COLÓQUIOS DE MATEMÁTICA DAS REGIÕES
REGIÃO SUL

IV Colóquio de Matemática
da Região Sul

TÓPICOS DE ÁLGEBRA LINEAR E PROBABILIDADE

JAIRO MENGUE

FURG

SBM
SOCIEDADE BRASILEIRA DE MATEMÁTICA

Tópicos de Álgebra Linear e Probabilidade

Tópicos de Álgebra Linear e Probabilidade

Capa: Pablo Diego Regino
Projeto gráfico: Cinthya Maria Schneider Meneghetti

Distribuição e vendas
Sociedade Brasileira de Matemática
Estrada Dona Castorina, 110 Sala 109 - Jardim Botânico

ISBN (eBook) 978-85-8337-096-3

COLÓQUIOS DE MATEMÁTICA DAS REGIÕES
REGIÃO SUL

IV Colóquio de Matemática
da Região Sul

TÓPICOS DE ÁLGEBRA LINEAR E PROBABILIDADE

JAIRO MENGUE

1ª EDIÇÃO
2016
RIO GRANDE

FURG

SBM
SOCIEDADE BRASILEIRA DE MATEMÁTICA

Sumário

Prefácio

Este texto foi preparado para utilização no IV Colóquio de Matemática da Região Sul a partir do texto já utilizado no III Colóquio de Matemática da Região Nordeste. O mesmo foi elaborado com a expectativa de poder ser lido e compreendido por alunos em início de graduação. Com este objetivo o trabalho com probabilidades ficou limitado a conjuntos finitos e alguns conceitos usuais em Teoria Ergódica, como probabilidade invariante e entropia, foram adaptados. Ainda assim, as ideias iniciais que aparecem no Formalismo Termodinâmico e na Otimização Ergódica foram preservadas, podendo ser este um primeiro contato do estudante com as mesmas, muito antes de um primeiro curso em teoria da medida.

Uma das motivações para a escolha deste tópico e forma de abordagem está no meu interesse em poder apresentar a alunos e orientandos de iniciação científica parte das ideias e conceitos destas áreas.

Além da utilização no colóquio, este texto pode servir de forma complementar a professores e alunos de álgebra linear, apresentando algumas das relações desta com o estudo de probabilidades. Aparecerão com frequência no texto os conceitos de autovalor e autovetor, por exemplo.

Alguns tópicos são apresentados de forma levemente informal, dependendo do ponto de vista do leitor. Por exemplo, no capítulo 1 o leitor mais experiente poderá sentir falta de uma discussão mais profunda sobre processos estocásticos e até mesmo sobre probabilidade condicional. Ainda assim estará convidado a avaliar se este texto pode servir como introdução ao estudo das Cadeias de Markov.

Cabe ressaltar ao leitor que uma discussão mais profunda dos conceitos apresentados requer alguns conhecimentos sobre espaços métricos e topologia, teoria da medida e análise funcional.

Jairo K. Mengue

Agradecimentos

Gostaria de agradecer aos colegas do grupo de pesquisa em sistemas dinâmicos do IME-UFRGS pelas diversas discussões, informais ou em seminários, de temas que serviram de inspiração para elaboração deste texto.

Agradeço também aos alunos, professores e demais participantes do III colóquio de matemática da região Nordeste que demonstraram interesse no assunto. Isso me motivou a propor este minicurso novamente.

Por fim, gostaria de agradecer a minha família pela paciência e compreensão nos momentos em que estive próximo, mas não dediquei a ela uma merecida atenção.

Capítulo 1

Matriz estocástica e vetor de probabilidade

Considere como modelo teórico uma caixa de base retangular dividida em três regiões denotados por R_1, R_2 e R_3. Vamos supor que dentro desta caixa existem 1000 pequenas bolas em movimento forçado e que a cada minuto uma câmera fotografa a caixa fornecendo informações sobre as posições das bolas. Não estando interessados em entender o comportamento de cada bola, mas a distribuição delas nas regiões R_1, R_2 e R_3, podemos considerar um vetor com 3 coordenadas indicando a quantidade de bolas em cada região em uma dada fotografia [1] . Por exemplo, se em uma dada fotografia temos R_1 com 125 bolas, R_2 com 250 bolas e R_3 com 625 bolas podemos escrever esta informação como

$$\begin{pmatrix} 125 \\ 250 \\ 625 \end{pmatrix}.$$

Se também não temos interesse na quantidade total de bolas podemos dizer, a partir dos números acima, que 12,5% das bolas estão em R_1, 25% das bolas estão em R_2 e 62,5% das bolas estão em R_3 ou, dividindo a quantidade de bolas de cada região pelo total de bolas da caixa, escrever esta informação como

$$P^{[0]} = \begin{pmatrix} 0,125 \\ 0,250 \\ 0,625 \end{pmatrix}. \tag{1.1}$$

Se para cada fotografia analisada associarmos um vetor, como este último, descrevendo a proporção de bolas de cada região em relação ao total de bolas, podemos observar que todos estes vetores possuem três coordenadas não negativas cuja soma resulta em 1.

[1] vamos desconsiderar neste modelo teórico os problemas que podem ser causados por fronteiras

Definição 1.1. *Dizemos que $P \in \mathbb{R}^n$ é um vetor de probabilidade se suas coordenadas $p_1, ..., p_n$ satisfazem:*
a) $p_i \geq 0, \ i \in \{1, ..., n\}$
b) $p_1 + p_2 + ... + p_n = 1.$

Suponhamos agora que para este modelo, independente da informação sobre as fotografias retiradas a mais de 1 minuto e da quantidade de fotografias já retiradas, a proporção de bolas que migram para a região R_i saindo da região R_j, ao compararmos fotografias consecutivas, seja estimada em a_{ij}, onde

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} 0,5 & 0,3 & 0,1 \\ 0,4 & 0,4 & 0,6 \\ 0,1 & 0,3 & 0,3 \end{pmatrix}. \tag{1.2}$$

Assim, a partir da segunda coluna desta matriz, temos como previsão:
- 30% das bolas em R_2 na fotografia atual estarão em R_1 na próxima fotografia.
- 40% das bolas em R_2 na fotografia atual estarão em R_2 na próxima fotografia.
- 30% das bolas em R_2 na fotografia atual estarão em R_3 na próxima fotografia.

As entradas da matriz dada em (1.2) são não negativas e a soma dos elementos de cada coluna resulta em 1. Assim, as colunas da matriz são vetores de probabilidade.

Definição 1.2. *Dizemos que uma matriz quadrada $A_{n\times n}$ com entradas $\{a_{ij}\}_{1\leq i,j\leq n}$ é coluna estocástica se satisfaz:*
a) $a_{ij} \geq 0$, *para quaisquer* $i, j \in \{1, ..., n\}$
b) $a_{1j} + ... + a_{nj} = 1$ *para qualquer* $j \in \{1, ..., n\}$.

Se para uma dada bola, denotarmos por X_n o número em $\{1, 2, 3\}$ que indica a região onde se localiza esta bola na n-ésima fotografia, a informação acima pode ser descrita como uma probabilidade condicional[2]:

$$Pr(X_{n+1} = i \,|\, X_n = j) = a_{ij}.$$

Assumimos que a probabilidade condicional satisfaz a relação (Regra de Bayes)

$$Pr(X_{n+1} = i) = \sum_{j=1}^{3} Pr(X_{n+1} = i \,|\, X_n = j) Pr(X_n = j).$$

Problema: Estando com o vetor dado em (1.1) para a fotografia atual, qual previsão podemos fazer para o vetor associado a fotografia seguinte?

[2]Estamos assumindo que para quaisquer n e $s_0, ..., s_{n-1}$ fixados,

$$Pr(X_{n+1} = i \,|\, X_n = j, X_{n-1} = s_{n-1}, ..., X_1 = s_1, X_0 = s_0)$$
$$= Pr(X_{n+1} = i \,|\, X_n = j) = Pr(X_1 = i \,|\, X_0 = j).$$

Solução: Supomos dado $P^{[0]}$, que indica a distribuição das bolas em tempo 0 (fotografia atual) e buscamos uma previsão de $P^{[1]}$, que indica a distribuição das bolas em tempo 1 (fotografia seguinte). Qual a probabilidade de uma dada bola estar em R_1 em tempo 1?

$$Pr(X_1 = 1) = \sum_{i=1}^{3} Pr(X_1 = 1|X_0 = i)Pr(X_0 = i)$$

$$= (0,5)(0,125) + (0,3)(0,25) + (0,1)(0,625) = 0,2.$$

De forma análoga temos

$$Pr(X_1 = 2) = \sum_{i=1}^{3} Pr(X_1 = 2|X_0 = i)Pr(X_0 = i)$$

$$= (0,4)(0,125) + (0,4)(0,25) + (0,6)(0,625) = 0,525.$$

e

$$Pr(X_1 = 3) = \sum_{i=1}^{3} Pr(X_1 = 3|X_0 = i)Pr(X_0 = i)$$

$$= (0,1)(0,125) + (0,3)(0,25) + (0,3)(0,625) = 0,275.$$

Concluímos que nossa previsão é dada por

$$P^{[1]} = \begin{pmatrix} 0,2 \\ 0,525 \\ 0,275 \end{pmatrix}.$$

Observe que as contas do problema acima podem ser resumidas na forma:

$$P^{[1]} = \begin{pmatrix} 0,2 \\ 0,525 \\ 0,275 \end{pmatrix} = \begin{pmatrix} 0,5 & 0,3 & 0,1 \\ 0,4 & 0,4 & 0,6 \\ 0,1 & 0,3 & 0,3 \end{pmatrix} \begin{pmatrix} 0,125 \\ 0,250 \\ 0,625 \end{pmatrix} = AP^{[0]}. \tag{1.3}$$

Para a previsão de P^1 podemos, como visto acima, calcular o produto dado em (1.3). Note que neste exemplo o produto de uma matriz coluna estocástica por um vetor de probabilidade resultou em um novo vetor de probabilidade.

Proposição 1.3. *Se $A_{n\times n}$ é uma matriz coluna estocástica e $P \in \mathbb{R}^n$ é um vetor de probabilidade, então (AP) é um vetor de probabilidade.*

Demonstração: Escrevendo $Q = AP$ onde $A = (a_{ij})_{1\leq i,j\leq n}$, $P = (p_i)_{1\leq i\leq n}$ e $Q = (q_i)_{1\leq i\leq n}$ temos

$$Q = \begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_n \end{pmatrix} = \begin{pmatrix} a_{11}p_1 + a_{12}p_2 + \ldots + a_{1n}p_n \\ a_{21}p_1 + a_{22}p_2 + \ldots + a_{2n}p_n \\ \vdots \\ a_{n1}p_1 + a_{n2}p_2 + \ldots + a_{nn}p_n \end{pmatrix} = AP,$$

o que resumidamente pode ser escrito como

$$q_i = \sum_{k=1}^{n} a_{ik} p_k, \ i \in \{1, ..., n\}.$$

Devemos provar que Q é um vetor de probabilidade. Como A é coluna estocástica e P é um vetor de probabilidade temos que $a_{ik} \geq 0$ e $p_k \geq 0$ para quaisquer $i, k \in \{1, ..., n\}$. Portanto $q_i = \sum\limits_{k=1}^{n} a_{ik} p_k \geq 0$ para todo $i \in \{1, ..., n\}$. Além disso, como A é coluna estocástica, sabemos que $\sum_{i=1}^{n} a_{ik} = 1, \ \ k \in \{1, ..., n\}$ e como P é um vetor de probabilidade, $\sum_{k=1}^{n} p_k = 1$. Assim

$$\begin{aligned} q_1 + ... + q_n = \sum_{i=1}^{n} q_i &= \sum_{i=1}^{n} \left(\sum_{k=1}^{n} a_{ik} p_k \right) = \sum_{k=1}^{n} \sum_{i=1}^{n} a_{ik} p_k \\ &= \sum_{k=1}^{n} p_k \left(\sum_{i=1}^{n} a_{ik} \right) = \sum_{k=1}^{n} p_k = 1. \end{aligned}$$

□

Voltando ao exemplo das bolas, dada a previsão de um vetor

$$P^{[n]} = \begin{pmatrix} Pr(X_n = 1) \\ Pr(X_n = 2) \\ Pr(X_n = 3) \end{pmatrix} = \begin{pmatrix} p_1^{[n]} \\ p_2^{[n]} \\ p_3^{[n]} \end{pmatrix},$$

qual será o vetor $P^{[n+1]}$ previsto?

Solução: Escrevendo $Pr(X_{n+1} = i | X_n = j) = a_{ij}$ onde os números a_{ij} são dados em (1.2), temos

$$P^{[n+1]} = \begin{pmatrix} Pr(X_{n+1} = 1) \\ \\ Pr(X_{n+1} = 2) \\ \\ Pr(X_{n+1} = 3) \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{3} Pr(X_{n+1} = 1 | X_n = j) Pr(X_n = j) \\ \\ \sum_{j=1}^{3} Pr(X_{n+1} = 2 | X_n = j) Pr(X_n = j) \\ \\ \sum_{j=1}^{3} Pr(X_{n+1} = 3 | X_n = j) Pr(X_n = j) \end{pmatrix}$$

$$= \begin{pmatrix} a_{11} p_1^{[n]} + a_{12} p_2^{[n]} + a_{13} p_3^{[n]} \\ a_{21} p_1^{[n]} + a_{22} p_2^{[n]} + a_{23} p_3^{[n]} \\ a_{31} p_1^{[n]} + a_{32} p_2^{[n]} + a_{33} p_3^{[n]} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} p_1^{[n]} \\ p_2^{[n]} \\ p_3^{[n]} \end{pmatrix} = AP^{[n]}.$$

Aplicando-se um argumento de indução matemática, podemos concluir que dados o vetor $P^{[0]}$ e a matriz A em (1.2) e (1.1), respectivamente, a estimativa para o vetor de probabilidade associado a fotografia em tempo m será $P^{[m]} = A^m P^{[0]}$, onde

$$A^m = \underbrace{A \cdot A \cdots A}_{m\, vezes}.$$

Com base nesta igualdade, dada uma matriz coluna estocástica $A_{n\times n}$ e um vetor de probabilidade $P^{[0]}$ em $\mathbb{R}^n$, podemos tentar entender o comportamento do vetor $P^{[m]} = A^m P^{[0]}$ "a longo prazo". Neste sentido, buscamos entender se por exemplo as coordenadas de $P^{[m]}$ se aproximam das coordenadas de algum vetor $\bar{P}$ quando $m \to +\infty$.

Antes de prosseguirmos com resultados gerais, vamos analisar dois exemplos.

Exemplo 1.4. *Considere a matriz coluna estocástica*

$$A = \begin{pmatrix} 1/3 & 1/4 \\ 2/3 & 3/4 \end{pmatrix}.$$

Os autovalores de A são $\lambda_1 = 1$ e $\lambda_2 = 1/12$, podendo ser verificado diretamente que

$$\begin{pmatrix} 1/3 & 1/4 \\ 2/3 & 3/4 \end{pmatrix}\begin{pmatrix} 3 \\ 8 \end{pmatrix} = \begin{pmatrix} 3 \\ 8 \end{pmatrix}$$

e

$$\begin{pmatrix} 1/3 & 1/4 \\ 2/3 & 3/4 \end{pmatrix}\begin{pmatrix} -1 \\ 1 \end{pmatrix} = \frac{1}{12}\begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

Note que

$$\bar{P} = \begin{pmatrix} 3/11 \\ 8/11 \end{pmatrix}$$

é um vetor de probabilidade que também é autovetor associado ao autovalor $\lambda_1 = 1$. Vamos denotar por $v = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$, o autovetor associado ao autovalor λ_2, que estamos considerando. Nenhum múltiplo de v é vetor de probabilidade.

Dado um vetor de probabilidade arbitrário $P = \begin{pmatrix} p \\ 1-p \end{pmatrix}$ onde $0 \leq p \leq 1$, podemos escrever

$$P = \begin{pmatrix} p \\ 1-p \end{pmatrix} = \begin{pmatrix} 3/11 \\ 8/11 \end{pmatrix} + (3/11 - p)\begin{pmatrix} -1 \\ 1 \end{pmatrix} = \bar{P} + (3/11 - p)v.$$

Assim

$$AP = A(\bar{P} + (3/11 - p)v) = A\bar{P} + (3/11 - p)Av = \bar{P} + \frac{(3/11 - p)}{12}v,$$

$$A^2P = A(AP) = A(\bar{P} + \frac{(3/11 - p)}{12}v) = A\bar{P} + \frac{(3/11 - p)}{12}Av = \bar{P} + \frac{(3/11 - p)}{(12)^2}v$$

e em geral, pode ser demonstrado com um argumento de indução matemática que

$$A^m P = \bar{P} + \frac{(3/11 - p)}{(12)^m}v.$$

Quando $m \to +\infty$, temos $\frac{(3/11-p)}{(12)^m} \to 0$ e por consequência

$$\lim_{m\to+\infty} A^m P = \bar{P}.$$

Concluímos que para qualquer vetor de probabilidade P, os vetores $A^m P$ se aproximam do vetor de probabilidade $\bar{P}$, onde $\bar{P}$ satisfaz $A\bar{P} = \bar{P}$. Como ilustração deste fato, note que se $P = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, então $A^m P$ corresponde a primeira coluna de A^m, enquanto para $P = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, $A^m P$ corresponde a segunda coluna de A^m. Portanto ambas as colunas de A^m devem convergir para $\bar{P}$ quando $m \to +\infty$. Isso está de acordo com o os dados abaixo:

$$P \approx \begin{pmatrix} 0,27273 \\ 0,72727 \end{pmatrix}, \quad A \approx \begin{pmatrix} 0,33333 & 0,25 \\ 0,66666 & 0,75 \end{pmatrix}, \quad A^2 \approx \begin{pmatrix} 0,27778 & 0,27083 \\ 0,72222 & 0,72917 \end{pmatrix}$$

$$A^3 \approx \begin{pmatrix} 0,27315 & 0,27257 \\ 0,72685 & 0,72743 \end{pmatrix}, \quad A^4 \approx \begin{pmatrix} 0,27276 & 0,27271 \\ 0,72724 & 0,72729 \end{pmatrix}.$$

Exemplo 1.5. *Dados*

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad P^{[0]} = \begin{pmatrix} a \\ 1-a \end{pmatrix} \quad e \quad P^{[1]} = \begin{pmatrix} 1-a \\ a \end{pmatrix},$$

temos que

$$P^{[1]} = AP = A^3P = A^5P = A^7P... \;\; e \;\; P^{[0]} = P = A^2P = A^4P = A^6P...$$

Portanto $A^m P$ não converge.

