UNIVERSIDADE FEDERAL DE SANTA CATARINA
CFM-CURSO DE PÓS-GRADUAÇÃO EM MATEMÁTICA

g-MEDIDAS E APLICAÇÕES EXPANSORAS NÃO ERGÓDICAS

DISSERTAÇÃO SUBMETIDA À UNIVERSIDADE FEDERAL DE SANTA CATARINA PARA A OBTENÇÃO DO GRAU DE MESTRE EM MATEMÁTICA E COMPUTAÇÃO CIENTÍFICA.

GENTIL LOPES DA SILVA

FLORIANÓPOLIS
SANTA CATARINA - BRASIL
AGOSTO - 1997

g-MEDIDAS E APLICAÇÕES EXPANSORAS NÃO ERGÓDICAS

GENTIL LOPES DA SILVA

Esta dissertação foi julgada adequada para obtenção do Título de **"Mestre em Matemática e Computação Científica"** e aprovada em sua forma final pelo Orientador e Membros da banca examinadora.

Aldrovando Araújo

Prof. Aldrovando Luiz Araújo, Dr.
Orientador

Etzel Ritter von Stocker

Prof. Etzel Ritter von Stocker, Dr.
Coordenador do Curso

Banca Examinadora :

Aldrovando Araújo

Prof. Aldrovando Luiz Araújo, Dr.
Orientador

Claudio Landim

Prof.Claudio Landim, Dr.

Prof. Eduardo Arbieto Alarcon, Dr.

Apalpamos as paredes como cegos;
sim, como os que não têm olhos andamos apalpando;
tropeçamos ao meio dia como no crepúsculo,
e entre os vivos somos como os mortos.

Isaías 59:10

DEDICATÓRIA

À MEU AVÔ

GERMANO LOPES DA SILVA (PAI VÉIO), in Memorium.

RECONHECIMENTO

À MINHA COMPANHEIRA

PATRÍCIA MELO DA SILVA

Por estes anos todos de peregrinação.

AGRADECIMENTOS

A **DEUS**, POR TUDO

Então saiu um espírito,
apresentou-se diante do Senhor, e disse:
Eu o induzirei. E o Senhor lhe perguntou: De que modo?
Respondeu ele: Eu sairei, e serei um espírito mentiroso
na boca de todos os seus profetas. Ao que disse o Senhor:
Tu o induzirás, e prevalecerás; saí, e faze assim.

1 **REIS** 22 : 21, 22

INDÍCE

Introdução

Nos últimos cinqüenta anos as idéias desenvolvidas pela Teoria dos Sistemas Dinâmicos Difereciáveis contribuíram enormemente na descrição de comportamentos muito irregulares que aparecem nos fenômenos naturais. De uma perspectiva mais determinista estas idéias convergiram para uma visão mais estatística dos fenômenos que combinando noções da Teoria das Probabilidades com as da Dinâmica resultaram numa das áreas mais férteis e produtivas em nossos dias. Esta área, a Teoria Ergódica Diferenciável, visa descrever propriedades "relevantes"de um certo sistema que ocorrem em determinados pontos do espaço de fase que são "visíveis"a partir de um determinado modo de medir os subconjuntos

do espaço. A forma ou proporção pela qual se estima o espaço de fase deve estar intrísecamente relacionada com a dinâmica do sistema e denomina-se *medida invariante.* Mais formalmente, consideremos um sistema dinâmico discreto, isto é (X, T), onde X é espaço métrico e $T: X \to X$ uma aplicação contínua. Sobre X está definida a σ-álgebra dos boreleanos de X e sobre esta σ-álgebra temos as medidas boreleanas. As medidas constituem a ferramenta pela qual estimamos (ou medimos) os subconjuntos de X. Dizemos que uma medida μ é invariante por T e notamos $\mu \in \mathcal{M}(T)$ se

$$\mu(T^{-1}(A)) = \mu(A) \qquad \forall A \in \mathcal{B}(X).$$

Essencialmente a Teoria Ergódica Diferenciável estuda a tripla (X, T, μ) onde X é uma variedade diferenciável, $T: X \to X$ uma aplicação C^r sobre X e μ uma medida invariante por T.

Especial atenção é dada àqueles sistemas denominados de *ergódicos.* Uma medida invariante μ pode ser decomposta em uma combinação de outras medidas invariantes (esta decomposição não é única). Se tal não for possível, dizemos que μ é ergódica e a posteriori que o sistema $(\mathcal{M}, T, \mu)$ é ergódico. A noção de ergodicidade foi introduzida por Boltzman como uma propriedade satisfeita por fluxos hamiltonianos em seus níveis de energia. No entanto a ergodicidade é uma propriedade extremamente fraca (suave) na hierarquia das propriedades estocásticas de um sistema dinâmico. O estudo de propriedades mais fortes (por exemplo, mixing, K-sistema, bernoullicidade) em sistemas dinâmicos diferenciáveis começou com os fluxos geodésicos em superfícies de curvatura negativa. Ao estudá-los Hopf [13] criou um método para provar a ergodicidade o qual se mostrou tão versátil que permitiu uma série de generalizações. Foi desenvolvido por Anosov e Sinai e aplicado aos sistemas de Anosov com uma medida invariante e absolutamente contínua com respeito à medida de Lebesgue da variedade.

A idéia chave nesta abordagem era a noção de "hiperbolicidade" de um sistema dinâmico. Por comportamento hiperbólico queremos significar a propriedade do afastamento exponencial de órbitas próximas. Na sua forma uniforme corresponde aos sistemas de Anosov e Smale e implicam em um comportamento topológico bem rígido.

Se $T: M \to M$ é um difeomorfismo C^r de uma variedade compacta, dizemos que T é Anosov se existem $c > 0, 0 < \lambda < 1$ e decomposição

$$TM = E^s \oplus E^u$$

tal que

$$(D_x T^n)E_x^t \subset E_{T^n(x)}^t \qquad t = s, u.$$

$$\|(D_x T^n)/E_x^s\| \leq c\lambda^n$$

$$\|(D_x T^{-n})/E_x^u\| \leq c\lambda^n$$

para todo $x \in M, n \in \mathbb{N}$.

A propriedade que define os difeomorfismos de Anosov consiste exatamente na hiperbolicidade uniforme sobre M. Exigir que esta propriedade seja válida em um subconjunto compacto e invariante por T, (Isto é, $T(\Lambda) = \Lambda$) corresponde a dizer que (Λ, T) é hiperbólico.

Denotemos por λ a medida de Lebesgue de M. Então se $\mu \in \mathcal{M}(T)$ dizemos que μ é absolutamente contínua com respeito a λ (notação: $\mu \ll \lambda$) se

$$A \in \mathcal{B}(M) \quad \text{satisfaz} \quad \lambda(A) = 0 \quad \Rightarrow \quad \mu(A) = 0.$$

Em 1966 Anosov demonstrou que fluxos e difeomorfismos de Anosov de classe $C^r, r > 1$, que presevam uma medida gerada por uma forma de volume são ergódicos [14]. Anosov havia essencialmente depurado o argumento de Hopf, pois os fluxos geodésicos em variedades de curvatura negativa são Anosov. Além disso ele notou que o argumento dependia basicamente da continuidade absoluta das transformações de Poincaré geradas pelas folheações estável e instável. Voltemos ao nosso Anosov $T: X \to X$.

Definição: Dado $x \in M$ definimos a variedade estável $W^s(x)$ do ponto x via

$$W^s(x) = \{y : \lim_{n\to\infty} d(T^n(x), T^n(y)) = 0\}$$

De modo análogo se define a variedade instável de x.

Quando T é Anosov estes conjuntos são imersões injetivas de $\mathbb{R}^k$ com a mesma diferenciabilidade de T e dimensão k igual a dimensão de E^s_x. A família de variedades $W^s(x), x \in M$ denomina-se folheação estável. Seja $\mathcal{F}^s$ a folheação estável de T. Se N_1, N_2 são subvariedades tranversais à $\mathcal{F}^s$, isto é $\forall x \in N_i$

$$T_xM = E^s_x \oplus T_xN_i \qquad i = 1, 2.$$

e se N_1, N_2 são próximas podemos definir uma aplicação de N_1 em N_2 associando a um dado $x \in N_1$ o ponto $y \in N_2$ tal que

$$y \in W^s(x)$$

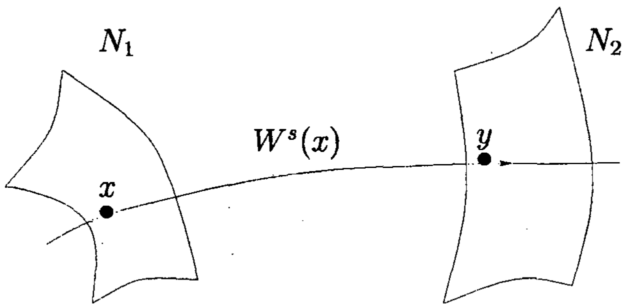

Figura 0.1: Aplicação de N_1 em N_2

Estas aplicações denominam-se *aplicações de Poincaré da folheação estável.*

Analisando a prova de Hopf, Anosov e depois Sinai perceberam que o argumento dependia apenas da propriedade das transformações de Poincaré das folheações possuírem um jacobiano. Esta descoberta de que as transformações de Poincaré de sistemas de Anosov C^r com $r > 1$ possuem um jacobiano, ainda que não tenham derivada, foi o que possibilitou aplicar o argumento de Hopf para provar que Anosov's $C^r, r > 1$, que preservam a medida de Lebesgue são ergódicos. Sinai levou mais além o argumento e usando partições de Markov observou que o argumento dependia em essência da expansão. A partir disto ele desenvolveu a mesma teoria para as tranformações expansoras de variedades, a saber; condições para exibirem uma única medida invariante e absolutamente contínua com respeito ao volume e as propriedades desta medida como ergodicidade, K-sistema e Bernoulli.

Por aplicação expansora entendemos uma aplicação de classe C^r , $T: M \to M$ para a qual existe $c > 1$ tal que $\forall x \in M$

$$\|(D_x T)V\| \geq c\|v\| \qquad \forall v \in T_x M.$$

Resultou que a propriedade fundamental que garante a existência, "unicidade" e implica a ergodicidade de medida invariante e absolutamente contínua com respeito à medida de Lebesgue era a transformação T ser de classe $C^{1+\varepsilon}$ para algum $\varepsilon > 0$.

Definição: Uma aplicação $T: X \to Y$, onde X e Y são espaços métricos é Hölder-contínua se existem $c > 0$ e $\gamma > 0$ tal que

$$d(T(x), T(y)) \leq c\, d^{\gamma}(x, y)$$

A constante γ denomina-se de constante de Hölder de T.

Definição: Uma aplicação $T: M \to M$, onde M é uma variedade é de classe $C^{1+\varepsilon}$ com $\varepsilon > 0$ se $\mathbf{det}(D_x T)$ é Hölder-contínuo com constante de Hölder ε.

No entanto para aplicações de classe apenas C^1 permaneceram em aberto todas estas questões. O primeiro resultado que indicava serem falsas as proposições com a eliminação da hipótese $C^{1+\varepsilon}$ foi devido a Clark Robinson e Lai-sang Young [15]. De modo extremamente engenhoso construíram um exemplo de um difeomorfismo de Anosov de classe C^1 cujas folheações instáveis e estáveis não eram absolutamente contínuas. Como as condições que permitiam o argumento de Hopf e Anosov não se verificavam, ou o resultado era falso ou outra prova deveria aparecer. No entanto, apesar das expectativas dos autores tal exemplo jamais pode ser modificado de modo a preservar volume, e este problema permanece até nossos dias, em aberto e pode ser formulado como conjectura.

Conjectura: Existe um difeomorfismo de Anosov de classe C^1 sobre uma variedade compacta preservando volume e não-ergódico.

Esta conjectura se revelou mais verossímel com os resultados sobre aplicações expansoras que se seguiram. Krzyzewski mostrou a existência de aplicação C^1 expansora sem nenhuma medida invariante e absolutamente contínua [16]. Um exemplo concreto veio a seguir com Góra e Schmitt [17]. Finalmente em 1994 Quas apresentou um exemplo de uma aplicação expansora do círculo unitário de classe C^1 preservando a medida de Lebesgue e não ergódica [18].

É nosso objetivo neste trabalho apresentar de modo completo e auto-suficiente a construção deste exemplo. Terminamos esta introdução dando ao leitor uma idéia esquemática da construção do exemplo e suas propriedades.

A ferramenta fundamental na construção do exemplo é a noção de g-medida que definiremos aqui de modo diferente do autor. Antes precisamos de uma definição.

Definição: Seja $T: M \to M$ uma aplicação localmente injetiva e $\mu \in \mathcal{M}(T)$. Dizemos que uma aplicação contínua

$$J: M \to \mathbb{R}$$

é o jacobiano de T com respeito a μ se

$$\mu(T(A)) = \int_A J\, d\mu$$

para todo boreleano A tal que $T\big|_A$ é injetiva.

Definição: Seja $T: S^1 \to S^1$ expansora. Seja $g: S^1 \to (0,1)$, contínua tal que

$$\sum_{y \in T^{-1}(\{x\})} g(y) = 1 \qquad \forall x \in S^1.$$

Uma medida ν sobre $\mathcal{B}(S^1)$ é uma g-medida se $\frac{1}{g}$ é o jacobiano de T com respeito a ν, isto é

$$J_\nu T = \frac{1}{g}$$

g-medidas são invariantes por T, positivas sobre abertos e sem átomos (isto será provado no capítulo 2). Usando o mesmo argumento de Sinai na prova da existência de uma conjugação entre homeomorfismos de S^1 com número de rotação irracional e transitivos e uma rotação irracional de mesmo número de rotação podemos definir

$$h: S^1 \to S^1$$

$$h(x) = \nu([0, x])$$

onde $S^1 \equiv [0,1)$ e provar que h é um homeomorfismo de S^1.

Além disso se

$$\mu = h^* \nu$$

então seja $\mu = \lambda$ a medida de Lebesgue de S^1 e definindo

$$H = h \circ T \circ h^{-1}$$

obtemos uma transformação de S^1 que é conjugada à T. Mas se $\frac{1}{g}$ é o jacobiano de T com repeito a ν, é fácil ver que $\frac{1}{goh^{-1}}$ é o jacobiano de g relativamente a μ que é a medida de Lebesgue. Logo g tem um jacobiano relativamente à medida de Lebesgue e como estamos em dimensão 1 ele coincide com a derivada provando que g é C^1. Como $0 < g(x) < 1$ segue que

$$H'(x) = \frac{1}{(g \circ h^{-1})(x)} > 1$$

para todo x e, portanto, H é expansora. Se ν é não-ergódica o mesmo vale para $\mu = \lambda$ e, neste caso a aplicação H seria o exemplo procurado.

Resumindo bastaria encontrar $k > 1$, $g: S^1 \to (0,1)$ nas condições acima mencionadas de modo que $T(z) = z^k$ tivesse mais de uma g-medida. Escolhendo uma combinação convexa, entre duas delas obeteríamos uma g-medida não ergódica.

Para tal usamos uma partição de Markov do S^1 e descemos para o shift. O argumento depende do shift unilateral ser Σ_{10} e portanto a construção é feita para $T(z) = z^{10}$. Identificando apropriadamente Σ_{10} com $I = [0,1]$ consideramos o problema de obter

$$g: I \to (0,1)$$

tal que

1) $gL = g$

2) $\exists\, \nu$ g-medida tal que
$$\nu([6]) > \frac{21}{40}$$
onde $L: [0,1] \to [0,1]$ definida por $L(x) = 1 - x$.

Se tal construção é possível consideramos

$$\mu = L^* \nu$$

e é fácil, usando a definição do autor, ver que μ é uma gL-medida. Como $gL = g$, μ é de fato uma g-medida. μ e ν sendo g-medidas restaria verificar que elas são distintas. Para tal fazemos

$$\mu([3]) = \nu(L([3])) = \nu([6]) > \frac{21}{40}$$

Se $\mu = \nu$, como $[6] \cap [3] = \emptyset$ segue que

$$\nu([6] \cup [3]) = \nu([6]) + \nu([3]) > \frac{21}{40} + \frac{21}{40} > 1$$

contradizendo o fato de ν ser uma medida de probabilidade. Segue que $\mu \neq \nu$.

Esta última construção é extremamente técnica, não podendo, portanto, ser apresentada aqui. Um último comentário no entanto se faz necessário: g-medidas podem ser definidas em um contexto mais geral o qual engloba os casos usados por Quas e também por Keane e outros. Para tal precisamos da definição de transformação expansora usada em [19] por aluno do professor Ricardo Mañé.

Definição: Seja K um espaço métrico compacto, $T: K \to K$ contínua é dita expansora se existirem $r > 0, 0 < \lambda < 1$ e $c > 0$ tal que

(a) $x \neq y$ e $T(x) = T(y) \quad \Rightarrow \quad d(x, y) > c$

(b) $\forall x \in K$ e $a \in T^{-1}(\{x\})$ existe

$$\varphi: B_r(x) \to K$$

tal que $\varphi(x) = a$ e

$$T(\varphi(y)) = y \quad \forall y \in B_r(x)$$

$$d(\varphi(z), \varphi(\omega)) \leq \lambda \, d(z, \omega) \qquad \forall z, \omega \in B_r(x).$$

Esta definição mais geral de aplicação expansora engloba todos os casos onde se definem g-medidas e permite de modo unificado e mais elegante se obter todas as propriedades usuais das g-medidas. No entanto, ainda que sua inclusão no nosso trabalho fosse mais esclarecedora envolveria reescrever inúmeros resultados já presentes na dissertação de Mestrado mencionada antes e tornaria as provas das propriedades mais longas e pesadas. O leitor interessado é convidado a ler [19].

Capítulo 1

Medida e Teoria Ergódica

1.1 - Elementos de Teoria da Medida

1.1.1 - *Álgebras, σ-álgebras, medidas*

Comprimento, área e volume, bem como probabilidade são exemplos do conceito de medida que vamos discutir. Uma medida é uma *função de conjunto*, isto é uma função que assinala um número $\mu(A)$ para cada conjunto A em uma certa classe. Alguma estrutura deve ser imposta sobre a classe de conjuntos na qual μ está definida, sendo assim considerações sobre probabilidade dão uma boa motivação para o tipo de estrutura requerido. Se Ω é um conjunto cujos pontos correspondem a possíveis resultados de um experimento aleatório, certos subconjuntos de Ω devem ser chamados de "eventos", aos quais assinalamos uma probabilidade. Intuitivamente, A é um evento se a questão "ω pertence a A?"tem uma resposta definida após o experimento ser realizado , i.e., sim ou não (o resultado corresponde a um ponto $\omega \in \Omega$). Agora se podemos responder à questão "$\omega \in A$?"podemos certamente responder a questão "$\omega \in A^c$?", e se, para cada $i = 1, \cdots, n$, podemos decidir quando ou não ω pertence a A_i, então podemos determinar quando ou não ω pertence a $\cup_{i=1}^{n} A_i$ (e similarmente para $\cap_{i=1}^{n} A_i$). Deste modo é natural exigir que a classe de eventos seja fechada sob complementação, união finita, e intersecção finita; outrossim, visto que a resposta à questão "$\omega \in \Omega$?"é sempre "sim", o espaço inteiro deve ser um evento. Fechamento sob uniões e intersecções enumeráveis é difícil justificar fisicamente, e talvez a mais convicente razão para se exigir isto é a rica teoria matemática obtida.

Definição. Seja $\mathcal{A}$ uma coleção de subconjuntos de um conjunto Ω. Então $\mathcal{A}$ é chamada uma álgebra se e sòmente se $\Omega \in \mathcal{A}$ e $\mathcal{A}$ é fechada sob complementação e uniões finitas, isto é,

(a) $\Omega \in \mathcal{A}$.

(b) Se $A \in \mathcal{A}$ então $A^c \in \mathcal{A}$.

(c) Se $A_1, A_2, \cdots, A_n \in \mathcal{A}$, então $\bigcup_{i=1}^{n} A_i \in \mathcal{A}$. Disto segue que $\mathcal{A}$ é fechada sob intersecções finitas. Pois se $A_1, A_2, \cdots, A_n \in \mathcal{A}$, então

$$\bigcap_{i=1}^{n} A_i = \left(\bigcup_{i=1}^{n} A_i^c\right)^c \in \mathcal{A}.$$

Se (c) é trocado por fechamento sob *união enumerável*, isto é

(d) Se $A_1, A_2, \cdots \in \mathcal{A}$, então $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$,

$\mathcal{A}$ é chamada uma σ-álgebra. Também como acima, $\mathcal{A}$ é fechada sob intersecção enumerável.

Exemplos: A maior σ-álgebra de subconjuntos de um conjunto fixo Ω é a coleção de todos os subconjuntos de Ω. A menor σ-álgebra é $\mathcal{A} = \{\emptyset, \Omega\}$.

Seja A um subconjunto próprio e não-vazio de Ω, e seja $\mathcal{A} = \{\emptyset, \Omega, A, A^c\}$. Então $\mathcal{A}$ é a menor σ-álgebra contendo A. De fato, se $\mathcal{B}$ é uma σ-álgebra e $A \in \mathcal{B}$, então por definição de uma σ-álgebra, $\emptyset, A^c$ e Ω pertencem a $\mathcal{B}$, conseqüentemente $\mathcal{A} \subset \mathcal{B}$. Mas $\mathcal{A}$ é uma σ-álgebra, assim se formamos complementações ou uniões de conjuntos em $\mathcal{A}$, invariavelmente obteremos conjuntos em $\mathcal{A}$. Deste modo $\mathcal{A}$ é uma σ-álgebra que está incluída em qualquer σ-álgebra contendo A, disto segue o resultado.

Se $A_1, \cdots, A_n$ são subconjuntos arbitrários de Ω, a menor σ-álgebra contendo $A_1, \cdots, A_n$ pode ser descrita explícitamente. A propósito, no caso de Ω finito (como ocorre em muitos problemas de probabilidade), não é difícil fazermos um programa computacional com esta finalidade, para isto a fórmula (4.4) (página 80-apêndice), para o cálculo de combinações certamente será útil.

A interseção de uma família qualquer de álgebras é uma álgebra. Portanto, dada uma classe não-vazia $\mathcal{H}$ de subconjuntos de Ω, podemos definir a *álgebra gerada por* $\mathcal{H}$, como a menor álgebra que contém $\mathcal{H}$, como é claro, coincide com a interseção de todas as álgebras que contém $\mathcal{H}$.

Se $\mathcal{H}$ é uma classe de conjuntos, a menor σ-álgebra contendo os conjuntos de $\mathcal{H}$ deve ser escrita como $\sigma(\mathcal{H})$, e ocasionalmente chamada a σ-álgebra *minimal* sobre $\mathcal{H}$.

Definição e Exemplos(Medida). Uma *medida* sobre uma σ-álgebra $\mathcal{A}$ é uma função μ sobre $\mathcal{A}$ a valores reais extendidos (i.e., a valores em $\mathbb{R}^* = \mathbb{R} \cup \{-\infty, +\infty\}$) e não-negativa, de modo que sempre que $A_1, A_2, \cdots$ formam uma coleção finita ou infinita enumerável de conjuntos disjuntos em $\mathcal{A}$ devemos ter

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

Se $\mu(\Omega) = 1$, μ é chamada uma *medida de probabilidade.*

Definição: Uma terna $(\Omega, \mathcal{A}, \mu)$, em que Ω é um conjunto, $\mathcal{A}$ é uma σ-álgebra contida no conjunto das partes de Ω e μ é uma medida sobre $\mathcal{A}$, é chamada um *espaço de medida.* O par $(\Omega, \mathcal{A})$ é chamado um *espaço mensurável.*

Exemplos: Seja Ω qualquer conjunto, e seja $\mathcal{A}$ consistindo de todos os subconjuntos de Ω. Defina $\mu(A)$ como o número de pontos em A. Deste modo se A tem n membros, $n = 1, 2, \cdots$, então $\mu(A) = n$; se A é um conjunto infinito, $\mu(A) = \infty$. A função de conjunto μ é uma medida sobre $\mathcal{A}$, chamada *medida da contagem* sobre Ω.

Uma medida proximamente relacionada a esta é definida como segue. Seja $\Omega = \{\omega_1, \omega_2, \cdots\}$ um conjunto finito ou infinito enumerável, e seja $p_1, p_2, \cdots$ números não-negativos. Tome $\mathcal{A}$ consistindo de todos os subconjuntos de Ω e defina

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

Deste modo se $A = \{\omega_{i_1}, \omega_{i_2}, \cdots\}$, então $\mu(A) = p_{i_1} + p_{i_2} + \cdots$. A função de conjunto μ é uma medida sobre $\mathcal{A}$ e $\mu\{\omega_i\} = p_i$, $i = 1, 2, \cdots$. Uma medida de probabilidade deve ser obtida sse $\Sigma_i p_i = 1$; se todo $p_i = 1$, então μ é a medida da contagem.

Agora se A é um subconjunto de $\mathbb{R}$, desejamos uma definição de *comprimento* de A. Se A é um intervalo (aberto, fechado, ou semi-fechado) com pontos extremos a e b, é razoável tomar o comprimento de A como $\mu(A) = b - a$. Se A é um conjunto complicado, podemos não ter qualquer intuição com respeito ao seu comprimento, mas pode-se mostrar que a exigência $\mu((a,b]) = b - a$ para todo $a, b \in R$,com $a < b$, determina μ sobre uma grande classe de conjuntos.

Específicamente, μ é determinada sobre a coleção dos *conjuntos de Borel* de $\mathbb{R}$, denotada por $\mathcal{B}(\mathbb{R})$ e definida como a menor σ-álgebra de subconjuntos de $\mathbb{R}$ contendo todos os intervalos $(a, b], a, b \in \mathbb{R}$.

Note que a existência de $\mathcal{B}(\mathbb{R})$ é garantida; e pode ser descrita como a intersecção de todas as σ-álgebras contendo os intervalos $(a, b]$. Também, se uma σ-álgebra contém digamos, todos os intervalos abertos, ela deve conter todos os intervalos $(a, b]$, e recíprocamente. Pois

$$(a,b] = \bigcap_{n=1}^{\infty} \left(a, b + \frac{1}{n}\right) \qquad (a,b) = \bigcup_{n=1}^{\infty} \left(a, b - \frac{1}{n}\right].$$

Sendo assim $\mathcal{B}(\mathbb{R})$ é a menor σ-álgebra contendo todos os intervalos abertos. De um modo geral se $\mathbf{X}$ é um espaço topológico denomina-se σ-álgebra de Borel de $\mathbf{X}$ à σ-álgebra gerada pela classe dos conjuntos abertos. Os conjuntos em $\mathcal{A}$ denominam-se boreleanos de $\mathbf{X}$. Similarmente podemos trocar as classes de intervalos $(a, b]$ por outras classes de intervalos, por exemplo,

todos os intervalos fechados,

todos os intervalos $[a, b), a, b, \in \mathbb{R}$,

todos os intervalos $(a, \infty), a \in \mathbb{R}$,

todos os intervalos $[a, \infty), a \in \mathbb{R}$,

todos os intervalos $(-\infty, b), b \in \mathbb{R}$,

todos os intervalos $(-\infty, b], b \in \mathbb{R}$.

Visto que uma σ-álgebra que contém todos os intervalos de um dado tipo contém todos os intervalos de qualquer outro tipo, $\mathcal{B}(\mathbb{R})$ pode ser descrita como a menor σ-álgebra que contém a classe de todos os intervalos de $\mathbb{R}$. Similarmente, $\mathcal{B}(\mathbb{R})$ é a menor σ-álgebra contendo todos os conjuntos abertos de $\mathbb{R}$. (Para ver isto, basta lembrar que um conjunto aberto é uma união enumerável de intervalos abertos). Visto que um conjunto é aberto se e sòmente se seu complementar é fechado, $\mathcal{B}(\mathbb{R})$ é a menor σ-álgebra contendo todos os intervalos fechados de $\mathbb{R}$. Finalmente se $\mathcal{A}_0$ é a álgebra de uniões finitas disjuntas de intervalos semi-fechados à direita, então $\mathcal{B}(\mathbb{R})$ é a menor σ-álgebra contendo os conjuntos de $\mathcal{A}_0$.

Intuitivamente, podemos imaginar a obtenção dos conjuntos de Borel da seguinte maneira: iniciando com os intervalos tomamos complementos, uniões e intersecções enumeráveis dos mesmos de todas as maneiras possíveis.

Definição: Uma função de conjunto μ definida sobre $\mathcal{A}$ é dita *finita* sse $\mu(A)$ é finita, isto é, não assume $\pm\infty$, para cada $A \in \mathcal{A}$.

