# TEORIA ERGÓDICA E ALGUMAS APLICAÇÕES AOS PROCESSOS ESTOCÁSTICOS ESTACIONÁRIOS

JOÃO BEAL VARGAS

DISSERTAÇÃO APRESENTADA AO

INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

DA

UNIVERSIDADE DE SÃO PAULO

PARA OBTENÇÃO DO GRAU DE MESTRE

EM

ESTATÍSTICA

ORIENTADOR:

PROF. DR. FLÁVIO WAGNER RODRIGUES

- SÃO PAULO, MAIO DE 1976 -

À Maria Fernanda, pela

compreensão e paciência.

## AGRADECIMENTOS

Para a realização deste trabalho, o autor contou com a colaboração das seguintes pessoas, a quem deseja agradecer:

- Prof.Dr. Flávio Wagner Rodrigues, que sugeriu o tema e orientou de forma precisa o trabalho;
- Prof.Dr. Pedro Jesus Fernandez que contribuiu com valiosas sugestões e esclarecimentos;
- Jandyra Maria Guimarães Fachel, colega e companheira de estudos durante muitos anos;
- Ronaldo Jack Eckstein, professor e posteriormente companheiro de profícuas discussões;
- todos os professores do Departamento de Estatística do IME-USP, que, direta ou indiretamente, transmitiram ao autor parte de seu conhecimento atual;
- todos os professores do Departamento de Estatística do IMPA-UFRGS, que, direta ou indiretamente, deram seu apoio para que o autor realizasse tranquilamente a parte final do trabalho;
- Prof. Ernesto Bruno Cossi e Prof. Willy Engel (Departamento de Matemática da UFRGS) e Profa. Magda Bercht (Centro de Processamento de Dados da UFRGS) que deram pequenos, porém importantes, esclarecimentos;
- Sr. João Baptista Esteves de Oliveira, pelo serviço de datilografia que, além de rápido, foi excelente.

Finalmente, o autor gostaria de fazer um agradecimento especial à FAPESP, cujo apoio financeiro permitiu que este trabalho fosse realizado.

## APRESENTAÇÃO

Esta dissertação pretende conter principalmente uma exposição da Teoria Ergódica completa em si mesmo, apresentando-a da maneira mais geral possível (isto é, fazendo suposições somente quando elas são realmente imprescindíveis). Além disso, pretende mostrar algumas de suas aplicações aos Processos Estocásticos Estacionários.

A Teoria da Medida é, infelizmente, utilizada em todo o trabalho, a não ser no Capítulo 6 e no Apêndice I.

O estudo da Teoria Ergódica originou-se na Mecânica Estatística. O Capítulo 1 pretende apresentar, junto com as primeiras definições e propriedades, as situações físicas que motivaram essa teoria.

O problema da recorrência é o mais antigo e um dos mais importantes da Teoria Ergódica. O Capítulo 2 ocupa-se exclusivamente dele.

O Capítulo 3 é, juntamente com o Capítulo 1 e o Capítulo 5, fundamental nesta dissertação. Nele encontram-se a noção de ergocidade e alguns dos resultados mais importantes com ela relacionados.

O Capítulo 4 apresenta algumas respostas à seguinte questão: "que acontece ao sistema quando sua medida é substituída por outra?".

O Capítulo 5 apresenta uma questão que é bastante pesquisada atualmente (o problema do isomorfismo), uma classe de sistemas muito importante (Sistemas de Markov) e uma subclasse muito especial (Sistemas de Bernoulli), uma propriedade definida apenas em sistemas probabilísticos ("mixing") e as aplicações mais importantes da Teoria Ergódica aos Processos Estocásticos Estacionários.

A noção de entropia é fundamental em Teoria da Informação. O Capítulo 6 pretende apresentá-la detalhadamente (seu significado heurístico, suas propriedades e algumas de suas aplicações na Teoria Ergódica), omitindo as demonstrações.

O Apêndice I completa os aspectos intuitivos da noção de entropia. Os outros apêndices são constituídos por tópicos relativamente interessantes, mas desnecessários para compreender o que foi apresentado no corpo do trabalho.

# ÍNDICE

# CAPÍTULO 1

## INTRODUÇÃO

DEFINIÇÃO 1.1 - Um ESPAÇO MENSURÁVEL é um par ordenado $(X,A)$, onde $X$ é um conjunto não vazio e $A$ é uma $\sigma$-álgebra de subconjuntos de $X$.

DEFINIÇÃO 1.2 - Um CONJUNTO MENSURÁVEL em um espaço mensurável $(X,A)$ é um subconjunto de $X$ que pertença a $A$.

DEFINIÇÃO 1.3 - Uma TRANSFORMAÇÃO MENSURÁVEL em um espaço mensurável $(X,A)$ é uma transformação $T:X \to X$ que satisfaz $T^{-1}A \subset A$.

DEFINIÇÃO 1.4 - Uma ESTRUTURA é uma terna ordenada $(X,A,T)$, onde $(X,A)$ é um espaço mensurável e $T$ é uma transformação mensurável neste espaço.

EXEMPLO 1.1 - Seja $X = R^n$, $A$ a $\sigma$-álgebra de Borel de $X$ e $c \in X$ um ponto arbitrário fixado. A transformação $T:X \to X$ definida por $Tx = x+c$ é mensurável, de modo que $(X,A,T)$ é uma estrutura. $\Delta$

Na definição 1.5, a palavra MEDIDA significa uma função da $\sigma$-álgebra no intervalo $[0;+\infty] = [0;+\infty) \cup \{+\infty\}$, contavelmente aditiva e que leva o conjunto vazio no número zero (ver [4] Halmos).

DEFINIÇÃO 1.5 - Um ESPAÇO DE MEDIDA é uma terna ordenada $(X,A,u)$, onde $(X,A)$ é um espaço mensurável e $u$ é uma medida em $A$.

<u>EXEMPLO 1.2</u> - Sejam X e $A$ como no exemplo 1.1 e u a medida de Lebesgue em $A$. A terna ordenada $(X,A,u)$ é um espaço de medida. $\Delta$

Consideremos um sistema físico que vai se transformando com o tempo. Os possíveis estados (as possíveis fases) do sistema físico formam um conjunto X, que é denominado ESPAÇO FASE. Cada ponto do espaço fase representa, com total exatidão, todas as propriedades do estado que ele caracteriza.

<u>EXEMPLO 1.3</u> - Consideremos k partículas livres no espaço euclidiano tri-dimensional. Assumindo que as massas destas partículas e as forças que elas exercem são completamente conhecidas, o estado deste sistema físico num determinando instante de tempo pode ser descrito por 3k coordenadas de posição junto com as 3k coordenadas de velocidade correspondentes. (Estas 6k coordenadas não são as únicas que podem ser usadas; para certos propósitos, posição e momento são muito mais convenientes que posição e velocidade.) Temos, considerando a posição e a velocidade, que $X = R^{6k}$. $\Delta$

Escolhemos um instante no tempo para origem $(t = 0)$ e uma certa quantidade de tempo para unidade (esta quantidade pode ser 1 segundo, 3 minutos, 7 horas, etc.); deste modo, cada instante no tempo é identificado com um número real. Vamos assumir que o sistema físico sempre existiu, existe neste instante e sempre existirá. Esta hipótese ajuda muito na construção de um modelo matemático para o sistema físico.

Enquanto o tempo vai passando, o estado do sistema físico vai mudando de acordo com as leis físicas (equações diferenciais) apropriadas; a história completa do sistema fí sico (passado, presente e futuro) fica representada por uma determinada trajetória no espaço fase, isto é, uma aplicação de $(-\infty;+\infty)$ em X. De acordo com a Mecânica Clássica, ciência de caráter determinístico, a trajetória é totalmente determinada por um de seus pontos (o estado do sistema físico em algum instante t).

Na prática, quase nunca há meios para a determinação exata e completa do estado do sistema físico em algum instante t. Dois fatores são básicos para isso:

(F1) as limitações da percepção humana (e dos aparelhos de medição),

(F2) a necessidade irrealizável de parar o tempo no instante t para analisar e medir (isto é, determinar) o estado do sistema físico no instante t.

Para transpor esse impasse, a Mecânica Estatística abandona o estudo determinístico de estados (isto é, de pontos do espaço fase) e passa a realizar o estudo estatístico de conjuntos de estados (isto é, de sub-conjuntos do espaço fase). A pergunta "qual o estado do sistema no instante t, dado que o estado do sistema no instante $t_0$ é $x_0$?" é substituída por "qual a distribuição de probabilidade no espaço fase X do estado do sistema no instante t, dado que o estado do sistema no instante $t_0$ pertence ao sub-conjunto $A_0$?". E neste ponto que se torna necessário uma $\sigma$-álgebra $A$ de sub-conjuntos de X, que contenha todos os sub-conjuntos do espaço fase que interessem, e uma medida u em $A$. Esta medida é para tornar precisa a noção de tamanho de um conjunto mensurável no espaço fase.

EXEMPLO 1.4 - Consideremos um resistor cuja resistência sofre influência da luz (light dependent resistor). O estado do sistema físico sendo a resistência do resistor (medida em ohms), temos que $X = [0;+\infty)$, com, digamos, a $\sigma$-álgebra de Borel e uma determinada medida de probabilidade da família gama. $\Delta$

DEFINIÇÃO 1.6 - Uma estrutura $(X,A,T)$ é uma ESTRUTURA INVERSÍVEL se, e somente se, a transformação $T$ é bijetora e $TA \subset A$ (isto é, $T^{-1}: X \to X$ é uma transformação mensurável).

Se $(X,A,T)$ é uma estrutura inversível, então $(X,A,T^{-1})$ é, também, uma estrutura; $(X,A,T^{-1})$ é denominada ESTRUTURA INVERSA de $(X,A,T)$.

A estrutura mencionada no exemplo 1.1 é uma estrutura inversível.

A composição de uma transformação por si mesmo repetidas vezes é representada sob a forma de potência:

$T^0 = Id$

$T^n = TT\ldots T$, para $n > 0$

$T^{-n} = T^{-1}T^{-1}\ldots T^{-1}$, para $n > 0$ (quando $T$ é bijetora).

PROPOSIÇÃO 1.1 - a) Se $(X,A,T)$ é uma estrutura, então $(X,A,T^n)$ é uma estrutura, $\forall n \in N$.

b) Se $(X,A,T)$ é uma estrutura inversível, então $(X,A,T\ )$ é uma estrutura, $\forall n \in Z$. $\square$

Nas condições da proposição 1.1, a estrutura $(X,A,T^n)$ é denominada n-ÉSIMA POTÊNCIA DA ESTRUTURA $(X,A,T)$.

OBSERVAÇÃO 1.1 - Somente as estruturas inversíveis tem potência negativas.

Voltemos ao modelo que representa fielmente a realidade sem considerar a inexistência de meios para a determinação exata e completa do estado do sistema físico em algum instante t, isto é, ao modelo da Mecânica Clássica. Como a trajetória (baseada na correspondência bijetora convencionada pelo instante 0 e a quantidade unitária de tempo já adotados) do sistema físico é totalmente determinada por um de seus pontos, vamos definir, para cada $t \in (-\infty;+\infty)$, a transformação $T_t : X \to X$ que a cada $x \in X$ associa o estado do sistema físico no instante t dado que o estado do sistema físico no instante 0 é x. Portanto

(C1) $$T_0 = Id.$$

Com a hipótese física do sistema físico ser conservativo, resulta

(Hf) $$T_s T_t = T_{t+s} \ , \ \forall t,s \in (-\infty;+\infty).$$

Faremos também uma hipótese necessária do ponto de vista matemático:

(Hm) $T_t$ é uma transformação mensurável, $\forall t \in (-\infty;+\infty)$.

Um sistema "matemático" seria, então, a quadrupla ordenada

(S1) $$(X, A, \{T_t : -\infty < t < +\infty\}, u).$$

Adotar-se-á, entretanto, outra definição de sistema, que será justificada por comentários a serem feitos.

<u>MODELO 2</u> - A situação física citada acima pode ser substituída por uma na qual se supõe que o sistema físico "nasce" em algum determinado instante no tempo. Alternativamente, podemos supor que não há interesse em analisar (e medir) o sis

tema físico nos instantes anteriores a algum determinado instante no tempo. Da mesma forma, podemos supor que, embora havendo interesse, não há informações que permitam analisar (e medir) o sistema físico nos instantes anteriores a algum determinado instante no tempo. Em qualquer destes casos, o mencionado instante no tempo é convencionado como $t = 0$ e

a) assumimos que o sistema físico existirá para sempre, de modo a identificar $[0;+\infty)$ com o tempo,
b) as trajetórias no espaço fase são aplicações de $[0;+\infty)$ em $X$,
c) assumimos as hipóteses (Hf) e (Hm) com $[0;+\infty)$ no lugar de $(-\infty;+\infty)$,
d) um sistema "matemático" seria, então, a quádrupla ordenada

(S2) $$(X, A, \{T_t : 0 \le t < +\infty\}, u).$$

Adotar-se-á, também neste modelo, a definição de sistema prometida após a expressão (S1).

O problema básico da Mecânica Estatística é o estudo assintótico, isto é, quando $t \to +\infty$, de (S1) ou (S2) para responder a questões sobre o futuro do sistema físico.

Faremos agora uma simplificação nos nossos modelos:

A PASSAGEM DE TEMPO CONTÍNUO PARA TEMPO DISCRETO. Ao invés de analisar (e medir) o estado do sistema físico em cada instante $t$, faremos isso apenas nos instantes do tempo identificados com os números inteiros. Esse procedimento é relativamente razoável quer do ponto de vista físico quer do ponto de vista matemático: procuramos informação sobre (S1) ou (S2) tomando observações periodicamente no tempo.

<u>OBSERVAÇÃO 1.2</u> - A passagem de tempo contínuo para tempo dis

creto deve influir na escolha da quantidade unitária de tempo. Cada intervalo de tempo $[n;n+1)$ deve ser suficiente para registrar a observação feita no instante n.

A hipótese (Hf) e a conclusão (C1) asseguram as seguintes conclusões:

(C2) $$T_n = T_1 T_1 \ldots T_1 \ , \ \forall n \geq 1.$$

(C3) $$T_{-n} = T_{-1} T_{-1} \ldots T_{-1} \ , \ \forall n \geq 1.$$

(C4) $$T_{-1} T_1 = T_1 T_{-1} = T_0 = Id.$$

De (C1), (C2) e (C3) segue que estudar

(S1a) $$(X, A, \{T_n : -\infty < n < +\infty\}, u)$$

é, na verdade, estudar

(S1b) $$(X, A, \{T_{-1}^n : 1 \leq n < +\infty\} \cup \{T_1^n : 0 \leq n < +\infty\}, u).$$

<u>OBSERVAÇÃO 1.3</u> - No modelo 2, as conclusões (C3) e (C4) não tem sentido. De (C1) e (C2) segue que estudar

(S2a) $$(X, A, \{T_n : 0 \leq n < +\infty\}, u)$$

é, na verdade, estudar

(S2b) $$(X, A, \{T_1^n : 0 \leq n < +\infty\}, u).$$

A conclusão (C4) e a hipótese (Hm) asseguram que as estruturas $(X, A, T_1)$ e $(X, A, T_{-1})$ são inversíveis e uma é a in versa da outra, de modo que estudar (S1a) ou (S1b) é, na verdade, estudar

(S1c) $(X, A, \{T_1^n : -\infty < n < +\infty\}, u)$.

Em virtude de (S1c) e (S1b) e denotando $T_1$ simplesmente por T, será enfim definido matematicamente o sistema.

DEFINIÇÃO 1.7 - Um SISTEMA é uma quádrupla ordenada $(X, A, T, u)$, onde $(X, A, T)$ é uma estrutura e u é uma medida σ-finita na σ-álgebra A que satisfaz $uX>0$.

OBSERVAÇÃO 1.4 - Um sistema é, neste trabalho, uma estrutura com uma médida σ-finita não nula na σ-álgebra A. A estrutura desempenha papel muito mais importante que a medida, pois esta nada tem a ver com o sistema físico (é uma convenção do observador), ao passo que a estrutura descreve totalmente o comportamento do sistema físico nos instantes de tempo identificados com os números inteiros. A importância da estrutura já pode ser vista nas definições e notações abaixo.

DEFINIÇÃO 1.8 - Um SISTEMA INVERSÍVEL é um sistema cuja estrutura é inversível.

Se $S = (X, A, T, u)$ é um sistema inversível, então $S^{-1} = (X, A, T^{-1}, u)$ é também, um sistema; $S^{-1}$ é denominado o SISTEMA INVERSO de $S$.

PROPOSIÇÃO 1.2 - a) Se $S = (X, A, T, u)$ é um sistema, então $S^n = (X, A, T^n, u)$ é um sistema, $\forall n \in N$.

b) Se $S = (X, A, T, u)$ é um sistema inversível, então $S^n = (X,A,T^n,u)$ é um sistema, $\forall n \in Z$.

□

Nas condições da proposição 1.2, o sistema $S^n$ é denominado n-ÉSIMA POTÊNCIA DO SISTEMA $S$.

OBSERVAÇÃO 1.5 - Somente os sistemas inversíveis têm potências negativas.

O estudo de (S1) ou (S2) quando $t \to +\infty$ reduz-se, após a passagem de tempo contínuo para tempo discreto, ao estudo dos sistemas $S^n$ quando $n \to +\infty$.

DEFINIÇÃO 1.9 - Um sistema $(X,A,T,u)$ é um SISTEMA PRESERVATIVO se, e somente se, $uT^{-1} = u$.

PROPOSIÇÃO 1.3 - Seja $S = (X,A,T,u)$ um sistema inversível. São equivalentes:

i) $S$ é preservativo;

ii) $S^{-1}$ é preservativo.

PROVA - Como há uma simetria, basta provar que (i $\to$ ii). Suponhamos que $S$ é preservativo. Seja $A \in A$. Temos:

$$u(T^{-1})^{-1}A = u(TA) = uT^{-1}(TA) = uA. \qquad \square$$

PROPOSIÇÃO 1.4 - As potências de um sistema preservativo são sistemas preservativos.

PROVA - A 0-ésima potência é claramente preservativa. A 1-ésima potência é preservativa, por hipótese. Suponhamos que alguma n-ésima $(n \geq 1)$ potência seja preservativa. Seja $A \in A$. Temos:

$$uT^{-(n+1)}A = uT^{-n}T^{-1}A = uT^{-1}A = uA,$$

o que prova que a $(n+1)$-ésima potência é preservativa.

Quando o sistema é inversível, o caso $n=-1$ segue da proposição 1.3 e a indução feita acima prova os casos em que $n < -1$. □

**OBSERVAÇÃO 1.6** - Mesmo que um sistema preservativo $(X,\mathcal{A},T,u)$ não seja inversível, a imagem $TX$ da transformação $T$ é essencialmente o espaço $X$. De fato: se $A \cap TX = \emptyset$ e $A \in \mathcal{A}$, então $uA = uT^{-1}A = u\emptyset = 0$. Em particular: se $TX \in \mathcal{A}$, então $u(TX)^c = 0$.

**OBSERVAÇÃO 1.7** - Nos exemplos, sempre que $X$ for um sub-conjunto de Borel de algum $R^n$ e omitimos a descrição da $\sigma$-álgebra de $\mathcal{A}$, fica subentendido que se trata da $\sigma$-álgebra de Borel de $X$, isto é, o traço da $\sigma$-álgebra de Borel do $R^n$ no conjunto $X$. Em particular: se $X$ é um sub-conjunto enumerável do $R^n$, então $\mathcal{A}$ é a $\sigma$-álgebra formada por todos os sub-conjuntos de $X$.

**EXEMPLO 1.5** - Seja $c \neq 0$ um número real. A estrutura com espaço fase $X = R$ e transformação definida por $Tx = c \cdot x$ é inversível.

O sistema formado pela estrutura acima com a medida de Lebesgue $u$ é preservativo se, e somente se, $|c| = 1$. Δ

O exemplo 1.5 fornece sistemas inversíveis preservativos e não preservativos. Fazendo $c = 0$ no exemplo 1.5, temos um sistema não inversível e não preservativo. O exemplo

1.6 fecha esta questão, fornecendo sistemas não inversíveis preservativos.

<u>EXEMPLO 1.6</u> - Seja $r \geq 2$ um número inteiro. Seja $X = [0;1)$, u a medida de Lebesgue e T a transformação definida por $Tx = r \cdot x$ (mód 1). Este sistema é denominado SISTEMA r-ÁDICO. $\Delta$

<u>DEFINIÇÃO 1.10</u> - Um sistema $(X, A, T, u)$ é um SISTEMA PROBABILÍSTICO se, e somente se, $uX = 1$.

Os sistemas do exemplo 1.6 são sistemas probabilísticos. É claro que sistemas probabilísticos podem ser inversíveis ou não, preservativos ou não; idem para os não probabilísticos.

Uma grandeza física escalar (tal como energia, temperatura ou pressão) descrevendo uma propriedade do sistema físico deve ser pensada como uma função (mensurável) de X em R. O conceito de FUNÇÃO MENSURÁVEL é o clássico, isto é, $f^{-1}\mathcal{B} \subset A$, onde $\mathcal{B}$ é a $\sigma$-álgebra de Borel de R.

Consideremos novamente (S1) ou (S2) e uma "função energia" f de X em R. Para cada $x \in X$, o caminho de $(-\infty;+\infty)$ ou $[0;+\infty)$ em X definido por

$$t \to T_t x$$

é denominado CAMINHO DA FASE x. Para cada $x \in X$, o caminho de $(-\infty;+\infty)$ ou $[0;+\infty)$ em R definido por

$$t \to fT_t x$$

é denominado CAMINHO DE ENERGIA DA FASE x. A palavra ERGÓDICA vem do Grego e significa CAMINHO DE ENERGIA.

Após a passagem de tempo contínuo para tempo discreto os caminhos de energia da fase x são:

$$n \in Z \to fT^n x \in R \text{ (modelo 1)} \qquad \text{e}$$

$$n \in N \to fT^n x \in R \text{ (modelo 2)}.$$

Em qualquer caso, para cada $n \geq 1$, o número

$$\frac{1}{n+1} \cdot \sum_{k=0}^{n} fT^k x \in R,$$

é denominado ENERGIA MÉDIA DA FASE x NO TRAJETO $\{0,1,\ldots,n\}$.

Vamos finalizar este capítulo com um exemplo importante, que generaliza o exemplo 1.5.

EXEMPLO 1.7 - Seja $S$ um sistema com espaço fase $X = R^n$, T uma transformação linear e u a medida de Lebesgue. O conjunto TX é um subespaço vetorial de X (portanto $TX \in A$) e, caso $\dim TX < n$ (isto é, caso $\det T = 0$), temos $u(TX) = 0$. Portanto, para $S$ ser inversível ou preservativo, é necessário que

$$\det T \neq 0.$$

Esta condição é suficiente para T ser inversível. A fórmula (válida para $\det T \neq 0$ e obtida de um caso particular da mudança de variáveis em integrais múltiplas)

$$uT^{-1}A = \frac{1}{|\det T|} . uA, \quad \forall A \in A$$

garante que $S$ é preservativo se, e somente se,

$$|\det T| = 1. \qquad \Delta$$

- 0 -

## CAPÍTULO 2

## RECORRÊNCIA

Se é conhecido que, no instante $t=0$, o sistema físico se encontra em um determinado estado $x \in X$, todo o futuro (em particular, o comportamento assintótico) do sistema físico está determinado. Na prática, não se consegue determinar exatamente qual o estado no instante $t=0$, mas se consegue determinar algum conjunto $A \in \mathcal{A}$ ao qual o estado no instante $t=0$ pertença (na pior das determinações, temos $A=X$).

EXEMPLO 2.1 - O comprimento de uma haste metálica varia com o tempo. O sistema físico sendo a haste e o estado num instante sendo o comprimento da haste neste instante, podemos escolher $X=[0;+\infty)$ e $\mathcal{A}$ de acordo com a observação 1.7. Como determinar que o comprimento da haste é $\pi$ cm? No entanto, é possível determinar que o comprimento em cm da haste pertence ao intervalo fechado $[3,0;3,3] \in \mathcal{A}$. $\Delta$

Uma das primeiras questões assintóticas surgidas na teoria ergódica é a questão de recorrência. A respeito desta questão existe um resultado clássico (o teorema 2.1) estabelecido por Poincaré em 1899. Se é determinado que, no instante $t=0$, o estado do sistema físico pertence a um determinado conjunto $A \in \mathcal{A}$, o sistema físico retomará ao conjunto A?

**DEFINIÇÃO 2.1** - Sejam $(X,\mathcal{A},T)$ uma estrutura e $E$ um sub-conjunto qualquer de $X$. Um ponto $x\in X$ é E-RECORRENTE se, e somente se, $x\in E$ e $T^n x\in E$ para algum $n>0$. Um ponto $x\in E$ é INFINITAMENTE-E-RECORRENTE se, e somente se, $x\in E$ e $T^n x\in E$ para infinitos $n>0$.

**TEOREMA 2.1** - Se $(X,\mathcal{A},T,u)$ é um sistema preservativo tal que $uX<\infty$ e $A\in\mathcal{A}$, então quase todo ponto de $A$ é A-recorrente.

**PROVA** - Seja $B$ o conjunto dos pontos de $A$ que nunca retomam a $A$, isto é, $B=A\cap T^{-1}A^c\cap T^{-2}A^c\cap\ldots$. Pela proposição 1.1(a), $B\in\mathcal{A}$. Precisamos provar que $uB=0$.

Se $x\in B$, então $Tx,T^2x,\ldots\notin B$. Logo

(1) $$B\cap T^{-n}B=\emptyset \quad , \forall n\geq 1$$

(2) $$T^{-k}B\cap T^{-(n+k)}B = T^{-k}(B\cap T^{-n}B) = \emptyset \quad , \forall n,k\geq 1.$$

Das expressões (1) e (2) segue:

(3) $B,\ T^{-1}B,\ T^{-2}B,\ \ldots$ são disjuntos dois a dois.

Como o sistema é preservativo, segue:

$$uX\geq u(B\cup T^{-1}B\cup T^{-2}B\cup\ldots) \overset{(3)}{=} uB+uT^{-1}B+uT^{-2}B+\ldots$$
$$\geq uB+uB+uB+\ldots$$

Se $uB>0$, a desigualdade acima implica que $uX=\infty$; logo $uB=0$. $\square$

**TEOREMA 2.2** - Nas condições do teorema 2.1, quase todo ponto de $A$ é infinitamente-A-recorrente.

**PROVA** - Para cada $n\geq 1$, seja $B_n = A\cap T^{-n}A^c\cap T^{-2n}A^c\cap\ldots$. Para ca da $n \geq 1$, os sistemas $(X, A, T^n, u)$ são preservativos, conforme a proposição 1.4; aplicando o teorema 2.1 a cada um deles, obtemos

$$B_n \in A \quad e \quad uB_n = 0 \ , \ \forall n \geq 1.$$

Daí segue que

$$u(\cup\{B_n : n \geq 1\}) = 0.$$

Seja $x \in A - \cup\{B_n : n \geq 1\}$. Como $x \in A - B_1$, existe $n_1 > 0$ tal que $T^{n_1} x \in A$. Como $x \in A - B_{n_1}$, existe $n_2 > 0$ tal que $T^{n_2 \cdot n_1} x \in A$. Como $x \in A - B_{n_2 \cdot n_1}$, existe $n_3 > 0$ tal que $T^{n_3 \cdot n_2 \cdot n_1} x \in A$. E assim sucessivamente. □

No teorema 2.1, a finitude de u e a preservação de u foram usadas em um sentido muito fraco: o essencial é a não existência de um conjunto mensurável B de medida positiva tal que (3) seja verificada. Em outras palavras, o essencial é a não existência de um conjunto mensurável relevante (isto é, de medida não nula) B com a propriedade: se o sistema físico em algum instante $n \geq 0$ está em B, então o sistema físico nunca voltará a B.

**DEFINIÇÃO 2.2** - Sejam $(X, A, T)$ uma estrutura e E um sub-conjunto qualquer de X. E é um CONJUNTO ESTÁVEL se, e somente se, $TE \subset E$ (ou equivalentemente, $E \subset T^{-1}E$).

Um conjunto estável E tem, então, a propriedade: se o sistema físico está em E em algum instante $n \geq 0$, o sistema físico permanecerá em E para sempre.