É possível[3] obtermos resultados como os do primeiro exemplo supondo que todas as entradas de A são positivas (estritamente maiores que zero).

Convenção: Dizemos que uma matriz é positiva se todas as suas entradas são positivas.

Lema 1.6. *Se $A_{n\times n}$ é uma matriz coluna estocástica, então $\lambda = 1$ é um autovalor de A.*

Demonstração: Seja $\mathbf{1}$ o vetor com todas as coordenadas iguais a 1. A matriz A^T é "linha estocástica", sendo fácil verificar que $A^T\mathbf{1} = \mathbf{1}$. Concluímos que $\lambda = 1$ é autovalor de A^T, portanto é um autovalor de A. □

[3]Seria possível obtermos resultados positivos com hipóteses mais fracas, como supor que alguma potência de A possui todas as entradas positivas, mas neste texto de caráter introdutório vamos simplificar a discussão.

Vamos generalizar alguns fatos que apareceram no primeiro exemplo acima, para uma matriz A coluna estocástica positiva. Antes de mostrarmos que A possui um autovetor de probabilidade associado ao autovalor 1 vamos analisar um pouco mais a matriz A^T.

Lema 1.7. *Seja $A_{n\times n}$ uma matriz coluna estocástica positiva e $v \in \mathbb{R}^n$ um vetor com entradas não negativas. Então $(A^T)^m v$ é convergente e o vetor limite terá todas as coordenadas iguais.*

Demonstração: Se $n = 1$ o resultado é imediato, portanto podemos supor $n \geq 2$. Escrevemos $A = (a_{ij})$ onde $a_{ij} > 0$ para quaisquer $i, j \in \{1, ..., n\}$ e $\sum_k a_{kj} = 1$ para qualquer $j \in \{1, ..., n\}$. Seja α a menor entrada da matriz A. Note que $0 < \alpha \leq \frac{1}{n} \leq \frac{1}{2}$. Denotamos por $|x|_{max}$ e $|x|_{min}$ a maior e menor entrada de um vetor x, respectivamente. Seja v um vetor com coordenadas não negativas $v_1, ..., v_n$. Suponha que $|v|_{min} = v_j$. Então a $i-$esima coordenada de $A^T v$ satisfaz

$$(A^T v)_i = \sum_{k=1}^{n} a_{ki} v_k = a_{ji} v_j + \sum_{k\neq j} a_{ki} v_k \leq a_{ji} v_j + \sum_{k\neq j} a_{ki} |v|_{max}$$

$$= a_{ji} v_j + (1 - a_{ji})|v|_{max} = |v|_{max} + a_{ji}(|v|_{min} - |v|_{max})$$
$$\leq |v|_{max} + \alpha(|v|_{min} - |v|_{max}).$$

Portanto

$$|A^T v|_{max} \leq |v|_{max} + \alpha(|v|_{min} - |v|_{max}). \tag{1.4}$$

Da mesma forma, suponha que $|v|_{max} = v_l$. Então a $i-$esima coordenada de $A^T v$ satisfaz

$$(A^T v)_i = \sum_{k=1}^{n} a_{ki} v_k \geq a_{li} v_l + (1 - a_{li})|v|_{min} = |v|_{min} + a_{li}(|v|_{max} - |v|_{min})$$

$$\geq |v|_{min} + \alpha(|v|_{max} - |v|_{min}).$$

Portanto

$$|A^T v|_{min} \geq |v|_{min} + \alpha(|v|_{max} - |v|_{min}). \tag{1.5}$$

Concluímos que para todo vetor v com coordenadas não negativas,

$$|A^T v|_{max} - |A^T v|_{min} \leq (1 - 2\alpha)(|v|_{max} - |v|_{min}). \tag{1.6}$$

Além disso, se

$$|(A^T)^{m-1} v|_{max} - |(A^T)^{m-1} v|_{min} \leq (1 - 2\alpha)^{m-1}(|v|_{max} - |v|_{min})$$

então, aplicando a desigualdade (1.6) para o vetor $(A^T)^{m-1} v$,

$$|(A^T)^m v|_{max} - |(A^T)^m v|_{min} = |(A^T)(A^T)^{m-1} v|_{max} - |(A^T)(A^T)^{m-1} v|_{min}$$

$\leq (1-2\alpha)\left[|(A^T)^{m-1}v|_{max} - |(A^T)^{m-1}v|_{min}\right] \leq (1-2\alpha)^m(|v|_{max} - |v|_{min})$.

Assim, com um argumento de indução matemática, concluímos que para $m = 1, 2, 3, ...$

$$|(A^T)^m v|_{max} - |(A^T)^m v|_{min} \leq (1-2\alpha)^m(|v|_{max} - |v|_{min}).$$

Como $0 < \alpha \leq \frac{1}{2}$,

$$\lim_{m\to+\infty}\left(|(A^T)^m v|_{max} - |(A^T)^m v|_{min}\right) = 0.$$

Além disso, por (1.4) e (1.5), $|(A^T)^m v|_{max}$ é decrescente em m enquanto $|(A^T)^m v|_{min}$ é crescente em m. Portanto quando $m \to +\infty$, $(A^T)^m v$ converge para um vetor constante. $\square$

Teorema 1.8. *Seja $A_{n\times n}$ uma matriz coluna estocástica e positiva. Então existe um único vetor de probabilidade $\bar{P}$ satisfazendo $A\bar{P} = \bar{P}$. Além disso, para qualquer vetor de probabilidade P, $\lim_{m\to+\infty} A^m P = \bar{P}$.*

Demonstração: Escrevemos $A = (a_{ij})$ onde $a_{ij} > 0$ para quaisquer $i, j \in \{1, ..., n\}$ e $\sum_k a_{kj} = 1$ para qualquer $j \in \{1, ..., n\}$. Denotamos por $e_1, e_2, ..., e_n$ os vetores que formam a base canônica do espaço $\mathbb{R}^n$. Aplicando-se o lema anterior concluímos que existem os limites

$$\psi_1 = \lim_{m\to+\infty}(A^T)^m e_1, ..., \psi_n = \lim_{m\to+\infty}(A^T)^m e_n,$$

onde os vetores ψ_i são constantes. Denotamos por p_i o valor que aparece em todas as entradas de ψ_i, $i = 1, 2, ..., n$. Ou seja

$$\psi_1 = \begin{pmatrix} p_1 \\ \vdots \\ p_1 \end{pmatrix}, \; \psi_2 = \begin{pmatrix} p_2 \\ \vdots \\ p_2 \end{pmatrix}, \ldots, \; \psi_n = \begin{pmatrix} p_n \\ \vdots \\ p_n \end{pmatrix}.$$

Seja

$$\bar{P} = \begin{pmatrix} p_1 \\ \vdots \\ p_n \end{pmatrix}.$$

Afirmamos que $\bar{P}$ é um vetor de probabilidade. De fato, como as entradas de A e de e_i são não negativas, obtemos que as entradas de ψ_i são não negativas, ou seja $p_i \geq 0$. Além disso, para mostrarmos que $p_1 + ... + p_n = 1$, basta mostrarmos que $\psi_1 + ... + \psi_n$ coincide com o vetor constante

$$\mathbf{1} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}.$$

Como $A^T\mathbf{1} = \mathbf{1}$, obtemos que

$$\psi_1 + ... + \psi_n = \sum_{i=1}^{n} \psi_i = \sum_{i=1}^{n} \lim_{m\to+\infty} (A^T)^m e_i$$

$$= \lim_{m\to+\infty} (A^T)^m \sum_{i=1}^{n} e_i = \lim_{m\to+\infty} (A^T)^m \mathbf{1} = \lim_{m\to+\infty} \mathbf{1} = \mathbf{1},$$

concluindo a prova da afirmação.

Afirmamos agora que para qualquer vetor de probabilidade $P = \begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_n \end{pmatrix}$,

$$\lim_{m\to+\infty} A^m P = \bar{P} = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{pmatrix}.$$

De fato, para cada $j \in \{1, ..., n\}$ temos que a $j-$ésima coordenada de $\lim_{m\to+\infty} A^m P$ satisfaz:

$$\langle e_j, \lim_{m\to+\infty} A^m P\rangle = \lim_{m\to+\infty} \langle e_j, A^m P\rangle = \lim_{m\to+\infty} \langle (A^T)^m e_j, P\rangle = \langle \lim_{m\to+\infty} (A^T)^m e_j, P\rangle$$

$$= \langle \psi_j, P\rangle = \left\langle \begin{pmatrix} p_j \\ \vdots \\ p_j \end{pmatrix}, \begin{pmatrix} q_1 \\ \vdots \\ q_n \end{pmatrix} \right\rangle = p_j(q_1 + ... + q_n) = p_j,$$

provando a afirmação.

Para mostrarmos que $A\bar{P} = \bar{P}$ basta observarmos que

$$\bar{P} = \lim_{m\to+\infty} A^{m+1}\bar{P} = \lim_{m\to+\infty} AA^m\bar{P} = A \lim_{m\to+\infty} A^m\bar{P} = A\bar{P}.$$

Por fim, se um vetor de probabilidade Q satisfaz $AQ = Q$, então $A^m Q = Q,\ m = 1, 2, 3, ...$ Portanto

$$Q = \lim_{m\to+\infty} A^m Q = \bar{P}.$$

Isso garante que $\bar{P}$ é o único vetor de probabilidade satisfazendo $A\bar{P} = \bar{P}$ e conclui a demonstração. $\square$

Definição 1.9. *Dada uma matriz A coluna estocástica positiva, o vetor de probabilidade $\bar{P}$ satisfazendo $A\bar{P} = \bar{P}$ é chamado vetor estacionário associado a matriz A.*

Note que se A é positiva então para estimativas de "longo prazo" do vetor $P^m = A^m P$, o vetor inicial P não é de fato relevante, ou seja quando $m \to \infty$, $A^m P$ converge ao vetor estacionário $\bar{P}$ independente do vetor de probabilidade inicial P.

Corolário 1.10. *O vetor estacionário associado a uma matriz coluna estocástica positiva é também positivo.*

Demonstração: Escrevemos $A = (a_{ij})$ e denotamos por $p_1, ..., p_n$ as coordenadas de $\bar{P}$. Lembramos que as entradas de A são positivas e as coordenadas de $\bar{P}$ são maiores ou iguais a zero. Suponhamos por absurdo que alguma coordenada p_k de $\bar{P}$ seja nula. Como $\sum_i a_{ki} p_i = p_k = 0$, concluímos que todas as coordenadas de $\bar{P}$ são nulas. Mas $p_1 + ... + p_k = 1$, garantindo uma contradição. $\square$

Exemplo 1.11. *Considere a matriz*

$$A = \begin{pmatrix} 0,5 & 0,3 & 0,1 \\ 0,4 & 0,4 & 0,6 \\ 0,1 & 0,3 & 0,3 \end{pmatrix}$$

dada em (1.2). Para determinarmos um autovetor x associado ao autovalor $\lambda = 1$ resolvemos a equação $Ax = x$ ou, equivalentemente, a equação linear homogênea $(A - I)x = 0$. Uma solução para esta equação é dada por

$$x = \begin{pmatrix} 12 \\ 17 \\ 9 \end{pmatrix}.$$

Portanto o vetor estacionário associado a matriz A é

$$\bar{P} = \begin{pmatrix} 12/38 \\ 17/38 \\ 9/38 \end{pmatrix} \approx \begin{pmatrix} 0,3158 \\ 0,4474 \\ 0,2368 \end{pmatrix}.$$

Dado

$$P^{[0]} = \begin{pmatrix} 0,125 \\ 0,250 \\ 0,625 \end{pmatrix},$$

pelo que vimos acima, os vetores $P^{[n]} = A^n P^{[0]}, \;\; n = 1, 2, 3, ...$ aproximam-se de $\bar{P}$. Mais precisamente, $\lim_{n\to\infty} P^{[n]} = \bar{P}$. Isso está de acordo com os dados abaixo:

$$P^{[1]} = AP^{[0]} = \begin{pmatrix} 0,200 \\ 0,525 \\ 0,275 \end{pmatrix}, \qquad P^{[2]} = AP^{[1]} = \begin{pmatrix} 0,285 \\ 0,455 \\ 0,260 \end{pmatrix},$$

$$P^{[3]} = AP^{[2]} = \begin{pmatrix} 0,305 \\ 0,452 \\ 0,243 \end{pmatrix}, \qquad P^{[4]} = AP^{[3]} = \begin{pmatrix} 0,3124 \\ 0,4486 \\ 0,2390 \end{pmatrix},$$

$$P^{[5]} = AP^{[4]} \approx \begin{pmatrix} 0,3147 \\ 0,4478 \\ 0,2375 \end{pmatrix}, \qquad P^{[6]} = AP^{[5]} \approx \begin{pmatrix} 0,3154 \\ 0,4475 \\ 0,2371 \end{pmatrix}.$$

Comentários: As ideias e resultados que apresentamos nesta seção são conhecidas do estudo de Cadeias de Markov em Processos Estocásticos. Em [4] o leitor encontrará uma discussão mais completa do tema, com aplicações. Outra referência, indicada aos leitores que estão tendo um primeiro contato com o assunto é [8]. Em [1] e [5] o leitor encontrará o tema discutido de forma elementar, com exemplos e cálculos do vetor estacionário.

Capítulo 2

Probabilidade invariante e Jacobiano

Nesta seção vamos considerar o conjunto

$$X = \{1, ..., n\} \times \{1, ..., n\} = \{(i, j) \,|\, 1 \le i \le n \text{ e } 1 \le j \le n\}.$$

Uma probabilidade sobre X será uma lista de números não negativos $\pi = (\pi_{ij})_{1 \le i,j \le n}$ satisfazendo $\sum_{i,j} \pi_{ij} = 1$. O número π_{ij} é o peso associado ao ponto (i, j) em X. Escrevemos $\pi((i, j)) = \pi_{ij}$. A probabilidade de um conjunto $A \subseteq X$ será por definição $\pi(A) = \sum_{(i,j) \in A} \pi_{ij}$. Convenientemente uma probabilidade sobre X poderá ser descrita por uma matriz do tipo $n \times n$ com entradas não negativas cuja soma resulta em 1. Denotamos por $\Pi(X)$ o conjunto das probabilidades sobre X.

Exemplo 2.1. *Para $n = 3$,*

$$X = \{(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (2, 3), (3, 1), (3, 2), (3, 3)\}.$$

A matriz

$$\pi = \begin{pmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{21} & \pi_{22} & \pi_{23} \\ \pi_{31} & \pi_{32} & \pi_{33} \end{pmatrix} = \begin{pmatrix} 0 & 0,1 & 0 \\ 0,2 & 0 & 0,2 \\ 0,1 & 0,3 & 0,1 \end{pmatrix}$$

representa uma probabilidade sobre X.

Dizemos que uma probabilidade π é positiva se a matriz que representa π é positiva, ou seja se $\pi((i, j)) > 0$ para qualquer ponto $(i, j) \in X$.

Fixado $k \in \{1, ..., n\}$ definimos

$$[\cdot, k] = \{(i, k) | i \in \{1, ..., n\}\} = \{(1, k), (2, k), ..., (n, k)\}$$

e

$$[k, \cdot] = \{(k, j) | j \in \{1, ..., n\}\} = \{(k, 1), (k, 2), ..., (k, n)\}.$$

Note que em geral $\pi([\cdot, k]) \neq \pi([k, \cdot])$. Por exemplo, se $X = \{1, 2\} \times \{1, 2\}$ e

$$\pi = \begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{pmatrix} = \begin{pmatrix} 1/2 & 1/2 \\ 0 & 0 \end{pmatrix}$$

então

$$\pi([\cdot, 1]) = \pi_{11} + \pi_{21} = 1/2 + 0 = 1/2.$$

enquanto

$$\pi([1, \cdot]) = \pi_{11} + \pi_{12} = 1/2 + 1/2 = 1.$$

Definição 2.2. *Uma probabilidade $\pi \in \Pi(X)$ será chamada invariante*[1] *se para todo $k \in \{1, ..., n\}$:*

$$\sum_{i=1}^{n} \pi_{ik} = \sum_{j=1}^{n} \pi_{kj}.$$

O conjunto das probabilidades invariantes será denotado por $\Pi(X, \sigma)$.

