Esta Tese foi julgada adequada para a obtenção do título de

"Mestre em Ciências"

especialidade em Matemática e aprovada em sua forma final pelo Curso de Pós-Graduação.

Prof. William Glenn Whitley, Ph.D.
Coordenador

BANCA EXAMINADORA:

Prof. Inder Jeet Taneja, Ph.D.
Orientador

Prof. Dr. Teófilo Abuabara Saad

Prof. William Glenn Whitley, Ph.D

ENTROPIAS GENERALIZADAS E CAPACIDADE

DO CANAL DISCRETO SEM MEMÓRIA

Fernando Guerra

Novembro - 1980

A Raquel (in Memória) e Délio, meus pais e a Maria Alice, minha esposa, dedico este trabalho.

AGRADECIMENTOS

Ao Professor Dr. Inder Jeet Taneja, pela segurança, dedicação, estímulo e compreenção durante a elaboração deste trabalho, os meus agradecimentos.

Estendo meus agradecimentos à Universidade Federal de Santa Catarina.

RESUMO

Neste trabalho apresentamos os teoremas da Capacidade de Ordem α e de Grau β de um Canal Discreto Sem Memória, usando Entropias Generalizadas (Entropia de Rényi e Entropia de Daróczy) através da Entropia de Shannon.

ABSTRACT

In this work, we present two theorems on Capacity of order α and Capacity of degree β of a discrete memoryless channel, using generalized entropies i.e., Rényi entropy of order α and Daróczy entropy of degree β. These entropies are generalization of Shannon entropy.

ÍNDICE

INTRODUÇÃO

No presente trabalho temos o objetivo de calcular a Capacidade do Canal Discreto Sem Memória usando Entropias Generalizadas através da Entropia de Shannon.

No Capítulo 1, apresentamos inicialmente o conceito de um Sistema de Comunicação e definimos a Entropia de Shannon. Definimos também a Informação Comunicada pelo Canal sobre X a partir de Y, bem como a Capacidade do Canal. Apresentamos o conceito de Canal Discreto Sem Memória e a seguir apresentamos as Entropias de Rényi [1961] e de Daróczy [1970].

No Capítulo 2, apresentamos o nosso trabalho, sobre a Entropia de Rényi e Doráczy para a Capacidade do Canal de Ordem α (Entropia de Rényi) e para a Capacidade do Canal de Grau β (Entropia de Daróczy).

CAPÍTULO 1

INTRODUÇÃO

Teoria de Informação é um ramo da Teoria de Probabilidade, com grande aplicação nos sistemas de comunicações, na qual teve uma origem física. Ela foi iniciada por cientistas da comunicação que estudavam a estrutura estatística de equipamentos elétricos de comunicações. Ela foi iniciada por Shannon em 1948.

1.1 - Sistema de Comunicação

Um "sistema clássico de comunicação" tem a forma representada na Fig. 1.1 abaixo

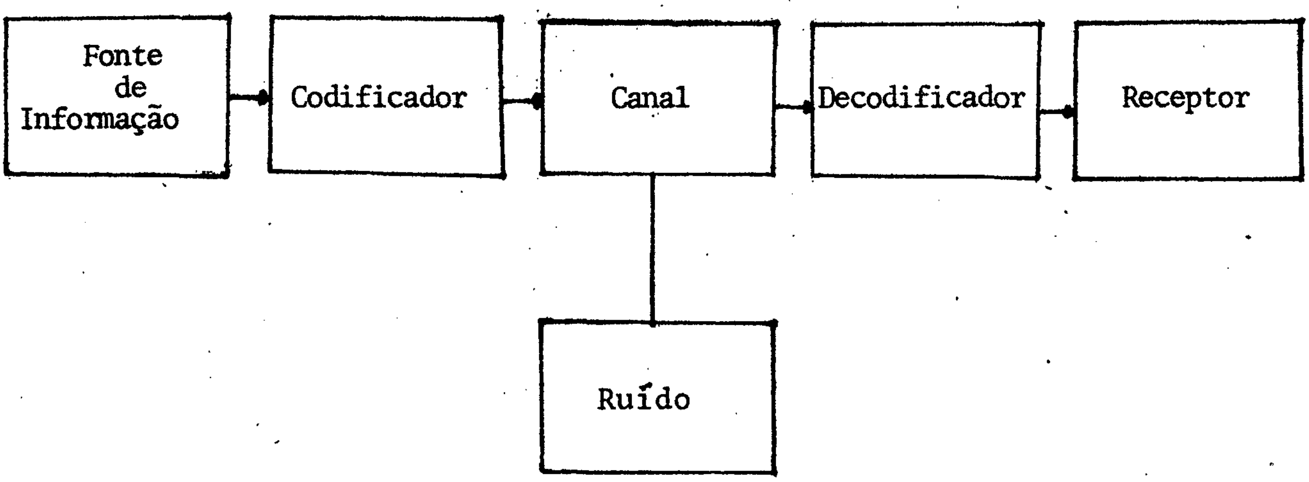

Fig. 1.1

onde

Fonte de Informação: É o componente do sistema que é capaz de produzir mensagens. A fonte pode ser uma pessoa ou máquina. Por exemplo, no computador é uma máquina, no telefone é uma pessoa, na televisão é uma máquina ou pessoa. Geralmente a fonte de Informação é representada por uma variável aleatória discreta ou contínua X.

Codificador: Transforma a linguagem da fonte para a linguagem do canal, mantendo inalterado o seu conteúdo, sendo portanto responsável pela mudança da forma de linguagem.

Canal: É o meio através do qual a mensagem é propagada.

Decodificador: Após receber a mensagem que foi transmitida pelo canal sua finalidade é de decifrá-la de modo que a mesma seja inteligível pelo receptor.

Receptor: É o ponto de destino da mensagem e representado pela variável aleatória discreta ou contínua Y.

A Teoria de Informação tenta construir um modelo matemático para cada um dos blocos da Fig. 1.1 acima. Uma aplicação de um modelo de Sistema de Comunicação, como na Fig. 1.1. acima, foi feito por Wiener e Shannon em suas discussões sobre a natureza estatística da comunicação de mensagens. Em tais modelos de comunicação, a fonte de informação, o codificador, o canal, o decodificador e o receptor devem estar estatisticamente definidos.

Os sistemas de comunicação, aqui citados, são de natureza estatística, isto é, o comportamento dos sistemas nunca podem ser descritos de uma forma determinística, ele é sempre dado em termos estatísticos. Por exemplo, uma fonte de informação é um aparato que seleciona e transmite seqüências de símbolos ao acaso, embora esta seleção possa ser baseada em alguma estatística.

Em Teoria de Informação temos o conhecido TEOREMA FUNDAMENTAL DA INFORMAÇÃO e enunciado da maneira como segue:

"É possível transmitir informação" pelo Canal Ruidoso, em qualquer velocidade menor do que a capacidade do canal com uma pequena e arbitrária probabilidade de erro, isto é, se

R - é a velocidade de transmissão da informação e

C - a capacidade do canal

então

se $R < C$, pode-se transmitir a informação com pequena probabilidade de erro, em qualquer canal ruidoso.