Uma função de conjunto μ não-negativa e enumerávelmente aditiva sobre $\mathcal{A}$ é dita σ-finita sobre $\mathcal{A}$ sse Ω pode ser escrito como uma $\cup_{n=1}^{\infty} A_n$ onde A_n pertence a $\mathcal{A}$ e $\mu(A_n) < \infty$ para todo n.

1.1.2 -*Medidas Sobre Espaços Produtos Infinitos*

Definição. Para cada $j = 1, 2, \cdots$, seja $(\Omega_j, \mathcal{A}_j)$ um espaço mensurável. Seja

$$\Omega = \prod_{j=1}^{\infty} \Omega_j,$$

o conjunto de todas as seqüências $(\omega_1, \omega_2, \cdots)$ sendo $\omega_j \in \Omega_j, j = 1, 2, \cdots$ Se

$$B^n \subset \prod_{j=1}^{n} \Omega_j,$$

definimos

$$B_n = \{\omega \in \Omega : (\omega_1, \cdots, \omega_n) \in B^n\}.$$

o conjunto B_n é chamado o *cilindro* com *base* B^n; o cilindro é dito *mensurável* se $B^n \in \prod_{j=1}^{n} \mathcal{A}_j$. Se $B^n = A_1 \times \cdots \times A_n$, onde $A_i \subset \Omega_i$ para cada i, B_n é chamado um *retângulo*, um *retângulo mensurável* se $A_i \in \mathcal{A}_i$ para cada i.

Os cilindros mensuráveis formam uma álgebra. É também verdade que uniões finitas disjuntas de retângulos mensuráveis formam uma álgebra.

Exemplo: Suponha que uma moeda é lançada repetidamente e que ω_n é igual a 1 se o n-ésimo lançamento é cara e 0 se corôa, deste modo o resultado do experimento é uma seqüência $\omega = (\omega_1, \omega_2, \cdots)$. Seja

$$\xi(\omega) = \sum_{n=1}^{\infty} \frac{\omega_n}{2^n}$$

Se $\lambda_1, \lambda_2, \cdots, \lambda_N$ são ou 0 ou 1, então a imagem sob ξ do conjunto dos ω para os quais $\omega_1 = \lambda_1, \omega_2 = \lambda_2, \cdots, \omega_N = \lambda_N$ é um subconjunto de Borel de $[0,1]$ com medida de Lebesgue $1/2^N$.

Proposição. Para cada $j = 1, 2, \cdots$, seja $(\Omega_j, \mathcal{A}_j, P_j)$ um espaço de probabilidade arbitrário. Sejam

$$\Omega = \prod_{j=1}^{\infty} \Omega_j \,, \; \mathcal{A} = \prod_{j=1}^{\infty} \mathcal{A}_j.$$

Existe uma única medida de probabilidade P sobre $\mathcal{A}$ tal que

$$P\{\omega \in \Omega : \omega_1 \in A_1, \cdots, \omega_n \in A_n\} = \prod_{j=1}^{n} P_j(A_j)$$

para todo $n = 1, 2, \cdots$ e todo $A_j \in \mathcal{A}_j, j = 1, 2, \cdots$

Prova: [8]

Chamamos P o *produto* de P_j, e escrevemos $P = \prod_{j=1}^{\infty} P_j$.

Definição(átomo). Seja $(\Omega, \mathcal{A}, \mu)$ um espaço de medida. Um conjunto $A \in \mathcal{A}$ é chamado um *átomo* se $\mu(A) > 0$ e para todo $B, B \subseteq A$ resulta $\mu(B) = 0$ ou $\mu(B) = \mu(A)$. μ é chamada *não-atômica* se não tem átomos.

Ainda: Um *átomo* de qualquer medida μ sobre $\mathcal{B}(\mathbb{R})$ é um conjunto unitário $\{x\}$ de modo que $\mu(\{x\}) > 0$.

Teremos ainda oportunidade de usar os seguintes resultados da teoria de medida:

Teorema 1.1. Se $\mathcal{A}$ é a σ-álgebra dos boreleanos de $\mathbb{R}^n$ existe uma única medida $\lambda: \mathcal{A} \to [0, +\infty]$ tal que, se $A = (a_1, b_1) \times \cdots \times (a_n, b_n)$,

$$\lambda(A) = (b_1 - a_1) \cdots (b_n - a_n).$$

Esta medida denomina-se *medida de Lebesgue.*

Se $(X, \mathcal{A}, \mu)$ é um espaço de medida, dizemos que um conjunto $A \subset X$ é de medida zero se existe $A_1 \in \mathcal{A}$ tal que $A \subset A_1$ e $\mu(A_1) = 0$.

Dizemos que dois conjuntos $A_1, A_2 \subset X$ coincidem mod(0), e denotamos $A_1 = A_2$ mod(0) se $A_1 \triangle A_2$ é de medida zero. Se $\mathcal{S}$ é uma família de subconjuntos de X, escrevemos $A \in \mathcal{S}$ mod(0) se $A = A_0$ mod(0) para algum $A_0 \in \mathcal{S}$ e definimos

$$\mathcal{S}mod(0) = \{A \subset X : A \in \mathcal{S}\ mod(0)\}.$$

Dizemos que $\mathcal{S}$ gera $\mathcal{A}$ mod(0) se $\mathcal{A} = \mathcal{A}_0$ mod(0), onde $\mathcal{A}_0$ é a σ-álgebra gerada por $\mathcal{S}$. $A \subset B$ mod(0) significa que $B - A$ é de medida zero. Uma propriedade aplicada a pontos de um conjunto $S \subset X$ vale em *quase todo ponto* $x \in S$ (abreviadamente q.t.p.) se o conjunto dos pontos de S onde é falsa é de medida zero.

Teorema de Extensão de Hahn-Kolmogorov. Seja X um conjunto, $\mathcal{A}_0$ uma álgebra de subconjuntos de X e $\mu_0: \mathcal{A}_0 \to [0,1]$ uma medida. Então se $\mathcal{A}$ é a σ-álgebra gerada por $\mathcal{A}_0$ existe uma e só uma medida $\mu: \mathcal{A} \to [0,1]$ tal que $\mu/\mathcal{A}_0 = \mu_0$.

1.1.3 -*Integração*

Seja $\mathcal{B}(\mathbb{R})$ denotando a σ-álgebra dos subconjuntos de Borel de $\mathbb{R}$. Esta é a σ-álgebra gerada pela coleção dos subconjuntos abertos de $\mathbb{R}$ e é também gerada pela coleção de todos os intervalos, ou ainda, pela coleção de todos os intervalos da forma (c, ∞).

Seja $(X, \mathcal{B}, \mu)$ um espaço de medida. Uma função $f: X \to \mathbb{R}$ é *mensurável* se $f^{-1}(A) \in \mathcal{B}$ sempre que $A \in \mathcal{B}(\mathbb{R})$ ou equivalentemente se $f^{-1}(c, \infty) \in \mathcal{B}$ para todo $c \in \mathbb{R}$. Uma função $f: X \to \mathbb{C}$ é mensurável se ambas as partes real e imaginária são mensuráveis. Se X é um espaço topológico e $\mathcal{B}$ é a σ-álgebra gerada pelos subconjuntos abertos de X, então qualquer função contínua $f: X \to \mathbb{C}$ é mensurável. Dizemos que $f = g$ q.t.p. se $\mu(\{x : f(x) \neq g(x)\}) = 0$. Suponha X um espaço topológico, $\mathcal{B}(X)$ a σ-álgebra dos conjuntos de Borel e μ uma medida sobre $(X, \mathcal{B})$ com a propriedade de que cada conjunto aberto não-vazio tem medida não-nula. Então para duas funções contínuas $f, g: X \to \mathbb{R}$, $f = g$ q.t.p. implica $f = g$ porque $\{x : f(x) - g(x) \neq 0\}$ é um conjunto aberto de medida nula.

Seja $(X, \mathcal{B}, \mu)$ um espaço de probabilidade. Uma função $f: X \to \mathbb{R}$ é uma *função simples* se pode ser escrita na forma $\sum_{i=1}^{n} a_i \chi_{A_i}$, onde $a_i \in \mathbb{R}$, $A_i \in \mathcal{B}$, os conjuntos A_i são subconjuntos disjuntos de X, e χ_{A_i} a função característica de A_i. Funções simples são mensuráveis. Definimos a integral para funções simples por:

$$\int f \, d\mu = \sum_{i=1}^{n} a_i \mu(A_i).$$

Este valor é independente da representação $\Sigma_i a_i \chi_{A_i}$.

Suponha $f: X \to \mathbb{R}$ mensurável $f \geq 0$. Então existe uma seqüência crescente de funções simples $f_n \nearrow f$. Por exemplo podemos tomar

$$f_n(x) = \begin{cases} \frac{i-1}{2^n}, & \frac{i-1}{2^n} \leq f(x) < \frac{i}{2^n} \quad i = 1, \cdots, n2^n \\ n, & f(x) \geq n. \end{cases}$$

Definimos

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

note que esta definição independe da seqüência escolhida $\{f_n\}$. Dizemos que f é integrável se $\int f \, d\mu < \infty$. Suponha $f: X \to \mathbb{R}$ mensurável. Então $f = f^+ - f^-$ onde $f^+(x) = max\{f(x), 0\} \geq 0$ e $f^-(x) = max\{-f(x), 0\} \geq 0$. Dizemos que f é integrável se $\int f^+ d\mu, \int f^- d\mu < \infty$ e então definimos

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

Dizemos que $f: X \to \mathbb{C}$ é integrável $(f = f_1 + i f_2)$ se f_1 e f_2 são integráveis e definimos

$$\int f \, d\mu = \int f_1 \, d\mu + i \int f_2 \, d\mu.$$

Observe que f é integrável se e somente se $|f|$ é integrável. Se $f = g$ (q.t.p.) então uma delas é integrável se e sòmente se a outra o é, e $\int f \, d\mu = \int g \, d\mu$.

Dois teoremas básicos sobre integração de seqüências de funções são os seguintes:

Teorema 1.2.(*Teorema da convergência dominada*). Suponha $f_1 \leq f_2 \leq f_3 \leq \cdots$ uma seqüência crescente de funções a valores reais integráveis sobre $(X, \mathcal{B}, \mu)$. Se $\{\int f_n \, d\mu\}$ é uma seqüência limitada de números reais então $\lim_{n \to \infty} f_n$ existe em q.t.p. e é integrável e $\int(\lim f_n)\,d\mu = \lim \int f_n\,d\mu$. Se $\{\int f_n\,d\mu\}$ é uma seqüência ilimitada então ou $\lim_{n\to\infty} f_n$ é infinito sobre um conjunto de medida positiva ou $\lim_{n\to\infty} f_n$ não é integrável.

Teorema 1.3.(*Lema de Fatou's*). Seja $\{f_n\}$ uma seqüência mensurável de funções a valores reais sobre $(X, \mathcal{B}, \mu)$ que é limitada abaixo por uma função integrável. Se $\liminf_{n\to\infty} \int f_n\,d\mu$ ∞ então $\liminf_{n\to\infty} f_n$ é integrável e $\int \liminf f_n d\mu \leq \liminf \int f_n\,d\mu$.

Corolário(*Teorema da convergência dominada*). Se $g: X \to \mathbb{R}$ é integrável e $\{f_n\}$ é uma seqüência de funções mensuráveis a valores reais com $|f_n| \leq g$ q.t.p $(n \geq 1)$ e $\lim_{n\to\infty} f_n = f$ q.t.p. então f é integrável e $\lim \int f_n\,d\mu = \int f\,d\mu$.

Denotamos por $\mathcal{L}^1(X, \mathcal{B}, \mu)$ (ou $\mathcal{L}^1(\mu)$) o espaço de todas as funções integráveis $f: X \to \mathbb{C}$ onde duas de tais funções são identificadas se são iguais em q.t.p. Contudo escrevemos $f \in \mathcal{L}^1(X, \mathcal{B}, \mu)$ para denotar que $f: X \to \mathbb{C}$ é integrável. O espaço $\mathcal{L}^1(X, \mathcal{B}, \mu)$ é um espaço de Banach com norma $\|f\|_1 = \int |f|\,d\mu$.

Se $f \in \mathcal{L}^1(X, \mathcal{B}, \mu)$, então $\int_A f\,d\mu$ denota $\int f.\chi_A\,d\mu$.

Medidas Absolutamente Contínuas

Seja $(X, \mathcal{B})$ um espaço mensurável e suponha μ, ν duas medidas de probabilidade sobre $(X, \mathcal{B})$. Dizemos que μ é *absolutamente contínua* com repeito a ν $(\mu \ll \nu)$ se $\nu(A) = 0$ implica $\mu(A) = 0$. Estas medidas são equivalentes se $\mu \ll \nu$ e $\nu \ll \mu$. O seguinte teorema caracteriza a continuidade absoluta.

Teorema 1.4.(Teorema de Radon-Nikodym). Sejam μ, ν duas medidas de probabilidade sobre o espaço mensurável $(X, \mathcal{B})$. Então $\mu \ll \nu$ se e sòmente se existe $f \in \mathcal{L}^1(\nu)$, com $f \geq 0$ e $\int f\,d\nu = 1$, de modo que $\mu(A) = \int_A f d\nu \;\; \forall A \in \mathcal{B}$. A função f é única q.t.p. (no sentido de que qualquer outra função com estas propriedades é igual a f q.t.p.).

A função f é chamada a derivada de Radon-Nikodym de μ com respeito a ν e é denotada por $f = d\mu/d\nu$.

A noção "oposta" de continuidade absoluta é como segue. Duas medidas de probabilidade μ, ν sobre $(X, \mathcal{B})$ são ditas *mutuamente singulares* $(\mu \perp \nu)$ se existe algum $A \in \mathcal{B}$ com $\mu(A) = 0$ e $\nu(X \setminus A) = 0$.

Partições e Derivação

Uma partição $\mathcal{P}$ de um espaço de probabilidade $(X, \mathcal{A}, \mu)$ é uma família de subconjuntos em $\mathcal{A}$ com medida $\neq 0$ tais que:

$$A \in \mathcal{P},\ B \in \mathcal{P} \Rightarrow \mu(A \cap B) = 0$$

$$\mu(X - \bigcup_{A \in \mathcal{P}} A) = 0.$$

Segue destas propriedades que $\mathcal{P}$ é uma família finita ou enumerável. Os conjuntos da família denominam-se átomos da partição. Se $\mathcal{P}_n$, $n = 1, \cdots, N$, são partições, definimos a partição $\bigvee_1^N \mathcal{P}_n$ como aquela cujos átomos são todos conjuntos da forma $A_1 \cap \cdots \cap A_n$ com $A_i \in \mathcal{P}_i$, $i = 1, \cdots, n$, e tais que

$$\mu(A_1 \cap \cdots \cap A_n) \neq 0.$$

Se $\mathcal{P}$ e Q são partições, dizemos que Q é mais fina que $\mathcal{P}$, o que denotamos $\mathcal{P} \leq Q$, se todo átomo de Q está contido mod(0) em algum átomo de $\mathcal{P}$. Isto implica que todo átomo de $\mathcal{P}$ pode ser escrito como uma união de átomos de Q mod(0).

Teorema 1.5. Seja $(X, \mathcal{A}, \mu)$ um espaço de probabilidade e $\mathcal{P}_1 \leq \mathcal{P}_2 \leq \cdots$ uma seqüência de partições tal que $\bigvee_{n \geq 1} \mathcal{P}_n = \mathcal{A}$ mod(0). Então, se $\mathcal{P}_n(x)$ denota o átomo de $\mathcal{P}_n$ que contém x, e $f \in \mathcal{L}^1(X, \mathcal{A}, \mu)$, para q.t.p. $\in X$ vale :

$$\lim_{n \to +\infty} \frac{1}{\mu(\mathcal{P}_n(x))} \int_{\mathcal{P}_n(x)} f \, d\mu = f(x)$$

e se $f_n: X \to \mathbb{C}$ é a função definida como:

$$f_n(x) = \frac{1}{\mu(\mathcal{P}_n(x))} \int_{\mathcal{P}_n(x)} f \, d\mu$$

então $f_n \to f$ em $\mathcal{L}^1(X, \mathcal{A}, \mu)$.

Prova: [2]

Teorema 1.6.(*Fórmula de mudança de variável*)

Sejam $\mathcal{U} \subseteq \mathbb{R}^n, V \subseteq \mathbb{R}^n$ dois conjuntos abertos e $g: \mathcal{U} \to V$ uma aplicação localmente lipschitziana com inversa g^{-1} localmente lipschitziana. Seja f uma função real mensurável sobre V não-negativa ou integrável. Então,existe uma função J_g definida sobre $\mathcal{U}$, mensurável, finita e localmente integrável, tal que

$$\int_V f(y) \, \lambda(dy) = \int_{\mathcal{U}} (f \circ g)(x) J_g(x) \, \lambda(dx).$$

Esta função tem a seguinte propriedade:

$$J_g(x) = \lim_{r \to 0} \frac{\lambda(g(B(x,r)))}{\lambda(B(x,r))} \qquad q.t.p.$$

Se g é um difeomorfismo de classe C^1, então

$$J_g(x) = |det(\frac{\partial g}{\partial x})(x)|,$$

onde o membro direito indica o valor absoluto do determinante da matriz jacobiana $(\frac{\partial g}{\partial x})$ no ponto x.

Prova: [4]

1.2 -Elementos de Teoria Ergódica

1.2.1 - *Transformações que Preservam medida*

A Teoria Ergódica dentre outros afazeres se preocupa em estudar a dinâmica das transformações que preservam medida.

Neste parágrafo devemos discutir transformações que preservam medida e algumas de suas propriedades básicas, bem como alguns exemplos.

Dada uma aplicação $T: X \to X$, podemos definir as iteradas T^n obtidas por composição de T consigo mesma n vezes. Por conveniência T^0 deve denotar a aplicação identidade, e T^{-n} deve ser definida como uma aplicação de conjunto

$$T^{-n}(E) = \{x : T^n(x) \in E\}$$

É importante frisar que o estudo de transformações que preservam medida tem início com certas considerações em mecânica estatística. Suponha que temos um sistema com k partículas cujos estado presente é descrito por um ponto no "espaço de fase"$\mathbb{R}^{6k}$ no qual cada partícula determina 3 coordenadas para posição e 3 coordenadas para o momento. Então a história inteira do sistema pode ser representada por uma trajetória no espaço de fase o qual é completamente determinado (assumindo as leis clássicas da mecânica) por um único ponto do mesmo. Deste modo para qualquer tempo t podemos definir uma transformação T_t invertível dizendo que, para $\mathbf{x}$ no espaço de fase, $T_t\mathbf{x}$ denota o estado do sistema no tempo t tendo o mesmo iniciado em $\mathbf{x}$. Um dos resultados básicos em mecânica estatística (devido a Liouville) estabelece que, se as coordenadas no espaço de fase são corretamente escolhidas, então o "fluxo" no espaço de fases leva todo volume (i.e., medida de Lebesgue em $\mathbb{R}^{6k}$) inalterado. Isto significa que T_t torna-se uma transformação que preserva medida em $(\mathbb{R}^{6k}, \mathcal{L}^{6k}, \mu)$. Na prática k é enorme, e não é possível observar qualquer momento de todas as partículas do sistema. Ao invés de perguntar, por exemplo, "qual a probabilidade de que no tempo t o estado do sistema pertença a um dado subconjunto do espaço de fase?", impomos condições que garantam que isto pode ser calculado por consideração do comportamento "médio" de $T_t\mathbf{x}$ quando $t \to \infty$. Para sermos mais precisos $T_{s+t} = T_s T_t$ de modo que $T_{nt} = T_t^n$ e, desta forma, possamos considerar um modelo discreto, contando a proporção de vezes (sob iterações n) em que $T^i\mathbf{x} \in E$ onde $T = T_{t_0}$ e $n \to \infty$. Na prática o conjunto E no espaço de fase é trocado por uma função $f(\mathbf{x})$ (representando alguma medida física) e consideramos o comportamento médio em termos da seqüência

$$\frac{1}{n}\sum_{i=0}^{n-1} f(T^i\mathbf{x}) \qquad (n = 1, 2, \cdots).$$

Definição. Suponha que $(X_1, \mathcal{B}_1, \mu_1)$, $(X_2, \mathcal{B}_2, \mu_2)$ são espaços de probabilidade.

(a) Uma transformação $T: X_1 \to X_2$ é mensurável se $T^{-1}(\mathcal{B}_2) \subset \mathcal{B}_1$ (i.e. se $B_2 \in \mathcal{B}_2 \Rightarrow T^{-1}B_2 \in \mathcal{B}_1$).

(b) Dizemos que uma transformação $T: X_1 \to X_2$ preserva medida se T é mensurável e $\mu_1(T^{-1}(B_2)) = \mu_2(B_2)$, $\forall B_2 \in \mathcal{B}_2$.

(c) Dizemos que uma transformação que preserva medida $T: X_1 \to X_2$ é invertível se T é bijetiva, e T^{-1} também preserva medida.

Obs: Estaremos interessados principalmente nos casos em que $(X_1, \mathcal{B}_1, \mu_1) = (X_2, \mathcal{B}_2, \mu_2)$ tendo em conta que queremos estudar as iteradas de T^n. Quando $T: X \to X$ é uma transformação que preserva medida de $(X, \mathcal{B}, \mu)$ dizemos que T preserva μ ou que μ é T-invariante.

Na prática é difícil checar, usando a definição anterior, quando uma dada transformação preserva medida ou não, visto que amiude não temos conhecimento explícito de todos os membros de $\mathcal{B}$. Contudo freqüentemente temos conhecimento de uma semi-álgebra $\mathcal{B}_0$ que gera $\mathcal{B}$ (por exemplo, quando X é o intervalo unitário $\mathcal{B}_0$ pode ser a semi-álgebra de todos os subintervalos de X, e quando X é o espaço produto direto $\mathcal{B}_0$ pode ser a coleção de todos os retangulos mensuráveis.) O seguinte resultado é por essa razão útil para checar quando ou não uma transformação preserva medida.

Teorema 1.7. Suponha que $(X_1, \mathcal{B}_1, \mu_1)$, $(X_2, \mathcal{B}_2, \mu_2)$ são espaços de probabilidade e $T: X_1 \to X_2$ é uma transformação. Seja $\mathcal{B}_{02}$ uma semi-álgebra que gera $\mathcal{B}_2$. Se para cada $A_2 \in \mathcal{B}_{02}$ tivermos $T^{-1}(A_2) \in \mathcal{B}_1$ e $\mu_1(T^{-1}(A_2)) = \mu_2(A_2)$ então T preserva medida.

Prova. Seja $\mathcal{C}_2 = \{B \in \mathcal{B}_2 : T^{-1}(B) \in \mathcal{B}_1, \mu_1(T^{-1}(B)) = \mu_2(B)\}$. Queremos mostrar que $\mathcal{C}_2 = \mathcal{B}_2$. Contudo $\mathcal{B}_{02} \subseteq \mathcal{C}_2$ e cada membro da álgebra $\mathcal{A}(\mathcal{B}_{02})$ gerada por $\mathcal{B}_{02}$ é uma união finita disjunta de membros de $\mathcal{B}_{02}$ temos $\mathcal{A}(\mathcal{B}_{02}) \subset \mathcal{B}_{02}$. Visto que $\mathcal{B}_{02}$ é uma classe monótona, o resultado segue visto que a σ-álgebra gerada por $\mathcal{A}(\mathcal{B}_{02})$ é a classe monótona gerada por $\mathcal{A}(\mathcal{B}_{02})$. □

Exemplos de Transformações que Preservam medidas

Em teoria da medida mostra-se, como propriedades geométricas da medida de Lebesgue, que transformações no espaço Euclidiano definidas em termos de translações, rotações ou reflexões são tranformações que preservam medida. Podemos também provar que transformações com matriz de determinante 1 definem transformações que preservam medida no espaço Euclidiano.

(1) A transformação identidade I sobre $(X, \mathcal{A}, \mu)$ obviamente preserva medida.

(2) Seja ρ um conjunto finito de r elementos. A medida de probabilidade geral sobre ρ -i.e. sobre a σ-álgebra de todos os subconjuntos de $\rho-$ é especificada pelo assinalamento de probabilidades não-negativas p_i para os elementos i de ρ de tal modo que $\Sigma_{i\in\rho} p_i = 1$. Seja $(\Omega, \mathcal{A}, P)$ o espaço produto resultante. O elemento geral de Ω é uma seqüência dupla infinita $\omega = (\cdots, \omega_{-1}, \omega_0, \omega_1, \cdots)$ de elementos de ρ. Seja x_n a n-ésima função coordenada- a aplicação de Ω para ρ cujo valor $x_n(\omega)$ no ponto ω é a n-ésima coordenada ω_n de ω. A σ-álgebra $\mathcal{A}$ é gerada pela álgebra(finitamente

aditiva) consistindo de cilindros, ou conjuntos da forma

$$\{\omega : (x_n(\omega)), \cdots, x_{n+k-1}(\omega)) \in E\} = \{\omega : (\omega_n, \cdots, \omega_{n+k-1}) \in E\}$$

onde E é um subconjunto do produto cartesiano ρ^k de k cópias de ρ. A σ-álgebra $\mathcal{F}$ é também gerada pelo que podemos chamar de cilindros "finos", isto é conjuntos da forma

$$\{\omega : x_l(\omega) = i_l,\ n \leq l < n+k\},$$

onde i_l são elementos de ρ; de fato, cada cilindro é uma união finita, disjunta de cilindros finos. Finalmente, P é especificada por seus valores sobre os cilindros finos:

$$P\{\omega : x_l(\omega) = i_l,\ n \leq l < n+k\} = \prod_{l=n}^{n+k-1} p_{i_l}.$$

Deste modo $\{\cdots, x_{-1}, x_0, x_1, \cdots\}$ é uma seqüência de variáveis randômicas com valores em ρ. Seja $T\colon \Omega \to \Omega$ que leva $\{\cdots, \omega_{-1}, \omega_0, \omega_1, \cdots\}$ para $\{\cdots, \omega_0, \omega_1, \omega_2, \cdots\}$. Mais precisamente, T está definida por $(T\omega)_n = \omega_{n+1}$, ou o que é a mesma coisa $x_n(T\omega) = x_{n+1}(\omega)$. Visto que $x_n(\omega) = x_0(T^n(\omega))$, qualquer afirmativa sobre a variável randômica x_n pode ser convertida a uma afirmação sobre x_0 e T. Se A é qualquer cilindro então claramente $T^{-1}A$ é também um cilindro e conseqüentemente mora em $\mathcal{A}$, e $P(T^{-1}A) = P(A)$. Que T é mensurável e preserva medida é uma conseqüência do teorema 1.7.

(3) Seja S^1 o círculo unitário no plano complexo, com $\mathcal{A}$ consistindo dos subconjuntos de Borel ordinários de S^1 ($\mathcal{A}$ é a σ-álgebra gerada pelos arcos), e seja λ a medida circular de Lebesgue sobre S^1, normalizada, i.e. tal que $\lambda(S^1) = 1$. Para um elemento c de S^1, seja $Tx = cx$. Visto que T é efetivamente uma rotação do círculo, sempre com angulo $arg\, c$, T preserva P.

(4) Seja $\mathcal{A}$ consistindo dos subconjuntos de Borel do intervalo unitário semi-aberto $I = [0, 1)$, com medida de Lebesgue λ, e seja $Tx = 2x (mod\, 1)$, isto é

$$Tx = \begin{cases} 2x, & x \in [0, \frac{1}{2}) \\ 2x - 1, & x \in [\frac{1}{2}, 1) \end{cases}$$

obs: T pode ser escrita ainda na forma $Tx = 2x - [2x]$.

A transformação T está associada com a expansão na base-2 de um ponto do intervalo unitário; de fato, se $f(x)$ é 0 para $x < \frac{1}{2}$ e 1 para $x \geq \frac{1}{2}$ (i.e., $f = \chi_{[1/2,1)}$) então $f(T^{n-1}x)$ é o n-ésimo dígito da expansão diádica (base-2) de

$$x = \sum_{n=1}^{\infty} \frac{f(T^{n-1}x)}{2^n}$$

Sendo assim, se x tem a expansão $x = .x_1x_2\cdots$, então $Tx = .x_2x_3\cdots$ Uma aplicação do teorema 1.7 mostra que T preserva medida. De fato (para a medida de Lebesgue) por consideração do efeito de T^{-1} sobre os intervalos $[p/2^q, (p+1)/2^q)$ os quais formam um semi-anel gerando $\mathcal{B}$.