Exemplo 2.3. *Se $X = \{1, 2\} \times \{1, 2\}$ então a probabilidade*

$$\pi = \begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{pmatrix} = \begin{pmatrix} 1/8 & 2/8 \\ 2/8 & 3/8 \end{pmatrix}$$

é invariante. De fato, a soma dos elementos da ***primeira linha*** *da matriz coincide com a soma dos elementos da* ***primeira coluna*** *e a soma dos elementos da* ***segunda linha*** *da matriz coincide com a soma dos elementos da* ***segunda coluna****.*

Se π é uma probabilidade invariante então para todo $k \in \{1, ..., n\}$ temos

$$\pi([\cdot, k]) = \sum_{i=1}^{n} \pi_{ik} = \sum_{j=1}^{n} \pi_{kj} = \pi([k, \cdot]).$$

Note que a matriz que representa π no exemplo anterior é simétrica. Se a matriz que representa a probabilidade π é simétrica, então π será invariante. A recíproca desta afirmação não é verdadeira.

Exemplo 2.4. *Se $X = \{1, 2, 3\} \times \{1, 2, 3\}$ então a probabilidade*

$$\pi = \begin{pmatrix} 0.5 & 0.1 & 0.1 \\ 0.2 & 0 & 0 \\ 0 & 0.1 & 0 \end{pmatrix}$$

é invariante, mas a matriz que representa π não é simétrica.

[1] a expressão holonômica (ver [3], [11], [14]) pode ser mais adequada. Apresentamos uma discussão no final do capítulo.

Definição 2.5. *O Jacobiano de uma probabilidade invariante $\pi \in \Pi(X,\sigma)$ é a matriz $J^\pi = (J^\pi_{ij})$, onde*

$$J^\pi_{ij} = \begin{cases} \frac{\pi((i,j))}{\pi([\cdot,j])} = \frac{\pi_{ij}}{\sum_{l=1}^n \pi_{lj}} & se\, \pi([\cdot,j]) > 0 \\ 1/n & se\, \pi([\cdot,j]) = 0 \end{cases}.$$

Exemplo 2.6. *Para $X = \{1,2,3\} \times \{1,2,3\}$. Temos:*

$$se\ \pi = \begin{pmatrix} 0.5 & 0.1 & 0.1 \\ 0.2 & 0 & 0 \\ 0 & 0.1 & 0 \end{pmatrix}\ então\ J^\pi = \begin{pmatrix} 5/7 & 1/2 & 1 \\ 2/7 & 0 & 0 \\ 0 & 1/2 & 0 \end{pmatrix},$$

$$se\ \pi = \begin{pmatrix} 0.2 & 0.2 & 0.1 \\ 0.1 & 0.1 & 0.1 \\ 0.2 & 0 & 0 \end{pmatrix}\ então\ J^\pi = \begin{pmatrix} 2/5 & 2/3 & 1/2 \\ 1/5 & 1/3 & 1/2 \\ 2/5 & 0 & 0 \end{pmatrix},$$

$$se\ \pi = \begin{pmatrix} 0.3 & 0.1 & 0 \\ 0.1 & 0.5 & 0 \\ 0 & 0 & 0 \end{pmatrix}\ então\ J^\pi = \begin{pmatrix} 3/4 & 1/6 & 1/3 \\ 1/4 & 5/6 & 1/3 \\ 0 & 0 & 1/3 \end{pmatrix}.$$

O que será de interesse neste texto é o valor que J^π assume nos pontos onde $\pi_{ij} > 0$. Nos demais pontos, o valor de J^π poderia ser definido de forma arbitrária. A nossa escolha se justifica pelo próximo lema.

Lema 2.7. *Dada uma probabilidade $\pi \in \Pi(X,\sigma)$, seu Jacobiano J^π será uma matriz coluna estocástica. Se π é uma probabilidade positiva, seu Jacobiano será uma matriz coluna estocástica positiva.*

Demonstração: Iniciamos supondo que π é uma probabilidade positiva. Neste caso $J^\pi_{ij} > 0$ para quaisquer $i,j \in \{1,...,n\}$. Além disso, para cada j, a soma das entradas da coluna j da matriz J^π satisfaz

$$\sum_{i=1}^n J^\pi_{ij} = \sum_{i=1}^n \frac{\pi_{ij}}{\sum_{l=1}^n \pi_{lj}} = 1.$$

Concluímos que J^π é uma matriz coluna estocástica positiva.

Se π não é positiva, por definição teremos $J^\pi_{ij} \geq 0$ para quaisquer $i,j \in \{1,...,n\}$. Além disso, se $\pi([\cdot,j]) > 0$, podemos repetir as contas acima e concluir que a soma dos elementos da coluna j de J^π resulta em 1. Caso contrário (se $\pi([\cdot,j]) = 0$) escrevemos

$$\sum_{i=1}^n J^\pi_{ij} = \sum_{i=1}^n \frac{1}{n} = 1.$$

Em qualquer caso concluímos que J^π é uma matriz coluna estocástica. $\square$

Dada uma matriz $A = (a_{ij})$ coluna estocástica e positiva, vimos na seção anterior que existe um único vetor de probabilidade $\bar{P}$ tal que $A\bar{P} = \bar{P}$ (vetor estacionário). Denotamos por $p_1,...,p_n$ as coordenadas de $\bar{P}$.

Definição 2.8. *Dada uma matriz $A = (a_{ij})_{1 \leq i,j \leq n}$ coluna estocástica e positiva com vetor estacionário $\bar{P} = (p_j)_{1 \leq j \leq n}$, definimos a partir de A uma probabilidade $\pi \in \Pi(X)$ por $\pi_{ij} = a_{ij} p_j$ que será chamada* ***auto-probabilidade*** *associada a matriz A.*

Abaixo vamos mostrar que a auto-probabilidade π de fato é uma probabilidade. Dependendo do contexto, poderíamos chamá-la de medida de Gibbs, equilíbrio ou de Markov. Note que $\pi = AD$ onde D é a matriz diagonal determinada por $p_1, ..., p_n$. Para $n = 3$, por exemplo

$$\begin{pmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{21} & \pi_{22} & \pi_{23} \\ \pi_{31} & \pi_{32} & \pi_{33} \end{pmatrix} = \begin{pmatrix} a_{11}p_1 & a_{12}p_2 & a_{13}p_3 \\ a_{21}p_1 & a_{22}p_2 & a_{23}p_3 \\ a_{31}p_1 & a_{32}p_2 & a_{33}p_3 \end{pmatrix}$$

$$= \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} p_1 & 0 & 0 \\ 0 & p_2 & 0 \\ 0 & 0 & p_3 \end{pmatrix}.$$

Proposição 2.9. *Se A é uma matriz coluna estocástica e positiva, então a auto-probabilidade associada é uma probabilidade invariante e positiva.*

Demonstração: Denotamos por (a_{ij}) as entradas da matriz A e por $p_1, ..., p_n$ as coordenadas do vetor estacionário $\bar{P}$. Como

$$\sum_j \sum_i \pi_{ij} = \sum_j \sum_i a_{ij} p_j = \sum_j p_j \sum_i a_{ij} = \sum_j p_j = 1,$$

concluímos que π é uma probabilidade. Para provarmos que π é positiva, observamos que A é uma matriz coluna estocástica positiva e $\bar{P}$ é um vetor de probabilidade positivo (corolário 1.10). Segue que $\pi_{ij} = a_{ij} p_j$ é positivo para quaisquer i, j.

Para cada $k \in \{1, ..., n\}$ fixado, como $A\bar{P} = \bar{P}$, temos que

$$p_k = a_{k1} p_1 + a_{k2} p_2 + ... + a_{kn} p_n = \sum_j a_{kj} p_j.$$

Então

$$\pi([\cdot, k]) = \sum_i \pi_{ik} = \sum_i a_{ik} p_k = p_k \sum_i a_{ik} = p_k = \sum_j a_{kj} p_j = \sum_j \pi_{kj} = \pi([k, \cdot]),$$

portanto π é invariante. □

Proposição 2.10. *Se π é a auto-probabilidade associada a uma matriz A coluna estocástica e positiva, então $J^{\pi} = A$.*

Demonstração: Como $\pi_{ij} = a_{ij}p_j$ é positivo para quaisquer i, j, pela definição de Jacobiano temos

$$J^{\pi}_{ij} = \frac{\pi_{ij}}{\sum_l \pi_{lj}} = \frac{a_{ij}p_j}{\sum_l a_{lj}p_j} = \frac{a_{ij}p_j}{p_j \sum_l a_{lj}} = \frac{a_{ij}p_j}{p_j} = a_{ij}.$$

□

Pelo que vimos até aqui, dada uma matriz A coluna estocástica positiva, podemos associar a ela uma probabilidade invariante e positiva π (auto-probabilidade). Por outro lado para esta probabilidade π podemos associar uma matriz coluna estocástica positiva J^{π} que resultará na própria matriz A.

$$A \longrightarrow \pi \longrightarrow J^{\pi} = A.$$

No que segue vamos analisar se vale um argumento deste tipo partindo-se de uma probabilidade invariante e positiva π. Neste sentido, partindo-se de uma probabilidade invariante e positiva π podemos construir uma matriz coluna estocástica positiva J^{π}. Será que a auto-probabilidade de J^{π} coincide com π?

$$\pi \longrightarrow J^{\pi} \longrightarrow (?)$$

Proposição 2.11. *Se π é uma probabilidade invariante e positiva, então π é a auto-probabilidade de J^{π}.*

Demonstração: Denotamos $\sum_l \pi_{lj}$ por q_j. Note que $q_1 + ... + q_n = 1$. Como π é positiva, pela definição de Jacobiano obtemos que $J^{\pi}_{ij} = \frac{\pi_{ij}}{\sum_l \pi_{lj}} = \frac{\pi_{ij}}{q_j}$. Ou seja

$$J^{\pi}_{ij} q_j = \pi_{ij}. \tag{2.1}$$

Então para concluirmos que π é a auto-probabilidade de J^{π}, basta mostrarmos que o vetor estacionário $\bar{P}$ que satisfaz $\bar{P} = J^{\pi}\bar{P}$ tem coordenadas $q_1, ..., q_n$. Assim a prova estará completa se mostrarmos que $q_k = \sum_j J^{\pi}_{kj} q_j$ para todo k. Como π é uma probabilidade invariante

$$\sum_j J^{\pi}_{kj} q_j \overset{(2.1)}{=} \sum_j \pi_{kj} \overset{\pi \text{ é invariante}}{=} \sum_i \pi_{ik} = q_k,$$

concluindo a demonstração. □

Resumimos estes resultados abaixo:

Teorema 2.12. -
i) Dada uma matriz coluna estocástica e positiva A, sua auto-probabilidade π é uma probabilidade invariante e positiva. Além disso $J^{\pi} = A$.
ii) Dada uma probabilidade invariante e positiva π, seu Jacobiano J^{π} é uma matriz coluna estocástica e positiva. Além disso π é a auto-probabilidade associada a J^{π}.

Comentários: A definição de probabilidade invariante apresentada neste capítulo é uma adaptação do conceito de probabilidade invariante em Sistemas Dinâmicos, [13], [16], [18], não correspondendo na íntegra a definição usual. No entanto ela está ligada ao conceito de probabilidade holonômica em sistemas de funções iteradas, [3], [11], [14]. Dados conjuntos $Y = \{1, ..., n\}$ e $Z = \{1, ..., m\}$, um sistema de funções iteradas (IFS) consiste em uma família de aplicações $\{\tau_i : Z \to Z \,|\, i \in Y\}$. Dado um IFS, dizemos que uma probabilidade $\pi = (\pi_{ij})$ sobre $Y \times Z$ é holonômica se para qualquer função $h : Z \to \mathbb{R}$

$$\sum_{i\in Y}\sum_{j\in Z} h(\tau_i(j))\pi_{ij} = \sum_{i\in Y}\sum_{j\in Z} h(j)\pi_{ij}.$$

Quando $Y = Z = \{1, ..., n\}$ e o IFS satisfaz $\tau_i(j) = i$ para quaisquer $i \in Y$,$j \in Z$ temos que π será holonômica se para qualquer função $h : \{1, ..., n\} \to \mathbb{R}$

$$\sum_{1\le i\le n}\sum_{1\le j\le n} h(i)\pi_{ij} = \sum_{1\le i\le n}\sum_{1\le j\le n} h(j)\pi_{ij}.$$

Afirmamos que isso equivale a dizer que para todo $k \in \{1, ..., n\}$,

$$\sum_{1\le j\le n} \pi_{kj} = \sum_{1\le i\le n} \pi_{ik}.$$

De fato, fixados $k \in \{1, ..., n\}$ e a função δ_k satisfazendo $\delta_k(j) = 1$ se $j = k$ e $\delta_k(j) = 0$ se $j \neq k$, se π é holonômica então

$$\sum_{1\le i\le n}\sum_{1\le j\le n} \delta_k(i)\pi_{ij} = \sum_{1\le i\le n}\sum_{1\le j\le n} \delta_k(j)\pi_{ij}$$

logo

$$\sum_{1\le j\le n} \pi_{kj} = \sum_{1\le i\le n} \pi_{ik}.$$

Por outro lado, se π satisfaz

$$\sum_{1\le j\le n} \pi_{kj} = \sum_{1\le i\le n} \pi_{ik},$$

então para qualquer função $h : \{1, ..., n\} \to \mathbb{R}$ temos

$$\sum_{1\le i\le n}\sum_{1\le j\le n} h(i)\pi_{ij} = \sum_{1\le i\le n}\left[h(i)\sum_{1\le j\le n}\pi_{ij}\right] = \sum_{1\le i\le n}\left[h(i)\sum_{1\le j\le n}\pi_{ji}\right]$$
$$= \sum_{1\le i\le n}\sum_{1\le j\le n} h(i)\pi_{ji} = \sum_{1\le i\le n}\sum_{1\le j\le n} h(j)\pi_{ij},$$

garantindo que π é holonômica para o IFS satisfazendo $\tau_i(j) = i$.

Capítulo 3

Entropia

Nesta seção continuamos considerando o espaço $X = \{1, ..., n\} \times \{1, ..., n\}$.

Definição 3.1. *Dadas uma matriz $A = (a_{ij})_{1\leq i,j\leq n}$ e uma probabilidade $\pi = (\pi_{ij})$ em $\Pi(X)$, a média de A em relação a π é dada por*

$$\langle A, \pi\rangle = \sum_{i,j} a_{ij}\pi_{ij}.$$

Note que $\langle A, \pi\rangle$ representa uma média ponderada das entradas de A com pesos dados pela probabilidade π.

Definição 3.2. *Se $\pi = (\pi_{ij})$ é uma probabilidade invariante, definimos sua entropia[1] por*

$$H(\pi) = -\sum_{i,j} \log(J^{\pi}_{ij})\pi_{ij}.$$

Acima estamos considerando o logaritmo natural $\log(x) = \ln(x)$. Note que se $J^{\pi}_{ij} = 0$ então $\pi_{ij} = 0$. Sempre que isso ocorrer assumimos que $0\log(0) = 0$, ou seja, desconsideramos esta parcela na soma acima. De fato, como já comentamos na seção anterior, as entradas de J^{π} que serão de interesse neste texto são aquelas onde $\pi_{ij} > 0$ e somente nestes casos teremos parcelas consideradas na soma acima.

Exemplo 3.3. *Se $\pi_{ij} = \frac{1}{n^2}$ para quaisquer i, j, então*

$$J^{\pi}_{ij} = \frac{\pi_{ij}}{\sum_s \pi_{sj}} = \frac{1/n^2}{1/n} = \frac{1}{n}.$$

Segue que

$$H(\pi) = -\sum_{i,j} \log(1/n)\frac{1}{n^2} = -n^2 \log(1/n)\frac{1}{n^2} = -\log(1/n) = \log(n).$$

[1] Esta definição não coincide com a definição de entropia de Shannon para probabilidades em conjuntos finitos. Também não coincide exatamente com a entropia de Kolmogorov-Sinai para probabilidades invariantes em sistemas dinâmicos. É equivalente as definições apresentadas em [11] e [14] para probabilidades holonômicas.

Exemplo 3.4. *Seja $\pi = (\pi_{ij})$, onde $\pi_{11} = 1$ (então todas as demais entradas de π são nulas, pois π é uma probabilidade). π é invariante, pois $\pi([i,\cdot]) = \pi([\cdot,i]) = 0$, se $i \neq 1$ e $\pi([1,\cdot]) = \pi([\cdot,1]) = 1$. Temos que $J^\pi_{11} = 1$ e*

$$H(\pi) = -\sum_{i,j} \log(J^\pi_{ij})\pi_{ij} = -\log(J^\pi_{11})\pi_{11} = -\log(1) = 0.$$

Como veremos abaixo a entropia de uma probabilidade invariante pertence ao intervalo $[0, \log(n)]$. No primeiro exemplo acima, consideramos a probabilidade com distribuição uniforme sobre X. Neste caso, a entropia obtida é máxima ($\log(n)$). No segundo exemplo, consideramos uma probabilidade concentrada em um único ponto de X ($\pi_{11} = 1$). Neste caso, a entropia obtida é mínima (zero).

Exemplo 3.5. *Suponha $n = 2$ e portanto $X = \{1,2\} \times \{1,2\}$. Sejam p_1 e p_2 números positivos satisfazendo $p_1 + p_2 = 1$. Considere*

$$\pi = \begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{pmatrix} = \begin{pmatrix} p_1p_1 & p_1p_2 \\ p_2p_1 & p_2p_2 \end{pmatrix}.$$

Como p_1 e p_2 são positivos concluímos que π é positiva. Além disso

$$\pi_{11} + \pi_{12} + \pi_{21} + \pi_{22} = p_1p_1 + p_1p_2 + p_2p_1 + p_2p_2 = p_1(p_1 + p_2) + p_2(p_1 + p_2)$$
$$= (p_1 + p_2)(p_1 + p_2) = (1)(1) = 1.$$

Isso mostra que π é uma probabilidade. Por ser simétrica, concluímos que π é invariante.