Este teorema foi provado a primeira vez por Shannon em 1948, usando uma medida de incerteza (ou entropia de Shannon).

1.2 - Entropia de Shannon

Seja $X = (x_1,\ldots,x_n)$ uma variável aleatória discreta com distribuição de probabilidades $(p(x_1),\ldots,p(x_n))$ tal que $p(x_i) \geqslant 0$, $i = 1,2,\ldots,n$ e $\sum_{i=1}^{n} p(x_i) = 1$.

A Entropia de Shannon é definida por

$$H(X) = H(p(x_1),\ldots,p(x_n)) = - \sum_{i=1}^{n} p(x_i) \log_2 p(x_i) \tag{1.1}$$

Suponhamos que X e Y são duas variáveis aleatórias discretas associadas com o mesmo experimento. X e Y têm probabilidade conjunta

$$p(x_i, y_j) = P\{X = x_i ; Y = y_j\} \tag{1.2}$$

para $i = 1,2,\ldots,n$ e

$j = 1,2,\ldots,m$

O resultado $\{X = x_i ; Y = y_j\}$ tem probabilidade conjunta $p(x_i, y_j)$. Então a entropia conjunta é definida por

$$H(X,Y) = - \sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i, y_j) \log_2 p(x_i, y_j) \tag{1.3}$$

Nota: - Salvo indicação em contrário, o logarítmo será sempre considerado com base 2.

Definição 1.2.1 - (Probabilidade Individual)

A probabilidade do evento $X = x_i$ no primeiro experimento indiferente a do segundo é definida por

$$p(x_i) = \sum_{j=1}^{m} p(x_i, y_j) \quad \text{para} \quad i = 1,2,\ldots,n \tag{1.4}$$

Definição 1.2.2 - (Probabilidade Individual)

A probabilidade do evento $Y = y_j$ no segundo experimento indiferente a do primeiro é dado por

$$p(y_j) = \sum_{i=1}^{n} p(x_i, y_j) \quad \text{para} \quad j = 1,2,\ldots,m \tag{1.5}$$

Definição 1.2.3 - (Probabilidade Condicional)

A probabilidade condicional de ocorrer o evento x_i do experimento X dado que o evento y_j do experimento Y ocorrer é definida por

$$p(x_i/y_j) = \frac{p(x_i, y_j)}{p(y_j)} \tag{1.6}$$

Definição 1.2.4 - (Probabilidade Condicional)

A probabilidade condicional de ocorrer o evento y_j do experimento Y dado que o evento x_i do experimento X ocorrer é definida por

$$p(y_j/x_i) = \frac{p(x_i, y_j)}{p(x_i)} \tag{1.7}$$

Quando X e Y são independentes temos

$$p(x_i/y_j) = \frac{p(x_i).p(y_j)}{p(y_j)} = p(x_i) \tag{1.8}$$

e

$$p(y_j/x_i) = \frac{p(x_i).p(y_j)}{p(x_i)} = p(y_j) \tag{1.9}$$

Definimos a incerteza condicional de Y dado que $X = x_i$, por

$$H(Y/X = x_i) = - \sum_{j=1}^{m} p(y_j/x_i) \log p(y_j/x_i) \tag{1.10}$$

para $i = 1,2,...,n$

Logo, definimos a incerteza condicional Y dado X como a incerteza média da $H(Y/X = x_i)$ com pesos $p(x_i)$, $i = 1,2,...,n$, por

$$H(Y/X) = \sum_{i=1}^{n} p(x_i)\ H(Y/X = x_i) \tag{1.11}$$

$$= - \sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i)\ p(y_j/x_i)\ \log\ p(y_j/x_i) \tag{1.12}$$

$$= - \sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i,\ y_j)\ \log\ p(y_j/x_i) \tag{1.13}$$

Similarmente, podemos definir

$$H(X/Y) = - \sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i,\ y_j)\ \log\ p(x_i/y_j) \tag{1.14}$$

Através das definições de incerteza obtemos os seguintes resultados:

$$H(X,Y) = H(X) + H(Y/X) = H(Y) + H(X/Y) \tag{1.15}$$

$$H(X,Y) \leq H(X) + H(Y) \tag{1.16}$$

$$H(Y/X) \leq H(Y) \tag{1.17}$$

Com a igualdade em (1.16) e (1.17) se e somente se X e Y são independentes.

1.2.1 - Informação Comunicada pelo Canal sobre X a partir de Y

Definimos a "informação comunicada pelo canal sobre X a partir de Y" como

$$I(X/Y) = H(X) - H(X/Y) \tag{1.18}$$

$$= - \sum_{i=1}^{n} p(x_i) \;\; \log p(x_i)$$

$$+ \sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i, y_j) \;\; \log p(x_i/y_j)$$

$$= - \sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i, y_j) \;\; \log p(x_i)$$

$$+ \sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i, y_j) \;\; \log p(x_i/y_j)$$

$$= - \sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i, y_j) \;\; \log \frac{p(x_i)}{p(x_i/y_j)} \qquad (1.19)$$

Esta medida de informação é as vezes chamada INFORMAÇÃO MÚTUA e satisfaz as seguintes propriedades:

$$I(X/Y) \geq 0 \qquad \text{(não negativa)} \qquad (1.20)$$

$$I(X/Y) = I(Y/X) \quad \text{(simetria)} \qquad (1.21)$$

isto é,

a informação comunicada sobre Y por X é a mesma informação sobre X por Y.

Definição 1.2.5 - (Capacidade do Canal)

A máxima informação comunicada pelo canal é chamada Capacidade do Canal. A maximização é tomada sobre o conjunto de todas as distribuições de probabilidades de entrada $\{p(x_i)\}$, $i = 1,2,\ldots,n$, isto é,

$$C = \max_{\substack{p(x_i) \\ i = 1,2,\ldots,n}} I(X/Y) \tag{1.22}$$

1.3 - Canal Discreto Sem Memória

Nesta secção começaremos a análise de canais de comunicação da qual as entradas estão sujeitas a pertubações na transmissão. Na situação usual de comunicação um objeto, de alguma maneira (por exemplo, uma letra de um alfabeto) é selecionado desde uma classe específica de entradas. O canal é um esquema que atua na entrada para produzir uma saida pertencente a outra classe específica. A classe variável de um canal, pode em muitos casos ser descrita dando uma distribuição de probabilidade sobre um conjunto das possíveis saídas. A distribuição dependerá, em geral, de uma particular entrada escolhida para transmissão e também do estado interno do canal para o tempo quando a entrada é aplicada. O estado interno do canal enquanto a entrada é aplicada afetará a transmissão de informação.

Interessaremos para a situação na qual à informação a ser transmitida consiste de uma seqüência de símbolos, cada símbolo pertencente a um alfabeto finito.

Definição 1.3.1.