Uma propriedade possuída por todas as transformações que preservam medida é a recorrência:

Teorema 1.8.(*Teorema de recorrência de Poincaré*). Seja $T: X \to X$ uma transformação que preserva medida de um espaço de probabilidade $(X, \mathcal{B}, \mu)$. Seja $E \in \mathcal{B}$ com $\mu(E) > 0$. Então quase todo ponto de E retorna infinitas vezes para E sob iterações positivas de T (i.e., existe $F \subset E$ com $\mu(F) = \mu(E)$ de modo que para cada $x \in F$ existe uma seqüência $n_1 < n_2 < n_3 < \cdots$ de números naturais com $T^{n_i}(x) \in F$ para cada i).

Prova. Para $N \geq 0$ seja $E_N = \bigcup_{n=N}^{\infty} T^{-n}E$. Então $\bigcap_{N=0}^{\infty} E_N$ é o conjunto de todos os pontos de X os quais entram em E infinitas vezes sob iterações positivas por T. Sendo assim o conjunto $F = E \cap \bigcap_{N=0}^{\infty} E_N$ consiste de todos os pontos de E que entram em E infinitas vezes sob iterações positivas por T. Se $x \in F$ então existe uma seqüência $0 < n_1 < n_2 < \cdots$ de números naturais com $T^{n_i}(x) \in E$ para todo i. Para cada i temos que $T^{n_i}(x) \in F$ porque $T^{n_j - n_i}(T^{n_i}x) \in E$ para todo j. Sendo assim resta mostrar que $\mu(F) = \mu(E)$.

Visto que $T^{-1}E_N = E_{N+1}$ temos que $\mu(E_N) = \mu(E_{N+1})$ e conseqüentemente $\mu(E_0) = \mu(E_N)$ para todo N. Visto que $E_0 \supset E_1 \supset E_2 \supset \cdots$ temos que $\mu(\bigcap_{N=0}^{\infty} E_N) = \mu(E_0)$. Sendo assim $\mu(F) = \mu(E \cap E_0) = \mu(E)$ visto que $E \subset E_0$.

1.2.2 -*Ergodicidade*

A começos do século os trabalhos de Boltzmann e Gibbs sobre Mecânica Estatística levantaram um problema matemático que em nosso contexto pode ser descrito da seguinte forma: dada uma transformação que preserva medida de um espaço $(X, \mathcal{A}, \mu)$ e uma função integrável $f: X \to \mathbb{R}$ dar condições para que o limite:

$$\lim_{n \to +\infty} \frac{f(x) + f(T(x)) + \cdots + f(T^{n-1}(x))}{n}$$

exista e seja o mesmo em q.t.p. De fato questões de mesmo teor já tinham aparecido em outras áreas da matemática, como por exemplo no problema do movimento médio do periélio em Mecânica Celeste. Em 1931 Birkhoff provou que quaisquer que sejam T e f o limite acima existe em q.t.p. A partir deste resultado mostrou que a condição necessária e suficiente para que o valor do limite seja o mesmo em quase todo ponto é que não exista nenhum conjunto $A \in \mathcal{A}$ tal que $0 < \mu(A) < 1$ e $T^{-1}(A) = A$. Sabendo que o limite é constante q.t.p. não é difícil mostrar que é exatamente a integral de f sobre X, e neste caso a transformação denomina-se ergódica. O resultado de Birkhoff não fechou o problema que o motivou, porque nas transformações da Mecânica Estatística não foi possível constatar a ergodicidade. Só nos anos 60 os trabalhos de Sinai, e mais recentemente os de Bunimovitch, provaram a ergodicidade de transformações análogas àquelas.

Seja T uma transformação que preserva medida de um espaço $(X, \mathcal{A}, \mu)$. Um conjunto $A \in \mathcal{A}$ é dito T-invariante se $T^{-1}A = A$.

Definição(Ergodicidade). Uma transformação T que preserva medida é dita ergódica (ou metricamente transitiva ou metricamente invariante) se para todo conjunto invariante A, $\mu(A) = 0$ ou $\mu(X \setminus A) = 0$.

De outra forma:

Definição. Seja $(X, \mathcal{A}, \mu)$ um espaço de probabilidade. Uma transformação T de $(X, \mathcal{A}, \mu)$ que preserva medida é chamada ergódica se os únicos membros A de $\mathcal{A}$ com $T^{-1}A = A$ satisfazem $\mu(A) = 0$ ou $\mu(A) = 1$.

Existem outras maneiras de definir ergodicidade como podemos ver do seguinte teorema:

Teorema 1.9. Se $T: X \to X$ é uma transformação que preserva medida em um espaço de probabilidade $(X, \mathcal{B}, \mu)$ então as seguintes afirmações são equivalentes:

(i) T é ergódica.

(ii) Os únicos membros B de $\mathcal{B}$ com $\mu(T^{-1}B \triangle B) = 0$ são aqueles com $\mu(B) = 0$ ou $\mu(B) = 1$.

(iii) Para todo $A \in \mathcal{B}$ com $\mu(A) > 0$ temos $\mu(\bigcup_{n=1}^{\infty} T^{-n}A) = 1$.

(iv) Para todo $A, B \in \mathcal{B}$ com $\mu(A) > 0$ existe $n > 0$ com $\mu(T^{-n}A \cap B) > 0$.

Prova: [5]

O primeiro grande resultado em teoria ergódica foi provado em 1931 por G.D.Birkhoff. Devemos enunciar este resultado para transformações que preservam medida de um espaço de medida σ-finito.

Teorema 1.10.(*Teorema ergódico de Birkoff.*) Suponha $T: (X, \mathcal{B}, \mu) \to (X, \mathcal{B}, \mu)$ uma transformação que preserva medida. (onde assumimos $(X, \mathcal{B}, \mu)$ σ-finito e $f \in \mathcal{L}^1(\mu)$). Então

$$\frac{1}{n}\sum_{i=0}^{n-1} f(T^i(x))$$

converge em q.t.p. para uma função $f^* \in \mathcal{L}^1(\mu)$. Também $f^* \circ T = f^*$ q.t.p. e se $\mu(X) < \infty$, então $\int f^* \, d\mu = \int f \, d\mu$.

Prova: [5]

Observações:

Se T é ergódica então f^* é constante em q.t.p. e deste modo se $\mu(X) < \infty$ temos $f^* = (1/\mu(X)) \int f \, d\mu$ q.t.p. Se $(X, \mathcal{B}, \mu)$ é um espaço de probabilidade e T é ergódica temos que

$$\forall f \in \mathcal{L}^1(\mu) \qquad \lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} f(T^i x) = \int f \, d\mu \quad q.t.p.$$

(i) Seja $(X, \mathcal{B}, \mu)$ um espaço de probabilidade e $T: X \to X$ preserva-medida. Seja $E \in \mathcal{B}$. Para $x \in X$ podemos perguntar com que freqüência os elementos do conjunto $\{x, T(x), T^2(x), \cdots\}$ estão no conjunto E?

Claramente $T^i(x) \in E$ se e sòmente se $\chi_E(T^i(x)) = 1$, sendo assim o número de elementos de $\{x, T(x), \cdots, T^{n-1}(x)\}$ em E é

$$\sum_{k=0}^{n-1} \chi_E(T^k(x)).$$

O percentual relativo de elementos de $\{x, T(x), \cdots, T^{n-1}(x)\}$ em E é igual

$$\frac{1}{n}\sum_{k=0}^{n-1}\chi_E(T^k(x)).$$

e se T é ergódica então

$$\frac{1}{n}\sum_{k=0}^{n-1}\chi_E(T^k(x)) \to \mu(E) \qquad q.t.p.$$

pelo teorema ergódico. Deste modo as órbitas de quase todos os pontos de X entram no conjunto E com freqüência relativa assintótica $\mu(E)$.

(ii) Reiteramos que o limite $lim_{n\to\infty}\sum_{i=0}^{n} f(T^i(x))$ sempre existe (é uma outra função $f^* \in \mathcal{L}^1(\mu)$) e que a condição necessária e suficiente para que o mesmo seja constante (q.t.p.) é que T seja ergódica. Nas aplicações do teorema ergódico $f \in \mathcal{L}^1(\mu)$ é escolhida de acôrdo com o problema em questão (i.e., escolhemos f adequada aos nossos propósitos).

- A Ergodicidade da Transformação Diádica

Considere a transformação $T(x) = 2x(mod\,1)$. Se $x = .x_1x_2x_3\cdots$ e $x' = x + 1/2(mod\,1)$ então $x' = .x_1x_2x_3\cdots$ com $x_1' = 1 - x_1$. Sendo assim, assumindo $A = T^{-1}A, x \in A$ é equivalente a $Tx \in A$, e $x' \in A$ é equivalente a $Tx' \in A$; visto que $Tx = Tx' = .x_2x_3\cdots, x \in A$ sse $x' \in A$. Sendo assim, se $E = [0, \frac{1}{2})$, o conjunto $A \cap E^c$ é exatamente o conjunto $A \cap E$ transladado para a direita de uma distância $\frac{1}{2}$. Disto segue que estes dois conjuntos têm a mesma medida (estamos considerando aqui a medida de Lebesgue λ), e conseqüêntemente $\lambda(A) = 2\lambda(A \cap E) = \lambda(A \cap E)/\lambda(E)$. Deste modo A e E são independentes (vistos como eventos de um espaço de probabilidades). Uma elaboração deste argumento mostra que também isto é verdade se E é qualquer intervalo diádico, ou união disjunta de intervalos diádicos. Dado um ε positivo, escolha uma união com E de maneira que $\lambda(A \cup E) < \varepsilon$. Então $|\lambda(A) - \lambda(E)| < \varepsilon$ e $|\lambda(A) - \lambda(A)\lambda(E)| = |\lambda(A) - \lambda(A \cap E)| < \varepsilon$, segue que $|\lambda(A) - \lambda(A)^2| < 2\varepsilon$. Visto que ε foi tomado arbitráriamente, $\lambda(A) = \lambda(A)^2$, e conseqü êntemente $\lambda(A)$ deve ser 0 ou 1. Sendo assim T é ergódica.

- Medidas Sobre Intervalo

Generalizamos a transformação diádica trocando a base 2 por uma base geral $r \geq 2$.

Exemplos:

(a) Seja λ a medida de Lebesgue sobre a classe $\mathcal{B}$ dos subconjuntos de Borel de $\Omega = [0, 1)$, mas defina T por $Tx = rx(mod\, 1)$. Se $f(x) = i$ sobre $[i/r, (i+1)/r)$, $\quad i = 0, 1, \cdots, r-1$, então x tem

$$\sum_{n=1}^{\infty} \frac{f(T^{n-1}x)}{r^n}$$

como sua expansão na base-r. Da mesma forma que no caso $r = 2$, T é ergódica. Uma aplicação do teorema ergódico mostra que quase todo número é normal na base r (contém todos os dígitos na mesma proporção). Se $X_n(x)$ é $f(T^{n-1}x)$, o n-ésimo dígito na expansão (base-r) de x, então a seqüência $\{X_1, X_2, \cdots\}$ de variáveis randômicas é independente com $\lambda\{X_n = i\} = 1/r, i = 0, 1, \cdots, r-1$. Podemos chamar T a *transformação r-ádica.*

(b) Seja $\Omega, \mathcal{A}$, e T como no exemplo precedente, mas seja μ qualquer medida preservada por T. Um r-ádico (diádico, se $r = 2$) intervalo tem a forma $[k/r^n, (k+1)/r^n)$; êle contém um ponto ω se e sòmente se os primeiros n dígitos $x_1(\omega), \cdots, x_n(\omega)$(na escala r)têm certos valôres especificados. Visto que uniões finitas disjuntas de intervalos diádicos constituem uma álgebra que gera $\mathcal{A}$, T preserva λ se e sòmente se preserva a medida de cada intervalo diádico, ou, equivalentemente, se e sòmente se $\{x_1, x_2, \cdots\}$ é um processo estacionário sob λ.

O significado deste exemplo, é que podemos converter fatos relativos à teoria ergódica e teoria de probabilidade sobre fatos relativos a intervalos unitários e vice-versa.

O Próximo teorema dá outra formulação da definição de ergodicidade.

Teorema 1.11. Seja $(X, \mathcal{B}, \mu)$ um espaço de probabilidade e seja $T: X \to X$ uma transformação que preseva medida. Então T é ergódica sse $\forall A, B \in \mathcal{B}$

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

Prova. Suponha que T é ergódica. Colocando $f = \chi_A$ no teorema ergódico temos

$$\frac{1}{n} \sum_{i=0}^{n-1} \chi_A(T^i x) \to \mu(A) \quad q.t.p.$$

Proposição 2.7. $\mu \in M_T(X, \mathcal{A})$ é ergódica se e sò se μ é ponto extremal de $M_T(X, \mathcal{A})$.

Prova: [2]

1.2.3 -*Medidas Invariantes Para Transformações Contínuas*

Neste parágrafo X deve denotar um espaço metrizável compacto e d deve denotar uma métrica sobre X. A σ-álgebra dos subconjuntos de Borel de X deve ser denotada por $\mathcal{B}(X)$. Deste modo $\mathcal{B}(X)$ é a menor σ-álgebra contendo todos os subconjuntos abertos de X e a menor σ-álgebra contendo todos os subconjuntos fechados de X. Devemos denotar por $M(X)$ a coleção de todas as medidas de probabilidade definidas sobre o espaço mensurável $(X, \mathcal{B}(X))$. Devemos chamar os membros de $M(X)$ de medidas de probabilidade de Borel sobre X. Cada $x \in X$ determina um membro δ_x de $M(X)$ definido por

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

Deste modo a aplicação $x \mapsto \delta_x$ imbute X dentro de $M(X)$. Note que $M(X)$ é um conjunto convexo onde $p\nu + (1-p)\mu$ está definida por

$$(p\nu + (1-p)\mu)(A) = p\nu(A) + (1-p)\mu(A) \qquad \text{se } p \in [0,1]$$

Nosso escopo neste parágrafo é estudar medidas invariantes para transformações contínuas $T: X \to X$. A seguir colecionamos alguns fatos conhecidos sobre $M(X)$.

O próximo resultado diz que cada $\nu \in M(X)$ é determinada como a integral de funções contínuas.

Teorema 1.12. Sejam ν, μ duas medidas de probabilidade de Borel sobre o espaço métrico X. Se

$$\int_X f\, d\nu = \int_X f\, d\mu \quad \forall f \in C(X)$$

então $\nu = \mu$.

Prova: [5]

obs: $C(X)$ =espaço de todas as funções contínuas sobre X, a valores complexos.

Definição: Um *funcional linear positivo* sobre $C(X)$ é um (não necessàriamente contínuo a priori) funcional linear L com $L(f) \geq 0$ para todo $f \geq 0$ pontualmente.

Definição: Sejam V e W espaços vetoriais, H um subconjunto convexo de V. Uma aplicação $T: H \to W$ é chamada uma *aplicação linear afim* sobre H se

$$T(tx + (1-t)y) = tTx + (1-t)Ty$$

para todo $x, y \in H$ e todo $0 \le t \le 1$.

O próximo teorema relaciona elementos de $M(X)$ com funcionais lineares sobre $C(X)$.

Teorema 1.13.(Teorema da Representação de Riesz). Seja X um espaço métrico compacto e $J: C(X) \to \mathbb{C}$ uma aplicação linear contínua de modo que J é um operador positivo. Então existe $\mu \in M(X)$ de maneira que

$$J(f) = \int_X f\, d\mu \qquad \forall f \in C(X).$$

Prova: [5]

Sendo assim a aplicação $\mu \to J$ é uma bijeção entre $M(X)$ e a coleção de todos os funcionais lineares positivos e normalizados sobre $C(X)$; a injetividade segue do teorema 1.12 e a sobrejetividade do teorema 1.13. Devemos denotar a imagem de μ sob esta aplicação por J_μ. Claramente esta bijeção é uma aplicação afim, isto é $J_{p\mu+(1-p)\nu} = pJ_\mu + (1-p)J_\nu$, $p \in [0,1]$; $\nu, \mu \in M(X)$. Sendo assim $M(X)$ pode ser identificado com um subconjunto convexo da bola unitária em $C(X)^*$. Isto permite darmos uma topologia sobre $M(X)$ devida a topologia $fraco^*$ sobre $C(X)^*$.

Definição. A topologia $fraco^*$ sobre $M(X)$ é a menor topologia fazendo cada aplicação $\mu \mapsto \int_X f\, d\mu \quad (f \in C(X))$ contínua. Uma base é dada pela coleção de todos os conjuntos da forma

$$V_\mu(f_1, \cdots, f_k; \varepsilon) = \{\nu \in M(X) : |\int f_i\, d\nu - \int f_i\, d\mu| < \varepsilon, 1 \le i \le k\}$$

onde $\mu \in M(X), k \ge 1, f_i \in C(X)$ e $\varepsilon > 0$.

Claramente esta topologia sobre $M(X)$ é independente de qualquer métrica escolhida sobre X.

Teorema 1.14. Se X é um espaço metrizável então o espaço $M(X)$ é metrizável na topologia $fraco^*$. Se $\{f_n\}_{n=1}^{\infty}$ é um subconjunto denso de $C(X)$ então

$$D(\nu, \mu) = \sum_{n=1}^{\infty} \frac{|\int f_n\, d\nu - \int f_n\, d\mu|}{2^n \|f_n\|}$$

é uma métrica sobre $M(X)$ dando a topologia $fraco^*$.

Prova: [5]

O seguinte importante resultado é uma fácil conseqüência da compacticidade da bola unitária de $C(X)^*$ na topologia $fraco^*$.

Teorema 1.15. Se X é um espaço metrizável compacto então $M(X)$ é compacto na topologia $fraco^*$.

Seja $T: X \to X$ uma transformação contínua do espaço metrizável e compacto X. Devemos mostrar nesta secção que sempre existe alguma $\mu \in M(X)$ para a qual T é uma transformação que preserva medida de $(X, \mathcal{B}(X), \mu)$.

Inicialmente note que $T^{-1}\mathcal{B}(X) \subset \mathcal{B}(X)$(i.e., T é mensurável) porque $\{E \in \mathcal{B}(X) : T^{-1}E \in \mathcal{B}(X)\}$ é uma σ-álgebra e contém todos os conjuntos abertos. Por essa razão temos uma aplicação

$$\tilde{T}: M(X) \to M(X), \text{ dada por } \qquad (\tilde{T}\mu)(B) = \mu(T^{-1}B).$$

Ocasionalmente escrevemos $\mu \circ T^{-1}$ ao invés de $\tilde{T}\mu$. Devemos necessitar do seguinte lema

Lema.

$$\int f\, d(\tilde{T}\mu) = \int f \circ T\, d\mu \qquad \forall f \in C(X).$$

Prova. É suficiente tratar com $f \in C(X)$ a valores-reais. Por definição de $\tilde{T}$ temos

$$\int \chi_B\, d(\tilde{T}\mu) = \int \chi_B \circ T\, d\mu \qquad \forall B \in \mathcal{B}(X).$$

Sendo assim $\int h\, d(\tilde{T}\mu) = \int h \circ T\, d\mu$ se h é uma função simples. A mesma fórmula se verifica quando h é uma função mensurável não-negativa, pela escolha de uma seqüência crescente de funções simples convergindo pontualmente para h. Por essa razão a fórmula se verifica para qualquer $f: X \to \mathbb{R}$ contínua por consideração das partes positiva e negativa de f.

Teorema 1.16. A aplicação $\tilde{T}: M(X) \to M(X)$ é contínua e afim.

Prova. Se $f \in C(X)$ então $\int f\, d\tilde{T}\mu = \int f \circ T\, d\mu$. Sendo assim se $\mu_n \to \mu$ em $M(X)$ então

$$\int f\, d\tilde{T}\mu_n = \int f \circ T\, d\mu_n \to \int f \circ T\, d\mu = \int f\, d\tilde{T}\mu \Rightarrow \tilde{T}\mu_n \to \tilde{T}\mu.$$

Isto prova que $\tilde{T}$ é contínua.

Se $\nu, \mu \in M(X)$ e $p \in [0,1]$ então

$$\tilde{T}(p\nu+(1-p)\mu)(B) = p\nu(T^{-1}B)+(1-p)\mu(T^{-1}B) = (p\tilde{T}\nu+(1-p)\tilde{T}\mu)(B) \quad \forall B \in \mathcal{B}(X).$$

Isto mostra que $\tilde{T}$ é afim. □

Estamos interessados em todos os membros de $M(X)$ que são medidas invariantes por T.

Seja

$$M(X,T) = \{\mu \in M(X) : \tilde{T}\mu = \mu\}.$$

Este conjunto consiste de todas as $\mu \in M(X)$ para as quais T é uma transformação de $(X, \mathcal{B}(X), \mu)$ que preserva medida.

Teorema 1.17. Se $T: X \to X$ é contínua e $\mu \in M(X)$ então $\mu \in M(X,T)$ se e sòmente se $\int f \circ T\, d\mu = \int f\, d\mu \quad \forall f \in C(X)$.

Prova. Isto é uma conseqüência imediata do último lema e teorema 1.12.

Visto que $\tilde{T}: M(X) \to M(X)$ é uma aplicação contínua e afim de um subconjunto compacto e convexo de $C(X)^*$ podemos usar o teorema de Markov-Kakutani para mostrar que $\tilde{T}$ tem um ponto fixo. Contudo devemos mostrar diretamente que $M(X,T)$ é não-vazio. O seguinte teorema fornece um método de construção de membros de $M(X,T)$.

Teorema 1.18. Seja $T: X \to X$ contínua. Se $\{\sigma_n\}_{n=1}^{\infty}$ é uma seqüência em $M(X)$ e formamos uma nova seqüência $\{\mu_n\}_{n=1}^{\infty}$ por

$$\mu_n = \frac{1}{n}\sum_{i=0}^{n-1} \tilde{T}^i \sigma_n$$

então qualquer ponto limite μ de $\{\mu_n\}$ é um membro de $M(X,T)$.(tais pontos limites existem pela compacticidade de $M(X)$.)

Prova. Seja $\mu_{n_j} \to \mu$ em $M(X)$. Seja $f \in C(X)$. Então

$$\left|\int f \circ T\, d\mu - \int f\, d\mu\right| = \lim_{j\to\infty}\left|\int f \circ T\, d\mu_{n_j} - \int f\, d\mu_{n_j}\right|$$

$$= \lim_{j\to\infty}\left|\frac{1}{n_j}\int \sum_{i=0}^{n_j-1} (f \circ T^{i+1} - f \circ T^i)\, d\sigma_{n_j}\right|$$

$$= \lim_{j\to\infty}\left|\frac{1}{n_j}\int (f\circ T^{n_j} - f)\, d\sigma_{n_j}\right|$$
$$\leq \lim_{j\to\infty}\frac{2\|f\|}{n_j} = 0.$$

Sendo assim $\mu \in M(X,T)$. □

Corolário.(Krylov e Bogolioubov). Se $T: X \to X$ é uma transformação contínua de um espaço métrico compacto X então $M(X,T) \neq \emptyset$.

Prova. Podemos fazer qualquer escolha para σ_n no Teorema 1.18; em particular escolha $y \in X$ e coloque $\sigma_n = \delta_y$ para cada n. □

Temos as seguintes propriedades de $M(X,T)$.

Teorema 1.19. Se T é uma transformação contínua do espaço métrico compacto X então

(i) $M(X,T)$ é um subconjunto compacto de $M(X)$.

(ii) $M(X,T)$ é convexo.

(iii) μ é um ponto extremo de $M(X,T)$ sse T é uma transformação que preserva medida e ergódica de $(X, \mathcal{B}, \mu)$.

(iv) Se $\mu, \nu \in M(X,T)$ são ambas ergódicas e $\nu \neq \mu$ então elas são mutuamente singulares.

Prova: [5]

Obs: Decorre da demonstração do teorema acima que se $\mu_1, \mu \in M(X,T), \mu_1 \ll \mu$ e μ é ergódica então $\mu_1 = \mu$.

Capítulo 2

g-Medidas

Neste capítulo apresentamos as propriedades fundamentais das g-medidas que constituem a ferramenta básica na prova do Teorema Principal (cap.3). Para uma abordagem mais completa e detalhada ver [10], [11] e [12]. Visamos sua definição nos casos do shift unilateral ou das aplicações expansoras do círculo S^1 (aplicações de grau $r > 1$).

Seja (X, d) um espaço métrico compacto.

Definição. Uma aplicação $T : X \to X$ é dita uma *transformação de cobertura* (minimal) de X se existem um inteiro $n \geq 2$ e $\rho > 1$ real tal que

(i) T é n-para-1 em todo o espaço X,

(ii) T é um homeomorfismo local,

(iii) para $\delta > 0$ suficientemente pequeno, $d(x, y) = \delta$ implica $d(Tx, Ty) \geq \rho\delta$, e,

(iv) para todo $\varepsilon > 0$, existe N tal que se $n \geq N$ e $x \in X$, $T^{-n}(\{x\})$ é ε-denso em X.

Exemplos:

(1) $T: S^1 \to S^1$, dada por $Tz = z^n$. $d(x, y) = |x - y|$ (distância euclidiana). $X = S^1$ é compacto e T é uma aplicação de cobertura.

(2) Se $\Sigma_r = \{0, 1, \cdots, r-1\}^{\mathbb{Z}^+}$ e $T: \Sigma_r \to \Sigma_r$, é dada por $Tx = rx$ (mod 1). $d(x, y) = |x - y|$ (distância euclidiana). $X = \Sigma_r$ é compacto e T é de cobertura.

(3) $X = \Sigma_r$ e T é a aplicação *shift* sôbre X, com a métrica

$$d(x,y) = \frac{1}{\inf\{i : x_i \neq y_i\} + 1}$$

X é compacto e T é uma aplicação de cobertura.

(4) Seja M variedade compacta sem bordo e $T: M \to M$ uma aplicação C^1. Então é expansora (métrico) se e somente se T é expansora no seguinte sentido: existe $\gamma > 1$ tal que

$$|T'(x)v| \geq \gamma |v| \qquad \forall x \in M, v \in T_x M$$

Sejam

$$\mathcal{C}(X) = \{f : X \to \mathbb{R} \ \text{, contínua}\}$$

$$\|f\| = \sup_{x \in X} |f(x)|$$

$\mathcal{C}^*(X)$ = medidas boreleanas finitas com sinal.

$\mathcal{P}(X)$ = medidas de probabilidades de $\mathcal{C}^*(X)$. Como já vimos no capítulo 1, $\mathcal{P}(X)$ é um subconjunto compacto e convexo de $\mathcal{C}^*(X)$. O mesmo vale para

$$\mathcal{P}_T(X) = \{\mu \in \mathcal{P}(X) : \mu \text{ é invariante por } T\}$$

Seja $\mu \in \mathcal{C}^*(X)$, denotamos por $Q\mu$ a medida definida via $Q\mu(A) = \mu(T(A))$ para A tal que $T\big|_A : A \to X$ é injetiva.