O Jacobiano de π é dado por

$$J^\pi = \begin{pmatrix} \frac{p_1p_1}{p_1p_1+p_2p_1} & \frac{p_1p_2}{p_1p_2+p_2p_2} \\ \\ \frac{p_2p_1}{p_1p_1+p_2p_1} & \frac{p_2p_2}{p_1p_2+p_2p_2} \end{pmatrix} = \begin{pmatrix} p_1 & p_1 \\ \\ p_2 & p_2 \end{pmatrix}.$$

Portanto

$$H(\pi) = -\sum_{i,j} \log(J^\pi_{ij})\pi_{ij}$$
$$= -[\log(p_1)p_1p_1 + \log(p_1)p_1p_2 + \log(p_2)p_2p_1 + \log(p_2)p_2p_2]$$
$$= -[\log(p_1)p_1 + \log(p_2)p_2].$$

Como $p_2 = 1 - p_1$, podemos escrever

$$H(\pi) = -[\log(p_1)p_1 + \log(1-p_1)(1-p_1)].$$

A função $\psi(t) = -\log(t)(t) - \log(1-t)(1-t)$, $t \in (0,1)$, tem derivada $\psi'(t) = -\log(\frac{t}{1-t})$, sendo fácil verificar que ψ é crescente no intervalo $(0, \frac{1}{2})$ e decrescente no intervalo $(\frac{1}{2}, 1)$. A entropia de π é máxima se $p_1 = p_2 = \frac{1}{2}$ e decresce conforme os números p_1 e p_2 se distanciam.

p_1	p_2	π	$H(\pi)$
$0,5$	$0,5$	$\begin{pmatrix} 0,25 & 0,25 \\ 0,25 & 0,25 \end{pmatrix}$	$0,6931$
$0,4$	$0,6$	$\begin{pmatrix} 0,16 & 0,24 \\ 0,24 & 0,36 \end{pmatrix}$	$0,5108$
$0,3$	$0,7$	$\begin{pmatrix} 0,09 & 0,21 \\ 0,21 & 0,49 \end{pmatrix}$	$0,3567$
$0,2$	$0,8$	$\begin{pmatrix} 0,04 & 0,16 \\ 0,16 & 0,64 \end{pmatrix}$	$0,2231$
$0,1$	$0,9$	$\begin{pmatrix} 0,01 & 0,09 \\ 0,09 & 0,81 \end{pmatrix}$	$0,1054$

Exemplo 3.6. *Generalizando o exemplo anterior, dados números positivos $p_1, ..., p_n$, tais que $p_1 + ... + p_n = 1$, defina uma probabilidade π sobre $X = \{1, ..., n\} \times \{1, ..., n\}$ por*

$$\pi_{ij} = p_i \cdot p_j.$$

π é de fato uma probabilidade positiva, pois $\pi_{ij} = p_i \cdot p_j > 0$ e

$$\sum_i \sum_j \pi_{ij} = \sum_i \big[\sum_j p_i \cdot p_j \big] = \sum_i \big[p_i \sum_j p_j \big] = \sum_i p_i = 1.$$

Além disso, π é invariante, pois é simétrica ($\pi_{ij} = \pi_{ji}$).

O Jacobiano de π é dado por

$$J_{ij}^{\pi} = \frac{\pi_{ij}}{\sum_l \pi_{lj}} = \frac{p_i \cdot p_j}{\sum_l p_l \cdot p_j} = \frac{p_i}{\sum_l p_l} = p_i$$

e a entropia de π é igual a

$$H(\pi) = -\sum_{i,j} \log(J_{ij}^{\pi}) \pi_{ij} = -\sum_i \sum_j \log(p_i) p_i p_j$$

$$= -\sum_i \log(p_i) p_i \sum_j p_j = -\sum_i \log(p_i) p_i.$$

Exemplo 3.7. *Dada uma matriz coluna estocástica positiva $A = (a_{ij})$ com vetor estacionário $\bar{P} = (p_j)$ e auto-probabilidade $\pi = (\pi_{ij}) = (a_{ij} p_j)$ temos que $J^{\pi} = A$. Portanto*

$$H(\pi) = -\sum_{i,j} \log(a_{ij}) a_{ij} p_j.$$

Vamos provar neste capítulo o seguinte

Teorema 3.8. *Dada uma probabilidade invariante π,*

$$H(\pi) = \inf\{-\sum_{i,j} \log(b_{ij})\pi_{ij} \mid B = (b_{ij})_{1\leq i,j\leq n} \text{ é coluna estocástica e positiva}\}.$$

Antes da prova vamos apresentar algumas de suas aplicações.

Corolário 3.9. *Para qualquer probabilidade invariante π, $0 \leq H(\pi) \leq \log(n)$.*

Demonstração: Como $J^{\pi}_{ij} \leq 1$ para quaisquer i, j, temos que $\log(J^{\pi}_{ij}) \leq 0$ e $\sum_{ij} \log(J^{\pi}_{ij})\pi_{ij} \leq 0$. Portanto $H(\pi) = -\sum_{ij} \log(J^{\pi}_{ij})\pi_{ij} \geq 0$. Por outro lado, aplicando-se o teorema anterior para a matriz constante $B = (1/n)$:

$$H(\pi) = -\sum_{i,j} \log(J^{\pi}_{ij})\pi_{ij} \leq -\sum_{i,j} \log(\frac{1}{n})\pi_{ij} = \log(n).$$

□

Corolário 3.10. *O conjunto das probabilidades invariantes é convexo e a entropia é uma função côncava sobre este conjunto. Mais precisamente, se π e η são probabilidades invariantes sobre X e $\lambda \in [0, 1]$, temos*
a. $\lambda\pi + (1 - \lambda)\eta$ é uma probabilidade invariante
b. $H(\lambda\pi + (1 - \lambda)\eta) \geq \lambda H(\pi) + (1 - \lambda)H(\eta)$.

Demonstração: Sejam $\pi = (\pi_{ij})$ e $\eta = (\eta_{ij})$ probabilidades invariantes sobre $X = \{1, ..., n\} \times \{1, ..., n\}$ e fixe $\lambda \in [0, 1]$. Denotamos por $\gamma = \lambda\pi + (1 - \lambda)\eta$, a combinação convexa de π e η com pesos λ e $(1 - \lambda)$. Queremos inicialmente mostrar que γ é uma probabilidade invariante. Para isso observamos que $\pi_{ij} \geq 0$ e $\eta_{ij} \geq 0$, então $\gamma_{ij} = \lambda\pi_{ij} + (1 - \lambda)\eta_{ij} \geq 0$. Além disso

$$\sum_{i,j} \gamma_{ij} = \sum_{i,j} (\lambda\pi_{ij} + (1 - \lambda)\eta_{ij}) = \lambda \sum_{i,j} \pi_{ij} + (1-\lambda) \sum_{i,j} \eta_{ij} = \lambda\cdot 1 + (1-\lambda)\cdot 1 = 1,$$

garantindo que γ é uma probabilidade.

Para mostrarmos que γ é invariante observamos que, fixado $k \in \{1, ..., n\}$, $\sum_i \pi_{ik} = \sum_j \pi_{kj}$ e $\sum_i \eta_{ik} = \sum_j \eta_{kj}$. Então

$$\sum_i \gamma_{ik} = \lambda \sum_i \pi_{ik} + (1 - \lambda) \sum_i \eta_{ik} = \lambda \sum_j \pi_{kj} + (1 - \lambda) \sum_j \eta_{kj} = \sum_j \gamma_{kj}.$$

Por fim, vamos mostrar que $H(\gamma) \geq \lambda H(\pi) + (1 - \lambda)H(\eta)$. Com este objetivo, fixamos $\epsilon > 0$. Pelo Teorema 3.8 existe uma matriz coluna estocástica positiva $B = (B_{ij})$ tal que $H(\gamma) > -\sum_{i,j} \log(B_{ij})\gamma_{ij} - \epsilon$. Usando a definição de γ e o Teorema 3.8 aplicado em π e η temos:

$$H(\gamma) > \lambda \left(-\sum_{i,j} \log(B_{ij})\pi_{ij}\right) + (1 - \lambda) \left(-\sum_{i,j} \log(B_{ij})\eta_{ij}\right) - \epsilon$$

$$> \lambda H(\pi) + (1-\lambda) H(\eta) - \epsilon.$$

Considerando ϵ arbitrariamente pequeno obtemos que

$$H(\gamma) \geq \lambda H(\pi) + (1-\lambda) H(\eta).$$

□

Exemplo 3.11. *Vamos mostrar um exemplo onde* $H(\frac{1}{2}\pi + \frac{1}{2}\eta) > \frac{1}{2}H(\pi) + \frac{1}{2}H(\eta)$. *Considere probabilidades em* $\{1,2\} \times \{1,2\}$ *definidas por*

$$\begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{pmatrix} = \begin{pmatrix} 0 & 1/2 \\ 1/2 & 0 \end{pmatrix} \; e \; \begin{pmatrix} \eta_{11} & \eta_{12} \\ \eta_{21} & \eta_{22} \end{pmatrix} = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}.$$

Neste caso

$$J^{\pi} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \; e \; J^{\eta} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Portanto $H(\pi) = -\sum_{i,j} \log(J^{\pi}_{ij})\pi_{ij} = 0$ *e* $H(\eta) = -\sum_{i,j} \log(J^{\eta}_{ij})\eta_{ij} = 0$. *No entanto*

$$\gamma = \frac{1}{2}\pi + \frac{1}{2}\eta = \begin{pmatrix} 1/4 & 1/4 \\ 1/4 & 1/4 \end{pmatrix}$$

satisfaz

$$J^{\gamma} = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}.$$

Como consequência temos

$$H(\gamma) = -\sum_{i,j} \log(J^{\gamma}_{ij})\gamma_{ij} = -\sum_{i,j} \log\left(\frac{1}{2}\right)\frac{1}{4}$$

$$= -4\log\left(\frac{1}{2}\right)\frac{1}{4} = -\log(1/2) = \log(2).$$

Podemos identificar uma probabilidade sobre $X = \{1,...,n\} \times \{1,...,n\}$ com um elemento de $\mathbb{R}^{n^2}$. Com esta identificação podemos induzir uma métrica no conjunto das probabilidades invariantes. Desta forma, dizemos que uma sequência de probabilidades $\{\pi^n\}_{n=1,2,...}$ converge para uma probabilidade π (quando $n \to +\infty$) se $\lim\limits_{n\to+\infty} \pi^n_{ij} = \pi_{ij}$ para quaisquer $i,j \in \{1,...,n\}$.

Proposição 3.12. *Dada uma sequência de probabilidades invariantes* $\pi^1, \pi^2, ...$ *convergindo para* π*, temos que*
a) π *é uma probabilidade invariante;*
b) $H(\pi) \geq \limsup\limits_{n\to+\infty} H(\pi^n)$.

Demonstração: a):

$$\pi_{ij} = \lim_{n\to+\infty} \pi_{ij}^n \geq 0$$

e

$$\sum_{i,j} \pi_{ij} = \sum_{i,j} \lim_{n\to+\infty} \pi_{ij}^n = \lim_{n\to+\infty} \sum_{i,j} \pi_{ij}^n = \lim_{n\to+\infty} 1 = 1,$$

garantindo que de fato π é uma probabilidade. Para mostrarmos que π é invariante escrevemos para cada $k \in \{1, ..., n\}$:

$$\sum_i \pi_{ik} = \sum_i \lim_{n\to+\infty} \pi_{ik}^n = \lim_{n\to+\infty} \sum_i \pi_{ik}^n$$

$$= \lim_{n\to+\infty} \sum_j \pi_{kj}^n = \sum_j \lim_{n\to+\infty} \pi_{kj}^n = \sum_j \pi_{kj}.$$

b): Dado $\epsilon > 0$, existe uma matriz coluna estocástica positiva $B = (B_{ij})$ tal que

$$H(\pi) > \left(- \sum_{i,j} \log(B_{ij})\pi_{ij} \right) - \epsilon = \lim_{n\to+\infty} - \sum_{i,j} \log(B_{ij})\pi_{ij}^n - \epsilon$$

$$\geq \limsup_{n\to+\infty} H(\pi^n) - \epsilon.$$

Considerando ϵ arbitrariamente pequeno concluímos a prova. $\square$

No que segue vamos apresentar resultados que nos auxiliam na prova do Teorema 3.8.

Lema 3.13. *Se $B = (b_{ij})$ e $A = (a_{ij})$ são matrizes $n \times n$ onde A é coluna estocástica e positiva com auto-probabilidade $\pi = (\pi_{ij})$ e vetor estacionário $\bar{P} = (p_i)$, então*

$$\langle B, \pi \rangle = \sum_{i,j} b_{ij}\pi_{ij} = \sum_{i,j} \left(\sum_l a_{li} b_{li} \right) \pi_{ij} = \sum_i \left(\sum_l a_{li} b_{li} \right) p_i.$$

Demonstração: Iniciamos observando que $\sum_j \pi_{ij} = \sum_j a_{ij} p_j = p_i$, pois $A \cdot \bar{P} = \bar{P}$. Então

$$\sum_{i,j} \sum_l a_{li} b_{li} \pi_{ij} = \sum_{i,l} a_{li} b_{li} p_i = \sum_{i,l} b_{li} a_{li} p_i = \sum_{i,l} b_{li} \pi_{li} = \sum_{i,j} b_{ij} \pi_{ij}.$$

$\square$

Lema 3.14. *Seja $\pi = (\pi_{ij})$ uma probabilidade invariante, não necessariamente positiva e $B = (b_{ij})$ uma matriz $n \times n$. Então*

$$\langle B, \pi \rangle = \sum_{i,j} b_{ij}\pi_{ij} = \sum_{i,j} \left(\sum_l J_{li}^\pi b_{li} \right) \pi_{ij} = \sum_{i,l} (J_{li}^\pi b_{li})\, \pi([., i]).$$

Demonstração: Se $\pi([i,\cdot]) = \pi([\cdot, i]) > 0$ então

$$J^{\pi}_{li} = \frac{\pi_{li}}{\sum_s \pi_{si}}.$$

Denotando por i' os indices i para os quais $\pi([i,\cdot]) = \pi([\cdot, i]) > 0$ temos:

$$\sum_{i,j}\left(\sum_l J^{\pi}_{li} b_{li}\right)\pi_{ij} = \sum_{i',j}\left(\sum_l J^{\pi}_{li'} b_{li'}\right)\pi_{i'j} = \sum_{i',j}\left(\sum_l \frac{\pi_{li'}}{\sum_s \pi_{si'}} b_{li'}\right)\pi_{i'j}$$

$$= \sum_{i',l}\left[\left(\frac{\pi_{li'} b_{li'}}{\sum_s \pi_{si'}}\right)\sum_j \pi_{i'j}\right] = \sum_{i',l}\pi_{li'} b_{li'} = \sum_{l,i}\pi_{li} b_{li} = \sum_{i,j} b_{ij}\pi_{ij}$$

$\square$

Definição 3.15. *Dizemos que uma função $f : I \to \mathbb{R}$ é estritamente côncava se para quaisquer valores $x_1 \neq x_2$ no intervalo I e $\lambda \in (0,1)$*

$$f(\lambda x_1 + (1-\lambda)x_2) > \lambda f(x_1) + (1-\lambda) f(x_2).$$

Lema 3.16. *A função $\psi : [0,1] \to \mathbb{R}$ definida por*

$$\psi(x) = \begin{cases} -x\log(x) & se\, x \in (0,1] \\ 0 & se\, x = 0 \end{cases} \tag{3.1}$$

é contínua e estritamente côncava.

Não iremos demostrar este resultado. O leitor é convidado, no entanto, a verificar que $\lim_{x\to 0^+} x\log(x) = 0$ (L'Hôpital) e que a derivada segunda de $x\log(x)$ é negativa em $(0, +\infty)$.

Lema 3.17 (Desigualdade de Jensen). *Dada uma função estritamente côncava $f : I \to \mathbb{R}$, números distintos $a_1, ..., a_k$ no intervalo I e números positivos $p_1, ..., p_k$, satisfazendo $p_1 + ... + p_k = 1$, onde $k \geq 2$:*

$$f\left(\sum_{i=1}^k a_i p_i\right) > \sum_{i=1}^k f(a_i) p_i.$$

Demonstração: Vamos provar o resultado por indução matemática. Para $k = 2$ o resultado segue da concavidade estrita da função f. Supondo verdadeira a afirmação para um determinado valor de k, vamos provar a afirmação para $k+1$:

$$f\left(\sum_{i=1}^{k+1} a_i p_i\right) = f\left(\sum_{i=1}^{k} a_i p_i + a_{k+1}p_{k+1}\right)$$

$$= f\left((1-p_{k+1})\sum_{i=1}^{k} a_i \frac{p_i}{(1-p_{k+1})} + a_{k+1}p_{k+1}\right)$$

(como $(1-p_{k+1})+p_{k+1}=1$, aplicando a Desigualdade de Jensen para 2 termos)

$$> (1-p_{k+1}) f\left(\sum_{i=1}^{k} a_i \frac{p_i}{(1-p_{k+1})}\right) + p_{k+1} f(a_{k+1})$$

(como $\frac{p_1+\ldots+p_k}{1-p_{k+1}}=1$, aplicando a Desigualdade de Jensen para k termos)

$$> (1-p_{k+1}) \sum_{i=1}^{k} f(a_i) \frac{p_i}{(1-p_{k+1})} + p_{k+1} f(a_{k+1}) = \sum_{i=1}^{k+1} f(a_i) p_i.$$

□

Lema 3.18. *Se $P=(p_1,\ldots,p_k)$ é um vetor de probabilidade positivo e $Q=(q_1,\ldots,q_k)$ é um vetor de probabilidade então*

$$\sum_{i=1}^{k} q_i \log(q_i) \geq \sum_{i=1}^{k} q_i \log(p_i).$$

A igualdade ocorre se e somente se $P=Q$.