Um Canal Discreto Sem Memória com alfabeto de entrada $X = (x_1,\ldots,x_n)$ e alfabeto de saída $Y = (y_1,\ldots,y_m)$ onde X é uma variável aleatória discreta com distribuição de probabilidades

$p(x_1),\ldots,p(x_n)$ e Y é uma variável aleatória discreta com distribuição de probabilidades $p(y_1),\ldots,p(y_m)$ é definido pela matriz n x m abaixo

$$\begin{bmatrix} p(y_1/x_1) & p(y_2/x_1) & \ldots\ p(y_m/x_1) \\ p(y_1/x_2) & p(y_2/x_2) & \ldots\ p(y_m/x_2) \\ p(y_1/x_3) & p(y_2/x_3) & \ldots\ p(y_m/x_3) \\ \ldots\ldots & \ldots\ldots & \ldots\ldots \\ \ldots\ldots & \ldots\ldots & \ldots\ldots \\ p(y_1/x_n) & p(y_2/x_n) & \ldots\ p(y_m/x_n) \end{bmatrix} \qquad (1.23)$$

ou

$$\{ p(y_j/x_i) \quad \text{com} \quad i = 1,2,\ldots,n \quad \text{e} \quad j = 1,2,\ldots,m \}$$

onde

$p(y_j/x_i)$ representa probabilidades condicionais de palavras códigos recebidas $y_j \in Y$ enquanto $x_i \in X$ é transmitida, com

$\sum_{j=1}^{m} p(y_j/x_i) = 1$ para $i = 1,2,\ldots,n$.

A matriz em (1.23) é chamada Matriz Canal.

Resumiremos agora o Teorema da Capacidade do Canal Discreto Sem Memória, (Ref., Gallager [1968]).

<u>Teorema 1.1</u>.

Um conjunto de condições necessárias e suficientes sob um vetor de probabilidade $P = (p(x_1),\ldots,p(x_n))$ para atingir a ca pacidade de um canal discreto sem memória, com entrada X e saída Y e matriz canal $[p(y_j/x_i)]$ é

$$I(X = x_i/Y) \begin{cases} = C & \text{se } p(x_i) > 0 \\ \leq C & \text{se } p(x_i) = 0 \end{cases} \tag{1.24}$$

para todo $i = 1,2,\ldots,n$

onde

$I(X = x_i/Y)$ é a Informação Comunicada pelo canal para a entrada x_i a partir da saída Y, dada por

$$I(X = x_i/Y) = \sum_{j=1}^{m} p(y_j/x_i) \log \frac{p(y_j/x_i)}{\sum_{k=1}^{n} p(x_k)\, p(y_j/x_k)} \tag{1.25}$$

para todo $i = 1,2,\ldots,n$

Sejam

X^N - o conjunto de todas as seqüências de entrada de comprimento N que pode ser transmitida, isto é,

$$X^N = \{ x = (x_1 x_2 x_3 \ldots x_N) \text{ tal que } x_t \in X \text{ para } t = 1,2,\ldots,N \}$$

e

Y^N - conjunto de todas as seqüências de saída de comprimento N que pode ser recebida, isto é,

$$Y^N = \{ y = (y_1 y_2 y_3 \ldots y_N) \text{ tal que } y_t \in Y \text{ para } t = 1,2,\ldots,N \}$$

Assumiremos que X^N e Y^N são ambos finitos.

Definição 1.3.2 - (Probabilidade de Erro)

A probabilidade de erro de codificação para um certo código $\mathcal{C}$, com M palavras códigos de comprimento N, denotada por $P_e(\mathcal{C})$ é definida da forma convencional

$$P_e(\mathcal{C}) = \sum_{x_m \in X^N} \frac{1}{M} \quad P_{em} \tag{1.26}$$

onde

P_{em} é a probabilidade de erro de codificação quando a seqüência x_m é transmitida.

Quando $x = (x_1 x_2 \ldots x_N) \in X^N$ é transmitida, a probabilidade de receber a seqüência $y = (y_1 y_2 \ldots y_N) \in Y^N$ é determinado por

$$p(y/x) = \prod_{t=1}^{N} p(y_t/x_t) \tag{1.27}$$

Em 1965, Gallager conseguiu uma elegante prova do teorema de codificação de Shannon e obteve uma nova limitação para a probabilidade de erro. O seu contrário foi recentemente apresentado por Arimoto [1973].

Resumiremos agora o teorema de codificação de Gallager e seu contrário da seguinte forma.

Teorema 1.2:

Seja

$$E_o(\rho, P) = -\log \left[\sum_{j=1}^{m} \left\{ \sum_{i=1}^{n} p(x_i) \; p(y_j/x_i)^{\frac{1}{1+\rho}} \right\}^{1+\rho} \right] \tag{1.28}$$

onde

$P = (p(x_1), \ldots, p(x_n))$ é uma distribuição de probabilidade arbitrária através do alfabeto de entrada X, ρ é uma constante fixada e fazendo

$$R = \frac{\log M}{N}$$

1) - (Gallager [1965]) - Existe pelo menos um bloco de código de comprimento N tal que a probabilidade de erro de codificação é limitada superiormente por

$$P_e(\mathcal{C}) \leq \exp\left[-N\left\{-\rho R + \max_{P} E_o(\rho, P)\right\}\right] \tag{1.29}$$

$$\text{para } 0 < \rho \leq 1$$

2) - (Arimoto [1973]) - Para algum bloco de código de comprimento N a probabilidade de erro de codificação é limitada inferiormente por

$$P_e(\mathcal{C}) \geq 1 - \exp\left[-N\left\{-\rho R + \min_{P} E_o(\rho, P)\right\}\right] \tag{1.30}$$

$$\text{para } -1 \leq \rho < 0$$

1.4 - Entropia de Ordem α

A Entropia de Ordem α obtida por Rényi [1961], para uma distribuição de probabilidade $(p(x_1), \ldots, p(x_n))$ de uma variável aleatória discreta $X = (x_1, \ldots, x_n)$ é definida por

$$H_\alpha(X) = \frac{1}{1-\alpha} \log\left\{\sum_{i=1}^{n} p^\alpha(x_i)\right\} \quad , \quad \alpha \neq 1 \ , \quad \alpha > 0 \tag{1.31}$$

Muitos pesquisadores realizam trabalhos sobre esta entropia de Rényi, especialmente o problema da construção axiomática desta entropia, (Ref. Aczel e Daróczy, [1975]).

