta seção vamos introduzir uma topologia importante no conjunto $\mathcal{M}_1(M)$ das probabilidades borelianas do espaço M, chamada topologia fraca* , que será muito útil para provar o Teorema 4.1. A idéia da definição é a seguinte: duas medidas estão próximas se dão integrais próximas para muitas funções contínuas. Procuremos exprimir esta idéia de modo preciso.

Dada uma medida $\mu \in \mathcal{M}_1(M)$, um conjunto finito $F = \{\phi_1, \ldots, \phi_N\}$ de funções contínuas $\phi_j : M \to \mathbb{R}$, e um número $\varepsilon > 0$, definimos

$$V(\mu, F, \varepsilon) = \{\eta \in \mathcal{M}_1(M) : \left| \int \phi_j \, d\eta - \int \phi_j \, d\mu \right| < \varepsilon \text{ para todo } \phi_j \in F\}.$$

Então a topologia fraca* é definida estipulando que estes conjuntos $V(\mu, F, \varepsilon)$, com F e ε variável, constituem uma base de vizinhanças da medida μ. O seguinte lema deveria ajudar a compreender o significado desta topologia:

Lema 4.2. *Uma sequência $(\mu_n)_{n\in\mathbb{N}}$ em $\mathcal{M}_1(M)$ converge para uma medida $\mu \in \mathcal{M}_1(M)$ na topologia* fraca* *se e somente se*

$$\int \phi \, d\mu_n \to \int \phi \, d\mu \quad \textit{para toda função contínua } \phi : M \to \mathbb{R}.$$

Demonstração. Para provar a parte "somente se", considere qualquer função contínua ϕ e forme o conjunto $F = \{\phi\}$. Como $\mu_n \to \mu$, temos que dado qualquer $\varepsilon > 0$ existe uma ordem a partir $\bar{n}$ da qual μ_n está na vizinhança $V(\mu, F, \varepsilon)$. Mas isto significa, precisamente, que

$$\left| \int \phi \, d\mu_n - \int \phi \, d\mu \right| < \varepsilon$$

para todo $n \geq \bar{n}$. Em outras palavras, a sequência $\int \phi \, d\mu_n$ converge para $\int \phi \, d\mu$.

A recíproca afirma que se $\int \phi \, d\mu_n$ converge para $\int \phi \, d\mu$, para toda função contínua, então dado qualquer F e ε existe uma ordem a partir da qual $\mu_n \in V(\mu, F, \varepsilon)$. Para ver isso, escrevemos $F = \{\phi_1, \dots, \phi_N\}$. A hipótese garante que para cada $1 \le j \le N$ existe $\bar{n}_j$ tal que

$$\left| \int \phi_j \, d\mu_n - \int \phi \, d\mu \right| < \varepsilon \quad \text{para todo } n \ge \bar{n}_j \, .$$

Tomando $\bar{n} = \max\{\bar{n}_1, \, , \dots, \bar{n}_N\}$, temos $\mu_n \in V(\mu, F, \varepsilon)$ para $n \ge \bar{n}$. □

Outra proposição muito útil que caracteriza a convergência de medidas é dada na:

Proposição 4.3. *Assuma que a sequência μ_n converge para μ na topologia* fraca* *. Então:*

1. $\limsup_{n\to\infty} \mu_n(K) \le \mu(K)$ *para cada conjunto compacto $K \subset M$;*
2. $\liminf_{n\to\infty} \mu_n(U) \ge \mu(U)$ *para cada conjunto aberto $U \subset M$.*

Em particular, se o bordo de A tem medida zero, temos que $\lim_{n\to\infty} \mu_n(A) = \mu(A)$.

Demonstração. Seja U um aberto e vamos mostrar o item (b). Tome K um compacto em U e escolha $\phi :\to [0,1]$ uma função contínua qualquer tal que $\phi|_K \equiv 1$ e $\phi|_{U^c} \equiv 0$. Por exemplo, basta tomar $\phi(x) = d(x, U^c)/(d(x,K) + d(x, U^c))$. Então:

$$\mu(K) \le \int \phi \, d\mu = \lim_{n\to\infty} \int \phi \, d\mu_n \le \liminf_{n\to\infty} \mu_n(U).$$

Como vale que $\mu(U) = \sup_K \mu(K)$, onde o supremo é tomado sobre todos os compactos $K \subset U$ provamos o item (b). O item (a) é inteiramente análogo, observando que $\mu(K) = \inf \mu(U)$, onde o ínfimo é tomado sobre todos os abertos U contendo K.

□

As principais propriedades desta topologia de que necessitamos estão dadas no seguinte

Teorema 4.4. *$\mathcal{M}_1$ munido da topologia* fraca* *é metrizável e compacto.*

Vamos começar por demonstrar a metrizabilidade, isto é, que existe uma distância d que gera a topologia fraca* em $\mathcal{M}_1(M)$. Para isso usamos o resultado seguinte, cuja prova pode ser encontrada em [Rud87]. Como é usual, denotamos por $C^0(M)$ o espaço das funções contínuas $\phi : M \to \mathbb{R}$, munido da norma da convergência uniforme:

$$\|\phi_1 - \phi_2\| = \sup\{|\phi_1(x) - \phi_2(x)| : x \in M\}.$$

Proposição 4.5. *Se M é um espaço métrico então $C^0(M)$ tem subconjuntos enumeráveis densos.*

Logo, podemos escolher um subconjunto enumerável $\mathcal{F} = \{\phi_n : n \in \mathbb{N}\}$ denso na bola unitária do espaço $C^0(M)$. Feito isso, definimos

$$d(\mu_1, \mu_2) = \sum_{n=1}^{\infty} \frac{1}{2^n} \left| \int \phi_n \, d\mu_1 - \int \phi_n \, d\mu_2 \right|, \tag{4.1}$$

para qualquer par de medidas μ_1 e μ_2.

Proposição 4.6. *A expressão d está bem definida, é uma distância, e gera a topologia* fraca* *em $\mathcal{M}_1(M)$.*

Demonstração. Como as funções ϕ estão na bola unitária de $C^0(M)$, ou seja, $\sup|\phi| \le 1$, e as medidas μ_i são probabilidades, o termo geral da soma é limitado por $2 \cdot 2^{-n}$. Isto garante que a série em (4.1) converge.

O único passo não trivial na prova de que d é uma distância é mostrar que

$$d(\mu_1, \mu_2) = 0 \quad \Rightarrow \quad \mu_1 = \mu_2 \, .$$

A hipótese $d(\mu_1, \mu_2) = 0$ significa que $\int \phi_j \, d\mu_1 = \int \phi_j \, d\mu_2$ para toda $\phi_j \in \mathcal{F}$. Agora, dada qualquer ϕ na bola unitária de $C^0(M)$ podemos encontrar uma sequência de elementos de $\mathcal{F}$ convergindo uniformemente para ϕ. Como consequência, a igualdade continua valendo para ϕ:

$$\int \phi \, d\mu_1 = \int \phi \, d\mu_2 \tag{4.2}$$

para toda ϕ na bola unitária de $C^0(M)$. Como todo elemento de $C^0(M)$ tem algum múltiplo na bola unitária, isto implica que a igualdade (4.2) é verdadeira para *toda* função contínua ϕ. Isto quer dizer que $\mu_1 = \mu_2$, como pretendíamos mostrar.

Para provar que d gera a topologia, devemos mostrar que toda bola $B(\mu,\delta) = \{\eta \in \mathcal{M}_1(M) : d(\mu,\eta) < \delta\}$ contém alguma vizinhança $V(\mu, F, \varepsilon)$ e reciprocamente. Dado $\delta > 0$ fixemos $N \geq 1$ suficientemente grande para que

$$\sum_{n=N}^{\infty} 2^{-n} < \frac{\delta}{2}$$

e consideremos $F = \{\phi_1, \ldots, \phi_N\}$ formado pelos primeiros N elementos do subconjunto enumerável denso. Além disso, consideremos $\varepsilon = \delta/2$. Afirmamos que $V(\mu, F, \epsilon) \subset B(\mu, \delta)$. De fato

$$\begin{aligned}
\nu \in V(\mu, F, \varepsilon) &\Rightarrow \Big| \int \phi_n \, d\mu - \int \phi_n \, d\nu \Big| < \varepsilon \quad \text{para todo } 1 \leq n \leq N \\
&\Rightarrow \sum_{n=1}^{\infty} 2^{-n} \Big| \int \phi_n \, d\mu - \int \phi_n \, d\nu \Big| < \\
&\qquad < \sum_{n=1}^{N} 2^{-n}\varepsilon + \sum_{n=N+1}^{\infty} 2 \cdot 2^{-n} < \delta,
\end{aligned}$$

o que prova a nossa afirmação.

Reciprocamente, dado $F = \{\psi_1, \ldots, \psi_N\}$ e $\varepsilon > 0$, selecionemos elementos $\phi_{n_1}, \ldots, \phi_{n_N}$ distintos de $\mathcal{F}$ tais que

$$\|\phi_{n_j} - \psi_j\| < \frac{\varepsilon}{4} \quad \text{para todo } 1 \leq j \leq N.$$

Fixemos $\delta > 0$ suficientemente pequeno para que $2^{n_j}\delta < \varepsilon/4$ para todo $1 \leq j \leq N$. Afirmamos que $B(\mu, \delta) \subset V(\mu, F, \varepsilon)$. De fato

$$\begin{aligned}
\nu \in B(\mu, \delta) &\Rightarrow \sum_{n=1}^{\infty} 2^{-n} \Big| \int \phi_n \, d\mu - \int \phi_n \, d\nu \Big| < \delta \\
&\Rightarrow \Big| \int \phi_{n_j} \, d\mu - \int \phi_{n_j} \, d\nu \Big| < 2^{n_j}\delta \quad \text{para todo } 1 \leq j \leq N \\
&\Rightarrow \Big| \int \psi_j \, d\mu - \int \psi_j \, d\nu \Big| < 2^{n_j}\delta + \frac{\varepsilon}{2} < \varepsilon \quad \text{for all } 1 \leq n \leq N,
\end{aligned}$$

e isto prova a nossa afirmação. $\square$

Resta provar que $(\mathcal{M}_1, \text{fraca}^*)$ é um espaço compacto. Na demonstração vamos utilizar o seguinte resultado clássico, que diz que as integrais são os únicos operadores lineares positivos no espaço das funções contínuas. Um operador linear diz-se positivo se $\Phi(\varphi) > 0$ para toda função ϕ positiva em todo ponto. Uma demonstração deste resultado pode ser encontrada em [Rud87].

Teorema 4.7 (Riesz-Markov). *Seja $\Phi : C^0(M) \to \mathbb{R}$ qualquer operador linear positivo. Então existe uma única medida boreliana μ em M tal que*

$$\Phi(\varphi) = \int \varphi \, d\mu \quad \textit{para toda } \varphi \in C^0(M).$$

Observe que μ é uma probabilidade se e somente se $\Phi(1) = 1$, já que $\mu(M) = \int 1 \, d\mu = \Phi(1)$.

Vamos então provar que o espaço $\mathcal{M}_1$ é compacto para esta topologia. Como já sabemos que o espaço é metrizável, basta provar

Proposição 4.8. *Toda sequência $(\mu_k)_{k\in\mathbb{N}}$ em $\mathcal{M}_1(M)$ admite alguma subsequência que é convergente na topologia* fraca* *.*

Demonstração. Seja $\mathcal{F} = \{\phi_n : n \in \mathbb{N}\}$ um subconjunto enumerável denso na bola unitária de $C^0(M)$. Para cada $n \in \mathbb{N}$, a sequência de números reais $\int \phi_n \, d\mu_k$, $k \in \mathbb{N}$ é limitada por 1. Portanto, para cada $n \in \mathbb{N}$ existe uma sequência $(k_j^n)_{j\in\mathbb{N}}$ tal que

$$\int \phi_n \, d\mu_{k_j^n} \text{ converge para algum número } \Phi_n \in \mathbb{R} \text{ quando } j \to \infty.$$

Além disso, cada sequência $(k_j^{n+1})_{j\in\mathbb{N}}$ pode ser escolhida como subsequência da anterior $(k_j^n)_{j\in\mathbb{N}}$. Definamos $\ell_j = k_j^j$ para cada $j \in \mathbb{N}$. Por construção, a menos de um número finito de termos, $(\ell_j)_{j\in\mathbb{N}}$ é uma subsequência de cada uma das $(k_j^n)_{j\in\mathbb{N}}$. Logo

$$\int \phi_n \, d\mu_{\ell_j}^n \to \Phi_n \quad \text{para todo } n \in \mathbb{N}.$$

Daqui se deduz facilmente que

$$\Phi(\varphi) = \lim_j \int \varphi \, d\mu_{\ell_j} \text{ existe, para toda função } \varphi \in C^0(M). \tag{4.3}$$

De fato, suponha primeiro que φ está na bola unitária de $C^0(M)$. Dado qualquer $\varepsilon > 0$ podemos encontrar $\phi_n \in \mathcal{F}$ tal que $\|\varphi - \phi_n\| \le \varepsilon$. Então

$$\left| \int \varphi \, d\mu_{\ell_j} - \int \phi_n \, d\mu_{\ell_j} \right| \le \varepsilon$$

para todo j. Como $\int \phi_n \, d\mu_{\ell_j}$ converge (para Φ_n), seque que

$$\limsup_j \int \varphi \, d\mu_{\ell_j} - \liminf_j \int \varphi \, d\mu_{\ell_j} \le 2\varepsilon.$$

Como ε é arbitrário, concluímos que $\lim_j \int \varphi \, d\mu_{\ell_j}$ existe. Isto prova (4.3) quando a função está na bola unitária. O caso geral reduz-se imediatamente a esse, substituindo φ por $\varphi/\|\varphi\|$. Assim, completamos a prova de (4.3).

Finalmente, é claro que o operador $\Phi : C^0(M) \to \mathbb{R}$ definido por (4.3) é linear e positivo: $\Phi(\varphi) \ge \min \varphi > 0$ para todo função $\varphi \in C^0(M)$ positiva em todo ponto. Além disso, $\Phi(1) = 1$. Logo, pelo Teorema 4.7, existe alguma probabilidade boreliana μ em M tal que $\Phi(\varphi) = \int \varphi \, d\mu$ para toda função contínua φ. Agora a igualdade em (4.3) pode ser reescrita

$$\int \varphi = \lim_j \int \varphi \, d\mu_{\ell_j} \quad \text{para toda } \varphi \in C^0(M).$$

De acordo com o Lema 4.2, isto quer dizer que a subsequência $(\mu_{\ell_j})_{j \in \mathbb{N}}$ converge para μ na topologia fraca* . Isto completa a demonstração do Teorema 4.4.