**DEFINIÇÃO 2.3** - Seja $(X, A, u)$ um espaço de medida e g, f funções mensuráveis. g é u-MAIOR QUE f se, e somente se:

a) $\qquad gx \geq fx \ , \ \forall x \in X.$

b) existe $K \in A$ tal que $uK > 0$ e

$$gx > fx \quad , \quad \forall x \in K.$$

<u>PROPOSIÇÃO 2.1</u> - Seja $(X,A,T,u)$ um sistema. São equivalentes:

i) Existe $B\in$ tal que $uB > 0$ e $B, T^{-1}B, T^{-2}B, \ldots$ são disjuntos dois a dois.

ii) Existe um conjunto estável $A \in A$ tal que

$$u(T^{-1}A - A) = u\{x \notin A : Tx \in A\} > 0.$$

iii) Existe uma função mensurável $f$ tal que

$fT$ é u-maior que $f$.

<u>PROVA</u> - $(i \to ii)$ Faz-se $A = (B \cup T^{-1}B \cup T^{-2}B \cup \ldots)^c$, isto é, $A$ é o conjunto dos pontos $x$ tais que $x, Tx, T^2x, \ldots \notin B$; portanto

$$x \in A \to Tx, TTx, T^2Tx, \ldots \notin B \to Tx \in A \to x \in T^{-1}A,$$

isto é, $A$ é estável. Além disso, $A$ é mensurável e

$$x \in B \to \begin{cases} x \in B \cup T^{-1}B \cup T^{-2}B \cup \ldots \to x \in A^c \\ \\ Tx \notin B \cup T^{-1}B \cup T^{-2}B \cup \ldots \to Tx \in A. \end{cases}$$

Logo, $B \subset T^{-1}A - A$, o que implica $u(T^{-1}A - A) \geq uB > 0$.

$(ii \to i)$ Faz-se $B = T^{-1}A - A$; então $B$ é mensurável, $uB > 0$ e

$$(*) \qquad x \in T^{-n}B \leftrightarrow T^n x \in B \leftrightarrow TT^n x \in A \text{ e } T^n x \notin A.$$

Como $A$ é estável, segue de $(*)$ que $T^{-n}B$ é o conjunto das fases que são "absorvidas" em $A$ no instante $n+1$; logo

$$B, T^{-1}B, T^{-2}B, \ldots \text{ são disjuntos dois a dois.}$$

(ii $\to$ iii) Seja $f = I_A$. Então $f$ é mensurável e, como $fT = I_A T = I_{T^{-1}A}$ e $A$ é estável, temos

$$I_{T^{-1}A}x \geq I_A x \quad , \quad \forall x \in X.$$

Faz-se $K = T^{-1}A - A$, então $K \in \mathcal{A}$, $uK > 0$ e

$$x \in K \to I_{T^{-1}A}x = 1 > 0 = I_A x.$$

(iii $\to$ ii) Como $fT$ é u-maior que $f$, então

$$fx \geq c \to fTx \geq c \quad , \quad \forall x \in X \text{ e } \forall c \in R.$$

Disso resulta

(1) $$f^{-1}[c;+\infty) \subset T^{-1}f^{-1}[c;+\infty) \quad , \quad \forall c \in R.$$

Para cada $r \in Q$, seja

$$D_r = \{x \in X : fTx > r > fx\} \in \mathcal{A}.$$

Se $x \in K$, então $fTx > fx$; logo, para algum $r \in Q$, tem-se $fTx > r > fx$. Portanto

(2) $$K \subset \cup \{D_r : r \in Q\}.$$

Como $uK > 0$, (2) implica a existência de $a \in Q$ tal que

(3) $$uD_a > 0.$$

Seja $A = f^{-1}[a;+\infty) \in \mathcal{A}$. A expressão (1) afirma, em particular, que $A$ é estável e, como

$$T^{-1}A - A = \{x \in X : fx < a \text{ e } fTx \geq a\} \supset D_a,$$

segue de (3) que $u(T^{-1}A - A) > 0$. □

__DEFINIÇÃO 2.4__ - Um SISTEMA COMPRESSÍVEL é um sistema que satisfaz uma (e portanto todas) das expressões (i), (ii) ou (iii) da proposição 2.1. Um SISTEMA INCOMPRESSÍVEL é um sistema não compressível.

__PROPOSIÇÃO 2.2__ - a) As potências positivas de um sistema com pressível são compressíveis.

b) As potências positivas de um sistema incompressíveis. □

**PROVA** - Seja $S = (X,A,T,u)$ o sistema e seja $n$ um inteiro positivo.

a) Decorre imediatamente da proposição 2.1 (i). Vamos, então provar (b).

b) Por hipótese, $(X,A,T,u)$ é incompressível. Suponhamos que $(X,A,T^n,u)$ seja compressível. Pela proposição 2.1 (iii), existe uma função mensurável $g$ tal que $gT^n$ é u-maior que $g$. Seja

$$f = \sum_{k=0}^{n-1} gT^k.$$

A função $f$ é mensurável e, como

$$fT - f = \sum_{k=0}^{n-1} gT^{k+1} - \sum_{k=0}^{n-1} gT^k = gT^n - g,$$

segue que $fT$ é u-maior que $f$, isto é, $(X,A,T,u)$ é compressível. □

**TEOREMA 2.3** - Se $(X,A,T,u)$ é um sistema incompressível e $A \in A$, então quase todo ponto de $A$ é A-recorrente.

**PROVA** - Se a conclusão não for válida, então

$$B = A \cap T^{-1}A^c \cap T^{-2}A^c \cap \ldots,$$

que é mensurável, tem medida positiva. Como já foi visto na prova do teorema 2.1,

$$B, T^{-1}B, T^{-2}B, \ldots \text{ são disjuntos dois a dois.}$$

Logo, $(X,A,T,u)$ é compressível. □

**TEOREMA 2.4** - Nas condições do teorema 2.3, quase todo ponto de A é infinitamente-A-recorrente.

PROVA - Para cada $n \geq 1$, seja $B_n = A \cap T^{-n}A^c \cap T^{-2n}A^c \cap \ldots$. Para cada $n \geq 1$, os sistemas $(X, \mathcal{A}, T^n, u)$ são incompressíveis, conforme a proposição 2.2 (b); aplicando o teorema 2.3 a cada um deles, obtemos

$$B_n \in \mathcal{A} \quad \text{e} \quad uB_n = 0 \ , \ \forall n \geq 1.$$

Daí segue que ... (ver a prova do teorema 2.2).□

TEOREMA 2.5 - Todo sistema preservativo de medida finita (em particular, todo sistema preservativo probabilístico) é incompressível.

PROVA - Seja $\mathcal{S} = (X, \mathcal{A}, T, u)$ o sistema. Se $\mathcal{S}$ é compressível, existe $B \in \mathcal{A}$ tal que $uB > 0$ e $B, T^{-1}B, T^{-2}B, \ldots$ são disjuntos dois a dois, conforme a proposição 2.1 (i). Pela proposição 1.4, $uB = uT^{-1}B = uT^{-2}B = \ldots$. Logo

$$uX \geq u(B \cup T^{-1}B \cup T^{-2}B \cup \ldots) = uB + uB + uB + \ldots = +\infty.□$$

O exemplo 2.2 mostra que um sistema preservativo de medida infinita pode ser compressível.

EXEMPLO 2.2 - Sejam $X = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$, $\mathcal{A}$ de acordo com a observação 1.7 e u a medida de contagem. Seja $k \in X$ tal que $k > 0$. Seja $T_k$ a transformação definida por $T_k x = x + k$.
O conjunto $A = \{k+1, k+2, k+3, \ldots\}$ é estável, mensurável e tal que

$$u(T_k^{-1}A - A) = u\{1, \ldots, k\} = k > 0.$$

Logo, o sistema $\mathcal{S}_k = (X, \mathcal{A}, T_k, u)$ é compressível.

$\Delta$

O teorema 2.4, que seria o resultado central deste capítulo, pode ser assim enunciado: se $(X,A,T,u)$ é um sistema incompressível e $A \in A$, então

$$u\{x \in A: \sum_{n \geq 0} I_A T^n x < +\infty\} = 0$$

$I_A$ é uma função real mensurável não negativa e $A = \{x \in X: I_A x > 0\}$. O teorema 2.6 é uma generalização do teorema 2.4.

TEOREMA 2.6 - Se $(X,A,T,u)$ é um sistema incompressível f é uma função real mensurável não negativa e $A = \{x \in X: fx > 0\}$, então

$$u\{x \in A: \sum_{n \geq 0} fT^n x < +\infty\} = 0 .$$

PROVA - Para cada inteiro $k > 0$, seja $A_k = \{x \in X: fx < 1/k\}$. Como f é mensurável, cada $A_k$ é um conjunto mensurável. O teorema 2.4, aplicado a cada conjunto $A_k$, garante a existência de um conjunto $B_k \subset A_k$ satisfazendo:

$$B_k \in A,\ uB_k = 0,\ x \in A_k - B_k \rightarrow\ x \text{ é}$$

infinitamente-$A_k$-recorrente.

Em outras palavras,

(1) $$x \in A_k - B_k \rightarrow \sum_{n \geq 0} fT^n x \geq +\infty \cdot (1/k) = +\infty,$$

pois f é não negativa.

Como $A = \cup\{A_k : k > 0\}$, tem-se

$$x \in A - \cup\{B_k : k>0\} \rightarrow x \in \cup\{A_k - B_k : k>0\} \xrightarrow{(1)} \sum_{n \geq 0} fT^n x = +\infty.$$

Além disso,

$$\cup\{B_k : k>0\} \in A \text{ e } u(\cup\{B_k : k>0\}) = 0. \qquad \square$$

EXEMPLO 2.3 - Sejam $X = \{0,1,2,\ldots\}$, $\mathcal{A}$ de acordo com a observação 1.7 e u a medida de contagem. Para cada $y \in X$, seja $T_y$ a transformação definida por

$$T_y x = \begin{cases} x-1, & \text{se } x > 0 \\ y, & \text{se } x = 0. \end{cases}$$

Consideremos um sistema $S_y = (X,\mathcal{A},T_y,u)$. Temos que $A = \{0,\ldots,y\}$ é mensurável, estável e

$$u(T^{-1}A-A) = u\{y+1\} = 1 > 0.$$

Portanto, $S_y$ é compressível. $\Delta$

OBSERVAÇÃO 2.1 - Se o sistema é inversível, os resultados obtidos neste capítulo aplicados ao sistema inverso podem informar sobre a questão: se é determinado que, no instante $t = 0$, o estado do sistema físico pertence a um determinado conjunto $A \in \mathcal{A}$, o sistema físico já esteve anteriormente em A?

Este capítulo não deve ser concluído sem um exemplo que mostre a existência de sistemas incompressíveis.

EXEMPLO 2.4 - Sejam $X = \{1,2,\ldots,r\}$, $\mathcal{A}$ de acordo com a observação 1.7, T e u definidos por

$$Tx = \begin{cases} x+1, & \text{se } x \neq r \\ 1, & \text{se } x = r. \end{cases} \quad \text{e} \quad u\{x\} = \frac{1}{x}.$$

Os únicos conjuntos estáveis são X e $\emptyset$. Temos

$$T^{-1}X = X \quad \text{e} \quad T^{-1}\emptyset = \emptyset$$

$$u(T^{-1}X-X) = 0 \quad \text{e} \quad u(T^{-1}\emptyset-\emptyset) = 0.$$

Logo, $(X,A,T,u)$ é um sistema incompressível e, apenas para recordar, não preservativo. $\Delta$

<u>OBSERVAÇÃO 2.2</u> - Qualquer medida $\sigma$-finita não nula (ver definição 1.7) colocada na estrutura do exemplo 2.4 produzirá um sistema incompressível.

- 0 -

## CAPÍTULO 3

## ERGODICIDADE

Neste capítulo, são introduzidas outras duas $\sigma$-álgebras e, portanto, diversas mensurabilidades são mencionadas. Vamos introduzir alguma notação:

$\mathcal{M}(S,\mathcal{F})$ - representará o conjunto de todas as funções (a valores reais) mensuráveis do espaço mensurável $(S,\mathcal{F})$;

$\mathcal{L}^1(S,\mathcal{F},m)$ - representará o conjunto de todas as funções (a valores reais) mensuráveis e integráveis do espaço de medida $(S,\mathcal{F},m)$;

$\mathcal{L}^2(S,\mathcal{F},m)$ - representará o conjunto de todas as funções (a valores reais) mensuráveis e de quadrado integrável do espaço de medida $(S,\mathcal{F},m)$.

<u>DEFINIÇÃO 3.1</u> - Sejam $(X,\mathcal{A},T)$ uma estrutura e $E$ um subconjunto qualquer de $X$. $E$ é um CONJUNTO INVARIANTE se, e somente se, $T^{-1}E = E$.

<u>PROPOSIÇÃO 3.1</u> - Sejam $(X,\mathcal{A},T)$ uma estrutura e $E$ um subconjunto qualquer de $X$. São equivalentes:

(i) $E$ é invariante;

(ii) $E$ é estável e $E^c$ é estável. $\square$

Um conjunto invariante tem, então, a propriedade mencionada após a definição 2.2 e também a seguinte: se o sistema físico não está em $E$ em algum instante $n \geq 0$, ele jamais "entrará" em E.

**PROPOSIÇÃO 3.2** - a) Se E é um conjunto invariante, então $E^c$ é um conjunto invariante.

b) A união e a interseção de uma classe (enumerável ou não) de conjuntos invariantes são conjuntos invariantes.

c) A classe formada por todos os conjuntos invariantes é uma $\sigma$-álgebra. □

**OBSERVAÇÃO 3.1** - Anotaremos por $I$ a $\sigma$-álgebra dos conjuntos invariantes em uma estrutura $(X,A,T)$. A classe $A \cap I$ é uma (sub-)$\sigma$-álgebra (de $A$), pois é uma interseção de $\sigma$-álgebras.

**DEFINIÇÃO 3.2** - Sejam $(X,A,T)$ uma estrutura e f uma função qualquer de X em R. f é uma FUNÇÃO INVARIANTE se, e somente se, $fT = f$.

**PROPOSIÇÃO 3.3** - Sejam $(X,A,T)$ uma estrutura e f uma função qualquer de X em R. São equivalentes:

i) f é uma função invariante;

ii) $f \in M(X,I)$.

**PROVA** - ($i \rightarrow ii$) como f é uma função invariante, segue

$fx \leq a \leftrightarrow fTx \leq a, \forall x \in X$ e $\forall a \in R$.

$f^{-1}(-\infty;a] = T^{-1}f^{-1}(-\infty;a], \forall a \in R$.

$f^{-1}(-\infty;a] \in I, \forall a \in R$.

$f \in M(X,I)$.

($ii \rightarrow i$) Se $E \in I$, então $I_E T = I_{T^{-1}E} = I_E$ e, portanto, $I_E$ é uma função invariante.

A classe $M$, formada por todas as funções invariantes, é fechada para combinações lineares:

$f \in M$, $g \in M \rightarrow (f+g)T = fT + gT = f+g$

$c \in R$, $f \in M \rightarrow (c \cdot f)T = c \cdot fT = c \cdot f$.

$M$ é fechada para limites de sequências não decrescentes de funções não negativas: se $f_1, f_2, \ldots \in M$, $0 \le f_1 \le f_2 \le \ldots$

$$\left.\begin{array}{l} fT = (\lim f_n)T \ge f_k T = f_k, \forall k \ge 1 \rightarrow fT \ge f \\ f = \lim f_n \ge f_k = f_k T, \ \forall k \ge 1 \rightarrow f \ge fT \end{array}\right\} \rightarrow fT = f.$$

O teorema B (pg.85 - [4]) assegura que $M(X, I) \subset M$. □

Para expressar a idéia de ergodicidade, é também necessária a idéia de conjunto trivial. Uma medida m em um espaço mensurável $(S, F)$ é uma MEDIDA NÃO NULA se, e somente se, $mS > 0$.

DEFINIÇÃO 3.3 - Sejam $(X, A, \mu)$ um espaço de medida e E um subconjunto qualquer de X. E é um CONJUNTO TRIVIAL se, e somente se,

a) $E \in A$

b) $\mu E = 0$ ou $\mu E^c = 0$.

PROPOSIÇÃO 3.4 - Seja $(X, A, \mu)$ um espaço de medida. A classe formada por todos os conjuntos triviais é uma (sub-)$\sigma$-álgebra (de $A$). □

OBSERVAÇÃO 3.2 - Anotaremos por $T$ a $\sigma$-álgebra dos conjuntos triviais em um espaço de medida $(X, A, \mu)$.

DEFINIÇÃO 3.4 - Sejam $(X, A, \mu)$ um espaço de medida não nula e f uma função qualquer de X em R. f é uma FUNÇÃO $\mu$-CONSTANTE se, e somente se:

a) $f \in M(X, A)$,

b) existe $a \in R$ tal que $\mu\{x \in X : fx \neq a\} = 0$.
Nestes casos, a constante $a \in R$ é o $\mu$-VALOR de $f$ e $\{x \in X : fx = a\} \in A$ é o $\mu$-SUPORTE de $f$.

PROPOSIÇÃO 3.5 - Sejam $(X, A, \mu)$ um espaço de medida não nula e $f$ uma função qualquer de $X$ em $R$. São equivalentes:

i) $f$ é uma função $\mu$-constante;

ii) $f \in M(X, T)$.

PROVA - (i $\rightarrow$ ii) Seja $a \in R$ o $\mu$-valor de $f$ e seja $B \in \mathcal{B}$.

$a \in B \rightarrow f^{-1}\{a\} \subset f^{-1}B \rightarrow \mu(f^{-1}B)^c = 0 \rightarrow f^{-1}B \in T$

$a \notin B \rightarrow f^{-1}B \subset (f^{-1}\{a\})^c \rightarrow \mu f^{-1}B = 0 \rightarrow f^{-1}B \in T$.

Logo $f^{-1}\mathcal{B} \subset T$.

(ii $\rightarrow$ i) Para cada inteiro $k$ e cada inteiro positivo $n$, seja

$$A_{n,k} = \left\{x \in X : \frac{k}{2^n} \leq fx < \frac{k+1}{2^n}\right\}.$$

Para cada $n$ fixado, existe um único $A_{n,k}$ de medida não nula. Representemos por $k(n)$ este inteiro. Segue que "toda a medida" está em $A = \cap\{A_{n,k(n)} : n \geq 1\}$ e, claramente, existe um único $a \in R$ tal que: $x \in A \rightarrow fx = a$. Logo, $f$ é $\mu$-constante. □

Se um sistema $S = (X, A, \mu)$ é tal que $X$ é a união de dois conjuntos disjuntos $A$-mensuráveis $X_1$ e $X_2$ que são invariantes e tais que $\mu X_1 > 0$ e $\mu X_2 > 0$, o estudo de $I$ é realmente o estudo de dois sistemas

$S_1 = (X_1, A_1, T_1, \mu_1)$ e $S_2 = (X_2, A_2, T_2, \mu_2)$, onde

$A_i = X_i \cap A$

$T_i$ é a restrição de $T$ a $X_i$

$\mu_i$ é a restrição de $\mu$ a $A_i$.

Quando isto ocorre, $S$ é um SISTEMA DISPONÍVEL e $\{S_1, S_2\}$ é uma DECOMPOSIÇÃO DO SISTEMA $S$; cada sistema da decomposição pode ser estudado separado do outro, daí os únicos sistemas relevantes serem os indecomponíveis, também denominados ergódicos.

DEFINIÇÃO 3.5 - Um sistema $S = (X, A, T, \mu)$ é um SISTEMA ERGÓDICO se, e somente se, todos os conjuntos $A$-mensuráveis invariantes são triviais (isto é, se $A \cap I \subset T$).

OBSERVAÇÃO 3.3 - Ergodicidade é uma formulação precisa, mas não a única, da seguinte idéia: a transformação T faz uma perfeita ( em relação a $\mu$) desordem no espaço fase X (considerado puramente como conjunto).

EXEMPLO 3.1 - Consideremos o sistema $S_1$ no exemplo 2.2. Os únicos conjuntos invariantes são $\emptyset$ e X, logo $S_1$ é um sistema ergódico. $\Delta$

EXEMPLO 3.2 - Consideremos o sistema $S_2$ no exemplo 2.2. $\{\ldots,-4,-2,0,2,4,\ldots\}$ é um $A$-mensurável invariante não trivial, logo $S_2$ é um sistema não ergódico. $\Delta$

EXEMPLO 3.3 - Consideremos o sistema do exemplo 2.4. Pela proposição 3.1, para E ser um conjunto invariante é necessário que E seja um conjunto estável, logo $I = \{\emptyset, X\}$. Portanto, o sistema considerado é ergódico. $\Delta$

EXEMPLO 3.4 - Sejam r,s inteiros positivos. Seja $S(r,s)$ o sis tema definido por $X=\{1,\dots,r,r+1,\dots,r+s\}$,

$$Tx = \begin{cases} 1 & , \text{ se } x=r \\ r+1, & \text{se } x=r+s \\ x+1, & \text{nos outros casos}, \end{cases}$$

$\mu\{x\} = \frac{1}{x}$.

Como $C_1 = \{1,\dots,r\}$ é um conjunto $A$-mensurável invariante não trivial, o sistema $S(r,s)$ é não ergódico.

Os únicos conjuntos estáveis são

$$C_1 \text{ e } C_2 = \{r+1,\dots,r+s\}.$$

Ambos são $A$-mensuráveis, porém

$$\mu(T^{-1}C_1 - C_1) = \mu(T^{-1}C_2 - C_2) = \mu\emptyset = 0.$$

Portanto, $S(r,s)$ é um sistema incompressível.

$\Delta$

É interessante "aplicar" a observação 3.3 aos exemplos 3.1 e 3.3.

Apesar da proposição 3.1, ergodicidade nada tem em comum com incompressibilidade como mostra o seguinte quadro

| | ERGÓDICO | DECOMPONÍVEL |
|---|---|---|
| Compressível | Exemplo 3.1 | Exemplo 3.2 |
| Incompressível | Exemplo 3.3 | Exemplo 3.4 |

Isto aconteceu porque as noções de conjunto invariante e conjunto estável independem da medida, o que não acontece com as noções de sistema ergódico e sistema incompressível.

<u>PROPOSIÇÃO 3.6</u> - Seja $S = (X,A,T,\mu)$ um sistema. São equivalentes:

i) $S$ é um sistema ergódico;

ii) $M(X,A\cap I)\subset M(X,T)$.

**PROVA** - (i → ii) $f\in M(X,A\cap I) \to f^{-1}B\subset A\cap I \to f^{-1}B\subset T \to f\in M(X,T)$.
(ii → i) $A\in A\cap I \to I_A\in M(X,A\cap I) \to I_A\in M(X,T) \to A\in T$. □

**PROPOSIÇÃO 3.7** - Seja $S=(X,A,T,\mu)$ um sistema tal que $\mu X<\infty$.
São equivalentes:
i) $I$ é um sistema ergódico.
ii) $M(X,A\cap I)\cap L^1(X,A,\mu)\subset M(X,T)$.

**PROVA** - (i → ii) Segue da proposição 3.6.
(ii → i)

$$A\in A\cap I \to \begin{cases} I_A\in M(X,A\cap I) \\ A\in A \to \int I_A\cdot d\mu=\mu A\le\mu X<\infty \to I_A\in L^1(X,A,\mu) \to \end{cases}$$
$$\to I_A\in (X,T) \to A\in T. \quad \square$$

**EXEMPLO 3.5** - Seja X o conjunto dos números complexos de módulo unitário e $A$ a σ-álgebra gerado pelos arcos. Para cada $c\in X$, seja $T_c:X\to X$ a transformação definida por $T_cx=c\cdot x$. Seja $\mu$ a medida que a cada arco associa o seu comprimento dividido por $2\pi$. Assim $S_c=(X,A,T_c,\mu)$ é um sistema probabilístico. Mais ainda: $T_c$ atua em X simplesmente como uma singela rotação do ângulo que é o argumento do número complexo c, daí $S_c$ é inversível e preservativo.

Se $c^k=1$, para algum $k>0$, então $I_c$ é não ergótico: a função $f:X\to R$ definida por $fx=Re(x^k)$ é uma função $A$-mensurável e invariante que não é $\mu$-constante, isto é,

$$f\in M(X,A\cap I) \quad e \quad f\notin M(X,T)$$

(ver proposições 3.3, 3.5 e 3.6). Δ

**TEOREMA 3.1** - Seja $X$ é um espaço topológico com uma base enumerável, $A$ é a $\sigma$-álgebra de Borel de $X$ (isto é, a menor $\sigma$-álgebra que contém todos os subconjuntos abertos de $X$), $\mu$ é uma medida em $A$ tal que cada aberto não vazio tem medida positiva e $T:X \to X$ é uma transformação tal que $S = (X,A,T,\mu)$ é um sistema inversível. Se $S$ é ergódico, então, para quase todo $x \in X$, o caminho da fase $x$ (isto é, o conjunto $\{T^n x: n \in Z\}$) é denso em $X$.

**PROVA** - Seja $\{G_1, G_2, \ldots\}$ uma base enumerável para a topologia de $X$. Para cada $k \geq 1$, seja

$$G_k^* = (\cup\{T^n G_k: n \in Z\})^c.$$

Tem-se:

1) $G_k^*$ é um $A$-mensurável invariante;

2) $G_k^* \in T$, pois $S$ é ergódico;

3) $G_k^* \cap G_k = \emptyset$;

4) $\mu G_k > 0$; se $G_k \neq \emptyset$;

5) $\mu G_k^* = 0$, se $G_k \neq \emptyset$, por (2), (3) e (4).

Logo,

$$G = \cup\{G_k^*: k \geq 1 \text{ e } G_k \neq \emptyset\} \in A \text{ e } \mu G = 0.$$

Seja $x \in X$ uma fase. Se o caminho de $X$ não é denso em $X$, existe um aberto básico não vazio $G_k$ tal que $\{T^n x: x \in Z\} \cap G_k = \emptyset$ e, portanto, $x \in G_k^*$; mais ainda, $x \in G$. $\square$

O teorema 3.1 pode ser utilizado como uma maneira alternativa de mostrar que o sistema $S_c$ no exemplo 3.5 é não ergódico se $c^k = 1$, para algum $k > 0$: para todo $x \in X$, o caminho de x é o conjunto finito (e,portanto, não denso em X)

$$\{x, c\cdot x, \ldots, c^{k-1}\cdot x\}.$$

Nos casos em que $c^k \neq 1$, para todo $k > 0$, o caminho de cada fase $x \in X$ é denso em X; entretanto, isto não garante a ergodicidade de $S_c$, pois a recíproca do teorema 3.1 é falsa,conforme um exemplo na pág. 27 - [ 5 ] Halmos.

O teorema 3.2 se encontra demonstrado em [ 5 ] Halmos, nas págs. 26-28.

<u>TEOREMA 3.2</u> - Seja $(X,\cdot)$ um grupo abeliano topológico compacto e com uma base enumerável. Sejam $A$ e $\mu$ como nas condições do teorema 3.1. Para cada $c \in X$, seja $T_c$ a transformação definida por

$$T_c x = c\cdot x.$$

São equivalentes:

i) $S_c = (X, A, T_c, \mu)$ é ergódico;

ii) $\{c^n : n \in Z\}$ é denso em X. □

<u>EXEMPLO 3.6</u> - Se $c^k \neq 1$, para todo $k > 0$,o sistema $S_c$ no exemplo 3.5 é ergódico, conforme o teorema 3.2. Δ

<u>EXEMPLO 3.7</u> - Seja $S$ o sistema formado pela estrutura do exemplo 1.1 com a medida de Lebesgue. Seja d a métrica euclidiana em X. Para cada fase y, seja

$$B(y) = \{x \in X : d(x,y) \leq 1\}.$$

O conjunto $\cup\{B(k\cdot c) : k \in Z\}$ é um A-mensurável invariante não trivial, logo $S$ é não ergódico.Δ

<u>OBSERVAÇÃO 3.4</u> - Há vários resultados relacionando a Teoria Ergódica com outros ramos da Matemática na

literatura especializada. Neste trabalho, os teoremas 3.1 e 3.2 foram mencionados apenas como uma ilustração deste fato; como o objetivo é relacionar a Teoria Ergódica com a Teoria das Probabilidades, este trabalho não prosseguirá nesta direção.

<u>EXEMPLO 3.8</u> - Sejam $X = [0;1)$, $\mu$ a medida de Lebesgue e, para cada $c \in X$, sejam $T_c$ a transformação definida por $Tx = c+x$ (mod.1) e $S_c = (X, A, T_c, \mu)$.