A entropia de ordem α pode ser reescrita na forma

$$H_\alpha(X) = \frac{\alpha}{1-\alpha} \log \left\{ \sum_{i=1}^{n} p^\alpha(x_i) \right\}^{\frac{1}{\alpha}} , \quad \alpha \neq 1, \alpha > 0 \tag{1.32}$$

A entropia condicional de ordem α, de uma variável aleatória discreta $X = (x_1,\ldots,x_n)$ é definida como

$$H_\alpha(X/Y) = \frac{\alpha}{1-\alpha} \log \left[\sum_{j=1}^{m} p(y_j) \left\{ \sum_{i=1}^{n} p^\alpha(x_i/y_j) \right\}^{1/\alpha} \right] ,$$

$$\alpha \neq 1 , \alpha > 0 \tag{1.33}$$

Usando (1.6) e (1.7) , (1.33) pode ser reescrita na forma

$$H_\alpha(X/Y) = \frac{\alpha}{1-\alpha} \log \left[\sum_{j=1}^{m} \left\{ \sum_{i=1}^{n} p^\alpha(x_i) \, p^\alpha(y_j/x_i) \right\}^{1/\alpha} \right]$$

$$\alpha \neq 1 , \alpha > 0 \tag{1.34}$$

Similarmente, podemos definir

$$H_\alpha(Y/X) = \frac{\alpha}{1-\alpha} \log \left[\sum_{i=1}^{n} \left\{ \sum_{j=1}^{m} p^\alpha(y_j) \, p^\alpha(x_i/y_j) \right\}^{1/\alpha} \right]$$

$$\alpha \neq 1 , \alpha > 0 \tag{1.35}$$

Podemos verificar simplesmente que

$$\lim_{\alpha \to 1} H_\alpha(X/Y) = H(X/Y) \tag{1.36}$$

e

$$\lim_{\alpha \to 1} H_\alpha(X) = H(X) \tag{1.37}$$

onde

$H(X)$ e $H(X/Y)$ representam, respectivamente, as entropias definidas em (1.1) e (1.14).

Geralmente

$$H_\alpha(X/Y) \leq H_\alpha(X) \tag{1.38}$$

com a igualdade se e somente se X e Y são independentes.

De (1.38) segue que

$$H_\alpha(X) - H_\alpha(X/Y) \geq 0 \tag{1.39}$$

1.4.1 - Informação Comunicada Pelo Canal de Ordem α

Definimos a informação comunicada pelo canal de ordem α sobre X a partir de Y por

$$I_\alpha(X/Y) = H_\alpha(X) - H_\alpha(X/Y) \tag{1.40}$$

a qual é não negativa por (1.39).

Logo,

$$I_\alpha(X/Y) = H_\alpha(X) - H_\alpha(X/Y)$$

$$= \frac{\alpha}{1-\alpha} \log \left\{ \sum_{i=1}^{n} p^\alpha(x_i) \right\}^{1/\alpha}$$

$$- \frac{\alpha}{1-\alpha} \log \left[\sum_{j=1}^{m} p(y_j) \left\{ \sum_{i=1}^{n} p^\alpha(x_i/y_j) \right\}^{1/\alpha} \right] ,$$

$$\alpha \neq 1 \ , \ \alpha > 0$$

$$= \frac{\alpha}{1-\alpha} \log \left\{ \sum_{i=1}^{n} p^{\alpha}(x_i) \right\}^{1/\alpha}$$

$$- \frac{\alpha}{1-\alpha} \log \left[\sum_{j=1}^{m} \left\{ \sum_{i=1}^{n} p^{\alpha}(x_i)\, p^{\alpha}(y_j/x_i) \right\}^{1/\alpha} \right]$$

$$= \frac{-\alpha}{1-\alpha} \log \frac{\sum_{j=1}^{m} \left\{ \sum_{i=1}^{n} p^{\alpha}(x_i)\, p^{\alpha}(y_j/x_i) \right\}^{1/\alpha}}{\left\{ \sum_{i=1}^{n} p^{\alpha}(x_i) \right\}^{1/\alpha}}, \quad \alpha \neq 1,\ \alpha > 0 \qquad (1.41)$$

Definição: Capacidade de Ordem α, de um Canal Discreto Sem Memória.

A capacidade de ordem α, de um canal discreto sem memória, denotada por C_α, é definida como sendo o máximo de $I_\alpha(X/Y)$. A maximização é tomada sobre o conjunto de todas as distribuições de probabilidades de entrada $\{p(x_i)\}$, $i = 1,2,\ldots,n$, isto é,

$$C_\alpha = \max_{\substack{p(x_i) \\ i=1,2,\ldots,n}} I_\alpha(X/Y) \qquad (1.42)$$

1.5 - Entropia de Grau β

Daróczy [1970], introduziu o conceito de uma função informação de grau β e definiu a entropia de grau β para uma distribuição de probabilidade $(p(x_1),\ldots,p(x_n))$ de uma variável aleatória discreta $X = (x_1,\ldots,x_n)$ por

$$H^{\beta}(X) = (2^{1-\beta} - 1)^{-1} \left[\sum_{i=1}^{n} p^{\beta}(x_i) - 1 \right] , \beta \neq 1 , \beta > 0 \tag{1.43}$$

Quando $\beta \to 1$, temos

$$\lim_{\beta \to 1} H^{\beta}(X) = H(X) = - \sum_{i=1}^{n} p(x_i) \log p(x_i)$$

na qual é a Entropia de Shannon.

Se X e Y são duas variáveis aleatórias discretas com distribuição de probabilidade conjunta $p(x_i, y_j)$ para $i = 1,2,\ldots,n$ e $j = 1,2,\ldots,m$, definimos a Entropia Conjunta de Grau β por

$$H^{\beta}(X,Y) = (2^{1-\beta} - 1)^{-1} \left[\sum_{i=1}^{n} \sum_{j=1}^{m} p^{\beta}(x_i, y_j) - 1 \right] , \beta \neq 1, \beta > 0 \tag{1.44}$$

Definimos a Entropia Condicional de grau β de X dado que $Y = y_j$, $j = 1,2,\ldots,m$, por

$$H^{\beta}(X/Y = y_j) = (2^{1-\beta} - 1)^{-1} \left[\sum_{i=1}^{n} p^{\beta}(x_i/y_j) - 1 \right] , \beta \neq 1 , \beta > 0 \tag{1.45}$$

Então definimos a Entropia Condicional de X dado Y como sendo

$$H^{\beta}(X/Y) = \sum_{j=1}^{m} p^{\beta}(y_j) \, H^{\beta}(X/Y = y_j)$$

$$= (2^{1-\beta} - 1)^{-1} \sum_{j=1}^{m} p^{\beta}(y_j) \left[\sum_{i=1}^{n} p^{\beta}(x_i/y_j) - 1 \right] ,$$

$$\beta \neq 1 , \beta > 0 \tag{1.46}$$

Similarmente, podemos definir a Entropia Condicional de Y dado X por

$$H^{\beta}(Y/X) = (2^{1-\beta} - 1)^{-1} \sum_{i=1}^{n} p^{\beta}(x_i) \left[\sum_{j=1}^{m} p^{\beta}(y_j/x_i) - 1 \right],$$

$$\beta \neq 1 \ , \ \beta > 0 \qquad (1.47)$$

Estas entropias, tem as seguintes propriedades, (Ref.: Sharma e Autar, [1973]):

i) $H^{\beta}(X,Y) = H^{\beta}(Y/X) + H^{\beta}(X)$

$= H^{\beta}(X/Y) + H^{\beta}(Y)$

ii) $H^{\beta}(X,Y) = H^{\beta}(X) + H^{\beta}(Y) + (2^{1-\beta} - 1)^{-1} H^{\beta}(X).H^{\beta}(Y)$

iii) $H^{\beta}(Y/X) = H^{\beta}(Y) + (2^{1-\beta} - 1) . H^{\beta}(X).H^{\beta}(Y)$

iv) $H^{\beta}(X/Y) \leq H^{\beta}(X)$

com a igualdade em iv) se e somente se X e Y são independentes.