□

4.3 Demonstração do Teorema de Existência

Comecemos por introduzir uma notação útil. Dado $f : M \to M$ e qualquer medida η em M denota-se por $f_*\eta$ e chama-se *imagem de η por f* a medida definida por

$$f_*\nu\big(E\big) = \nu\big(f^{-1}(E)\big) \quad \text{para cada conjunto mensurável } E \subset M.$$

Note que η é invariante por f se e somente se $f_*\eta = \eta$.

Lema 4.9. *A aplicação $f_* : \mathcal{M}_1(M) \to \mathcal{M}_1(M)$ é contínua relativamente à topologia* fraca* .

Demonstração. Para mostrarmos o lema acima, basta mostrar que se μ_n converge para μ na topologia fraca* , então para toda função contínua ϕ temos que

$$\lim_{n\to\infty} \phi \, df_*\mu_n = \int \phi \, df_*\mu.$$

De fato, se η uma medida qualquer, afirmamos que

$$\int \phi \, df_*\eta = \int \phi \circ f \, d\eta.$$

Com efeito, podemos aproximar ϕ por uma sequência de funções simples ϕ_n com $\|\phi_n\| \leq \|\phi\|$. Observe que isso implica, em particular, que $\|\phi_n \circ f\| \leq \|\phi \circ f\|$. Observe que se χ_A é função característica, então

$$\int \chi_A \, df_*\eta = \eta(f^{-1}(A)) = \int \chi_A \circ f \, d\eta.$$

Por linearidade, a igualdade acima se estende para as funções simples ϕ_n. Para finalizar, temos que pelo Teorema da Convergência Dominada,

$$\int \phi \, df_*\eta = \lim_{n\to\infty} \int \phi_n \, df_*\eta = \lim_{n\to\infty} \int \phi_n \circ f \, d\eta = \int \phi \circ f \, d\eta,$$

o que termina a prova da afirmação. Para completar a prova do Lema, basta observar que a função $\phi \circ f$ também é contínua, uma vez que f é contínua. Assim,

$$\lim_{n\to\infty} \int \phi \, df_*\mu_n = \lim_{n\to\infty} \int \phi \circ f \, d\mu_n \int \phi \circ f \, d\mu = \int \phi \, df_*\mu,$$

como queríamos provar.

$\square$

Voltando a prova do Teorema de Existência, considere qualquer probabilidade ν em M: por exemplo, a medida de Dirac em um ponto qualquer. Forme a sequência de probabilidades

$$\mu_n = \frac{1}{n} \sum_{j=0}^{n-1} f_*^j \nu \tag{4.4}$$

onde $f_*^j\nu$ é a imagem de ν pelo iterado f^j. Pelo Teorema 4.4, esta sequência tem algum ponto de acumulação: existe alguma subsequência $(n_j)_{j\in\mathbb{N}}$ e alguma probabilidade $\mu \in \mathcal{M}_1(M)$ tais que

$$\mu = \lim_k \mu_{n_k} = \lim_k \frac{1}{n_k}\sum_{j=0}^{n_k-1} f_*^j\nu\,. \tag{4.5}$$

Agora é suficiente provar o seguinte

Lema 4.10. *Todo ponto de acumulação de uma sequência $(\mu_n)_{n\in\mathbb{N}}$ é uma probabilidade invariante por f.*

Demonstração. A partir de (4.5), e usando o Lema 4.9, obtemos que

$$f_*\mu = f_*\big(\lim_k \frac{1}{n_k}\sum_{j=0}^{n_k-1} f_*^j\nu\big) = \lim_k f_*\big(\frac{1}{n_k}\sum_{j=0}^{n_k-1} f_*^j\nu\big) = \lim_k \frac{1}{n_k}\sum_{j=1}^{n_k} f_*^j\nu\,.$$

A expressão do lado direito pode ser reescrita como

$$\lim_k \big(\frac{1}{n_k}\sum_{j=0}^{n_k-1} f_*^j\nu - \nu + f_*^{n_k}\nu\big).$$

Afirmamos que $\lim_k \frac{1}{n_k}\nu = 0$ e $\lim_k \frac{1}{n_k} f_*^{n_k}\nu = 0$. A primeira afirmação é óbvia, e para a segunda basta observar que

$$\frac{1}{n_k} f_*^{n_k}\nu(E) = \frac{1}{n_k}\nu(f^{-n_k}(E)) \le \frac{1}{n_k}$$

para todo conjunto mensurável $E \subset F$. Deste modo obtemos que

$$f_*\mu = \lim_k \frac{1}{n_k}\sum_{j=0}^{n_k-1} f_*^j\nu = \mu$$

e portanto μ é invariante por f. □

Isto completa a demonstração do Teorema de Existência 4.1.

Corolário 4.11 (Teorema de Recorrência de Birkhoff). *Se $f : M \to M$ é uma transformação contínua num espaço métrico compacto então f tem algum ponto recorrente.*

Demonstração. Pelo Teorema 4.1, existe alguma probabilidade f-invariante μ. Por outro lado, todo espaço métrico compacto admite uma base enumerável de abertos (verifique!). Portanto, podemos aplicar o Teorema 2.3, para concluir que μ-quase todo ponto é recorrente. Em particular, o conjunto dos pontos recorrentes é não vazio, conforme foi afirmado. $\square$

4.4 Exercícios

4.1. Prove a seguinte generalização do Lema 4.10: Seja $f : M \to M$ uma transformação contínua num espaço compacto, ν uma probabilidade em M e $(I_n)_n$ uma sequência de intervalos de números naturais tais que $\# I_n$ converge para infinito quando n vai para infinito. Então qualquer ponto de acumulação da sequência

$$\mu_n = \frac{1}{\# I_n} \sum_{j \in I_n} f_*^j \nu$$

é uma probabilidade f-invariante.

4.2. Dizemos que uma sequência $(\mu_n)_{n\in\mathbb{N}}$ de probabilidades converge pontualmente (ou fortemente) para $\mu \in \mathcal{M}_1(M)$

$$\mu_n(E) \to \mu(E) \quad \text{para todo conjunto mensurável } E \subset M.$$

1. Mostre que se $(\mu_n)_{n\in\mathbb{N}}$ converge pontualmente para μ então também converge para μ na topologia fraca* . Mostre, através de um exemplo, que a recíproca é falsa.

2. Mostre que $(\mu_n)_{n\in\mathbb{N}}$ converge para μ na topologia fraca* se e somente se $\mu_n(E) \to \mu(E)$ para todo conjunto mensurável $E \subset M$ cujo bordo ∂E satisfaz $\mu(\partial E) = 0$.

Dica para (2): Dado o mensurável E e $\varepsilon > 0$ encontre funções contínuas φ_1 e φ_2 tais que $\varphi_1 \geq \mathcal{X}_E \geq \varphi_2$ e $\int \varphi_1 \, d\mu - \int \varphi_2 \, d\mu < \varepsilon$.

4.3. Fixe um subconjunto enumerável denso $\mathcal{F} = \{\phi_n : n \in \mathbb{N}\}$ da bola unitária de $C^0(M)$. Mostre que uma sequência $(\mu_k)_{k\in\mathbb{N}}$ de probabilidades em M converge na topologia fraca para alguma $\mu \in \mathcal{M}_1(M)$ se e somente se, para todo $n \in \mathbb{N}$,

$$\int \phi_n \, d\mu_k \text{ converge para } \int \phi_n \, d\mu.$$

4.4. Seja $f_1, f_2, \dots, f_N : M \to M$ uma família finita qualquer de transformações contínuas num espaço métrico compacto que comutam entre si: $f_i \circ f_j = f_j \circ f_i$ para todo i e todo j em $\{1, 2, \dots, N\}$. Prove que existe alguma probabilidade μ que é invariante por f_i para todo $i \in \{1, 2, \dots, N\}$.

Definição 4.12. Dizemos que uma transformação $f : M \to M$ é *unicamente ergódica* se admite exatamente uma probabilidade invariante.

Os exercícios a seguir tratam de transformações unicamente ergódicas. Esta terminologia é justificada pelo Exercício 4.7 abaixo, que afirma que nesse caso a probabilidade invariante é necessariamente ergódica. No que segue suporemos que M é um espaço métrico compacto e $f : M \to M$ é contínua.

4.5. Seja R_α é uma rotação irracional do círculo. Mostre que R_α é unicamente ergódica.

4.6. Seja $f : M \to M$ uma transformação unicamente ergódica. Mostre que se $\varphi : S^1 \to \mathbb{R}$ é uma função contínua qualquer, então:

$$\tilde{\varphi}(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \varphi(R_\alpha^j(x))$$

existe em *todo* ponto e, de fato, o limite é uniforme. Justifique que $\tilde{\varphi}$ é constante em todo ponto.

Dica: Verifique que a sequência do lado direito é equicontínua e use o teorema de Ascoli-Arzela.

4.7. Mostre que f é uma transformação unicamente ergódica se e somente se

$$\lim \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f^j(x)) = \int \varphi \, d\mu$$

para toda função contínua $\varphi : M \to \mathbb{R}$ e todo $x \in M$. Obtenha que, se uma transformação é unicamente ergódica então a sua probabilidade invariante é ergódica.

Capítulo 5

Teorema Ergódico de Birkhoff

O teorema fundamental da Teoria Ergódica afirma que, para qualquer subconjunto mensurável e para quase todo ponto, existe um *tempo médio de permanência* da órbita do ponto nesse conjunto. Este resultado é devido a von Neumann, que provou um enunciado mais fraco, e sobretudo a Birkhoff, que o provou na forma definitiva que iremos estudar.

Em muitos casos, esse tempo médio de permanência é precisamente igual à medida do subconjunto, ou seja, órbitas típicas passam em cada subconjunto um tempo que é exatamente igual à "importância"que a probabilidade invariante atribui ao conjunto. Isto é o que se chama de *ergodicidade*, uma propriedade que remonta a Boltzmann, e que estudaremos mais tarde.

5.1 Enunciados e comentários

Comecemos por explicar o que entendemos por tempo médio de permanência de uma órbita num conjunto. Dado $x \in M$ e um conjunto mensurável $E \subset M$, vamos tomar um certo número (grande) de iterados iniciais da órbita de x e vamos considerar a fração desses iterados que estão em E:

$$\tau_n(E,x) = \frac{1}{n}\#\big\{j \in \{0,1,\dots,n-1\} : f^j(x) \in E\big\}.$$

Observe que isto é o mesmo que

$$\tau_n(E,x) = \frac{1}{n}\sum_{j=0}^{n-1} \mathcal{X}_E(f^j(x)),$$

onde $\mathcal{X}_E$ designa a função característica do conjunto E, isto é, $\mathcal{X}_E(x) = 1$ se $x \in E$ e $\mathcal{X}_E(x) = 0$ caso contrário.

Em seguida, fazemos n ir para infinito e chamamos *tempo médio de permanência* da órbita de x em E o limite destas frações:

$$\tau(E,x) = \lim_{n\to\infty} \tau_n(E,x).$$

Em geral, este limite pode não existir. Iremos ver um exemplo desse fato daqui a pouco. No entanto, o teorema ergódico afirma que, relativamente a qualquer probabilidade invariante, o limite realmente existe para quase todo ponto:

Teorema 5.1. *Seja $f : M \to M$ uma transformação mensurável e μ uma probabilidade invariante por f. Dado qualquer conjunto mensurável $E \subset M$, o tempo médio de permanência $\tau(E,x)$ existe em μ-quase todo ponto $x \in M$. Além disso,*

$$\int \tau(E,x)\, d\mu(x) = \mu(E).$$

Antes de passarmos à demonstração deste resultado notável, e a algumas das suas aplicações, vamos fazer alguns comentários relacionados. O primeiro deles é que se $\tau(E,x)$ existe para um certo ponto $x \in M$ então

$$\tau(E,f(x)) = \tau(E,x). \tag{5.1}$$

De fato, por definição,

$$\begin{aligned}\tau(E, f(x)) &= \lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} \mathcal{X}_E(f^j(x)) \\ &= \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \mathcal{X}_E(f^j(x)) - \frac{1}{n}\big[\mathcal{X}_E(x) - \mathcal{X}_E(f^n(x))\big] \\ &= \tau(E, x) + \lim_{n\to\infty} \frac{1}{n}\big[\mathcal{X}_E(x) - \mathcal{X}_E(f^n(x))\big]\end{aligned}$$

Como a função característica é limitada, o último limite é igual a zero. Isto prova a igualdade (5.1).

O teorema ergódico pode ser enunciado de modo um pouco mais geral:

Teorema 5.2. *Seja $f : M \to M$ uma transformação mensurável e μ uma probabilidade invariante por f. Dada qualquer função integrável $\varphi : M \to \mathbb{R}$ o limite*

$$\tilde{\varphi}(x) = \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \varphi(f^j(x))$$

existe em μ-quase todo ponto $x \in M$. Além disso,

$$\int \tilde{\varphi}(x)\, d\mu(x) = \int \varphi(x)\, d\mu(x).$$

Observe que o Teorema 5.1 é o caso particular φ = função característica $\mathcal{X}_E$ do conjunto E. Este enunciado mais geral pode ser provado usando uma versão um pouco mais elaborada do argumento da seção 5.2, que não apresentaremos aqui.

5.2 Demonstração do teorema ergódico

A estratégia da prova é a seguinte. Seja $E \subset M$ um conjunto mensurável qualquer. Para cada $x \in M$, definimos

$$\begin{aligned}\overline{\tau}(E, x) &= \limsup \frac{1}{n}\#\Big\{j \in \{0, \dots, n-1\} : f^j(x) \in E\Big\} \\ \underline{\tau}(E, x) &= \liminf \frac{1}{n}\#\Big\{j \in \{0, \dots, n-1\} : f^j(x) \in E\Big\}.\end{aligned}$$

Note que, para todo $x \in M$,

$$\overline{\tau}(E, f(x)) = \overline{\tau}(E, x) \quad \text{e} \quad \underline{\tau}(E, f(x)) = \underline{\tau}(E, x) \tag{5.2}$$

A justificação é análoga à da relação (5.1).

O principal passo da demonstração consiste em mostrar que

$$\overline{\tau}(E, x) = \underline{\tau}(E, x) \quad \text{para } \mu\text{-quase todo ponto } x. \tag{5.3}$$

É claro que $\overline{\tau}(E, x)$ é sempre maior ou igual que $\underline{\tau}(E, x)$. Portanto, para mostrar (5.3) será suficiente que provemos

$$\int \overline{\tau}(E, x)\, d\mu(x) \leq \mu(E) \leq \int \underline{\tau}(E, x)\, d\mu(x). \tag{5.4}$$

Vamos provar a primeira desigualdade em (5.4). A segunda segue de um argumento inteiramente análogo [1].