Se $k \cdot c = 0$ (mod.1) para algum $k > 0$, então $S_c$ é não ergódico: a função $f: X \to R$ definida por $fx = k \cdot x$ (mod.1) é uma função A-mensurável e invariante que não é $\mu$-constante (ver proposições 3.3, 3.5 e 3.6).

Se $k \cdot c \neq 0$ (mod.1) para todo $k > 0$, então $S_c$ é ergódico, conforme o teorema 3.2.

Em qualquer caso, $S_c$ é um sistema preservativo e inversível. $\Delta$

<u>OBSERVAÇÃO 3.5</u> - Se $(X, A, T, \mu)$ é um <u>ESPAÇO DE MEDIDA NÃO ATÔMICO</u> (i.é, $A \in A$ e $\mu A > 0 \Rightarrow$ existe $B \in A$ tal que $\mu(A \cap B) > 0$ e $\mu(A \cap B^c) > 0$) e $\mu X = \infty$, não é fácil construir uma transformação mensurável $T$ no espaço mensurável $(X, A)$ de modo que $(X, A, T, \mu)$ seja um sistema ergódico.

O exemplo 3.9 mostra que existem sistemas ergódicos nas condições da observação 3.5, nos casos em que

$$\sum_{n \geq 1} a_n = \infty .$$

<u>EXEMPLO 3.9</u> - Seja $a_1, a_2, \ldots$ uma sequência estritamente de-

crescente de reais positivos menores que 1 tal que $\lim a_n = 0$. Seja $X_0 = \{(x;0) \in R^2 : 0 \le x < 1\}$ e, para cada $n \ge 1$, seja $X_n = \{(x,n) \in R^2 : 0 \le x < a_n\}$. Seja

$$X = \cup\{X_n : n \ge 0\}.$$

Seja $A$ a $\sigma$-álgebra de Borel e seja $\mu$ a medida que a cada segmento de reta contido em X associa o seu comprimento.

Sejam $A_0 = X_0 \cap A$, $\mu_0$ a restrição de $\mu$ a $A_0$, $T_0$: $X_0 \to X_0$ uma transformação tal que

$$S_0 = (X_0, A_0, T_0, \mu_0)$$

seja um sistema preservativo, inversível, ergódico. (Estas transformações existem, conforme o exemplo 3.8.) Vamos definir T pela expressão

$$T(x,n) = \begin{cases} (x,n+1), \text{ se } 0 \le x < a_{n+1} \\ T_0(x,0), \text{ se } a_{n+1} \le x < a_n. \end{cases}$$

Temos que $(X,A,T,\mu)$ é um sistema preservativo e inversível. Se $A \in A \cap I$, seja $A_0 = A \cap X_0$. Como $A \in I$, $A = \{(x,n) \in X: (x,0) \in A_0\}$ e $A_0$ é invariante em $S_0$. Portanto, ou $\mu_0 A_0 = 0$ ou $\mu_0 A_0^c = 1$, donde conclui-se que $\mu A = 0$ ou $\mu A^c = 0$, isto é $A \in T$. $\Delta$

O teorema 2.4 afirma: num sistema incompressível quase todo ponto de cada $A \in A$ retorna a A infinitas vezes. É natural perguntar durante que "fração do tempo" o estado do sistema físico ocupa um determinado conjunto $A \in A$. Mais precisamente: dada uma fase $x \in X$ (para o presente propósito não é necessário exigir que $x \in A$), formamos, para cada inteiro não negativo n, a razão

$$\frac{1}{n+1} \cdot \sum_{k=0}^{n} I_A T^k x$$

e procuramos obter o limite desta sequência de razões quando $n \to \infty$.

A existência deste limite (em algum dos conceitos de limite) e, quando ele existe, suas propriedades constituem os chamados TEOREMAS ERGÓDICOS. O problema do tempo médio de residência em A é, então, o problema da convergência de Cesaro para a sequência

$$I_A x,\ I_A Tx,\ I_A T^2 x,\ \ldots$$

e o primeiro passo significante nesta direção foi dado quando se procurou resolver este problema para uma função

$$f \in M(X,A)$$

no lugar da particular função $I_A \in M(X,A)$.

Existe uma boa quantidade de Teoremas Ergódicos. O teorema 3.3 é um deles, denominado TEOREMA ERGÓDICO INDIVIDUAL, devido a G.D.Birkhoff e F.Riesz, sem dúvida alguma o mais importante numa introdução à Teoria Ergódica. A demonstração pode ser encontrada em [ 5 ] Halmos, pág. 18-21.

<u>DEFINIÇÃO 3.6</u> - Sejam $(X,A,\mu)$ um espaço de medida e $f \in M(X,A)$, Uma VERSÃO DE f e qualquer função $g \in M(X,A)$ tal que $\mu\{x \in X: fx \neq gx\} = 0$.

<u>TEOREMA 3.3</u> - Se $S = (X,A,T,\mu)$ é um sistema preservativo e $f \in L^1(X,A,\mu)$, então:

a) $X_f = \left\{x \in X: \frac{1}{n+1} \cdot \sum_{k=0}^{n} fT^k x \text{ converge}\right\} \in A$ e $\mu X_f^c = 0$,

b) $f^*: X \to R$ definida por:

$$f^* x = \begin{cases} \lim_{n\to\infty} \frac{1}{n+1} \cdot \sum_{k=0}^{n} fT^k x, & \text{se } x \in X_f \\ 0 & \text{, caso contrário} \end{cases}$$

é tal que $f^* \in L^1(X,A,\mu)$,

c) $f^*$ possue uma versão $f'$ em $M(X,A\cap I)$,

d) $\mu X < \infty \rightarrow \int f\cdot d\mu = \int f^*\cdot d\mu$. □

A conclusão obtida no teorema 3.3.(d) pode não ser válida se omitimos a hipótese adicional $\mu X < \infty$, como mostra o exemplo 3.10.

**EXEMPLO 3.10** - Sejam $X = R$, $T$ a transformação definida por $Tx = x+1$, $\mu$ a medida de Lebesgue, $f = I_{[0;1]}$. Claramente, $f^* = 0$ e

$$\int f\cdot d\mu = 1 \neq 0 = \int f^*\cdot d\mu. \qquad \Delta$$

**TEOREMA 3.4** - Se $S = (X,A,T,\mu)$ é um sistema preservativo ergódico e se $f \in L^1(X,A,\mu)$, então $f^*$ (definida no teorema 3.3) é $\mu$-constante com $\mu$-valor:

a) $\frac{1}{\mu X}\int f\cdot d\mu$, se $\mu X < \infty$;

b) $0$ , se $\mu X = \infty$.

**PROVA** - O teorema 3.3(c) assegura que $f^*$ possue uma versão $f'$ em $M(X,A\cap I)$. Como $S$ é ergódico, a proposição 3.6 assegura que $f' \in M(X,T)$, isto é, $f'$ é $\mu$-constante. Seja $a \in R$ o $\mu$-valor de $f'$. Como

$$\{f^* \neq a\} \subset \{f^* = f',\ f' \neq a\} \cup \{f^* \neq f'\}$$

segue $\mu\{f^* \neq a\} = 0$ e, portanto, $f^*$ é $\mu$-constante.

a) Do teorema 3.3(d) segue

$$\int f\cdot d\mu = \int f^*\cdot d\mu = \int a\cdot d\mu = a\cdot \mu X \rightarrow a = \frac{1}{\mu X}\int f\cdot d\mu.$$

b) Se o $\mu$-valor de $f^*$ é $a \neq 0$, então

$$\left|\int f^*\cdot d\mu\right| = \left|\int a\cdot d\mu\right| = |a\cdot\mu X| = \infty \quad e \quad f^* \notin L^1(X,A,\mu)$$

contrariando o teorema 3.3(b). □

Em outras palavras, o teorema 3.4(b) afirma: a ENERGIA MÉDIA ASSINTÓTICA $f^*x$ DA FASE $x \in X$ é igual, a menos que $x$ pertença a um conjunto de medida nula, à ENERGIA MÉDIA DO SISTEMA $\frac{1}{\mu X}\int f\cdot d\mu$. Esta afirmação tem grande utilidade nas aplicações da Teoria Ergódica à Física.

Condições suficientes para que um sistema seja ergódico apresentam bastante interesse, quer do ponto de vista matemático quer das aplicações.

TEOREMA 3.5 - Seja $S = (X,A,T,\mu)$ um sistema preservativo tal que $\mu X < \infty$. São equivalentes:

i) $S$ é um sistema ergódico;

ii) $f \in L^1(X,A,\mu) \Rightarrow f^*$ é $\mu$-constante.

PROVA - A afirmação (i → ii) é válida até sem a hipótese da medida ser finita, conforme o teorema 3.4. A afirmação (ii → i) será provada utilizando a proposição 3.7: seja $f \in M(X,A \cap I) \cap L^1(X,A,\mu)$, então $f = fT$ e existe $f^*$ (pela proposição 3.3 e pelo teorema 3.3), logo $f = f^*$ e, por hipótese, $f$ é $\mu$-constante, isto é, $f \in M(X,T)$. □

Se $S = (X,A,T,\mu)$ é um sistema preservativo tal que $\mu X = \infty$ e as energias médias assintóticas $f^*$ sejam $\mu$-constantes para toda $f \in L^1(X,A,\mu)$, o sistema $S$ não é necessariamente ergódico. Em outras palavras, se omitirmos a hipótese da medida ser finita no teorema 3.5, (ii) não é uma condição suficiente para (i), como mostra o exemplo 3.11.

EXEMPLO 3.11 - Sejam $X = Z\times Z$, onde $Z = \{\dots,-2,-1,0,1,2,\dots\}$, T a transformação definida por $T(x,y) = (x+1,y)$, $\mu$ a medida de contagem. Claramente, $\mathcal{S} = (X,\mathcal{A},T,\mu)$ é um sistema preservativo e $\mu X = \infty$. Como as "retas horizontais" são conjuntos $\mathcal{A}$-mensuráveis invariantes não triviais, $\mathcal{S}$ é um sistema não ergódico. Mais ainda:

a) qualquer função não possue nenhuma versão distinta de si mesma,

b) as "retas horizontais" são os átomos de $\mathcal{I}$,

c) $g\in\mathcal{M}(X,\mathcal{A}\cap\mathcal{I})\cap\mathcal{L}^1(X,\mathcal{A},\mu) \xrightarrow{(b)} g = 0$.

Portanto:

$$f\in\mathcal{L}^1(X,\mathcal{A},\mu) \xrightarrow{T.3.3 \text{ e } (a)}$$

$$f^*\in\mathcal{L}^1(X,\mathcal{A},\mu)\cap\mathcal{M}(X,\mathcal{A}\cap\mathcal{I}) \xrightarrow{(c)} f^* = 0$$

$\longrightarrow f^*$ é $\mu$-constante. Δ

EXEMPLO 3.12 - Sejam $X = [0;1)\times[0;1)$, T a transformação definida por $T(x,y) = (x,x+y[\text{mod}.1])$, $\mu$ a medida de Lebesgue. Com um pequeno esforço geométrico, vê-se que $\mathcal{S} = (X,\mathcal{A},T,\mu)$ é um sistema preservativo. Como todo conjunto da forma $B\times[0;1)$, onde B é um boreliano de $[0;1)$, é $\mathcal{A}$-mensurável e invariante, segue que $\mathcal{S}$ é não ergódico.

Observe-se que $\mu X = 1 < \infty$. Portanto, pelo teorema 3.5, existe $f\in\mathcal{L}^1(X,\mathcal{A},\mu)$ tal que $f^*$ não é $\mu$-constante: $f = I_A$, onde $A = [0;1/2)\times[0,1)$, por exemplo, é tal que $f^* = f$ não é $\mu$-constante. Δ

OBSERVAÇÃO 3.6 - Se $\mathcal{S} = (X,\mathcal{A},T,P)$ é um sistema probabilístico, preservativo e ergódico e $A\in\mathcal{A}$, o teore ma 3.4(a) afirma que a $\mu$-constante de $I_A^*$ é precisamente PA. (Este fato possue uma interpretação bem interessante).

<u>DEFINIÇÃO 3.7</u> - Sejam $(X,A,T)$ uma estrutura e $E$ um subconjunto não vazio de $X$. A função

$$\pi_E: E \to \{1,2,\ldots,+\infty\}$$

definida por

$$\pi_E x = \begin{cases} \min\{n>0: T^n x \in E\}, & \text{se este mínimo existe} \\ +\infty & \text{, caso contrário} \end{cases}$$

é denominada 1º TEMPO DE RECORRÊNCIA DE E.

<u>PROPOSIÇÃO 3.8</u> - (a) Se $(X,A,T)$ é uma estrutura e $A \in A$, então $\pi_A$ é mensurável (considerando em $\{1,2,\ldots,+\infty\}$ a $\sigma$-álgebra formada por todos os subconjuntos).

(b) Se $(X,A,T,\mu)$ é um sistema incompessível e $A \in A$, então $\mu\{x \in A: \pi_A x = \infty\} = 0$.

<u>PROVA</u> - (a) Se $1 \leq n < +\infty$, então

$$A_n \overset{not}{=} \pi_A^{-1}\{n\} = A \cap T^{-1}A^c \cap \ldots \cap T^{-(n-1)}A^c \cap T^{-n}A \in A.$$

Por outro lado,

$$\pi_A^{-1}\{+\infty\} = (\cup\{A_n: 1 \leq n < \infty\})^c \in A.$$

(b) É outra forma do teorema 2.3. $\square$

O seguinte teorema é devido a Kac (ver o boletim da A.M.S., 1947, pg. 1006).

<u>TEOREMA 3.6</u> - Se $S = (X,A,T,P)$ é um sistema probabilístico, preservativo, inversível e ergódico, $A \in A$ e

$PA > 0$, então

$$\int_A \pi_A \cdot dP = 1.$$

Do teorema 3.6, obtemos

$$\frac{1}{PA}\int_A \pi_A \cdot dP = \frac{1}{PA},$$

cuja interpretação (bem interessante) é:o tempo médio para um ponto de A retomar pela 1ª vez ao conjunto A é precisamente 1/PA.

A ergodicidade de $S$ não pode ser omitida no teorema 3.6: se $B \in \mathcal{A}$ é um conjunto invariante ($\pi_B x = 1$, $\forall x \in B$) não trivial ($0 < PB < 1$), então

$$\int_B \pi_B \cdot dP = PB \neq 1. \qquad \square$$

OBSERVAÇÃO 3.7 - Nos teoremas deste capítulo e do capítulo 2, a hipótese do sistema ser preservativo é frequentemente assumida. Também é bastante assumida a hipótese da medida ser finita.

## CAPÍTULO 4

## RELAÇÕES ENTRE MEDIDAS

Neste capítulo serão consideradas várias medidas num espaço mensurável $(X,A)$ ou numa estrutura $(X,A,T)$ e certas relações entre elas (rever a observação 1.4). Vamos introduzir alguma notação:

$M(X,A)$ representará o conjunto de todas as medidas $\sigma$-finitas não nulas no espaço mensurável $(X,A)$.

$M^b(X,A) = \{\mu \in M(X,A): \mu X < \infty\}$

$M^1(X,A) = \{\mu \in M(X,A): \mu X = 1\}$

$M(X,A,T) = \{\mu \in M(X,A): \mu T^{-1} = \mu\}$

$M^b(X,A,T) = \{\mu \in M(X,A): \mu T^{-1} = \mu \text{ e } \mu X < \infty\}$

$M^1(X,A,T) = \{\mu \in M(X,A): \mu T^{-1} = \mu \text{ e } \mu X = 1\}$.

DEFINIÇÃO 4.1 - Seja $(X,A,T)$ uma estrutura. Uma medida $\mu \in M(X,A)$ é uma MEDIDA ERGÓDICA (respectivamente, COMPRESSÍVEL, INCOMPRESSÍVEL, PRESERVATIVA, PROBABILÍSTICA) se, e somente se, o sistema $(X,A,T,\mu)$ é um sistema ergódico (respectivamente, compressível, incompressível, preservativo, probabilístico).

DEFINIÇÃO 4.2 - Um SUPORTE DE UMA MEDIDA $\mu \in M(X,A)$ é um conjunto $S \in A$ tal que $\mu S^c = 0$.

É óbvio que cada medida $\mu \in M(X,A)$ tem pelo menos um suporte: o conjunto X.

<u>DEFINIÇÃO 4.3</u> - Sejam $\mu,\nu \in M(X,A)$.

a) $\nu$ é uma MEDIDA ABSOLUTAMENTE CONTÍNUA COM RESPEITO A $\mu$ (anote-se $\nu \ll \mu$) se, e somente se, $A \in A$ e $\mu A = 0 \to \nu A = 0$.

b) $\nu$ é uma MEDIDA EQUIVALENTE A $\mu$ (anote-se $\nu \sim \mu$) se, e somente se,

$$\nu \ll \mu \quad \text{e} \quad \mu \ll \nu.$$

c) $\nu$ é uma MEDIDA SINGULAR A $\mu$ (anote-se $\nu \perp \mu$) se, e somente se, existe um suporte de $\nu$ disjunto de algum suporte de $\mu$.

<u>NOTAÇÃO</u> - $K(\mu) = \{\nu : \nu \in M(X,A) \text{ e } \nu \ll \mu\}$.

<u>TEOREMA 4.1</u> - Sejam $(X,A,T)$ uma estrutura e $\mu,\nu \in M(X,A)$.

a) $\nu \ll \mu$ e $\nu$ compressível $\to$ $\mu$ compressível.

b) $\nu \ll \mu$ e $\mu$ incompressível $\to$ $\nu$ incompressível.

c) $\nu \ll \mu$ e $\mu$ ergódica $\to$ $\nu$ ergódica.

d) $\mu \in M^b(X,A,T) \to \mu$ incompressível.

<u>PROVA</u> - a) Segue imediatamente da proposição 2.1(i).

b) Segue imediatamente da proposição 2.1(ii).

c) Segue imediatamente de $T_\mu \subset T_\nu$.

d) É outra forma do teorema 2.5. □

<u>TEOREMA 4.2</u> - Sejam $(X,A,T)$ uma estrutura inversível e $\mu \in M(X,A)$ uma medida preservativa e ergódica. São equivalentes:

i) $\nu \in M(X,A,T) \cap K(\mu)$;

ii) existe $c > 0$ tal que $\nu = c \cdot \mu$.

<u>PROVA</u> - (ii $\to$ i) Trivial.

(i $\rightarrow$ ii) Pelo teorema de Radon-Nikodym, existe uma função $f \in \mathcal{M}(X,\mathcal{A})$ tal que

$$(*) \qquad \nu A = \int_A f \cdot d\mu, \ \forall A \in \mathcal{A}.$$

Seja $A \in \mathcal{A}$. Temos que

$$\nu(TA) = \int_{TA} f \cdot d\mu.$$

Como $\nu T^{-1} = \nu$, segue

$$\nu T^{-1}(TA) = \nu(TA) \ \therefore \ \nu A = \nu(TA) \ \therefore \ \int_A f \cdot d\mu = \int_{TA} f \cdot d\mu$$

e, como $\mu T^{-1} = \mu$, segue

$$\int_A f \cdot d\mu = \int_{TA} f \cdot d(\mu T^{-1}).$$

Pela "mudança de variável", resulta

$$\int_A f \cdot d\mu = \int_A (fT) \cdot d\mu.$$

Da unicidade (em relação a $\mu$) da derivada de Radon-Nikodym, segue que $f$ é uma $\mu$-versão de $fT$. Logo:

$$D \overset{\text{not}}{=\!=} \{x \in X : fx \neq fTx\} \in \mathcal{A} \text{ e } \mu D = 0.$$

$$\overline{D} \overset{\text{not}}{=\!=} \cup\{T^n D : n \in Z\} \in \mathcal{A} \cap \mathcal{I}$$

e, pela proposição 1.4, $\mu\overline{D} = 0$. A função $f' : X \rightarrow R$ definida por

$$f'x = \begin{cases} fx, & \text{se } x \notin \overline{D} \\ 0, & \text{se } x \in \overline{D} \end{cases}$$

satisfaz $f'T = f'$:

$$x \notin \overline{D} \rightarrow Tx \notin \overline{D} \rightarrow f'(Tx) = f(Tx) = fx = f'x$$

$$x \in \overline{D} \rightarrow Tx \in \overline{D} \rightarrow f'(Tx) = 0 = f'x.$$

Portanto, $f' \in \mathcal{M}(X, \mathcal{A} \cap \mathcal{I})$.

Como $\mu$ é ergódica, a proposição 3.6 afirma que $f'$ é

$\mu$-constante com, digamos, $\mu$-valor c. De (*) segue:

$$\nu A = c\cdot\mu A,\ \forall A\in\mathcal{A}.$$

Como $\nu\in M(X,\mathcal{A})$, c é positiva. □

<u>TEOREMA 4.3</u> - Se $(X,\mathcal{A},T)$ é uma estrutura inversível e

$$\mu\in M^b(X,\mathcal{A})$$

é uma medida preservativa e ergódica, então, para cada $c>0$, a equação $\nu X = c$ possui uma única solução em $M^b(X,\mathcal{A},T)\cap K(\mu)$:

$$\nu_c = \frac{c}{\mu X}\cdot\mu.$$

<u>PROVA</u> - $\nu_c\in M^b(X,\mathcal{A},T)\cap K(\mu)$ e $\nu_c X = c$. A unicidade decorre do teorema 4.2. □

<u>TEOREMA 4.4</u> - Se $(X,\mathcal{A},T)$ é uma estrutura inversível,

$$\mu\in M^b(X,\mathcal{A})$$

é uma medida preservativa e ergódica, $\mu_0$ é a restrição de $\mu$ a $\mathcal{A}\cap\mathcal{I}$,

$$\nu_0\in M^b(X,\mathcal{A}\cap\mathcal{I}) \quad \text{e} \quad \nu_0 \ll \mu_0,$$

então existe uma única extensão de $\nu_0$ em

$$M^b(X,\mathcal{A},T)\cap K(\mu).$$

<u>PROVA</u> - $\nu_1 = \frac{\nu_0 X}{\mu X}\cdot\mu$ é, pelo teorema 4.3, a única medida em $M^b(X,\mathcal{A},T)\cap K(\mu)$ que satisfaz $\nu X = \nu_0 X$. Falta provar que $\nu_1$ é uma extensão de $\nu_0$, isto é $B\in\mathcal{A}\cap\mathcal{I} \rightarrow \nu_1 B = \nu_0 B$.

Seja $B\in\mathcal{A}\cap\mathcal{I}$. Como $\mu$ é ergódica, ou $\mu B = 0$ ou $\mu B^c = 0$.

$$\mu B = 0 \rightarrow \begin{cases} \mu_0 B = 0 \rightarrow \nu_0 B = 0 \\ \nu_1 B = 0. \end{cases}$$

$$\mu B^c = 0 \to \begin{cases} \mu_0 B^c = 0 \to \nu_0 B^c = 0 \to \nu_0 B = \nu_0 X \\ \mu B = \mu X \to \nu_1 B = \nu_0 X. \end{cases} \qquad \square$$

TEOREMA 4.5 - Se $\mu_1, \mu_2 \in M^b(X, A, T)$ são ergódicas, então ou $\mu_1 \sim \mu_2$ ou $\mu_1 \perp \mu_2$.

PROVA - Suponha que $\mu_1$ não é uma medida equivalente a $\mu_2$, isto é, existe $B \in A$ tal que

$$(*) \qquad \begin{cases} \mu_1 B = 0 \quad e \quad \mu_2 B > 0 \text{ ou} \\ \mu_1 B > 0 \quad e \quad \mu_2 B = 0 \text{ .} \end{cases}$$

Basta provar que $\mu_1 \perp \mu_2$. Para cada $i \in \{1,2\}$, temos

$$\int I_B \cdot d\mu_i = \mu_i B \leq \mu_i X < \infty \quad \therefore \quad I_B \in L^1(X, A, \mu_i).$$

Do teorema 3.4, segue que

$$S_i = \left\{ x \in X : I_B^* x = \frac{\mu_i B}{\mu_i X} \right\}$$

é um suporte da medida $\mu_i$. Por (*), $S_1 \cap S_2 = \emptyset$. $\square$

TEOREMA 4.6 - Seja $\mu \in M^b(X, A, T)$. São equivalentes:

i) $\mu$ é ergódica;

ii) $\nu \in K(\mu) \to \nu$ é ergódica;

iii) $\nu \in M^b(X, A, T) \cap K(\mu) \to \nu$ é ergódica.

PROVA - (i $\to$ ii) Resulta do teoerema 4.1(c).

(ii $\to$ iii) Óbvio.

(iii $\to$ i) Trivial, pois $\mu \in M^b(X, A, T) \cap K(\mu)$. $\square$

TEOREMA 4.7 - Se $\mu$ e $\nu$ são medidas em $M^b(X, A, T)$, $\mu$ é ergódica e $\nu << \mu$, então $\nu \sim \mu$.

PROVA - O teorema 4.6 assegura que $\nu$ é ergódica. Segue do teo rema 4.5 que ou $\nu \sim \mu$ ou $\nu \perp \mu$ (esta é impossível, pois $\nu << \mu$). □

Os três teoremas seguintes são os análogos dos três teoremas anteriores para medidas de probabilidade. Suas demonstrações são, respectivamente, muito semelhantes.

TEOREMA 4.5' - Se $P_1, P_2 \in M^1(X,A,T)$ são ergódicas, então ou $P_1 = P_2$ ou $P_1 \perp P_2$. □

TEOREMA 4.6' - Seja $P \in M^1(X,A,T)$. São equivalentes:

i) $P$ é ergódica;

ii) $\nu \in K(P) \to \nu$ é ergódica;

iii) $Q \in M^1(X,A,T)\ K(P) \to Q$ é ergódica. □

TEOREMA 4.7' - Se $P$ e $Q$ são medidas em $M^1(X,A,T)$, $P$ é ergódica e $Q << P$, então $Q = P$. □

NOTAÇÃO - $V(X,A)$ representará o conjunto de todas as *medidas com sinal* finitas no espaço mensurável $(X,A)$.

$V(X,A,T) = \{\mu \in V(X,A):\ \mu T^{-1} = \mu\}$.

Sejam $V$ um espaço vetorial sobre $R$, $\mu_1 \in V$, $\mu_2 \in V$.

$$[\mu_1;\mu_2] \overset{\text{Not}}{=} \{a \cdot \mu_1 + (1-a) \cdot \mu_2 :\ 0 \le a \le 1\}.$$

DEFINIÇÃO 4.4 - Seja $V$ um espaço vetorial sobre $R$. $E \subset V$ é um CONJUNTO CONVEXO se, e somente se,

$$\mu_1 \in E,\ \mu_2 \in E \to [\mu_1;\mu_2] \subset E.$$

DEFINIÇÃO 4.5 - Seja $E$ um conjunto convexo. $\mu \in E$ é um PONTO EXTREMO se, e somente se,

$$\mu_1 \in E,\ \mu_2 \in E,\ \mu \in [\mu_1;\mu_2] \to \mu = \mu_1 \text{ ou } \mu = \mu_2.$$

As afirmações abaixo são fáceis de serem verificadas:

(A1) $V(X,\mathcal{A})$, com as operações abaixo, é um espaço vetorial sobre $\mathbb{R}$:

$$(\mu_1+\mu_2)A = \mu_1 A + \mu_2 A, \text{p/qq } \mu_1 \in V(X,\mathcal{A}), \mu_2 \in V(X,\mathcal{A}), A \in \mathcal{A}$$

$$(a\cdot\mu)A = a\cdot\mu A, \text{ p/qq } a \in \mathbb{R},\ \mu \in V(X,\mathcal{A}),\ A \in \mathcal{A};$$

(A2) $V(X,\mathcal{A},T)$ é um subespaço vetorial de $V(X,\mathcal{A})$;

(A3) $\left\{\begin{array}{l} V \text{ é um espaço vetorial sobre } \mathbb{R} \\ E \text{ é um conjunto convexo em } V \\ W \text{ é um subespaço vetorial de } V \end{array}\right\} \to E \cap W$ é um conjunto convexo em $V$;

(A4) $M^b(X,\mathcal{A})$ é um conjunto convexo em $V(X,\mathcal{A})$;

(A5) $M^b(X,\mathcal{A},T)$ é um conjunto convexo em $V(X,\mathcal{A})$;

(A6) $M^1(X,\mathcal{A})$ é um conjunto convexo em $V(X,\mathcal{A})$;

(A7) $M^1(X,\mathcal{A},T)$ é um conjunto convexo em $V(X,\mathcal{A})$.