Além disso, estas entropias de grau β satisfazem muitas outras propriedades, (Ref. Aczel e Daróczy, [1975] e Taneja [1979])

1.5.1 - Informação Comunicada Pelo Canal de Grau β

Definimos a informação comunicada pelo canal de grau β sobre X a partir de Y por

$$I^{\beta}(X/Y) = H^{\beta}(X) - H^{\beta}(X/Y) \qquad (1.48)$$

na qual satisfaz as seguintes propriedades

$$I^{\beta}(X/Y) \geq 0 \qquad \text{(não negativa)} \tag{1.49}$$

$$I^{\beta}(X/Y) = I^{\beta}(Y/X) \quad \text{(simetria)}, \tag{1.50}$$

isto é, a informação comunicada pelo canal de grau β sobre Y por X é a mesma informação sobre X por Y.

Definição: Capacidade de Grau β De Um Canal Discreto Sem Memória.

A capacidade de grau β, de um canal discreto sem memória, denotada por C^{β}, é definida como sendo a máxima informação comunicada pelo canal de grau β. A maximização é tomada sobre o conjunto de todas as distribuições de probabilidades de entradas, isto é,

$$C^{\beta} = \max_{\substack{p(x_i) \\ i = 1,2,\ldots,n}} I^{\beta}(X/Y) \tag{1.51}$$

CAPÍTULO 2

ENTROPIAS GENERALIZADAS E CAPACIDADE DO CANAL DISCRETO SEM MEMÓRIA

Neste capítulo apresentamos o teorema da capacidade do canal de ordem α e de grau β, objetivo de nosso trabalho. Para o teorema da capacidade do canal de ordem α, precisamos a convexidade da função $I_\alpha(X/Y)$, sua maximização e sua continuidade, que citamos na secção 2.1.. Na secção 2.2. apresentamos o teorema da capacidade do canal de grau β.

2.1 - Capacidade do Canal de Ordem α

A capacidade do canal de ordem α, de um canal discreto sem memória, denotada por C_α, já foi definida no Capítulo 1, como sendo a máxima informação de ordem α comunicada pelo canal. A maximização é tomada sobre o conjunto de todas as distribuições de probabilidades de entradas $(p(x_1),\ldots,p(x_n))$, isto é,

$$C_\alpha = \max_{\substack{p(x_i)\\ i=1,2,\ldots,n}} I_\alpha(X/Y)$$

$$= \max_{\substack{p(x_i)\\ i=1,2,\ldots,n}} \frac{-\alpha}{1-\alpha} \log \frac{\sum_{j=1}^{m}\left\{\sum_{i=1}^{n} p^\alpha(x_i)\, p^\alpha(y_j)\; x_i)\right\}^{1/\alpha}}{\left\{\sum_{i=1}^{n} p^\alpha(x_i)\right\}^{1/\alpha}}$$

$$\alpha \neq 1,\ \alpha > 0 \qquad (2.1)$$

Seja a função $E_o(\rho, P)$, como no teorema 1.2, definimos agora

$$F(\rho, P) = \sum_{j=1}^{m} \left[\sum_{i=1}^{n} p(x_i)\, p(y_j/x_i)^{1/1+\rho} \right]^{1+\rho} \tag{2.2}$$

Logo

$$E_o(\rho, P) = - \log \left[F(\rho, P) \right] \tag{2.3}$$

<u>**Definição 2.1.1**</u> - (<u>Convexidade de uma Função</u>)

Uma função (contínua) $f: [a,b] \to R$ é dita convexa para baixo quando

$$f\left(\sum_{i=1}^{n} \lambda_i x_i\right) \leqslant \sum_{i=1}^{n} \lambda_i\, f(x_i), \tag{2.4}$$

quaisquer que sejam $x_1, x_2, \ldots, x_n \in [a,b]$ e quaisquer números reais não negativos $\lambda_1, \lambda_2, \ldots, \lambda_n$ tais que $\sum_{i=1}^{n} \lambda_i = 1$. Para convexa para cima temos "$\geqslant$" em (2.4).

<u>**Teorema 2.1**</u>:-

A função $F(\rho, P) = \sum_{j=1}^{m} \left[\sum_{i=1}^{n} p(x_i)\, p(y_j/x_i)^{1/1+\rho} \right]^{1+\rho}$

é convexa para baixo de P sobre a região

$A^n = \{P : P = (p(x_1), \ldots, p(x_n)),\ p(x_i) \geqslant 0,\ i = 1,2,\ldots,n,$

$\sum_{i=1}^{n} p(x_i) = 1\}$ na qual P é um vetor de probabilidade e $\rho \geqslant 0$.

<u>**Prova**</u>:

A função $F(\rho, P)$, pode ser reescrita como

$$F(\rho, P) = \sum_{j=1}^{m} t_j^{1+\rho} \tag{2.5}$$

onde

$$t_j = \sum_{i=1}^{n} p(x_i)\, p(y_j/x_i)^{1/1+\rho} \tag{2.6}$$

Sejam P e Q elementos de A^n, $P = (p(x_1),\ldots,p(x_n))$ e $Q = (q(x_1),\ldots,q(x_n))$ tal que

$$t_j = \sum_{i=1}^{n} p(x_i)\, p(y_j/x_i)^{1/1+\rho} \quad \text{e}$$

$$k_j = \sum_{i=1}^{n} q(x_i)\, p(y_j/x_i)^{1/1+\rho} \quad .$$

Para algum $\lambda \in (0,1)$, temos

$$F(\rho, \lambda P + (1-\lambda)Q) =$$

$$= \sum_{j=1}^{m} \left[\sum_{i=1}^{n} (\lambda p(x_i) + (1-\lambda)\, q(x_i)).p(y_j/x_i)^{1/1+\rho} \right]^{1+\rho} \tag{2.7}$$

$$= \sum_{j=1}^{m} \left[\lambda \sum_{i=1}^{n} p(x_i)\, p(y_j/x_i)^{1/1+\rho} + (1-\lambda) \sum_{i=1}^{n} q(x_i)\, p(y_j/x_i)^{1/1+\rho} \right]^{1+\rho}$$

$$= \sum_{j=1}^{m} (\lambda\, t_j + (1-\lambda)\, k_j)^{1+\rho}$$

Como os t_j e k_j, para $j = 1,2,\ldots,m$, são não negativos e como a função $W(x) = x^{1+\rho}$, para $x \geqslant 0$ e $\rho \geqslant 0$ é convexa para baixo, temos que

$$F(\rho, \lambda P + (1-\lambda)\, Q) = \sum_{j=1}^{m} (\lambda\, t_j + (1-\lambda)\, k_j)^{1+\rho}$$

$$\leqslant \sum_{j=1}^{m} \lambda t_j^{1+\rho} + \sum_{j=1}^{m} (1 - \lambda) k_j^{1+\rho}$$

(Ref.: Gallager, [1968], pág. 523)

$$= \lambda \sum_{j=1}^{m} t_j^{1+\rho} + (1 - \lambda) \sum_{j=1}^{m} k_j^{1+\rho} \quad , \quad \rho \geqslant 0$$

$$= \lambda F(\rho, P) + (1 - \lambda) F(\rho, Q)$$

Logo

$$F(\rho, \lambda P + (1 - \lambda) Q) \leqslant \lambda F(\rho, P) + (1 - \lambda) F(\rho, Q) \tag{2.8}$$

Portanto, $F(\rho, P)$ é uma função convexa para baixo de P para $\rho \geqslant 0$.