Seja $P_1, P_2, \cdots, P_n$ uma partição de X tal que

$$P_i \bigcap P_j = \emptyset \ (i \neq j) \ , \bigcup_{i=1}^{n} P_i = X$$

e de modo que

$$T\big|_{P_i} : P_i \to X \quad \text{é injetiva}$$

X

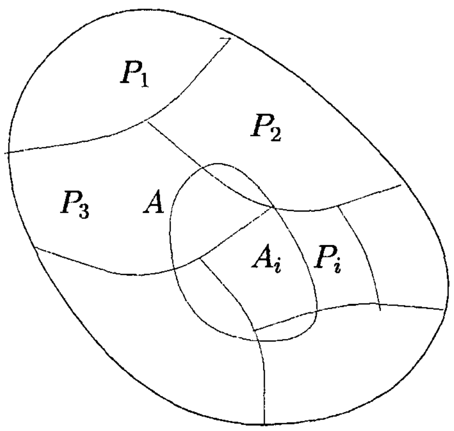

Figura 2.1: Partição de X

Então se $\mu \in \mathcal{P}_T(X) \Rightarrow \mu \ll Q\mu$. De fato, seja $A \in \beta(\mathrm{X})$ tal que $Q\mu(A) = 0$.

$$Q\mu(A) = \sum_{i=1}^{n} \mu(TA_i) = 0$$

onde $(A_i = A \cap P_i)$, logo $\mu(TA_i) = 0\,, i = 1, \cdots, n$. Mas

$$\mu(TA_i) = \mu(T^{-1}TA_i) = 0$$

e

$$A_i \subset T^{-1}(TA_i)$$

implicando que $\mu(A_i) = 0$, $\forall i \leq n$, isto é $\mu(A) = 0$. Segue que $\mu \ll Q\mu$, e portanto, pelo teorema de Radon-Nikodym existe $g \in \mathcal{L}^1(Q\mu)$ tal que $g: X \to \mathbb{R}^+$, e

$$\mu(A) = \int_A g dQ\mu \qquad \forall A \in \beta(X)$$

Lema: As seguintes propriedades valem para g:

(a) $0 \leq g(x) \leq 1 \qquad q.t.p. \in X$

(b) $\sum_{y \in T^{-1}(\{x\})} g(y) = 1 \qquad q.t.p. \in X$

Prova:

(a) Seja $A \in \beta(X)$ tal que $\forall x \in A \quad g(x) > 1$. Suponhamos por absurdo que $\mu(A) > 0$. Então, sem perda de generalidade, podemos supor que $A \subset P_i$ para algum $1 \leq i \leq n$.

$$\mu(A) = \int_A g dQ\mu > Q\mu(A) = \mu(TA) = \mu(T^{-1}TA) \geq \mu(A) \quad \text{i.e.} \quad \mu(A) > \mu(A) \quad \rightarrow\leftarrow .$$

Segue que

$$\mu(\{x: g(x) > 1\}) = 0$$

(b) Seja $\varepsilon > 0$ tal que para todo $x \in X$,

$$T\Big|_{B_\varepsilon(x)}: B_\varepsilon(x) \to T(B_\varepsilon(x))$$

é um homeomorfismo. Seja $\mathcal{P}$ uma partição de X tal que os diâmetros dos átomos de $\mathcal{P}$ sejam $< \varepsilon$. Seja

$$\mathcal{P}^k = T^{-k}\mathcal{P} \bigvee \mathcal{P}^{k-1} \bigvee \cdots \bigvee \mathcal{P}.$$

Então,

$$\mathcal{P} \leq \mathcal{P}^1 \leq \mathcal{P}^2 \leq \cdots$$

e se $\varepsilon > 0$ é suficientemente pequeno, temos

$$\bigvee \mathcal{P}^k = \mathcal{B}(X) \quad \mu(mod(0))$$

Seja $P_k(x)$ o átomo de $\mathcal{P}^k$ que contém x. Então pelo teorema de derivação de Lebesgue vale

$$(*) \qquad \lim_{k\to\infty} \frac{1}{Q\mu(P_k(x))} \int_{P_k(x)} g\, dQ\mu = g(x) \quad q.t.p.$$

Seja

$$X_0 = \{x \in X : (*) \text{ é válida em cada } x_i \in T^{-1}(\{x\})\}$$

Dado $x \in X_0$ sejam $\{x_1, \cdots, x_n\} = T^{-1}(\{x\})$. Então $\forall i$, temos

$$\lim_{k\to\infty} \sum_{i=1}^{n} \frac{1}{Q\mu(P_{k+1}(x_i))} \int_{P_{k+1}(x_i)} g\, dQ\mu = \sum_{i=1}^{n} g(x_i)$$

Mas

$$Q\mu(P_{k+1}(x_i)) = \mu(T(P_{k+1}(x_i))) = \mu(T^{-1}T(P_{k+1}(x_i))) =$$
$$= \mu(\bigcup_{i=1}^{n} P_{k+1}(x_i))$$

Por outro lado,

$$\sum_{i=1}^{n} \frac{1}{Q\mu(P_{k+1}(x_i))} \int_{P_{k+1}(x_i)} g\, dQ\mu$$
$$= \sum_{i=1}^{n} \frac{1}{\mu(\cup_{i=1}^{n} P_{k+1}(x_i))} \int_{P_{k+1}(x_i)} g\, dQ\mu$$
$$= \frac{1}{\mu(\cup_{i=1}^{n} P_{k+1}(x_i))} \sum_{i=1}^{n} \int_{P_{k+1}(x_i)} g\, dQ\mu$$
$$= \frac{1}{\mu(\cup_{i=1}^{n} P_{k+1}(x_i))} \sum_{i=1}^{n} \mu(P_{k+1}(x_i)) = 1$$

pois se ε é pequeno $\big(P_{k+1}(x_i)\big)_i$ é uma família disjunta.

Seja pois,

$$\mathcal{G} = \{g: X \to [0,1] : g \text{ mensurável, e } \sum_{y \in T^{-1}(\{x\})} g(y) = 1, \quad \forall x \in X\}$$

Definição: Uma medida de probabilidade μ sobre os boreleanos de X é uma g-medida para uma dada $g \in \mathcal{G}$ se

$$\frac{d\mu}{dQ\mu} = g \qquad (mod\, \mu)$$

Antes de provarmos a existência de g-medidas vamos introduzir uma noção que nos permitirá entender mais claramente o significado das mesmas.

Definição: Seja $f: X \hookleftarrow$ localmente injetiva e $\mu \in \mathcal{C}^*(X)$. Dizemos que $F \in \mathcal{L}^1(\mu)$ é um jacobiano de f, relativamente a μ, se

$$\mu(f(A)) = \int_A F\, d\mu$$

para todo boreleano A tal que $f\big|_A$ é injetiva. O jacobiano se existir é único (q.t.p.) e denota-se por $J_\mu f$.

Voltemos ao nosso caso. Suponhamos $g > 0$. Então seja $J_\mu T$ o jacobiano de T, relativamente a μ, isto é,

$$\mu(TA) = \int_A J_\mu T\, d\mu$$

para todo A tal que $T\big|_A$ é injetiva. Se μ é uma g-medida então

$$Q\mu(A) = \int_A 1\, dQ\mu = \mu(TA) = \int_A J_\mu T\, d\mu = \int_A J_\mu T . g\, dQ\mu$$

isto é,

$$\int_A 1\, dQ_\mu = \int_A J_\mu T.g\, dQ_\mu$$

para todo $A \in \beta(X)$ tal que $T\Big|_A$ é injetiva. Segue que

$$1 = J_\mu T.g \qquad (q.t.p.)$$

isto é,

$$J_\mu T = \frac{1}{g(x)} \qquad \mu\,(q.t.p.)$$

Resumindo: Encontrar uma g-medida é encontrar uma medida tal que o jacobiano de T, relativamente a esta medida, seja $\frac{1}{g}$ em q.t.p.

Dada $g \in \mathcal{G}$, seja o operador

$$L_g^*: \mathcal{C}^*(X) \to \mathcal{C}^*(X)$$

dado por

$$(L_g^*)(\mu)(A) = \int_A g\, dQ\mu$$

isto é $(L_g^*)(\mu)$ é a medida que satisfaz

$$\frac{dL_g^*(\mu)}{dQ\mu} = g$$

Afirmação: Se $\mu \in \mathcal{P}(X)$ então $L_g^*(\mu) \in \mathcal{P}(X)$

Prova: Seja a família de abertos $\{B_r(x) : x \in X\}$. Como X é compacto existem $x_1, \cdots, x_s \in X$ tais que

$$\bigcup_{i=1}^{s} B_r(x_i) = X$$

Tomemos

$$B_1 = B_r(x_1)$$

e

$$B_k = B_r(x_k) \setminus \bigcup_{j=1}^{k-1} B_r(x_j)$$

Então

$$\bigcup_{k=1}^{s} B_k = X, \quad B_j \cap B_i = \emptyset \quad \text{se} \quad i \neq j$$

Além disso dado $1 \leq i \leq s$ existem

$$\varphi_k^i: B_i \to \varphi_k^i(B_i)$$

ramos da inversa de T. Sejam

$$B_k^j = \varphi_k^j(B_j)$$

então dado $W \subset B_k^j$ temos que

$$\int\limits_{B_k^j} \chi_W \, dQ\mu = Q\mu(W) = \mu(TW) = \int_{B_k} \chi_{TW} \, d\mu = \int_{B_k} \chi_W \circ \varphi_k^j \, d\mu$$

$$(\chi_W \circ \varphi_k^j)(x) = \chi_W(\varphi_k^j(x)) = \begin{cases} 1, \text{se} & \varphi_k^j(x) \in W; \\ 0, \text{se} & \varphi_k^j(x) \notin W \end{cases}$$

isto é,

$$\chi_W \circ \varphi_k^j(x) = \begin{cases} 1, \text{se} & \varphi_k^j(x) \in W; \\ 0, \text{se} & \varphi_k^j(x) \notin W \end{cases}$$

Segue que se $h \in \mathcal{C}(B_k^j)$

$$\int_{B_k^j} h \, dQ\mu = \int_{B_k} h \circ \varphi_k^j \, d\mu$$

Logo

$$\begin{aligned} \sum_j \int_{B_k^j} g \, dQ\mu &= \sum_j \int_{B_k} g \circ \varphi_k^j \, d\mu \\ &= \int_{B_k} \sum_j g \circ \varphi_k^j \, d\mu = \int_{B_k} 1 \, d\mu \end{aligned}$$

Pois $\sum_{y \in T^{-1}(\{x\})} g(y) = 1$.

Finalmente,

$$\begin{aligned} \int_X 1 \, dL_g^*(\mu) &= \sum_{k,j} \int_{B_k^j} 1 \, dL_g^*(\mu) \\ &= \sum_{k,j} \int_{B_k^j} g \, dQ\mu = \sum_k \int_{B_k} 1 \, d\mu \\ &= \sum_{k=1}^{s} \mu(B_k) = \mu(x) = 1. \end{aligned}$$

Teorema 2.1. μ é T-invariante se e somente se existe $g \in \mathcal{G}$ tal que μ é uma g-medida. Se $g \in \mathcal{G}(X)$ então existe uma g-medida.

Prova: A primeira afirmação é imediata. A segunda seguirá de L_g^* ser um operador afim e contínuo sobre $\mathcal{P}$(X) que é um compacto convexo.

(i) L_g^* é contínuo. Seja $\mu_n \to \mu$ então

$$dL_{\mu_n}^* = g \, dQ\mu_n$$

e

$$dL^*_\mu = g\,dQ\mu$$

logo para provarmos que se f é contínua

$$\int_X f dL^*_{\mu_n} \to \int_X f dL^*_\mu$$

basta provarmos que

$$\int_X fg\,dQ\mu_n \to \int_X fg\,dQ\mu$$

o que seguirá de provarmos que para todo $\Phi \in \mathcal{C}(X)$ vale

$$\int_X \Phi\,dQ\mu_n \to \int_X \Phi\,dQ\mu$$

$L^*_g : \mathcal{P}(X) \hookleftarrow$ é contínuo. Seja $\mu_n \to \mu$, então como

$$dL^*_g(\mu_n) = g\,dQ\mu_n$$

e

$$dL^*_g(\mu) = g\,dQ\mu$$

precisamos provar que para toda ϕ contínua

$$\int_X \phi g\,dQ\mu_n \to \int_X \phi g\,dQ\mu$$

para o qual é suficiente provar que $dQ\mu_n \to dQ\mu$. Seja então A com $dQ\mu(\partial A) = 0$ e tal que $T\Big|_A$ seja injetiva. Então

$$\mu(\partial A) = 0$$

pois $\mu \ll Q\mu$. Então

$$Q\mu_n(A) = \mu_n(TA) \to \mu(TA) = Q\mu(A)$$

observe que como $\mu(\partial A) = 0$ e μ é T-invariante, $\mu(\partial TA) = 0$.

(ii) Existência de ponto fixo. Seja $\mu \in \mathcal{P}(X)$ e considere a sequência

$$\mu_n = \frac{1}{n+1}\sum_{k=0}^{n}(L^*_g)^k(\mu)$$

então como L_g^* é afim, é fácil ver que

$$L_g^*(\mu_n) = \frac{1}{n+1}\sum_{k=0}^{n}(L_g^*)^{k+1}(\mu)$$

Como $\mathcal{P}(X)$ é um compacto convexo

$$(L_g^*)(\mu_n) \in \mathcal{P}(X)$$

e existe $n_j \to +\infty$ tal que

$$\mu_{n_j} \to \nu \in \mathcal{P}(X)$$

Por continuidade

$$(L_g^*)(\mu_{n_j}) \to (L_g^*)(\nu)$$

Por outro lado

$$(L_g^*)(\mu_{n_j}) = \frac{1}{n_j+1}(L_g^*(\mu) + \cdots + (L_g^*)^{n_j+1}(\mu))$$

$$= \frac{1}{n_j+1}(\mu + (L_g^*)(\mu) + \cdots + (L_g^*)^{n_j}(\mu)) + \frac{1}{n_j+1}[(L_g^*)^{n_j+1}(\mu) - \mu]$$

$$= \mu_{n_j} + \frac{1}{n_j+1}[(L_g^*)^{n_j+1}(\mu) - \mu] \to \nu$$

Segue que

$$(L_g^*)(\mu_{n_j}) \to \nu$$

isto é,

$$(L_g^*)(\nu) = \nu$$

ν é uma g-medida.

Exemplos:

1. Shift de Bernoulli. Seja $X = \prod_0^\infty\{0, 1, \cdots, n-1\}$ e seja T a transformção shift sobre X. Sob a métrica definida por

$$d(\theta, \hat{\theta}) = (\inf\{i : \theta(i) \neq \hat{\theta}(i)\} + 1)^{-1}$$

X é compacto e T é uma transformação de cobertura sobre X.

Se $p = (p_0, p_1, \cdots, p_{n-1})$ satisfaz $p_k > 0$ e $\sum p_k = 1$, seja μ^p a medida produto sôbre X com distribuição p em cada componente. Um cálculo fácil mostra que μ^p é uma g^p-medida com

$$g^p(\theta) = p_k \qquad (\theta \in X, \theta_0 = k).$$

Visto que g^p é uma função contínua, positiva e localmente constante sôbre X.

2. Medidas de Markov. Seja (X, T), tal que T é n-para-1, e seja $P = (p_{ij})$ um núcleo de Markov sobre $\{0, 1, \cdots, n-1\}$. Escolha um vetor de probabilidade $\pi = (\pi_0, \cdots, \pi_{n-1})$ com $\pi P = \pi$ e denote por m a medida de Markov sobre X gerada pela distribuição inicial π e transição de probabilidade P.

   Se $a_0 a_1 \cdots a_k$ é uma seqüência de estados, seja

$$[a_0 a_1 \cdots a_k] = \{\theta \in X : \theta_i = a_i \quad \text{para} \quad 0 \leq i \leq k\}$$

   Agora, $m \in \mathcal{P}_T(X)$ porque $\pi P = \pi$, isto é m é uma g-medida para alguma $g \in \mathcal{G}$. Podemos calcular g como segue. Fixando os estados i e j, a razão

$$q_{ji} = \frac{m([ija_2 \cdots a_k])}{m([ja_2 \cdots a_k])} = \frac{\pi_i p_{ij}}{\pi_j}$$

   é independente de $a_2, \cdots, a_k$. Sendo assim se

$$g(\theta) = \frac{\pi_{\theta_0} p_{\theta_0 \theta_1}}{\pi_{\theta_1}} \qquad (\theta \in X),$$

   m é uma g-medida.

3. Seja $X = \mathbb{R}/\mathbb{Z}$ e $Tx = nx (\mathrm{mod}\ 1)$ para algum $n \geq 2$. Para algum α com $|\alpha| \leq 1$, seja

$$g(x) = \frac{1 + \alpha \cos 2\pi x}{n} \qquad (x \in X)$$

   então obviamente $0 \leq g(x) \leq 1$, e

$$\sum_{y \in T^{-1}(\{x\})} g(y) = 1 + \frac{\alpha}{n} \sum_{j=0}^{n-1} \cos 2\pi \frac{y+j}{n} = 1.$$

   Por essa razão $g \in \mathcal{G}(X) \cap \mathcal{C}^1(X)$, logo existe uma g-medida associada a esta g.

4. Seja $X = \mathbb{R}/\mathbb{Z}$ e $Tx = 2x(\text{mod } 1)$. Se $g \in \mathcal{G}$ com $g(0) = g(\frac{1}{3}) = g(\frac{2}{3}) = 1$, então a massa pontual ε_0 em zero e a medida $\frac{1}{2}(\varepsilon_{\frac{1}{3}} + \varepsilon_{\frac{2}{3}})$ são ambas g-medidas.

5. Seja $X = \mathbb{R}/\mathbb{Z}$ e $Tx = 2x(\text{mod } 1)$. Seja

$$g(x) = \begin{cases} \frac{1+\cos 2\pi x}{2}, & (x \neq 0, \frac{1}{2}) \\ \frac{1}{2}, & (x = 0 \quad ou \quad \frac{1}{2}) \end{cases}$$

então, $g \in \mathcal{G}$ e para $f \in \mathcal{C}(X)$, $(L_g^*)^n f$ converge para $f(0)$, mas o leitor pode verificar fácilmente que ϵ_0 não é uma g-medida devido a descontinuidade de g em 0.

Lema: Se $A \in \beta(X)$ satisfaz $\mu_g(A) = 0$ então $\mu_g(T^n(A)) = 0$ para todo n.

Prova: (indução)

$\mu_g(T(A)) = 0$.

Seja $P_1, P_2, \cdots, P_k$ uma partição de X, isto é tal que

$$P_i \cap P_j = \emptyset, \qquad i \neq j$$

$$\bigcup_1^k P_i = X$$

e de modo que $T\big|_{P_i} : P_i \to X$ é injetiva.

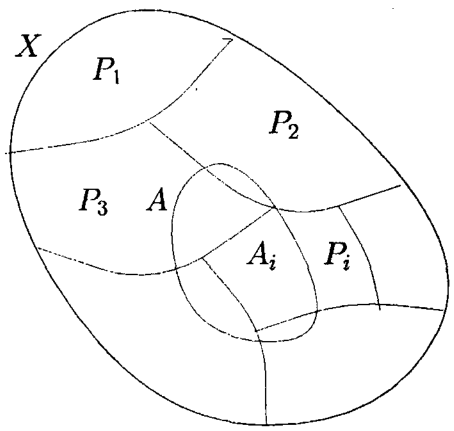

Figura 2.2: Partição de X

Seja $A = \bigcup_{i=1}^k A_i$, onde $A_i = A \cap P_i$, então

$$\mu_g(T(A_i)) = \int_{A_i} \frac{1}{g}\, d\mu \leq \int_A \frac{1}{g}\, d\mu = 0$$

pois $\mu_g(A) = 0$.

$$T(A) \subset \bigcup_{i=1}^{k} T(A_i) \Rightarrow \mu_g(T(A)) \leq \sum_i \mu_g(T(A_i)) = 0.$$

Observe que isto implica $\mu_g(T^n(A)) = 0, \forall n \geq 0$. Senão vejamos:

(i) $\mu_g(A) = 0$

(ii) $\mu_g(T^n(A)) = 0$ (hipótese)

tomando $B = T^n(A)$, temos $\mu_g(T(B)) = 0 \Rightarrow \mu_g(T^{n+1}(A)) = 0$.

Proposição: Se $g > 0$ então $\mu_g(A) > 0$ sobre abertos. Isto é se $g > 0$ então a g-medida associada é positiva sobre abertos.

Prova: Sem perda de generalidade, suponhamos $A = B_\varepsilon(x_0)$. Considere, por um momento, $X \neq \bigcup_{j\geq 0} T^j(A)$. Então se $x \notin \bigcup_{j\geq 0} T^j(A)$ existe $n \in \mathbb{N}$ tal que $T^{-n}(\{x\})$ é ε-denso em X. Logo existe $y \in T^{-n}(\{x\})$ tal que $d(y, x_0) < \varepsilon$.

Então

$$T^n(y) = x, \quad \text{e} \quad y \in B_\varepsilon(x_0)$$

isto é, $x \in T^n(A)$ $\rightarrow\leftarrow$, pois $X \neq \bigcup_{j\geq 0} T^j(A)$. Logo $\bigcup_{j\geq 0} T^j(A) = X$. Já vimos no lema anterior que $\mu_g(T^j(A)) = 0$, logo

$$\mu(X) = \mu(\bigcup_{j\geq 0} T^j(A)) \leq \sum_j \mu(T^j(A)) = 0 \qquad \rightarrow\leftarrow.$$

Lema: g-medidas com $g \in \mathcal{G}, g > 0$, são não-atômicas.

Prova: Suponha que exista $x \in X$ e $\delta > 0$ tal que

$$\mu(\{x\}) = \delta$$

Então se $\forall i, j \quad (i \neq j)$

$$T^{-i}(\{x\}) \cap T^{-j}(\{x\}) = \emptyset$$

temos que os conjuntos

$$A_n = T^{-n}(\{x\})$$

satisfazem

$$A_n \cap A_m = \emptyset \qquad \forall n \neq m$$

e portanto, da invariância de μ obtemos

$$\mu(A_n) = \delta \qquad , \forall n$$

e

$$\mu(\bigcup_{n=1}^{N} A_n) = N\delta$$

Tomando N tal que $N\delta > 1$ obtemos

$$1 = \mu(X) > \mu(\bigcup_{n=1}^{N} A_n) > 1 \qquad \rightarrow\leftarrow .$$

Segue que existem $i > j$ com

$$T^{-i}(\{x\}) \cap T^{-j}(\{x\}) \neq \emptyset$$

Seja $y \in T^{-i}(\{x\}) \cap T^{-j}(\{x\})$. Então

$$T^{j}(\{y\}) = x$$

$$T^{i}(\{y\}) = x$$

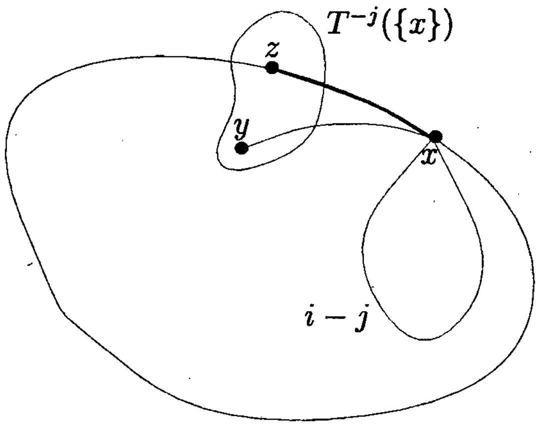

Figura 2.3: Órbitas

Seja $z = T^{i-j}(\{y\})$. Então

$$T^{j}(\{z\}) = T^{j}(T^{i-j}(\{y\})) = T^{i}(\{y\}) = x$$

isto é, $z \in T^{-j}(\{x\})$

Finalmente:

$$T^{i}(\{x\}) = T^{j}(T^{i-j}(\{x\})) = T^{j}(\{z\}) = x$$

isto é, $T^i(\{x\}) = x$ e x é periódico. Seja N o menor período de x, isto é

$$T^N(\{x\}) = x \quad \text{e} \quad T^j(\{x\}) \neq x, \quad 0 < j < N$$

Como

$$\mu(T(A)) = \int_A \frac{1}{g} d\mu$$

segue que

$$\mu(T^N(A)) = \int_A \prod_{j=0}^{N-1} \frac{1}{g(T^j(\{y\}))} d\mu(y)$$

Portanto se $G(\{y\}) = \prod_{j=0}^{N-1} g(T^j(\{y\}))$, $G \in \mathcal{G}$ e μ é uma G-medida para T^n. Segue que basta supor x_0 ponto fixo. Logo seja x_0 tal que

$$Tx_0 = x_0$$

e

$$\mu(\{x_0\}) = c > 0$$

Então,

$$\mu(T^{-1}(\{x_0\})) = \mu(\{x_0, x_2, \cdots, x_n\}) = c$$

$$= \mu(\{x_0\}) + \mu(\{x_2, \cdots, x_n\})$$

$$\Rightarrow \mu(T^{-n}(\{x_0\}) \setminus \{x_0\}) = 0.$$

$$\mu(\{x_0\}) = \mu(\{Tx_0\}) = \int_{\{x_0\}} \frac{1}{g} d\mu = \frac{1}{g(x_0)} c$$

Deste modo,

$$c = \mu(\{x_0\}) = (g(x_0))^{-1} c \quad \Rightarrow (g(x_0))^{-1} = 1, \text{ isto é}, g(x_0) = 1$$

Mas

$$\sum_{y \in T^{-1}(\{x_0\})} g(y) = 1$$

logo,

$$g(x_0) + \sum_{y \in T^{-1}(\{x_0\}) \setminus \{x_0\}} g(y) = 1 + \sum_{y \in T^{-1}(\{x_0\}) \setminus \{x_0\}} g(y)$$

logo,

$$g(y) = 0 \quad \forall y \in T^{-1}(\{x_0\}) \setminus \{x_0\} \quad \rightarrow\leftarrow, \text{ pois}, g > 0$$

.

Capítulo 3

Aplicações expansoras C^1 não ergódicas e g-medidas

Neste capítulo demonstraremos uma equivalência entre certas proposições que nos permitirá construir (no próximo capítulo) uma aplicação expansora C^1 do círculo que preserva a medida de Lebesgue mas para a qual a medida de Lebesgue não é ergódica. Ao mesmo tempo o exemplo a ser construído responde na negativa uma questão alhures levantada, sobre a unicidade de g-medidas.

Devemos considerar neste capítulo aplicações diferenciáveis do círculo. Uma tal aplicação T é chamada *expansora* se existe uma constante $c > 1$ tal que $|T'(x)| \geq c$ para todo $x \in S^1$. Uma aplicação expansora do círculo é uma aplicação de grau r para algum $|r| > 1$. Isto significa que a aplicação é uma $|r|$-cobertura do círculo por si mesmo (ver apêndice). Consideraremos tão sòmente aplicações que preservam orientação (i.e., aquelas com $r > 1$), e sempre devemos identificar o círculo com o intervalo $[0, 1)(\text{mod } 1)$. Devemos estar interessados na existência e no número de medidas invariantes e absolutamente contínuas (abreviação MIAC) para estas aplicações. Trabalharemos também com um espaço de símbolos. O *shift unilateral* sobre r símbolos está definido sobre o conjunto $\Sigma_r = \{0, 1, \cdots, r-1\}^{\mathbb{Z}^+}$, com a métrica

$$d(x,y) = \begin{cases} 2^{-n}, & \text{se } x \text{ e } y \text{ diferem pela primeira vez na posição n} \\ 0, & se \quad x = y \end{cases}$$

A aplicação shift (ver apêndice) sobre Σ_r será denotada por σ. Se $x \in \Sigma_r$ e $a \in \{0, 1, \cdots, r-1\}^k$, escremos ax para a seqüência em Σ_r definida por

$$(ax)_i = \begin{cases} a_i & se \quad i < k \\ x_{i-k} & se \quad i \geq k \end{cases}$$

Observemos que o produto ax "cola"a seqüência finita a na frente da seqüência infinita x.

No espaço métrico (Σ_r, d), de seqüências, a bola aberta de centro x e raio 2^{-n} (também conhecida em nosso contexto como n-cilindro relativo a x) é o conjunto

$$[x]^n = \{y \in \Sigma_r : d(x,y) < 2^{-n}\},$$

é fácil mostrar (ver apêndice) que este é o conjunto de todas as seqüências $y \in \Sigma_r$ cujos símbolos concordam com x da posição 0-ésima até n-ésima. Defina

$$Var_n f = \sup_{\{x,y:d(x,y)<2^{-n}\}} |f(x) - f(y)|.$$

Conforme vimos no capítulo 2 para descrever o conceito de uma g-medida devemos considerar um homeomorfismo local T ($r-$para-1), de um espaço métrico compacto X sobre si mesmo. Devemos considerar, como já enfatizamos, os casos onde T é uma aplicação expansora de grau r do círculo (mais precisamente, $T_r(x) = rx(\text{mod } 1)$), ou $T = \sigma$ é a aplicação shift sobre Σ_r.

Seja $\mathcal{G}(X)$ a classe de todas as funções contínuas $g: X \to (0,1)$ sobre X de maneira que para todo $x \in X, \Sigma_{y \in T^{-1}(\{x\})} g(y) = 1$. Escremos $|A|$ para o diâmetro do conjunto A.

Enunciaremos agora uma formulação equivalente de uma g-medida, a qual será usada em nosso contexto:

Uma g-medida ν sobre X é uma medida de probabilidade de Borel satisfazendo

$$\lim_{\substack{|A| \to 0 \\ x \in A}} \frac{\nu(A)}{\nu(TA)} = g(x) \tag{3.1}$$

para todo $x \in X$. Em particular, se $X = \Sigma_r$, temos

$$\lim_{\substack{|A| \to 0 \\ x \in A}} \frac{\nu([ix]^{n+1})}{\nu([x]^n)} = g(ix)$$

Provemos (3.1) da seguinte forma:

Temos $\frac{d\nu}{dQ\nu} = g(x)$

logo,

$$\frac{\nu(A)}{\nu(TA)} = \frac{\int_A g\, dQ\nu}{\nu(TA)} = \frac{1}{Q\nu(A)}\int_A g\, dQ\nu = \frac{1}{Q\nu(A)} g(\xi) Q\nu(A)$$

tomando na igualdade anterior limite quando o diâmetro de A tende a zero, temos,

$$\lim_{\substack{|A|\to 0 \\ x\in A}} \frac{\nu(A)}{\nu(TA)} = g(x)$$

uma vez que g é contínua.

Obs: Usamos na demonstração o teorema do valor médio para integrais e consideramos $T\Big|_A : A \to X$ injetiva.