<u>TEOREMA 4.8</u> - Se $\mu$ é um ponto extremo no conjunto convexo $M^b(X,\mathcal{A})$, então $\mu$ é ergódica.

<u>PROVA</u> - Se $\mu$ não é ergódica, existe $B \in \mathcal{A} \cap \mathcal{I}$ tal que $\mu B > 0$ e $\mu B^c > 0$. As medidas $\mu_1$ e $\mu_2$ definidas por

$$\mu_1 A = \frac{\mu X}{\mu B}\cdot\mu(B\cap A) \quad \text{e} \quad \mu_2 A = \frac{\mu X}{\mu B^c}\cdot\mu(B^c\cap A)$$

satisfazem

$$\mu_1 \in M^b(X,\mathcal{A}), \mu_2 \in M^b(X,\mathcal{A}),$$

$$\mu = \frac{\mu B}{\mu X}\cdot\mu_1 + \frac{\mu B^c}{\mu X}\cdot\mu_2 \in [\mu_1;\mu_2],$$

$$\mu \neq \mu_1 \text{ e } \mu \neq \mu_2.$$

Logo, $\mu$ não é um ponto extremo. □

<u>TEOREMA 4.8'</u> - Se $\mu$ é um ponto extremo no conjunto convexo $M^b(X,A,T)$, então $\mu$ é ergódica. □

<u>TEOREMA 4.8''</u> - Se $\mu$ é um ponto extremo no conjunto convexo $M^1(X,A)$, então $\mu$ é ergódica. □

<u>TEOREMA 4.8'''</u> - Se $\mu$ é um ponto extremo no conjunto convexo $M^1(X,A,T)$, então $\mu$ é ergódica. □

<u>EXEMPLO 4.1</u> - Seja $X = \{1,2,\ldots,r\}$, T a transformação definida por

$$Tx = \begin{cases} x+1, & \text{se } x \neq r \\ 1, & \text{se } x = r. \end{cases}$$

Como $A \cap I = \{\emptyset, X\}$, qualquer medida em $M(X,A)$ é ergódica.

Para cada $c > 0$, seja $\mu_c$ a medida que a cada conjunto unitário associa o número $c > 0$. Então

1) $\mu_c \in M^b(X,A,T) \subset M^b(X,A)$, $\forall c > 0$

2) $\mu_c = \frac{1}{4}\cdot\mu_{3c} + \frac{3}{4}\cdot\mu_{c/3} \in [\mu_{3c};\mu_{c/3}]$, $\forall c > 0$

3) $\mu_c \neq \mu_{3c}$ e $\mu_c \neq \mu_{c/3}$, $\forall c > 0$.

Portanto, $\mu_c$ não é um ponto extremo, quer em $M^b(X,A,T)$ quer em $M^b(X,A)$, apesar de ser ergódica. Δ

<u>EXEMPLO 4.2</u> - Sejam $X = \{1,2\}$. T a transformação definida por $T1 = 2$ e $T2 = 1$.

Como $A \cap I = \{\emptyset, X\}$, qualquer medida em $M(X,A)$ é ergódica.

Para cada $c \in [0;1]$, seja $P_c$ a probabilidade em $M(X,A)$ tal que $P_c\{1\} = c$, $P_c\{2\} = 1-c$.

Se $c \in (0;1/3)$, então:

1) $P_c = \frac{1}{4} \cdot P_{3c} + \frac{3}{4} \cdot P_{c/3} \in [P_{3c}; P_{c/3}]$

2) $P_c \neq P_{3c}$ e $P_c \neq P_{c/3}$.

Portanto, para qualquer $c \in (0;1/3)$, $P_c$ não é um ponto extremo em $M^1(X,A)$, apesar de ser ergódica. $\Delta$

Os exemplos 4.1 e 4.2 mostram que as recíprocas dos teoremas 4.8, 4.8' e 4.8" são falsas, mas não há exemplos para mostrar que a recíproca do teorema 4.8''' é falsa.

<u>TEOREMA 4.9</u> - Se P é uma medida ergódica em $M^1(X,A,T)$, então P é um ponto extremo no conjunto convexo $M^1(X,A,T)$.

<u>PROVA</u> - Se P não é um ponto extremo em $M^1(X,A,T)$, existem $P_1, P_2 \in M^1(X,A,T)$ tais que $P_1 \neq P_2$ e existe $a \in (0;1)$ tal que $P = a \cdot P_1 + (1-a) \cdot P_2$. Portanto

1) $P_1 \ll P$

2) $P_1 \neq P$ (pois $P_1 = P \rightarrow P_1 = P_2$).

Pelo teorema 4.7', (1) e (2) constituem um absurdo. $\square$

# CAPÍTULO 5

## ALGUNS TÓPICOS RELEVANTES

Os tópicos abordados neste capítulo podem ser vistos como "aplicações" da Teoria Ergódica aos Processos Estocásticos Estacionários.

O termo "aplicações" está entre aspas pela seguinte nota no rodapé da pg. 1002 do Boletim da A.M.S. (1947):

*"That the theory of stationary stochastic processes is mathematically equivalent with an "ergodic" theory (to wich one is also led by dynamical considerations) was clearly recognized by J.L.Doob, Stochastic Processes and Statistics, Proc.Nat.Acad.Sci.U.S.A., vol.20 (1934), pg. 376-379."*

### 5.1 - O PROBLEMA DO ISOMORFISMO

O ente matemático básico neste trabalho é um SISTEMA. Há alguns sistemas que são formalmente distintos, mas são idênticos do ponto de vista matemático, isto é, isomorfos.

**DEFINIÇÃO 5.1.1** - $(X_1,A_1,T_1)$ e $(X_2,A_2,T_2)$ são ESTRUTURAS ISOMORFAS se, e somente se, existe uma aplicação $\phi: X_1 \to X_2$ tal que:

a) $\phi$ é injetora e sobrejetora

b) $\phi A_1 \subset A_2$ e $\phi^{-1} A_2 \subset A_1$

c) $\phi T_1 x = T_2 \phi x,\ \forall x \in X_1$.

Para definir sistemas isomorfos é necessário negligenciar eventuais conjuntos de medida nula.

DEFINIÇÃO 5.1.2 - $(X_1, A_1, T_1, \mu_1)$ e $(X_2, A_2, T_2, \mu_2)$ são SISTEMAS ISOMORFOS se, e somente se, existem dois conjuntos, $Y_1$ e $Y_2$, e uma aplicação $\phi: Y_1 \to Y_2$ tais que:

a) $Y_i \in A_i \cap I_i,\ \forall i \in \{1,2\}$

b) $\mu_i Y_i^c = 0,\ \forall i \in \{1,2\}$

c) $\phi$ é injetora e sobrejetora

d) $\phi(Y_1 \cap A_1) \subset Y_2 \cap A_2$ e $\phi^{-1}(Y_2 \cap A_2) \subset Y_1 \cap A_1$

e) $\phi T_1 y = T_2 \phi y,\ \forall y \in Y_1$

f) $\mu_2 A = \mu_1 \phi^{-1} A,\ \forall A \in Y_2 \cap A_2$.

Nas condições da definição 2, as estruturas

$$(Y_i,\ Y_i \cap A_i,\ S_i),$$

onde $S_i$ denota a restrição da transformação $T_i$ ao conjunto $Y_i$, $i \in \{1,2\}$, são isomorfas.

EXEMPLO 5.1.1 - Seja $r \geq 2$ um número inteiro. Seja $S_1$ o sistema r-ádico (ver exemplo 1.6). Seja $S_2$ o sistema cujo espaço fase é o conjunto de todos os números complexos de módulo unitário, cuja $\sigma$-álgebra é a gerada pelos arcos, cuja transformação é definida por

$$T_2 z = z^r$$

e a medida é aquela que a cada arco associa o seu comprimento dividido por $2\pi$, de modo que $\mu_2 X_2 = 1$.

É fácil ver que $S_1$ e $S_2$ são isomorfos com $Y_1 = X_1$, $Y_2 = X_2$ e $\phi$ definido por

$$\phi x = \cos(2\pi x) + i \operatorname{sen}(2\pi x). \qquad \Delta$$

EXEMPLO 5.1.2 - Sejam $(X_1, A_1, \mu_1)$ e $(X_2, A_2, \mu_2)$ como no exemplo 5.1.1. Seja $c \in X_2$. Sejam $T_1$ e $T_2$ definidas por

$$T_1 x = x + \frac{\arg c}{2\pi} (\text{mód}.1) \quad \text{e} \quad T_2 z = c \cdot z.$$

Com $Y_1$, $Y_2$ e $\phi$ como no exemplo 5.1.1, temos novamente dois sistemas isomorfos. $\Delta$

OBSERVAÇÃO 5.1.1 - É bastante trivial verificar que isomorfismo, quer de sistemas quer de estruturas, possui as três propriedades de uma relação de equivalência: reflexividade, simetria, transitividade.

OBSERVAÇÃO 5.2.2 - Para cada inteiro $n \geq 0$, as n-ésimas potências de duas estruturas isomorfas (dois sistemas isomorfos) são estruturas isomorfas (sistemas isomorfos). Se as estruturas (os sistemas) são inversíveis, este resultado estende-se para todo inteiro n.

## 5.2 - SISTEMAS DE BERNOULLI

Nesta seção será apresentada uma classe de sistemas muito importantes para a Teoria Ergódica, os Sistemas de Bernoulli. Esses sistemas podem ser apresentados de diversas formas; serão aqui introduzidos com um enfoque da Teoria da Informação para dar uma pequena ilustração das aplicações da Teoria Ergódica a esta teoria.

Consideremos um ALFABETO finito com r LETRAS (isto é, sinais distintos) $\Omega$. A esse alfabeto associamos a $\sigma$-álgebra $2^{\Omega}$, formada por todos os subconjuntos de $\Omega$.

Escolhemos um instante no tempo para origem $(t = 0)$ e uma certa quantidade de tempo para unidade (essa quantidade pode ser 1 segundo, 3 minutos, 7 horas, etc.).

Imaginemos uma FONTE GERADORA que, em cada intervalo unitário $[n-1;n)$ do tempo, escolhe uma letra $w_n \in \Omega$ de acordo com uma lei de probabilidade p em $(\Omega, 2^{\Omega})$ que, além de não considerar as letras já escolhidas em períodos anteriores, não se modifica com a passagem do tempo. O experimentador recebe as letras escolhidas pela fonte geradora através de um CANAL DE TRANSMISSÃO no instante n.

Consideremos somente dois casos:

a) a fonte geradora sempre existiu e assim procedeu, existe e assim procede, existirá e assim procederá para sempre $(I = Z = \{\ldots,-2,-1,0,1,2,\ldots\})$;

b) a fonte geradora "nasceu" no instante $t = 0$, existirá para sempre, procedeu assim desde seu nascimento e procederá assim para sempre $(I = N = \{1,2,\ldots\})$.

Cada sequência I-infinita, $I \in \{Z,N\}$, de realizações do pequeno experimento $(\Omega, 2^{\Omega}, p)$ pode ser vista como uma realização de um grande experimento $(X,A,P)$: o espaço de probabilidade produto do pequeno experimento (ver [ 4 ] Halmos). Assim, um particular resultado do grande experimento (isto é, um ponto $x \in X$) é uma sequência I-infinita de pontos de $\Omega$ (uma aplicação de I em $\Omega$), denominada uma MENSAGEM I-infinita:

$$(1) \qquad x = \begin{cases} (\ldots, w_{-2}, w_{-1}, w_0, w_1, w_2, \ldots), & \text{se } I = Z \\ (w_1, w_2, w_3, \ldots) & \text{, se } I = N. \end{cases}$$

O grande experimento não pode ser encaixado na teoria frequentista da probabilidade, uma vez que uma de suas realizações utiliza todo o tempo futuro.

Para cada $n \in I$, a aplicação mensurável $h_n: X \to \Omega$ definida, usando a representação (1) acima, por

$$(2) \qquad h_n(x) = w_n$$

é denominada n-ÉSIMA PROJEÇÃO.

Vamos intercalar nesta exposição uma interpretação que pode ajudar a compreender o que vai ser apresentado.

<u>INTERPRETAÇÃO</u> - Uma mensagem $x \in X$ é escolhida pela fonte geradora de acordo com a lei de probabilidade P antes de ser iniciado o tempo. Em cada período unitário $[n-1;n)$ do tempo, ela envia ao experimentador através do canal de transmissão uma letra da mensagem x que ela escolheu: $h_n(x)$, a n-ésima projeção da mensagem x. A mensagem x está determinada (ela foi escolhida antes de iniciar o tempo), mas é desconhecida pelo experimentador, que vai conhecendo suas projeções na medida em que o tempo se escoa. Ao tempo n, o experimentador conhece apenas as projeções $h_k(x)$, para $k \leq n$. A mensagem x será inteiramente conhecida somente após todo o tempo passar. $\square$

Desloquemos o tempo uma unidade. Seja $T: X \to X$ a transformação definida por

$$(3) \qquad h_n(Tx) = h_{n+1}(x).$$

Na representação (1) acima, tem-se:

$$Tx = \begin{cases} (\ldots, w_{-1}, w_0, w_1, w_2, w_3, \ldots), & \text{se } I = Z \\ (w_2, w_3, w_4, \ldots) & \text{, se } I = N. \end{cases}$$

A transformação T possui as seguintes propriedades:

(p1) T é A-mensurável

(p2) $PT^{-1} = P$

(p3) $\begin{cases} I = Z \to T \text{ é 1-1 e sobrejetora} \\ I = N \to T \text{ é r-1 é sobrejetora,} \end{cases}$

portanto (X,A,T,P) é um sistema probabilístico preservativo (inversível, no caso em que $I = Z$). Como

$$h_n(x) = h_1(T^{n-1}x), \ \forall n \in I \text{ e } \forall x \in X, \tag{4}$$

a 1-ésima projeção e T permitem conhecer todas as projeções e, portanto, a mensagem x (qualquer que ela seja).

INTERPRETAÇÃO (cont.) - Conhecer a mensagem x que a fonte geradora escolheu é conhecer as projeções $h_k(x)$, $k \in I$. O deslocamento T transforma x de modo que, ao tempo n, o experimentador conhece as projeções $h_k(x)$, para $k \leq n+1$. Para cada $m > 1$, a transformação $T^m$ adianta o tempo m unidades de modo que, ao tempo n, o experimentador conhece as projeções $h_k(x)$, para $k \leq n+m$. □

Quando $I = Z (I = N)$, o sistema (X,A,T,P) é denominado SISTEMA BILATERAL (UNILATERAL) DE BERNOULLI GERADO POR $(\Omega, 2^\Omega, p)$.

OBSERVAÇÃO 5.2.1 - Se $\Delta$ é também um alfabeto finito com r letras e q é uma probabilidade em $(\Delta, 2^\Delta)$ que "atribui os mesmos valores que p", isto é, existe uma aplicação $\alpha: \Omega \to \Delta$ tal que

i) $\alpha$ é 1-1 (e, portanto, sobrejetora)
ii) $p\{w\} = q\{\alpha w\}$, $\forall w \in \Omega$,

então os sistemas bilaterais (unilaterais) de Bernoulli gerados por $(\Omega, 2^\Omega, p)$ e $(\Delta, 2^\Delta, q)$ são claramente isomorfos. Em particular, se substituirmos $p$ por uma probabilidade $p'$ em $(\Omega, 2^\Omega)$ que atribui os mesmos valores que $p$, os sistemas bilaterais (unilaterais) de Bernoulli gerados por $(\Omega, 2^\Omega, p)$ e $(\Omega, 2^\Omega, p')$ são claramente isomorfos.

NOTAÇÃO - Vamos representar $(\Omega, 2^\Omega, p)$ pela matriz

$$(5) \qquad \begin{bmatrix} \Omega \\ p \end{bmatrix} = \begin{bmatrix} w_1 & w_2 & \dots & w_r \\ p_1 & p_2 & \dots & p_r \end{bmatrix}$$

e, em virtude da observação 5.2.1, vamos anotar os sistemas (bilateral e unilateral) de Bernoulli gerado por $(\Omega, 2^\Omega, p)$ somente por

$$SBZ[p_1; p_2; \dots; p_r] \text{ e } SBN[p_1; p_2; \dots; p_r],$$

respectivamente, omitindo qual é o alfabeto $\Omega$ (que não é importante do ponto de vista matemático) e a $\sigma$-álgebra $2^\Omega$ (pois sempre é ela a $\sigma$-álgebra associada aos conjuntos finitos).

EXEMPLO 5.2.1 - Sejam

$$\begin{bmatrix} \Omega \\ p \end{bmatrix} = \begin{bmatrix} a & b & c & d \\ \text{zero} & 0,2 & 0,4 & 0,4 \end{bmatrix},$$

$$\begin{bmatrix}\Omega\\ p'\end{bmatrix} = \begin{bmatrix}a & b & c & d\\ \text{zero} & 0{,}4 & 0{,}4 & 0{,}2\end{bmatrix},$$

$$\begin{bmatrix}\Delta\\ q\end{bmatrix} = \begin{bmatrix}1 & 2 & 3 & 4\\ 0{,}4 & 0{,}4 & 0{,}2 & \text{zero}\end{bmatrix}.$$

Quaisquer dois dos sistemas bilaterais (unilaterais) de Bernoulli gerados por

$$(\Omega,2^{\Omega},p),\ (\Omega,2^{\Omega},p'),\ (\Delta,2^{\Delta},q)$$

são isomorfos e representados por

SBZ[zero;0,2;0,4;0,4]

(SBN[zero;0,2;0,4;0,4]). $\Delta$

O exemplo 5.2.1 sugere as perguntas, cujas respostas são afirmativas, abaixo:

SBZ[zero;0,2;0,4;0,4] é isomorfo a SBZ[0,2;0,4;0,4]?

SBN[zero;0,2;0,4;0,4] é isomorfo a SBN[0,2;0,4;0,4]?

Prosseguindo nesta direção, são isomorfos:

SBZ[1/2;1/2] e SBZ[1/3;1/3;1/3]?

SBZ[1/2;1/2] e SBN[1/2;1/2]?

SBZ[1/2;1/2] e SBN[1/3;1/3;1/3]?

A resposta a questões deste tipo serão dadas utilizando o conceito de entropia (Capítulo 6).

Vejamos dois exemplos de isomorfismos menos triviais que os da seção 5.1.

<u>EXEMPLO 5.2.2</u> - Seja $r \geq 2$ um número inteiro. Sejam $\Omega = \{0,1,\ldots,r-1\}$, $p$ a probabilidade em $(\Omega,2^{\Omega})$ definida por $p\{0\} = p\{1\} = \ldots = p\{r-1\} = \frac{1}{r}$, $S_1 = (X_1,A_1,T_1,P_1)$ o sistema unilateral de Bernoulli gerado por $(\Omega,2^{\Omega},p)$, $S_2=(X_2,A_2,T_2,P_2)$

o sistema r-ádico (ver exemplo 1.6).

Um número $x \in X_2 = [0;1)$ é denominado um NÚMERO RACIONAL r-ÁDICO se, e somente se, existem $n \in \{1,2,\ldots\}$ e $k \in \{0,1,\ldots,r^n-1\}$ tais que $x = \frac{k}{r^n}$. Esses números admitem duas representações na base r. Para exemplificar, consideremos $r = 10$.

$$\frac{1}{10} = \begin{cases} 0{,}0999\ldots \\ 0{,}1000\ldots \end{cases} \qquad \frac{26}{10^2} = \begin{cases} 0{,}25999\ldots \\ 0{,}26000\ldots \end{cases}$$

$$\frac{999}{10^3} = \begin{cases} 0{,}998999\ldots \\ 0{,}999000\ldots \end{cases}$$

Uma das duas representações de cada número racional r-ádico é constituída por no máximo uma quantidade finita de algarismos diferentes de zero.

Seja $Y_1$ o conjunto cujo complementar é formado por todas as mensagens constituídas por no máximo uma quantidade finita de algarismos diferentes de zero. $Y_1^c$ é uma união enumerável de conjuntos finitos, portanto enumerável, portanto $Y_1^c \in A_1$ e $P_1 Y_1^c = 0$. Além disso, $T_1^{-1} Y_1^c = Y_1^c$, portanto $Y_1^c \in \mathcal{I}_1$. Logo $Y_1 \in A_1 \cap \mathcal{I}_1$.

Seja $Y_2 = X_2$. A aplicação $\psi : Y_1 \to Y_2$ definida por

$$\psi(w_1, w_2, \ldots) = \sum_{k=1}^{\infty} \frac{w_k}{r^k}$$

satisfaz as condições (c),(d),(e),(f) da definição 5.1.2; portanto, $\mathcal{S}_1$ e $\mathcal{S}_2$ são sistemas isomorfos. (Para verificar as condições acima, deve-se usar o fato que os racionais r-ádicos constituem um conjunto denso enumerável em $[0;1)$ e, portanto, a classe

$C = \{[0;q]: q$ é um racional r-ádico$\}$

gera $A_2$ e determina $P_2$.) $\Delta$

EXEMPLO 5.2.3 - Sejam $r, \Omega, p, X_1, Y_1, X_2, A_2, P_2, \psi$ como no exemplo 5.2.2; seja $S_3 = (X_3, A_3, T_3, P_3)$ o sistema bilateral de Bernoulli gerado por

$$(\Omega, 2^\Omega, p).$$

A aplicação $\alpha: X_3 \to X_1 \times X_1$ definida por

$$\alpha(\ldots, w_{-2}, w_{-1}, w_0, w_1, w_2, \ldots) = ((w_1, w_2, \ldots), (w_0, w_{-1}, w_{-2}, \ldots))$$

é claramente bijetora. Seja $Y_3 = \alpha^{-1}(Y_1 \times Y_1)$. Temos que $Y_3^c \in A_3$, $P_3 Y_3^c = 0$, $Y_3^c \in I_3$, $Y_3 \in A_3 \cap I_3$.

Vamos montar um sistema $S_4 = (X_4, A_4, T_4, P_4)$ isomorfo a $S_3$. A aplicação $\alpha$ e o exemplo 5.2.2 sugerem que $(X_4, A_4, P_4)$ seja o espaço de probabilidade produto $(X_2, A_2, P_2) \times (X_2, A_2, P_2)$, isto é, $X_4 = [0;1)^2$, $A_4$ a $\sigma$-álgebra de Borel de $X_4$ e $P_4$ a medida de Lebesgue em $(X_4, A_4)$; sugerem também que $Y_4 = X_4$. Sejam

$$\alpha_1, \alpha_2: X_3 \to X_1$$

as aplicações definidas por

$$\alpha_1(\ldots, w_{-2}, w_{-1}, w_0, w_1, w_2, \ldots) = (w_1, w_2, \ldots)$$

$$\alpha_2(\ldots, w_{-2}, w_{-1}, w_0, w_1, w_2, \ldots) = (w_0, w_{-1}, w_{-2}, \ldots).$$

Continuando as sugestões anteriores, seja $\phi: Y_3 \to Y_4$ definida por

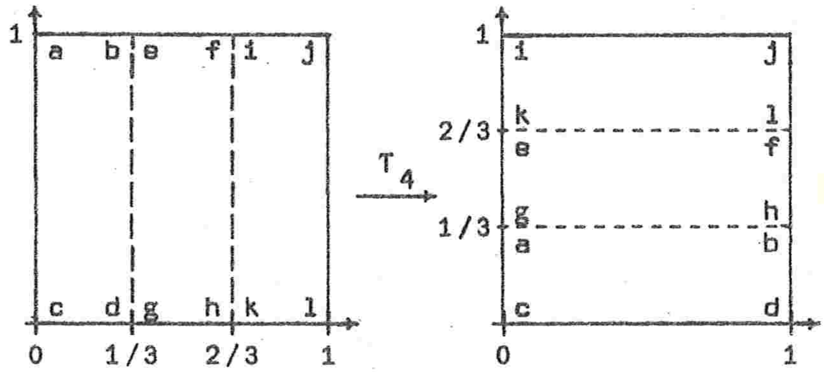
$$\phi y = (\psi\alpha_1 y, \psi\alpha_2 y).$$

A definição de $\phi$ e a condição (e) da defini ção 5.1.2 determinam $T_4: X_4 \to X_4$, cuja fórmula é

$$T_4(a,b) = \left( r\cdot a (\text{mód.}1),\ \frac{[r\cdot a]}{r} + \frac{b}{r} \right)$$

onde $[r\cdot a]$ é a parte inteira do produto $r\cdot a$.

O desenho abaixo mostra como atua a transformação $T_4$ no caso em que $r = 3$.




O movimento feito por um padeiro quando está fazendo a massa para um pão é semelhante à transformação $T_4$ no caso em que $r = 2$, daí $S_4$ ser denominado um SISTEMA DO PADEIRO.

$\Delta$

<u>OBSERVAÇÃO 5.2.2</u> - Em virtude da simetria e da transitividade dos isomorfismos, o sistema $S_1$ do exemplo 5.2.2 é isomorfo ao sistema $S_2$ do exemplo 5.1.1, pois cada um deles é isomorfo ao sistema r-ádico.

Seja k um inteiro positivo. Sejam $\Omega_1,\ldots,\Omega_k \in 2^\Omega$ subconjuntos não vazios de $\Omega$ e $n_1,\ldots,n_k \in I$. O conjunto

(6) $$B = \{x \in X : h_{n_1}(x) \in \Omega_1, \ldots, h_{n_k}(x) \in \Omega_k\} \in A$$

é chamado um RETÂNGULO. Os números $n_1, \ldots, n_k \in I$ são denominados COORDENADAS DO RETÂNGULO B. Anote-se:

$$C(B) = \{n_1, \ldots, n_k\}.$$

<u>NOTAÇÃO</u> - Vamos representar por $A_0$ a classe formada pelo conjunto vazio e por todos os subconjuntos de X que são uma união finita de retângulos disjuntos.

A classe $A_0$ é uma álgebra geradora do σ-álgebra $A$.

A medida P é definida de modo a satisfazer a seguinte propriedade: se A e B são retângulos tais que

$$C(A) \cap C(B) = \emptyset, \text{ então } P(A \cap B) = PA \cdot PB.$$

Se, na expressão (6), todos os conjuntos $\Omega_1, \ldots, \Omega_k$ são conjuntos unitários, então B é denominado RETÂNGULO ATÔMICO.

## 5.3 - "MIXING"

<u>DEFINIÇÃO 5.3.1</u> - Um sistema probabilístico $(X,A,T,P)$ é um SISTEMA "MIXING" se, e somente, se

$$\lim_{n\to\infty} P(A \cap T^{-n}B) = PA \cdot PB, \ \forall A,B \in A.$$

<u>TEOREMA 5.3.1</u> - Se $S = (X,A,T,P)$ é um sistema "mixing", então $S$ é um sistema ergódico.

<u>PROVA</u> - Seja $B \in A \cap I$. Então

$$T^{-n}B = B \text{ e } P(B \cap T^{-n}B) = P(B \cap B) = PB, \ \forall n \geq 1.$$

Por hipótese:

$$\lim_{n\to\infty} P(B \cap T^{-n}B) = PB \cdot PB.$$

Portanto:

$$PB = (PB)^2 \rightarrow \text{ou } PB = 0 \text{ ou } PB = 1 \rightarrow B \in \mathcal{T}. \quad \square$$

<u>TEOREMA 5.3.2</u> - Seja $S = (X,A,T,P)$ um sistema probabilístico preservativo inversível. São equivalentes:

i) $S$ é um sistema ergódico;

ii) $\lim\limits_{n\to\infty} \frac{1}{n+1} \cdot \sum\limits_{k=0}^{n} P(A \cap T^k B) = PA \cdot PB, \ \forall A,B \in A.$ $\square$

<u>TEOREMA 5.3.3</u> - Seja $S = (X,A,T,P)$ um sistema probabilístico preservativo inversível. São equivalentes:

i) $S$ é um sistema "mixing";

ii) $\lim\limits_{n\to\infty} P(A \cap T^{-n} A) = (PA)^2, \ \forall A \in A.$ $\square$

O conceito de sistema "mixing" já foi bastante generalizado. Definições, resultados e exemplos encontram-se em alguns dos recentes trabalhos de Chacon, S. Kakutani, N. A. Friedman, D. Ornstein, England e Martin, V.A.Rokhlin, entre outros. Uma descrição sumária desse desenvolvimento pode ser encontrada, junto as demonstrações dos teoremas 5.3.2 e 5.3.3, em [6] Petersen.