Para calcular a capacidade de ordem α, de um canal discreto sem memória, precisamos dos lemas seguintes.

Lema 2.1: (Convexidade de $I_\alpha(X/Y)$)

A informação comunicada pelo canal de ordem α é uma função convexa para baixo das probabilidades de entradas, para $\frac{1}{2} \leqslant \alpha < 1$.

Prova:

Sabemos que

$$I_\alpha(X/Y) = H_\alpha(X) - H_\alpha(X/Y)$$

$$= \frac{\alpha}{1 - \alpha} \log \left\{ \sum_{i=1}^{n} p^\alpha(x_i) \right\}^{1/\alpha}$$

$$- \frac{\alpha}{1-\alpha} \log \sum_{j=1}^{m} \left\{ \sum_{i=1}^{n} p^{\alpha}(x_i)\, p^{\alpha}(y_j/x_i) \right\}^{1/\alpha} , \quad \alpha \neq 1, \; \alpha > 0$$

$$= \frac{-\alpha}{1-\alpha} \log \frac{\sum_{j=1}^{m} \left\{ \sum_{i=1}^{n} p^{\alpha}(x_i)\, p^{\alpha}(y_j/x_i) \right\}^{1/\alpha}}{\left\{ \sum_{i=1}^{n} p^{\alpha}(x_i) \right\}^{1/\alpha}}$$

$$= \frac{-\alpha}{1-\alpha} \log \left[\sum_{j=1}^{m} \left\{ \sum_{i=1}^{n} p'(x_i)\, p^{\alpha}(y_j/x_i) \right\}^{1/\alpha} \right]$$

onde

$$p'(x_i) = \frac{p^{\alpha}(x_i)}{\sum_{k=1}^{n} p^{\alpha}(x_k)} , \quad \text{com } i = 1,2,\ldots,n \tag{2.9}$$

Portanto,

$$I_{\alpha}(X/Y) = \frac{-\alpha}{1-\alpha} \; E_o(\alpha, P') \tag{2.10}$$

onde

$$E_o(\alpha, P') = -\log \left[\sum_{j=1}^{m} \left\{ \sum_{i=1}^{n} p'(x_i)\, p^{\alpha}(y_j/x_i) \right\}^{1/\alpha} \right] \tag{2.11}$$

com $P' = (p'(x_1),\ldots,p'(x_n))$ um vetor de probabilidade de entra - da, $\alpha \neq 1$, $\alpha > 0$.

A função $E_o(\alpha, P')$ é a função $E_o(\rho, P)$ como no teorema 1.2, com a convenção

$$\alpha = \frac{1}{1+\rho} , \quad \frac{1}{2} \leq \alpha < +\infty.$$

Logo,

$$I_\alpha(X/Y) = \frac{1}{\rho} E_o(\rho, P) \quad , \quad \rho > 0 \tag{2.12}$$

Por (2.3), temos que

$$I_\alpha(X/Y) = \frac{-1}{\rho} \log \left[F(\rho, P)\right] \quad , \quad \rho > 0 \tag{2.13}$$

Como $F(\rho, P)$ é uma função convexa para baixo (convexa para cima) para $\rho \geq 0$ ($\rho < 0$). Logo, $E_o(\rho, P) = -\log\left[F(\rho, P)\right]$ é uma função convexa para cima (convexa para baixo) para $\rho \geq 0(\rho < 0)$, portanto,

$I_\alpha(X/Y) = \frac{-1}{\rho} \log\left[F(\rho, P)\right]$ é convexa para cima (convexa para baixo) para $\rho > 0$ ($\rho < 0$) onde $\alpha = \frac{1}{1+\rho}$. Isto implica que $I_\alpha(X/Y)$ é uma função convexa para cima para $\alpha \geq 1$ e convexa para baixo para $\alpha < 1$. Logo, $I_\alpha(X/Y)$ é uma função convexa para baixo para $\frac{1}{2} \leq \alpha < 1$.

Lema 2.2: (Lema da Maximização)
(Ref.: Gallager, [1968], pág. 87)

Seja $f(P)$ uma função convexa para baixo de $P = (p(x_1),\ldots,p(x_n))$ na região

$$A^n = \{P : P = (p(x_1),\ldots,p(x_n)), p(x_i) \geq 0,\ i = 1,2,\ldots,n \ ,\ \sum_{i=1}^{n} p(x_i) = 1\}.$$

Suponha que as derivadas parciais $\frac{\partial f(P)}{\partial p(x_k)}$ são definidas e contínuas na região A^n, com a possível excessão que

$$\lim_{p(x_k) \to 0} \frac{\partial f(P)}{\partial p(x_k)} = +\infty$$

Então as condições

$$\frac{\partial f(P)}{\partial p(x_k)} \begin{cases} = \lambda & \text{se } p(x_k) > 0 \\ \leq \lambda & \text{se } p(x_k) = 0 \end{cases} \qquad (2.14)$$

$$\text{para } k = 1,2,\ldots,n$$

são necessárias e suficientes sob o vetor de probabilidade P para maximizar f na região A^n.

Proposição 2.1

$I_\alpha(X/Y)$ é uma função continuamente diferenciável nas variáveis $p(x_1),\ldots,p(x_n)$, para $\alpha > 0$. (Ref. Arimoto, [1975]).

Teorema 2.2:

Um conjunto de condições necessárias e suficientes sobre um vetor $P = (p(x_1),\ldots,p(x_n))$ que minimiza a função convexa para baixo continuamente diferenciável

$$F(\rho, P) = \sum_{j=1}^{m} \left[\sum_{i=1}^{n} p(x_i)\, p(y_j/x_i)^{1/1+\rho} \right]^{1+\rho} , \quad \rho \geq 0$$

(e maximiza $E_o(\rho, P) = -\log\,[F(\rho, P)]$)

na região $A^n = \{P : P = (p(x_1),\ldots,p(x_n)),\ p(x_i) \geq 0$,

$i = 1,2,\ldots,n$, $\sum_{i=1}^{n} p(x_i) = 1\}$

é

$$\sum_{j=1}^{m} p(y_j/x_i)^{1/1+\rho} \, t_j^{\rho} \begin{cases} = \sum_{j=1}^{m} t_j^{1+\rho} & \text{se } p(x_i) > 0 \\ \geq \sum_{j=1}^{m} t_j^{1+\rho} & \text{se } p(x_i) = 0 \end{cases} \qquad (2.15)$$

para $i = 1,2,\ldots,n$

onde

$$t_j = \sum_{i=1}^{n} p(x_i) \, p(y_j/x_i)^{1/1+\rho} .$$

A prova deste teorema é dada em Jelinek, [1968], p. 199.