Demonstração: Vamos supor $P \neq Q$. Considere a função ψ estritamente côncava dada em 3.1. Temos:

$$\sum_{i=1}^{k} q_i \log(q_i) - \sum_{i=1}^{k} q_i \log(p_i) = \sum_{i=1}^{k} q_i \log(\frac{q_i}{p_i}) = \sum_{i=1}^{k} p_i \frac{q_i}{p_i} \log(\frac{q_i}{p_i})$$

$$= -\sum_{i=1}^{k} p_i \psi(\frac{q_i}{p_i}) \overset{Jensen}{>} -\psi\left(\sum_{i=1}^{k} p_i \frac{q_i}{p_i}\right) = -\psi(1) = 0.$$

□

Lema 3.19. *Seja $B=(b_{ij})$ uma matriz $n \times n$ coluna estocástica positiva e π uma probabilidade invariante. Então*

$$H(\pi) = -\sum_{i,j} \log(J_{ij}^{\pi}) \pi_{ij} \leq -\sum_{i,j} \log(b_{ij}) \pi_{ij}.$$

A igualdade ocorre se e somente se $b_{ij} = J_{ij}^{\pi}$ sempre que $\pi_{ij} > 0$.

Demonstração: Segue do lema anterior que para cada i fixado

$$\sum_{s} J_{si}^{\pi} \log(J_{si}^{\pi}) \geq \sum_{s} J_{si}^{\pi} \log(b_{si}).$$

Supondo $b_{si} \neq J_{si}^{\pi}$ em algum ponto (s,i) satisfazendo $\pi_{si} > 0$, como a desigualdade acima será estrita, obtemos

$$\sum_{i,s} J_{si}^{\pi} \log(J_{si}^{\pi}) \pi([\cdot, i]) > \sum_{i,s} J_{si}^{\pi} \log(b_{si}) \pi([\cdot, i]).$$

A partir do Lema 3.14 obtemos

$$\sum_{i,j} \log(J_{ij}^{\pi})\pi_{ij} = \sum_{i,s} J_{si}^{\pi} \log(J_{si}^{\pi})\pi([\cdot, i])$$

e

$$\sum_{i,j} \log(b_{ij})\pi_{ij} = \sum_{i,s} J_{si}^{\pi} \log(b_{si})\pi([\cdot, i]).$$

Portanto, da desigualdade anterior,

$$\sum_{i,j} \log(J_{ij}^{\pi})\pi_{ij} > \sum_{i,j} \log(b_{ij})\pi_{ij}.$$

□

Prova do Teorema 3.8: Como consequência do lema anterior, para toda matriz coluna estocástica e positiva $B = (b_{ij})$,

$$H(\pi) \leq -\sum_{i,j} \log(b_{ij})\pi_{ij}.$$

Portanto

$$H(\pi) \leq \inf\{-\sum_{i,j} \log(b_{ij})\pi_{ij} \mid B = (b_{ij}) \text{ é coluna estocástica e positiva}\}.$$

Queremos provar a igualdade. Se J^{π} é positiva basta tomarmos $B = J^{\pi}$.

Suponhamos que J^{π} não seja positiva. Fixado $0 < \epsilon < 1$ vamos construir uma matriz coluna estocástica B^{ϵ} pelo procedimento a seguir: Para cada $j \in \{1, ..., n\}$ fixado, analisamos a coluna j de J^{π}. Se todas as entradas desta coluna são positivas, então a $j-$ésima coluna de B^{ϵ} coincide com a $j-$ésima coluna de J^{π}. Se alguma entrada da coluna j é nula então para $J_{ij}^{\pi} \neq 0$ definimos $B_{ij}^{\epsilon} = (1-\epsilon)J_{ij}^{\pi}$ e para as demais entradas da coluna colocamos o número positivo que garanta que $\sum_i B_{ij}^{\epsilon} = 1$. Segue desta construção que se $\pi_{ij} > 0$ então

$$B_{ij}^{\epsilon} = \begin{cases} J_{ij}^{\pi}, & \text{se todas as entradas da coluna j de } J^{\pi} \text{ são positivas} \\ (1-\epsilon)J_{ij}^{\pi}, & \text{se alguma entrada da coluna j de } J^{\pi} \text{ é nula} \end{cases}.$$

Portanto

$$-\sum_{i,j} \log(B_{ij}^{\epsilon})\pi_{ij} \leq -\sum_{i,j} \log((1-\epsilon)J_{ij}^{\pi})\pi_{ij} = -\sum_{i,j} \log(J_{ij}^{\pi})\pi_{ij} - \sum_{i,j} \log((1-\epsilon))\pi_{ij}$$
$$= H(\pi) - \log(1-\epsilon).$$

Assim

$$H(\pi) \geq -\sum_{i,j} \log(B_{ij}^{\epsilon})\pi_{ij} + \log(1-\epsilon),$$

onde $\log(1-\epsilon) < 0$. Consequentemente obtemos que

$$H(\pi) \geq \inf\{-\sum_{i,j} \log(b_{ij})\pi_{ij} \mid B = (b_{ij}) \text{ é coluna estocástica e positiva}\} + \log(1-\epsilon).$$

Como ϵ pode ser arbitrariamente pequeno, concluímos a prova.

□

Capítulo 4

Discussão relacionada ao Formalismo Termodinâmico

Nesta seção continuamos considerando o conjunto $X = \{1, ..., n\} \times \{1, ..., n\}$. Uma função $F : X \to \mathbb{R}$ pode ser identificada a uma matriz de entradas (F_{ij}), onde $F_{ij} = F(i, j)$. Da mesma forma uma função $\psi : \{1, ..., n\} \to \mathbb{R}$ pode ser identificada a um vetor de R^n. Vamos considerar a função $e^{F(i,j)}$ sobre X. Note que e^F pode ser representada por uma matriz cujas entradas são $(e^{F(i,j)})$.

Exemplo 4.1. *Para $X = \{1, 2\} \times \{1, 2\}$, considere a função $F : X \to \mathbb{R}$ satisfazendo:*

$$F(1,1) = 3,\ F(1,2) = 5,\ F(2,1) = -7,\ F(2,2) = -11.$$

Então a função composta e^F pode ser representada pela matriz

$$\begin{pmatrix} e^{F(1,1)} & e^{F(1,2)} \\ e^{F(2,1)} & e^{F(2,2)} \end{pmatrix} = \begin{pmatrix} e^3 & e^5 \\ e^{-7} & e^{-11} \end{pmatrix}.$$

Vamos também considerar o operador linear L_F agindo em $\mathbb{R}^n$ e definido por $(L_F(\psi))_j = \sum_i e^{F(i,j)} \psi_i$. Podemos interpretar $L_F(\psi)$ como o vetor linha dado por $\psi \cdot e^F$ onde ψ é também um vetor linha. Se $X = \{1, 2, 3\} \times \{1, 2, 3\}$, por exemplo, e $\varphi = L_F(\psi)$ então

$$(\varphi_1 \ \ \varphi_2 \ \ \varphi_3) = (\psi_1 \ \ \psi_2 \ \ \psi_3) \begin{pmatrix} e^{F(1,1)} & e^{F(1,2)} & e^{F(1,3)} \\ e^{F(2,1)} & e^{F(2,2)} & e^{F(2,3)} \\ e^{F(3,1)} & e^{F(3,2)} & e^{F(3,3)} \end{pmatrix}.$$

Temos assim uma aplicação linear dada por uma matriz agindo pela direita em vetores linha.

Podemos também escrever $(\varphi)^T = (\psi \cdot e^F)^T = (e^F)^T (\psi)^T$. Neste caso a matriz transposta $(e^F)^T$ age pela esquerda em vetores coluna. No exemplo anterior obtemos

$$\begin{pmatrix}\varphi_1\\ \varphi_2\\ \varphi_3\end{pmatrix} = \begin{pmatrix} e^{F(1,1)} & e^{F(2,1)} & e^{F(3,1)}\\ e^{F(1,2)} & e^{F(2,2)} & e^{F(3,2)}\\ e^{F(1,3)} & e^{F(2,3)} & e^{F(3,3)}\end{pmatrix}\begin{pmatrix}\psi_1\\ \psi_2\\ \psi_3\end{pmatrix}.$$

Teorema 4.2 (Perron-Frobenius: parte I)**.** *Se A é uma matriz quadrada e positiva, então A admite um autovalor positivo associado a um autovetor positivo.*

Demonstração: Vamos escrever $B > 0$ se todas as entradas da matriz (ou vetor) B forem positivas e $B \geq 0$ se todas as entradas de B forem maiores ou iguais a zero. $B > C$ se $(B - C) > 0$ e $B \geq C$ se $(B - C) \geq 0$. Denotamos por $\mathbb{I}$ o vetor de $\mathbb{R}^n$ com todas as coordenadas iguais a 1 e por $\mathbb{H}$ a matriz $n \times n$ com todas as entradas iguais a 1. Fixamos $A_{n\times n}$, $A > 0$.

Note que

$$\mathbb{H}\mathbb{I} = n\mathbb{I}.$$

Seja I o conjunto dos números λ satisfazendo a seguinte propriedade: Existe algum vetor não nulo $x \geq 0$ (que depende de λ) tal que $Ax \geq \lambda x$.

Afirmação 1: existe um número $\lambda_0 > 0$ tal que $I = (-\infty, \lambda_0]$.
De fato, se $\lambda_1 < \lambda_2$ e $\lambda_2 \in I$ então $\lambda_1 \in I$ (pode ser tomado o mesmo x que se aplica a λ_2). Seja a o valor da menor entrada de A. Então

$$A\mathbb{I} \geq a\mathbb{H}\mathbb{I} = na\mathbb{I}.$$

Isso mostra que $na \in I$ e portanto I contém o intervalo $(-\infty, na]$. Seja b o valor da maior entrada de A. Se $x \geq 0$ é um vetor não nulo e x_i é sua maior entrada então

$$Ax \leq (b\mathbb{H})(x_i\mathbb{I}) = nbx_i\mathbb{I}.$$

Em particular a coordenada i de Ax é limitada por nbx_i. Isso mostra que se $\lambda > nb$ então λ não pertence a I. Concluímos que existe um número $\lambda_0 \in [na, nb]$ tal que $I = (-\infty, \lambda_0)$ ou $I = (-\infty, \lambda_0]$. Suponhamos por absurdo que $I = (-\infty, \lambda_0)$. Seja $\lambda_1 < \lambda_2 < ...$ uma sequência crescente de números convergindo para λ_0. Como $\lambda_i \in I, i = 1, 2, ...$ podemos escolher um vetor não nulo $x^i \geq 0$ tal que $Ax^i \geq \lambda_i x^i$. Note que esta desigualdade se mantém se multiplicarmos ambos os lados por uma constante positiva e portanto podemos supor que os vetores x^i são unitários. Considere uma subsequência convergente de $x^1, x^2, x^3, ...$ Seja x o vetor limite desta subsequência. Como $Ax^i - \lambda_i x^i \geq 0$, concluímos que $Ax - \lambda_0 x \geq 0$. Isso garante que $\lambda_0 \in I$, contrariando a hipótese de ser $I = (-\infty, \lambda_0)$. Concluímos assim a prova da afirmação.

Seja $x_0 \geq 0$ um vetor não nulo tal que $Ax_0 \geq \lambda_0 x_0$. Desejamos mostrar que vale a igualdade. Suponhamos então por absurdo que $Ax_0 - \lambda_0 x_0 \neq 0$. Seja $c \geq 0$, o valor da maior entrada do vetor $(Ax_0 - \lambda_0 x_0)$. Suponhamos que c ocorra na coordenada j. Neste caso,

$$A(Ax_0 - \lambda_0 x_0) \geq (a\mathbb{H})(ce_j) = ac\mathbb{I}.$$

Assim $A(Ax_0) \geq \lambda_0(Ax_0) + ac\mathbb{I}$. Seja h a maior entrada do vetor Ax_0. Então

$$A(Ax_0) \geq \lambda_0(Ax_0) + ac\mathbb{I} = \lambda_0(Ax_0) + \frac{ac}{h}h\mathbb{I}$$

$$\geq \lambda_0(Ax_0) + \frac{ac}{h}(Ax_0) = (\lambda_0 + \frac{ac}{h})(Ax_0).$$

Isso mostra que $(\lambda_0 + \frac{ac}{h}) \in I$ contrariando a Afirmação 1. A contradição foi um resultado da hipótese $Ax_0 - \lambda_0 x_0 \neq 0$, portanto devemos ter $Ax_0 = \lambda_0 x_0$. Sabemos que $\lambda_0 > 0$. Resta observarmos que $x_0 > 0$. Como $A > 0$ e $x_0 \geq 0$ é não nulo temos que $Ax_0 > 0$. Da igualdade $Ax_0 = \lambda_0 x_0$ concluímos que $x_0 > 0$. $\square$

Corolário 4.3. *Existe um autovalor positivo λ associado a um autovetor positivo h para L_F.*

Vamos chamar h de autovetor ou autofunção, uma vez que temos uma identificação entre vetores $h \in \mathbb{R}^n$ e funções $h : \{1, ..., n\} \to \mathbb{R}$.

Como $\lambda > 0$ e $h > 0$ (onde λ e h estão dados no corolário acima) podemos considerar a função $\bar{F}(i,j) = F(i,j) + \log(h(i)) - \log(h(j)) - \log(\lambda)$. Existe uma relação entre os autovalores e autovetores de L_F e $L_{\bar{F}}$. Observe que para $\psi \in \mathbb{R}^n$ e $\beta \in \mathbb{R}$ fixados,

$$\sum_i e^{\bar{F}(i,j)}\psi(i) = \beta\psi(j) \Longleftrightarrow \sum_i e^{F(i,j)+\log(h(i))-\log(h(j))-\log(\lambda)}\psi(i) = \beta\psi(j)$$

$$\Longleftrightarrow \sum_i e^{F(i,j)} \frac{h(i)}{\lambda h(j)}\psi(i) = \beta\psi(j) \Longleftrightarrow \frac{1}{\lambda h(j)} \sum_i e^{F(i,j)} h(i)\psi(i) = \beta\psi(j)$$

$$\Longleftrightarrow \sum_i e^{F(i,j)}(h(i)\psi(i)) = (\lambda\beta)(h(j)\psi(j)).$$

Concluímos que β é autovalor de $L_{\bar{F}}$ com autovetor ψ se e somente se $\lambda\beta$ é autovalor de L_F com autovetor $h\psi$, onde a i-ésima coordenada de $h\psi$ é dada por $h(i)\psi(i)$. Em particular o vetor constante $\psi = 1$ é um autovetor de $L_{\bar{F}}$ associado ao autovalor 1. Segue desta afirmação que $\sum_i e^{\bar{F}(i,j)} = 1$ para todo j.

Definição 4.4. *Uma função $F : X \to \mathbb{R}$ está normalizada se para todo $j \in \{1, ..., n\}$, $\sum_i e^{F(i,j)} = 1$.*

Como consequência desta definição, a função F está normalizada se a matriz associada a função e^F é coluna estocástica. Se F não está normalizada, a função $\bar{F}$ definida acima por $\bar{F}(i,j) = F(i,j) + \log(h(i)) - \log(h(j)) - \log(\lambda)$ será normalizada.

Se L_F admitir dois autovalores positivos associados a duas autofunções positivas, então teremos duas formas distintas de normalizarmos F. No entanto o lema abaixo garante que isso não ocorrerá.

Lema 4.5. *Existe um único autovalor positivo λ que pode ser associado a um autovetor positivo para L_F. O auto-espaço associado a λ tem dimensão 1.*

Demonstração: Suponhamos por absurdo existirem dois autovalores positivos λ, β associados a autofunções positivas ψ, φ respectivamente. Seja $\bar{F}(i,j) = F(i,j) + \log(\psi(i)) - \log(\psi(j)) - \log(\lambda)$. Como β é autovalor de L_F concluímos que $\alpha := \frac{\beta}{\lambda}$ é autovalor para $L_{\bar{F}}$ com autovetor $h := \frac{\varphi}{\psi}$ (ou seja $h(i) = \frac{\varphi(i)}{\psi(i)}$, $i = 1, 2, ..., n$). Como $\bar{F}$ está normalizada, denotando por $h(j)$ a maior coordenada de h, obtemos

$$\alpha h(j) = \sum_i e^{\bar{F}(i,j)} h(i) \leq \sum_i e^{\bar{F}(i,j)} h(j) = h(j) \sum_i e^{\bar{F}(i,j)} = h(j).$$

Portanto $\alpha \leq 1$. De forma análoga, denotando por $h(l)$ a menor coordenada de h, obtemos

$$\alpha h(l) = \sum_i e^{\bar{F}(i,l)} h(i) \geq \sum_i e^{\bar{F}(i,l)} h(l) = h(l) \sum_i e^{\bar{F}(i,l)} = h(l).$$

Portanto $\alpha \geq 1$. Concluímos que $\alpha = 1$, garantindo que $\beta = \lambda$. Além disso, se h não for um vetor constante teremos uma contradição porque neste caso, denotando por $h(j)$ a maior coordenada de h, teremos

$$h(j) = \alpha h(j) = \sum_i e^{\bar{F}(i,j)} h(i) < \sum_i e^{\bar{F}(i,j)} h(j) = h(j) \sum_i e^{\bar{F}(i,j)} = h(j).$$

Concluímos que h é um vetor constante, portanto ψ é um múltiplo de φ.