Fixemos qualquer $\varepsilon > 0$. Por definição de lim sup, para cada $x \in M$ existem inteiros $t \geq 1$ tais que

$$\frac{1}{t} \#\Big\{ j \in \{0, \ldots, t-1\} : f^j(x) \in E \Big\} \geq \overline{\tau}(E, x) - \varepsilon. \tag{5.5}$$

Representaremos por $t(x)$ o menor inteiro com esta propriedade. Para tornar a demonstração mais transparente, consideraremos primeiro o caso particular em que a função $x \mapsto t(x)$ é limitada, isto é,

Caso particular: Existe $T \in \mathbb{N}$ tal que $t(x) \leq T$ para todo $x \in M$.

Dado qualquer $x \in M$, definimos uma sequência $x_0, x_1, \ldots, x_s$ de pontos em M e uma sequência $t_0, t_1, \ldots, t_s$ de números naturais, do seguinte modo:

1. Primeiramente, tomamos $x_0 = x$.
2. Supondo que x_i já foi definido, tomamos $t_i = t(x_i)$ e $x_{i+1} = f^{t_i}(x_i)$.
3. Terminamos quando encontramos x_s tal que $t_0 + t_1 + \cdots + t_{s-1} + t_s \geq n$.

[1] Alternativamente, a segunda desigualdade pode ser deduzida da primeira, aplicada ao complementar E^c, observando que $\mu(E) = 1 - \mu(E^c)$ e $\underline{\tau}(E, x) = 1 - \overline{\tau}(E^c, x)$.

Note que todo x_i é iterado do ponto x: de fato $x_i = f^{t_0+\cdots+t_{i-1}}(x)$. Aplicando (5.2) concluímos que $\overline{\tau}(E, x_i) = \overline{\tau}(E, x)$ para todo i. A definição de $t(x_i)$ implica que, dos t_i primeiros iterados de x_i, pelo menos

$$t_i\big(\overline{\tau}(E, x_i) - \varepsilon\big) = t_i\big(\overline{\tau}(E, x) - \varepsilon\big) \tag{5.6}$$

estão em E. Isto vale para cada $i = 0, 1, \ldots, s-1$. Portanto, pelo menos

$$(t_0 + t_1 + \cdots + t_{s-1})(\overline{\tau}(E, x) - \varepsilon)$$

dos n primeiros iterados de x, estão em E. Além disso, a última regra na definição das nossas sequências implica que

$$t_0 + t_1 + \cdots + t_{s-1} \geq n - t_s \geq n - T.$$

Deste modo, mostramos que pelo menos $(n - T)(\overline{\tau}(E, x) - \varepsilon)$ dos n primeiros iterados de x estão em E. Em outras palavras,

$$\sum_{j=0}^{n-1} \mathcal{X}_E(f^j(x)) \geq (n - T)(\overline{\tau}(E, x) - \varepsilon) \tag{5.7}$$

para todo $x \in M$ e todo $n \geq 1$. Integrando a relação (5.7), obtemos que

$$\sum_{j=0}^{n-1} \int \mathcal{X}_E(f^j(x))\, d\mu(x) \geq (n - T) \int \overline{\tau}(E, x)\, d\mu(x) - (n - T)\varepsilon.$$

Todas as parcelas no membro da esquerda são iguais a $\mu(E)$, uma vez que a probabilidade μ é invariante por f. Portanto, esta desigualdade pode ser escrita como

$$n\mu(E) \geq (n - T) \int \overline{\tau}(E, x)\, d\mu(x) - (n - T)\varepsilon.$$

Dividindo os dois termos por n e fazendo n ir para infinito, concluímos que

$$\mu(E) \geq \int \overline{\tau}(E, x)\, d\mu(x) - \varepsilon$$

Como $\varepsilon > 0$ é qualquer, isto implica a primeira desigualdade em (5.4). Isto termina a demonstração neste caso.

Caso geral: Vamos indicar as modificações que devem ser feitas relativamente ao caso particular.

Dado $\varepsilon > 0$, começamos por fixar $T \geq 1$ suficientemente grande, de modo que a medida do

$$B = \{y \in M : t(y) > T\}$$

seja menor que ε. Em seguida, na definição das sequências substituímos a regra 2 por

2a. Se $t(x_i) \leq T$, tomamos $t_i = t(x_i)$ e $x_{i+1} = f^{t_i}(x_i)$.

2b. Se $t(x_i) > T$, tomamos $t_i = 1$ e $x_{i+1} = f(x_i)$.

As regras 1 e 3 permanecem inalteradas. A estimativa referente a (5.6) continua válida, *para os valores de i aos quais se aplica a regra 2a:*

$$\sum_{j=0}^{t_i-1} \mathcal{X}_E(f^j(x_i)) \geq t_i\big(\overline{\tau}(E,x) - \varepsilon\big).$$

É claro que esta desigualdade implica a seguinte:

$$\sum_{j=0}^{t_i-1} \mathcal{X}_E(f^j(x_i)) \geq t_i\big(\overline{\tau}(E,x) - \varepsilon\big) - \sum_{j=0}^{t_i-1} \mathcal{X}_B(f^j(x_i)). \tag{5.8}$$

A vantagem é que (5.8) *é válida também para os valores de i aos quais se aplica a regra 2b.* De fato, nesse caso tem-se $t_i = 1$, o membro da esquerda é maior ou igual que zero e o membro da direita é menor que zero, uma vez que $\overline{\tau}(E,x)$ é sempre menor ou igual que 1. Isso significa que, no lugar de (5.7), tem-se

$$\sum_{j=0}^{n-1} \mathcal{X}_E(f^j(x)) \geq (n-T)(\overline{\tau}(E,x) - \varepsilon) - \sum_{j=0}^{n-1} \mathcal{X}_B(f^j(x)).$$

Integrando, como fizemos anteriormente, obtemos

$$n\mu(E) \geq (n-T)\int \overline{\tau}(E,x)\,d\mu(x) - (n-T)\varepsilon - n\mu(B).$$

Dividindo por n e fazendo $n \to \infty$, deduzimos que (lembre que $\mu(B) < \varepsilon$)

$$\mu(E) \geq \int \overline{\tau}(E, x)\, d\mu(x) - \varepsilon - \mu(B) \geq \int \overline{\tau}(E, x)\, d\mu(x) - 2\varepsilon.$$

Como $\varepsilon > 0$ é arbitrário, segue que

$$\mu(E) \geq \int \overline{\tau}(E, x)\, d\mu(x).$$

Isto completa a demonstração do Teorema 5.1.

5.3 Exercícios

5.1. Considere a transformação $f : M \to M$, $f(x) = 10x - [10x]$ introduzida na seção 3.1. Considere

$$x = 0,335533335555555533333333333333335\ldots.$$

Ou seja: a expansão decimal de x consiste de blocos de 3s e 5s, alternados, cada bloco (exceto o segundo) com duas vezes mais dígitos que o anterior. Considere também $E = [0,3,\ 0,4)$. Mostre que

$$\tau_2(E,x) = 1, \quad \tau_8 = \frac{3}{4}, \quad \ldots \quad \tau_{2^{2k-1}}(E,x) \to \frac{2}{3}\,,$$

enquanto que

$$\tau_4(E,x) = \frac{1}{2}, \quad \tau_{16} = \frac{3}{8}, \quad \ldots \quad \tau_{2^{2k}}(E,x) \to \frac{1}{3}\,,$$

e portanto o tempo médio de permanência da órbita de x em E não existe.

5.2. Mostre que, para qualquer função integrável φ, a média temporal $\tilde{\varphi}$ satisfaz $\tilde{\varphi} \circ f = \tilde{\varphi}$ em μ-quase todo ponto.

Capítulo 6

Ergodicidade

Uma transformação $f : M \to M$ diz-se *ergódica* para uma probabilidade invariante μ (também dizemos que a medida μ é ergódica para f, ou que o sistema (f, μ) é ergódico) se as médias temporais dadas pelo Teorema de Birkhoff 5.2 coincidem em quase todo ponto com as respectivas médias espaciais:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f^j(x)) = \int \varphi \, d\mu \,,$$

para toda função μ-integrável $\varphi : M \to \mathbb{R}$ e μ-quase todo $x \in M$.

Na próxima proposição vamos reescrever esta condição de várias maneiras equivalentes, para ajudar a entender o seu significado. Um conjunto mensurável $A \subset M$ diz-se *invariante* se $f^{-1}(A) = A$. Uma função mensurável $\psi : M \to \mathbb{R}$ diz-se *invariante* se $\psi \circ f = \psi$.

Proposição 6.1. *Seja μ uma probabilidade invariante de uma transformação $f : M \to M$ mensurável. As seguintes condições são equivalentes:*

1. *O sistema (f, μ) é ergódico.*

2. *Para todo subconjunto invariante A tem-se $\mu(A) = 0$ ou $\mu(A) = 1$.*

3. *Toda função invariante ψ é constante num conjunto de medida total.*

Demonstração. **(1) implica (2):** Considere $\varphi = \mathcal{X}_A$. Por um lado, a hipótese (1) significa que

$$\tilde{\varphi}(x) = \int \varphi \, d\mu = \mu(A)$$

para quase todo $x \in M$. Por outro lado, como A é invariante, temos que $x \in A$ se e somente se $f(x) \in A$. Isto implica que $\varphi(f^j(x)) = \varphi(x)$ para todo $j \geq 0$ e para todo x. Portanto,

$$\tilde{\varphi}(x) = \varphi(x) = \mathcal{X}_A(x)$$

para todo $x \in M$. Como a função característica só toma os valores 0 e 1, estas duas igualdades implicam que $\mu(A) = 0$ ou $\mu(A) = 1$, como é afirmado em (2).

(2) implica (3): Seja ψ uma função invariante qualquer. Então, a pré-imagem $\psi^{-1}(I)$ de qualquer intervalo $I \subset \mathbb{R}$ é um conjunto invariante. Portanto, pela hipótese (2), essa pré-imagem tem medida zero ou um. Como o intervalo I é qualquer, isto prova que ψ é constante num conjunto com probabilidade μ total.

(3) implica (1): Seja φ uma função integrável qualquer. Como vimos no exercício 5.2, a média temporal $\tilde{\varphi}$ é uma função invariante. Logo, pela hipótese (3), $\tilde{\varphi}$ é constante em quase todo ponto. Então, usando o teorema ergódico,

$$\tilde{\varphi}(x) = \int \tilde{\varphi} \, d\mu = \int \varphi \, d\mu$$

em quase todo ponto. Isto é, o sistema é ergódico. □

6.1 Exemplos e aplicações

Nesta seção descrevemos diversos exemplos de sistemas ergódicos.

6.1.1 Expansão decimal

Considere a transformação $f : [0,1] \to [0,1]$, $f(x) = 10x - [10x]$ da seção 3.1. Afirmamos que f é ergódica para a medida de Lebesgue

μ. Tendo em vista a proposição 6.1, para mostrar isto só temos que provar que se A é um conjunto invariante com medida positiva então A tem medida total.

Suponhamos então que A é invariante e $\mu(A) > 0$. O ingrediente principal é o teorema de derivação 1.25. No nosso caso, como estamos tratando com subconjuntos de $\mathbb{R}$, a condição (1.2) torna-se

$$\lim_{\varepsilon\to 0} \inf \Big\{ \frac{\mu(I\cap A)}{\mu(I)} : I \subset (a-\varepsilon, a+\varepsilon) \text{ intervalo contendo } a \Big\} = 1\,. \tag{6.1}$$

Fixemos um ponto de densidade $a \in A$ qualquer. Consideremos a sequência de intervalos

$$I_k = \Big(\frac{m_k}{10^k}, \frac{m_k+1}{10^k}\Big), \quad m_k \in \mathbb{Z}, k \in \mathbb{N},$$

que contêm o ponto a. Como a é um ponto de densidade de A, a propriedade (6.1) implica que

$$\frac{\mu(I_k \cap A)}{\mu(I_k)} \to 1 \quad \text{quando } k \to \infty.$$

Observe também que cada f^k é uma bijeção *afim* de I_k sobre o intervalo $(0,1)$. Isso tem a seguinte consequência, que é crucial para o nosso argumento:

$$\frac{\mu(f^k(E_1))}{\mu(f^k(E_2))} = \frac{\mu(E_1)}{\mu(E_2)} \tag{6.2}$$

para quaisquer subconjuntos mensuráveis E_1 e E_2 de I_k. Aplicando este fato a $E_1 = I_k \cap A$ e $E_2 = I_k$ obtemos que

$$\frac{\mu\big(f^k(I_k \cap A)\big)}{\mu\big((0,1)\big)} = \frac{\mu(I_k \cap A)}{\mu(I_k)}\,.$$

Claro que $\mu\big((0,1)\big) = 1$. Além disso, como estamos supondo que A é invariante, $f^k(I_k \cap A)$ está contido em A. Deste modo obtemos que

$$\mu(A) \geq \frac{\mu(I_k \cap A)}{\mu(I_k)}\,.$$

Como a sequência do lado direito converge para 1, segue que $\mu(A) = 1$, como queríamos demonstrar. Ficou provado que a transformação f é ergódica para a medida de Lebesgue μ.

Em seguida vamos dar uma aplicação deste fato no contexto da Teoria dos Números. Dizemos que um número $x \in \mathbb{R}$ é *balanceado* se todo dígito aparece com a mesma frequência, $1/10$, na sua expansão decimal. É fácil dar exemplos de números balanceados. Mas em geral é muito difícil decidir se um dado número irracional é balanceado ou não. Por exemplo, não é sabido até hoje se o número π é balanceado.

No entanto, a conclusão da seção anterior nos permite deduzir que quase todo número é balanceado:

Proposição 6.2. *O conjunto dos números $x \in \mathbb{R}$ não balanceados tem medida de Lebesgue nula.*

Demonstração. Como o fato de ser balanceado é independente da parte inteira do número, só precisamos mostrar que quase todo $x \in [0,1]$ é balanceado. Considere $f : [0,1] \to [0,1]$ definida por $f(x) = 10x - [10x]$. Para cada dígito $j \in \{0,1,\ldots,9\}$ considere o intervalo $E_j = [j/10, (j+1)/10)$. Recorde que se $x = 0, a_0 a_1 \cdots a_k a_{k+1} \cdots$ então $f^k(x) = 0, a_k a_{k+1} \cdots$. Portanto, $f^k(x) \in E_j$ se e somente se o k-ésimo dígito da expansão decimal de x é igual a j. Consequentemente, o tempo médio de permanência $\tau(E_j, x)$ é exatamente a frequência do dígito j na expansão decimal de x. Usando o teorema ergódico e o fato de que a transformação é ergódica para a medida de Lebesgue μ, concluímos que para cada $j \in \{0,1,\ldots,9\}$ existe um subconjunto B_j de M com $\mu(B_j) = 1$ tal que

$$\tau(E_j, x) = \mu(E_j) = \frac{1}{10} \quad \text{para todo } x \in B_j.$$

Então $B = B_0 \cap B_1 \cap \cdots \cap B_9$ também tem $\mu(B) = 1$, e todo número $x \in B$ é balanceado. $\square$

6.1.2 Deslocamentos (“shifts”) de Bernoulli

Vamos agora voltar a discussão dos deslocamentos de Bernoulli, introduzidos na Secção 3.3 do Capítulo 3. Mostraremos que as medidas de Bernoulli são ergódicas. Para isso, a seguinte propriedade das medidas de Bernoulli vai ser útil :

Lema 6.3. *Se A e B são elementos da álgebra $\mathcal{B}_0$, isto é, uniões finitas de cilindros disjuntos, então tem-se*

$$\mu(A \cap f^{-m}(B)) = \mu(A)\mu(f^{-m}(B)) = \mu(A)\mu(B),$$

para todo m suficientemente grande.