Apresentaremos agora, o primeiro teorema de nosso trabalho que nos fornece a capacidade de ordem α de um canal discreto sem memória.

Teorema 2.3: (Capacidade de Ordem α)

Um conjunto de condições necessárias e suficientes sob um vetor de probabilidade $P = (p(x_1),\ldots,p(x_n))$ para alcançar a capacidade C_α de um canal discreto sem memória com entrada X e saída Y e matriz canal $[p(y_j/x_i)]$, $i = 1,2,\ldots,n$ e $j = 1,2,\ldots,m$. é

$$\frac{-\alpha}{1-\alpha} \log \sum_{j=1}^{m} p^{\alpha}(y_j/x_i) \left\{ \sum_{i=1}^{n} p(x_i) \, p^{\alpha}(y_j/x_i) \right\}^{\frac{\alpha-1}{\alpha}} \begin{cases} = C_\alpha & \text{se } p(x_i) > 0 \\ \leq C_\alpha & \text{se } p(x_i) = 0 \end{cases} \qquad (2.16)$$

para $1/2 \leq \alpha < 1$

<u>**Prova**</u>:

A função que nós queremos maximizar é da forma

$$I_\alpha(X/Y) = H_\alpha(X) - H_\alpha(X/Y)$$

$$= \frac{\alpha}{1-\alpha} \log \left\{ \sum_{i=1}^{n} p^\alpha(x_i) \right\}^{1/\alpha}$$

$$- \frac{\alpha}{1-\alpha} \log \left[\sum_{j=1}^{m} \left\{ \sum_{i=1}^{n} p^\alpha(x_i)\, p^\alpha(y_j/x_i) \right\}^{1/\alpha} \right],$$

$$\alpha \neq 1 \ , \ \alpha > 0$$

$$= \frac{-\alpha}{1-\alpha} \log \frac{\sum_{j=1}^{m} \left\{ \sum_{i=1}^{n} p^\alpha(x_i)\, p^\alpha(y_j/x_i) \right\}^{1/\alpha}}{\left\{ \sum_{i=1}^{n} p^\alpha(x_i) \right\}^{1/\alpha}}$$

$$= \frac{-\alpha}{1-\alpha} \log \left[\sum_{j=1}^{m} \left\{ \sum_{i=1}^{n} p'(x_i)\, p^\alpha(y_j/x_i) \right\}^{1/\alpha} \right] \qquad (2.17)$$

onde

$$p'(x_i) = \frac{p^\alpha(x_i)}{\sum_{k=1}^{n} p^\alpha(x_k)} \qquad \text{para} \quad i = 1,2,\ldots,n \ .$$

Portanto, como anterior, podemos escrever

$$I_\alpha(X/Y) = \frac{\alpha}{1-\alpha} \quad E_0(\alpha, P')$$

onde

$$E_0(\alpha, P') = -\log \left[\sum_{j=1}^{m} \left\{ \sum_{i=1}^{n} p'(x_i)\, p^\alpha(y_j/x_i) \right\}^{1/\alpha} \right]$$

e $\quad P' = (p'(x_1),\ldots,p'(x_n))$.

A função $E_o(\alpha, P')$ é da mesma forma que $E_o(\rho, P)$ como no teorema 1.2 com a convenção $\alpha = \dfrac{1}{1+\rho}$, $\frac{1}{2} \leqslant \alpha < +\infty$.

Então

$$I_\alpha(X/Y) = \frac{1}{\rho} E_o(\rho, P)$$

$$= \frac{-1}{\rho} \log \left[\sum_{j=1}^{m} \left\{ \sum_{i=1}^{n} p(x_i)\, p(y_j/x_i)^{1/1+\rho} \right\}^{1+\rho} \right]$$

$$= \frac{-1}{\rho} \log \left[F(\rho, P) \right]$$

onde

$$F(\rho, P) = \sum_{j=1}^{m} \left[\sum_{i=1}^{n} p(x_i)\, p(y_j/x_i)^{1/1+\rho} \right]^{1+\rho}$$

Logo

$$I_\alpha(X/Y) = \frac{\alpha}{1-\alpha} E_o(\alpha, P')$$

$$= \frac{-1}{\rho} \log \left[F(\rho, P) \right] = \frac{1}{\rho} E_o(\rho, P) \ , \ \rho > 0 \ .$$

A expressão (2.17) é definida para os valores reais não negativos das variáveis $p(x_i)$, $i = 1,2,\dots,n$. Com efeito a expressão (2.17) é a informação comunicada de ordem α por um único canal somente para $0 \leqslant p(x_i) \leqslant 1$, $i = 1,2,\dots,n$, $\sum_{i=1}^{n} p(x_i) = 1$ e $I_\alpha(X/Y)$ é uma função contínuamente diferenciável e convexa para baixo, para $\frac{1}{2} \leqslant \alpha < 1$ nas variáveis $p(x_1),\dots,p(x_n)$, quando $p(x_1),\dots,p(x_n)$ são maiores que zero. Observe que as quantidades $p(y_j/x_i)$ são fixadas para um dado canal.

Pelo teorema 2.2, temos

$$\sum_{j=1}^{m} p(y_j/x_i)^{1/1+\rho}\ t_j^{\rho}\ \begin{cases} = \sum_{j=1}^{m} t_j^{1+\rho} & \text{se} \quad p(x_i) > 0 \\ \geqslant \sum_{j=1}^{m} t_j^{1+\rho} & \text{se} \quad p(x_i) = 0 \end{cases}$$

para $i = 1,2,\ldots,n$

onde

$$t_j = \sum_{i=1}^{n} p(x_i)\ p(y_j/x_i)^{1/1+\rho}$$

ou

$$\frac{-1}{\rho} \log \sum_{j=1}^{m} p(y_j/x_i)^{1/1+\rho}\ t_j^{\rho}\ \begin{cases} = \frac{-1}{\rho} \log \sum_{j=1}^{m} t_j^{1+\rho} & \text{se } p(x_i) > 0 \\ \leqslant \frac{-1}{\rho} \log \sum_{j=1}^{m} t_j^{1+\rho} & \text{se } p(x_i) = 0 \end{cases} \qquad (2.18)$$

para $i = 1,2,\ldots,n.$

A expressão (2.18) pode ser escrita como.