- O conjunto das g-medidas para uma dada g é um subconjunto compacto afim do conjunto das medidas sobre X (na toplogia $fraco^*$), e os pontos extremos deste conjunto são precisamente g-medidas ergódicas. Outrossim, pelo teorema de Krein-Mil'man, o conjunto das g-medidas é o invólucro convexo fechado do conjunto das g-medidas ergódicas. Disto segue que se existe uma g com uma g-medida não-ergódica então existe ao menos duas g-medidas para esta g.

A proposição principal deste capítulo é o seguinte

Teorema 3.1.(*Teorema Principal*). Existe uma aplicação expansora C^1 que preserva a medida de Lebesgue, para a qual a medida de Lebesgue não é ergódica.

Antes de demonstrarmos o teorema, vejamos alguns preliminares necessários:

- Formulações Equivalentes Para o Teorema 3.1

Para estabelecer os resultados desta secão, necessitaremos de algumas definições. Se $X = \Sigma_r$, sejam $u = 000\cdots$ e $v = r-1, r-1, r-1, \cdots$. Definimos uma relação de equivalência sobre X gerada por

$$\begin{cases} u \sim v \\ aju \sim aiv \quad \text{para} \quad i < r-1, j = i+1 \end{cases}$$

onde a é qualquer palavra finita, isto é, $a \in \{0, \cdots, r-1\}^n$.

Note que as classes de equivalência para a relação acima consistem precisamente dos elementos de Σ_r, que descrevem o mesmo número (mod 1) quando considerados expandidos na base r.

Uma interpretação últil da relação acima é que se, por exemplo, $r = 10$ (sistema de numeração de base 10) cada número entre 0 e 1 (representados na base $r = 10$) tem uma expansão decimal infinita, e tal expressão deve ser única exceto pelo fato de que

$$.a_1a_2\cdots(a_n+1)000\cdots = .a_1a_2\cdots a_n999\cdots$$

Seja $\mathcal{G}(X)$ =classe de todas as funções contínuas $g: X \to (0,1)$ sobre X tal que para todo $x \in X, \Sigma_{y \in T^{-1}(\{x\})} g(y) = 1$. Onde estamos considerando T como a aplicação expansora de grau r do círculo($X = S^1$), $x \mapsto rx(\text{mod } 1)$, ou a plicação shift sobre $X = \Sigma_r$.

Definição(*função compatível*)

Se $g \in \mathcal{G}$, dizemos que g é *compatível* se $x \sim y$ implica $g(x) = g(y)$. Seja $\mathcal{G}^{comp}$ denotando a classe das funções $g \in \mathcal{G}$ que são compatíveis.

Antes de enunciarmos o próximo teorema vejamos algumas definições úteis
para a demonstração do mesmo:

(i) $h: X \to X$ é um isomorfismo mensurável se existe $A \subset X$ com $\lambda(A) = 0$ e tal que $h: X \setminus A \to X \setminus h(A)$ é uma bijeção que preserva medida.

(ii) Um diagrama de aplicações como

$$\begin{tikzcd} & A \arrow[dl, "f"'] \arrow[dr, "h"] & \\ B \arrow[rr, "g"'] & & C \end{tikzcd}$$

Figura 3.1: Diagrama comutativo

é chamado *comutativo* se $g \circ f = h$.

(iii) (semi-conjugação)

Sejam X e Y dois espaços topológicos e $f: X \to X$ e $g: Y \to Y$ duas aplicações contínuas. Uma aplicação $h: X \to Y$ contínua e sobrejetiva(não necessáriamente um-a-um) é dita uma *semi-conjugação* se $h \circ f = g \circ h$, isto é, se o seguinte diagrama comuta:

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[d, "h"'] & X \arrow[d, "h"] \\ Y \arrow[r, "g"] & Y \end{tikzcd}$$

Figura 3.2: Semi-conjugação

Se, além disso, h é um homeomorfismo de X sobre Y, então h é chamada uma *conjugação topológica*, e f e g são ditas *topológicamentes conjugadas.*

Note que $h \circ f = g \circ h$ implica $h \circ f^n = g^n \circ h$. De fato, $hf^2 = (hf)f = g(hf) = g^2h$, etc. A imagem de uma órbita de f por uma semi-conjugação é, deste modo, uma órbita de g, enquanto uma conjugação topológica envia órbitas de f para órbitas de g e preserva suas propriedades topológicas. Uma conjugação h pode ser interpretada como uma mudança de variáveis contínua a qual identifica f e g. Fácilmente vemos que uma conjugação topológica é uma relação de equivalência.

(iv) Se $\varphi: X \to \mathbb{R}$ é uma função sobre um espaço topológico X, o *suporte* de φ é o subconjunto de X definido por $supp\,\varphi = \overline{\varphi^{-1}(\mathbb{R} - \{0\})}$

Teorema 3.2. As seguintes afirmações são equivalentes:

(1) Existe uma aplicação expansora e C^1 de S^1 a qual, preserva orientação, preserva a medida de Lebesgue, mas para a qual a medida de Lebesgue não é ergódica.

(2) Existe uma $g \in \mathcal{G}^{comp}$ que tem mais que uma g-medida.

(3) Existe uma $g \in \mathcal{G}(S^1)$, que tem mais que uma g-medida.

Prova: (1) $\Rightarrow$ (2) : Suponha que (1) se verifica. Então suponha que a aplicação expansora T em questão, é uma aplicação de grau r. Por uma mudança de coordenadas da forma $x \mapsto x + \alpha$, podemos assumir que $T(0) = 0$. Considere então $0, a_1, \cdots, a_{r-1}$, como sendo as pré-imagens de 0. Sejam os conjuntos $I_0 = [0, a_1], I_1 = [a_1, a_2], \cdots, I_{r-1} = [a_{r-1}, 1]$. Estes conjuntos formam uma partição de Markov do círculo. Por um argumento padrão, existe uma semi-conjugação π_1 de (σ, Σ_r) para (T, S^1). Pela condição expansora, podemos verificar que cada ponto do círculo tem uma única pré-imagem sob π_1, exceto para uma quantidade enumerável de pontos que são pré-imagens de 0 sob T. Em particular, $\pi_1(x) = \pi_1(y)$ se e sò se $x \sim y$. Por conseguinte podemos definir uma medida sobre Σ_r colocando $\nu(A) = \lambda(\pi_1(A))$, onde λ denota a medida de Lebesgue, visto que λ não tem átomos. Agora fixe um ponto $x \in \Sigma_r$ e considere conjuntos A contendo x. Temos

$$\frac{\nu(A)}{\nu(\sigma(A))} = \frac{\lambda(\pi_1(A))}{\lambda(T(\pi_1(A)))}$$

Quando o diâmetro de A torna-se arbitráriamente pequeno, a razão do lado direito aproxima-se do limite $1/T'(\pi_1(x))$, deste modo temos que ν é uma g-medida, onde $g(x) = 1/T'(\pi_1(x))$. Mas π_1 é um isomorfismo mensurável

$$(\sigma, \Sigma_r, \nu) \to (T, S^1, \lambda)$$

por conseguinte, visto que λ não é ergódica, devemos ter ν não-ergódica, conseqüêntemente existe mais que uma g-medida. Esta g-função é claramente compatível, deste modo temos mostrado que (2) decorre.

(2) $\Rightarrow$ (3) : Suponhamos que (2) se verifica para $g \in \mathcal{G}^{comp}(\Sigma_r)$. Então defina

$$\pi_2 : \Sigma_r \to S^1, \quad \text{por} \quad x \mapsto \sum_{i=0}^{\infty} \frac{x_i}{r^{i+1}}$$

(por exemplo, para $r = 2 \Rightarrow \pi_2(010101\cdots) = 1/2^2 + 1/2^4 + 1/2^6 + \cdots = 1/3$, se expandirmos $1/3$ na base 2 teremos exatamente $01010101\cdots$). Esta aplicação é uma semi-conjugação de (σ, Σ_r) para (T_r, S^1) com a propriedade de que $x \sim y$ se e sò se $\pi_2(x) = \pi_2(y)$. Então seja ν uma g-medida não-ergódica .

Defina uma medida μ sobre S^1 por:

$$\mu(A) = \nu(\pi_2^{-1}(A))$$

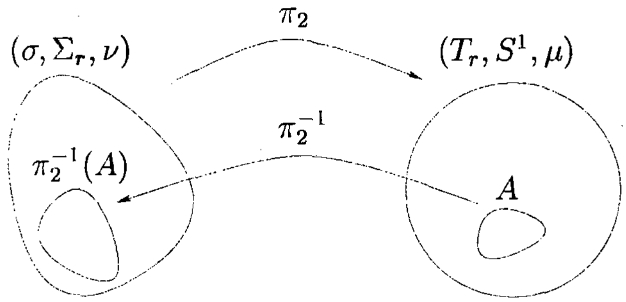

Figura 3.3: Definição de μ

- Dado um ponto $x \in S^1$, considere A um conjunto ao qual x pertence. Temos:

$$\frac{\mu(A)}{\mu(T_r(A))} = \frac{\nu(\pi_2^{-1}(A))}{\nu(\sigma(\pi_2^{-1}(A)))}$$

Se x não é uma pré-imagem de 0 sob T_r, então esta expressão claramente converge para $g(\pi_2^{-1}(x))$. Mas se x é uma pré-imagem de 0 sob T_r, a mesma conclusão se verifica devido a compatibilidade de g. Segue que μ é uma h-medida, onde $h(x) = g(\pi_2^{-1}(x))$. Esta função é contínua e está bem definida, tendo em conta a compatibilidade de g. Novamente, π_2 é um isomorfismo mensurável $(\sigma, \Sigma_r, \nu) \to (T_r, S^1, \mu)$, por conseguinte segue que μ não é ergódica, e conseqüêntemente existe mais que uma h-medida. Isto prova (3).

(3) $\Rightarrow$ (1) : Finalmente, suponha que (3) se verifica. Seja $g \in \mathcal{G}(S^1)$ uma g-função que tem mais que uma g-medida. Então seja ν uma g-medida não-ergódica. Defina $h(x) = \nu[0, x]$. Este é um homeomorfismo do círculo que preserva orientação (visto que g-medidas têm como suporte todo o conjunto X e não têm átomos). Seja μ a medida puxada para frente por h. Então temos:

$$\mu[0, x] = \nu(h^{-1}[0, x]) = \nu([0, h^{-1}(x)]) = h(h^{-1}(x)) = x$$

Disto segue que μ é a medida de Lebesgue.

Seja $T = h \circ T_r \circ h^{-1}$. Então temos que (T_r, S^1, ν) é um isomorfismo mensurável para (T, S^1, λ) por h e, deste modo concluimos que λ não é ergódica para T. Agora suponha que $x \in S^1$ está fixado e $y < x < z$. Sendo assim, temos

$$\frac{T(z) - T(y)}{z - y} = \frac{\lambda(T[y, z])}{\lambda([y, z])} = \frac{\nu(T_r[h^{-1}(y), h^{-1}(z)])}{\nu([h^{-1}(y), h^{-1}(z)])}$$

Tomando olimite quando y cresce para x e z decresce para x, obtemos convergência para $1/g(h^{-1}(x))$. Segue que T é diferenciável e expansora, preservando a medida de Lebesgue (a qual não é ergódica para T), deste modo temos mostrado que (1) se verifica, completando assim a prova do teorema.

OBS: Vamos justificar que: $\lambda(T[y, z]) = \nu(T_r[h^{-1}(y), h^{-1}(z)])$

$$T = h \circ T_r \circ h^{-1} \Rightarrow T[y, z] = (h \circ T_r)(h^{-1}([y, z]))$$

$$\Rightarrow T[y, z] = (h \circ T_r)([h^{-1}(y), h^{-1}(z)]) = h(T_r([h^{-1}(y), h^{-1}(z)])) \qquad (*)$$

Sabemos que, por definição, $\mu(A) = \lambda(A) = \nu(h^{-1}(A))$, logo aplicando λ á igualdade $(*)$ acima, temos

$$\lambda(T[y, z]) = \nu(h^{-1}\{h(T_r([h^{-1}(y), h^{-1}(z)]))\}) = \nu(T_r([h^{-1}(y), h^{-1}(z)]))$$

Ainda:

$$\lim_{x\to z}\frac{T(z)-T(y)}{z-y}=\frac{1}{g(h^{-1}(x))}\Rightarrow \quad \text{T é diferenciável}$$

observe que

$$|T'(x)|=\left|\frac{1}{g(h^{-1}(x))}\right|>1\Leftrightarrow g(h^{-1}(x))<1$$

o que é sempre verdade pois $g(y)\in(0,1)$. Isto é T é expansora.

Do exposto acima, concluímos que para demonstrar o Teorema 3.1 (Teorema Principal), é suficiente achar um exemplo satisfazendo a condição (3) do Teorema 3.2.

Capítulo 4

Prova do Teorema Principal

Desejamos dar uma interpretação mais probabilística de uma g-medida sobre Σ_r. Devemos considerar seqüências de variáveis randômicas $(X_n)_{n\in\mathbb{Z}}$ tomando valôres no conjunto $\{0,\cdots,r-1\}$ (isto é, este é o espaço de estados- ver apêndice). Estritamente, devemos considerar X_n uma aplicação de algum espaço de probabilidade Ω para $\{0,\cdots,r-1\}$, e escrevemos $X_n(\omega)$ para X_n, devemos sempre usar o mesmo espaço de probabilidade, amiúde preferimos escrever simplesmente X_n.

Desejamos ver a evolução das variáveis randômicas especificando os vários resultados do experimento "presente" (isto é, X_0) condicionado sobre o "passado" (isto é, $(X_n)_{n<0}$). Exemplos simples e não triviais de tais experimentos são dados por cadeias de Markov com transição de probabilidade estacionária, onde as probabilidades dos resultados do experimento presente são completamente determinados pelo resultado prévio apenas (isto é, $P\{X_n = i/X_{n-1} = j_1, X_{n-2} = j_2, \cdots\}$ é independente de $j_2, j_3, \cdots$ e n). Podemos similarmente considerar os assim chamados processos de "imagem finita" ou *cadeias de Markov de k-passos*, onde as probabilidades são determinadas pelos resultados dos k experimentos prévios.

Desejamos generalizar os processos de imagem finita para os de "imagem infinita". Seja $(X_n)_{n\in\mathbb{Z}}$ uma seqüência de variáveis randômicas tomando valores no conjunto $\{0,\cdots,r-1\}$. Suponha que esta seqüência satisfaça

$$P\{X_n = i/X_{n-1} = a_1, X_{n-2} = a_2, \cdots\} = g(ia_1a_2\cdots) \tag{4.1}$$

onde $g \in \mathcal{G}$. Agora fixe um n, então temos uma aplicação natural

$$\rho_n \colon \Omega \to \Sigma_r$$

dada por $\rho_n(\omega)_i = X_{n-i}(\omega)$. Isto é $\rho_n(\omega) = (X_n(\omega), X_{n-1}(\omega), \cdots)$.

Seja $\nu_n = \rho_n^*(P)$, isto é

$$\nu_n(A) = P\{\rho_n^{-1}(A)\} \quad \forall A \in \mathcal{B}(\Sigma_r).$$

Lema 1. P é estacionária se e sòmente se $\nu_n = \nu_0$ é independente de n.

Prova: A propriedade $\nu_n = \nu_{n+k}$ basta ser verificada em cilindros. Logo seja,

$$\mathcal{C} = (j; a_1, \cdots, a_l) = \{x : x(j+i) = a_i,\ i = 1, \cdots, l\}$$

Seja $n \in \mathbb{Z}$, $k \in \mathbb{Z}$. Então

$$\rho_n^{-1}(\mathcal{C}) = \{\omega : \rho_n(\omega) \in \mathcal{C}\} = \{\omega : \rho_n(\omega)(j+i) = a_i\} =$$

$$= \{\omega : X_{n-j-i}(\omega) = a_i,\ i = 1, \cdots, l\} = \{X_{n-j-1} = a_1, X_{n-j-2} = a_2, \cdots, X_{n-j-l} = a_l\}$$

Logo,

$$\rho_{n+k}^{-1}(\mathcal{C}) = \{X_{n+k-j-1} = a_1, X_{n+k-j-2} = a_2, \cdots, X_{n+k-j-l} = a_l\}$$

Agora,

$$\nu_n(\mathcal{C}) = P\{\rho_n^{-1}(\mathcal{C})\} = P\{X_{n-j-1} = a_1, X_{n-j-2} = a_2, \cdots, X_{n-j-l} = a_l\}$$

e

$$P\{X_{n+k-j-1} = a_1, X_{n+k-j-2} = a_2, \cdots, X_{n+k-j-l} = a_l\} = \nu_{n+k}(\mathcal{C})$$

Segue que $\nu_{n+k}(\mathcal{C}) = \nu_n(\mathcal{C})$ se e sòmente se P é estacionária.

Definição. Denotamos por ν a medida de probabilidade invariante em Σ_r dada por

$$\nu = \nu_n \qquad \forall n.$$

Consideremos os seguintes cilindros:

$$[ix]^{n+1} = (ix_0x_1x_2 \cdots x_n)$$

$$[x]^n = (x_0x_1 \cdots x_n)$$

Observe que $\sigma(ix) = x$. Sejam as bolas de raios $\frac{1}{2^{k+1}}$ e $\frac{1}{2^k}$ com centro em ix e x denotadas por

$$[ix]^{k+1}\ ,\ [x]^k$$

Observe ainda que $\sigma([ix]^{k+1}) = [x]^k$

Como

$$\nu([ix]^{k+1}) = P\{X_n = i, X_{n-1} = x_0, \cdots, X_{n-k} = x_{k+1}\}$$

$$\nu([x]^k) = P\{X_{n-1} = x_0, \cdots, X_{n-k} = x_k\}$$

Logo,

$$\frac{\nu([ix]^{k+1})}{\nu([x]^k)} = \frac{P\{X_n = i, X_{n-1} = x_0, \cdots, X_{n-k} = x_{k+1}\}}{P\{X_{n-1} = x_0, \cdots, X_{n-k} = x_k\}} = g(ix)$$

isto é,

$$\lim_{k\to\infty} \frac{\nu([ix]^{k+1})}{\nu(\sigma([ix]^{k+1}))} = g(ix)$$

Segue do teorema de Radon-Nikodym que

$$\nu([ix])^{n+1} = \int_{[x]^n} g(i\gamma)\, d\nu(\gamma)$$

Sejam os conjuntos $[ix]^{n+1}$ e $[x]^n$. Então

$$\sigma \colon [ix]^{n+1} \to [x]^n$$

é uma bijeção. Logo a fórmula que provamos diz que $g(iy)$ é o jacobiano de σ^{-1} relativo à medida ν, isto é

$$\nu([ix]^{n+1}) = \nu((\sigma\big|_{[ix]^n})^{-1}([x]^n)) = \int_{[x]^n} g(iy)\, d\nu(y)$$

segue trivialmente da igualdade provada, usando o fato de que g é contínua, que ν é uma g-medida.

Estaremos também interessados no operador

$$\mathcal{L} \colon C(\Sigma_r) \to C(\Sigma_r) \text{ definido por } \mathcal{L}f(x) = \Sigma_{y\in T^{-1}(\{x\})} g(y)f(y)$$

Ocasionalmente escrevemos $\mathcal{L}_g$ para enfatizar a dependência sobre g. Por simples cálculo mostra-se que

$$\mathcal{L}^m\, f(x) = \Sigma_{y\in T^{-m}(\{x\})} g^{(m)}(y)f(y), \text{ onde } g^{(m)}(x) = g(x)g(\sigma(x))\cdots g(\sigma^{m-1}(x)).$$

A interpretação deste operador é que $\mathcal{L}^m f(x)$ é a esperança de $f(X_m, X_{m-1}, \cdots)$ condicionada sobre $X_{-i} = x_i$, para todo $i \geq 0$, dado que (X_n) satisfaz (4.1).

Keane demonstrou que se g é Lipschitz então existe uma única g-medida, quando $X = S^1$. Walters mostrou que a mesma conclusão se verifica quando $X = \Sigma_r$ se g tem variação somàvel (isto é, $\sum_{i=0}^{\infty} var_i\, g < \infty$), ou em particular se g é *Hölder*. Êle também mostrou que nesta situação, se a única g-medida é ν então para qualquer função contínua f, teremos $\mathcal{L}^m f(x)$ convergindo uniformemente para $\int f d\nu$ quando $m \to \infty$.

Antes de passarmos aos aspectos técnicos da construção de uma g-função em Σ_r

($r = 10$) compatível com mais de uma g-medida, daremos uma idéia geral do que pretendemos.

Seja a aplicação

$$\pi : \Sigma_{10} \to I$$

definida por

$$\pi(x) = \sum_{i=0}^{\infty} \frac{x_i}{10^{i+1}}$$

Estaremos sempre identificando Σ_{10} com I, omitindo portanto π na maioria das vezes.

Seja $L: I \to I$ dado por $L(x) = 1 - x$. Nosso objetivo é construir $g: \Sigma_{10} \hookleftarrow$ compatível satisfazendo

(i) $gL = g$

(ii) existe ν g-medida tal que $\nu([6]) > 21/40$.

Assumindo, por um momento, tal construção possível, definimos

$$\mu = L^* \nu$$

isto é

$$\mu(A) = \nu(L(A)) \qquad \forall A \in \beta(\Sigma_{10}).$$

Proposição. μ é uma g-medida.

Prova: De fato, se $A \in \beta(\Sigma_{10})$ e $x \in A$, temos

$$\frac{\mu(A)}{\mu(\sigma A} = \frac{\nu(L(A))}{\nu(L\sigma(A))} = \frac{\nu(L(A))}{\nu(\sigma LA)} \to g(Lx), \qquad |A| \to 0.$$

pois ν é g-medida e onde usamos que L e σ comutam, isto é

$$L\sigma = \sigma L$$

Mas $g(Lx) = g$, logo

$$\lim_{\substack{|A|\to 0\\ x\in A}} \frac{\mu(A)}{\mu(\sigma A)} = g(Lx) = g(x) \quad \mu(q.t.p.)$$

isto é μ é uma g-medida.

Agora provaremos que $\mu \neq \nu$ concluindo a argumentação. Observe que

$$\mu([3]) = \nu(L([3])) = \nu([6]) > \frac{21}{40}$$

isto é, se $\mu = \nu$ então como

$$[6] = [3] = \emptyset$$

temos,

$$\nu([6] \cup [3]) = \nu([6]) + \nu([3]) > \frac{21}{40} + \frac{21}{40} = \frac{21}{20} > 1$$

contradizendo o fato de que ν é uma medida de probabilidade. Sendo assim segue $\mu \neq \nu$.

Devemos trabalhar no que segue com $r = 10$, deste modo produziremos uma aplicação de dez-ramos.

Definiremos uma ordem parcial sobre $\{0, 1, \cdots, 9\}$ por $3 \preceq i \preceq 6$, para todo i. Então definimos uma ordem parcial sobre Σ_{10} por $x \preceq y$ se $x_i \preceq y_i$, para todo $i \in \mathbb{Z}^+$.

Uma g-função é chamada *atrativa* ou *monotônica* se

$$\sum_{i\geq j} g(ix) \leq \sum_{i\geq j} g(iy), \text{ sempre que } x \preceq y$$

Escrevemos π para a aplicação $\Sigma_{10} \to I$, definida por

$$x \mapsto \sum_{i=0}^{\infty} \frac{x_i}{10^{i+1}}.$$

Nos voltaremos para a construção de duas funções Hölder's, g e h, sobre I tal que $g(0) = 0, g(1) = 1, h(0) = 1, h(1) = 0$ e $x \preceq y \Rightarrow g(x) \leq h(y), h(x) \leq h(y)$. Devemos usar esta função para escrever uma g-função posteriormente.

Defina um operador $\Phi: C(I) \to C(I)$ como segue:

$$\Phi f(x) = \begin{cases} 0, & 0 \leq x < 0.4 \\ \frac{1}{2} f(\sigma(x)), & 0.4 \leq x < 0.5 \\ \frac{1}{2}(f(\sigma(x)) + 1), & 0.5 \leq x < 0.6 \\ 1, & 0.6 \leq x \leq 1 \end{cases}$$

onde $C(I) = \{f : I \to I, f(0) = 0, f(1) = 1, f \text{ contínua}\}$. É fácil ver que este operador aplica $C(I)$ sobre si mesmo, e êle pode ser visto também como uma contração (com respeito à norma uniforme). Disto segue que existe um único ponto fixo g. Resta mostrar que esta função respeita a ordem e é Hölder. Para isto, suponha que x e y moram no mesmo n-cilindro de Σ_{10} (onde $n \geq 0$). Então visto que g é um ponto fixo de Φ, segue que $|g(x) - g(y)| \leq \frac{1}{2}|g(\sigma(x)) - g(\sigma(y))|$.

Por indução, segue que $|g(x) - g(y)| \leq 2^{-n}$. Disto segue que g é Hölder.

Proposição. Suponha $x \prec y$. Então $g(x) \leq g(y)$.

Prova: Por indução sobre o primeiro dígito onde x e y diferem. Note que por definição de $\preceq$, se x e y diferem na posição 0 então ou x inicia com um 3 (caso em que $g(x) = 0 \leq g(y)$) ou y começa em 6 (caso em que $g(x) \leq 1 = g(y)$). Claramente, um ou outro caso se confirma. Suponha que a proposição se verifica para todo x e y que diferem pela primeira vez na posição n (com $n > 0$), mas suponha que são dados $x \prec y$ com $x_i = y_i$, para todo $i \leq n$ e $x_{n+1} \prec y_{n+1}$. Se $x_0 < 4$ ou $x_0 > 6$, então $g(x) = g(y)$. Por outro lado, $g(x) - g(y) = \frac{1}{2}(g(\sigma(x)) - g(\sigma(y)))$. Mas $\sigma(x) \prec \sigma(y)$ e êles diferem pela primeira vez na posição n, deste modo pela hipótese de indução, vemos $g(x) - g(y) \leq 0$ e a proposição se verifica. □

Comentário sobre o tipo de demonstração usado: Se $x \prec y$ então $x \neq y$, isto é, existe um dígito i_0 em que x e y diferem (além é claro, de $x_i \preceq y_i, \quad \forall i \in \mathbb{N} \setminus \{i_0\}$).Então deve-se mostrar que qualquer que seja este $i_0 \in \mathbb{N}$ a tese se verifica. De outra forma: qualquer que seja o natural (i.e., a posição) em que x e y difiram pela primeira vez, a tese deve se verificar.

Agora devemos construir h. Seja D dado por $\{f: I \to I; f(0) = 1, f(1) = 0, \ f \text{ contínua}\}$. Então defina $\Psi: D \to D$ por

$$\Psi f(x) = \begin{cases} \frac{1}{2}(1+f(\sigma(x))), & 0 \leq x < 0.1 \\ \frac{1}{2}f(\sigma(x)), & 0.1 \leq x < 0.2 \\ g(x) & 0.2 \leq x < 0.8 \\ \frac{1}{2}(1+f(\sigma(x))), & 0.8 \leq x < 0.9 \\ \frac{1}{2}f(\sigma(x)), & 0.9 \leq x \leq 1 \end{cases}$$

Então Ψ é também uma contração. Argumentos similares àqueles dados anteriormente mostram que o único ponto fixo h também respeita a ordem parcial e é Hölder (com a mesma constante).

Notamos que $h(1-x) = 1 - h(x)$ e $g(1-x) = 1 - g(x)$ essas duas funções revertem a ordem parcial e devem também ser usadas na construção da g-função.

Dado $m \in \mathbb{N}$, seja S_m denotando $\{0, 1, \cdots, 9\}^m$, um conjunto de palavras de comprimento m. Dado $a \in S_m$, defina $a-1 \in S_m$ e $a+1 \in S_m$ da maneira óbvia (isto é tal $13246+1 = 13247$ e $00000 - 1 = 99999$ por exemplo). Defina

$$\psi^i : S_m \to \{0, 1, \cdots, 9\}$$

tomando $\psi^i(a)$ igual ao número de ocorrências do símbolo i em a. Definamos então $\delta(a)$ dado por $\delta(a) = \psi^6(a) - \psi^3(a)$.