<u>LEMA 1</u>-Se $(X,A,\mu)$ é um espaço de medida tal que $\mu X < \infty$, então

$$|\mu A - \mu B| \leq \mu(A \Delta B), \ \forall A,B \in A.$$

<u>PROVA</u> - $|\mu A - \mu B| = |\mu(A-B) + \mu(A\cap B) - \mu(B-A) - \mu(B\cap A)| =$
$= |\mu(A-B) - \mu(B-A)| \leq \mu(A-B) + \mu(B-A) = \mu(A\Delta B).$ $\square$

<u>LEMA 2</u>-Seja $\varepsilon>0$. Se $x,y,a,b \in [0;1]$ são tais que

$$|x-a| < \varepsilon/2 \quad \text{e} \quad |y-b| < \varepsilon/2, \text{ então } |x \cdot y - a \cdot b| < \varepsilon.$$

PROVA - $|x\cdot y-a\cdot b| = |x\cdot y-x\cdot b+x\cdot b-a\cdot b| = |x\cdot(y-b)+(x-a)\cdot b| \leq$
$\leq |x\cdot(y-b)|+|(x-a)\cdot b| = x\cdot|y-b|+|x-a|\cdot b \leq |y-b|+|x-a| <$
$< \varepsilon/2 + \varepsilon/2 = \varepsilon.$ □

TEOREMA 5.3.4 - Sejam $S = (X, A, T, P)$ um sistema probabilístico preservativo, $F$ uma álgebra geradora de $A$. São equivalentes:

i) $S$ é um sistema "mixing";

ii) $\lim_{n\to\infty} P(A_0 \cap T^{-n}B_0) = PA_0 \cdot PB_0$, $\forall A_0, B_0 \in F$.

PROVA - (i→ii) Trivial.

(ii→i) Sejam $A, B \in A$. (Precisamos provar que, para cada $\varepsilon > 0$, existe um inteiro positivo $n_\varepsilon$ tal que

$$n \geq n_\varepsilon \rightarrow |P(A\cap T^{-n}B) - PA\cdot PB| < \varepsilon.$$

Seja $\varepsilon > 0$ arbitrário.

Pelo teorema D ([ 4 ] Halmos, pg.56), existem $A_0, B_0 \in F$ tais que

1) $P(A\Delta A_0) < \varepsilon/8$ e $P(B\Delta B_0) < \varepsilon/8$.

Por (1) e pela proposição 1.4:

2) $P(T^{-n}B\Delta T^{-n}B_0) = P[T^{-n}(B\Delta B_0)] = P(B\Delta B_0) < \varepsilon/8$, $\forall n\geq 1$.

Apenas pelo fato de $A, A_0, T^{-n}B, T^{-n}B_0$ serem subconjuntos de $X$:

3) $(A\cap T^{-n}B)\Delta(A_0\cap T^{-n}B_0) \subset (A\Delta A_0)\cup(T^{-n}B\Delta T^{-n}B_0)$, $\forall n\geq 1$.

Por (1), (2) e (3):

4) $P[(A\cap T^{-n}B)\Delta(A_0\cap T^{-n}B_0)] < \varepsilon/8 + \varepsilon/8 = \varepsilon/4$, $\forall n\geq 1$.

Por (4) e pelo lema 1:

5) $|P(A\cap T^{-n}B) - P(A_0\cap T^{-n}B_0)| < \varepsilon/4$, $\forall n\geq 1$.

Por hipótese, para cada $\delta > 0$, existe um inteiro positivo $n_\delta$ tal que

6) $n \geq n_\delta \rightarrow |P(A_0\cap T^{-n}B_0) - PA_0\cdot PB_0| < \delta.$

Por (5) e (6):

7) $n \geq n_{\varepsilon/4} \rightarrow |P(A \cap T^{-n}B) - PA_0 \cdot PB_0| < \varepsilon/4 + \varepsilon/4 = \varepsilon/2.$

Por (1) e pelo lema 1:

8) $|PA_0 - PA| < \varepsilon/8$ e $|PB_0 - PB| < \varepsilon/8.$

Por (8) e pelo lema 2:

9) $|PA_0 \cdot PB_0 - PA \cdot PB| < \varepsilon/4 < \varepsilon/2.$

Por (7) e (9):

$$n \geq n_{\varepsilon/4} \rightarrow |P(A \cap T^{-n}B) - PA \cdot PB| < \varepsilon/2 + \varepsilon/2 = \varepsilon. \ \square$$

<u>TEOREMA 5.3.5</u> - Os sistemas de Bernoulli são sistemas "mixing" (portanto ergódicos, pelo teorema 5.3.1).

<u>PROVA</u> - Vamos aplicar o teorema 5.3.4 na álgebra $A_0$: sejam $A, B \in A_0$. Inicialmente, vamos supor que A e B são dois retângulos não vazios. Seja $n_A$ a maior coordenada de A e seja $n_B$ a menor coordenada de B. Então

$$n > 1 + |n_A| + |n_B| \rightarrow C(A) \cap C(T^{-n}B) = \emptyset,$$

portanto

$$P(A \cap T^{-n}B) = PA \cdot PT^{-n}B = PA \cdot PB.$$

Logo,

$$\lim_{n \to \infty} P(A \cap T^{-n}B) = PA \cdot PB.$$

Se $A, B \in A_0$, $A \neq \emptyset$ e $B \neq \emptyset$, então

$$A = \bigcup_{i=1}^{m} A_i \quad \text{e} \quad B = \bigcup_{j=1}^{k} B_j,$$

onde os $A_i$'s, e os $B_j$'s, são retângulos (atômicos, se necessário) *disjuntos*. Portanto:

$$\lim_{n\to\infty} P(A\cap T^{-n}B) = \lim_{n\to\infty} \sum_{i=1}^{m} \sum_{j=1}^{k} P(A_i\cap T^{-n}B_j) =$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{k} \lim_{n\to\infty} P(A_i\cap T^{-n}B_j) = \sum_{i=1}^{m} \sum_{j=1}^{k} PA_i\cdot PB_j = PA\cdot PB.$$

Se $A,B\in A_0$, $A=\emptyset$ ou $B=\emptyset$, então é óbvio que

$$\lim_{n\to\infty} P(A\cap T^{-n}B) = PA\cdot PB. \qquad \square$$

Todo sistema probabilístico não ergódico não é "mixing" (teorema 5.3.1). Entretanto há também sistemas probabilísticos ergódicos que não são "mixing", como mostram os exemplos 5.3.1 e 5.3.2.

EXEMPLO 5.3.1 - Seja $r\geq 2$ um número inteiro. Sejam $X=\{1,2,\ldots,r\}$, $A$ de acordo com a observação 1.7, $T$ e $P$ definidas por

$$Tx = \begin{cases} x+1, & \text{se } x\neq r \\ 1, & \text{se } x=r \end{cases} \quad \text{e} \quad P\{x\} = \frac{1}{r}.$$

Como os únicos invariantes são $X$ e $\emptyset$, este sistema é ergódico. Por outro lado, fazendo $A=B=\{1\}$, tem-se

$$P(A\cap T^{-n}B) = \begin{cases} P\{1\} = \frac{1}{r}, & \text{se } n\in\{r,2\cdot r,3\cdot r,\ldots\} \\ P\emptyset = 0 & \text{, caso contrário.} \end{cases}$$

Portanto, não existe $\lim_{n\to\infty} P(A\cap T^{-n}A)$; logo, esse sistema não é "mixing". $\Delta$

EXEMPLO 5.3.2 - Seja $S_c$ o sistema considerado nos exemplos 3.5 e 3.6 com $c^k\neq 1$, para todo $k>0$.

Seja $A = B$ = semicircunferência superior. Como $\{c^n : n \geq 1\}$ é denso em X, há infinitos valores $c^n$ próximos de 1; para esses valores,

$$P(A \cap T^{-n}A) \cong PA = 1/2.$$

Portanto,

$$\lim_{n\to\infty} P(A \cap T^{-n}A),$$

se existir, não será $PA \cdot PA = 1/4$, logo esse sistema não é "mixing", apesar de ser ergódico. Δ

## 5.4 - SISTEMAS DE MARKOV

Nesta seção será apresentada uma classe de sistemas mais geral que a classe dos sistemas de Bernoulli. Eles podem ser introduzidos e interpretados como na seção 5.2; para evitar repetições, aqueles aspectos não serão explicitados novamente, mas é conveniente tê-los em mente.

Sejam $r, \Omega, 2^\Omega, Z, N, I, X, A, h_n, T$ como na seção 5.2. A medida P é, entretanto, diferente: há possibilidade de cada letra escolhida pela fonte influenciar na escolha da próxima letra da mensagem, mas a passagem do tempo (a transformação T) não altera a lei de probabilidade que rege o grande experimento. Em outras palavras,

(1) $$P\{h_n \in \Omega_0, h_{n+1} \in \Omega_1, \ldots, h_{n+k} \in \Omega_k\} =$$
$$= P\{h_{m+n} \in \Omega_0, h_{m+n+1} \in \Omega_1, \ldots, h_{m+n+k} \in \Omega_k\},$$
$$\forall n \in I,\ \forall k, m > 0,\ \forall \Omega_0, \Omega_1, \ldots, \Omega_k \in 2^\Omega.$$

Vamos introduzir a medida P. Seja $\pi: \Omega\times\Omega \to [0;1]$ uma MATRIZ ESTOCÁSTICA, isto é,

(2) $$\sum_{w\in\Omega} \pi(w_0,w) = 1, \ \forall w_0 \in \Omega.$$

Seja $p\colon \Omega \to [0;1]$ um VETOR ESTOCÁSTICO invariante sobre $\pi$, isto é

(3) $$\sum_{w\in\Omega} p(w) = 1$$

(4) $$\sum_{w\in\Omega} p(w)\cdot\pi(w,w_0) = p(w_0), \ \forall w_0 \in \Omega.$$

Seja $P$ a única medida em $(X,A)$ que satisfaz:

(5) $$P\{h_n = w_0, h_{n+1} = w_1, \ldots, h_{n+k} = w_k\} =$$
$$= p(w_0)\cdot\pi(w_0,w_1)\cdot\ldots\cdot\pi(w_{k-1},w_k),$$
$$\forall n \in I, \ \forall k>0, \ \forall w_0,w_1,\ldots,w_k \in \Omega.$$

Quando $I = Z$ $(I = N)$, o sistema $(X,A,T,P)$ é denominado SISTEMA BILATERAL (UNILATERAL) DE MARKOV GERADO POR $(\Omega, 2^\Omega, \pi, p)$.

PROPOSIÇÃO 5.4.1 - Os sistemas bilaterais (unilaterais) de Markov são sistemas preservativos. Em particular, satisfazem a expressão (1). □

OBSERVAÇÃO 5.4.1 - A estrutura de um sistema bilateral (unilateral) de Markov e a estrutura de um sistema bilateral (unilateral) de Bernoulli são as mesmas (isto é, ismorfas), quando ambos os alfabetos tem a mesma quantidade de letras.

PROPOSIÇÃO 5.4.2 - Os sistemas bilaterais (unilaterais) de Bernoulli são sistemas bilaterais (unila terais) de Markov.

PROVA - Basta escolher para $\pi$ a matriz estocástica definida por

$$\pi(w_1,w) = \pi(w_2,w) = p(w), \ \forall w, w_1, w_2 \in \Omega. \qquad \square$$

OBSERVAÇÃO 5.4.2 - Substituir $\Omega$ por outro alfabeto $\Delta$ com $r$ letras, estabelecer uma aplicação bijetora $\alpha$: $\Omega \to \Delta$, escolher a matriz estocástica $\pi'$ definida por

$$(6) \quad \pi'(\delta_1,\delta_2) = \pi(\alpha^{-1}\delta_1, \alpha^{-1}\delta_2), \ \forall \delta_1, \delta_2 \in \Delta$$

e escolher o vetor estocástico $p'$ definido por

$$(7) \quad p'(\delta) = p(\alpha^{-1}\delta), \qquad \forall \delta \in \Delta$$

é uma atividade talvez interessante, mas o sistema bilateral (unilateral) de Markov gerado por $(\Delta, 2^{\Delta}, \pi', p')$ é claramente isomorfo ao sistema bilateral (unilateral) de Markov gerado por $(\Omega, 2^{\Omega}, \pi, p)$. Em particular, se $\Omega = \Delta$ e $(\pi, p)$ for substituído por $(\pi', p')$ dados pelas expressões (6) e (7), o sistema bilateral (unilateral) de Markov gerado por $(\Omega, 2^{\Omega}, \pi', p')$ é claramente isomorfo ao sistema bilateral (unilateral) de Markov gerado por $(\Omega, 2^{\Omega}, \pi, p)$.

NOTAÇÃO - Para cada $n \geq 0$, seja $\pi^n$ a n-ésima potência da matriz $\pi$ e, para cada $(w_1, w_2) \in \Omega \times \Omega$, seja

$$q(w_1,w_2) = \lim_{n\to\infty} \frac{1}{n+1} \cdot \sum_{k=0}^{n} \pi^k(w_1,w_2).$$

(Omitiremos a prova que este limite existe para todo $(w_1,w_2)\in\Omega\times\Omega$, pois é uma decorrência imediata dos teoremas básicos sobre Cadeias de Markov).

As provas das proposições 5.4.3 e 5.4.4 podem ser encontradas em [ 2 ] Billingsley, pg. 31-33.

<u>PROPOSIÇÃO 5.4.3</u> - Seja $S$ um sistema de Markov tal que $p(w) > 0,\ \forall w\in\Omega$.

São equivalentes:

i) $S$ é um sistema ergódico;

ii) $\pi$ é irredutível;

iii) $q(w_1,w_2) = p(w_2),\ \forall\ w_1,w_2\in\Omega$;

iv) $q(w_1,w) = q(w_2,w),\ \forall w,w_1,w_2\in\Omega$;

v) $q(w_1,w_2) > 0,\ \forall w_1,w_2\in\Omega$. □

<u>PROPOSIÇÃO 5.4.4</u> - Seja $S$ um sistema de Markov tal que $p(w) > 0,\ \forall w\in\Omega$.

São equivalentes:

i) $S$ é um sistema "mixing";

ii) $\pi$ é irredutível e aperiódica;

iii) $\lim_{n\to\infty} \pi^n(w_1,w_2) = p(w_2),\ \forall w_1,w_2\in\Omega$. □

N.A.Friedman e D.Ornstein (Advanced in Mathematics, 1970) provaram que cada sistema bilateral de Markov "mixing" é isomorfo a um sistema bilateral de Bernoulli.

<u>OBSERVAÇÃO 5.4.3</u> - Para cada inteiro $r \geq 2$ fixado:

a) a estrutura obtida dos sistemas unilaterais de Bernoulli (ou de Markov) após a retirada de um conjunto enumerável apropriado (ver exemplo 5.2.2) é isomorfa à estrutura do sistema r-ádico: $X = [0;1)$ e $Tx = r\cdot x \pmod{1}$;

b) a estrutura obtida dos sistemas bilaterais de Bernoulli (ou de Markov) após a retirada de um conjunto não enumerável apropriado (ver exemplo 5.2.3) é isomorfa à "estrutura do padeiro": $X = [0;1)^2$ e $T =$ transformação do padeiro.

## 5.5 - PROCESSOS ESTOCÁSTICOS ESTACIONÁRIOS

Seja $(\Omega, \mathcal{A}, P)$ um espaço de probabilidade e seja $X = (X_0, X_1, \ldots)$ uma sequência de VARIÁVEIS ALEATÓRIAS, isto é, uma sequência de funções mensuráveis de $(\Omega, \mathcal{A})$ em $(R^1, \mathcal{B}^1)$, onde $R^1$ denota o conjunto dos números reais e $\mathcal{B}^1$ sua $\sigma$-álgebra de Borel.

Seja $(R^\infty, \mathcal{B}^\infty) = (R^1, \mathcal{B}^1) \times (R^1, \mathcal{B}^1) \times \ldots$. A sequência $X$ considerada como uma aplicação de $(\Omega, \mathcal{A})$ em $(R^\infty, \mathcal{B}^\infty)$, é mensurável.

A quádrupla ordenada $(\Omega, \mathcal{A}, P, X)$ é denominada um PROCESSO ESTOCÁSTICO. A probabilidade $P_X$ em $(R^\infty, \mathcal{B}^\infty)$ definida por

$$(1) \qquad P_X B = PX^{-1}(B) = P(X^{-1}B) \overset{\text{not.}}{=\!=} P\{X \in B\}, \quad \forall B \in \mathcal{B}^\infty$$

é denominada DISTRIBUIÇÃO DO PROCESSO ESTOCÁSTICO X (a terna $(\Omega, \mathcal{A}, P)$ será omitida, como é usual em Teoria das Probabilidades).

OBSERVAÇÃO 5.5.1 - Algumas vezes é conveniente definir um processo estocástico como uma sequência $X = (\ldots, X_{-1}, X_0, X_1, \ldots)$. Em consequência, defini-se
$(R^\infty, \mathcal{B}^\infty) = \ldots(R^1, \mathcal{B}^1)\times(R^1, \mathcal{B}^1)\times(R^1, \mathcal{B}^1)\times\ldots.$
Um desenvolvimento paralelo ao que será feito nesta seção pode ser feito neste caso. Para evitar repetições, tais processos estocásticos bilaterais serão omitidos.

Seja $S: R^\infty \to R^\infty$ a transformação definida por

(2) $S(x_0, x_1, \ldots) = (x_1, x_2, \ldots), \ \forall (x_0, x_1, \ldots) \in R^\infty.$

Temos:

(3) $(R^\infty, \mathcal{B}^\infty, S)$ é uma estrutura não inversível,

(4) para cada $n \geq 0$, $S^n X$ é um processo estocástico,

(5) $(R^\infty, \mathcal{B}^\infty, S, P_X)$ é um sistema probabilístico não inversível, denominado SISTEMA INDUZIDO POR X.

Dizemos que X é um PROCESSO ESTOCÁSTICO ESTACIONÁRIO se, e somente se, o sistema induzido por X é um sistema preservativo, ou equivalentemente,

(6) $P\{X_0, X_1, \ldots) \in B\} = P\{(X_1, X_2, \ldots) \in B\}, \ \forall B \in \mathcal{B}^\infty.$

Como as potências de um sistema preservativo são sistemas preservativos (proposição 1.4), temos o seguinte resultado: X é um processo estocástico estacionário se, e somente se,

(7) $P\{(X_0, X_1, \ldots) \in B\} = P\{(X_n, X_{n+1}, \ldots) \in B\}, \ \forall B \in \mathcal{B}^\infty$ e $\forall n \geq 0.$

Heuristicamente, processos estocásticos estacionários são aqueles que determinam, não importa quando as observações (mediações) são iniciadas, a mesma distribuição.

EXEMPLO 5.5.1 - Cada sistema unilateral de Markov é, convenientemente enfocado, um processo estocástico estacionário. Neste enfoque, deve-se:

a) tomar o alfabeto formado por alguns números reais,

b) considerar o grande experimento como espaço de probabilidade,

c) considerar as projeções como variáveis aleatórias.

Em particular, os sistemas unilaterais de Bernoulli são, enfocados desta maneira, processos estocásticos estacionários. Δ

EXEMPLO 5.5.2 - Se $X_0, X_1, \ldots$ são variáveis aleatórias independentes identicamente distribuídas, então $X = (X_0, X_1, \ldots)$ é um processo estocástico estacionário. Δ

O teorema 5.5.1 fornece uma maneira de se obter, de cada processo estocástico estacionário, uma infinidade de processos estocásticos estacionários.

TEOREMA 5.5.1 - Se X é um processo estocástico estacionário e g é uma função mensurável de $(R^\infty, \mathcal{B}^\infty)$ em $(R^1, \mathcal{B}^1)$, então é estacionário o processo estocástico $Y = (Y_0, Y_1, \ldots)$ definido por

$$Y_n = gS^n X, \quad \forall n \geq 0.$$

PROVA - Para cada $n \geq 0$, seja $g_n$: $R^\infty \to R^1$ definida por $g_n = gS^n$. Seja G: $R^\infty \to R^\infty$ definida por $G = (g_0, g_1, \ldots)$. As mensurabilidades das funções $g_0, g_1, \ldots$ implicam na mensurabilidade da transformação G. Além disso,

$$(*) \qquad Y = GX \quad \text{e} \quad GS = SG.$$

Para cada $B \in \mathcal{B}^\infty$, temos:

$$P_Y B = P\{Y\in B\} \overset{(*)}{=} P\{GX\in B\} = P\{X\in G^{-1}B\} = P_X(G^{-1}B) \overset{(hip)}{=\!=}$$
$$= P_X S^{-1}(G^{-1}B) = P\{\, GSX\in B\} \overset{(*)}{=} P\{SGX\in B\} \overset{(*)}{=} P\{SY\in B\}$$
$$= P\{Y\in S^{-1}B\} = P_Y(S^{-1}B) = (P_Y S^{-1})B,$$

o que prova que $(R^\infty, \mathcal{B}^\infty, S, P_Y)$ é um sistema preservativo. □

Se $X$ é um processo estocástico estacionário, então $(R^\infty, \mathcal{B}^\infty, S, P_X)$ é um sistema probabilístico preservativo por definição. O teorema 5.5.2 fornece uma maneira de se obter, de cada sistema probabilístico preservativo, uma infinidade de processos estocásticos estacionários.

<u>TEOREMA 5.5.2</u> - Se $T\colon \Omega \to \Omega$ é tal que $(\Omega, \mathcal{A}, T, P)$ é um sistema probabilístico preservativo e $f$ é uma função mensurável de $(\Omega, \mathcal{A})$ em $(R^1, \mathcal{B}^1)$, então é estacionário o processo estocástico

$$Y = (Y_0, Y_1, \ldots)$$

definido por

$$Y_n = fT^n, \ \forall n \geq 0.$$

<u>PROVA</u> - Para cada $B \in \mathcal{B}^\infty$, temos:

$$Y^{-1}B = \{w\in\Omega\colon (fw, fTw, \ldots) \in B\} \in \mathcal{A},$$
$$Y^{-1}S^{-1}B = \{w\in\Omega\colon (fTw, fT^2w, \ldots) \in B\} \in \mathcal{A},$$
$$w\in Y^{-1}S^{-1}B \leftrightarrow Tw\in Y^{-1}B, \text{ isto é, } Y^{-1}S^{-1}B = T^{-1}Y^{-1}B.$$

Portanto:

$$P_Y B = P(Y^{-1}B) \overset{(hip)}{=\!=} PT^{-1}(Y^{-1}B) = P(T^{-1}Y^{-1}B) =$$
$$= P(Y^{-1}S^{-1}B) = P_Y(S^{-1}B) = (P_Y S^{-1})B. \quad □$$

Para cada $n \geq 0$, seja $h_n\colon R^\infty \to R^1$ definida por

$$(8) \qquad h_n(x_0,x_1,\ldots) = x_n, \ \forall (x_0,x_1,\ldots) \in R^\infty.$$

Seja $H\colon R^\infty \to R^\infty$ definida por $H = (h_0,h_1,\ldots)$. É claro que $h_0$, $h_1,\ldots,H$ são mensuráveis. O processo estocástico

$$(R^\infty, \mathcal{B}^\infty, P_X, H)$$

é denominado PROCESSO ESTOCÁSTICO INDUZIDO POR X. O processo estocástico induzido por X é estacionário se, e somente se, o processo estocástico X é estacionário (ambos induzem o mesmo sistema).

Se $f \in L^1(\Omega,\mathcal{A},P)$, o número real

$$\int_\Omega f\cdot dP$$

é denominado ESPERANÇA DA FUNÇÃO f e anotado Ef.

Se $f \in L^1(\Omega,\mathcal{A},P)$ e $\mathcal{G}$ é uma sub-$\sigma$-álgebra de $\mathcal{A}$, a medida com sinal

$$(9) \qquad A \in \mathcal{G} \longmapsto \int_A f\cdot dP$$

é absolutamente contínua com respeito à restrição de P a $\mathcal{G}$. Pelo teorema Radon-Nikodyn, existe uma única função de $\Omega$ em $R^1$ (esta função é única no seguinte sentido: se duas funções tem as propriedades (10) e (11), então uma é P-versão da outra), anotada $E(f|\mathcal{G})$, satisfazendo:

$$(10) \qquad E(f|\mathcal{G}) \text{ é } \mathcal{G}\text{-mensurável}$$

$$(11) \qquad \int_A f\cdot dP = \int_A E(f|\mathcal{G})\cdot dP, \ \forall A \in \mathcal{G}.$$

A função $E(f|\mathcal{G})$ é denominada ESPERANÇA CONDICIONAL DA FUNÇÃO f DADO $\mathcal{G}$.

Se $f,f_0,f_1,\ldots$ são funções mensuráveis de $\Omega$ em $R^1$, então

$$\{w\in\Omega:\ f^*w \overset{\text{not}}{=\!=} \lim_{n\to\infty} f_n w \text{ existe}\}\in A$$

e, portanto,

$$\{f^*=f\} \overset{\text{not}}{=\!=} \{w\in\Omega: f^*w=fw\}\in A.$$

Se $P\{f^*=f\}=1$, dizemos que $f_0, f_1, \ldots$ CONVERGE QUASE CERTAMENTE a f e anotamos

$$f_n \xrightarrow{q.c.} f.$$

TEOREMA 5.5.3 - Se $T:\ \Omega \to \Omega$ é tal que $(\Omega,A,T,P)$ é um sistema probabilístico preservativo e

$$f\in L^1(\Omega,A,P),$$

então

$$\frac{1}{n+1}\cdot\sum_{k=0}^{n} fT^k \xrightarrow{q.c.} E(f|A\cap I). \qquad \square$$

TEOREMA 5.5.4 - Se $T:\ \Omega \to \Omega$ é tal que $(\Omega,A,T,P)$ é um sistema probabilístico preservativo ergódico e

$$f\in L^1(\Omega,A,P),$$

então

$$\frac{1}{n+1}\cdot\sum_{k=0}^{n} fT^k \xrightarrow{q.c.} Ef. \qquad \square$$

Os teoremas 5.5.3 e 5.5.4 são, respectivamente, adaptações dos teoremas 3.3 e 3.4 ao contexto desta seção. É conveniente reler: Obsevação 3.6, Definição 3.7, Proposição 3.8, Teorema 3.6 e Observação 3.7.

Queremos resultados análogos aos teoremas 5.5.3 e 5.5.4 aplicados ao contexto desta seção: um processo estocástico $(\Omega,A,P,X)$ e o seu sistema induzido $(R^\infty,B^\infty,S,P_X)$.

Seja $I^\infty = \{B\subset R^\infty:\ B = S^{-1}B\}$. Segue que:

(12) $$B \in \mathcal{I}^\infty \leftrightarrow B = S^{-n}B, \ \forall n \geq 0.$$

A classe $\mathcal{I}^\infty$ é uma $\sigma$-álgebra, logo $\mathcal{B}^\infty \cap \mathcal{I}^\infty$ é uma $\sigma$-álgebra. Pode-se pensar que $\mathcal{B}^\infty \cap \mathcal{I}^\infty$ é trivial (isto é, igual a $\{\phi, R^\infty\}$), por causa do modo como S atua, mas os exemplos 5.5.3 e 5.5.4 mostram que este pensamento é incorreto.

<u>EXEMPLO 5.5.3</u> - Para cada $B \in \mathcal{B}^\infty$ e para cada $n \geq 0$, seja

$$K(B,n) = \underbrace{R^1 \times \ldots \times R^1}_{n \text{ vezes}} \times B$$

Para cada $B \in \mathcal{B}^\infty$,

$$\hat{B} \overset{\text{not}}{=} \bigcup_{n=0}^{\infty} K(B,n) \in \mathcal{B}^\infty \cap \mathcal{I}^\infty.$$

Em particular, podemos considerar $A \in \mathcal{B}^1 - \{\phi, R^1\}$ e $B = A \times A \times \ldots$; escolhendo $a \in A$ e $b \notin A$ temos

$$(a,b,a,b,a,b,\ldots) \notin \hat{B},$$

logo $\hat{B} \neq R^\infty$. É claro que $\hat{B} \neq \emptyset$. $\Delta$

<u>EXEMPLO 5.5.4</u> - Para cada $A \in \mathcal{B}^1 - \{\phi, R^1\}$,

$$\bar{A} \overset{\text{not}}{=} \bigcap_{n=0}^{\infty} \bigcup_{k=n}^{\infty} h_k^{-1} A \in \mathcal{B}^\infty \cap \mathcal{I}^\infty. \quad \Delta$$

Podemos trabalhar com o sistema $(R^\infty, \mathcal{B}^\infty, S, P_X)$. Há, entretanto, muitos processos estocásticos relevantes nos quais utilizamos apenas parte da $\sigma$-álgebra $\mathcal{B}^\infty$ (por exemplo os processos estocásticos com espaço de estados enumerável ). Para que considerar os conjuntos $B \in \mathcal{B}^\infty$ tais que $Xw \notin B$, $\forall w \in \Omega$?