Demonstração. Expliquemos porque esta propriedade é verdadeira quando A e B são cilindros, $A = [k, l; a_k, \ldots, a_l]$ e $B = [u, v; b_u, \ldots, b_v]$. Para cada m tem-se $f^{-m}(B) = [u+m, v+m; b_u, \ldots, b_v]$. Escolhendo m suficientemente grande garantimos que $u + m > l$ e, então,

$$\begin{aligned} A \cap f^{-m}(B) &= \{\underline{\alpha} : \alpha_k = a_k, \ldots, \alpha_l = a_l, \alpha_{u+m} = b_u, \ldots, \alpha_{v+m} = b_v\} \\ &= \bigcup [k, v+m; a_k, \ldots, a_l, c_{l+1}, \ldots, c_{u+m-1}, b_u, \ldots, b_v], \end{aligned}$$

onde a união é sobre todos os valores possíveis de $c_{l+1}, \ldots, c_{u+m-1}$. Usando (3.6), concluímos que $\mu(A \cap f^m(B)) = \mu(A)\mu(B)$. Isto prova o lema quando os conjuntos envolvidos são cilindros. O caso geral segue pelo fato de μ ser finitamente aditiva. $\square$

Proposição 6.4. *Seja $f : M \to M$ um deslocamento e μ uma medida de Bernoulli em M, como antes. Então o sistema (f, μ) é ergódico.*

Demonstração. Seja A um conjunto mensurável invariante qualquer. Queremos mostrar que $\mu(A) = 0$ ou $\mu(A) = 1$. Para tornar a ideia da prova mais clara, comecemos por um caso particular: suponhamos que A está na álgebra $\mathcal{B}_0$ das uniões finitas de cilindros disjuntos dois-a-dois. Nesse caso podemos aplicar o lema anterior, com $B = A$. Concluímos que $\mu(A \cap f^{-m}(A)) = \mu(A)^2$ sempre que tomemos m suficientemente grande. Mas, como A é invariante, $f^{-m}(A) = A$ para todo m. Então a igualdade anterior quer dizer que $\mu(A) = \mu(A)^2$, o que só pode acontecer se $\mu(A) = 0$ ou $\mu(A) = 1$.

Agora vamos fazer a prova quando $A \in \mathcal{B}$ é um conjunto invariante qualquer. A ideia é aproximar A por elementos da álgebra $\mathcal{B}_0$, usando o Teorema de Aproximação 1.11: dado qualquer $\varepsilon > 0$ existe $A_0 \in \mathcal{B}_0$ tal que $\mu(A \Delta A_0) < \varepsilon$. Escolha m como no caso anterior, de modo que

$$\mu(A_0 \cap f^{-m}(A_0)) = \mu(A_0)\mu(f^{-m}(A_0)) = \mu(A_0)^2. \tag{6.3}$$

Observe que

$$\begin{aligned}(A \cap f^{-m}(A))\Delta(A_0 \cap f^{-m}(A_0)) &\subset (A\Delta A_0) \cup (f^{-m}(A)\Delta f^{-m}(A_0)) \\ &\subset (A\Delta A_0) \cup f^{-m}(A\Delta A_0).\end{aligned}$$

Isto, junto com o fato de que μ é invariante por f, implica que

$$\big|\mu(A \cap f^{-m}(A)) - \mu(A_0 \cap f^{-m}(A_0))\big| \leq 2\mu(A\Delta A_0) < 2\varepsilon. \tag{6.4}$$

Além disso,

$$\big|\mu(A)^2-\mu(A_0)^2\big| \leq \big|(\mu(A)+\mu(A_0))(\mu(A)-\mu(A_0))\big| \leq 2\big|\mu(A)-\mu(A_0)\big| < 2\varepsilon. \tag{6.5}$$

Juntando as relações (6.3), (6.4), (6.5), concluímos que $|\mu(A)-\mu(A)^2| < 4\varepsilon$. Como ε é arbitrário, deduzimos que $\mu(A) = \mu(A)^2$ e então, do mesmo modo que antes, concluímos que $\mu(A) = 0$ ou $\mu(A) = 1$. □

6.1.3 Rotação irracional no círculo

Para nós o círculo S^1 será o conjunto dos números complexos com módulo igual a 1. Dado $\alpha \in \mathbb{R}$, a *rotação* de ângulo α é a multiplicação pelo número complexo $e^{\alpha i}$

$$R_\alpha : S^1 \to S^1, \quad R_\alpha(z) = e^{\alpha i} z\,.$$

É claro que R_α preserva o comprimento dos intervalos (segmentos) de S^1. Usando o Lema 3.2 se deduz que a medida de Lebesgue (comprimento de arco) é invariante por qualquer R_α .

O comportamento dinâmico e ergódico de R_α depende muito da natureza de α, como vamos ver. Dizemos que a rotação é *irracional* se o número $\alpha/(2\pi)$ é irracional, e dizemos que a rotação é racional no caso contrário.

A recíproca é muito mais interessante:

Proposição 6.5. *Se R_α é rotação irracional então R_α é ergódica para a medida de Lebesgue.*

Vamos mencionar duas demonstrações diferentes deste fato. A primeira, que detalharemos a seguir, usa fatos simples de análise de Fourier. A segunda, que deixaremos como exercício, é baseada num argumento de ponto de densidade semelhante ao que usamos no caso da expansão decimal.

Seja μ a medida de Lebesgue no círculo. Chama-se $L^2(\mu)$ o espaço das funções [1] mensuráveis $\psi : S^1 \to \mathbb{C}$ cujo quadrado é integrável:

$$\int |\psi|^2 \, d\mu < \infty.$$

É claro que este espaço contém todas as funções mensuráveis limitadas e, em particular, todas as funções características de conjuntos mensuráveis. Outro fato de que necessitamos é que a família de funções $\{\phi_k(z) = z^k : k \in \mathbb{Z}\}$ é uma base (de Hilbert) desse espaço: dada qualquer $\varphi \in L^2(\mu)$ existe uma única sequência $(c_k)_{k\in\mathbb{Z}}$ de números complexos tais que

$$\varphi(z) = \sum_{k\in\mathbb{Z}} c_k z^k \quad \text{para quase todo } z \in S^1.$$

Demonstração. Pela proposição 6.1, basta mostrar que toda função integrável φ que é invariante é constante em μ-quase todo ponto. Observe que se φ é integrável, então automaticamente $\varphi \in L^2(\mu)$ (verifique! Utilize que μ é finita). Usando a expansão de Fourier $\varphi(z) = \sum_{k\in\mathbb{Z}} c_k z^k$, a condição de ser invariante $\varphi \circ R_\alpha = \varphi$ se escreve

$$\sum_{k\in\mathbb{Z}} c_k e^{ki\alpha} z^k = \sum_{k\in\mathbb{Z}} c_k z^k$$

Por unicidade dos coeficientes da expansão em série de Fourier, obtemos que

$$c_k(e^{ki\alpha} - 1) = 0 \quad \text{para todo } k \in \mathbb{Z}.$$

A hipótese de que a rotação é irracional significa que $e^{ki\alpha} - 1 \neq 0$ para todo $k \neq 0$, e portanto, $c_k = 0$ para todo $k \neq 0$. Ou seja, $\varphi(z) = c_0$ para μ-quase todo $z \in S^1$, como queríamos provar. $\square$

De fato as rotações irracionais satisfazem uma propriedade muito mais forte do que ergodicidade: elas são *unicamente ergódicas*, o que quer dizer que têm uma única probabilidade invariante (que é a medida de Lebesgue, claro).

[1] Quando lidamos com $L^2(\mu)$ sempre identificamos funções que diferem apenas num conjunto de medida nula.

Observação 6.6. A noção de rotação irracional se estende para dimensões maiores. Dado qualquer $d \geq 1$ chamamos *d-toro* o produto $\mathbb{T}^d = S^1 \times \cdots \times S^1$ do círculo por si mesmo d vezes. A rotação de ângulo $\alpha = (\alpha_1, \ldots, \alpha_d)$ é a aplicação $R_\alpha : \mathbb{T}^d \to \mathbb{T}^d$, $R_\alpha(z_1, \ldots, z_d) = (e^{i\alpha_1} z_1, \ldots, e^{i\alpha_d} z_d)$. A rotação é irracional se os números $\alpha_j/(2\pi)$ são incomensuráveis:

$$m_0 + m_1 \frac{\alpha_1}{2\pi} + \cdots + m_d \frac{\alpha_d}{2\pi} = 0 \quad \Rightarrow \quad m_0 = m_1 = \cdots = m_d = 0,$$

quaisquer que sejam os inteiros $m_0, m_1, \ldots, m_d$. Usando uma versão multidimensional das idéias anteriores, se prova que uma rotação é ergódica se e somente se ela é irracional.

6.1.4 Transformação de Gauss

Como vimos na seção 3.4, a transformação de Gauss $G(x) = 1/x - [1/x]$ admite uma probabilidade invariante que é equivalente à medida de Lebesgue, nomeadamente,

$$\mu(E) = \frac{1}{\log 2} \int_E \frac{dx}{1+x}$$

Temos também que o sistema (G, μ) é ergódico. Este fato pode ser demonstrado pelo mesmo tipo de argumento que usamos na seção 6.1.1. Vamos esboçar o argumento neste caso, explicando qual é a principal dificuldade adicional.

Seja A um conjunto invariante com medida positiva. Em primeiro lugar, continua sendo verdade que para quase todo ponto $a \in [0,1]$ existe uma sequência de intervalos I_k contendo a e tais que f^k envia I_k bijetivamente e diferenciavelmente sobre $(0,1)$. O diâmetro desses intervalos converge para zero. Logo, tomando para a um ponto de densidade qualquer de A, temos que

$$\frac{\mu(I_k \cap A)}{\mu(I_k)} \to 1 \quad \text{quando } k \to +\infty. \tag{6.6}$$

Por outro lado embora f^k seja uma bijeção restrita a cada I_k, ela não é afim. Por essa razão não temos o análogo da relação (6.2) neste caso. Esta dificuldade é contornada através do seguinte resultado, que é um exemplo de controle de *distorção*: é muito importante notar que a constante K é independente de k, I_k , E_1 , e E_2 .

Lema 6.7. *Existe uma constante $K > 1$ tal que para todo $k \geq 1$, todo intervalo I_k tal que G restrita a I_k é uma bijeção diferenciável, tem-se*

$$\frac{\mu(f^k(E_1))}{\mu(f^k(E_2))} \leq K \frac{\mu(E_1)}{\mu(E_2)}$$

para quaisquer subconjuntos mensuráveis E_1 e E_2 de I_k .

Antes de demonstrarmos o Lema 6.7, explicamos como a ergodicidade de (G, μ) pode ser obtida a partir dele. Observe que $f^k(I_k \cap A^c) = A^c$, porque o conjunto A é invariante. Lembre também que $f^k(I_k) = (0,1)$, que tem medida total. Tomando $E_1 = I_k \cap A^c$ e $E_2 = I_k$ no lema 6.7, concluímos que

$$\mu(A^c) \leq \frac{\mu(f^k(I_k \cap A^c)}{\mu(A^c)} \leq K \frac{\mu(I_k \cap A^c)}{\mu(I_k} .$$

De acordo com (6.6), a expressão do lado direito converge para zero quando $k \to \infty$. Logo $\mu(A^c) = 0$, como queríamos demonstrar.

Prova do Lema 6.7. Usaremos os seguintes fatos sobre a transformação f que podem ser facilmente verificados pelo leitor:

1. Para todo $x \in (0,1)$ vale que $|f'(x)| > 1$ e $|(f^2)'(x)| \geq 4$.

2. Existe $C_1 > 0$ tal que $|\frac{f''(x)}{f'(x)}| < C_1$.

Observe que a partir do item (1) acima, podemos mostrar que se $x, y \in I_k$ então

$$|f^i(x) - f^i(y)| \leq \frac{1}{2^{k-i}} |f^k(x) - f^k(y)| \text{ se } i = 0, 1, 2, \dots k. \tag{6.7}$$

Observe também que se $x, y \in I_k$ temos que

$$|\log \frac{(f^k)'(x)}{(f^k)'(y)}| \leq \sum_{i=0}^{k-1} |\log f'(f^i(x)) - \log f'(f^i(y))|.$$

O item (2) nos garante que a função $x \to \log f'(x)$ tem derivada limitada por C, logo pelo Teorema do Valor Médio temos que $|\log f'(a) - \log f'(b)| \leq C_1|a-b|$. Aplicando este fato na desigualdade acima e observando a equação 6.7:

$$|\log \frac{(f^k)'(x)}{(f^k)'(y)}| \leq C_1 \sum_{i=0}^{k-1} |f^i(x) - f^i(y)| \leq \sum_{i=0}^{k-1} \frac{1}{2^{k-i}} C_1 |f^k(x) - f^k(y)| \leq C_2,$$

onde C_2 é uma constante propriamente escolhida. Logo, tomando $K = \exp C_2$, vem que para todos $x, y \in I_k$ vale:

$$\frac{(f^k)'(x)}{(f^k)'(y)} < C_3.$$

Note que a constante C_3 escolhida não depende de k nem de I_k. Observe ainda que se $A \subset [0,1]$ é um conjunto mensurável, então

$$\frac{1}{2\log 2} m(A) \leq \mu(A) \leq \frac{1}{\log 2} m(A),$$

onde m representa a medida de Lebesgue de [0,1].

Assim, para concluir a prova do Lema 6.7, basta observar que se E_1 e E_2 são subconjuntos mensuráveis de I_k, então:

$$\frac{\mu(f^k(E_1))}{\mu(f^k(E_2))} = 2(\log 2)^2 \frac{m(f^k(E_1))}{m(f^k(E_2))} \leq \frac{\int_{E_1} (f^k)'(x)\, dm}{\int_{E_2} (f^k)'(y)\, dm} \leq$$
$$2(\log 2)^2 (C_3)^2 \frac{m(E_1)}{m(E_2)} \leq 4(\log 2)^4 C_3 \frac{\mu(E_1)}{\mu(E_2)}.$$

Assim, basta tomar $K = 4(\log 2)^4 (C_3)^2$ e o lema está provado.