Observe que $\delta(a)$ é a diferença entre o número de $6's$ e $3's$ na palavra a. Por exemplo, $\delta(0011223344) = \psi^6(0011223344) - \psi^3(0011223344) = 0 - 2 = -2$

Proposição. Se $a \in S_m$, então $|\delta(a) - \delta(a+1)| \leq 1$

Prova: (Por absurdo): Suponhamos $a \in S_m$ e $|\delta(a) - \delta(a+1)| > 1$.Observe que

$$\delta(a) - \delta(a+1) = \psi^6(a) - \psi^6(a+1) + \psi^3(a+1) - \psi^3(a)$$

tome, por exemplo, $a = 111 \cdots 1$(m posições). Então $a+1 = 111 \cdots 2$, logo

$$\psi^6(a) = \psi^6(a+1) = 0 \quad , \quad \psi^3(a) = \psi^3(a+1) = 0$$

$\Rightarrow |\delta(a) - \delta(a+1)| = 0 > 1 \rightarrow\leftarrow$

Escrevemos $[a]$ para o m-cilindro consistindo de todos os elementos de Σ_{10} cujos primeiros m termos estão no bloco a. Então definimos

$$V^6_{m,n} : S^1 \to [0, 1]$$

cilindro por cilindro :

$$V^6_{m,n}\Big|_{[a]}(x) = \begin{cases} 1, & \delta(a) > n \\ 0, & \delta(a) < n \\ 0, & \delta(a) = n, \delta(a-1) \leq n, \delta(a+1) \leq n \\ g(\sigma^m(x)), & \delta(a) = n, \delta(a-1) \leq n, \delta(a+1) > n \\ h(\sigma^m(x)), & \delta(a) = n, \delta(a-1) > n, \delta(a+1) \leq n \\ 1, & \delta(a) = n, \delta(a-1) > n, \delta(a+1) > n \end{cases}$$

Então é imediato verificar que $V^6_{m,n}$ é Hölder e respeita a ordem parcial.

Defina então

$$V^3_{m,n}(x) = V^6_{m,n}(1-x)$$

Por observações anteriores, vemos que esta aplicação reverte a ordem e de fato satisfaz

$$V^3_{m,n}(x) = 1 - V^6_{m,-n}(x)$$

Agora seja

$$W^6_{m,n}(x) = \frac{1}{10} + \frac{1}{2}V^6_{m,n}(x)$$
$$W^3_{m,n}(x) = \frac{1}{10} + \frac{1}{2}V^3_{m,n}(x)$$
$$W^i_{m,n}(x) = \frac{1}{10} - \frac{1}{16}(V^6_{m,n}(x) + V^3_{m,n}(x)), \quad i \neq 3, 6$$

Temos que

$$\Sigma_i W^i_{m,n}(x) \equiv 1$$

De fato,

$$\sum_{i=0}^{9} W^i_{m,n}(x) = (\frac{1}{10} + \frac{1}{2}V^6_{m,n}(x)) + (\frac{1}{10} + \frac{1}{2}V^3_{m,n}(x)) +$$

$$+8\{\frac{1}{10} - \frac{1}{16}(V^6_{m,n}(x) + V^3_{m,n}(x))\} = 1.$$

Consideraremos sempre o caso onde $n > 0$ e neste caso, é fácil ver que para cada x, um dos valores $V^6_{m,n}$ e $V^3_{m,n}(x)$ é 0 (pois, se $n > 0 \Rightarrow -n < 0 \Rightarrow \delta(a) \geq 0 > -n \Rightarrow V^6_{m,-n}(x) = 1 \Rightarrow V^3_{m,n}(x) = 1 - V^6_{m,-n}(x) = 0$). Segue que

$$\frac{1}{10} \leq W^i_{m,n}(x) \leq \frac{6}{10} \quad \text{para}\, i = 3, 6 \quad \text{e} \quad \frac{3}{80} \leq W^i_{m,n}(x) \leq \frac{1}{10}$$

caso contrário. Nossa g-função deve ser construída como uma combinação afim infinita de $W_{m,n}$.

-Construção do Exemplo

Seja $q_j = \frac{1}{2}(\frac{2}{3})^j$, deste modo $\sum_{j=1}^{\infty} q_j = 1$. Devemos escolher n_j e m_j tal que tomando

$$g(ix) = \sum_{j=1}^{\infty} q_j W^i_{m_j,n_j}(x)$$

desejamos dar uma g contínua e compatível com mais que uma g-medida. A escolha de m_j e n_j deve ser feita indutivamente, por considerações de certas truncações Hölder's da g-função final. Suponha $m_1, \cdots, m_{k-1}$ e $n_1, \cdots, n_{k-1}$ são escolhidos. Então defina vetores como segue:

$$u = (\frac{3}{80}, \frac{3}{80}, \frac{3}{80}, \frac{3}{5}, \frac{3}{80}, \frac{3}{80}, \frac{1}{10}, \frac{3}{80}, \frac{3}{80}, \frac{3}{80})$$

$$v = (\frac{3}{80}, \frac{3}{80}, \frac{3}{80}, \frac{1}{10}, \frac{3}{80}, \frac{3}{80}, \frac{3}{5}, \frac{3}{80}, \frac{3}{80}, \frac{3}{80})$$

$$z = (\frac{3}{80}, \frac{3}{80}, \frac{3}{80}, \frac{7}{20}, \frac{3}{80}, \frac{3}{80}, \frac{7}{20}, \frac{3}{80}, \frac{3}{80}, \frac{3}{80}),$$

onde os índices varrem de 0 a 9 (é útil observar que, com exceção dos termos de posições 3 e 6 todos os demais são iguais).

Agora defina

$$g^1_k(ix) = \sum_{j=1}^{k-1} q_j W^i_{m_j,n_j}(x) + \sum_{j=k}^{\infty} q_j z_i$$

$$g^2_k(ix) = \sum_{j=1}^{k-1} q_j W^i_{m_j,n_j}(x) + q_k u_i + \sum_{j=k+1}^{\infty} q_j v_i$$

$$g^3_k(ix) = \sum_{j=1}^{k-1} q_j W^i_{m_j,n_j}(x) + q_k W^i_{M,N}(x) + \sum_{j=k+1}^{\infty} q_j v_i,$$

onde $M > N > 0$. Estas são todas g-funções Hölder's e tal que têm g-medidas únicas, as quais chamaremos μ^e_k onde $e = 1, 2, 3$. Primeiro note que g^1_k é simétrica: $g^1_k(1-x) = g^1_k(x)$. Isto significa que a única medida invariante deve ser preservada sob a involução $x \mapsto 1 - x$. Segue que $\mu^1_k([6]) = \mu^1_k([3])$. Deveremos usar a propriedade de preservação de ordem de g para mostrar que $\mu^3_k([6]) \geq mu^2_k([6]) > \mu^1_k([6])$ e $\mu^3_k([3]) \leq \mu^2_k([3]) < \mu^1_k([3])$. Seja $\alpha_k = \frac{1}{16}(\frac{2}{3})^k$.

Antes do próximo lema lembremos brevemente o que vem a ser um 'coupling' em processos estocásticos:

Suponha que (X_n) e (Y_n) são seqüências de variáveis randômicas com distribuições P_1 e P_2 respectivamente, então um *coupling* é uma distribuição de probabilidade conjunta P cuja distribuição marginal sobre (X_n) é P_1 e sobre (Y_n) é P_2.

Lema 2. Temos que

$$\mu_k^2([6]) \geq \mu_k^1([6]) + 2\alpha_k \quad \text{e} \quad \mu_k^2([3]) \leq \mu_k^1([3]) - 2\alpha_k.$$

Além disso, suponha que é dado $x \in \Sigma_{10}$. Então existe um coupling dos dois processos (Y_n) e (Z_n) evoluindo sob g_k^2 e g_k^3 respectivamente, condicionado sobre $Y_i = Z_i = x_{-i}$, para todo $i \leq 0$ tal que $Y_n \preceq Z_n$ com probabilidade 1 para todo n.

Prova. A prova trabalha encontrando couplings de dois processos evoluindo sob diferentes g-funções, as quais obviamente fazem as desigualdades verdadeiras. É fácil checar que

$$g_k^2(6x) - g_k^1(6x) = 2\alpha_k \quad e \quad g_k^2(3x) - g_k^1(3x) = -2\alpha_k,$$

enquanto

$$g_k^2(ix) = g_k^1(ix), \forall i \neq 3,6 \,\text{e todo}\, x.$$

(na verdade vale, $g_k^2(ix) = g_k^1(jx), \forall i,j \neq 3,6 \,\text{e todo}\, x.$)

Usamos isto para dar explícitamente um coupling dos dois processos ramdômicos (X_n) e (Y_n) evoluindo sob g_k^1 e g_k^2 respectivamente como em (4.1). Escrevemos $P(ix, jy)$ em que i é adicionado a x e j é adicionado a y. A probabilidade de transição deve ser definida sòmente quando $x \preceq y$, e deve-se por essa razão ter $P(ix \preceq jy) = 1$, naturalmente que isto pode ser aplicado repetidamente. Suponha $x \preceq y$. Então defina

$$P(ix, jy) = \begin{cases} g_k^1(6x), & i = j = 6 \\ max(0, g_k^1(ix) - g_k^2(iy)), & i \neq 3,6 \,\text{e}\, j = 6 \\ min(g_k^1(ix), g_k^2(iy)), & i = j \neq 3,6 \\ max(0, g_k^2(iy) - g_k^1(ix)), & i = 3 \,\text{e}\, j \neq 3,6 \\ min(g_k^2(6y) - g_k^1(6x), g_k^1(3x) - g_k^2(3y)), & i = 3 \,\text{e}\, j = 6 \\ g_k^2(3y), & i = j = 3 \end{cases}$$

Note que todas as probabilidades de transição são não-negativas, e devemos checar que as marginais destes couplings são exigidos. Calcularemos um exemplo como ilustração. Mostraremos que sob P, a probabilidade de que x vá para $3x$ é $g_k^1(3x)$ como requerido.

Antes do exemplo faremos o seguinte comentário (ver apêndice): Sejam X e Y variáveis rândomicas tomando valores inteiros, e seja,

$$p_{ij} = P(X = i, Y = j)$$

a distribuição conjunta de probabilidade. A *distribuição marginal* de X e Y são dadas por

$$P_{i.} = \Sigma_j p_{ij} \quad , \quad P_{.j} = \Sigma_i p_{ij}$$

respectivamente. Em nosso contexto $X = X_n, Y = Y_n$ e $p_{ij} = P(X = i, Y = j) = P(ix, jy)$. Em nosso exemplo $i = 3$, logo $p_{3.} = \Sigma_j p_{3j}$, onde esta soma se verifica sobre todos os índices j que satisfazem as restrições na definição de $P(ix, jy)$, isto é

$$p_{3.} = \Sigma_j p_{3j} = p_{30} + p_{31} + p_{32} + p_{33} + p_{34} + p_{35} + p_{36} + p_{37} + p_{38} + p_{39}.$$

Por observação, vemos que a probabilidade de que x vá para $3x$ é

$$g_k^2(3y) + min(g_k^2(6y) - g_k^1(6x), g_k^1(3x) - g_k^2(3y)) + 8max(0, g_k^2(iy) - g_k^1(ix))$$

$$= g_k^1(3x) + min(g_k^2(6y) - g_k^1(6x) - g_k^1(3x) + g_k^2(3y), 0) + 8max(0, g_k^2(iy) - g_k^1(ix))$$

$$= g_k^1(3x) + min(8g_k^1(ix) - 8g_k^2(ix), 0) + 8max(0, g_k^2(iy) - g_k^1(ix)) = g_k^1(3x)$$

como requerido, onde usamos o fato de que $g_k^2(ix) = g_k^1(jx), \forall i, j \neq 3, 6$ e todo x. Isto mostra que dado $x \preceq y$, podemos escolher i e j tal que y evolui portanto para g_k^2 e x também evolui para g_k^1 de tal forma que a probabilidade é 1, $ix \leq jy$. Além disso observando no coupling, vemos que a probabilidade de que y é precedido por um 6 e x não é precedido por um 6 dado que $x \preceq y$ é $g_k^2(6y) - g_k^1(6x)$, mas $g_k^2(6y) - g_k^1(6y) = 2\alpha_k$ e $g_k^1(6y) \geq g_k^1(6x)$, deste modo segue que como (x, y) vai para $(ix, 6y)$ para algum $i \neq 6$ com probabilidade no mínimo $2\alpha_k$. Disto segue $\mu_k^2([6]) \geq \mu_k^1([6]) + 2\alpha_k$. Um argumento similar mostra que $\mu_k^2([3]) \leq \mu_k^1([3]) - 2\alpha_k$.

Para provar a parte restante do Lema, é necessário considerar um coupling do processo (Y_n) evoluindo sob g_k^2 e (Z_n) evoluindo sob g_k^3. Isto é feito por um coupling exatamente similar ao coupling acima, com g_k^2 substituindo g_k^1 e g_k^3 substituindo g_k^2. A conclusão então é que dado $y \preceq z$, então y pode ser deixado evoluir sob g_k^2 e z sob g_k^3 de tal modo que a ordem é preservada.

□

Agora descreveremos a escolha indutiva de m_k e n_k. Em cada caso, n_k é dado por $[\alpha_k m_k]$. Suponha que escolhemos $m_1, m_2, \cdots, m_{k-1}$ e conseqüentemente $n_1, n_2, \cdots, n_{k-1}$. Seja

$$\eta(x) = \chi_{[6]}(x) - \chi_{[3]}(x) \quad \text{e} \quad \Delta_j(x) = \sum_{i=0}^{m_j - 1} \eta(\sigma^i(x)).$$

Seja G_j denotando $\{x : \Delta_j(x) \geq n_j\}$ e H_j denotando $\{x : \Delta_j(x) \geq 3n_j\}$. Então note que $\sigma^{-n_j}(H_j) \subseteq G_j$. Note também que se $x \in H_j$ e $y \succeq x$, então $y \in H_j$. Assuma agora que existem $t_1, t_2, \cdots, t_{k-1}$ tal que

$$P\{(X_{t_j}, X_{t_j - 1}, \cdots) \in H_j \backslash X_{-i} = x_i, \ \forall i \geq 0\} \geq 1 - 4^{-j} \tag{4.2}$$

para todo $j < k$ e $x \in \Sigma_{10}$, onde X_n evolui conseqüentemente para g_j^3.

Seja $A_m = \{x : \Delta_m(x) > 3\alpha_k m\}$. Sabemos que $\int \eta(x) d\mu_k^2(x) \geq 4\alpha_k$ e devemos usar isto para mostrar que $\mu_k^2(A_m) \to 1$ quando $m \to \infty$.

Lema 3. Temos $\mu_k^2(A_m) \to 1$ quando $m \to \infty$.

Prova. Suponha que a exigência não se verifica. Visto que temos $\mu_k^2(A_m) \leq 1$ para todo m, a única maneira de a exigência falhar é se existe um $\varepsilon > 0$ e uma seqüência M_i de modo que $\mu_k^2(A_M) < 1 - \varepsilon$ para todo i. Neste caso, temos

$$\mu_k^2\left(\bigcup_{i>j} A_i^c\right) > \varepsilon \text{ para todo } j,$$

deste modo

$$\mu_k^2\left(\bigcap_j \bigcup_{i>j} A_i^c\right) \geq \varepsilon$$

Seja $S = \bigcap_j \bigcup_{i>j} A_i^c$. Se $x \in S$ então

$$\lim_{n\to\infty} \inf \frac{1}{n} \sum_{i=0}^{n-1} \eta(\sigma^i(x)) \leq 3\alpha_k.$$

Temos contudo que μ_k^2 é ergódica, deste modo para quase todo x (com respeito a μ_k^2), temos

$$\lim_{n\to\infty} \inf \frac{1}{n} \sum_{i=0}^{n-1} \eta(\sigma^i(x)) \geq 4\alpha_k.$$

Isto é uma contradição. □

A seguir, escolha m_k de modo que $\mu_k^2(A_{m_k}) > 1 - 4^{-k}$ e $\alpha_k m_k > t_{k-1}$. Agora $H_k = A_{m_k}$. Visto que g_k^2 é Hölder, podemos aplicar o teorema de Walter's [12] para concluir que $\mathcal{L}^n_{g_k^2} \chi_{H_K(x)}$

converge uniformemente para $\mu_k^2(H_k)$, o qual é maior que $1 - 4^{-k}$. Segue que existe um t_k de modo que $\mathcal{L}_{g_k^2}^{t_k}\chi_{H_K(x)} \geq 1 - 4^{-k}$, para todo $x \in \Sigma_{10}$. Isto diz que para todo $x \in \Sigma_{10}$,

$$P\{(X_{t_k}, X_{t_k-1}, \cdots) \in H_k \backslash X_{-i} = x_i,\ \forall i \geq 0\} \geq 1 - 4^{-k},$$

onde X_n evolui conseqüêntemente para g_k^2, mas pela segunda afirmação do lema 2, disto segue que a mesma equação se verifica quando a evolução se processa para g_k^3. Isto é precisamente a afirmação de (4.2) quando tomamos j sendo igual a k. Isto completa o passo indutivo.

Para completar a construção indutiva do exemplo, resta sòmente especificar um exemplo inicial para a indução. Tomando $t_0 > 1$, aplicando os passos de indução anteriores produzimos m_1, n_1 e t_1 que podem ser usados como ponto de partida para indução.

-COMPLETAMENTO DA PROVA

Prova do Teorema Principal. Na seção acima, m_i e n_i foram construídos indutivamente, sendo assim a g-função é agora dada por

$$g(ix) = \sum_{j=1}^{\infty} q_j W_{m_j,n_j}^i(x).$$

Note que para checar a compatibilidade, deve-se checar $g(000\cdots) = g(999\cdots)$ e $g(i999\cdots) = g(j000\cdots)$ para $i < 9$ e $j = i+1$. Isto é contudo direto já que todas estas quantidades são iguais a $\frac{1}{10}$. A continuidade de g é também clara já que é o limite uniforme de funções contínuas. Por essa razão resta mostrar que existem duas g-medidas distintas para esta g.

Consideremos os eventos E_k^t para os quais $(X_t, X_{t-1}, \cdots) \in H_k$. Escreva $I\!P_6$ para a distribuição de probabilidade X_n condicionada a $X_i = 6$ para todo $i < 0$ com evolução subseqüente sob g. Informalmente, E_k^t é o evento do processo tendo "uma grande majoração de 6s sobre 3s em cada escala m_k vezes t."Então considerando o processo evoluindo de uma condição inicial para longe de todos 6s (deste modo $I\!P_6\{E_k^0\} = 1$, para todo k). Mostraremos indutivamente, por indução sobre t, que os eventos E_k^t têm uma alta probabilidade, usaremos o resultado anterior o qual diz que se o processo tem uma grande majoração de 6s sobre escalas $m_{k+1}, m_{k+2}, \cdots$ no tempo $t - t_k$, então com alta probabilidade, o processo deve ter uma majoração de 6s sobre escala m_k no tempo t.

Lema 4. Temos

$$I\!P_6\{E_k^t\} \geq 1 - \zeta_k \tag{4.3}$$

para todo $t \in \mathbb{Z}$ e $k \in \mathbb{N}$ onde $\zeta_k = \frac{3}{2}4^{-k}$.

Prova. A prova é por indução sobre t. Note que a hipótese é automáticamente verdadeira para todo k se $t < 0$, deste modo necessitamos provar apenas o passo indutivo. Suponha que (4.3) se verifica para todo $t < s$ então tome $k \in \mathbb{N}$. Seja $S = \bigcap_{j>k} E_j^{s-t_k}$. Então por hipótese de indução,

$$\mathbb{P}_6\{S^c\} \leq \sum_{j>k} \zeta_j = \frac{1}{2}4^{-k}$$

Agora decompomos $(E_k^s)^c$ como $((E_k^s)^c) \bigcap S) \bigcup ((E_k^s)^c \bigcap S^c))$. Então temos

$$\mathbb{P}_6\{(E_k^s)^c\} \leq \mathbb{P}_6\{(E_k^s)^c \cap S\} + \mathbb{P}_6\{S^c\} \leq \mathbb{P}_6\{(E_k^s)^c \backslash S\} + \frac{1}{2}4^{-k}.$$

Agora suponha $\omega \in S$. Então seja $x = (X_{s-t_k}, X_{s-t_k-1}, \cdots)$ e $z = (X_s, X_{s-1}, \cdots)$. Então $x \in \bigcap_{j>k} H_j$. Disto segue que se $y \in \sigma^{-t}(x)$, para algum $t \leq t_k$ então $y \in \bigcap_{j>k} G_j$. Em particular, $g(y) = g_k^3(y)$, onde M e N em g_k^3 são tomados como sendo m_k e n_k. Segue que a evolução de x para t_k passos toma lugar sob g_k^3, mas por (4.2), a probabilidade de que $z \in (E_k^s)^c$ não é maior que 4^{-k}. Em particular, temos mostrado que $\mathbb{P}_6\{(E_k^s)^c\} \leq \zeta_k$ como requerido. Isto completa a prova do passo indutivo e conseqüentemente do lema. □

Devemos calcular $\mathbb{P}_6\{X_n = 6\}$. Usando o lema acima, isto está limitado abaixo por

$$\sum_{j \geq 1} q_j((1-\zeta_j)\frac{3}{5} + \zeta_j\frac{1}{10}).$$

Isto fica sendo igual $\frac{21}{40}$. Seja $\mu_n = \rho_n^*(\mathbb{P}_6)$, como definido anteriormente. Então temos

$$\mu_{n+1}([ix]^{m+1}) = \int_{[x]^m} g(iy)\, d\mu_n(y).$$

Agora seja $\nu_n = \frac{1}{n}\sum_{j=0}^{n-1} \mu_j$. Então vemos que

$$\left| \nu_n([ix]^{m+1}) - \int_{[x]^m} g(iy)\, d\nu_n(y) \right| \leq \frac{2}{n}.$$

Tomando uma subseqüência $\nu_{n_k} \to \nu$ de ν_n $fraco^*$-convergente, achamos

$$\nu([ix]^{m+1}) = \int_{[x]^m} g(iy)\, d\nu_n(y)$$

Como notado no início deste capítulo isto implica que ν é uma g-medida. Contudo $\mu_n([6]) \geq \frac{21}{40}$, para todo n, sendo assim segue que $\nu([6]) \geq \frac{21}{40}$. Visto que a g-função com a qual construímos tem simetria sob $x \mapsto 1 - x$, existe uma segunda g-medida dando peso no mínimo $\frac{21}{40}$ para o símbolo 3. Visto que estas são medidas de probabilidades, elas não podem ser iguais. Disto segue que existem duas g-medidas para esta g. Isto completa a prova do Teorema Principal. □

Apêndice

1.1 -Elementos de Dinamica Simbolica (Shift e Subshift)

Nosso objetivo nesta secção é dar um modelo rico em aplicações em sistemas dinâmicos.

Definição. $\Sigma_2 = \{x = (x_0 x_1 x_2 \cdots) : x_i = 0\, ou\, 1\}$.

Σ_2 é chamado o *espaço de seqüências* sobre os dois símbolos 0 e 1. Σ_2 também é denotado por $\Sigma_2 = \{0,1\}^{\mathbb{Z}^+}$. Mais geralmente, podemos considerar o espaço Σ_n consistindo de seqüências infinitas de inteiros entre 0 e $n-1$. Os elmentos de Σ_2 são strings infinitos de inteiros, como por exemplo os termos da seqüência $x_n = (\delta_{mn})_{m\geq 0} \quad (n = 0,1,2,\cdots)$, onde

$$\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

é o delta de Kronecker. Por exemplo,

$$x_0 = (\delta_{m0})_{m\geq 0} = (1,0,0,0,\cdots)$$

$$x_1 = (\delta_{m1})_{m\geq 0} = (0,1,0,0,\cdots)$$

etc,

Proposição. Σ_2 é não enumerável.

Prova. Suponha o contrário, $\Sigma_2 = \{(x_{1m})_{m\geq 1}, (x_{2m})_{m\geq 1}, \cdots\}$, ou seja enumeramos os elementos de Σ_2. Agora construímos a seguinte seqüência

$$y = (y_n), \quad onde \quad y_n = \begin{cases} 1, & x_{nn} = 0 \\ 0, & x_{nn} = 1 \end{cases}$$

obviamente $y \in \Sigma_2$ e $y \neq (x_{nm})$ para $n = 1,2,\cdots$. □

Podemos tornar Σ_2 um espaço métrico como segue. Para duas seqüências $x = (x_0 x_1 x_2 \cdots)$ e $y = (y_0 y_1 y_2 \cdots)$, defina a distância entre ambas por

$$d(x,y) = \sum_{i=0}^{\infty} \frac{|x_i - y_i|}{2^i}$$

Visto que $|x_i - y_i|$ é 0 ou 1, a série infinita é dominada pela série geométrica

$$\sum_{i=0}^{\infty} \frac{1}{2^i} = 2$$

e sendo assim converge.

Por exemplo, se $x = (000\cdots)$ e $y = (111\cdots)$, então $d(x, y) = 2$. Se $z = (1010\cdots)$, então

$$d(x, z) = \sum_{i=0}^{\infty} \frac{1}{2^{2i}} = \frac{1}{1 - \frac{1}{4}} = \frac{4}{3}.$$

Ainda, considerando novamente a seqüência $x_n = (\delta_{mn})_{m \geq 0}$ $(n = 0, 1, 2, \cdots)$, é fácil ver que

$$d(x_m, x_n) = \frac{1}{2^m} + \frac{1}{2^n}, \qquad m \neq n$$

a propósito, a seqüência em questão é de Cauchy.

Proposição. d é uma métrica sobre Σ_2.

Prova: Claramente, $d(x, y) \geq 0$ para quaisquer $x, y \in \Sigma_2$, e $d(x, y) = 0$ sse $x_i = y_i$ para todo i. Visto que $|x_i - y_i| = |y_i - x_i|$, disto segue que $d(x, y) = d(y, x)$. Finalmente, se x, y e $z \in \Sigma_2$, então $|x_i - y_i| + |y_i - z_i| \geq |x_i - z_i|$ do que se deduz que $d(x, y) + d(y, z) \geq d(x, z)$.

A métrica d acima permite decidir quais subconjuntos de Σ_2 são abertos e quais são fechados, bem como quando duas seqüências estão próximas uma da outra.

Proposição. Sejam $x, y \in \Sigma_2$ e suponha $x_i = y_i$ para $i = 0, 1, \cdots, n$. Então $d(x, y) \leq 1/2^n$. Reciprocamente, se $d(x, y) < 1/2^n$, então $x_i = y_i$ para $i \leq n$.

Prova; Se $x_i = y_i$ para $i \leq n$, então

$$d(x, y) = \sum_{i=0}^{n} \frac{|x_i - x_i|}{2^i} + \sum_{i=n+1}^{\infty} \frac{|x_i - y_i|}{2^i}$$

$$\leq \sum_{i=n+1}^{\infty} \frac{1}{2^i} = \frac{1}{2^n}.$$

Por outro lado, se $x_i \neq y_i$ para algum $j \leq n$, devemos ter

$$d(x, y) = \sum_{i=0}^{\infty} \frac{|x_i - y_i|}{2^i} \geq \sum_{i=0}^{n} \frac{|x_i - y_i|}{2^i} \geq \frac{1}{2^j} \geq \frac{1}{2^n}$$

conseqüêntemente, se $d(x, y) < 1/2^n$, então $x_i = y_i$ para $i \leq n$. □

A importância deste resultado é que podemos decidir de imediato quando ou não duas seqüências estão próximas uma da outra. Intuitivamente, este resultado diz que duas seqüências em Σ_2 estão próximas se suas "primeiras"entradas concordam. Em particular a bola aberta de centro x e raio $\varepsilon = 1/2^n$ será denotada por

$$[x]^n = \{y \in \Sigma_2 : d(x,y) < \frac{1}{2^n}\}$$

Este é o conjunto de seqüências em Σ_2 cujas entradas concordam com as de x até a posição n, inclusive.

Agora definiremos o conceito mais importante na dinâmica simbólica, a aplicação shift sobre Σ_2.

Definição. A aplicação shift $\sigma: \Sigma_2 \to \Sigma_2$ é dada por

$$\sigma(x_0 x_1 x_2 \cdots) = (x_1 x_2 x_3 \cdots).$$

A aplicação shift "deleta"a primeira entrada em uma seqüência, e desloca todas as outras entradas de uma posição para a esquerda. Claramente, σ é uma aplicação 2-para-1 de Σ_2, por exemplo x_0 pode ser 0 ou 1. Outrossim, na métrica definida acima, σ é uma aplicação contínua.

Proposição. $\sigma: \Sigma_2 \to \Sigma_2$ é contínua.

Prova: Seja $\varepsilon > 0$ e $x = x_0 x_1 x_2 \cdots$ Tome n de modo que $1/2^n < \varepsilon$. Seja $\delta = 1/2^{n+1}$. Se $y = y_0 y_1 y_2 \cdots$ satisfaz $d(x,y) < \delta$, então pela proposição anterior temos $x_i = y_i$ para $i \leq n+1$. Sendo assim as i-ésimas entradas de $\sigma(x)$ e $\sigma(y)$ concordam para $i \leq n$. Isto é $d(\sigma(x), \sigma(y)) \leq 1/2^n < \varepsilon$. $\square$

Subshift de tipo finito.