Para contornar esse problema vamos definir novos entes com as mesmas 3 palavras (invariante, trivial, ergódi co) e com os mesmos significados, porém, convém repetir, são outros entes.

Um conjunto $A \in \mathcal{A}$ é denominado um EVENTO. Dizemos que um evento A é um EVENTO TRIVIAL se, e somente se, ou $PA = 0$ ou $PA = 1$. Sejam

$$\mathcal{F}^{\infty} = \{B \in \mathcal{B}^{\infty}: X^{-1}B = (S^n X)^{-1}B, \ \forall n \geq 0\} \quad \text{e}$$

$$\mathcal{F} = X^{-1}\mathcal{F}^{\infty} = \{X^{-1}B: B \in \mathcal{F}^{\infty}\} \ .$$

A classe $\mathcal{F}^{\infty}$ é uma $\sigma$-álgebra tal que

$$\mathcal{I}^{\infty} \subset \mathcal{F}^{\infty} \subset \mathcal{B}^{\infty}. \tag{13}$$

A mensurabilidade de X implica, portanto, que $\mathcal{F}$ é uma sub-$\sigma$-álgebra de $\mathcal{A}$. Dizemos que um evento A é um EVENTO INVARIANTE se, e somente se, $A \in \mathcal{F}$. Dizemos que o processo estocástico X é um PROCESSO ESTOCÁSTICO ERGÓDICO se, e somente se, todo evento invariante é um evento trivial.

A expressão (13) implica: se X é um processo estocástico ergódico, então o sistema induzido por X é um sistema ergódico.

<u>TEOREMA 5.5.1'</u> - Se X é um processo estocástico ergódico e g é uma função mensurável de $(R^{\infty}, \mathcal{B}^{\infty})$ em $(R^1, \mathcal{B}^1)$, então é ergódico o processo estocástico $Y = (Y_0, Y_1, \ldots)$ definido por

$$Y_n = gS^n X, \ \forall n \geq 0.$$

<u>PROVA</u> - A transformação $G: R^{\infty} \to R^{\infty}$, definida na prova do teorema 5.5.1, é mensurável e além disso, $Y = GX$ e $GS = SG$. Seja $A \in \mathcal{F}_Y$; existe $B \in \mathcal{F}_Y^{\infty} \subset \mathcal{B}^{\infty}$ tal que

$$A = Y^{-1}B = (S^nY)^{-1}B, \ \forall n \geq 0$$

$$\therefore A = (GX)^{-1}B = (S^nGX)^{-1}B, \ \forall n \geq 0$$

$$\therefore A = X^{-1}(G^{-1}B) = (S^nX)^{-1}(G^{-1}B), \ \forall n \geq 0.$$

Como $G^{-1}B \in F_X^{\infty}$, $A \in F_X$ e, portanto, A é trivial. □

<u>TEOREMA 5.5.2'</u> - Se $T: \Omega \to \Omega$ é tal que $(\Omega, A, T, P)$ é um sistema probabilístico ergódico e f é uma função mensurável de $(\Omega, A)$ em $(R^1, B^1)$, então é ergódico o processo estocástico

$$Y = (Y_0, Y_1, \ldots)$$

definido por

$$Y_n = fT^n, \ \forall n \geq 0.$$

<u>PROVA</u> - Na prova do teorema 5.5.2, obtivemos:

$$Y^{-1}S^{-1}B = T^{-1}Y^{-1}B, \ \forall B \in B^{\infty}.$$

Seja $A \in F_Y$; existe $B \in F_Y^{\infty} \subset B^{\infty}$ tal que

$$A = Y^{-1}B = (S^nY)^{-1}B, \ \forall n \geq 0$$

$$\therefore A = Y^{-1}B = (SY)^{-1}B = Y^{-1}S^{-1}B = T^{-1}Y^{-1}B = T^{-1}A$$

$$\therefore A \in A \cap I \ \therefore A \in I \ \therefore \text{ ou } PA = 0 \text{ ou } PA = 1. \quad \square$$

<u>TEOREMA 5.5.3'</u> - Se X é um processo estocástico estacionário tal que $X_0 \in L^1(\Omega, A, P)$, então

$$\frac{1}{n+1} \cdot \sum_{k=0}^{n} X_k \xrightarrow{q.c.} E(X_0 | F).$$

<u>PROVA</u> - Ver [ 3 ] Breiman, é o teorema 6.28. □

<u>TEOREMA 5.5.4'</u> - Se X é um processo estocástico estacioná-

rio ergódico tal que $X_0 \in L^1(\Omega,A,P)$, então

$$\frac{1}{n+1} \sum_{k=0}^{n} X_k \xrightarrow{q.c.} EX_0.$$

PROVA - Decorre do teorema 5.5.3'. Como X é ergódico, as únicas funções $F$-mensuráveis são P-constantes (ver proposição 3.5). Por (11), fazendo $A = \Omega \in F$ segue que o P-valor de $E(X_0|F)$ é precisamente $EX_0$. □

Para cada $n \geq 0$, seja $B_n = (S^n X)^{-1} B^\infty$, isto é, $B_n$ é a menor σ-álgebra em relação a qual as variáveis aleatórias $X_n, X_{n+1}, \ldots$ são mensuráveis. É claro que: $B_n \subset A$, $\forall n \geq 0$. A σ-álgebra

$$C = \bigcap_{n=0}^{\infty} B_n \tag{14}$$

é denominada σ-ÁLGEBRA CAUDAL. Um evento A é um EVENTO CAUDAL se, e somente se, $A \in C$.

TEOREMA 5.5.5 - Cada evento invariante é um evento caudal, isto é, $F \subset C$.

PROVA - Seja $A \in F$; existe $B \in F^\infty \subset B^\infty$ tal que

$$A = X^{-1}B = (S^n X)^{-1}B, \ \forall n \geq 0$$

$$\therefore A \in (S^n X)^{-1} B^\infty = B_n, \ \forall n \geq 0 \ \therefore A \in C. \qquad \square$$

TEOREMA 5.5.6 - Se $X_0, X_1, \ldots$ são variáveis aleatórias independentes, então $X = (X_0, X_1, \ldots)$ é um processo estocástico ergódico.

PROVA - A independência assegura que os eventos caudais são triviais (é um teorema da Teoria das Probabilidades).

Pelo teorema 5.5.5, os eventos invariantes são triviais. □

<u>TEOREMA 5.5.7</u> - Se $X_0, X_1, \ldots$ são variáveis aleatórias independentes identicamente distribuídas, então $X = (X_0, X_1, \ldots)$ é um processo estocástico estacionário ergódico.

<u>PROVA</u> - Estacionário: trivial. Ergódico: teorema 5.5.6.□

<u>TEOREMA 5.5.8</u> - Se $X_0, X_1, \ldots$ são variáveis aleatórias independentes identicamente distribuídas e

$$X_0 \in \mathcal{L}^1(\Omega, A, P),$$

então

$$\frac{1}{n+1} \cdot \sum_{k=0}^{n} X_k \xrightarrow{q.c.} EX_0 .$$

<u>PROVA</u> - O resultado segue imediatamente dos teoremas 5.5.7 e 5.5.4'. □

<u>OBSERVAÇÃO 5.5.2</u> - O teorema 5.5.4' é uma generalização da LEI FORTE DOS GRANDES NÚMEROS DE KOLMOGOROV, como é conhecido o teorema 5.5.8. O teorema 5.5.9 é ainda mais geral: o teorema 5.5.4' é o caso particular que se obtém quando $g = h_0$.

<u>TEOREMA 5.5.9</u> - Se X é um processo estocástico estacionário ergódico e g é uma função mensurável de $(R^\infty, \mathcal{B}^\infty)$ em $(R^1, \mathcal{B}^1)$ tal que $gX \in \mathcal{L}^1(\Omega, A, P)$, então

$$\frac{1}{n+1} \cdot \sum_{k=0}^{n} g S^k X \xrightarrow{q.c.} E[gX].$$

PROVA - Pelos teoremas 5.5.1 e 5.5.1', $Y = (Y_0, Y_1, \ldots)$ definido por $Y_k = gS^k X$, $\forall k \geq 0$ é um processo estocástico estacionário ergódico. Como $Y_0 = gX$, o resultado segue do teorema 5.5.4'. □

Vamos usar um pouco do teorema 5.5.9: suponhamos que $X = (X_0, X_1, \ldots)$ é um processo estocástico estacionário ergódico. Se $A \in \mathcal{B}^1$, definindo $g: R^\infty \to R^1$ por

$$g(x_0, x_1, \ldots) = \begin{cases} 1, & \text{se } x_0 \in A \\ 0, & \text{caso contrário} \end{cases},$$

obtemos

$$(15) \qquad \frac{1}{n+1} \cdot \sum_{k=0}^{n} I_{\{X_k \in A\}} \xrightarrow{q.c.} P\{X_0 \in A\}.$$

Se $A_0, A_1 \in \mathcal{B}^1$, definindo $g: R^\infty \to R^1$ por

$$g(x_0, x_1, \ldots) = \begin{cases} 1, & \text{se } (x_0, x_1) \in A_0 \times A_1 \\ 0, & \text{caso contrário} \end{cases},$$

obtemos

$$(16) \quad \frac{1}{n+1} \cdot \sum_{k=0}^{n} I_{\{X_k \in A_0, X_{k+1} \in A_1\}} \xrightarrow{q.c.} P\{X_0 \in A_0, X_1 \in A_1\};$$

no caso em que $X_0, X_1, \ldots$ são variáveis aleatórias independentes identicamente distribuídas, temos

$$(16') \quad \frac{1}{n+1} \cdot \sum_{k=0}^{n} I_{\{X_k \in A_0, X_{k+1} \in A_1\}} \xrightarrow{q.c.} P\{X_0 \in A_0\} \cdot P\{X_0 \in A_1\}.$$

Se $f: R^1 \to R^1$ é mensurável e $fX_0 \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$, definindo $g: R^\infty \to R^1$ por $g(x_0, x_1, \ldots) = fx_0$, obtemos

$$(17) \qquad \frac{1}{n+1} \cdot \sum_{k=0}^{n} fX_k \xrightarrow{q.c.} E[fX_0].$$

Se $f: R^2 \to R^1$ é mensurável e $f(X_0, X_1) \in L^1(\Omega, A, P)$, definindo $g: R^\infty \to R^1$ por $g(x_0, x_1, \ldots) = f(x_0, x_1)$, obtemos

$$(18) \qquad \frac{1}{n+1} \cdot \sum_{k=0}^{n} f(X_k, X_{k+1}) \xrightarrow{q.c.} E[f(X_0, X_1)].$$

As interpretações de (15),(16),(17),(18) são muito interessantes. Naturalmente, o tereoma 5.5.9 pode ser utilizado para obter vários outros resultados particulares.

Os exemplos 5.5.5 e 5.5.6 são alguns dos muitos resultados obtidos através da teoria ergódica para particulares processos estocásticos estacionários.

<u>EXEMPLO 5.5.5</u> - Seja $X_1, X_2, \ldots$ uma sequência de variáveis aleatórias independentes identicamente distribuídas tomando valores no conjunto

$$Z = \{\ldots, -1, 0, 1, \ldots\}.$$

Para cada $w \in \Omega$, sejam

$$S_0 w = 0, \quad S_1 w = X_1 w, \quad S_2 w = X_1 w + X_2 w, \ldots .$$

Para cada $w \in \Omega$ e cada $n > 0$, seja

$$Y_n w = \operatorname{card}\{S_1 w, \ldots, S_n w\}.$$

Seja

$$p = P\left(\bigcap_{n=1}^{\infty} \{S_n \neq 0\}\right).$$

Aplicando os conteúdos desta seção (ver [3] Breiman, capítulo 6, seção 8), obtém-se:

a) $$\lim_{n \to \infty} \frac{EY_n}{n} = p;$$

b) $$\frac{Y_n}{n} \xrightarrow{q.c.} p. \qquad \Delta$$

EXEMPLO 5.5.6 - Sejam $X = (X_0, X_1, \ldots)$ um processo estocástico estacionário, $M \in \mathcal{B}^1$, $B_M = M \times R^1 \times R^1 \times \ldots \in \mathcal{B}^\infty$. Como o sistema $(R^\infty, \mathcal{B}^\infty, S, P_X)$ é preservativo, o teorema 2.2 assegura que quase todo ponto de $B_M$ é infinitamente-$B_M$-recorrente, isto é,

$$P_X\left(B_M \cap \left[\limsup_{n\to\infty} S^{-n} B_M\right]\right) = P_X B_M.$$

Disso resulta:

$$P(\{X_0 \in M\} \cap \{X_n \in M \text{ i.v.}\}) = P\{X_0 \in M\}.$$

Seja $\Omega_M = \{X_0 \in M\} \cap \{X_n \in M \text{ i.v.}\}$. Para cada $w \in \Omega_M$, sejam

$$\begin{array}{ll}
Y_1 w = \min\{n > 0 : X_n w \in M\} & T_1 w = Y_1 w \\
Y_2 w = \min\{n > Y_1 w : X_n w \in M\} & T_2 w = Y_2 w - Y_1 w \\
Y_3 w = \min\{n > Y_2 w : X_n w \in M\} & T_3 w = Y_3 w - Y_2 w \\
\vdots \qquad \vdots & \vdots \quad \vdots \quad \vdots
\end{array}$$

Para cada $n > 0$, a função $Y_n : \Omega_M \to \{1, 2, \ldots\}$ é denominada TEMPO DO n-ÉSIMO RETORNO A M e a função $T_n : \Omega_M \to \{1, 2, \ldots\}$ é denominada n-ÉSIMO TEMPO DE RECORRÊNCIA DE M (é fácil verificar que $Y_n$ e $T_n$ são mensuráveis).

Para o que se segue, é necessário supor que $P\,\Omega_M = P\{X_0 \in M\} > 0$. Seja $\mathcal{A}_M = \Omega_M \cap \mathcal{A} = \{\Omega_M \cap A : A \in \mathcal{A}\}$ e seja $P_M$ a probabilidade em $(\Omega_M, \mathcal{A}_M)$ definida por

$$P_M A = \frac{P A}{P \Omega_M}, \quad \forall A \in A_M.$$

Aplicando os conteúdos desta seção (ver [3] Breiman, capítulo 6, seção 9), obtém-se

a) se $P\left(\bigcup_{n=0}^{\infty} \{X_n \in M\}\right) = 1$, então

$$(\Omega_M, A_M, P_M, (T_1, T_2, \ldots))$$

é um processo estocástico estacionário e

$$ET_1 = EY_1 = \frac{1}{P\{X_0 \in M\}} ;$$

b) se X é ergódico, então

$$(\Omega_M, A_M, P_M, (T_1, T_2, \ldots))$$

é um processo estocástico ergódico. $\Delta$

Pelo que foi visto nesta seção, a Teoria Ergódica é um caminho para chegar a Lei Forte dos Grandes Números de Kolmogorov, bastante conhecida e bastante útil; mais ainda, através dela chegamos inclusive a uma generalização relativamente ampla dessa lei.

- $\square$ -

# CAPITULO 6

# ENTROPIA

O problema do isomorfismo é fundamental na Teoria Ergódica, como em qualquer outra teoria matemática. É uma das partes da Teoria Ergódica que oferece atualmente muitos problemas em aberto. O conceito de entropia,introduzido por Kolmogorov em 1958, permitiu que se obtivesse um bom desenvolvimento nessa área.

Neste capítulo, procura-se conceituar as idéias fundamentais, dar algumas interpretações e apresentar, sem demonstrações, os resultados mais relevantes envolvendo entropia; embora algumas demonstrações sejam bem interessantes, as que predominam são essencialmente técnicas ou então muito extensas.

Os tópicos apresentados neste capítulo podem ser encontrados (com as demonstrações) em [1] Ash, [2] Billingsley, [6] Petersen, [7] Reza e nos trabalhos publicados por Donald Ornstein em 1970 e 1971 na revista Advanced in Mathematics.

## 6.1 - INVARIANTES

Neste trabalho vimos que um sistema pode satisfazer, ou não, as seguintes propriedades: ser inversível, ser preservativo, ser probabilístico, ser incompressível,ser ergódico, ser "mixing". Há, é claro, outras propriedades que não foram apresentadas.

DEFINIÇÃO 6.1.1 - Uma propriedade p é denominada uma PROPRIEDADE INVARIANTE POR ISOMORFISMO (uma PIPI) se, e somente se,

$$(*)\quad \left\{\begin{matrix} S_1 \text{ é um sistema que satisfaz p} \\ S_2 \text{ é um sistema isomorfo a } S_1 \end{matrix}\right\} \rightarrow S_2 \text{ satisfaz p.}$$

As PIPI resolvem o problema de verificar se dois sistemas dados são isomorfos ou não,apenas no seguinte caso: um dos sistemas satisfaz uma PIPI e o outro não a satisfaz, segue que eles não são isomorfos.Outra utilidade das PIPI é o uso direto de sua definição (ver exemplo 6.1.3).

PROPOSIÇÃO 6.1.1 - As seguintes propriedades são PIPI: ser preservativo, ser probabilístico, ser incompressível,ser ergódico, ser "mixing".

□

EXEMPLO 6.1.1 - Pelo teorema 5.3.5, os sistemas de Bernoulli são "mixing". Pelo exemplo 5.3.2, as rotações no círculo unitário, quando c não é raiz da unidade, não são "mixing" (o mesmo argumento lá empregado mostra que,também nos casos em que c é raiz da unidade, esses sistemas não são "mixing"). Pela proposição 6.1.1, ser "mixing" é PIPI; portanto,nenhum sistema de Bernoulli, quer unilateral quer bilateral, é isomorfo a alguma rotação no círculo unitário (com a medida de Lebesgue normalizada, é claro). Δ

EXEMPLO 6.1.2 - Como ser ergódico é uma PIPI, nenhum sistema de Markov com matriz redutível é isomorfo a algum sistema de Markov com matriz irredutível, conforme a proposição 5.4.3.Δ

EXEMPLO 6.1.3 - Os sistemas de Bernoulli são "mixing" (teorema 5.3.5), ser "mixing" é PIPI (proposição 6.1.1), os sistemas r-ádicos e os sistemas do padeiro são isomorfos a sistemas de Bernoulli (exemplos 5.2.2 e 5.2.3). Por definição de PIPI, temos que os sistemas r-ádicos e os sistemas do padeiro são "mixing". Logo, pelo teorema 5.3.1, são ergódicos.Δ

OBSERVAÇÃO 6.1.1 - Ser inversível não é PIPI, isto é, existem sistemas isomorfos tais que um deles é inversível e outro não (ver exemplo 6.1.4). Os culpados pela existência desse fato estarrecedor são somente as definições de sistema inversível e de sistemas isomorfos (mudando estas definições, talvez ...).

EXEMPLO 6.1.4 - Consideremos os sistemas indicados nas tabelas abaixo:

| $X_1$ | a | b | c | d | e |
|---|---|---|---|---|---|
| $T_1$ | b | a | e | e | e |
| $\mu_2$ | 1/2 | 1/2 | 0 | 0 | 0 |

| $X_2$ | $\alpha$ | $\beta$ |
|---|---|---|
| $T_2$ | $\beta$ | $\alpha$ |
| $\mu_2$ | 1/2 | 1/2 |

Temos: $S_1$ não é inversível, $S_2$ é inversível, $S_1$ e $S_2$ são isomorfos. $\Delta$

DEFINIÇÃO 6.1.2 - Seja p uma PIPI e seja $\tilde{S}$ uma determinada classe de sistemas. A propriedade de p é denominada uma PIPI COMPLETA EM $\tilde{S}$ se e somente se,

$$(**)\qquad \left.\begin{array}{l} S_1 \in \tilde{S} \text{ e } S_1 \text{ satisfaz p} \\ \\ S_2 \in \tilde{S} \text{ e } S_2 \text{ satisfaz p} \end{array}\right\} \to S_1 \text{ e } S_2 \text{ são isomorfos.}$$

Se $S_1$ e $S_2$ são dois sistemas que satisfazem p, uma PIPI, não se pode dizer se eles são isomorfos ou não. Entretanto, se p é uma PIPI completa em alguma classe de sistemas contendo $\{S_1, S_2\}$, então eles são isomorfos.

EXEMPLO 6.1.5 - Ser ergódico *não* é uma PIPI completa na classe dos sistemas probabilísticos preservativos inversíveis. As rotações no círculo complexo unitário (com $c^k \neq 1$, $\forall k>0$) com a medida de Lebesgue normalizada são ergódicos (exemplo 3.6), assim como os sistemas do padeiro (exemplo 6.1.3).

Se uma rotação fosse isomorfa a algum sistema do padeiro, seria isomorfa a algum sistema de Bernoulli (pela transitividade do isomorfismo e pelo exemplo 5.2.3). Isto é uma contradição (ver exemplo 6.1.1). $\Delta$

São isomorfos os sistemas SBZ$\left[\frac{1}{2};\frac{1}{2}\right]$ e SBZ$\left[\frac{1}{3};\frac{1}{3};\frac{1}{3}\right]$? Pelo teorema 5.3.5, são ambos "mixing" (isto é, ambos satisfazem uma propriedade PIPI), mas isto não é suficiente para responder a pergunta acima, quer afirmativamente quer negativamente. Esta pergunta ficou durante longo tempo sem resposta. Em 1958, Kolmogorov introduziu um INVARIANTE NUMÉRI-

CO, hoje denominado entropia de um sistema e a resposta foi finalmente obtida: como os dois sistemas tem entropias diferentes, eles não são isomorfos.

Ao introduzir o conceito de entropia na Teoria Ergódica, Kolmogorov modificou alguns aspectos do conceito de entropia que Shannon já havia introduzido na Teoria da Informação. Esses dois conceitos são variações do conceito clássico de entropia pertencente à Física.

## 6.2 - DECOMPOSIÇÕES

Em toda esta seção, seja $(X,A)$ um espaço mensurável.

DEFINIÇÃO 6.2.1 - Seja $r \geq 1$ um número inteiro. Uma aplicação $D$ de $\{1,\ldots,r\}$ em $A$ é denominada uma DECOMPOSIÇÃO DE TAMANHO $r$ se, e somente se:

a) $i \neq j \rightarrow D_i \cap D_j = \emptyset$

b) $\bigcup_{i=1}^{r} D_i = \Omega$,

onde $D_i$ denota o ponto imagem da aplicação $D$ em $i \in \{1,\ldots,r\}$. Cada $D_i$ é denominado um ÁTOMO DA DECOMPOSIÇÃO $D$.

NOTAÇÃO - Uma decomposição $D$ é anotada como uma lista ordenada: $D = (D_1,\ldots,D_r)$.

Seria natural exigir que os átomos de uma decomposição fossem não vazios. Essa exigência foi omitida porque acarretaria complicações formais e também impediria certas considerações.

NOTAÇÃO - $\mathcal{D}^r$ denotará o conjunto de todas as decomposições de tamanho r. Quando necessário, será utilizado $\mathcal{D}^r(X,A)$.

NOTAÇÃO - $\mathcal{D} = \bigcup_{r=1}^{\infty} \mathcal{D}^r$.

OBSERVAÇÃO 6.2.1 - Uma decomposição de tamanho r desempenha o papel de uma discretização, isto é, a informação (determinação, precisão) máxima que se pode ter sobre qualquer ponto $x \in X$ é qual o átomo da decomposição a que x pertence (é como se todos os pontos pertencentes a um mesmo átomo fosse indistinguíveis). Assim, uma decomposição $D \in \mathcal{D}^r$ pode ser considerada uma função

$$x \in X \xmapsto{D} \sum_i i \cdot I_{D_i}(x).$$

NOTAÇÃO - Seja $D \in \mathcal{D}^r$ e seja $E \in \mathcal{D}^s$. $D \vee E$ denotará a decomposição $(D_1 \cap E_1, \ldots, D_1 \cap E_s, \ldots, D_r \cap E_1, \ldots, D_r \cap E_s) \in \mathcal{D}^{r+s}$.

OBSERVAÇÃO 6.2.2 - $D \vee E$ é praticamente a conjugação (produto cartesiano) das decomposições D e E. Conhecer o átomo de $D \vee E$ a que x pertence equivale conhecer o átomo de D a que x pertence e o átomo de E a que x pertence.

NOTAÇÃO - Sejam $D, E \in \mathcal{D}$. Anotaremos $E \leq D$ se, e somente se, D é um REFINAMENTO de E (isto é, cada átomo não vazio de E é uma união de átomos de D).

PROPOSIÇÃO 6.2.1 - Se $D, E \in \mathcal{D}$, então $E \leq D \vee E$. □

PROPOSIÇÃO 6.2.2 - Se T é uma transformação mensurável em $(X,\mathcal{A})$ e $D\in\mathcal{D}^r$, então

$$T^{-n}D \overset{\text{not}}{=\!=} (T^{-n}D_1,\ldots,T^{-n}D_r)\in\mathcal{D}^r,\ \forall n\geq 0.$$

□

DEFINIÇÃO 6.2.2 - Seja P uma probabilidade em $(X,\mathcal{A})$ e sejam $D,E\in\mathcal{D}$. D e E são DECOMPOSIÇÕES INDEPENDENTES se, e somente se,

$$P(D_i\cap E_j) = PD_i\cdot PE_j,\ \forall i \text{ e } \forall j.$$

EXEMPLO 6.2.1 - Sejam $X=\{1,\ldots,6\}$ e P a probabilidade que a cada evento unitário associa o número 1/6. Sejam

$D = (\{1,2,3\},\{4,5,6\})$
$E = (\{1,3,5\},\{2,4,6\})$
$F = (\{1,2\},\{3,4\},\{5,6\})$.

Temos:

$D\vee E = (\{1,3\},\{2\},\{5\},\{4,6\})$
$E\vee F = (\{1\},\{3\},\{5\},\{2\},\{4\},\{6\})$
$D\vee F = (\{1,2\},\{3\},\ \emptyset\ ,\ \emptyset\ ,\{4\},\{5,6\})$.

O único par de decomposições independentes é E e F, pois D e E e também D e F não o são.

Δ

OBSERVAÇÃO 6.2.3 - A independência de classes finitas e de classes infinitas de decomposições são definidas de maneira análoga a independência de classes finitas e de classes infinitas de sub-σ-álgebras.

É importante ter em mente que cada decomposição "quebra" X em uma quantidade *finita* de pedaços mensuráveis disjuntos dois a dois.

OBSERVAÇÃO 6.2.4 - Sejam r,s dois inteiros positivos tais que $r < s$. A correspondência

$$(D_1,\dots,D_2)\in\mathcal{D}^r \longmapsto (D_1,\dots,D_r,\emptyset,\dots,\emptyset)\in\mathcal{D}^s$$

é a imersão natural de $\mathcal{D}^r$ em $\mathcal{D}^s$. Esta imersão é consistente com todas as definições e notações introduzidas nesta seção. Desta forma, podemos escrever:

$$\mathcal{D}^1\subset\mathcal{D}^2\subset\dots\ .$$

$(\Omega)$ é denominada DECOMPOSIÇÃO TRIVIAL. É imediato verificar as seguintes propriedades:

$$D\in\mathcal{D}^1 \longleftrightarrow D = (\Omega)$$

$$D\vee(\Omega) = (\Omega)\vee D = D,\ \forall D\in\mathcal{D}$$

$$(\Omega) \le D,\ \forall D\in\mathcal{D}$$

$$T^{-n}(\Omega) = (\Omega),\ \forall n\ge 0$$

$(\Omega)$ e D são independentes, $\forall D\in\mathcal{D}$.

## 6.3 - ENTROPIA EM ESPAÇOS DE PROBABILIDADE

Em toda esta seção, seja $(X,A,P)$ um espaço de probabilidade.

Há dois enfoques para apresentar a noção de entropia de uma decomposição: um deles está no Apêndice I (aquele que é o mais "bonito"), o outro será ligeiramente explanado a seguir.