$\square$

6.1.5 Máquina de somar (“adding machine”)

A máquina de somar modela sistemas tais como o contador de quilometragem de um carro ou o registro de consumo de gás (em algumas cidades): a dinâmica consiste em fazer avançar o contador de uma unidade. A principal diferença com relação à realidade é que este contador idealizado comporta infinitos dígitos.

Fixe $d \geq 2$, que representa a base de numeração (por exemplo, $d = 10$). Consideramos o espaço M de todas as sequências

$$\beta = (\beta_0\,, \beta_1\,, \beta_2\,, \ldots, \beta_{k-1}\,, \beta_k\,, \ldots)$$

com $\beta_j \in \{0, 1, \ldots, d-1\}$. Munimos este espaço da seguinte distância

$$d(\beta, \beta') = 2^{-N(\beta,\beta')} \quad \text{onde } N(\beta, \beta') = \min\{j \geq 0 : \beta_j \neq \beta'_j\}.$$

Também consideramos a transformação $f : M \to M$ "soma uma unidade":

- Para toda sequência com $\beta_0 < d - 1$, definimos

  $$f(\beta_0\,, \beta_1\,, \beta_2\,, \ldots, \beta_{k-1}\,, \beta_k\,, \ldots) = (\beta_0{+}1, \beta_1\,, \beta_2\,, \ldots, \beta_{k-1}\,, \beta_k\,, \ldots).$$

- Se $\beta_0 = d - 1$ mas $\beta_1 < d - 1$, definimos

  $$f(\beta_0\,, \beta_1\,, \beta_2\,, \ldots, \beta_{k-1}\,, \beta_k\,, \ldots) = (0, \beta_1{+}1, \beta_2\,, \ldots, \beta_{k-1}\,, \beta_k\,, \ldots).$$

- Mais geralmente, se $\beta_0 = \cdots = \beta_{k-1} = d - 1$ mas $\beta_k < d - 1$, definimos

  $$f(\beta_0\,, \beta_1\,, \beta_2\,, \ldots, \beta_{k-1}\,, \beta_k\,, \ldots) = (0, 0, 0, \ldots, 0, \beta_k + 1, \ldots).$$

- Se $\beta_j = d - 1$ para todo $j \geq 0$, definimos

  $$f(\beta_0\,, \beta_1\,, \beta_2\,, \ldots, \beta_{k-1}\,, \beta_k\,, \ldots) = (0, 0, 0, \ldots, 0, 0, \ldots).$$

M munido da distância definida em (4.1) é um espaço métrico compacto, e a transformação f é contínua nesse espaço. O exercício 6.7 pede para mostrar que f é unicamente ergódica e para calcular a (única) probabilidade invariante.

6.2 Propriedades de medidas ergódicas

Fixemos uma transformação $f : M \to M$ qualquer. Lembre que uma medida ν diz-se absolutamente contínua com relação a outra medida μ se $\mu(E) = 0$ implica $\nu(E) = 0$. O próximo lema afirma que probabilidades ergódicas são minimais para a relação $\ll$:

Lema 6.8. *Se μ e ν são probabilidades invariantes tais que μ é ergódica e ν é absolutamente contínua com relação a μ então $\mu = \nu$.*

Demonstração. Seja $\varphi : M \to \mathbb{R}$ uma função mensurável limitada qualquer, e seja

$$\tilde{\varphi}(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f^j(x))$$

a sua média temporal. Como μ é invariante e ergódica, a média temporal é constante

$$\tilde{\varphi}(x) = \int \varphi \, d\mu$$

para μ-quase todo ponto. Segue que isto é verdade para ν-quase todo ponto, já que $\nu \ll \mu$. Em particular,

$$\int \tilde{\varphi} \, d\nu = \int \varphi \, d\mu \, .$$

Por outro lado, pelo teorema ergódico,

$$\int \tilde{\varphi} \, d\nu = \int \varphi \, d\nu \, .$$

Portanto, as integrais de φ com relação a μ e em relação a ν coincidem, qualquer que seja a função mensurável limitada φ. Logo, considerando funções características, $\mu = \nu$. $\square$

Naturalmente, se μ_1 e μ_2 são probabilidades invariantes com respeito à f a probabilidade $\mu_1 + t(\mu_2 - \mu_1)$ ainda é invariante. Isso siginfica que o conjunto das probabilidades invariantes é um conjunto *convexo*. Veremos que dentro deste conjunto, as medidas ergódicas desempenham um papel destacado:

Definição 6.9. Seja X um conjunto convexo. Um ponto $p \in X$ é dito *extremal*, se para quaisquer $x, y \in X$ e $t \in [0, 1]$, $x + t(y - x) = p$ implica que $t = 0$ ou 1.

O lema seguinte afirma que uma probabilidade invariante é ergódica se e somente se é ponto extremal no conjunto das probabilidades invariantes:

Lema 6.10. *Uma probabilidade invariante μ é ergódica se e somente se não é possível escrevê-la na forma*

$$\mu = c_1\mu_1 + c_2\mu_2$$

com c_1, c_2 maiores que zero e μ_1, μ_2 probabilidades invariantes distintas.

Demonstração. Para provar a parte "se", suponha que μ não seja ergódica. Então existe algum conjunto invariante A com $0 < \mu(A) < 1$. Defina μ_1 e μ_2 como sendo as restrições normalizadas de μ a A e ao seu complementar, respectivamente:

$$\mu_1(E) = \frac{\mu(E \cap A)}{\mu(A)} \qquad \mu_2(E) = \frac{\mu(E \cap A^c)}{\mu(A^c)}\,.$$

Como A e A^c são conjuntos invariantes e μ é medida invariante, μ_1 e μ_2 são também probabilidades invariantes. Além disso, $\mu = \mu(A)\mu_1 + \mu(A^c)\mu_2$ e portanto μ não é extremal.

Para provar a recíproca, suponha que μ é ergódica e temos $\mu = c_1\mu_1 + c_2\mu_2$ com c_1, $c_2 > 0$. É claro que $\mu(E) = 0$ implica $\mu_1(E) = \mu_2(E) = 0$, ou seja, μ_1 e μ_2 são absolutamente contínuas com relação a μ. Logo, pelo lema 6.8, $\mu_1 = \mu = \mu_2$. Isto prova que μ é extremal. $\square$

Em seguida vamos mostrar que medidas ergódicas distintas "vivem" em subconjuntos disjuntos do espaço M:

Lema 6.11. *Sejam $\mu_1, \ldots, \mu_N$ probabilidades invariantes e ergódicas, todas distintas. Então existem subconjuntos mensuráveis $P_1, \ldots, P_N$ invariantes disjuntos tais que*

$$\mu_j(P_k) = \begin{cases} 1 & \text{se } j = k \\ 0 & \text{se } j \neq k \end{cases}$$

Demonstração. Fixe qualquer par j, k de números distintos em $\{1, \ldots, N\}$. Pelo lema 6.8, a medida μ_j não pode ser absolutamente contínua em relação a μ_k . Em outras palavras, existe algum subconjunto mensurável E tal que $\mu_j(E) > 0$ mas $\mu_k(E) = 0$. Então

$$\mu_j\Big(\bigcup_{j=0}^{\infty} f^{-j}(E)\Big) \geq \mu_j(E) > 0. \quad \text{e} \quad \mu_k\Big(\bigcup_{j=0}^{\infty} f^{-j}(E)\Big) = 0$$

Defina $P_{j,k} = \bigcap_{m=0}^{\infty} \bigcup_{j=m}^{\infty} f^{-j}(E)$. Como a sequência de conjuntos na interseção é decrescente com m,

$$\mu_j\big(P_{j,k}\big) = \lim_{m\to\infty} \mu_j\big(\bigcup_{j=m}^{\infty} f^{-j}(E)\big) \tag{6.8}$$

e, analogamente para μ_k. Como as medidas μ_j e μ_k são invariantes, e

$$\bigcup_{j=m}^{\infty} f^{-j}(E) = f^{-m}\big(\bigcup_{j=0}^{\infty} f^{-j}(E)\big),$$

a sequência no lado direito de (6.8) é constante. Concluímos que

$$\mu_j(P_{j,k}) = \mu_j\big(\bigcup_{j=0}^{\infty} f^{-j}(E)\big) > 0 \quad \text{e} \quad \mu_k(P_{j,k}) = \mu_k\big(\bigcup_{j=0}^{\infty} f^{-j}(E)\big) = 0.$$

Além disso, $P_{j,k}$ é um conjunto invariante por f. Portanto $\mu_j(P_{j,k}) = 1$, uma vez que μ_j é ergódica. Agora defina

$$\tilde{P}_j = \bigcap_{k\neq j} P_{j,k} \quad \text{e} \quad P_j = \tilde{P}_j \setminus \bigcup_{k\neq j} \tilde{P}_k\,.$$

Primeiramente, $\mu_j(\tilde{P}_j) = 1$ e $\mu_k(\tilde{P}_j) = 0$ para todo $k \neq j$. Segue que $\mu_j(P_j) = 1$ e $\mu_k(P_j) = 0$ para todo $k \neq j$. Além disso, os P_j são disjuntos dois-a-dois. □

6.3 Teorema de decomposição ergódica

Na sequência dos resultados da seção anterior, é natural perguntar se toda medida invariante é uma combinação linear de medidas ergódicas. O teorema que vamos enunciar nesta seção afirma que a resposta é afirmativa, exceto que o número de "parcelas" nesta combinação não é necessariamente finito, nem mesmo enumerável, em geral.

Teorema 6.12. *Seja $f : M \to M$ uma transformação contínua num espaço compacto. Então existe um conjunto mensurável $M_0 \subset M$, uma partição $\mathcal{P}$ de M_0 e uma família de probabilidades $\{\nu_P : P \in \mathcal{P}\}$ satisfazendo*

- $\nu_P(P) = 1$ *para todo elemento* P *de* $\mathcal{P}$*;*
- *a aplicação* $P \mapsto \nu_P$ *é mensurável;*
- *toda* ν_P *é invariante e ergódica para* f*;*

tais que, dada qualquer probabilidade f*-invariante* μ*, o conjunto* M_0 *satisfaz* $\mu(M_0) = 1$ *e, além disso,*

$$\mu(E) = \int \nu_P(E)\, d\hat{\mu}(P) \quad \textit{para todo conjunto mensurável } E \subset M \tag{6.9}$$

onde $\hat{\mu}$ *é a medida projeção de* μ *em* $\mathcal{P}$*.*

A relação (6.9) significa que μ é uma combinação convexa das várias probabilidades ergódicas ν_P, em que cada ν_P entra com "coeficiente"igual a $\hat{\mu}(P)$. Dada qualquer partição $\mathcal{P}$ de M fica definida a projeção natural $\pi : M \to \mathcal{P}$ que associa a cada ponto $x \in M$ o elemento $P(x)$ da partição que o contém. Isto permite definir o que é um subconjunto mensurável da partição: $\mathcal{Q} \subset \mathcal{P}$ é mensurável se e somente

$$\pi^{-1}(\mathcal{Q}) = \text{união dos } P \in \mathcal{Q}$$

é um subconjunto mensurável de M. É fácil ver que esta definição está correta: a família dos subconjuntos mensuráveis é uma σ-álgebra em $\mathcal{P}$. A medida projeção de μ está definida nesta σ-álgebra, por

$$\hat{\mu}(\mathcal{Q}) = \mu(\pi^{-1}(\mathcal{Q})).$$

A demonstração do Teorema 6.12 está fora dos objetivos do presente texto e pode ser encontrada em [Wal82] ou [Mañ87].

6.4 Exercícios

6.1. Considere o espaço $M = \{1, 2, \ldots, d\}^{\mathbb{Z}}$ das sequências com valores num conjunto $\{1, 2, \ldots, d\}$. Fixe qualquer número $\theta \in (0, 1)$. Para cada $\underline{\beta} = (\beta_n)_{n\in\mathbb{Z}}$ e $\underline{\gamma} = (\gamma_n)_{n\in\mathbb{Z}}$ em M, defina

$$N(\underline{\beta}, \underline{\gamma}) = \max\left\{N \geq 0 : \beta_n = \gamma_n \text{ para todo } n \in \mathbb{Z} \text{ com } |n| < N\right\}$$

e $d(\underline{\beta}, \underline{\gamma}) = \theta^{N(\underline{\beta}, \underline{\gamma})}$. Verifique que d é uma métrica em M e gera a mesma topologia que a família dos cilindros. Em particular, (M, d) é um espaço métrico compacto. Tem-se um resultado análogo para $M = \{1, 2, \ldots, d\}^{\mathbb{N}}$.

6.2. Suponha que R_α é uma rotação irracional.

1. Mostre que a órbita $\{R_\alpha^n(z) : n \in \mathbb{Z}\}$ de todo $z \in S^1$ é densa em S^1.

2. Seja A um conjunto invariante com medida positiva. Mostre que *nenhum* ponto de S^1 é ponto de densidade de A^c. Conclua que $\mu(A) = 1$.

   Dica: considere um ponto de densidade de A e use o item (1).

6.3. A rotação R_α é racional se e somente se $e^{\alpha i}$ é uma raiz da unidade, isto é, se existe $k \neq 0$ tal que $e^{ki\alpha} = 1$.

6.4. Se R_α é rotação racional então R_α não é ergódica para a medida de Lebesgue.

No exercício a seguir propomos outra demonstração para a proposição 6.5:

6.5. Suponha que R_α é uma rotação irracional.

1. Mostre que a órbita $\{R_\alpha^n(z) : n \in \mathbb{Z}\}$ de todo $z \in S^1$ é densa em S^1.

2. Seja A um conjunto invariante com medida positiva. Mostre que *nenhum* ponto de S^1 é ponto de densidade de A^c. Conclua que $\mu(A) = 1$.

   Dica: considere um ponto de densidade de A e use o item (1).

6.6. Suponha que R_α é uma rotação irracional.

1. Seja $\varphi : S^1 \to \mathbb{R}$ uma função contínua qualquer. Mostre que

$$\tilde{\varphi}(x) = \lim_{n \to \infty} \frac{1}{n} \sum_{j=0}^{n-1} \varphi(R_\alpha^j(z))$$

existe em *todo* ponto e, de fato, o limite é uniforme. Justifique que $\tilde{\varphi}$ é constante em todo ponto.

Dica: Verifique que a sequência do lado direito é equicontínua e use o teorema de Ascoli-Arzela.