Agora definiremos o shift sobre r símbolos. Seja Σ_r o conjunto de todas as possíveis seqüências de números naturais entre 1 e r, isto é,

$$\Sigma_r = \{x = (x_0 x_1 x_2 \cdots) : x_i \in \mathbb{Z} \ , 1 \leq x_i \leq r\}.$$

Como anteriormente, existe uma métrica natural sobre Σ_r definida por

$$d_r(x,y) = \sum_{i=0}^{\infty} \frac{|x_i - y_i|}{r^i}$$

onde $x = (x_0x_1x_2\cdots)$ e $y = (y_0y_1y_2\cdots)$. A prova da seguinte proposição é similar à feita anteriormente e por essa razão é deixada como exercício.

Proposição.

(a) d_r é uma métrica sobre Σ_r

(b) Se $x_i = y_i$ para $i = 0, \cdots, k$, então $d_r(x, y) \leq 1/r^k$.

(c) Se $d_r(x, y) < 1/r^k$, então $x_i = y_i$ para $i \leq k$.

Igualmente ao caso de Σ_2, existe uma aplicação dada por

$$\sigma(x_0x_1x_2\cdots) = (x_1x_2x_3\cdots)$$

O mesmo raciocínio feito para Σ_2 pode ser feito aqui para mostrar que σ é contínua.

Nosso objetivo é descrever certos subconjuntos de Σ_r que surgem naturalmente em sistemas dinâmicos. Seja A uma matriz $r \times r$ cujas entradas denotadas por a_{ij} são 0 ou 1. Isto é A é um arranjo quadrado de 0's e 1's. A é chamada *matriz de transição* para o sistema. Devemos usar A para descrever quais seqüências em Σ_r pertencem a um subconjunto que será denotado por Σ_A. A seqüência $x = (x_0x_1x_2\cdots)$ mora em Σ_A se obedece à seguinte regra: cada par adjacente de entradas da seqüência x determina uma locação na matriz A, a entrada $a_{x_ix_{i+1}}$. A seqüência mora em Σ_A se e sòmente se todas tais entradas são 1. Mais concisamente,

$$\Sigma_A = \{x = (x_0x_1x_2\cdots) : a_{x_ix_{i+1}} = 1 \text{ para todo } i\}.$$

Isto é, usamos a matriz de transição como contrôle de certos pares de entradas de seqüências que moram em Σ_A.

Exemplos.

(a) Seja

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Visto que $a_{12} = a_{21} = 0$, disto segue que 1 e 2 nunca podem ser adjacentes em um elemento de Σ_A. Sendo assim, existem apenas duas seqüências permitidas em Σ_A, as seqüências constantes $(111\cdots)$ e $(222\cdots)$.

(b) Seja

$$\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Neste exemplo, 2 pode seguir 1, mas não o contrário. Deste modo, Σ_A consiste das seqüências constantes mais qualquer seqüência da forma $(11\cdots11222\cdots)$ onde existe um número arbitrário 1's seguido por um número arbitrário de 2's.

(c) Seja

$$\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}.$$

Quaisquer combinações de 1's e 2's sãos permitidas em uma seqüência em Σ_A, exceto um par de 2's adjacentes.

Agora denotamos por σ_A a restrição de σ ao conjunto Σ_A. A seguinte proposição garante que isto faz sentido.

Proposição. Σ_A é um subconjunto fechado de Σ_r que é invariante sob σ_A.

Prova: A invariância é clara. Para provar que Σ_A é fechado, suponha que x_n é uma seqüência de elementos de Σ_A, i.e., uma seqüência de seqüências, que converge para y. Se $y \notin \Sigma_A$, existe um menor inteiro α para o qual $a_{y_\alpha y_{\alpha+1}} = 0$.

Visto que x_n converge para y, existe um outro inteiro K tal que, se $i > K$, então $d_r(x_n, y) < 1/r^{\alpha+1}$. Pela proposição anterior, isto força $y_0, y_1, \cdots, y_{\alpha+1}$ a concordar com as correspondentes entradas de x_n para $n \geq K$. Em particular, devemos ter $a_{y_\alpha y_{\alpha+1}} = 1$, visto que $x_n \in \Sigma_A$. Esta contradição estabelece o resultado. □

Chamamos σ_A um *um subshift de tipo finito*, visto que êle é determinado por um número finito de condições impostas pela matriz de transição A. Existem subshifts que não são do tipo finito, mas não serão abordados neste trabalho.

1.2 -Variáveis Randômicas e Processos Estocásticos

Vamos fazer um resumo sobre variáveis randômicas e processos estocásticos. Inicialmente lembremos as seguinte definições da teoria das probabilidades :

1. Sejam A, B eventos, isto é subconjuntos mensuráveis de (Ω, P). Então definimos a **probabilidade condicional** de A dado que B ocorreu por

$$P(A \backslash B) = \frac{P(A \cap B)}{P(B)}$$

Observe que a definição só é válida para $P(B) > 0$.

2. Dois eventos A e B são ditos independentes se

$$P(A \cap B) = P(A)P(B)$$

Exemplo: Defina um espaço de probabilidade $(\Omega, \mathcal{A}, \mathbf{P})$ tomando $\Omega =]0, 1[$, $\mathcal{A} = \mathcal{L}$, a σ-álgebra dos subconjuntos de Ω mensuráveis Lebesgue, e $\mathbf{P}$ a medida de Lebesgue. Seja a expansão binária de $\omega \in \Omega$

$$\omega = \sum_{n=1}^{\infty} \omega_n 2^{-n},$$

onde $\omega_n = 0$ ou 1. Se m e n são inteiros distintos, e a e b são iguais a 0 ou 1, então os eventos $\{\omega : \omega_m = a\}$ e $\{\omega : \omega_n = b\}$ são independentes.

- Seja $(\Omega, \mathcal{A}, \mathbf{P})$ um espaço de probabilidade, uma coleção de eventos $\mathcal{C} \subset \mathcal{A}$ é dita independente se, para qualquer subcoleção finita

$$\{E_1, E_2, \cdots, E_n\} \subset \mathcal{C},$$

$$\mathbf{P}\{\cap_{j=1}^{n}\} = \prod_{j=1}^{n} \mathbf{P}(E_j).$$

-Variáveis randômicas e funções mensuráveis.

Seja $(\Omega, \mathcal{A}, \mathbf{P})$ um espaço de probabilidade. Um ponto ω do conjunto Ω representa um "resultado" de um "experimento randômico". Os elementos da σ-álgebra $\mathcal{A}$ representam eventos relativo aos quais podemos atribuir uma probabilidade. A probabilidade de $A \in \mathcal{A}$, sendo $\mathbf{P}(A)$, e $\mathbf{P}$ uma medida de probabilidade sobre $(\Omega, \mathcal{A})$.

Para tomar um exemplo muito simples, suponha que uma moeda honesta é lançada três vezes. Este é um experimento que tem oito resultados possíveis:

$$\omega_1 = CCC, \quad \omega_2 = CCK, \quad \omega_3 = CKC, \omega_4 = CKK,$$

$$\omega_5 = KCC, \quad \omega_6 = KCK, \quad \omega_7 = KKC, \omega_8 = KKK.$$

Cada resultado tem probabilidade $\frac{1}{8}$. Deste modo

$$\Omega = \{\omega_1, \omega_2, \cdots, \omega_8\},$$

$\mathcal{A}$ é a classe de todos os subconjuntos de Ω, e $\mathbf{P}$ atribui a um conjunto A contendo r elementos a probabilidade $\frac{1}{8}r$.

Podemos estar interessados não no resultado do experimento em si, mas digamos, no número X de caras. Esta é uma quantidade que pode tomar qualquer um dos valôres, $0, 1, 2, 3$, de acordo com o resultado do experimento. Isto é um exemplo do que é chamado uma *variável randômica.* X é uma função assinalando para cada possível resultado ω um número $X(\omega)$, o número de $C's$ em ω. Mais específicamente, X é uma função de Ω para $\mathbb{R}$, dada por

$$X(\omega_1) = 3,$$

$$X(\omega_2) = X(\omega_3) = X(\omega_5) = 2,$$

$$X(\omega_4) = X(\omega_6) = X(\omega_7) = 1,$$

$$X(\omega_8) = 0.$$

Uma propriedade importante das variáveis randômicas é que afirmações de probabilidades podem ser feitas com respeito a elas. Por exemplo, no exemplo acima

$$\mathbf{P}\{\omega : X(\omega) = 2\} = \mathbf{P}\{\omega_2, \omega_3, \omega_5\} = \frac{3}{8},$$

afirmação que é usualmente abreviada por

$$\mathbf{P}\{X = 2\} = \frac{3}{8};$$

a probabilidade de dá duas caras em 3 lançamentos é $\frac{3}{8}$

Mais geralmente, se $(\Omega, \mathcal{A}, \mathbf{P})$ é um espaço de probabilidade, consideremos funções $X : \Omega \to \mathbb{R}^*$ (onde $\mathbb{R}^* = \mathbb{R} \cup \{-\infty, +\infty\}$) que assinalam para cada resultado ω, um número $X(\omega)$. Tais funções não podem, contudo, ser completamente arbitrárias. Isto porque desejamos tratar de probabilidade como por exemplo, "a probabilidade de que X esteja no intervalo (a, b)", ou simbólicamente

$$\mathbf{P}\{\omega : a < X < b\}$$

Sendo assim exigimos que conjuntos da forma

$$\{\omega : a < X < b\}$$

pertençam a $\mathcal{A}$, isto é, que $X:\Omega \to \mathbb{R}^*$ seja mensurável. Estas considerações levam à seguinte definição.

Definição(*variável randômica*). Uma variável randômica sobre um espaço de probabilidade $(\Omega, \mathcal{A}, \mathbf{P})$ é uma função mensurável de $(\Omega, \mathcal{A})$ para $\mathbb{R}^*$.

Muitas vezes é conveniente eliminar o símbolo ω das afirmações de probabilidade envolvendo variáveis randômicas, de maneira que, por exemplo, a afirmação

$$\mathbf{P}\{\omega : a < X(\omega) < b\} = p$$

pode ser escrito mais sucintamente como

$$\mathbf{P}\{\omega : a < X < b\} = p$$

De fato, devemos ver que muitas manipulações envolvendo variáveis randômicas podem (e amiúde são) se dar sem referência explícita ao espaço Ω compreendido.

Exercício. Seja Ω consistindo dos enteiros positivos, com

$$\mathbf{P}\{\omega\} = \frac{1}{2^\omega},$$

e seja $X(\omega)$ o resíduo de ω módulo k. Mostre que X é uma variável randômica, e ache

$$\mathbf{P}\{X = r\} \qquad (r = 0, 1, 2, \cdots, k-1).$$

-Distribuição de variáveis randômicas

Se uma variável randômica X sobre $(\Omega, \mathcal{A}, \mathbf{P})$ é finita com probabilidade 1, então

$$F(x) = \mathbf{P}\{\omega : X(\omega) \leq x\}$$

determina uma função $F:\mathbb{R} \to [0,1]$ a qual é não decrescente, contínua à direita, e satisfaz

$$\lim_{x\to-\infty} F(x) = 0, \qquad \lim_{x\to+\infty} F(x) = 1.$$

Para qualquer conjunto de Borel $B \subset \mathbb{R}$,

$$\mathbf{P}\{\omega : X(\omega) \in B\} = \int_B dF(x).$$

A função $F(x)$ é chamada *função distribuição* da variável randômica X. O exposto acima mostra que, do conhecimento da função distribuição de uma variável randômica, podemos deduzir a probabilidade de que a variável more em qualquer subconjunto de Borel de $\mathbb{R}$.

-Distribuição conjunta

Seja $(\Omega, \mathcal{A}, \mathbf{P})$ um espaço de probabilidade sobre o qual está definida uma coleção $\mathbf{X} = (X_1, X_2, \cdots, X_n)$ de variáveis randômicas. Devemos assumir por simplicidade que cada X_r é quase certamente finita, deste modo X_r é uma função mensurável de $(\Omega, \mathcal{A})$ para $(\mathbb{R}, \mathcal{B})$. Segue que a função $\mathbf{X}$ a qual leva $\omega \in \Omega$ para o ponto

$$(X_1(\omega), X_2(\omega), \cdots, X_n(\omega)) \in \mathbb{R}^n$$

é uma função mensurável de $(\Omega, \mathcal{A})$ para $(\mathbb{R}^n, \mathcal{B}^n)$.

A medida $\mathbf{P}$ induz uma nova medida $\mathbf{P}'$ sobre $(\mathbb{R}^n, \mathcal{B}^n)$ por

$$\mathbf{P}'(B) = \mathbf{P}\{\omega : \mathbf{X}(\omega) \in B\} \qquad (B \in \mathcal{B}^n).$$

Se $\mathcal{A}'$ é a completação de $\mathcal{B}^n$ com respeito a $\mathbf{P}'$, então $(\mathbb{R}^n, \mathcal{A}', \mathbf{P}')$ é um espaço de probabilidade. A medida $\mathbf{P}'$ é chamada **distribuição conjunta** da coleção $\mathbf{X}$.

-Distribuição marginal

Suponha que conhecemos uma distribuição conjunta de $\mathbf{X}_n = (X_1, X_2, \cdots, X_n)$, a qual por uma mudança de notação denotamos por $\mathbf{P}_n$. A fortiori conhecemos a distribuição $\mathbf{P}_{n-1}$ da subcoleção

$$\mathbf{X}_{n-1} = (X_1, X_2, \cdots, X_{n-1})$$

De fato, se $B \in \mathcal{B}^{n-1}$,

$$\mathbf{P}_{n-1}\{B\} = \mathbf{P}\{\mathbf{X}_{n-1} \in B\} = \mathbf{P}\{\mathbf{X}_n \in B \times \mathbb{R}\} = \mathbf{P}\{B \times \mathbb{R}\}.$$

$\mathbf{P}_{n-1}$ é chamada uma *distribuição marginal* obtida de $\mathbf{P}_n$.

Exercício. Sejam X, Y variáveis randômicas tomando valores sobre os inteiros, e seja

$$p_{ij} = \mathbf{P}(X = i, Y = j).$$

Mostre que as distribuições marginais de X e Y são dadas por

$$p_{i.} = \sum_j p_{ij} \ , \quad p_{.j} = \sum_i pij \ ,$$

respectivamente. Mostre que X e Y são independentes se e sòmente se existem números a_i, b_j de modo que, para todo i e j,

$$p_{ij} = a_i b_j$$

- **Variáveis randômicas independentes**

O conceito de independência de eventos pode ser extendido a variáveis randômicas.

Inicialmente vejamos como pode surgir a necessidade de tratarmos com "seqüências" de variáveis randômicas:

Em muitos problemas de probabilidade queremos considerar uma seqüência infinita de experimentos. Como pode parecer à primeira vista, uma tal seqüência infinita não é necessáriamente uma abstração de nossa imaginação. Suponha que consideremos o experimento escolher aleatoriamente um número entre 0 e 1 e tomar o n-ésimo dígito de sua expansão decimal (na base 2, por exemplo). Então, uma única escolha randômica de um número determina um resultado

A propósito, temos:

$$X_n(\omega) = f(T^{n-1}(\omega)) \qquad (n = 1, 2, \cdots)$$

onde:

$$T(\omega) = 2\omega \text{ (mod 1)}, \qquad f = \chi_{[1/2,1)}$$

X_n nos dá o n-ésimo dígito da expansão (na base 2) de um $\omega \in \Omega = [0, 1)$.

Suponha que $(\Omega, \mathcal{A}, \mathbf{P})$ é um espaço de probabilidade e que X e Y são duas variáveis randômicas definidas sobre este espaço. Para quaisquer números $x, y \in \mathbb{R}^*$, considere os dois eventos

$$\{\omega : X(\omega) \leq x\}, \quad \{\omega : Y(\omega) \leq y\}.$$

Se, para toda escolha de x, y, estes dois eventos são independentes, então X e Y são ditos independentes. Mais geralmente, podemos fazer a seguinte definição.

Independência

Uma coleção de variáveis randômicas $\{X_i : i \in I\}$ sobre o espaço de probabilidade $(\Omega, \mathcal{A}, \mathbf{P})$ é dito independente se, para qualquer $x_i \in \mathbb{R}^*(i \in I)$, a coleção de eventos

$$\{\omega : X_i(\omega) \leq x_i\} \qquad (i \in I),$$

é independente.

Se E é qualquer evento, então a função indicador χ_E definida por

$$\chi_E(\omega) = \begin{cases} 1, & \omega \in E \\ 0, & \omega \notin E \end{cases}$$

é claramente uma variável randômica. Outrossim, dois eventos A e B são independentes se e sòmente se χ_A e χ_B são variáveis randômicas independentes.

Seja $(\Omega^n, \mathcal{A}^n, \mathbf{P}^n)$ o produto de n cópias de $(\Omega, \mathcal{A}, \mathbf{P})$, sendo assim os eventos elementares são n-nuplas $(\omega_1, \omega_2, \cdots, \omega_n)$ de elementos de Ω, e se $X_1, X_2, \cdots, X_n$ são variáveis randômicas sobre Ω^n com a propriedade de que X_j é uma função de ω_j sòmente, então é fácil verificar que $X_1, X_2, \cdots, X_n$ são independentes.

Definição (Processos Estocásticos)

Um *processo estocástico* é uma família de variáveis randômicas sobre um mesmo espaço amostral Ω. Se os membros da família são enumeráveis, o processo deve ser denotado por $X_1, X_2, X_3, \cdots$ se são não enumeráveis, o processo deve ser denotado por $\{X_t : t \geq 0\}$ ou $\{X_t\}_{t \geq 0}$. No primeiro caso o processo é chamado um *processo de tempo-discreto*, enquanto no segundo caso êle é chamado um *processo de tempo contínuo.*

Definição(Espaço de Estados, Cadeia)

O conjunto de valôres distintos assumidos por um processo estocástico é chamado o *espaço de estados.* Se o espaço de estados de um processo estocástico é finito, ou enumerável, o processo deve ser chamado uma *cadeia.*

Um processo estocástico é ocasionalmente visto como uma função de duas variáveis $X_t(\omega) = X(t, \omega)$. Para t fixado a função é uma variável randômica. Para ω fixado, a função de t, resultante é chamada um *caminho amostral.*

Os conceitos acima colocados podem ocorrer nos mais simples experimentos,como

Exemplo (*lançamento de uma moeda*)

Seja $\{Y_i\}$, $i = 1, 2, \cdots, 6$, uma seqüência de variáveis randômicas independentes com $P\{Y_i = 1\} = P\{Y_i = -1\} = \frac{1}{2}$ para $i = 1, 2, \cdots, 6$. Lembramos que $\{Y_i = 1\} = \{\omega \in \Omega : Y_i = 1\}$. Defina $X_0 = 0$ e $X_n = \sum_{i=1}^{n} Y_i$. Então $X_n \equiv X_n(\omega) \equiv X(n, \omega)$ é um processo estocástico discreto no tempo com espaço de estados $S = \{-6, -5, \cdots, -1, 0, 1, \cdots, 6\}$. Este processo estocástico aparece naturalmente do seguinte experimento: uma moeda honesta é lançada seis vezes, um jogador ganha um real quando uma cara (C) aparece e perde um real quando uma coroa (K) aparece. Logo X_n denota o ganho obtido na brincadeira no tempo n. Um típico resultado ω para este experimento é $\omega^* = (C, C, K, C, K, C)$. O correspondente caminho amostral para este resultado ω^* é dado na figura seguinte (o gráfico consiste dos vértices da poligonal, esta foi incluída para facilitar a interpretação).

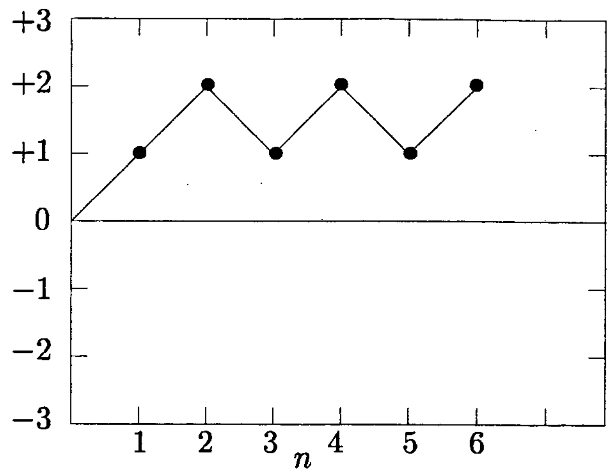

Figura 4.1: Um típico caminho amostral para $X(n, \omega)$

- Observe que durante a brincadeira qualquer dos jogadôres pode desde perder 6 reais até ganhar 6 reais; isto é o que nos diz o espaço de estados $S = \{-6, -5, \cdots, -1, 0, 1, \cdots, 6\}$.

O espaço amostral $\Omega = \{\omega : \omega = (\omega_1, \cdots, \omega_6), \omega_i \in \{C, K\}\}$ para este experimento consiste de $2^6 = 64$ combinações possíveis de cara e coroa nos seis lançamentos.

Se quiséssemos obter tôdos os caminhos amostrais deste experimento, ou mesmo o espaço amostral, teríamos certamente de tratar com combinações. Forneceremos agora uma fórmula bastante útil que nos permite obter combinações:

Dado um conjunto com n elementos, $A = \{a_1, a_2, \cdots, a_n\}$, o número de subconjuntos do mesmo é 2^n. Mais ainda, o número de combinações dos n elementos tomados r a r, é:

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

Mas esta fórmula não nos diz **quais** são estas combinações. A fórmula seguinte tem precisamente esta finalidade:

$$\boxed{a_{ij} = (-1)^{\left[\frac{i-1}{2^{j-1}}\right]}} \tag{4.4}$$

Onde: $[x]$ é o maior inteiro que não supera x.

As combinações de um conjunto com N elementos (2^N subconjuntos) são obtidas apartir da matriz de ordem $2^N \times N$ calculada pela fórmula (4.4). Vamos tomar como exemplo um conjunto com $N = 3$ elementos, $A = \{a_1, a_2, a_3\}$. Tôdas as combinações possíveis podem ser obtidas da seguinte matriz de ordem $2^3 \times 3$:

1	1	1
-1	1	1
1	-1	1
-1	-1	1
1	1	-1
-1	1	-1
1	-1	-1
-1	-1	-1

Agora convencionamos que $+1$ significa que o elemento entra na combinação e que -1 significa que o elemento não entra na combinação. Desta forma teremos a seguinte tabela:

a_1	a_2	a_3	$\{a_i\}$	$\binom{n}{r}$
1	1	1	$\{a_1, a_2, a_3\}$	$\binom{3}{2}$
-1	1	1	$\{a_2, a_3\}$	$\binom{3}{2}$
1	-1	1	$\{a_1, a_3\}$	$\binom{3}{2}$
-1	-1	1	$\{a_3\}$	$\binom{3}{1}$
1	1	-1	$\{a_1, a_2\}$	$\binom{3}{2}$
-1	1	-1	$\{a_2\}$	$\binom{3}{1}$
1	-1	-1	$\{a_1\}$	$\binom{3}{1}$
-1	-1	-1	$\emptyset$	$\binom{3}{0}$

Obs: Com a fórmula acima podemos obter programas computacionais bastante compactos para o cálculo de combinações.

1.2.1 Processos de Markov

Considere novamente o exemplo da moeda, onde agora consideramos o domínio do parâmetro n os naturais (evidentemente a brincadeira agora só pode acontecer entre os anjos).

X_n denota, como já vimos, o saldo no tempo n. Seja por exemplo, $X_{n-1} = 10$. É fácil ver que X_n deve ser 9 ou 11 visto que $P\{Y_n = 1\} = \frac{1}{2} = P\{Y_n = -1\}$. De fato, isto pode ser expressado em termos de probabilidade condicional como,

$$P\{X_n = 9/X_{n-1} = 10\} = \frac{1}{2} \quad \text{e} \quad P\{X_n = 11/X_{n-1} = 10\} = \frac{1}{2}$$

Outrossim o ganho total no tempo n depende sòmente de X_{n-1}, o ganho total no tempo $n-1$, e o valor de Y_n, no tempo n. Os valôres de $X_{n-2}, X_{n-3}, \cdots, X_1$ não afetam o valor de X_n. Expressando isto em termos de probabilidade condicional, temos

$$P\{X_n = 9/X_{n-1} = 10\} = P\{X_n = 9/X_{n-1} = 10, X_{n-2} = i_{n-2}, \cdots, X_1 = i_1\}$$

Este é um exemplo de um processo estocástico dito processo de Markov.

Definição (*Processo de Markov*)

Um processo estocástico é dito Markov se a seguinte igualdade é válida:

$$P\{X_n = i_n/X_{n-1} = i_{n-1}, X_{n-2} = i_{n-2}, \cdots, X_1 = i_1\} = P\{X_n = i_n/X_{n-1} = i_{n-1}\}$$

isto é a probabilidade do resultado atual do processo ser i_n depende apenas do estado anterior do processo.

Vejamos mais um exemplo de um processo markoviano: Consideremos um rato R que se move em um reticulado como na figura

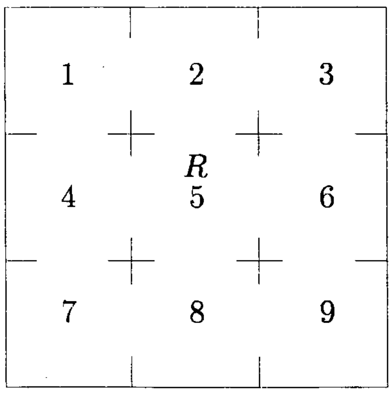

Figura 4.2: Reticulado

e que a cada $n = 1, 2, \cdots$ segundos o rato se move. Seja X_n=número do compartimento ocupado pelo rato no tempo n. $X_n \in \{1, 2, \cdots, 9\}$. É fácil concluir que este processo estocástico é Markov.

- Repetiremos aqui uma abordagem de processos estocásticos, ligeiramente diferente da anterior, e mais adequada aos nossos propósitos que se seguem:

Processos estocásticos

Seja $(\Omega, \mathcal{A}, \mathbf{P})$ um espaço de probabilidade, e T qualquer conjunto, uma coleção $\{X(t) : t \in T\}$ de variáveis randômicas sobre Ω é chamado um processo estocástico com conjunto de parâmetros T.

Em muitas aplicações, como na seção prévia, t representa o parâmetro tempo, deste modo T é algum subconjunto da linha real. Este nem sempre é o caso, contudo; existem importantes exemplos nos quais T tem uma estrutura mais complexa.

Seja $\{X(t) : t \in T\}$ um processo estocástico sobre Ω, então $X(t)$, sendo uma variável randômica sobre Ω, tem uma distribuição de probabilidade. Mais geralmente, se $t_1, t_2, \cdots, t_n$ são elementos distintos de T, as variáveis randômicas

$$X(t_1), X(t_2), \cdots, X(t_n)$$

devem ter uma distribuição conjunta, a qual deve ser uma medida de probabilidade

$$\mathbf{P}_{t_1 t_2 \cdots t_n}$$

sobre $\mathbb{R}^n$. Para diferentes valôres de $n, t_1, t_2, \cdots, t_n$ estas medidas são chamadas *distribuições finito-dimensional* do processo estocástico.

-Processos estacionários

Seja $X(t)$ um processo estocástico cujo conjunto de parâmetros T é um subconjunto da linha real. Então t é uma variável real, é intuitivamente conveniente pensar assim, como de fato se dá em muitas aplicações, como medição de tempo. Pode ser que, conquanto $X(t)$ varie com t, o mesmo não acontece com sua distribuição de probabilidade. Mais geralmente, a distribuição conjunta de $X(t_1), X(t_2), \cdots, X(t_n)$ pode ter dependencia sòmente sobre as diferenças $t_j - t_1$. Esta é uma propriedade que é possuída por muitos processos os quais representam sistemas de algum modo em equilíbrio estatístico, e é conhecida como estacionária.