Vamos agora provar que o auto-espaço associado a λ tem dimensão 1. Seja x um autovetor qualquer (não necessariamente positivo) associado ao autovalor λ para o operador L_F. Então $b = \frac{x}{\psi}$ é autovetor associado ao autovalor 1 para $L_{\bar{F}}$. Para qualquer constante C temos,

$$C + b(j) = C + \sum_i e^{\bar{F}(i,j)} b(i)$$

$$= C \sum_i e^{\bar{F}(i,j)} + \sum_i e^{\bar{F}(i,j)} b(i) = \sum_i e^{\bar{F}(i,j)} (C + b(i)),$$

ou seja $C + b$ é um autovetor associado ao autovalor 1 para $L_{\bar{F}}$. Se C é suficientemente grande obtemos que $C + b$ é positivo. Repetindo o argumento anterior aplicado a h obtemos que $C + b$ é um vetor constante. Em particular b é um vetor constante e x é um múltiplo de ψ. Isso mostra que o auto-espaço associado a λ para o operador L_F tem dimensão 1. $\square$

Se λ e h são o autovalor e o autovetor positivos para L_F, então qualquer múltiplo ch do autovetor é também positivo, se $c > 0$. A função normalizada $\bar{F}$ não depende de c porque

$$\bar{F}(i,j) = F(i,j) + \log(ch(i)) - \log(ch(j)) - \log(\lambda)$$

$$= F(i,j) + \log(c) + \log(h(i)) - \log(c) - \log(h(j)) - \log(\lambda)$$
$$= F(i,j) + \log(h(i)) - \log(h(j)) - \log(\lambda).$$

Denotamos por $\Pi(X, \sigma)$ o conjunto das probabilidades invariantes sobre X. Se $\pi \in \Pi(X, \sigma)$, a notação $\langle F, \pi \rangle$ representa a média de F segundo π, dada por

$$\langle F, \pi \rangle = \sum_{1 \leq i,j \leq n} F(i,j)\pi_{ij}.$$

Definição 4.6. *A pressão de uma função $F : X \to \mathbb{R}$ é dada por*

$$P(F) = \sup_{\pi \in \Pi(X,\sigma)} \langle F, \pi \rangle + H(\pi).$$

As probabilidades invariantes que estamos considerando estão definidas sobre um conjunto finito. A definição de entropia neste texto não coincide com a entropia de Kolmogorov-Sinai e consequentemente a definição de pressão acima não coincide exatamente com a usada em Formalismo Termodinâmico [17]. No entanto como será mostrado abaixo, esta definição é equivalente.

Lema 4.7. *Dada $V : \{1, ..., n\} \to \mathbb{R}$, defina $W : X \to \mathbb{R}$ por $W(i,j) = V(i) - V(j)$. Então $\langle W, \pi \rangle = 0$ para todo $\pi \in \Pi(X, \sigma)$.*

Demonstração:

$$\langle W, \pi \rangle = \sum_{i,j} W(i,j)\pi_{ij} = \sum_{i,j} (V(i) - V(j))\pi_{ij} = \sum_{i,j} V(i)\pi_{ij} - \sum_{i,j} V(j)\pi_{ij}$$

$$= \sum_i V_i \sum_s \pi_{is} - \sum_j V(j) \sum_s \pi_{sj} = \sum_i V_i \sum_s \pi_{si} - \sum_j V(j) \sum_s \pi_{sj} = 0.$$

□

Corolário 4.8. *Se λ e h são o autovalor e o autovetor positivos para L_F e $\bar{F}(i,j) = F(i,j) + \log(h(i)) - \log(h(j)) - \log(\lambda)$ é a função normalizada associada, então $\langle \bar{F}, \pi \rangle = \langle F, \pi \rangle - \log(\lambda)$ para todo $\pi \in \Pi(X, \sigma)$. Em particular $P(\bar{F}) = P(F) - \log(\lambda)$.*

Como $e^{\bar{F}}$ é coluna estocástica e positiva, pelo Teorema 1.8 existe um vetor estacionário P associado e seguindo a discussão da seção 2, podemos construir uma probabilidade invariante $\pi_{ij} = e^{\bar{F}(i,j)} P_j$ (auto-probabilidade associada). Pelo teorema 2.12 temos que $J^\pi = e^{\bar{F}}$. Então $H(\pi) = -\langle \log(e^{\bar{F}}), \pi \rangle = -\langle \bar{F}, \pi \rangle$. Portanto

$$P(\bar{F}) = \sup_{\eta \in \Pi(X,\sigma)} \langle \bar{F}, \eta \rangle + H(\eta) \geq \langle \bar{F}, \pi \rangle + H(\pi) = 0.$$

Por outro lado, pelo Lema 3.19 para qualquer probabilidade invariante η temos que $H(\eta) \leq -\langle \bar{F}, \eta \rangle$. Segue que

$$P(\bar{F}) = \sup_{\eta \in \Pi(X,\sigma)} \langle \bar{F}, \eta \rangle + H(\eta) \leq 0.$$

Concluímos que $P(\bar{F}) = 0$ e o supremo que define a pressão é atingido pela auto-probabilidade π.

Teorema 4.9. $P(F) = \log(\lambda)$ *onde* λ *é o único autovalor positivo associado a uma autofunção positiva* h *para* L_F*. O supremo de* $\langle F, \pi \rangle + H(\pi)$ *é atingido apenas pela auto-probabilidade associada a* $e^{\bar{F}}$*, onde* $\bar{F}(i,j) = F(i,j) + \log(h(i)) - \log(h(j)) - \log(\lambda)$.

Demonstração: Como vimos acima, para qualquer probabilidade invariante η temos que

$$\langle \bar{F}, \eta \rangle = \langle F, \eta \rangle - \log(\lambda)$$

e $P(F) = P(\bar{F}) + \log(\lambda) = \log(\lambda)$. Se π é a auto-probabilidade associada a $e^{\bar{F}}$ então

$$\langle F, \pi \rangle + H(\pi) = \langle F, \pi \rangle - \langle \bar{F}, \pi \rangle = \log(\lambda).$$

Resta provarmos que π é a única probabilidade invariante atingindo o supremo. Seja η outra probabilidade invariante. Se η é positiva e seu jacobiano não coincide com $e^{\bar{F}}$, pelo Lema 3.19, $H(\eta) < -\langle \bar{F}, \eta \rangle$. Portanto

$$\langle \bar{F}, \eta \rangle + H(\eta) < 0$$

e η não atinge o supremo. Se η é positiva e seu jacobiano coincide com $e^{\bar{F}}$ então, pelo lema 2.11, $\eta = \pi$. Se η não é positiva então $\eta_{ij} = 0$ para algum par i, j. Podemos supor que algum elemento da coluna j de η é não nulo. De fato, se η for nula em toda a coluna j então pela invariância, η será nula em toda a linha j. Neste caso podemos escolher um outro elemento da linha j, digamos η_{jm} onde η não é identicamente nula na coluna m (pois η não pode ser identicamente nula em todas as colunas). Supondo então $\eta_{ij} = 0$ para algum par i, j e que η não é identicamente nula na coluna j, concluímos que $J^{\eta}_{ij} = 0$ enquanto $e^{\bar{F}(i,j)} > 0$. Como

$$\sum_s J^{\eta}_{sj} = 1 = \sum_s e^{\bar{F}(s,j)},$$

em algum elemento (l, j), onde $\eta_{lj} > 0$ teremos $J^{\eta}_{lj} \neq e^{\bar{F}(l,j)}$. Segue do lema 3.19 que $H(\eta) < -\langle \bar{F}, \eta \rangle$. Portanto $\langle \bar{F}, \eta \rangle + H(\eta) < 0$ e $\langle F, \eta \rangle + H(\eta) < \log(\lambda)$. Concluímos que π é a única probabilidade que atinge o supremo. □

Seguindo a terminologia adotada no formalismo termodinâmico, a auto-probabilidade associada a $e^{\bar{F}}$ pode ser chamada de **medida de equilíbrio** associada a função F, uma vez que $P(F) = \langle F, \pi\rangle + H(\pi)$. Note que neste texto uma probabilidade invariante π será o equilíbrio de alguma função se e somente se π é positiva (teorema 2.12). Neste sentido poderíamos dizer também que uma probabilidade invariante π é uma medida de Gibbs se e somente se π é positiva.

Observação 4.10. *Para $X = \{1,2\} \times \{1,2\}$ temos que toda probabilidade invariante π satisfaz $\pi_{12} = \pi_{21}$ pois $\pi_{11} + \pi_{12} = \pi([1,\cdot]) = \pi([\cdot,1]) = \pi_{11} + \pi_{21}$. Assim, toda probabilidade invariante é da forma*

$$\pi = \begin{pmatrix} a & b \\ b & c \end{pmatrix}, \quad a + 2b + c = 1, \quad a, b, c \geq 0.$$

Em particular toda probabilidade invariante pode ser escrita como combinação convexa (média ponderada) das probabilidades

$$\eta^1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \ \eta^2 = \begin{pmatrix} 0 & 1/2 \\ 1/2 & 0 \end{pmatrix} \ e \ \eta^3 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix},$$

ou seja

$$\begin{pmatrix} a & b \\ b & c \end{pmatrix} = a \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + 2b \begin{pmatrix} 0 & 1/2 \\ 1/2 & 0 \end{pmatrix} + c \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad a + 2b + c = 1, \quad a, b, c \geq 0.$$

Identificando as probabilidades em X com um conjunto do $\mathbb{R}^4$, obtemos que as probabilidades invariantes formam um conjunto convexo com vértices η^1, η^2, η^3. Estas 3 probabilidades são as únicas extremais (vértices), mas não são medidas de equilíbrio. Dada uma função F estas probabilidades não são realizadoras do supremo que define $P(F)$, pois não são positivas. Convém observar que a entropia que definimos neste texto não corresponde na íntegra a entropia de Kolmogorov-Sinai que aparece com frequência em textos de Teoria Ergódica e Formalismo Termodinâmico. No corolário 3.10 foi provado que a entropia é côncava (ver também o exemplo que o seguiu). Neste sentido a função $\pi \to \langle F, \pi\rangle + H(\pi)$ é uma função côncava e o supremo não precisa ser (e de fato não será) atingido em uma medida extremal.

Se, no entanto, $\Psi : \Pi(X, \sigma) \to \mathbb{R}$ é uma função convexa, ou seja $\Psi(\lambda\pi + (1-\lambda)\eta) \leq \lambda\Psi(\pi) + (1-\lambda)\Psi(\eta)$ para quaisquer π, η invariantes e $\lambda \in [0,1]$, então o supremo de Ψ é atingido por uma das medidas extremais. De fato, para qualquer probabilidade invariante

$$\pi = \begin{pmatrix} a & b \\ b & c \end{pmatrix},$$

temos que

$$\Psi(\pi) = \Psi(a\eta^1 + 2b\eta^2 + c\eta^3) \leq a\Psi(\eta^1) + 2b\Psi(\eta^2) + c\Psi(\eta^3)$$

e como $a + 2b + c = 1$*, pelo menos um dos números* $\Psi(\eta^1), \Psi(\eta^2)$*ou* $\Psi(\eta^3)$ *terá que ser maior ou igual a* $\Psi(\pi)$*. Assim*

$$\Psi(\pi) \leq \max\{\Psi(\eta^1),\ \Psi(\eta^2)\ \Psi(\eta^3)\}.$$

Como isso é verificado para qualquer $\pi \in \Pi(X, \sigma)$ *obtemos*

$$\max\{\Psi(\eta^1),\ \Psi(\eta^2)\ \Psi(\eta^3)\} = \sup\{\Psi(\pi) | \pi \in \Pi(X, \sigma)\}.$$

Proposição 4.11. *A pressão é uma função convexa.*

Demonstração: Dados um número $\lambda \in [0, 1]$ e funções F e G sobre X. Seja $B = \lambda F + (1 - \lambda)G$ e π_0 a medida de equilíbrio associada a função B. Então

$$P(\lambda F + (1 - \lambda)G) = P(B) = \langle B, \pi_0 \rangle + H(\pi_0) = \lambda \langle F, \pi_0 \rangle + (1 - \lambda)\langle G, \pi_0 \rangle + H(\pi_0)$$

$$= \lambda[\langle F, \pi_0 \rangle + H(\pi_0)] + (1 - \lambda)[\langle G, \pi_0 \rangle + H(\pi_0)]$$

$$\leq \lambda \left[\sup_{\pi \in \Pi(X,\sigma)} \langle F, \pi \rangle + H(\pi) \right] + (1 - \lambda) \left[\sup_{\pi \in \Pi(X,\sigma)} \langle G, \pi \rangle + H(\pi) \right]$$

$$= \lambda P(F) + (1 - \lambda) P(G).$$

□

Em [17] o leitor encontrará uma boa discussão sobre o formalismo termodinâmico e as relações entre pressão e equilíbrio com as informações obtidas a partir das órbitas periódicas de um sistema dinâmico. Isso inclui a discussão sobre as funções zeta em sistemas dinâmicos, que possuem semelhanças com a função zeta de Riemann, sendo que no teorema 6.9 em [17], por exemplo, é apresentado o "Teorema das Órbitas Primas".

Capítulo 5

Otimização sobre probabilidades invariantes

Nesta seção continuamos escrevendo $X = \{1, ..., n\} \times \{1, ..., n\}$. Dada uma função $F : X \to \mathbb{R}$, poderíamos tentar entender qual o valor máximo de $\langle F, \pi \rangle$ entre todas as probabilidades π sobre X. No entanto, se F atinge seu maior valor em (i, j) então o máximo de $\langle F, \pi \rangle$ será atingido pela probabilidade que concentra peso total no ponto (i, j) e o valor máximo será $F(i, j)$. Podemos considerar o problema de maximizar $\langle F, \pi \rangle$ para probabilidades π satisfazendo algumas restrições. Nesta seção vamos considerar como restrição a invariância de π. A discussão que faremos está relacionada à Otimização Ergódica, sendo [2] uma boa referência para este tema.

Definição 5.1. *Dada uma função $F : X \to \mathbb{R}$, denotamos por*

$$M(F) = \sup_{\pi \in \Pi(X,\sigma)} \langle F, \pi \rangle.$$

Exemplo 5.2. *Seguindo a discussão da Observação 4.10, se $X = \{1, 2\} \times \{1, 2\}$ então as probabilidades*

$$\eta^1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \ \eta^2 = \begin{pmatrix} 0 & 1/2 \\ 1/2 & 0 \end{pmatrix} \text{ e } \eta^3 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

são as probabilidades extremais de $\Pi(X, \sigma)$. Afirmamos que o supremo na definição de $M(F)$ é atingido em pelo menos uma das probabilidades η^1, η^2 ou η^3. De fato, se $\pi = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$ é uma probabilidade invariante qualquer, então como $a, b, c \geq 0$ e $a + 2b + c = 1$,

$$\begin{aligned} \langle F, \pi \rangle = a\langle F, \eta^1 \rangle + 2b\langle F, \eta^2 \rangle + c\langle F, \eta^3 \rangle &\leq (a+2b+c) \max\{\langle F, \eta^1 \rangle, \langle F, \eta^2 \rangle, \langle F, \eta^3 \rangle\} \\ &= \max\{\langle F, \eta^1 \rangle, \langle F, \eta^2 \rangle, \langle F, \eta^3 \rangle\}. \end{aligned}$$

Portanto

$$M(F) = \max\{\langle F, \eta^1 \rangle, \langle F, \eta^2 \rangle, \langle F, \eta^3 \rangle\}.$$

Esse exemplo contém uma ideia de abordagem comum neste tipo de problema. O conjunto $\Pi(X,\sigma)$ é convexo e para F fixada a aplicação $\pi \to \langle F,\pi\rangle$ é convexa (de fato $\langle F, \lambda\pi + (1-\lambda)\eta\rangle = \lambda\langle F,\pi\rangle + (1-\lambda)\langle F,\eta\rangle$, $\lambda \in [0,1]$). Seguindo a discussão da Observação 4.10, o supremo

$$M(F) = \sup_{\pi\in\Pi(X,\sigma)} \langle F,\pi\rangle$$

será atingido em pelo menos uma probabilidade extremal do conjunto $\Pi(X,\sigma)$. Neste sentido, uma melhor compreensão das probabilidades extremais de $\Pi(X,\sigma)$ nos ajuda na compreensão de $M(F)$.

Outra abordagem comum consiste no estudo do "problema dual" associado. Dados um número real m e uma função $V : \{1,...,n\} \to \mathbb{R}$ satisfazendo

$$m \geq F(i,j) + V(i) - V(j)\ \forall (i,j) \in X,$$

pelo Lema 4.7, denotando por $W(i,j) = V(i) - V(j)$ temos que

$$M(F) = \sup_{\pi\in\Pi(X,\sigma)} \langle F,\pi\rangle = \sup_{\pi\in\Pi(X,\sigma)} \langle F+W,\pi\rangle \leq m.$$

Portanto, para cada função V fixada, obtemos que $M(F)$ é menor ou igual a qualquer possível m satisfazendo a desigualdade

$$m \geq F(i,j) + V(i) - V(j)\ \forall (i,j) \in X.$$

Em particular, podemos considerar

$$m = \max_{i,j}(F(i,j) + V(i) - V(j)).$$

Assim, para V fixada,

$$M(F) \leq \max_{i,j}(F(i,j) + V(i) - V(j)).$$

Esta desigualdade é satisfeita para qualquer escolha de função $V : \{1,...,n\} \to \mathbb{R}$. Portanto

$$M(F) \leq \inf_V \max_{i,j}(F(i,j) + V(i) - V(j)). \tag{5.1}$$

Veremos abaixo que de fato a igualdade é verificada.