2. Deduza que R_α tem uma única probabilidade invariante.

6.7. Seja $f : M \to M$ uma máquina de somar, definida na secção 6.1.5.

1. Seja $\varphi : M \to \mathbb{R}$ uma função contínua qualquer. Mostre que

$$\tilde{\varphi}(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f^j(z))$$

existe em todo ponto e o limite é uniforme.

2. Justifique que $\tilde{\varphi}$ é constante em todo ponto e deduza que f tem uma única probabilidade invariante.

3. Calcule essa probabilidade, encontrando uma expressão explícita para a medida de qualquer subconjunto $[b_0, b_1, \ldots, b_k]$ das sequências β com $\beta_0 = b_0$, $\beta_1 = b_1$, $\ldots$, $\beta_k = b_k$.

6.8 (Teorema de Kac). Seja μ uma medida ergódica para uma transformação $f : M \to M$ e A um conjunto com $\mu(A) > 0$. Considere $n_A : A \to \mathbb{N} \cup \{+\infty\}$ como o menor número $n_A(x) > 0$ tal que $f^{n_A(x)}(x) \in A$. Caso este número não exista, definimos $n_A(x) = +\infty$.

1. Mostre que n_A é integrável com respeito a μ.

2. Mostre que se $\mu_A(B) = \frac{\mu(A\cap B)}{\mu(A)}$ então:

$$\int_A n_A(x)\, d\mu_A(x) = \frac{1}{\mu(A)}.$$

6.9. Seja $f : M \to M$ definida no espaço topológico M tal que existe alguma medida ergódica μ tal que para todo A aberto, $\mu(A) > 0$. Mostre que f e transitiva e a órbita de μ-quase todo ponto é densa.

Capítulo 7

Aplicações em Teoria dos Números

Neste capítulo apresentamos duas aplicações da Teoria Ergódica no domínio da Teoria dos Números: o Teorema de S. Szemerédi [Sze75] sobre existência de progressões aritméticas dentro de subconjuntos suficientemente "densos" do conjunto $\mathbb{Z}$ dos números inteiros, e o teorema de H. Weyl [Wey16] sobre equidistribuição da parte fracionária dos valores de funções polinomiais restritas a $\mathbb{Z}$.

7.1 Teorema de Szemerédi

A nossa apresentação é inspirada por Furstenberg [Fur81], onde o leitor pode encontrar muita informação adicional sobre este tema.

7.1.1 Densidade superior

Chamamos *intervalo* do conjunto $\mathbb{Z}$ dos números inteiros qualquer subconjunto I da forma $\{n \in \mathbb{Z} : a \leq n < b\}$, para quaisquer $a \leq b$ em $\mathbb{Z}$. O seu *cardinal* é $\#I = b - a$.

Definição 7.1. A *densidade superior* $D_s(S)$ de um subconjunto S

de $\mathbb{Z}$ é

$$D_s(S) = \limsup_{\#I\to\infty} \frac{\#(S\cap I)}{\#I}$$

onde I representa qualquer intervalo em $\mathbb{Z}$. Do mesmo modo se define a *densidade inferior* $D_i(S)$, trocando limite superior por limite inferior.

Em outras palavras, $D_s(I)$ é o maior número D tal que existe uma sequência de intervalos $I_j \subset \mathbb{Z}$ tais que

$$\#I_j \to \infty \quad \text{e} \quad \frac{\#(S\cap I_j)}{\#I_j} \to D$$

e $D_i(S)$ é o menor número nessas condições. Note que $0 \le D_i(S) \le D_s(S) \le 1$. No Exercício 7.1 também veremos que $D_i(S) = D_s(\mathbb{Z}\setminus S)$ para todo $S \subset \mathbb{Z}$.

Exemplo 7.2. Seja S o conjunto dos números pares. Dado qualquer intervalo $I \subset \mathbb{Z}$, temos que $\#(S\cap I) = \#I/2$ se o cardinal de I é par e $\#(S\cap I) = \#(I\pm 1)/2$ se o cardinal de I é ímpar, onde o sinal $\pm$ é positivo se o menor elemento de I é um número par, e é negativo caso contrário. Desta observação segue, imediatamente, que $D_s(S) = D_i(S) = 1/2$.

Exemplo 7.3. Seja S o seguinte subconjunto de $\mathbb{Z}$:

$$\{1,3,4,7,8,9,13,14,15,16,21,22,23,24,25,31,32,33,34,35,36,42,\ldots\}.$$

Isto é, para cada $k \ge 1$ incluímos em S um bloco de k inteiros consecutivos e omitimos os k inteiros seguintes. Este conjunto contém intervalos com comprimento arbitrariamente grande. Portanto $D_s(S) = 1$. Por outro lado, o complementar de S também contém intervalos com comprimento arbitrariamente grande. Portanto, $D_i(S) = 1 - D_s(\mathbb{Z}\setminus S) = 0$.

Exemplo 7.4. Seja S o seguinte subconjunto de $\mathbb{Z}$:

$$\{1,3,4,5,6,9,10,11,12,14,15,16,17,18,\\ 22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,42,\ldots\}.$$

Ou seja, para cada $k \geq 1$ incluímos em S um bloco de k^2 inteiros consecutivos e depois excluímos os k inteiros seguintes. Neste caso temos

$$\lim_{n\to\infty} \frac{\#(S \cap [1, 2, \ldots, n])}{n} = 1.$$

Isto implica $D_s(S) = 1$. Mas, tal como no caso anterior, $D_i(S) = 0$.

7.1.2 Enunciados

Nos anos 30, Erdös e Turan [ET36] conjecturaram que todo subconjunto de $\mathbb{Z}$ com densidade superior positiva contém sequências aritméticas finitas com comprimento arbitrariamente grande. Esta conjectura foi demonstrada por Szemerédi [Sze75], quase quatro décadas mais tarde:

Teorema 7.5 (Szemerédi). *Se S é um subconjunto de $\mathbb{Z}$ com densidade superior positiva, então para todo $k \in \mathbb{N}$ existem $m \in \mathbb{Z}$ e $n \in \mathbb{N}$ tais que m, $m+n$, $m+2n$, $\ldots$, $m+kn$ pertencem a S.*

Em geral, não podemos esperar que S contenha progressões aritméticas com comprimento infinito, como mostram os Exemplos 7.3 e 7.4.

A demonstração original do Teorema 7.5 usa argumentos combinatórios bastante intrincados. No entanto, poucos anos depois Furstenberg [Fur77] deu uma nova demonstração, utilizando idéias de Teoria Ergódica. Na verdade, ele deduziu o Teorema 7.5 de uma generalização do Teorema 2.1 para famílias de transformações que comutam entre si:

Teorema 7.6 (Recorrência Simultânea de Poincaré). *Sejam $f_i : M \to M$, $i = 1, 2, \ldots, k$ transformações que preservam uma probabilidade μ em M e tais que $f_i \circ f_j = f_j \circ f_i$ para todo $i, j = 1, 2, \ldots, k$. Então, para qualquer conjunto $E \subset M$ tal que $\mu(E) > 0$, existe algum $n \geq 1$ tal que*

$$\mu\big(E \cap f_1^{-n}(E) \cap f_2^{-n}(E) \cap \cdots \cap f_k^{-n}(E)\big) > 0.$$

Em outras palavras, este teorema afirma que existe algum tempo n tal que os iterados de um subconjunto com medida positiva de pontos de E, por todas as transformações f_i, regressam a E *simultaneamente* nesse momento n.

A demonstração do Teorema 7.6 escapa ao âmbito deste texto. Mas, na Seção 7.1.6, explicaremos porquê ele implica o Teorema 7.5. Além disso, vamos discutir versões um pouco mais fracas destes resultados, que chamamos teorema de van der Waerden e teorema de Recorrência Simultânea de Birkhoff, respectivamente.

O teorema de van der Waerden [vdW27] afirma que dada qualquer partição do conjunto $\mathbb{Z}$ num número finito de subconjuntos, algum desses subconjuntos deve conter progressões aritméticas com comprimento arbitrariamente grande:

Teorema 7.7 (van der Waerden). *Sejam $S_1, S_2, \ldots, S_q$ subconjuntos dois-a-dois disjuntos de $\mathbb{Z}$ tais que $S_1 \cup S_2 \cup \cdots \cup S_q = \mathbb{Z}$. Dado qualquer $k \geq 1$ existe algum S_i que contém alguma progressão aritmética com comprimento $k+1$. Em particular, algum elemento S_j da partição contém progressões aritméticas com comprimento arbitrariamente grande.*

Na Seção 7.1.3 veremos que este resultado é uma consequência simples do Teorema 7.5. Também veremos, na Seção 7.1.4, que ele pode ser deduzido da seguinte extensão do Teorema 4.11:

Teorema 7.8 (Recorrência Simultânea de Birkhoff). *Seja M um espaço métrico compacto e $f_i : M \to M$, $i = 1, 2, \ldots, k$ transformações contínuas tais que $f_i \circ f_j = f_j \circ f_i$ para todo $i, j = 1, 2, \ldots, k$. Então existe algum $x \in M$ e alguma sequência $n_j \to \infty$ tal que*

$$f_i^{n_j}(x) \to x \quad \textit{quando } j \to \infty, \quad \textit{para todo } i = 1, 2, \ldots, k.$$

Em outras palavras, as trajetórias de x por todas as transformações recorrem para x *simultaneamente* nos momentos n_j.

Na Seção 7.1.5 veremos que o Teorema 7.8 é uma consequência simples do Teorema 7.6. Aliás, como veremos na Seção 7.1.6, este último teorema também implica o Teorema 7.5. Portanto, o diagrama a seguir resume as relações lógicas entre os quatro enunciados, que serão discutidas nas próximas seções:

$$\begin{array}{ccc}
\text{T. Szemerédi 7.5} & \Leftarrow & \text{T. R. S. Poincaré 7.6} \\
\Downarrow & & \Downarrow \\
\text{T. van der Waerden 7.7} & \Leftarrow & \text{T. R. S. Birkhoff 7.8.}
\end{array}$$

7.1.3 T. de Szemerédi implica T. van der Waerden

Seja S_1, S_2, ..., S_N uma partição finita de $\mathbb{Z}$ qualquer. De acordo com o Exercício 7.1, pelo menos um dos elementos S_i da partição deve ter densidade superior positiva. Aplicando o Teorema 7.5 a $S = S_i$ concluímos que ele contém progressões aritméticas com comprimento arbitrariamente grande. Isto prova o Teorema 7.7.

7.1.4 T. de Birkhoff implica T. de van der Waerden

Vamos começar por traduzir o Teorema 7.7 num enunciado sobre o deslocamento ("shift") $f : M \to M$ no espaço $M = \{1, 2, \ldots, q\}^{\mathbb{Z}}$ das sequências bilaterais com valores em $\{1, 2, \ldots, q\}$. Observe que cada sequência $\underline{\alpha} = (\alpha_n)_{n\in\mathbb{Z}}$ em M define uma partição de $\mathbb{Z}$ em subconjuntos

$$S_i = \{n \in \mathbb{Z} : \alpha_n = i\}, \quad i = 1, 2, \ldots, q$$

e, reciprocamente, toda partição de $\mathbb{Z}$ em q subconjuntos determine uma sequência $\underline{\alpha} \in M$. Portanto, o teorema pode ser reformulado do seguinte modo: para todo $\underline{\alpha} \in M$ e todo $k \geq 1$, existem $m \in \mathbb{Z}$ e $n \geq 1$ tais que

$$\alpha_m = \alpha_{m+n} = \cdots = \alpha_{m+nk}. \tag{7.1}$$

Para provarmos este fato, vamos munir M da métrica $d(\underline{\beta}, \underline{\gamma}) = \theta^{-N(\underline{\beta},\underline{\gamma})}$,

$$N(\underline{\beta}, \underline{\gamma}) = \max\left\{N \geq 0 : \beta_n = \gamma_n \text{ para todo } n \in \mathbb{Z} \text{ com } |n| < N\right\},$$

que foi definida no Exercício 6.1, sendo θ um número qualquer em $(0, 1)$. Note que

$$d(\underline{\beta}, \underline{\gamma}) < 1 \quad \text{se e somente se } \alpha_0 = \beta_0. \tag{7.2}$$

Como o espaço métrico (M, d) é compacto, o fecho $A = \overline{\left\{f^n(\underline{\alpha}) : n \in \mathbb{Z}\right\}}$ da trajetória de $\underline{\alpha}$ é também um compacto, para a métrica induzida. Lembre que o deslocamento $f : M \to M$ é definido por

$$f\left((\alpha_n)_{n\in\mathbb{Z}}\right) = (\alpha_{n+1})_{n\in\mathbb{Z}}. \tag{7.3}$$

Consideremos as transformações $f_1 = f$, $f_2 = f^2$, ..., $f_k = f^k$ definidas de A em A. É claro que as f_i comutam entre si. Portanto, podemos aplicar o Teorema 7.8 e concluir desta maneira que existe $\underline{\sigma} \in A$ e uma sequência $n_j \to \infty$ tal que

$$f_i^{n_j}(\underline{\sigma}) \to \underline{\sigma} \quad \text{para todo } i = 1, 2, \ldots, k.$$

Observe que $f_i^{n_j} = f^{i\,n_j}$. Em particular, podemos fixar $n = n_j$ tal que os iterados $f^n(\underline{\sigma})$, $f^{2n}(\underline{\sigma})$, ..., $f^{kn}(\underline{\sigma})$ estão todos a distância menor que $1/3$ do ponto $\underline{\sigma}$. Logo, os pontos

$$\underline{\sigma}, f^n(\underline{\sigma}), f^{2n}(\underline{\sigma}), \ldots, f^{kn}(\underline{\sigma})$$

estão todos a distância menor que $2/3$ uns dos outros. Então, como $\underline{\sigma}$ está no fecho A da órbita de $\underline{\alpha}$, podemos encontrar $m \in \mathbb{Z}$ tal que $f^m(\underline{\alpha})$ está tão próximo de $\underline{\sigma}$ que os pontos

$$f^m(\underline{\alpha}), f^{m+n}(\underline{\alpha}), f^{m+2n}(\underline{\alpha}), \ldots, f^{m+kn}(\underline{\alpha})$$

estão a distância menor que 1 uns dos outros. Tendo em conta a observação (7.2) e a definição (7.3) da transformação f, isto quer dizer que

$$\alpha_m = \alpha_{m+n} = \cdots = \alpha_{m+kn},$$

como pretendíamos provar. Isto completa a demonstração do teorema de van der Waerden a partir do teorema de Recorrência Simultânea de Birkhoff.