Processos estacionários

Um processo estocástico $X(t)$ com conjunto de parâmetros $T \subset \mathbb{R}$ é dito estacionário se sempre que $t_1, t_2, \cdots, t_n, t_1 + h, t_2 + h, \cdots, t_n + h$ todos pertencem a T, suas distribuições finito-dimensionais satisfazem

$$\mathbf{P}_{t_1+h,t_2+h,\cdots,t_n+h} = \mathbf{P}_{t_1t_2\cdots t_n}$$

Seja $\{X_1, X_2, \cdots\}$ uma seqüência de variáveis randômicas, então, isto induz uma medida de probabilidade $\mathbf{P}'$ sobre $(\mathbb{R}^T, \mathcal{B}^T)$ (onde $T = \{1, 2, \cdots\}$) por

$$\mathbf{P}'\{B\} = \mathbf{P}\{\{X_1, X_2, \cdots\} \in B\}$$

Considere agora $S: \mathbb{R}^T \to \mathbb{R}^T$ uma função definida por

$$S\{x_1, x_2, \cdots\} = \{x_2, x_3, \cdots\}$$

S é chamada a *função shift*. Então S e $\mathbf{P}'$ têm uma importante propriedade. Para ver isto, seja $B \in \mathcal{B}^T$ um cilindro, isto é, um conjunto da forma

$$B = \{x \in \mathbb{R}^T : (x_1, x_2, \cdots, x_n) \in A\}$$

para $A \in \mathcal{B}^n$. Então

$$\begin{aligned}
\mathbf{P}'\{S^{-1}B\} &= \mathbf{P}\{\{X_1, X_2, \cdots\} \in S^{-1}B\} \\
&= \mathbf{P}\{S\{X_1, X_2, \cdots\} \in B\} \\
&= \mathbf{P}\{\{X_2, X_3, \cdots\} \in B\} \\
&= \mathbf{P}\{(X_2, X_3, \cdots, X_{n+1}) \in A\} \\
&= \mathbf{P}_{2,3,\cdots,n+1}(A) \\
&= \mathbf{P}_{1,2,\cdots,n}(A) \\
&= \mathbf{P}\{(X_1, X_2, \cdots, X_n) \in A\} \\
&= \mathbf{P}\{\{X_1, X_2, \cdots\} \in B\} \\
&= \mathbf{P}'\{B\}
\end{aligned}$$

Deste modo a medida de probabilidade $\mathbf{P}'S^{-1}$ coincide com $\mathbf{P}'$ sobre todos os cilindros, e conseqüentemente sobre a σ-álgebra $\mathcal{B}^T$ gerada pelos cilindros. Sendo assim temos provado o seguinte resultado

Teorema. Seja uma seqüência estacionária

$$\{X(t) : t \in T = \{1, 2, \cdots\}\} = \{X_1, X_2, \cdots\}$$

a qual induz uma medida $\mathbf{P}'$ sobre $(\mathbb{R}^T, \mathcal{B}^T)$. Então a função shift S definida anteriormente é uma transformação que *preserva medida* de $(\mathbb{R}^T, \mathcal{B}^T, \mathbf{P}')$ sobre si mesmo.

1.3-**Preliminares Para o Entendimento da Definição de uma g-Medida**

Definição (*Homeomorfismo Local*). Sejam X, Y espaços topológicos. Uma aplicação $T\colon X \to Y$ chama-se um *homeomorfismo local* quando cada ponto $x \in X$ está contido num aberto $\mathcal{U}$ tal que $\mathcal{V} = T(\mathcal{U})$ é aberto em Y e $T\big|_{\mathcal{U}}$ é um homeomorfismo de $\mathcal{U}$ sobre $\mathcal{V}$.

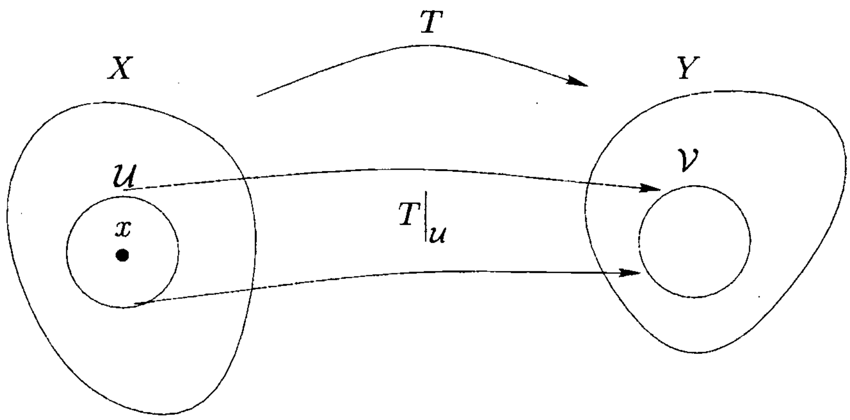

Figura 4.3: Homeomorfismo Local

Obs: para um exmplo de homeomorfismo local ver exmplo 1 (pg-87)

Proposição 1. Se $T\colon X \to Y$ é uma aplicação contínua localmente injetiva (em particular, um homeomorfismo local) então a imagem inversa $T^{-1}(\{y\})$ de cada ponto $y \in Y$ é um subconjunto discreto de X.

Prova: Cada ponto $x \in T^{-1}(\{y\})$ possui uma vizinhança $\mathcal{U}$, na qual ele é o único ponto que se aplica por T em y. então $\mathcal{U} \cap T^{-1}(\{y\}) = x$. Ou seja, todo ponto $x \in T^{-1}(\{y\})$ é isolado em $T^{-1}(\{y\})$.

corolário. Sejam X compacto, Y um espaço de Hausdorff e $T : X \to Y$ localmente injetiva e contínua. Então $T^{-1}(\{y\})$ é finito para cada $y \in Y$.

Prova: Num espaço de Hausdorff, o conjunto $\{x\} \subset X$ é fechado em X. Como a pré-imagem de um fechado, por uma transformação contínua é fechado, temos que $T^{-1}(\{y\})$ é fechado em

X. Como X é compacto segue-se que $T^{-1}(\{y\})$ é finito- uma vez que não podemos ter um conjunto discreto e infinito que seja compacto (pois se M é um espaço discreto e infinito ($\mathbb{N}$ e $\mathbb{Z}$, por exemplo) seus pontos constituem uma cobertura aberta infinita que não admite subcobertura própria, logo não se pode extrair dela uma subcobertura finita).

Obs: Embora para cada $y \in Y$, $T^{-1}(\{y\})$ seja finito, isto não implica que para diferentes $y_1, y_2 \in Y$ tenhamos o mesmo número de elementos nas respectivas pré-imagens, isto é $\# T^{-1}(\{y_1\}) \neq \# T^{-1}(\{y_2\})$ em geral.

Definição (*Espaços de Recobrimento*). Uma aplicação $p : \tilde{X} \to X$ chama-se ***aplicação de recobrimento***(ou, simplesmente, um *recobrimento*) quando cada ponto $x \in X$ pertence a um aberto $\mathcal{V} \subset X$ tal que

$$p^{-1}(\mathcal{V}) = \bigcup_{\alpha} \mathcal{U}_\alpha$$

é uma reunião de abertos $\mathcal{U}_\alpha$, dois a dois disjuntos, cada um dos quais se aplica homeomorficamente sobre $\mathcal{V}$ (isto é, existe um homeomorfismo $p' : p\big|_{\mathcal{U}_\alpha} \to \mathcal{V}$)

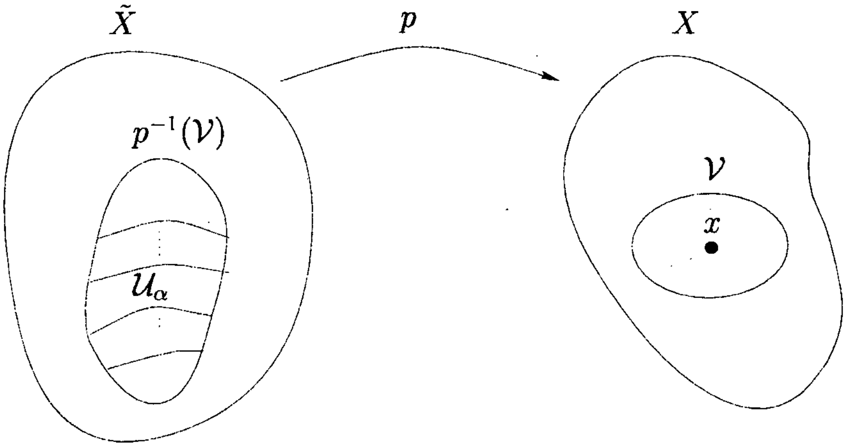

Figura 4.4: Aplicação de recobrimento

-Cada aberto $\mathcal{V}$ desse tipo chama-se uma *vizinhança distinguida.* O espaço $\tilde{X}$ chama-se um *espaço de recobrimento* de X e, para cada $x \in X$, o conjunto $p^{-1}(\{x\})$ chama-se a *fibra* sobre X. Às vezes, X chama-se a *base*.

Obs: Uma aplicação de recobrimento $p : \tilde{X} \to X$ é um homeomorfismo local de $\tilde{X}$ sobre X (nem todo homeomorfismo local é uma aplicação de recobrimento).

Exemplos:

(1) Considere a aplicação $p\colon \mathbb{R} \to S^1$ da linha real sobre o círculo

$$p(t) = e^{2\pi i t} = (cos 2\pi t, sen 2\pi t), \qquad t \in \mathbb{R}$$

Então p é um recobrimento de S^1. Qualquer intervalo aberto próprio ou arco sobre S^1 pode servir como uma vizinhança distinguida. Para o particular ponto 1 em S^1, seja $\mathcal{V}$ denotando o intervalo aberto direito sobre S^1 de $-i$ a i. Então

$$p^{-1}(\mathcal{V}) = \bigcup_{-\infty}^{+\infty}\left(n - \frac{1}{4}, n + \frac{1}{4}\right)$$

onde os abertos $\mathcal{U}_\alpha$ componentes de $p^{-1}(\mathcal{V})$ são os intervalo reais $(n - \frac{1}{4}, n + \frac{1}{4})$. Note que p aplica cada um desses intervalos homeomorficamente sobre $\mathcal{V}$, como ilustrado na figura.

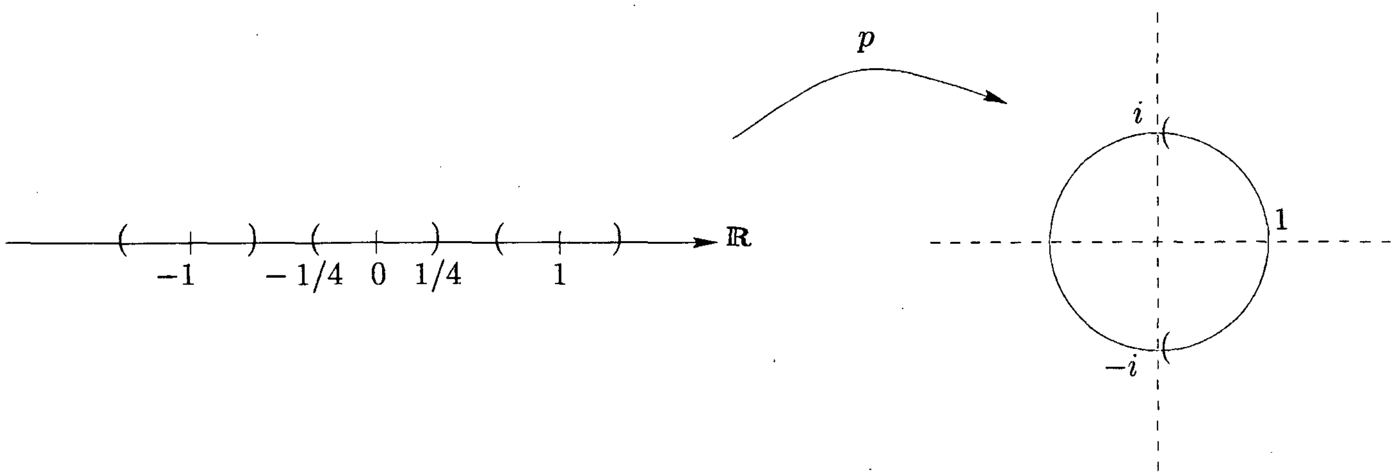

Figura 4.5: Recobrimento

Resolvendo o sistema trigonométrico abaixo

$$p(t) = (\cos 2\pi t, \sin 2\pi t) = (1, 0)$$

temos,

$$p^{-1}((1,0)) = \{x \in \mathbb{R} \; : x = 2k\pi \; , \; k \in \mathbb{Z} \; \}$$

Obs: Como p é um homeomorfismo local, pela proposição 1 $p^{-1}(\{y\})$ é discreto em $X = \mathbb{R}$ para todo $y \in S^1$. Ainda, como $\mathbb{R}$ não é compacto $p^{-1}(\{y\})$ não é finito, como se constata no exemplo em consideração.

(2) Para qualquer inteiro positivo n, seja $p_n : S^1 \to S^1$ a aplicação definida por $p_n(z) = z^n$, $z \in S^1$, onde z^n é a n-ésima potência do número complexo z. Então (S^1, p_n) é um espaço de cobertura de S^1. Representando o círculo em coordenadas polares, a ação de p_n é descrita como segue: p_n toma qualquer ponto $(1, \theta)$ e o aplica em $(1, n\theta)$. Seja $\mathcal{V}$ um intervalo aberto sobre S^1 subentendido por um ângulo θ $(0 \leq \theta \leq 2\pi)$, e contendo um ponto x. Então $p^{-1}(\mathcal{V})$ consiste de n intervalos abertos cada um determinando um ângulo θ/n e cada contendo uma raíz n-ésima de x. Estes n intervalos são as componentes $\mathcal{U}_\alpha$ de $p^{-1}(\mathcal{V})$ e cada um é mapeado por q_n homeomorficamente sobre $\mathcal{V}$.

Proposição 2. Se a base X de um recobrimento $p: \tilde{X} \to X$ é conexa então todas as fibras $p^{-1}(\{x\}), x \in X$, possuem o mesmo número cardinal, que se chama o *número de folhas* do recobrimento.

Prova: Para todos os pontos x de uma vizinhança distinguida $\mathcal{V}$, o número cardinal da fibra $p^{-1}(\{x\})$ é o mesmo. Segue-se que o conjunto dos pontos $x \in X$ tais que $p^{-1}(\{x\})$ tem um número cardinal dado é aberto. Isto determina uma decomposição de X como reunião de abertos disjuntos, em cada um dos quais o cardinal de $p^{-1}(\{x\})$ é constante. Como X é conexo só pode existir um desses abertos.

Obs: Pelo corolário da proposição 1, quando $\tilde{X}$ é compacto e X é conexo de Hausdorff, toda aplicação de recobrimento $p: \tilde{X} \to X$ tem um número finito de folhas. Neste caso, $X = p(\tilde{X})$ é necessáriamente compacto.

- Caracterizaremos os recobrimentos com um número finito de folhas:

Definição: Uma aplicação $T: X \to Y$ diz-se *fechada* quando a imagem $T(F)$ de todo subconjunto $F \subset X$ é um subconjunto fechado de Y.

Definição: Uma aplicação contínua $T: X \to Y$ chama-se *própria* quando é fechada e, para todo $y \in Y$, a imagem inversa $T^{-1}(\{y\})$ é compacta.

(Por exemplo, se X é compacto e Y é de Hausdorff, toda aplicação contínua $T: X \to Y$ é própria)

Proposição 3. Sejam X um espaço de Hausdorff e $T : X \to Y$ um homeomorfismo local. Cada uma das afirmações abaixo implica a seguinte:

(1) Existe $n \in \mathbb{N}$ tal que cada imagem inversa $T^{-1}(\{y\}), y \in Y$, possui n elementos;

(2) T é própria e sobrejetiva;

(3) T é uma aplicação de recobrimento cujas fibras $T^{-1}(\{y\})$ são finitas.

Se Y for conexo então as três afirmações são equivalentes.

Prova: [9]

Obs: Se X é compacto toda aplicação contínua $T: X \to Y$ é fechada.

Conclusão: Se $T: X \to Y$ é um homeomorfismo local $(T^{-1}(\{y\})$ é discreto em $X)$ e se X é compacto e Y é de Hausdorff $(T^{-1}(\{y\})$ é finito) e se, em adição Y for conexo, então as afirmações da proposição anterior se equivalem.

Observe que T é uma aplicação de recobrimento.

- Continuando os preliminares para o entendimento da definição de g-medida, temos:

Recordamos: se $(A_n)_{1,2,\ldots}$ é uma seqüência de subconjuntos de X, seja $E_0 = \emptyset$ e para todo $n \in \mathbb{N}$, seja

$$E_n = \bigcup_{k=1}^{n} A_k \quad , \quad F_n = A_n / E_{n-1}$$

então: (E_n) é uma seqüência monótona crescente e (F_n) é uma seqüência disjunta tais que

$$\bigcup_{n=1}^{\infty} E_n = \bigcup_{n=1}^{\infty} F_n = \bigcup_{n=1}^{\infty} A_n$$

Observemos que $F_n \subset A_n$, $\forall n \in \mathbb{N}$

Seja $(X, \beta(X), \mu)$ um espaço de medida onde X é compacto e $T: X \to X$ é uma aplicação de cobertura n-para-1, (ou seja, o par (X, T) é um espaço de cobertura do próprio espaço X)

Da definição de transformação de recobrimento, temos

$$T^{-1}(\mathcal{V}_\alpha) = \bigcup_{i=1}^{n} \mathcal{U}_{\alpha_i}$$

Pois T é $n - para - 1$.

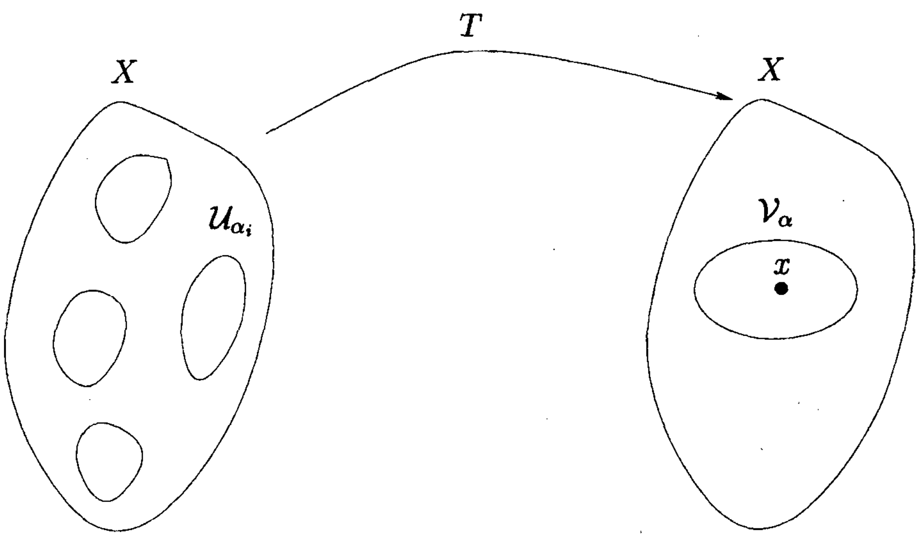

Figura 4.6: Transformação de recobrimento

Obs: α percorre um conjunto de indíces o suficiente para cobrir todo o espaço X. A cada $x \in X$, temos uma vizinhança $\mathcal{V}_\alpha$ de x. Isto é

$$X = \bigcup_\alpha \mathcal{V}_\alpha$$

Obs: para cada α a pré-imagem se constitui de n abertos disjuntos em X, cada um dos quais se aplica homeomorficamente a $\mathcal{V}_\alpha$, isto é

$$T^{-1}(\mathcal{V}_\alpha) = \bigcup_{i=1}^{n} \mathcal{U}_{\alpha_i}$$

Sendo $X = \bigcup_\alpha \mathcal{V}_\alpha$, temos

$$T^{-1}(X) = X = T^{-1}(\bigcup_\alpha \mathcal{V}_\alpha) = \bigcup_\alpha T^{-1}(\mathcal{V}_\alpha) = \bigcup_\alpha (\bigcup_{i=1}^{n} \mathcal{U}_{\alpha_i})$$

Então:

$$X = \bigcup_\alpha (\bigcup_{i=1}^{n} \mathcal{U}_{\alpha_i})$$

a princípio, é claro, esta cobertura aberta é qualquer, como X é compacto, podemos obter uma cobertura finita, isto é

$$X = \bigcup_{\alpha=1}^{m} (\bigcup_{i=1}^{n} \mathcal{U}_{\alpha_i})$$

observe que temos uma seqüência $\bigcup_{\alpha=1}^{m}(\bigcup_{i=1}^{n} \mathcal{U}_{\alpha_i})$ com $m \times n$ conjuntos:

$$\mathcal{U}_{11}, \mathcal{U}_{12}, \cdots, \mathcal{U}_{1n}; \mathcal{U}_{21}, \mathcal{U}_{22}, \cdots, \mathcal{U}_{2n}; \cdots; \mathcal{U}_{m1}, \mathcal{U}_{m2}, \cdots, \mathcal{U}_{mn}$$

reenumerando-os, temos $\mathcal{U}_1, \mathcal{U}_2, \cdots, \mathcal{U}_{m\times n}$

- Uma observação pertinente é a de que para cada α fixo $\mathcal{U}_{\alpha_i} \cap \mathcal{U}_{\alpha_j} = \emptyset$ (pela definição de espaço de recobrimento) e cada $\mathcal{U}_{\alpha_i}$ se aplica homeomorficamente a $\mathcal{V}_\alpha$.

A partir da seqüência acima obtemos uma seqüência disjunta (P_i) tal que $X = \bigcup_{i=1}^{n} P_i$. Como cada $\mathcal{U}_i$ $(i = 1, 2, \cdots, m \times n)$ se aplica homeomorficamente a $\mathcal{V}_\alpha$ e $P_i \subset \mathcal{U}_i$, pois $P_i = \mathcal{U}_i / E_{i-1}$, onde $E_i = \cup_{k=1}^{i} \mathcal{U}_k$. Temos uma aplicação injetiva(mas não bijetiva necessáriamente) de P_i em $\mathcal{V}_\alpha$

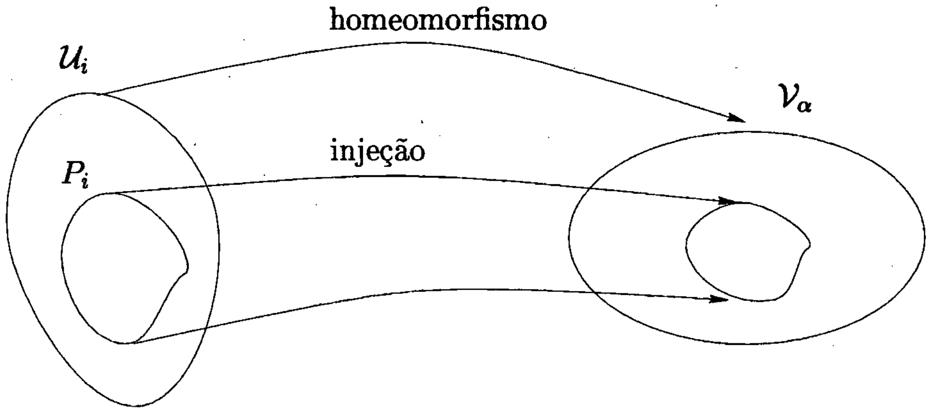

Figura 4.7: Aplicação injetiva de P_i em $\mathcal{V}_\alpha$

Então X foi decomposto em n pedaços disjuntos $P_1, P_2, \cdots, P_n$ tal que

$$T\Big|_{P_i} : P_i \to X$$

é injetiva.

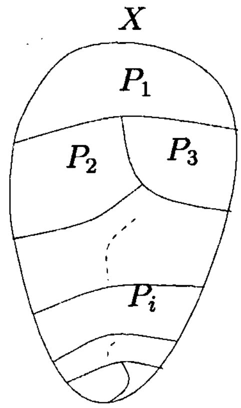

Figura 4.8: Decomposição de X

Obs: Dados um homeomorfismo local $T: X \to Y$ e um subconjunto $A \subset X$, a restrição $T\Big|_A$ é um homeomorfismo local de A sobre $T(A)$(ver [9], pg-111)

Seja $A \in X$ tal que $T\Big|_A: A \to TA$ seja um homeomorfismo. Logo para todo subconjunto aberto $A \subset X$ temos que $T\Big|_A: A \to TA$ é um homeomorfismo local.

Obs: os P_i da decomposição de X não necessáriamente são abertos, são apenas a diferença entre abertos $P_i = \mathcal{U}_i / (\bigcup_{n=1}^{i} E_{n-1})$

Em particular estão nesta situação (i.e., $T\Big|_A: A \to TA$ é um homeomorfismo) todos os $\mathcal{U}_{\alpha_i}$ $(1 = 1, \cdots, n)$ pois cada um se aplica homeomorficamente a $\mathcal{V}_\alpha$

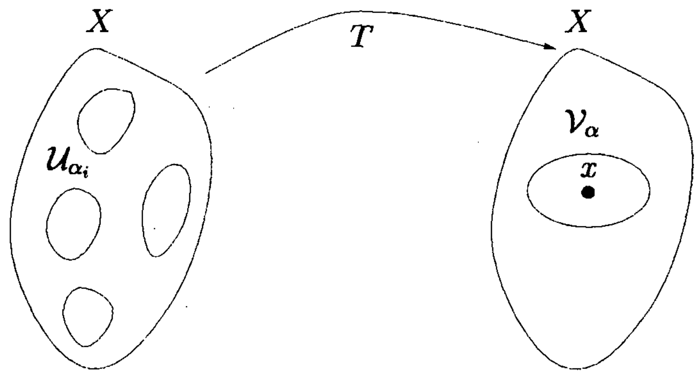

Figura 4.9: Aplicação homeomorfica de $\mathcal{U}_{\alpha_i}$ em $\mathcal{V}_\alpha$

Para os $A \in \beta(X)$ tal que $T\Big|_A: A \to TA$ é um homeomorfismo definimos

$$Q\mu(A) = \mu(TA)$$

Se $Y \in \mathcal{B}(X)$ é qualquer, definimos $Y_i = Y \cap P_i$. Observe que

$$\bigcup_{i=1}^{n} Y_i = \bigcup_{i=1}^{n}(Y \cap P_i) = Y \cap (\bigcup_{i=1}^{n} P_i) = Y \cap X = Y$$

então

$$Q\mu(Y) = Q\mu(\bigcup_{i=1}^{n} Y_i) = \sum_{i=1}^{n} Q\mu(Y_i)$$

pois, $Y_i \cap Y_j = \emptyset$ (pois $P_i \cap P_j = \emptyset$).

Referências

[**1**] BILLINGSLEY, P. **Ergodic teory and information**. New York: John Wiley and Sons, 1965.

[**2**] MAÑÉ, R. **Teoria ergódica**. Rio de Janeiro: IMPA, CNPq, 1983.

[**3**] DEVANEY, Robert L. **An introduction to chaotic dynamical systems**. Canada: Addison-Wesley, 1989.

[**4**] FERNANDEZ, P. **Medida e integração**. Rio de Janeiro: IMPA, CNPq, 1976.

[**5**] WALTERS, P. **An introduction to ergodic theory**. New York: Springer-Verlag, 1982.

[**6**] M. E. Munroe. **Introduction to measure and integration**. Massachusetts: Addison-Wesley, 1953.

[**7**] J.F.C. KINGMANN and S.J.TAYLOR. **Introduction to measure and probability**. New York: Cambridge University, 1966.

[**8**] ROBERT B. Ash. **Measure, integration, and functional analysis**. Academic Press, 1972.

[**9**] LIMA, Elon Lages. **Grupo fundamental e espaços de recobrimento**. Rio de Janeiro: IMPA,CNPq, 1993.

[**10**] P. Hulse. Uniqueness and ergodic properties of ergodic g-measures. Ergodic Theory Dynamical Systems **11** (1991), 65-77.

[**11**] M. Keane. Strongly mixing g-measures. Invent.Math **16** (1972), 309-324.

[**12**] P. Walters. Ruelle's operator theorem e g-measures. Trans.Amer.Math.Soc. **214** (1975), 375-387.

[**13**] Hopf, E. Statistik der geodätischen liner in Mannigfaltigreiten negativer. Krümmung, Ber. verch. sochs. Akad. wiss. Leipzig 91 (1939), 261-304.

[**14**] Anosov, D.V. Geodesic flows on closed Riemannian manifold with negative curvature, Proc Inst. Steklov, 90(1967) 1-235.

[**15**] Robinson, C., Young, L.S, Nonabsolutely Continuous Foliations for Anosov Diffeomorphisms.

[**16**] K. Krzyżewski. On expanding mappings. Bull. Acad. Polon. Sci. Sér. Sci. Math. **XIX**(1971),
23-24.

[**17**] P. Góra and B. Schimitt. Un exemple de transformation dilatante et C^1 par morceaux de l'intervalle, sans probabilité absolument continue invariante. Ergodic Theory Dynamical Systems **9**(1989), 101-113.

[**18**] A.N.QUAS. Non-ergodicity for C^1 expanding maps and g-measures.

[**19**] Craizer, M. Teoria ergódica das transformações expansoras. Informes de Matemática-Série E
Rio de Janeiro,IMPA-CNPq (1985).