Deseja-se "medir" a aleatorização de uma decomposi ção. Por exemplo, decompondo um espaço de probabilidade de 3 modos:

$$(D_1, D_2) \text{ tal que } PD_1 = 0,50 \quad PD_2 = 0,50$$
$$(E_1, E_2) \text{ tal que } PE_1 = 0,75 \quad PE_2 = 0,25$$
$$(F_1, F_2) \text{ tal que } PF_1 = 0,99 \quad PF_2 = 0,01$$

sentimos que a decomposição D é a mais aleatória, E é relativamente aleatória, F tem pequena aleatoriedade (é quase determinística). Postula-se algumas propriedades desejadas para uma "medida" com essa finalidade e, baseando-se nesses postulados, procura-se determinar como ela é.

Esse enfoque pode ser encontrado em [7] Reza, onde chega-se a conclusão que a "medida" procurada é do tipo

$$(D_1, \ldots, D_r) \in \mathcal{D} \xmapsto{\lambda} c \cdot \sum_i [(PD_i) \cdot \log_a (PD_i)],$$

com $c<0$, $a>0$, $a \neq 1$ arbitrários. Escolhemos $c=-1$ e $a=2$. Os motivos desta escolha se encontram no Apêndice I, cuja leitura, embora interessante e agradável, não é necessária. Entretanto é indispensável conhecer as convenções e notações que lá se encontram.

<u>CONVENÇÃO 3</u> - Em todo este trabalho log t denotará $\log_2 t$.

<u>EXEMPLO 6.3.1</u> - Consideremos seis decomposições de tamanho 2 cujas probabilidades estão dados na tabela abaixo.

| i | $PD_1$ | $PD_2$ | $\lambda = \eta(D_1) + \eta(D_2)$ |
|---|---|---|---|
| 1 | 0,50 | 0,50 | 1,000 |
| 2 | 0,75 | 0,25 | 0,811 |
| 3 | 0,99 | 0,01 | 0,080 |
| 4 | 0,51 | 0,49 | 0,999 |
| 5 | 0,76 | 0,24 | 0,795 |
| 6 | UM | ZERO | ZERO |

Observemos que as entropias calculadas usando a fórmula

$$\sum_i \eta(PD_i)$$

estão em perfeita harmonia com o senso comum: a 6.ª decomposição é determinística (logo não há aleatoriedade), a 2.ª e a 5.ª têm praticamente a mesma aleatoriedade, a 1.ª é a mais aleatória de todas, etc. Δ

<u>EXEMPLO 6.3.2</u> - Consideremos o exemplo 6.2.1. Temos:

$$\lambda(D) = \lambda(E) = 2\cdot\eta(1/2) = 2\times 0,500 = 1,000$$
$$\lambda(F) = 3\cdot\eta(1/3) = 3\times 0,528 = 1,584$$
$$\lambda(D\vee E) = \lambda(D\vee F) = \eta(1/3)+\eta(1/6)+\eta(1/6)+\eta(1/3)$$
$$= 0,528+0,435+0,435+0,528 = 1,926$$
$$\lambda(E\vee F) = 6\cdot\eta(1/6) = 6\times 0,435 = 2,610.$$

Observemos que, quando refinamos uma decomposição (ver proposição 6.2.1), a entropia não decresceu (até aumentou). Δ

<u>DEFINIÇÃO 6.3.1</u> - O número real extendido não negativo

$$\lambda = \sup\{\lambda(D):\ D\in\mathcal{D}\}$$

é denominado ENTROPIA DO ESPAÇO DE PROBABILIDADE.

<u>EXEMPLO 6.3.3</u> - Sejam $X = [0;1)$ e $P$ a medida de Lebesgue. Para cada $k \geq 1$, seja $D(k)$ a decomposição

$$([0;1/k),[1/k;2/k),\ldots,[(k-1)/k;1)).$$

$$\lambda(D(k)) = k\cdot\eta(1/k) = k\cdot[(-\log 1/k)\cdot 1/k] = \log k$$

$$\lambda \geq \sup\{\lambda(D(k)):k\geq 1\} = +\infty \ \therefore\ \lambda = +\infty.\ \Delta$$

OBSERVAÇÃO 6.3.1 - A entropia de uma decomposição $D \in \mathcal{D}^r$ não depende propriamente da lista $(D_1, \ldots D_r)$, mas sim da lista $(t_1, \ldots, t_r)$, onde $t_i = PD_i$, $\forall i$. Portanto:

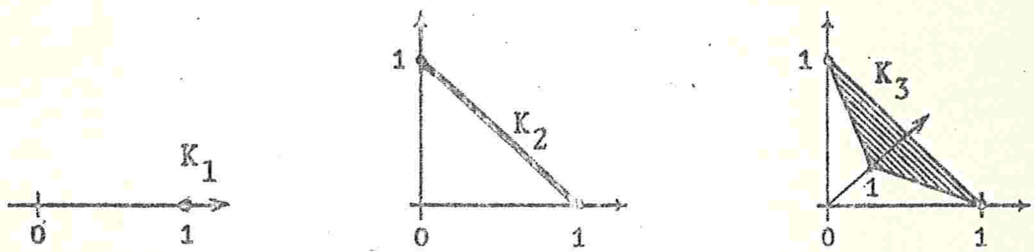
a) $t_i \geq 0$, $\forall i$;

b) $\sum_i t_i = 1$.

Assim, quando for mencionada uma lista deste tipo, fica subentendida alguma decomposição de algum espaço de probabilidade que "forneceu" a lista.

Para cada $r \geq 1$, seja $K_r = \{(t_1, \ldots, t_r) \in R^r : t_1 + \ldots + t_r = 1$ e $t_i \geq 0$, $\forall i\}$ e seja $g_r : K_r \longrightarrow [0;+\infty)$ a função definida por

$$g_r(t_1, \ldots, t_r) = \sum_i \eta(t_i).$$




PROPOSIÇÕES 6.3 - Seja $r$ um inteiro positivo.

1. $g_r \geq 0$.
2. $K_r$ é convexo (ver definição 4.4).
3. $g_r(t) = 0 \leftrightarrow t \in \{(1,0,\ldots,0), (0,1,\ldots,0), \ldots, (0,0,\ldots,1)\}$.
4. $g_r$ é simétrica.
5. $g_r$ é contínua.
6. $K_r$ é compacto.
7. $\sup\{g_r(t) : t \in K_r\} = \log r$.
8. $g_r(t) = \log r \leftrightarrow t = (1/r, 1/r, \ldots, 1/r)$. □

PROPOSIÇÃO 6.3.9 - Para cada $r \geq 1$, a aplicação

$$d_r: \mathcal{D}^r \times \mathcal{D}^r \longrightarrow [0;+\infty)$$

definida por

$$d_r(D,E) = \sum_i P(D_i \Delta E_i)$$

é uma pseudo-métrica. □

PROPOSIÇÃO 6.3.10 - Para cada $r \geq 1$, sejam $(\mathcal{D}^r_*, d^*_r)$ o espaço métrico obtido de $(\mathcal{D}^r, d_r)$ e $\pi_r: \mathcal{D}^r_* \longrightarrow K_r$ definida por

$$\pi_r(D_1,\ldots,D_r) = (PD_1,\ldots,PD_r).$$

A aplicação $\pi_r$ é contínua. □

É fácil obter exemplos onde $\pi_r$ não é sobrejetora: basta escolher espaços de probabilidade discretos. Em espaços de probabilidade não atômicos, $\pi_r$ é claramente sobrejetora (a recíproca é falsa). O exemplo 6.3.4 mostra um caso em que $\pi_r$ não é injetora.

EXEMPLO 6.3.4 - Sejam $B = \{1,2,3\}$, $X = B \times B$, $P$ a probabilidade que atribui a cada evento unitário o valor 1/9. Sejam

$$D = (\{1\}\times B, \{2\}\times B, \{3\}\times B) \quad \therefore \quad \pi_3(D) = (\tfrac{1}{3},\tfrac{1}{3},\tfrac{1}{3})$$

$$E = (B\times\{1\}, B\times\{2\}, B\times\{3\}) \quad \therefore \quad \pi_3(E) = (\tfrac{1}{3},\tfrac{1}{3},\tfrac{1}{3})$$

Como

$$d_3(D,E) = \sum_i P(D_i \Delta E_i) = \frac{4}{9} + \frac{4}{9} + \frac{4}{9} \neq 0,$$

temos $D^* \neq E^*$ e, portanto, $\pi_3$ não é injetora. Δ

Considerando a entropia como uma função $\lambda_r$ de $\mathcal{D}_*^r$ em $[0;+\infty)$, obtem-se as proposições 6.3.11 e 6.3.12.

PROPOSIÇÕES 6.3 -Seja r um inteiro positivo.

11. $\lambda_r = g_r \pi_r$.

12. $\lambda_r$ é contínua. □

No exemplo 6.2.1, as únicas decomposições relevantes são as de tamanho seis, pois as de tamanho menor que seis também estão em $\mathcal{D}^6$ (conforme a observação 6.2.4). As proposições 6.3.7, 6.3.8 e 6.3.11 asseguram que

$$\lambda = \lambda_6(E \vee F) = \log 6 \cong 2{,}610.$$

## 6.4 - ENTROPIA CONDICIONAL

Em toda esta seção, seja $(X,\mathcal{A},P)$ um espaço de probabilidade.

Para cada $B \in \mathcal{A}$ tal que $PB > 0$, $P(\cdot|B)$ denota a probabilidade em $(X,\mathcal{A})$ definida por

$$P(A|B) = \frac{P(B \cap A)}{PB}.$$

Heuristicamente, o experimento $(X,\mathcal{A},P)$ é realizado e revela-se apenas que ocorreu B; não se dispõe de mais nenhuma informação sobre o ponto ocorrido. Nessas condições, a probabilidade P é substituída por $P(\cdot|B)$.

DEFINIÇÃO 6.4.1 - Para cada $B \in \mathcal{A}$ tal que $PB > 0$ e para cada $E \in \mathcal{D}$, a variável aleatória $\phi_B(E,\cdot)$ definida por

$$\phi_B(E,x) = \sum_j \phi[P(E_j|B)] \cdot I_{E_j}(x)$$

é denominada INDETERMINAÇÃO DA DECOMPOSIÇÃO E DADO O EVENTO B.

DEFINIÇÃO 6.4.2 - Nas condições da definição 6.4.1, o número real não negativo

$$\lambda_B(E) = E_{P(\cdot|B)}[\phi_B(E,\cdot)] = \sum_j \eta[P(E_j|B)]$$

é denominado ENTROPIA DA DECOMPOSIÇÃO E DADO O EVENTO B.

DEFINIÇÃO 6.4.3 - Para cada $B \in A$ tal que $PB > 0$, o número real extendido não negativo

$$\lambda_B = \sup\{\lambda_B(E) : E \in \mathcal{D}\}$$

é denominado ENTROPIA DO ESPAÇO DE PROBABILIDADE DADO O EVENTO B.

CONVENÇÃO 4 - $P(A|B) = 0$, quando $PB = 0$.

CONVENÇÃO 5 - $\lambda_B(E) = +\infty$, quando $PB = 0$.

DEFINIÇÃO 6.4.4 - Para cada $D \in \mathcal{D}$ e cada $E \in \mathcal{D}$ a variável aleatória $\phi(E,\cdot|D)$ definida por

$$\phi(E,x|D) = \sum_i \left[\sum_j \phi[P(E_j|D_i)] \cdot I_{E_j}(x)\right] \cdot I_{D_i}(x)$$

é denominada INDETERMINAÇÃO DA DECOMPOSIÇÃO E DADA A DECOMPOSIÇÃO D.

DEFINIÇÃO 6.4.5 - Nas condições da definição 6.4.4, o número real não negativo

$$\lambda(E|D)=E[\phi(E,\cdot|D)]=\sum_i\sum_j\phi[P(E_j|D_i)]\cdot P(E_j\cap D_i)$$

$$=\sum_i PD_i\cdot\left(\sum_j\eta[P(E_j|D_i)]\right)=\sum_i\lambda_{D_i}(E)\cdot PD_i$$

é denominado ENTROPIA DA DECOMPOSIÇÃO E DADA A DECOMPOSIÇÃO D.

Heuristicamente, o experimento probabilístico $(X,A,P)$ será realizado e, após isso feito, será revelado apenas qual o átomo de D a que x, o ponto ocorrido, pertence. Isto não basta para determinar qual o átomo de E a que x pertence. A aleatorização da decomposição E, após revelado qual o $D_i$ a que x pertence, é expressa por $\lambda_{D_i}(E)$. Assim,

$$\lambda(E|D)=\sum_i\lambda_{D_i}(E)\cdot PD_i$$

é a média destas aleatorizações.

<u>DEFINIÇÃO 6.4.6</u> - Para cada $D\in\mathcal{D}$, o número real extendido não negativo

$$\sup\{\lambda(E|D):E\in\mathcal{D}\}$$

é denominado ENTROPIA DO ESPAÇO DE PROBABILIDADE DADA A DECOMPOSIÇÃO D.

<u>OBSERVAÇÃO 6.4.1</u> - A entropia de uma decomposição $E\in\mathcal{D}^s$ dada uma decomposição $D\in\mathcal{D}^r$ não depende propriamente das listas $(E_1,\ldots,E_s)$ e $(D_1,\ldots,D_r)$, mas sim da matriz

$$\begin{bmatrix} t_{11} & t_{12}\ldots & t_{1s} \\ t_{21} & t_{22}\ldots & t_{2s} \\ \vdots & \vdots & \vdots \\ t_{r1} & t_{r2}\ldots & t_{rs} \end{bmatrix},\text{onde } t_{ij}=P(E_j\cap D_i),\ \forall i \text{ e } \forall j.$$

Portanto

a) $t_{ij} \geq 0$, $\forall i$ e $\forall j$;

b) $\sum_i \sum_j t_{ij} = 1$.

Assim, ao ser mencionada tal matriz, fica subentendido um para de decomposições que "forneceu" a matriz.

EXEMPLO 6.4.1 -

| | $E_1$ | $E_2$ | $E_3$ | |
|---|---|---|---|---|
| $D_1$ | 0,42 | 0,18 | 0 | 0,6 |
| $D_2$ | 0,28 | 0,12 | 0 | 0,4 |
| | 0,7 | 0,3 | 0 | 1 |

D e E são decomposições independentes.

$$\lambda(D \vee E) = \eta(0,42) + \eta(0,18) + \eta(0,28) + \eta(0,12) = 1,852.$$

$$\lambda(D) = \eta(0,6) + \eta(0,4) = 0,971.$$

$$\lambda(E) = \eta(0,7) + \eta(0,3) = 0,881.$$

$$\lambda(D|E) = 0,7\left[\eta\left(\frac{0,42}{0,7}\right) + \eta\left(\frac{0,28}{0,7}\right)\right] + 0,3\left[\eta\left(\frac{0,18}{0,3}\right) + \eta\left(\frac{0,12}{0,3}\right)\right] = 0,971.$$

$$\lambda(E|D) = 0,6\left[\eta\left(\frac{0,42}{0,6}\right) + \eta\left(\frac{0,18}{0,6}\right)\right] + 0,4\left[\eta\left(\frac{0,28}{0,4}\right) + \eta\left(\frac{0,12}{0,4}\right)\right] = 0,881.$$

Observemos que

$$\lambda(D) = \lambda(D|E) = \lambda_{E_j}(D), \ \forall j$$

$$\lambda(E) = \lambda(E|D) = \lambda_{D_i}(E), \ \forall i$$

$$\lambda(D \vee E) = \lambda(D) + \lambda(E). \qquad \Delta$$

EXEMPLO 6.4.2 -

| | $E_1$ | $E_2$ | $E_3$ | |
|---|---|---|---|---|
| $D_1$ | 0,10 | 0,20 | 0,15 | 0,45 |
| $D_2$ | 0,20 | 0,10 | 0,15 | 0,45 |
| $D_3$ | 0,10 | 0 | 0 | 0,10 |
| | 0,40 | 0,30 | 0,30 | 1 |

D e E não são independentes.

$$\lambda(D \vee E) = 3 \cdot \eta(0{,}10) + 2 \cdot \eta(0{,}15) + 2 \cdot \eta(0{,}20) = 2{,}746.$$

$$\lambda(D) = \eta(0{,}10) + 2 \cdot \eta(0{,}45) = 1{,}368.$$

$$\lambda(E) = \eta(0{,}40) + 2 \cdot \eta(0{,}30) = 1{,}571.$$

$$\lambda(D|E) = 0{,}40\left[2 \cdot \eta\left(\frac{0{,}10}{0{,}40}\right) + \eta\left(\frac{0{,}20}{0{,}40}\right)\right] + 0{,}30\left[\eta\left(\frac{0{,}20}{0{,}30}\right) + \eta\left(\frac{0{,}10}{0{,}30}\right)\right] + 0{,}30\left[2 \cdot \eta\left(\frac{0{,}15}{0{,}30}\right)\right] = 1{,}175.$$

$$\lambda(E|D) = 0{,}45\left[\eta\left(\frac{0{,}10}{0{,}45}\right) + \eta\left(\frac{0{,}20}{0{,}45}\right) + \eta\left(\frac{0{,}15}{0{,}45}\right)\right] + 0{,}45\left[\eta\left(\frac{0{,}20}{0{,}45}\right) + \eta\left(\frac{0{,}10}{0{,}45}\right) + \eta\left(\frac{0{,}15}{0{,}45}\right)\right] + 0{,}10 \cdot \eta\left(\frac{0{,}10}{0{,}10}\right) = 1{,}377.$$

Observemos que

$$\lambda(D|E) < \lambda(D)$$
$$\lambda(E|D) < \lambda(E)$$
$$\lambda(D\vee E) < \lambda(D) + \lambda(E). \qquad \Delta$$

PROPOSIÇÕES 6.4 - Sejam $D,E,F \in \mathcal{D}$.

1. $\lambda(E|(\Omega)) = \lambda(E)$.
2. $E \leq D \longrightarrow \lambda(E|D) = 0$.
3. $\lambda(D\vee E) = \lambda(D)+\lambda(E|D)$. (FÓRMULA FUNDAMENTAL)
4. $\lambda(D\vee E|F) = \lambda(D|F) + \lambda(E|D\vee F)$.
5. $E \leq D \longrightarrow \lambda(E) \leq \lambda(D)$.
6. $E \leq D \longrightarrow \lambda(E|F) \leq \lambda(D|F)$.
7. $\lambda(E|D) \leq \lambda(E)$.
8. $E \leq D \longrightarrow \lambda(F|D) \leq \lambda(F|E)$.
9. $\lambda(D\vee E) \leq \lambda(D) + \lambda(E)$.
10. $\lambda(D\vee E|F) \leq \lambda(D|F) + \lambda(E|F)$.
11. D e E independentes, $PD_i > 0 \rightarrow \lambda_{D_i}(E) = \lambda(E)$.
12. D e E independentes $\longrightarrow \lambda(E|D) = \lambda(E)$.
13. D e E independentes $\rightarrow \lambda(D\vee E) = \lambda(D)+\lambda(E)$.

$\square$

Em [2] Billingsley, pg. 80-81, há dois teoremas que tornam precisa a seguinte idéia: se $d_r(D,E)$ é pequena, então $\lambda(E|D)$ é pequena.

PROPOSIÇÃO 6.4 - Seja T uma transformação em X tal que $(X,\mathcal{A},T,P)$ é um sistema probabilístico preservativo. Sejam $D,E \in \mathcal{D}$.

14. $\lambda(T^{-n}D) = \lambda(D)$, $\forall n \geq 0$.
15. $\lambda(T^{-n}E|T^{-n}D) = \lambda(E|D)$, $\forall n \geq 0$. $\square$

PROPOSIÇÃO 6.4.16 - A aplicação $\delta : \mathcal{D}\times\mathcal{D} \longrightarrow [0;+\infty)$ definida por

$$\delta(D,E) = \lambda(E|D) + \lambda(D|E)$$

é uma pseudo-métrica. □

PROPOSIÇÃO 6.4.17 - Seja r um inteiro positivo. Sejam $D,E \in \mathcal{D}^r$. Então:

$$d_r(D,E) = 0 \longrightarrow \delta(D,E) = 0. \quad \square$$

Em relação à proposição 6.4.17, é interessante observar que $d_r$ considera a ordem das decomposições, ao passo que $\delta$ é insensível a elas.

## 6.5 - ENTROPIA EM SISTEMAS

Em todas esta seção, seja $S = (X,A,T,P)$ um sistema probabilístico.

Em nossas considerações, um ponto $x \in X$ representa algum estado do sistema ao tempo zero e $T^n x$ representa o estado do sistema ao tempo n quando, ao tempo zero, o estado do sistema é representado por $x \in X$.

O número $PT^{-n}A$, $A \in A$, é interpretado como a probabilidade de, ao tempo n, o estado do sistema pertencer a A.

Uma decomposição $D \in \mathcal{D}^r = \mathcal{D}^r(X,A)$ pode ser "aplicada" ao sistema em qualquer instante n, indicando qual o átomo de D a que o estado do sistema ao tempo n pertence. Como

$$T^n x \in D_i \leftrightarrow x \in T^{-n} D_i \leftrightarrow \sum_j j \cdot I_{T^{-n}D_j}(x) = i,$$

"aplicar" a decomposição D ao tempo n equivale a "aplicar" a decomposição $T^{-n}D$ (ver proposição 6.2.2) ao tempo zero.

OBSERVAÇÃO 6.5.1 - Para cada n, seja $D(n,\cdot)$ a função

$$x \in X \xmapsto{D(n,\cdot)} \sum_i i \cdot I_{T^{-n}D_i}(x),$$

conforme a observação 6.2.1. As probabilidades induzidas por $D(n,\cdot)$ no espaço mensurável

$$(\{1,\ldots,r\}, 2^{\{1,\ldots,r\}})$$

pode depender de n; entretanto, quando $S$ é preservativo, isso não acontece (a recíproca é falsa).

<u>OBSERVAÇÃO 6.5.2</u> - As decomposições $D, T^{-1}D, \ldots, T^{-n}D$ não são necessariamente independentes. Em muitos casos, conhecer o D-átomo a que o estado do sistema pertence ao tempo k ajuda a determinar (ou até determina) qual o D-átomo a que o estado do sistema pertence ao tempo k+1, e vice-versa.

<u>DEFINIÇÃO 6.5.1</u> - Seja $D \in \mathcal{D}$. Para cada $n \geq 0$, a decomposição

$$D^n = D \vee T^{-1}D \vee \ldots \vee T^{-n}D$$

é denominada n-ÉSIMA POTÊNCIA DA DECOMPOSIÇÃO D.

$D^n$ é a conjugação (produto cartesiano) das decomposições $D$, $T^{-1}D$, ..., $T^{-n}D$, ou equivalentemente, das variáveis aleatórias $D(0,\cdot)$, $D(1,\cdot)$, ..., $D(n,\cdot)$, que não precisam ser nem independentes nem identicamente distribuídas, conforme as observações 6.5.1 e 6.5.2. $\lambda(D^n)$ é a aleatorização existente na decomposição $D^n$ e

$$\frac{1}{n+1} \cdot \lambda(D^n)$$

é a aleatorização média por repetição da decomposição D no "intervalo" $\{0,1,\ldots,n\}$.

<u>DEFINIÇÃO 6.5.2</u> - Seja $D\in\mathcal{D}$. O número real não negativo

$$\Lambda(D) = \limsup_{n\to\infty} \frac{1}{n+1}\cdot\lambda(D^n)$$

é denominado ENTROPIA DA DECOMPOSIÇÃO D.

<u>EXEMPLO 6.5.1</u> - Seja $\Omega=\{1,\ldots,r\}$ e p uma probabilidade em $(\Omega,2^\Omega)$. Vamos representar $p\{i\}$ por $p_i$, $\forall i\in\Omega$. Seja $(X,A,T,P)$ o sistema (bilateral ou unilateral, aqui não importa) de Bernoulli gerado por $(\Omega,2^\Omega,p)$.

Seja D a DECOMPOSIÇÃO DO TEMPO 1, isto é,

$$D_i = \{x\in X:\ h_1(x)=i\}.$$

$\lambda(D)=\sum_i \eta(p_i)$ representa a aleatorização com que a fonte emite uma letra isolada em qualquer tempo n (aqui $D(0,\cdot), D(1,\cdot),\ldots$ são variáveis aleatórias independentes identicamente distribuídas). A proposição 6.4.13 assegura que $\lambda(D^n)=(n+1)\cdot\lambda(D)$, $\forall n\geq 0$. Portanto,

$$\Lambda(D) = \lambda(D) = \sum_i \eta(p_i). \qquad \Delta$$

<u>DEFINIÇÃO 6.5.3</u> - O número real extendido não negativo

$$\Lambda = \sup\{\Lambda(D):\ D\in\mathcal{D}\}$$

é denominado ENTROPIA DO SISTEMA.

<u>OBSERVAÇÃO 6.5.3</u> - O cálculo da entropia de um sistema pro-

babilístico é geralmente difícil, conforme se pode ver pela definição. No exemplo 6.5.1, embora a decomposição D seja "natural", nada impede, em princípio, que exista uma decomposição E tal que

$$\Lambda(E) > \Lambda(D).$$

TEOREMA 6.5.1 - A entropia de sistemas probabilísticos é um INVARIANTE NUMÉRICO POR ISOMORFISMO, isto é,

$$S_1 \text{ e } S_2 \text{ são sistemas isomorfos} \rightarrow \Lambda(S_1) = \Lambda(S_2).$$

□

EXEMPLO 6.5.2 - Seja $r \geq 1$ e seja $S_r = (X, A, T, P)$ o sistema formado por $X = \{1, \ldots, r\}$, $A = 2^X$,

$$Tx = \begin{cases} x+1, & \text{se } x \neq r \\ 1, & \text{se } x = r \end{cases},$$

$$P\{x\} = 1/r, \ \forall x.$$

$S_r$ é um sistema probabilístico preservativo inversível.

Seja $D = (\{1\}, \ldots, \{r\})$. Como $\lambda(D) = \log r$ e $\lambda(T^{-k}D|D) = 0$, $\forall k \geq 0$, temos

$$\lambda(D \vee T^{-1}D \vee \ldots \vee T^{-n}D) = \log r, \ \forall n \geq 0;$$

portanto $\Lambda(D) = 0$.

É intuitivamente claro que D é a decomposição mais aleatória. De qualquer forma, o raciocínio acima utilizado mostra que $\Lambda(E) = 0$, $\forall E \in \mathcal{D}$. Portanto $\Lambda = 0$. Δ

O exemplo 6.5.2 mostra que a entropia não é um INVARIANTE NUMÉRICO POR ISOMORFISMO COMPLETO (isto é, mesma entropia não implica isomorfismo) na classe dos sistemas probabilísticos preservativos inversíveis, já que os $\mathcal{S}_r$ são claramente não isomorfos entre si.

**PROPOSIÇÕES 6.5** - Seja $S$ um sistema preservativo e sejam $D,E\in\mathcal{D}$.

1. $\Lambda(D) = \lim_{n\to\infty} \lambda(T^{-n}D|DvT^{-1}Dv\ldots vT^{-(n-1)}D)$.
2. $\Lambda(D) = \lim_{n\to\infty} \lambda(D|T^{-1}Dv\ldots vT^{-n}D)$.
3. $\Lambda(D) = \lim_{n\to\infty} \frac{1}{n+1}\cdot\lambda(D^n)$.
4. As sequências envolvidas em 1, 2, 3 são monótonas não crescentes.
5. $E \le D \longrightarrow \Lambda(E) \le \Lambda(D)$.
6. $0 \le j \le k \longrightarrow \Lambda(T^{-j}Dv\ldots vT^{-k}D) = \Lambda(D)$.
7. $k \ge 1 \longrightarrow$ a entropia da decomposição
$$D^{k-1} = DvT^{-1}Dv\ldots vT^{-(k-1)}D$$
no sistema $S^k = (X,A,T^k,D)$ é igual a $k\cdot\Lambda(D)$.
8. $\Lambda(E) \le \Lambda(D) + \lambda(E|D)$.
9. $T = Id \longrightarrow \Lambda = 0$.
10. $D\in\mathcal{D}^r \longrightarrow \Lambda(D) \le \log r$. □

As proposições 6.5.1 e 6.5.2 têm interpretações bem interessantes.

O teorema 6.5.2 é devido a Kolmogorov e Sinai, os teoremas 6.5.3 e 6.5.4 são corolários, o teorema 6.5.5 é uma aplicação.