De fato, a conclusão do Teorema 7.7 ainda vale para partições de subconjuntos finitos de $\mathbb{Z}$, desde que sejam suficientemente grandes:

Teorema 7.9 (van der Waerden). *Dados $k \geq 1$ e $q \geq 2$ existe $N \geq 1$ tal que, dada qualquer partição do intervalo $\{1, 2, \ldots, N\}$ em q subconjuntos, algum desses subconjuntos contém progressões aritméticas com comprimento $k + 1$.*

É fácil ver que o Teorema 7.9 implica o Teorema 7.7. No Exercício 7.2 veremos que a recíproca também é verdadeira.

7.1.5 T. de Poincaré implica T. de Birkhoff

Comecemos por lembrar (Exercício 4.4) que se $f_i : M \to M$, $i = 1, 2, \ldots, k$ são transformações contínuas num espaço métrico compacto que comutam entre si, então existe alguma probabilidade invariante μ comum a todas essas transformações.

Em seguida, observemos que o Teorema 7.6 tem a seguinte consequência:

Corolário 7.10. *Sejam $f_i : M \to M$, $i = 1, 2, \ldots, k$ transformações que preservam uma probabilidade μ em M e tais que $f_i \circ f_j = f_j \circ f_i$ para todo $i, j = 1, 2, \ldots, k$. Então, para qualquer conjunto $E \subset M$ tal que $\mu(E) > 0$, e para quase todo $x \in E$ existe uma sequência $n_j \to \infty$ tal que $f_i^{n_j}(x) \in E$ para todo $i = 1, \ldots, k$ e todo $j \geq 1$.*

Deixaremos a prova desta proposição como exercício para o leitor (Exercício 7.6). Compare também com o Exercício 2.1. Este corolário será útil na:

Prova do Teorema de Recorrência Múltipla de Birkhoff. Considere uma base enumerável de abertos U_j de M com o diâmetro de U_j indo a zero quando $j \uparrow \infty$. Seja μ alguma medida invariante simultaneamente para todos os f_i (note que o Exercício 4.4 nos garante a existência de alguma destas medidas).

Para cada j representamos por D_j o conjunto dos pontos $x \in U_j$ tais que existe $n > j$ tal que $f_i^n(x) \in U_j$ para *todo* $i = 1, 2, \ldots, k$. Observe que de acordo com o Corolário 7.10, o conjunto $U_j \setminus D_j$ tem medida μ igual a zero, uma vez que quase todo ponto retorna simutaneamente a U_j em algum momento. Consequentemente, como $\{U_j\}_{j \in \mathbb{N}}$ é uma cobertura de M, temos que o conjunto:

$$D = \bigcap_{n=1}^{\infty} \bigcup_{m \geq n} D_m,$$

tem medida $\mu(D) = 1$. Em particular, $D \neq \emptyset$.

Mostraremos que todo ponto $x \in D$ é simultaneamente recorrente para as transformações $f_1, f_2, \ldots, f_k$. Ora, se $x \in D \Rightarrow x \in \bigcup_{m \geq n} D_m$ para todo $n \in \mathbb{N}$. Logo, existe $m \geq n$ tal que $x \in D_m$. De acordo com a definição de D_m, existe algum $n_m > m$ tal que $f_i^{n_m}(x) \in U_m$ para todo $i = 1, 2, \ldots, k$. Logo, $d(x, f_i^{n_m}(x)) \to 0$, uma vez que os pontos x e $f_i^{m_n}(x)$ pertencem a U_m e o diâmetro de U_m vai a zero quando $m \uparrow \infty$. Isto encerra a prova do Teorema de Recorrência Múltipla de Birkhoff.

$\square$

7.1.6 Prova do Teorema de Szemerédi

Mostraremos nesta secção como deduzir o Teorema de Szemerédi (Teorema 7.5) a partir do Teorema de Recorrência Simultânea (Teorema 7.6). Novamente, utilizaremos o dicionário entre partições de $\mathbb{Z}$ e sequências de inteiros, como já fizemos na prova do Teorema de Van der Waerden (Teorema 7.7).

Prova do Teorema de Szemerédi. Considere S um conjunto com densidade superior positiva qualquer. Vamos associar a S uma sequência $\alpha = (\alpha_n)_{n \in \mathbb{Z}} \in M = \{0, 1\}^{\mathbb{Z}}$ definida por:

$$\alpha_n = 1 \Leftrightarrow n \in S.$$

Como S possui densidade superior positiva, existe $c > 0$ e uma sequência de intervalos $I_n = [a_n, b_n)$ de $\mathbb{Z}$ com $\lim \# I_n = \infty$ e tais que

$$D_s(S) = \lim_{\# I_n \to \infty} \frac{\#(S \cap I_n)}{\# I_n} > c > 0.$$

Considere o deslocamento $f : M \to M$ e defina o subconjunto $A \subset M$ por

$$A = \{y \in M; y_0 = 1\}.$$

Note que o fato de $f^j(\alpha) \in A$ equivale a dizer que $\alpha_j = 1$, ou seja, $j \in S$. Resumindo,

$$f^j(\alpha) \in A \Leftrightarrow j \in S \tag{7.4}$$

O conjunto A é um aberto e ao mesmo tempo um fechado de M, considerando a topologia dada onde os cilindros são abertos, pois A é um cilindro de comprimento 1 em M e seu complementar é uma união de cilindros. Deste modo, tendo em vista (7.4), mostrar que $m+in \in S$ equivale a mostrar que $f^{m+in}(\alpha) \in A$. Logo, para mostrar o Teorema de Szemerédi, basta provar que para todo $k \in \mathbb{N}$ existem $m \in \mathbb{Z}$ e $n \in \mathbb{N}$ tais que

$$f^m(\alpha), f^{m+n}(\alpha), f^{m+2n}(\alpha), \dots, f^{m+kn}(\alpha) \in A.$$

Para mostrar este fato, vamos definir a sequência μ_n de probabilidades em M por:

$$\mu_n = \frac{1}{\# I_n} \sum_{i=a_n}^{b_n-1} \delta_{f^i(\alpha)} \tag{7.5}$$

Como já vimos no Teorema 4.4, o conjunto das probabilidades $\mathcal{M}_1(M)$ munido com a topologia fraca* é compacto. Assim, podemos garantir que alguma subsequência μ_{n_i} converge para uma probabilidade μ de M. Para não carregar a notação, vamos supor que a própria sequência μ_n converge para μ na topologia fraca* . Observe que μ é uma probabilidade f-invariante, pois para toda função contínua $\phi : M \to \mathbb{R}$, vale

$$\int \phi \circ f \, d\mu = \lim_{n\to\infty} \int \phi \circ f \, d\mu_n = \lim_{n\to\infty} \frac{1}{\# I_n} \sum_{i=a_n}^{b_n-1} \phi(f^i(\alpha))$$
$$+ \lim_{n\to\infty} \frac{\phi(f^{b_n}(\alpha)) - \phi(f^{a_n}(\alpha))}{\# I_n} = \lim_{n\to\infty} \int \phi \, d\mu_n = \int \phi \, d\mu.$$

Para utilizar o Teorema de Recorrência Simultânea de Poincaré para o conjunto A, precisamos mostrar inicialmente que $\mu(A) > 0$. De fato, observe que A é um conjunto fechado e aberto de M. Logo, pela Proposição 4.3 temos que

$$\mu(A) \geq \mu_n(A) = \lim_{\# I_n \to \infty} \frac{\#(S \cap I_n)}{\# I_n} > c > 0.$$

Dado $k \in \mathbb{N}$, considerando as funções $f, f^2, f^3, \dots, f^k$ (que claramente comutam entre si) o Teorema de Recorrência Simultânea de Poincaré nos garante que existe algum $n \geq 1$ tal que

$$\mu\big(A \cap f^{-n}(A) \cap f^{-2n}(A) \cap \dots \cap f_k^{-kn}(A)\big) > 0.$$

Em particular, como A é aberto, existe algum l tal que

$$\mu_l\big(A \cap f^{-n}(A) \cap f^{-2n}(A) \cap \dots \cap f^{-kn}(A)\big) > 0.$$

Como $\mu_l = (1/\#I_l)\sum_{i=a_l}^{b_l-1}\delta_{f^i(\alpha)}$, podemos garantir que pelo menos para algum $a_n \le m \le b_l - 1$, o ponto $f^m(\alpha)$ pertence a $A \cap f^{-n}(A) \cap \cdots \cap f^{-kn}(A)$. Assim, $f^{m+in}(\alpha) \in A$, para $i = 0, 1, \dots, k$, como queríamos provar.

□

7.2 Teorema de Weyl

Vamos descrever outra bela aplicação da Teoria Ergódica à Teoria dos Números, devida a H. Weyl [Wey16]. Consideramos funções polinomiais

$$P(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_d x^d,$$

com coeficientes reais e grau $d \ge 1$. Para cada inteiro positivo n, calculamos o valor da função P em n e chamamos z_n a parte fracionária do valor $P(n)$ obtido. De maneira mais formal,

$$z_n = \{P(n)\} = P(n) - [P(n)]$$

onde $\{x\}$ = parte fracionária e $[x]$ = parte inteira de x. Observe que $z_n \in [0,1)$ para cada n. Mas podemos, igualmente, considerar que a sequência toma valores no círculo $S^1 = \mathbb{R}/\mathbb{Z}$, e faremos isso no que segue. Estamos interessados em entender como se distribui a sequência z_n no círculo.

Definição 7.11. Dizemos que uma sequência $x_n \in S^1$ é *equidistribuída* se para qualquer função contínua $\varphi : S^1 \to \mathbb{R}$ tem-se

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} \varphi(x_j) = \int \varphi(x)\, dx.$$

Veremos no Exercício 7.4 que isto equivale a dizer que, para todo intervalo $I \subset S^1$, a fração dos termos da sequência que estão em I é igual ao comprimento $m(I)$ desse intervalo.

Teorema 7.12 (Weyl). *Se algum dos coeficientes $a_1, a_2, \dots, a_d$ é irracional então a sequência $z_n = \{P(n)\}$ é equidistribuída.*

Podemos, sem restrição, supor que o coeficiente a_d é irracional. De fato, a sequência z_n sempre pode ser decomposta numa soma

$$z_n = x_n + y_n, \quad x_n = \{a_d n^d\}, \quad y_n = \{Q(n)\}$$

onde $Q(x) = a_0 + a_1 x + \cdots + a_{d-1} x^{d-1}$. Suponha que a_d é racional, isto é, que existem inteiros p e q tais que $a_d = p/q$. Então a primeira parcela x_n toma no máximo q valores distintos. De fato esta sequência é periódica com período q:

$$x_{n+p} = \Big\{\frac{p}{q}(n+q)^d\Big\} = \Big\{\frac{p}{q}n^d\Big\} = x_n \quad \text{para todo } n \in \mathbb{Z}.$$

Por outro lado, a segunda parcela y_n é do mesmo tipo que z_n, exceto que o polinômio Q que lhe está associado tem grau $d-1$. Portanto, por indução no grau, podemos supor que y_n é equidistribuída. Mais que isso, podemos supor que as subsequências

$$y_{qn+r} = \{Q(qn+r)\}, \quad n \in \mathbb{Z}$$

são equidistribuídas para todo $r \in \{0, 1, \ldots, q-1\}$. Na verdade, estas sequências podem ser escritas como $y_{nq+r} = \{Q_r(n)\}$ para algum polinômio Q_r com o mesmo grau que Q (verifique), e portanto a hipótese de indução se aplica a elas também. Destas duas observações segue que a soma z_n também é equidistribuída, porque cada uma das subsequências z_{qn+r}, $n \in \mathbb{Z}$ é equidistribuída.

7.2.1 O caso afim

Para desenvolvermos alguma intuição sobre o problema, comecemos por considerar o caso especial $d = 1$. Neste caso a função polinomial resume-se a $P(x) = a_0 + a_1 x$. Estamos supondo que o coeficiente a_1 é irracional. Consideremos a transformação

$$f : S^1 \to S^1, \quad f(\theta) = \theta + a_1 \mod \mathbb{Z}.$$

Foi visto na Proposição 4.6 que esta transformação f admite uma única probabilidade invariante, que é a medida de Lebesgue m. Consequentemente, dada qualquer função contínua $\varphi : S^1 \to \mathbb{R}$, e dado qualquer ponto $\theta \in S^1$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \varphi(f^j(\theta)) = \int \varphi \, dm.$$

Considere $\theta = a_0 \mod \mathbb{Z}$. Então, $f^j(\theta) = a_0 + a_1 j \mod \mathbb{Z} = P(j) \mod \mathbb{Z}$ e isto significa que podemos identificar $z_j = \{P(j)\}$ com a sequência dos iterados $f^j(a_0)$. Então a relação anterior dá

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \varphi(z_j) = \int \varphi \, dm.$$

Isto é precisamente o que significa dizer que z_j é equidistribuída.

7.2.2 Ergodicidade

Vamos estender os argumentos acima para provar o caso geral do Teorema 7.8. Seja $\mathbb{T}^d$ o toro d-dimensional, isto é,

$$\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d = S^1 \times \cdots \times S^1 \quad (d \text{ vezes}).$$

Introduzimos a transformação $f : \mathbb{T}^d \to \mathbb{T}^d$

$$f(\theta_1, \theta_2, \ldots, \theta_d) = (\theta_1 + \alpha, \theta_2 + \theta_1, \ldots, \theta_d + \theta_{d-1}), \tag{7.6}$$

onde α é um número irracional que será escolhido mais tarde. Observe que f preserva a medida de Lebesgue m em $\mathbb{T}^d$. Isto pode ser visto usando as idéias da Seção 3.2: a derivada de f em cada ponto vem dada pela matriz

$$\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 & 1 \end{pmatrix}$$

cujo determinante é 1. Observe que a transformação f é invertível.