Dados pontos $l_1, l_2, ..., l_k \in \{1,...,n\}$, considere os pontos $x_1 = (l_1,l_k)$, $x_2 = (l_2,l_1), ..., x_{k-1} = (l_{k-1},l_{k-2})$, $x_k = (l_k,l_{k-1})$ em X e defina uma probabilidade π^k sobre X por

$$\pi^k_{ij} = \frac{\text{número de pontos } x_s = (i,j),\ s \in \{1,...,k\}}{k}.$$

Note que π^k é de fato uma probabilidade pois $\pi^k_{ij} \geq 0$ e

$$\sum_{i,j} \frac{\text{número de pontos}\, x_s = (i,j),\ s \in \{1,...,k\}}{k} = \frac{\text{número de pontos}\, x_s,\ s \in \{1,...,k\}}{k} = 1.$$

Afirmamos que π^k é invariante. De fato:

$$\sum_a \pi^k_{aj} = \frac{\text{número de pontos}\, x_s = (\cdot, j),\ s \in \{1,...,k\}}{k}$$

$$= \frac{\text{número de pontos}\, l_s = j,\ s \in \{1,...,k\}}{k}$$

$$= \frac{\text{número de pontos}\, x_s = (j,\cdot),\ s \in \{1,...,k\}}{k} = \sum_a \pi^k_{ja}.$$

Observamos também que para qualquer função $F : X \to \mathbb{R}$ temos que

$$\langle F, \pi^k \rangle = \sum_{i,j} F(i,j) \frac{\text{número de pontos}\, x_s = (i,j),\ s \in \{1,...,k\}}{k}$$

$$= \sum_{x_s} \frac{F(x_s)}{k} = \frac{F(l_1,l_k) + F(l_2,l_1) + ... + F(l_k,l_{k-1})}{k}.$$

Lema 5.3. *Para qualquer número k e quaisquer pontos $l_0, l_1 ..., l_k \in \{1, ..., n\}$ temos que*

$$F(l_k,l_{k-1}) + ... + F(l_2,l_1) + F(l_1,l_0) - kM(F) \leq 2 \max_{i,j} F(i,j).$$

Demonstração: Fixados os pontos $l_0, l_1, ..., l_k$, desconsiderando o ponto l_0 construa com os demais pontos a probabilidade invariante π^k, como discutido acima. Então $M(F) \geq \langle F, \pi^k \rangle = \frac{F(l_1,l_k)+F(l_2,l_1)+...+F(l_k,l_{k-1})}{k}$. Assim

$$F(l_k,l_{k-1}) + ... + F(l_2,l_1) + F(l_1,l_0) - kM(F)$$

$$\leq F(l_k,l_{k-1}) + ... + F(l_2,l_1) + F(l_1,l_0) - k\frac{F(l_1,l_k) + F(l_2,l_1) + ... + F(l_k,l_{k-1})}{k}$$

$$= F(l_k,l_{k-1}) + ... + F(l_2,l_1) + F(l_1,l_0) - [F(l_1,l_k) + F(l_2,l_1) + ... + F(l_k,l_{k-1})]$$

$$= F(l_1,l_0) - F(l_1,l_k) \leq 2 \max_{i,j} F(i,j).$$

□

Proposição 5.4 (Dualidade)**.**

$$\sup_{\pi \in \Pi(X,\sigma)} \langle F, \pi \rangle = M(F) = \inf_V \max_{i,j} (F(i,j) + V(i) - V(j)).$$

Demonstração: A primeira igualdade segue por definição. Por (5.1) sabemos que

$$M(F) \leq \inf_{V} \max_{i,j}(F(i,j) + V(i) - V(j)),$$

portanto basta apresentarmos uma função $V : \{1, ..., n\} \to \mathbb{R}$ satisfazendo

$$M(F) \geq \max_{i,j}(F(i,j) + V(i) - V(j)).$$

Para cada j fixado defina

$$V(j) = \limsup_{k \to +\infty} \sup_{l_k, \ldots, l_1} [F(l_k, l_{k-1}) + ... + F(l_2, l_1) + F(l_1, j) - kM(F)].$$

Pelo lema acima $V(j)$ está bem definida e $V(j) \leq 2 \max_{i,j} F(i,j)$.

Para cada i, j temos:

$$F(i,j) + V(i) - V(j) - M(F) = F(i,j) - M(F) + V(i) - V(j)$$

$$= F(i,j) - M(F) + \limsup_{k \to +\infty} \sup_{l_k, \ldots, l_1} [F(l_k, l_{k-1}) + ... + F(l_2, l_1) + F(l_1, i) - kM(F)] - V(j)$$

$$= \limsup_{k \to +\infty} \sup_{l_k, \ldots, l_1} [F(l_k, l_{k-1}) + ... + F(l_2, l_1) + F(l_1, i) + F(i,j) - (k+1)M(F)] - V(j)$$

$$\overset{t_{s+1} = l_s}{=} \limsup_{k \to +\infty} \sup_{t_{k+1}, \ldots, t_2} [F(t_{k+1}, t_k) + ... + F(t_2, i) + F(i,j) - (k+1)M(F)] - V(j)$$

$$\leq \limsup_{k \to +\infty} \sup_{t_{k+1}, \ldots, t_1} [F(t_{k+1}, t_k) + ... + F(t_2, t_1) + F(t_1, j) - (k+1)M(F)] - V(j) = 0.$$

Portanto

$$\max_{i,j} F(i,j) + V(i) - V(j) - M(F) \leq 0,$$

como desejado. □

Uma função V satisfazendo

$$M(F) = \max_{i,j}(F(i,j) + V(i) - V(j)) \tag{5.2}$$

será chamada de **sub-ação**. Por (5.1) obtemos que

$$M(F) \leq \max_{i,j}(F(i,j) + V(i) - V(j))$$

é verificada por qualquer função V. Portanto V é uma sub-ação se e somente se

$$F(i,j) + V(i) - V(j) \leq M(F) \ \forall i, j.$$

Exemplo 5.5. *Vejamos como a dualidade acima pode ser útil nas estimativas de $M(F)$. Considere como exemplo a função*

$$F = \begin{pmatrix} F(1,1) & F(1,2) & F(1,3) \\ F(2,1) & F(2,2) & F(2,3) \\ F(3,1) & F(3,2) & F(3,3) \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Por um lado, como as probabilidades

$$\pi_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \ e \ \pi_2 = \begin{pmatrix} 0 & 1/2 & 0 \\ 1/2 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

são invariantes, temos $M(F) \geq \langle F, \pi_1 \rangle = 1$ e $M(F) \geq \langle F, \pi_2 \rangle = 1,5$. Note que qualquer probabilidade invariante nos fornece uma cota inferior para $M(F)$. Ou seja, já sabemos que $M(F) \geq 1,5$. Por outro lado, dada por exemplo a função V satisfazendo $V(1) = 0,\ V(2) = 0$ e $V(3) = 1$ temos que

$$M(F) \leq \max_{i,j}[F(i,j) + V(i) - V(j)] = \max\{1,1,0,2,1,0,1,1,1\} = 2.$$

Note que qualquer função V nos fornece uma cota superior para $M(F)$. Das contas acima obtemos que $1,5 \leq M(F) \leq 2$. Podemos buscar métodos eficientes para fazer a estimativa de $M(F)$ seguindo esta ideia. Neste exemplo observamos que para V satisfazendo $V(1) = \frac{1}{2}, V(2) = 0, V(3) = 1$ temos

$$M(F) \leq \max_{i,j}[F(i,j) + V(i) - V(j)] = \max\{1, \frac{3}{2}, \frac{1}{2}, \frac{3}{2}, 1, 0, \frac{1}{2}, 1, 1\} = 1,5.$$

Portanto $1,5 \leq M(F) \leq 1,5$, ou seja $M(F) = 1,5$.

A discussão que faremos agora busca relacionar os conceitos apresentados na seção anterior com os desta seção, através do chamado caso de temperatura zero em Formalismo Termodinâmico.

Fixado $\beta > 0$, considere a função[1] βF. Seguindo a discussão da seção anterior podemos obter uma autofunção h_β positiva e um autovalor λ_β positivo para o operador $L_{\beta F}$. A função $\bar{F}_\beta(i,j) = \beta F(i,j) + \log(h_\beta(i)) - \log(h_\beta(j)) - \log(\lambda_\beta)$ está normalizada, ou seja

$$\sum_i e^{\beta F(i,j) + \log(h_\beta(i)) - \log(h_\beta(j)) - \log(\lambda_\beta)} = 1.$$

Lema 5.6.

$$\lim_{\beta \to +\infty} \frac{1}{\beta} \log(\lambda_\beta) = \lim_{\beta \to +\infty} \frac{P(\beta F)}{\beta} = M(F).$$

[1] O parâmetro β em formalismo termodinâmico costuma ser chamado de inverso da temperatura, $\beta = \frac{1}{T}$. Desta forma, quando $\beta \to +\infty$ temos que $T \to 0$.

Demonstração:

$$\lim_{\beta\to+\infty}\frac{P(\beta F)}{\beta}=\lim_{\beta\to+\infty}\frac{\sup_{\pi\in\Pi(X,\sigma)}\beta\langle F\,\pi\rangle+H(\pi)}{\beta}=\lim_{\beta\to+\infty}\sup_{\pi\in\Pi(X,\sigma)}\langle F\,\pi\rangle+\frac{H(\pi)}{\beta}.$$

Como $0\le H(\pi)\le\log(n)$ para qualquer probabilidade invariante, obtemos que

$$\lim_{\beta\to+\infty}\sup_{\pi\in\Pi(X,\sigma)}\langle F\,\pi\rangle+\frac{H(\pi)}{\beta}\ge\lim_{\beta\to+\infty}\sup_{\pi\in\Pi(X,\sigma)}\langle F\,\pi\rangle=\sup_{\pi\in\Pi(X,\sigma)}\langle F,\pi\rangle$$

e

$$\lim_{\beta\to+\infty}\sup_{\pi\in\Pi(X,\sigma)}\langle F\,\pi\rangle+\frac{H(\pi)}{\beta}\le\lim_{\beta\to+\infty}\sup_{\pi\in\Pi(X,\sigma)}\langle F\,\pi\rangle+\frac{\log(n)}{\beta}=\sup_{\pi\in\Pi(X,\sigma)}\langle F,\pi\rangle.$$

□

Lema 5.7. *Para cada $\beta>0$ seja π_β a medida de equilíbrio de βF. Suponha que π_β convirja a uma probabilidade π_∞, quando $\beta\to+\infty$. Então $M(F)=\langle F,\pi_\infty\rangle$. O mesmo vale se π_∞ for um ponto de acumulação de π_β, $\beta\to+\infty$.*

Demonstração: Seja π uma probabilidade invariante qualquer. Então

$$P(\beta F)=\beta\langle F,\pi_\beta\rangle+H(\pi_\beta)\ge\beta\langle F,\pi\rangle+H(\pi).$$

Como consequência

$$\langle F,\pi_\beta\rangle+\frac{H(\pi_\beta)}{\beta}\ge\langle F,\pi\rangle+\frac{H(\pi)}{\beta}.$$

Quando $\beta\to+\infty$, obtemos

$$\langle F,\pi_\infty\rangle\ge\langle F,\pi\rangle.$$

Como π é uma probabilidade invariante qualquer concluímos que $\langle F,\pi_\infty\rangle=M(F)$.
□

Lema 5.8. *Para cada $\beta>0$ seja h_β a autofunção positiva associada ao autovalor positivo λ_β de $L_{\beta F}$. Suponha que $\frac{1}{\beta}\log(h_\beta)$ convirja a uma função V, quando $\beta\to+\infty$. Então V é uma sub-ação de F. O mesmo vale se V for um ponto de acumulação de $\frac{1}{\beta}\log(h_\beta)$, $\beta\to+\infty$.*

Demonstração: A função $\bar{F}_\beta(i,j)=\beta F(i,j)+\log(h_\beta(i))-\log(h_\beta(j))-\log(\lambda_\beta)$ está normalizada, ou seja , para cada j,

$$\sum_i e^{\beta F(i,j)+\log(h_\beta(i))-\log(h_\beta(j))-\log(\lambda_\beta)}=1.$$

Assim,

$$\frac{1}{\beta}\log\left(\sum_i e^{\beta F(i,j)+\log(h_\beta(i))-\log(h_\beta(j))-\log(\lambda_\beta)}\right)=0.$$

Fixado j, para qualquer i_1 também fixado

$$\begin{aligned}
0 &= \lim_{\beta\to+\infty}\frac{1}{\beta}\log\left(\sum_i e^{\beta F(i,j)+\log(h_\beta(i))-\log(h_\beta(j))-\log(\lambda_\beta)}\right)\\
&\geq \lim_{\beta\to+\infty}\frac{1}{\beta}\log\left(e^{\beta F(i_1,j)+\log(h_\beta(i_1))-\log(h_\beta(j))-\log(\lambda_\beta)}\right)\\
&= \lim_{\beta\to+\infty}\left(F(i_1,j)+\frac{1}{\beta}\log(h_\beta(i_1))-\frac{1}{\beta}\log(h_\beta(j))-\frac{1}{\beta}\log(\lambda_\beta)\right)\\
&= F(i_1,j)+V(i_1)-V(j)-M(F).
\end{aligned}$$

Concluímos que para qualquer j e i_1

$$F(i_1,j)+V(i_1)-V(j)-M(F)\leq 0,$$

portanto

$$\sup_{i,j} F(i,j)+V(i)-V(j)\leq M(F).$$

Como a desigualdade oposta é sempre verificada, concluímos que vale a igualdade.
□

Capítulo 6

Outro tópico: Transporte Ótimo

Este capítulo difere um pouco dos anteriores. Não iremos trabalhar aqui com probabilidades invariantes, por exemplo. No entanto, como o leitor irá perceber, algumas das ideias empregadas aqui são semelhantes às discutidas anteriormente. Em [19] o leitor encontrará uma boa exposição do assunto que será brevemente introduzido neste capítulo. Relações entre o que discutiremos aqui e o que foi discutido anteriormente podem ser encontradas por exemplo em [9], [10] e [14].

Exemplo 6.1. *Suponha que uma empresa tenha 2 filiais x_1 e x_2 em uma região denotada por X e 2 filiais y_1 e y_2 em uma região denotada por Y. Devemos transportar 1000 toneladas de um determinado produto de X para Y, sendo que 300 toneladas partirão de x_1 e 700 toneladas partirão de x_2, 400 toneladas deverão ser recebidas por y_1 e 600 toneladas deverão ser recebidas por y_2. Vamos chamar de plano de transporte uma lista de quatro números não negativos*

$$\pi = \begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{pmatrix}$$

onde π_{ij} representa a quantidade de toneladas transportadas de x_i para y_j. Devemos ter compatibilidade com os dados fornecidos. Por exemplo, $\pi_{11}+\pi_{12}$ (quantidade de toneladas transportadas de x_1 para y_1 somada a quantidade de toneladas transportadas de x_1 para y_2) deve resultar em 300 (quantidade de toneladas partindo de x_1). Portanto devemos ter

$$\pi_{11} + \pi_{12} = 300,\ \pi_{21} + \pi_{22} = 700,\ \pi_{11} + \pi_{21} = 400,\ \ \pi_{12} + \pi_{22} = 600.$$

Existem vários possíveis planos de transporte, como por exemplo

$$\begin{pmatrix} 300 & 0 \\ 100 & 600 \end{pmatrix},\ \begin{pmatrix} 200 & 100 \\ 200 & 500 \end{pmatrix}\ e\ \begin{pmatrix} 150 & 150 \\ 250 & 450 \end{pmatrix}.$$

Note que, em qualquer matriz acima, a soma das entradas nas linhas resultam em 300 e 700 (respectivamente), enquanto a soma das entradas nas colunas resultam em 400 e 600 (respectivamente).

Agora, suponha ser fornecido o custo para transportarmos cada tonelada do produto de uma filial x_i para uma filial y_j, denotado por C_{ij}. Então para cada plano de transporte $\pi = (\pi_{ij})$, o custo total de transporte associado será dado por

$$\langle C, \pi \rangle = C_{11}\pi_{11} + C_{12}\pi_{12} + C_{21}\pi_{21} + C_{22}\pi_{22}.$$

Considere o seguinte problema: Fixado o custo (C_{ij}), determinar o plano de transporte π que resulte no menor custo total possível.

O problema dado acima serve de inspiração para o que iremos desenvolver neste capítulo. Sejam $X = \{1, 2, ..., n\}$ e $Y = \{1, 2, ..., m\}$. Dadas probabilidades $p = (p_1, ..., p_n)$ sobre X e $q = (q_1, ..., q_m)$ sobre Y, um plano de transporte de marginais p e q é uma probabilidade $\pi = (\pi_{ij})_{i \in X, j \in Y}$ sobre $X \times Y$ satisfazendo:

$$\sum_{j=1}^{m} \pi_{ij} = p_i \ \ e \ \ \sum_{i=1}^{n} \pi_{ij} = q_j.$$

"Se escrevermos π como uma matriz do tipo $n \times m$, então a soma das entradas da linha i de π deve resultar em p_i e a soma das entradas da coluna j deve resultar em q_j."