**TEOREMA 6.5.2** - Se $S$ é preservativo inversível e $D\in\mathcal{D}^r$ é uma decomposição tal que $\{T^{-n}D_i: 1\le i\le r, -\infty<n<+\infty\}$ gera $A$, então $\Lambda = \Lambda(D)$. □

**TEOREMA 6.5.3** - Se $S$ é preservativo, $D \in \mathcal{D}^r$ é uma decomposição tal que $\{T^{-n}D_i : 1 \leq i \leq r,\ 0 \leq n < +\infty\}$ gera $A$, então $\Lambda = \Lambda(D)$. □

**TEOREMA 6.5.4** - Nas condições do teorema 6.5.3, se $S$ é inversível, então $\Lambda = \Lambda(D) = 0$. □

**TEOREMA 6.5.5** - Se $\Omega = \{1,\ldots,r\}$ e $S$ é o sistema (bilateral ou unilateral) de Markov gerado por $(\Omega, 2^{\Omega}, \pi, p)$, então

$$\Lambda = -\sum_{j\in\Omega} [p(j) \cdot \sum_{i\in\Omega} \pi(j,i) \cdot \log \pi(j,i)].$$

Em particular, se $S$ é um sistema de Bernoulli, então

$$\Lambda = -\sum_{i\in\Omega} [p(i) \cdot \log p(i)].$$ □

**EXEMPLO 6.5.3** -

| $S$ | $\Lambda$ |
|---|---|
| SBZ[1/2;1/2] | log 2 |
| SBN[1/2;1/2] | log 2 |
| SBZ[1/3;1/3;1/3] | log 3 |
| SBN[1/3;1/3;1/3] | log 3 |

Temos:

$SBZ\left[\frac{1}{2};\frac{1}{2}\right]$ e $SBZ\left[\frac{1}{3}\ \frac{1}{3}\ \frac{1}{3}\right]$ não são isomorfos;

$SBN\left[\frac{1}{2};\frac{1}{2}\right]$ e $SBN\left[\frac{1}{3};\frac{1}{3};\frac{1}{3}\right]$ não são isomorfos.

Da mesma forma:

$SBZ\left[\frac{1}{2};\frac{1}{2}\right]$ e $SBN\left[\frac{1}{3};\frac{1}{3};\frac{1}{3}\right]$ não são isomorfos;

$SBZ\left[\frac{1}{3};\frac{1}{3};\frac{1}{3}\right]$ e $SBN\left[\frac{1}{2};\frac{1}{2}\right]$ não são isomorfos.

Não se pode responder, por enquanto, às perguntas:

$SBZ\left[\frac{1}{2};\frac{1}{2}\right]$ e $SBN\left[\frac{1}{2};\frac{1}{2}\right]$ são isomorfos?

$SBZ\left[\frac{1}{3};\frac{1}{3};\frac{1}{3}\right]$ e $SBN\left[\frac{1}{3};\frac{1}{3};\frac{1}{3}\right]$ são isomorfos?

Se (ver observação 6.1.1) ser inversível fosse uma PIPI, então poderíamos responder (negativamente, é claro) a essas duas perguntas. $\Delta$

<u>EXEMPLO 6.5.4</u> - Seja $S_c$ uma rotação no círculo unitário com $c^k \neq 1$, para todo $k>0$ (ver exemplo 3.5). Sejam

$$D_1 = \{(\cos t, \text{sen } t): 0 \le t < \pi\} \text{ e } D_2 = D_1^c.$$

A classe $\{T^{-n}D_i: 1\le i\le 2, 0\le n<\infty\}$ gera $A$ e, como $S_c$ é inversível, o teorema 6.5.4 assegura que $\Lambda(I_c) = \Lambda(D) = 0$, onde $D = (D_1, D_2)$. $\Delta$

<u>TEOREMA 6.5.6</u> - Se $S$ é um sistema preservativo, então

$$\Lambda(S^k) = k\cdot\Lambda(S), \ \forall k\ge 0. \qquad \square$$

<u>TEOREMA 6.5.7</u> - Se $S$ é um sistema preservativo inversível, então $\Lambda(S) = \Lambda(S^{-1})$. $\square$

<u>EXEMPLO 6.5.5</u> - Seja $S_c$ uma rotação no círculo unitário com $c^k=1$, para algum $k>0$ (ver exemplo 3.5). Como $T^k x = c^k\cdot x = x$, $\forall x \in X$, isto é, $T^k = Id$, temos, conforme proposição 6.5.10, $\Lambda(S_c^k) = 0$.
Pelo teorema 6.5.6, $k\cdot\Lambda(S_c) = 0$; logo,

$$\Lambda(S_c) = 0. \qquad \Delta$$

Os sistemas do exemplo 6.5.4 são ergódicos e os do exemplo 6.5.5 não são. Como ergodicidade é PIPI, eles não são isomorfos, apesar de terem a mesma entropia.

TEOREMA 6.5.8 - A entropia é um invariante numérico por isomorfismo completo na classe dos sistemas bilaterais de Bernoulli. □

EXEMPLO 6.5.6 - $SBZ\left[\frac{1}{4};\frac{1}{4};\frac{1}{4};\frac{1}{4}\right]$ e $SBZ\left[\frac{1}{2};\frac{1}{8};\frac{1}{8};\frac{1}{8};\frac{1}{8}\right]$ são isomorfos, pois ambos tem a mesma entropia. Δ

O resultado do exemplo 6.5.6 foi obtido por Meshalkin em 1959, anteriormente ao teorema 6.5.8, obtido por Ornstein em 1970, que encerrou o problema do isomorfismo na classe dos sistemas bilaterais de Bernoulli.

Entre os trabalhos de Kolmogorov (1958 e 1959) e os de Ornstein (1970 e 1971), Sinai (1962) foi responsável por uma importante contribuição intermediária ao problema do isomorfismo na classe dos sistemas bilaterais de Bernoulli, envolvendo o conceito de ISOMORFISMO FRACO (ver Apêndice II).

EXEMPLO 6.5.7 - Sejam

$$\Omega_1=\{1,2\} \quad \pi_1=\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \quad p_1=\left(\frac{1}{2},\frac{1}{2}\right)$$

$$\Omega_2=\{1,2,3,4\} \quad \pi_2=\begin{bmatrix} 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{0} & \frac{1}{0} & 0 & 0 \end{bmatrix} \quad p_2=\left(\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4}\right)$$

$$\Omega_3=\{1,2,3,4\} \quad \pi_3=\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \quad p_3=\left(\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4}\right).$$

Para cada $i\in\{1,2,3\}$, seja $S_i$ o sistema bilateral de Markov gerado por $(\Omega_i, 2^{\Omega_i}, \pi_i, p_i)$.

| $i$ | $S_i$ é "mixing" | $S_i$ é ergódico | $\Lambda(S_i)$ |
|---|---|---|---|
| 1 | sim | sim | log 2 |
| 2 | não | sim | log 2 |
| 3 | não | não | log 2 |

Como ser "mixing" e ser ergódico são PIPI, nenhum deles é isomorfo a algum dos outros dois, apesar de terem a mesma entropia. $\Delta$

O exemplo 6.5.7 mostra que a entropia não é um invariante numérico por isomorfismo completo na classe dos sistemas bilaterais de Markov, nem mesmo na classe dos sistemas bilaterais de Markov ergódicos. O comentário feito no parágrafo imediatamente anterior à observação 5.4.3 assegura que a entropia é um invariante numérico por isomorfismo completo na classe dos sistemas bilaterais de Markov "mixing" (essa classe é, na verdade, a classe dos sistemas bilaterais de Bernoulli).

<u>OBSERVAÇÃO 6.5.4</u> - O problema geral do isomorfismo e o estudo específico da entropia são áreas que estão recebendo muita atenção pelos atuais pesquisadores da Teoria Ergódica.

- □ -

## APÊNDICE I

## UMA "MEDIDA" DE ALEATORIZAÇÃO

Seja $(X,\mathcal{A},P)$ um espaço de probabilidade. A probabilidade de um evento é uma "medida" da *certeza* que se possui, antes da realização do experimento, em sua ocorrência. Deseja-se construir uma "medida" $\rho$ que, considerando a probabilidade de um evento como o seu tamanho, expresse quão *pequeno* o evento é (isto é, quão difícil é determinar algum de seus elementos na imensidão do todo). Posto isto, é razoável postular:

(p1) $A \in \mathcal{A} \longrightarrow \rho A \geq 0$

(p2) $A \in \mathcal{A},\ B \in \mathcal{B},\ PA = PB \longrightarrow \rho A = \rho B$

(p3) $A \in \mathcal{A},\ B \in \mathcal{A},\ PA < PB \longrightarrow \rho A > \rho B$

(p4) $A \in \mathcal{A},\ PA = 1 \longrightarrow \rho A = 0$

(p5) $A \in \mathcal{A},\ PA = 0 \longrightarrow \rho A = +\infty$

Qual deve ser a unidade (o BIT) de pequenez (incerteza, indeterminação, aleatoridade)? O primeiro espaço de probabilidade que as pessoas geralmente conhecem é aquele que representa o lançamento de uma moeda equilibrada. Não apenas por esse motivo nostálgico, é razoável postular:

(p6) $A \in \mathcal{A},\ PA = 1/2 \longrightarrow \rho A = 1.$

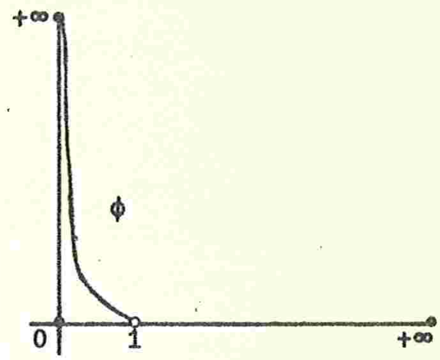
Seja $n \geq 2$ um número inteiro. A incerteza na ocorrência de n caras no lançamento independente de n moedas equilibradas é exatamente n vezes a incerteza na ocorrência de 1 cara no lançamento de 1 moeda equilibrada. Posto isto, é razoável postular:

(p7) $A \in \mathcal{A}$, $PA = 1/2^n \longrightarrow \rho A = n$.

É também razoável postular:

(p8) $\rho$ depende continuamente de P.

<u>CONVENÇÃO 1</u> - $\log_2 0 = -\infty$.




<u>NOTAÇÃO</u> - $\phi$: $[0;1] \longrightarrow [0;+\infty]$

$$t \longrightarrow -\log_2 t$$

A função $\phi P : \mathcal{A} \longrightarrow [0;+\infty]$ satisfaz (p1),...,(p8) e será adotada neste trabalho como "medida" de indeterminação. Para cada $A \in \mathcal{A}$, o *número real extendido* $\phi PA$ é denominado INDETERMINAÇÃO DO EVENTO A; heuristicamente, $\phi PA$ é um indicador numérico da insegurança (intranqüilidade, angústia), antes da realização do experimento, de um jogador que apostou sua fortuna na ocorrência de A.

Para cada $D \in \mathcal{D}$, a *variável aleatória* $\phi(D,\cdot)$ definida por

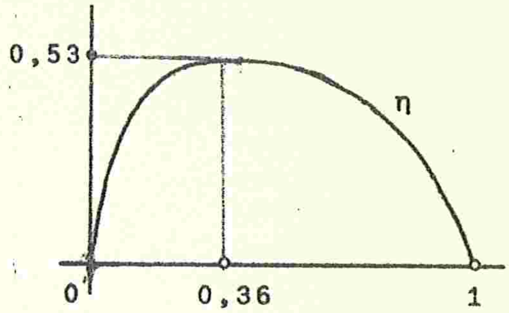
$$\phi(D,x) = \sum_i \phi(PD_i)\, I_{D_i}(x)$$

é denominada INDETERMINAÇÃO DA DECOMPOSIÇÃO D. Para cada $x \in X$, $\phi(D,\cdot)$ associa a pequenez (incerteza, indeterminação, aleatoriedade) do D-átomo a que x pertence. Portanto, essa variável aleatória assume valores em $[0;+\infty]$.

<u>CONVENÇÃO 2</u> - (zero)$\cdot$($+\infty$) = (zero). (Essa convenção permitirá, quando estivermos calculando alguma entropia, desprezar os átomos de probabilidade zero; em particular, os átomos vazios.)

<u>NOTAÇÃO</u> - $\eta : [0;1] \longrightarrow R^1$

$$t \longmapsto \phi(t)\cdot t$$




<u>Propriedades</u>:

1. $\eta \geq 0$
2. $\eta(0) = \eta(1) = 0$
3. $\eta$ é contínua
4. $\eta$ é estritamente côncava.

A função $\eta$ está tabelada na página seguinte. Uma aproximação razoável para um valor $\phi(t)$ pode ser obtido dividindo $\eta(t)$ por $t$. Há tabelas bem mais precisas em [7] Reza.

Para cada $D \in \mathcal{D}$, o *número real não negativo*

$$\lambda(D) = E[\phi(D,\cdot)] = \sum \phi(PD_i)\cdot PD_i = \sum \eta(PD_i)$$

é denominado ENTROPIA DA DECOMPOSIÇÃO D. A entropia de uma decomposição sempre existe e é finita, além de ser não negativa; heuristicamente, $\lambda(D)$ é a indeterminação média (a indeterminação esperada) da decomposição D.

| t | η(t) | t | η(t) | t | η(t) | t | η(t) |
|---|---|---|---|---|---|---|---|
| 0,01 | 0,066 | 0,26 | 0,505 | 0,51 | 0,495 | 0,76 | 0,301 |
| 0,02 | 0,113 | 0,27 | 0,510 | 0,52 | 0,491 | 0,77 | 0,290 |
| 0,03 | 0,152 | 0,28 | 0,514 | 0,53 | 0,485 | 0,78 | 0,280 |
| 0,04 | 0,186 | 0,29 | 0,518 | 0,54 | 0,480 | 0,79 | 0,269 |
| 0,05 | 0,216 | 0,30 | 0,521 | 0,55 | 0,474 | 0,80 | 0,258 |
| 0,06 | 0,244 | 0,31 | 0,524 | 0,56 | 0,468 | 0,81 | 0,246 |
| 0,07 | 0,269 | 0,32 | 0,526 | 0,57 | 0,462 | 0,82 | 0,235 |
| 0,08 | 0,292 | 0,33 | 0,528 | 0,58 | 0,456 | 0,83 | 0,223 |
| 0,09 | 0,313 | 0,34 | 0,529 | 0,59 | 0,449 | 0,84 | 0,211 |
| 0,10 | 0,332 | 0,35 | 0,530 | 0,60 | 0,442 | 0,85 | 0,199 |
| 0,11 | 0,350 | 0,36 | 0,531 | 0,61 | 0,435 | 0,86 | 0,187 |
| 0,12 | 0,367 | 0,37 | 0,531 | 0,62 | 0,428 | 0,87 | 0,175 |
| 0,13 | 0,383 | 0,38 | 0,530 | 0,63 | 0,420 | 0,88 | 0,162 |
| 0,14 | 0,397 | 0,39 | 0,530 | 0,64 | 0,412 | 0,89 | 0,150 |
| 0,15 | 0,411 | 0,40 | 0,529 | 0,65 | 0,404 | 0,90 | 0,137 |
| 0,16 | 0,423 | 0,41 | 0,527 | 0,66 | 0,396 | 0,91 | 0,124 |
| 0,17 | 0,435 | 0,42 | 0,526 | 0,67 | 0,387 | 0,92 | 0,111 |
| 0,18 | 0,445 | 0,43 | 0,524 | 0,68 | 0,378 | 0,93 | 0,097 |
| 0,19 | 0,455 | 0,44 | 0,521 | 0,69 | 0,369 | 0,94 | 0,084 |
| 0,20 | 0,464 | 0,45 | 0,518 | 0,70 | 0,360 | 0,95 | 0,070 |
| 0,21 | 0,473 | 0,46 | 0,515 | 0,71 | 0,351 | 0,96 | 0,057 |
| 0,22 | 0,481 | 0,47 | 0,512 | 0,72 | 0,341 | 0,97 | 0,043 |
| 0,23 | 0,488 | 0,48 | 0,508 | 0,73 | 0,331 | 0,98 | 0,029 |
| 0,24 | 0,494 | 0,49 | 0,504 | 0,74 | 0,321 | 0,99 | 0,014 |
| 0,25 | 0,500 | 0,50 | 0,500 | 0,75 | 0,311 | 1,00 | zero |

## APÊNDICE II

## ISOMORFISMO FRACO

Sejam $S_1 = (X_1, A_1, T_1, \mu_1)$ e $S_2 = (X_2, A_2, T_2, \mu_2)$ dois sistemas. Dizemos que $S_2$ é um FATOR DE $S_1$ se, e somente se, existe um HOMOMORFISMO DE $S_1$ EM $S_2$, isto é, existem dois conjuntos, $Y_1$ e $Y_2$, e uma aplicação $\phi: Y_1 \longrightarrow Y_2$ satisfazendo todas as condições da definição 5.1.2, exceto a condição de $\phi$ ser injetora.

Dois sistemas são denominados SISTEMAS FRACAMENTE ISOMORFOS se, e somente se, cada um deles é um fator do outro. É imediato que dois sistemas isomorfos são sistemas fracamente isomorfos.

É fácil verificar: se $S_1$ e $S_2$ são sistemas probabilísticos e $S_2$ é um fator de $S_1$, então $\Lambda(S_2) \leq \Lambda(S_1)$. Consequentemente: se $S_1$ e $S_2$ são sistemas probabilísticos fracamente isomorfos, eles tem a mesma entropia; em outras palavras: a entropia de sistemas probabilísticos é um INVARIANTE NUMÉRICO POR ISOMORFISMO FRACO.

Sinai (1962) provou que: dois sistemas bilaterais de Bernoulli com a mesma entropia são fracamente isomorfos. O resultado de Ornstein (teorema 6.5.8) assegura, então, que isomorfismo e isomorfismo fraco são noções equivalentes para sistemas bilaterais de Bernoulli.

Ornstein (Advanced in Mathematics, pg. 349-364, 1971) provou que: se $S$ é um fator de um sistema bilateral de Bernoulli, então $S$ é um isomorfo a algum sistema bilateral de Bernoulli.

É trivial verificar que: se $S$ é um sistema (quer bilateral quer unilateral) de Bernoulli, então $S^k$, $k>1$, é um sistema de Bernoulli. Para isso basta substituir $(\Omega, 2^{\Omega}, p)$ por $(\Omega^k, 2^{\Omega^k}, p^k)$. No trabalho mencionado no parágrafo anterior, Ornstein obtém um resultado suficiente para mostrar que: se $S^k$ é um sistema bilateral de Bernoulli para algum $k>1$, então $S$ é isomorfo a um sistema bilateral de Bernoulli.

- 0 -

## APÊNDICE III

## SISTEMAS DE KOLMOGOROV

Sejam $r, \Omega, 2^{\Omega}, Z, N, I, X, A, h_n, T$ como na seção 5.2. Para cada $n \geq 0$, seja $A_n$ a menor $\sigma$-álgebra em relação a qual as aplicações

$$h_n,\ h_{n+1},\ \ldots$$

são mensuráveis, isto é, $A_n$ é a $\sigma$-álgebra gerada pelos retângulos que tem todas as suas coordenadas em $\{n, n+1, \ldots\}$. Seja

$$A_\infty = \bigcap_{n=0}^{\infty} A_n .$$

Dizemos que um sistema $(X,A,T,P)$ é um SISTEMA DE KOLMOGOROV se, e somente se, a medida P satisfaz

$$A \in A_\infty \longrightarrow PA = 0 \quad \text{ou} \quad PA = 1.$$

Quando $I = Z$ $(I=N)$, o sistema $(X,A,T,P)$ é denominado SISTEMA BILATERAL (UNILATERAL) DE KOLMOGOROV.

PROPOSIÇÃO III-1 - Se $S$ é um sistema de Kolmogorov, então $S$ é um sistema "mixing". □

PROPOSIÇÃO III-2 - Se $S$ é um sistema de Bernoulli, então $S$ é um sistema de Kolmogorov. □

<u>PROPOSIÇÃO III-3</u> - Se $S$ é um sistema de Markov com matriz estocástica $\pi$ irredutível e aperiódica e tal que $p(w)>0$, $\forall w \in \Omega$, então $S$ é um sistema de Kolmogorov. $\square$

- $\square$ -

## APÊNDICE IV

## COMPLEMENTOS

Este apêndice é uma miscelânea de alguns fatos que foram omitidos nos capítulos 1,...,6.

**A - EXEMPLO** - Seja $\Omega$ um conjunto finito com r elementos e seja $2^\Omega$ a $\sigma$-álgebra formada por todos os subconjuntos de $\Omega$. Seja X o conjunto de todas as sequências bilaterais de elementos de $\Omega$ e, para cada $n \in Z$, seja $h_n: X \longrightarrow \Omega$ a n-ésima projeção (conforme seção 5.2).

Seja $\mathcal{A}$ a menor $\sigma$-álgebra em relação a qual as aplicações $h_1, h_2, \ldots$ são mensuráveis e seja T o deslocamento definido na seção 5.2. Assim:

i) T é injetora e sobrejetora
ii) T é mensurável
iii) $T^{-1}$ não é mensurável.

Portanto, a estrutura $(X, \mathcal{A}, T)$ não é inversível. $\Delta$

**B - EXEMPLO** - Sejam $X = \{a, b, c, d, e\}$, $\mathcal{A} = 2^X$, T a permutação definida por

Se $\mu$ é uma medida preservativa (definição 4.1), então

$$\mu\{a\} = \mu\{b\} = \mu\{c\} \quad e \quad \mu\{d\} = \mu\{e\}. \quad \Delta$$

<u>C - EXEMPLO</u> - Sejam $X,\mathcal{A}$ como no exemplo B acima e seja $T$ a permutação definida por

$$\begin{tikzcd} a \arrow[r] & b \arrow[r] & c \arrow[dl] \\ e \arrow[u] & d \arrow[l] & \end{tikzcd}$$

Se $P$ é uma probabilidade preservativa, então

$$P\{x\} = 1/5, \ \forall x \in X. \qquad \Delta$$

<u>D - PROPOSIÇÃO</u> - Seja $(X,\mathcal{A},T)$ uma estrutura. Para cada $A \in \mathcal{A}$,

$$A^* \overset{\text{not}}{=} \limsup_{n\to\infty} T^{-n}A \in \mathcal{A}\cap\mathcal{I}.$$

<u>PROVA</u> -

$$\limsup_{n\to\infty} T^{-n}A \overset{\text{def}}{=} \bigcap_{n\geq 0}\ \bigcup_{k\geq n} T^{-k}A \in \mathcal{A},$$

sem dúvida. Para cada $n \geq 0$, seja

$$A_n = \bigcup_{k\geq n} T^{-k}A.$$

Dessa forma:

(1) $A_1 \supset A_2 \supset \ldots$ e

(2) $A^* \overset{\text{def.}}{=} \bigcap_{n\geq 0} A_n \overset{(1)}{=} \bigcap_{n\geq 1} A_n.$

Portanto:

$$(3)\ T^{-1}A^* = T^{-1}\left(\bigcap_{n\geq 0}\ \bigcup_{k\geq n} T^{-k}A\right) = \bigcap_{n\geq 0}\ \bigcup_{k\geq n} T^{-(k+1)}A =$$

$$= \bigcap_{n\geq 1}\ \bigcup_{k\geq n} T^{-k}A = \bigcap_{n\geq 1} A_n \overset{(2)}{=} A^*\ \therefore\ A^* \in I. \square$$

E - PROPOSIÇÃO - Seja $S = (X,A,T,\mu)$ inversível. São equivalentes:

i) $S$ é ergódico.

ii) $S^{-1}$ é ergódico.

PROVA - Vamos anotar a classe dos conjuntos invariantes de $S$ por $I$ e a classe dos conjuntos invariantes de $S^{-1}$ por $I^{-1}$. Acontece que eles são iguais, pois:
$E \in I \leftrightarrow T^{-1}E = E \leftrightarrow T(T^{-1}E) = TE \leftrightarrow E = TE \leftrightarrow E \in I^{-1}$.
$\square$

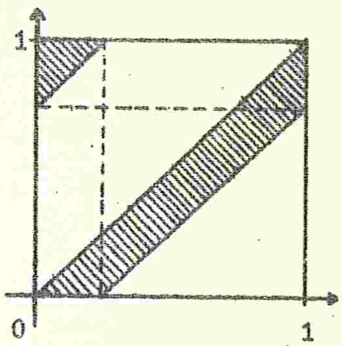
F - OBSERVAÇÃO - Seja $(X,A,T)$ uma estrutura da seção 5.2, e $P$ uma probabilidade qualquer no espaço mensurável $(X,A)$. Embora não se possa pensar que há uma fonte geradora emitindo uma letra em cada intervalo unitário do tempo, a interpretação da seção 5.2 (a mensagem é escolhida antes do início do tempo) se aplica a esse caso geral.

G - DECOMPOSIÇÃO E PROCESSOS ESTOCÁSTICOS - Seja $(\Omega,A,T,P)$ um sistema probabilístico e $D \in \mathcal{D}^r(\Omega,A)$. Para cada $n\geq 0$, seja

$$X_n(w) = \sum_i i \cdot I_{D_i}(T^n w),\ \forall w \in \Omega.$$

Dessa forma, $X = (X_0, X_1, \ldots)$ é um processo estocástico com espaço de estados $\{1,\ldots,r\}$, que pode ser denominado PROCESSO ESTOCÁSTICO INDUZIDO PELA DECOMPOSIÇÃO D. Se o sistema é preservativo, esse processo estocástico é estacionário.

H - EXEMPLO - Sejam $X=[0;1)^2$, $A$ a $\sigma$-álgebra de Borel, $\mu$ a medida de Lebesgue. Para cada $a=(a_1,a_2)\in X$, seja $T_a$ a transformação definida por




$$T_a(x_1,x_2) = (x_1+a_1(\text{mód. } 1),\ x_2+a_2(\text{mód. } 2)).$$

É imediato que: $S_a=(X,A,T_a,\mu)$ é preservativo e inversível. A parte hachurada da figura representa um conjunto $A$-mensurável não trivial, que é invariante quando $a=(1/2,1/2)$. $\Delta$

Está provado em [6] Petersen, pg. 65-66, que são equivalentes:

i) $S_a$ é ergódico

ii) $k_1\in Z$, $k_2\in Z$, $k_1\cdot a_1+k_2\cdot a_2=0 \longrightarrow k_1=k_2=0$.

Esse resultado vale também para "quadrados unitários" de dimensão maior que dois.

I - EXEMPLO - Sejam $X=\{\ldots,-2,-1,0,1,2\ldots\}$ uma população onde os negativos são indivíduos solteiros e $\{0,1\}$, $\{2,3\}$, $\{4,5\}$,... são os casais. Seja $A$ a $\sigma$-álgebra formada por todos os subconjuntos da população que não separam casais. Assim, $\{-4,-3,0,1,8,9\}\in A$, $\{1,2\}\notin A$ e $\{6,7,15\}\notin A$.

Seja T a transformação definida por $Tx = x+2$ e $\mu$ a medida de contagem. A transformação T é bijetora e mensurável. O sistema $(X, A, T, \mu)$ é preservativo. Entretanto

$$\{-1,0,1\} \in A \text{ e } T\{-1,0,1\} = \{1,2,3\} \notin A,$$

o que mostra que $T^{-1}$ é não mensurável. $\Delta$

Esse exemplo mostra que um sistema preservativo com transformação bijetora (em particular, uma estrutura com transformação bijetora) pode não ser inversível.

- ▯ -

# REFERÊNCIAS BIBLIOGRÁFICAS


[1] - ASH, R.B. - Information Theory. Ed. Wiley - Interscience (1965).

[2] - BILLINGSLEY, P.P. - Ergodic Theory and Information. Ed. Wiley (1965).

[3] - BREIMAN, L. - Probability. Ed. Addison-Wesley (1968).

[4] - HALMOS, P.R. - Measure Theory. Ed. Van Nostrand (1950).

[5] - HALMOS, P.R. - Lectures on Ergodic Theory. The Mathematical Society of Japan (1956).

[6] - PETERSEN, K.E. - Introductory Ergodic Theory. Depto. Math. Univ. North Carolina (1971).

[7] - REZA, F.M. - An Introduction to Information Theory. Ed. McGraw-Hill (1961).

[8] - WRIGHT, F.B. - Ergodic Theory. Ed. Academic Press, (1963).


- § -