Proposição 7.13. *A transformação f é ergódica relativamente à medida de Lebesgue no toro $\mathbb{T}^d$.*

Demonstração. O método é análogo ao da Proposição 6.5, baseado em análise de Fourier. Seja $\varphi : \mathbb{T}^d \to \mathbb{R}$ uma função em $L^2(m)$. Escrevemos

$$\varphi(\theta) = \sum_{n\in\mathbb{Z}^d} a_n e^{2\pi i n\cdot\theta}$$

onde $\theta = (\theta_1, \ldots, \theta_d)$, $n = (n_1, \ldots, n_d)$, $n \cdot \theta = n_1\theta_1 + \cdots + n_d\theta_d$, e

$$\sum_{n\in\mathbb{Z}^d} |a_n|^2 = \int |\varphi(\theta)|^2 \, d\theta_1 \cdots d\theta_d < \infty. \tag{7.7}$$

Suponhamos que a função φ é invariante, isto é, $\varphi \circ f = \varphi$ em quase todo ponto. Observe que

$$\begin{aligned}\varphi(f(\theta)) &= \sum_{n\in\mathbb{Z}^d} a_n e^{2\pi i(n_1(\theta_1+\alpha)+n_2(\theta_2+\theta_1)+\cdots n_d(\theta_d+\theta_{d-1}))} \\ &= \sum_{n\in\mathbb{Z}^d} a_n e^{2\pi i n_1 \alpha} e^{2\pi i L(n)\cdot\theta}\end{aligned}$$

onde $L(n) = (n_1+n_2, n_2+n_3, \ldots, n_{d-1}+n_d, n_d)$. Portanto, a relação de invariância $\varphi \circ f = \varphi$ se traduz por

$$a_n e^{2\pi i n_1 \alpha} = a_{L(n)} \quad \text{para todo } n \in \mathbb{Z}^d. \tag{7.8}$$

Isto implica que a_n e $a_{L(n)}$ têm o mesmo valor absoluto. Por outro lado, a relação de integrabilidade (7.7) implica que existe no máximo um número finito de termos com um dado valor absoluto não-nulo. Concluímos que $a_n = 0$ para todo $n \in \mathbb{Z}^d$ cuja órbita $L^j(n)$, $j \in \mathbb{Z}$ seja infinita. Observando a expressão de L deduzimos que $a_n = 0$ exceto, possivelmente, se $n_2 = \cdots = n_d = 0$. Além disso, para os valores de n restantes, ou seja, para $n = (n_1, 0, \ldots, 0)$, tem-se que $L(n) = n$ e portanto a relação (7.8) torna-se

$$a_n = a_n e^{2\pi i n_1 \alpha}.$$

Como α é irracional, o último fator é diferente de 1 sempre que n_1 é não-nulo. Portanto esta relação dá que $a_n = 0$ também para $n = (n_1, 0, \ldots, 0)$ com $n_1 \neq 0$. Deste modo, mostramos que se φ é uma função invariante então todos os termos da sua expansão de Fourier se anulam exceto, possivelmente, o termo constante. Isto mostra que φ é constante, e isso prova que f é ergódica. $\square$

7.2.3 Unicidade ergódica

O próximo passo da demonstração do Teorema 7.12 é a seguinte

Proposição 7.14. *A transformação f é unicamente ergódica, isto é, a medida de Lebesgue no toro é a sua única medida invariante.*

Demonstração. A demonstração será por indução no grau d do polinômio P. O caso de grau 1 já foi tratado na Seção 7.2.1, portanto só precisamos explicar como o caso de grau d pode ser deduzido do caso de grau $d-1$. Para isso, escrevemos $\mathbb{T}^d = \mathbb{T}^{d-1} \times S^1$ e

$$f : \mathbb{T}^{d-1} \times S^1 \to \mathbb{T}^{d-1} \times S^1 \quad f(\theta_0, \eta) = (f_0(\theta_0), \eta + \theta_{d-1}) \tag{7.9}$$

onde $\theta_0 = (\theta_1, \dots, \theta_{d-1})$ e $f_0(\theta_0) = (\theta_1 + \alpha, \theta_2 + \theta_1, \dots, \theta_{d-1} + \theta_{d-2})$. Vamos representar por $\pi : \mathbb{T}^d \to \mathbb{T}^{d-1}$ a projeção $\pi(\theta) = \theta_0$. Por indução, a transformação

$$f_0 : \mathbb{T}^{d-1} \to \mathbb{T}^{d-1}$$

é unicamente ergódica. Para mostrar que f é unicamente ergódica só precisamos mostrar que a medida de Lebesgue m é a sua única probabilidade invariante *ergódica*.

Lema 7.15. *Se μ é uma probabilidade invariante por f então a projeção $\pi_*\mu$ coincide com a medida de Lebesgue m_0 em $\mathbb{T}^{d-1}$.*

Demonstração. Dado qualquer conjunto mensurável $E \subset \mathbb{T}^{d-1}$,

$$(\pi_*\mu)(f_0^{-1}(E)) = \mu(\pi^{-1} f_0^{-1}(E)).$$

Usando $\pi \circ f = f_0 \circ \pi$ e o fato de que μ é f-invariante, se verifica que a expressão do lado direito é igual a

$$\mu(f^{-1}\pi^{-1}(E)) = \mu(\pi^{-1}(E)) = (\pi_*\mu)(E).$$

Portanto $(\pi_*\mu)(f_0^{-1}(E)) = (\pi_*\mu)(E)$ para todo subconjunto mensurável E, ou seja, $\pi_*\mu$ é probabilidade f_0-invariante. Como supomos que f_0 é unicamente ergódico, segue que $\pi_*\mu$ coincide com a medida de Lebesgue m_0 em $\mathbb{T}^{d-1}$. □

Agora suponhamos que μ é ergódica. Pelo Teorema de Birkhoff 5.2, o conjunto G_μ dos pontos $\theta \in \mathbb{T}^d$ tais que

$$\lim \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f^j(\theta)) = \int \varphi \, d\mu \quad \text{para toda função contínua } \varphi : \mathbb{T}^d \to \mathbb{R} \tag{7.10}$$

tem medida total. Seja $G_0(\mu)$ o conjunto dos $\theta_0 \in \mathbb{T}^{d-1}$ tais que $G(\mu)$ intersecta $\{\theta_0\} \times S^1$. Em outras palavras, $G_0(\mu) = \pi(G_\mu)$. É claro que $\pi^{-1}(G_0(\mu))$ contém G_μ e portanto tem medida μ igual a 1. Logo, usando o Lema 7.15,

$$m_0(G_0(\mu)) = \mu(\pi^{-1}(G_0(\mu))) = 1.$$

Em particular, isto vale para a medida de Lebesgue:

$$m_0(G_0(m)) = m(\pi^{-1}(G_0(m))) = 1.$$

Uma consequência direta destas relações é que a intersecção de $G_0(\mu)$ e $G_0(m)$ tem medida m_0 total e, portanto, estes conjuntos não podem ser disjuntos. Seja θ_0 um ponto qualquer na intersecção. Por definição, $G(\mu)$ intersecta $\{\theta_0\} \times S^1$. Mas o próximo resultado afirma que $G(m)$ *contém* $\{\theta_0\} \times S^1$:

Lema 7.16. *Se $\theta_0 \in G_0(m)$ então $\{\theta_0\} \times S^1$ está contido em $G(m)$.*

Demonstração. A observação crucial é que a medida m é invariante por toda a transformação da forma

$$R_\beta : \mathbb{T}^{d-1} \times S^1 \to \mathbb{T}^{d-1} \times S^1, \quad (\zeta, \eta) \mapsto (\zeta, \eta + \beta).$$

A hipótese $\theta_0 \in G_0(m)$ significa que existe algum $\eta \in S^1$ tal que $(\theta_0, \eta) \in G(m)$, ou seja,

$$\lim \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f^j(\theta_0, \eta)) = \int \varphi \, dm$$

para toda função contínua $\varphi : \mathbb{T}^d \to \mathbb{R}$. Qualquer outro ponto de $\{\theta_0\} \times S^1$ pode ser escrito como $(\theta_0, \eta + \beta) = R_\beta(\theta_0, \eta)$ para algum $\beta \in S^1$. Recordando (7.6), vemos que

$$f\big(R_\beta(\tau_0, \zeta)\big) = (\tau_1 + \alpha, \tau_2 + \tau_1, \ldots, \tau_{d-1} + \tau_{d-2}, \zeta + \beta + \tau_{d-1}) = R_\beta\big(f(\tau_0, \zeta)\big)$$

para todo $(\tau_0, \zeta) \in \mathbb{T}^{d-1} \times S^1$. Logo, por indução,

$$f^j(\theta_0, \eta + \beta) = f^j\big(R_\beta(\theta_0, \eta)\big) = R_\beta\big(f^j(\theta_0, \eta)\big)$$

para todo $j \geq 1$. Portanto, dada qualquer função contínua $\varphi : \mathbb{T}^d \to \mathbb{R}$,

$$\begin{aligned}\lim \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f^j(\theta_0, \eta + \beta)) &= \lim \frac{1}{n} \sum_{j=0}^{n-1} (\varphi \circ R_\beta)(f^j(\theta_0, \eta)) \\ &= \int (\varphi \circ R_\beta)\, dm = \int \varphi\, dm.\end{aligned}$$

Isto prova que $(\theta_0, \eta + \beta)$ está em G_m para todo $\beta \in S^1$, conforme afirmado. □

Segue do que dissemos até agora que $G(\mu)$ e $G(m)$ se intersectam em algum ponto de $\{\theta_0\} \times S^1$. Tendo em vista a definição (7.10), isto implica que as duas medidas têm a mesma integral para cada função contínua. De acordo com o Teorema de Riesz-Markov 4.7, isto implica que $\mu = m$, como queríamos demonstrar. □

Corolário 7.17. *A órbita de todo ponto $\theta \in \mathbb{T}^d$ é equidistribuída no toro: para toda função contínua $\psi : \mathbb{T}^d \to \mathbb{R}$ tem-se*

$$\lim \frac{1}{n} \sum_{j=0}^{n-1} \psi(f^j(\theta)) = \int \psi\, dm.$$

Demonstração. Isto é uma consequência imediata da Proposição 7.14 e da Proposição 4.7. □

7.2.4 Demonstração do Teorema de Weyl

Para completarmos a demonstração do Teorema 7.12, introduzimos os polinômios definidos por $p_d(x) = P(x)$ e

$$p_{j-1}(x) = p_j(x+1) - p_j(x) \quad \text{for } j = 2, \ldots, d. \tag{7.11}$$

Lema 7.18. *1. O polinômio $p_j(x)$ tem grau j, para todo $1 \leq j \leq d$.*

2. $p_1(x) = \alpha x + \beta$ onde $\alpha = a_d d!$ é irracional.

Deixamos a demonstração deste lema para o Exercício 7.5.

Lema 7.19. *Para todo* $n \geq 0$,

$$f^n\big(p_1(0), p_2(0), \ldots, p_d(0)\big) = \big(p_1(n), p_2(n), \ldots, p_d(n)\big).$$

Demonstração. A demonstração será por indução em n. Como o caso $n = 0$ é óbvio, só precisamos tratar do passo indutivo. Lembre que f foi definida em (7.6). Se

$$f^{n-1}(p_1(0), p_2(0), \ldots, p_d(0)) = (p_1(n-1), p_2(n-1), \ldots, p_d(n-1))$$

então $f^n(p_1(0), p_2(0), \ldots, p_d(0))$ é igual a

$$(p_1(n-1) + \alpha, p_2(n-1) + p_1(n-1), \ldots, p_d(n-1) + p_{d-1}(n-1)).$$

Usando a definição (7.11) e o Lema 7.18, obtemos que esta expressão é igual a

$$(p_1(n), p_2(n), \ldots, p_d(n)),$$

e isto prova o lema. $\square$

Finalmente, estamos prontos para provar que a sequência $z_n = \{P(n)\}$ é equidistribuída, conforme afirma o Teorema 7.12. Seja $\varphi : S^1 \to \mathbb{R}$ uma função contínua qualquer. Considere $\psi : \mathbb{T}^d \times \mathbb{R}$ definida por

$$\psi(\theta_1, \theta_2, \ldots, \theta_d) = \varphi(\theta_d).$$

Fixemos $\theta = (p_1(0), p_2(0), \ldots, p_d(0))$. Usando o Lema 7.19 e o Corolário 7.17,

$$\lim \frac{1}{n} \sum_{j=0}^{n-1} \varphi(z_n) = \lim \frac{1}{n} \sum_{j=0}^{n-1} \psi(f^n(\theta)) = \int \psi \, dm = \int \varphi \, dx.$$

Isto termina a demonstração do Teorema 7.12.

7.3 Exercícios

7.1. Prove que

1. $D_i(S) = D_s(\mathbb{Z} \setminus S)$ para qualquer subconjunto S de $\mathbb{Z}$.

2. Se $S_1, S-2, \dots, S_N$ é uma partição de $\mathbb{Z}$ então

$$D_s(S_1) + D_s(S_2) + \cdots + D_s(S_N) \geq 1.$$

7.2. Deduza o Teorema 7.9 a partir do Teorema 7.7.

7.3. Mostre que dadas quaisquer transformações contínuas $f_i : M \to M$, $i = 1, 2, \dots, k$ num espaço métrico compacto M, tais que $f_i \circ f_j = f_j \circ f_i$ para todo $i, j = 1, 2, \dots, k$, existe alguma medida de probabilidade μ que é invariante por todas essas transformações.

7.4. Mostre que uma sequência x_n é equidistribuída se e somente se, dado qualquer intervalo $I \subset [0, 1]$, tem-se

$$\lim_{n\to\infty} \frac{1}{n} \#\{j = 1, 2, \dots, n : x_j \in I\} = m(I)$$

onde m representa a medida de Lebesgue em $[0, 1]$.

7.5. Demonstre o Lema 7.18.

7.6. Sejam $f_i : M \to M$, $i = 1, 2, \dots, k$ transformações que preservam uma probabilidade μ em M e tais que $f_i \circ f_j = f_j \circ f_i$ para todo $i, j = 1, 2, \dots, k$. Então, para qualquer conjunto $E \subset M$ tal que $\mu(E) > 0$, e para quase todo $x \in E$ existe uma sequência $n_j \to \infty$ tal que $f_i^{n_j}(x) \in E$ para todo $i = 1, \dots, k$ e todo $j \geq 1$.

Bibliografia

[Cas04] A. A. Castro. *Teoria da medida.* Projeto Euclides. IMPA, 2004.

[ET36] P. Erdös and P. Turán. On some sequences of integers. *J. London. Math. Soc.*, 11:261–264, 1936.

[Fer02] R. Fernandez. *Introdução à teoria da medida.* Projeto Euclides. IMPA, 2002.

[Fur77] H. Furstenberg. Ergodic behavior and a theorem of Szemerédi on arithmetic progressions. *J. d'Analyse Math.*, 31:204–256, 1977.

[Fur81] H. Furstenberg. *Recurrence in ergodic theory and combinatorial number theory.* Princeton Univertsity Press, 1981.

[Mañ87] R. Mañé. *Ergodic theory and differentiable dynamics.* Springer Verlag, 1987.

[Rud87] W. Rudin. *Real and complex analysis.* McGraw-Hill, 3 edition, 1987.

[Sze75] S. Szemerédi. On sets of integers containing no k elements in arithmetic progression. *Acta Arith.*, 27:199–245, 1975.

[vdW27] B. van der Waerden. Beweis eibe Baudetschen Vermutung. *Nieuw Arch. Wisk.*, 15:212–216, 1927.

[Wal82] P. Walters. *An introduction to ergodic theory.* Springer Verlag, 1982.

[Wey16] H. Weyl. Uber die Gleichverteilungen von Zahlen mod Eins. *Math. Ann.*, 77:313–352, 1916.