Dizemos que p é a $X-$marginal de π e q é a $Y-$marginal de π. Denotamos por $\Pi(p, q)$ o conjunto dos planos de transporte de marginais p e q.

Exemplo 6.2. *Com os dados do exemplo anterior podemos supor que $X = \{1, 2\}$, $Y = \{1, 2\}$, $p = (\frac{3}{10}, \frac{7}{10})$ e $q = (\frac{4}{10}, \frac{6}{10})$.*

Exemplo 6.3. *Para $X = \{1, 2, 3\}$, $Y = \{1, 2\}$, $p = (\frac{1}{2}, \frac{1}{3}, \frac{1}{6})$ e $q = (\frac{1}{2}, \frac{1}{2})$ temos que*

$$\pi = \begin{pmatrix} 1/4 & 1/4 \\ 1/6 & 1/6 \\ 1/12 & 1/12 \end{pmatrix}$$

é um plano de transporte com marginais p e q pois

$$\begin{array}{cccc} 1/4 & 1/4 & \rightarrow & 1/2 \\ 1/6 & 1/6 & \rightarrow & 1/3 \\ 1/12 & 1/12 & \rightarrow & 1/6 \\ \downarrow & \downarrow & & \\ 1/2 & 1/2 & & \end{array}.$$

Outro plano de transporte é dado por

$$\begin{array}{cccc} 8/24 & 4/24 & \rightarrow & 1/2 \\ 3/24 & 5/24 & \rightarrow & 1/3 \\ 1/24 & 3/24 & \rightarrow & 1/6 \\ \downarrow & \downarrow & & \\ 1/2 & 1/2 & & \end{array}.$$

Podemos considerar o problema de minimização a partir de uma função custo.

Definição 6.4. *Dada uma função (custo) $C : X \times Y \to \mathbb{R}$, definimos*

$$I(C) = \inf_{\pi \in \Pi(p,q)} \langle C, \pi \rangle.$$

Dizemos que $\pi_0 \in \Pi(p,q)$ é um plano de transporte ótimo se $I(C) = \langle C, \pi_0 \rangle$.

Poderíamos supor que o custo é uma função não negativa. Isso não alteraria significativamente o problema porque, somando-se uma constante α ao custo C, obtemos que para toda probabilidade π

$$\langle C + \alpha, \pi \rangle = \langle C, \pi \rangle + \langle \alpha, \pi \rangle = \langle C, \pi \rangle + \alpha.$$

Note que a definição acima é parecida com a utilizada no capítulo anterior. Como primeira diferença, aqui estamos interessados em minimizar $\langle C, \pi \rangle$. Isso pode ser facilmente contornado se considerarmos que $C = -D$ para alguma função D. Neste caso minimizar $\langle C, \pi \rangle$ e maximizar $\langle D, \pi \rangle$ são problemas análogos. Uma segunda diferença está no conjunto de probabilidades considerado. No capítulo anterior considerávamos probabilidades invariantes enquanto neste momento estamos considerando planos de marginais p e q. Anteriormente vimos que as probabilidades invariantes formavam um conjunto convexo e que pelo menos uma probabilidade extremal deste convexo atingia o supremo.

Lema 6.5. *O conjunto $\Pi(p,q)$ é convexo.*

Demonstração: Dados planos de transporte π e η em $\Pi(p,q)$ e um número $\lambda \in [0,1]$, seja $\gamma = \lambda\pi + (1-\lambda)\eta$. Como π e η são probabilidades sobre $X \times Y$ temos que γ é também uma probabilidade. Além disso

$$\sum_{j=1}^{m} \gamma_{ij} = \sum_{j=1}^{m} [\lambda\pi_{ij} + (1-\lambda)\eta_{ij}] = \lambda \sum_{j=1}^{m} \pi_{ij} + (1-\lambda) \sum_{j=1}^{m} \eta_{ij} = \lambda p_i + (1-\lambda) p_i = p_i.$$

De forma análoga obtemos que $\sum_{i=1}^{n} \gamma_{ij} = q_j$. Portanto γ tem marginais p e q. $\square$

Para o custo C fixado, a aplicação $\pi \to \langle C, \pi \rangle$ satisfaz

$$\langle C, \lambda\pi + (1-\lambda)\eta \rangle = \lambda \langle C, \pi \rangle + (1-\lambda) \langle C, \eta \rangle, \lambda \in [0,1],$$

garantindo que o ínfimo em $I(C)$ será atingido por uma probabilidade extremal de $\Pi(p,q)$.

Exemplo 6.6. *Se $X = \{1,2\}$, $Y = \{1,2\}$, $p = (\frac{1}{2}, \frac{1}{2})$ e $q = (\frac{1}{2}, \frac{1}{2})$, qualquer plano de transporte é dado por*

$$\begin{pmatrix} a & 1/2 - a \\ 1/2 - a & a \end{pmatrix}, \; a \in \left[0, \frac{1}{2}\right].$$

Os planos extremais são

$$\eta^1 = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix} \ \ e \ \ \eta^2 = \begin{pmatrix} 0 & 1/2 \\ 1/2 & 0 \end{pmatrix}.$$

Se $\pi = \begin{pmatrix} a & 1/2-a \\ 1/2-a & a \end{pmatrix}$ *é um plano de transporte, então* $\pi = (2a)\eta^1 + (1-2a)\eta^2$. *Usando que* $a \in [0, 1/2]$, *para um custo* C *fixado, temos:*

$$\langle C, \pi\rangle = 2a\langle C, \eta^1\rangle + (1-2a)\langle C, \eta^2\rangle$$

$$\geq 2a \min\{\langle C, \eta^1\rangle,\ \langle C, \eta^2\rangle\} + (1-2a)\min\{\langle C, \eta^1\rangle,\ \langle C, \eta^2\rangle\}$$
$$= \min\{\langle C, \eta^1\rangle,\ \langle C, \eta^2\rangle\}.$$

Como π *é um plano de transporte qualquer obtemos*

$$I(C) \geq \min\{\langle C, \eta^1\rangle,\ \langle C, \eta^2\rangle\}.$$

Como a desigualdade oposta é trivial, concluímos que

$$I(C) = \min\{\langle C, \eta^1\rangle,\ \langle C, \eta^2\rangle\},$$

ou seja, o ínfimo em $I(C)$ *será atingido por* η^1 *ou* η^2.

Lema 6.7. *Dadas funções* $A : X \to \mathbb{R}$, $B : Y \to \mathbb{R}$ *e uma probabilidade* $\pi \in \Pi(p, q)$ *temos*

$$\langle A, \pi\rangle = \langle A, p\rangle, \ \ e \ \langle B, \pi\rangle = \langle B, q\rangle.$$

Demonstração:

$$\langle A, \pi\rangle = \sum_{i,j} A(i)\pi_{ij} = \sum_i \left(A(i) \sum_j \pi_{ij} \right) = \sum_i A(i) p_i = \langle A, p\rangle$$

e

$$\langle B, \pi\rangle = \sum_{i,j} B(j)\pi_{ij} = \sum_j \left(B(j) \sum_i \pi_{ij} \right) = \sum_j B(j) q_j = \langle B, q\rangle.$$

□

Assim como na seção anterior, podemos buscar entender o problema dual associado para o cálculo de $I(C)$. Dadas duas funções $A : X \to \mathbb{R}$ e $B : Y \to \mathbb{R}$ satisfazendo $C(i, j) \geq A(i) + B(j)\, \forall i \in X,\, \forall j \in Y$ temos:

$$I(C) = \inf_{\pi \in \Pi(p,q)} \langle C, \pi\rangle \geq \inf_{\pi \in \Pi(p,q)} \langle A + B, \pi\rangle$$

$$= \inf_{\pi \in \Pi(p,q)} [\langle A, \pi \rangle + \langle B, \pi \rangle] = \inf_{\pi \in \Pi(p,q)} [\langle A, p \rangle + \langle B, q \rangle] = \langle A, p \rangle + \langle B, q \rangle.$$

Essa desigualdade é verificada por qualquer par de funções A, B satisfazendo $C(i,j) \geq A(i) + B(j)$, portanto

$$I(C) \geq \sup_{C(i,j) \geq A(i)+B(j)} \{\langle A, p \rangle + \langle B, q \rangle\}. \tag{6.1}$$

Uma interpretação para a desigualdade acima é dada no exemplo abaixo.

Exemplo 6.8. *Retomamos o problema do início deste capítulo. Suponha que uma empresa tenha filiais x_1 e x_2 em X e filiais y_1 e y_2 em Y. Devemos transportar 1000 toneladas de um determinado produto de X para Y, sendo que 300 toneladas partirão de x_1 e 700 toneladas partirão de x_2, 400 toneladas deverão ser recebidas por y_1 e 600 toneladas deverão ser recebidas por y_2. Com estes dados podemos supor que $X = \{1, 2\}$, $Y = \{1, 2\}$, $p = (\frac{3}{10}, \frac{7}{10})$ e $q = (\frac{4}{10}, \frac{6}{10})$. Suponha termos duas opções de escolha:*
1) usar uma função custo $C(i,j)$ proposta, que representa o custo de transportarmos cada tonelada de x_i para y_j.
2) usar um par de funções custos $A(i)$ e $B(j)$ propostos, onde $A(i)$ é o valor pago para cada tonelada retirada de x_i independente de seu destino ser y_1 ou y_2 e $B(j)$ é o valor pago para cada tonelada entregue a y_j independente de sua origem ser x_1 ou x_2.
No caso 1) dado qualquer plano de transporte π_{ij} de marginais p e q temos que o custo total associado será $\langle C, \pi \rangle$. Como este número depende de π podemos tentar determinar o plano que resulte no menor custo. No caso 2) independente do plano adotado o custo total será $\langle A, p \rangle + \langle B, q \rangle$. A desigualdade (6.1) garante que se $C(i,j) \geq A(i) + B(j)$, $\forall i, j$ então independente da escolha do plano para o caso 1) teremos um custo maior ou igual ao obtido no caso 2). Mesmo se mudarmos as funções A e B no caso 2), se a relação $C(i,j) \leq A(i) + B(j)$ permanecer, então a desigualdade continuará sendo verificada.

Note que cada par de funções A, B satisfazendo $C(i,j) \geq A(i) + B(j)$ nos fornece uma cota inferior de $I(C)$ e cada plano de transporte nos fornece uma cota superior de $I(C)$. Fixados C, p e q, se $C(i,j) \geq A(i) + B(j) \, \forall \, i, j$ e $\pi \in \Pi(p,q)$ então

$$\langle A, p \rangle + \langle B, q \rangle \leq I(C) \leq \langle C, \pi \rangle.$$

Exemplo 6.9. *Se $X = \{1, 2\}$, $Y = \{1, 2\}$, $p = (\frac{1}{2}, \frac{1}{2})$ e $q = (\frac{1}{2}, \frac{1}{2})$, os planos de transporte extremais são*

$$\eta^1 = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix} \;\; e \;\; \eta^2 \begin{pmatrix} 0 & 1/2 \\ 1/2 & 0 \end{pmatrix}.$$

Dado um custo $C = (C_{ij})$, onde $C_{ij} = C(i,j)$, um dos planos extremais será um plano ótimo.

1) Se ambos η^1 e η^2 são ótimos, então

$$\frac{C_{11}+C_{22}}{2} = \langle C, \eta^1\rangle = I(C) = \langle C, \eta^2\rangle = \frac{C_{12}+C_{21}}{2}.$$

Portanto

$$C_{11}+C_{22} = C_{12}+C_{21}.$$

Defina

$$A_1 = C_{11}, \quad A_2 = C_{21}, \quad B_1 = 0, \quad B_2 = C_{12} - C_{11}.$$

Então

$$A_2 + B_2 = (C_{21}) + (C_{12} - C_{11}) = C_{22}$$

e os números A_1, A_2, B_1, B_2 satisfazem

$$\begin{cases} A_1 + B_1 = C_{11} \\ A_1 + B_2 = C_{12} \\ A_2 + B_1 = C_{21} \\ A_2 + B_2 = C_{22} \end{cases}$$

Portanto $C(i,j) \geq A(i) + B(j) \,\forall\, i,j$. Além disso

$$\langle A, p\rangle + \langle B, q\rangle = \frac{A_1+A_2}{2} + \frac{B_1+B_2}{2}$$

$$= \frac{(C_{11})+(C_{21})}{2} + \frac{0+(C_{12}-C_{11})}{2} = \frac{C_{12}+C_{21}}{2} = I(C).$$

2) Suponha agora que η^1 é um plano ótimo e que η^2 não. Então

$$I(C) = \frac{C_{11}+C_{22}}{2} < \frac{C_{12}+C_{21}}{2}.$$

Defina

$$A_1 = 0, \quad A_2 = C_{21} - C_{11}, \quad B_1 = C_{11}, \quad B_2 = C_{22} + C_{11} - C_{21}.$$

Neste caso temos:

$$\begin{cases} A_1 + B_1 = C_{11} \\ A_1 + B_2 < C_{12} \\ A_2 + B_1 = C_{21} \\ A_2 + B_2 = C_{22} \end{cases}$$

Portanto $C(i,j) \geq A(i) + B(j) \,\forall\, i,j$. Além disso

$$\langle A, p\rangle + \langle B, q\rangle = \frac{A_1+A_2}{2} + \frac{B_1+B_2}{2}$$

$$= \frac{0+(C_{21}-C_{11})}{2} + \frac{(C_{11})+(C_{22}+C_{11}-C_{21})}{2} = \frac{C_{11}+C_{22}}{2} = I(C).$$

No exemplo anterior apresentamos um par de funções A e B verificando a igualdade em (6.1). Esta é a direção do resultado abaixo.

Teorema 6.10 (Dualidade de Kantorovich)**.**

$$\sup_{C(i,j)\geq A(i)+B(j)\,\forall\, i,j} \{\langle A,p\rangle + \langle B,q\rangle\} = \inf_{\pi\in\Pi(p,q)} \langle C,\pi\rangle,$$

onde o supremo acima deve ser tomado entre todos os possíveis pares de funções A e B verificando a desigualdade $C(i,j) \geq A(i) + B(j)\,\forall\, i,j$.

Não iremos apresentar a prova deste resultado neste texto. O leitor interessado poderá encontrar provas em [19] e [15].

Referências Bibliográficas

[1] ANTON, H. *Álgebra linear : com aplicações*. 8. ed. Porto Alegre : Bookman, 2001-2002.

[2] CONTRERAS, G.; LOPES, A.; THIEULLEN, PH. Lyapunov minimizing measures for expanding maps of the circle, *Ergodic Theory and Dynamical Systems*, V. 21, p. 1379-1409, 2001.

[3] GARIBALDI, E; LOPES, A. On the Aubry - Mather theory for symbolic dynamics. *Ergodic Theory and Dynamical Systems*, Vol 28, p. 791-815, 2008.

[4] KEMENY, J.; SNELL, J. *Finite Markov Chains*. Princeton : D. van Nostrand, 1963.

[5] LAY, D. *Álgebra linear e suas aplicações*. 2. ed. Rio de Janeiro : Livros Técnicos e Científicos, 1999.

[6] LIMA, E. *Álgebra linear*. 7. ed. Rio de Janeiro : Impa/CNPq, 2006.

[7] LIMA, E. *Análise real: volume 1*. 8. ed. Rio de Janeiro : IMPA, 2006.

[8] LOPES, A. Introdução à probabilidade e aos processos estocásticos: uma exposição para quem não sabe nada do assunto. *Matemática Universitária*, V. 38/39, p. 35-68, 2005.

[9] LOPES, A.; MENGUE, J. Duality Theorems in Ergodic Transport. *Journal of Statistical Physics*, Vol 149, Issue 5, p. 921-942, 2012.

[10] LOPES, A.; MENGUE, J.; MOHR, J.; SOUZA, R. Entropy, pressure and duality for Gibbs plans in Ergodic Transport. *Bull. Braz. Math. Soc.* v. 46, p. 353-389, 2015.

[11] LOPES, A.; OLIVEIRA, E. Entropy and variational principles for holonomic probabilities of IFS. *Discrete Contin. Dyn. Syst.* v. 23, no. 3, p. 937 - 955, 2009.

[12] LUENBERGER, D. *Introduction to dynamic systems : theory, models, and applications*. New York : Wiley, 1979.

[13] MANE, R. *Introdução à teoria ergódica*. Rio de Janeiro : Impa, 1983.

[14] MENGUE, J.; OLIVEIRA, E. Duality results for Iterated Function Systems with a general family of branches. (arXiv:1404.7801 [math.DS])

[15] OLIVEIRA, A. *O teorema da dualidade de Kantorovich para o transporte ótimo*. Dissertação de mestrado, UFRGS. Porto Alegre, 2011. (http://hdl.handle.net/10183/32470)

[16] OLIVEIRA, K.; VIANA, M. *Fundamentos da Teoria Ergódica*. 1ed. Rio de Janeiro: SBM, 2014.

[17] PARRY, W.; POLLICOTT, M. *Zeta functions and the periodic orbit structure of hyperbolic dynamics*. Astérisque Vol 187-188, 1990.

[18] POLLICOTT, M.; YURI, M. *Dynamical systems and ergodic theory*. Cambridge : Cambridge University Press, 1998.

[19] VILLANI, C. *Topics in optimal transportation*. Providence : American Mathematical Society, 2003.

[20] ROCKAFELLAR, R. *Convex analisys*. Princeton : Princeton University, 1970.

ISBN 978-85-8337-096-3

SBM
SOCIEDADE BRASILEIRA DE MATEMÁTICA

FURG