$$\frac{-1}{\rho}\log\left[\sum_{j=1}^{m} p(y_j/x_i)^{1/1+\rho}\left\{\sum_{i=1}^{n} p(x_i)\, p(y_j/x_i)^{1/1+\rho}\right\}^{\rho}\right] \begin{cases} = \dfrac{-1}{\rho}\log \displaystyle\sum_{j=1}^{m}\sum_{i=1}^{n} p(x_i)\, p(y_j/x_i)^{1/1+\rho\;\;1+\rho} & \text{se } p(x_i) > 0 \\ \leqslant \dfrac{-1}{\rho}\log \displaystyle\sum_{j=1}^{m}\sum_{i=1}^{n} p(x_i)\, p(y_j/x_i)^{1/1+\rho\;\;1+\rho} & \text{se } p(x_i) = 0 \end{cases}$$

ou

$$\frac{-1}{\rho} \log \left[\sum_{j=1}^{m} p(y_j/x_i)^{1/1+\rho} \left\{ \sum_{i=1}^{n} p(x_i)\, p(y_j/x_i)^{1/1+\rho} \right\}^{\rho} \right] \begin{cases} = \frac{1}{\rho} E_o(\rho\,,\, P) & \text{se} \quad p(x_i) > 0 \\ \leqslant \frac{1}{\rho} E_o(\rho\,,\, P) & \text{se} \quad p(x_i) = 0 \end{cases}$$

Como $P = (p(x_1),\ldots,p(x_n))$ é um vetor que maximiza $E_o(\rho, P)$, logo

$$C_\alpha = \max_{\substack{p(x_i)\\ i=1,2,\ldots,n}} I_\alpha(X/Y) = \frac{1}{\rho} E_o(\rho, P) \ , \quad \text{onde}$$

$\alpha = \dfrac{1}{1+\rho}$, temos

$$\frac{-\alpha}{1-\alpha} \log \sum_{j=1}^{m} p^\alpha(y_j/x_i) \left\{ \sum_{i=1}^{n} p(x_i)\, p^\alpha(y_j/x_i) \right\}^{\frac{\alpha-1}{\alpha}} \begin{cases} = C_\alpha & \text{se } p(x_i) > 0 \\ \leq C_\alpha & \text{se } p(x_i) = 0 \end{cases}$$

onde

C_α é a capacidade de ordem α do canal.

Teorema de Codificação:

O teorema de codificação e seu contrário, para um canal discreto sem memória, podem ser expressos em termos de capacidade de ordem α, isto é, do teorema 1.2, capítulo 1, temos

$$P_e(\mathcal{C}) \leq \exp\left[-N \{-\rho R + \max_P E_o(\rho, P)\}\right] \quad , \ 0 < \rho \leq 1$$

Portanto,

$$- N \{-\rho R + \max_P E_o(\rho, P)\}$$

$$= -N \{-\rho R + \rho \max_P \frac{1}{\rho} E_o(\rho, P)\}$$

$$= -N \{-\rho R + \rho\, C_\alpha\} = -N \{\rho\, (C_\alpha - R)\}$$

$$= N \{- \frac{1-\alpha}{\alpha} (C_\alpha - R)\} = N \ \{\frac{\alpha - 1}{\alpha} (C_\alpha - R)\}$$

Logo

$$-N\ \{-\rho R + \max_{P} E_o(\rho,\ P)\} = N\ \{\frac{\alpha - 1}{\alpha}\ (C_\alpha - R)\} \tag{2.19}$$

Então

$$P_e(\mathcal{C}) \leq \exp\left[N\ \{\frac{\alpha - 1}{\alpha}\ (C_\alpha - R)\}\right]\ ,\ \frac{1}{2} \leq \alpha < 1 \tag{2.20}$$

Similarmente temos

$$P_e(\mathcal{C}) \geq 1 - \exp\left[N\ \{\frac{\alpha - 1}{\alpha}\ (C_\alpha - R)\}\right]\ ,\ 1 < \alpha < +\infty \tag{2.21}$$

Caso Particular:

$$\lim_{\alpha \to 1} C_\alpha = C \tag{2.22}$$

a qual é a capacidade do canal, conforme definida na secção 1.2, definição 1.2.5.

2.2 - Capacidade do Canal de Grau β

Nesta secção provamos o segundo teorema de nosso trabalho que dá a capacidade de grau β de um canal discreto sem memória.

Lema 2.3 - (Convexidade de $I^\beta(X/Y)$)

A função $I^\beta(X/Y)$ é convexa para baixo das probabilidades de entrada, para $\beta \geq 1$, isto é, se $a_1, a_2, \ldots, a_r$ são números não negativos tal que $\sum_{k=1}^{r} a_k = 1$ e definimos uma distribuição de probabilidade de entrada dada por

$$p_o(x) = \sum_{k=1}^{r} a_k\, p_k(x) \qquad (x = x_1,\ldots,x_n)$$

e também definimos uma distribuição de probabilidade de saída dada por

$$p_o(y) = \sum_{k=1}^{r} a_k\, p_k(y) \qquad (y = y_1,\ldots,y_m).$$

Então a informação $I_o^\beta(X/Y)$ correspondente a $p_o(x)$ e $p_o(y)$ satisfaz a

$$I_o^\beta(X/Y) \geq \sum_{k=1}^{r} a_k\, I_k^\beta(X/Y) \quad , \quad \beta \geq 1 \tag{2.23}$$

onde

$I^\beta(X/Y)$ é a informação de grau β quando a distribuição de probabilidade de entrada é

$$P\,\{X = x\} = p_k(x) \quad \text{para} \quad k = 1,2,\ldots,r \quad \text{e} \quad x = x_1,\ldots,x_n$$

e a distribuição de probabilidade de saída é

$$P\,\{Y = y\} = p_k(y) \quad \text{para} \quad k = 1,2,\ldots,r \quad y = y_1,\ldots,y_m .$$

<u>**Prova**</u>:

Seja

$$\Delta I^\beta = I_o^\beta(X/Y) - \sum_{k=1}^{r} a_k\, I_k^\beta(X/Y) \tag{2.24}$$

$$= H_o^\beta(X) - H_o^\beta(X/Y) - \sum_{k=1}^{r} a_k \left[H_k^\beta(X) - H_k^\beta(X/Y)\right]$$

$$= H_o^\beta(Y) - H_o^\beta(Y/X) - \sum_{k=1}^{r} a_k \left[H_k^\beta(Y) - H_k^\beta(Y/X)\right]$$

(por (1.50))

$$= (2^{1-\beta}-1)^{-1}\left[\sum_{j=1}^{m} p_o^\beta(y_j)-1\right] - (2^{1-\beta}-1)^{-1}\sum_{i=1}^{n} p_o^\beta(x_i)\left[\sum_{j=1}^{m} p_o^\beta(y_j/x_i)-1\right]$$

$$- \sum_{k=1}^{r} a_k \left\{(2^{1-\beta}-1)^{-1}\left[\sum_{j=1}^{m} p_k^\beta(y_j) - 1\right] - (2^{1-\beta}-1)^{-1}\sum_{i=1}^{n} p_k^\beta(x_i) \cdot \right.$$

$$\left. \cdot \left[\sum_{j=1}^{m} p_k^\beta(y_j/x_i) - 1\right]\right\}$$

$$= (2^{1-\beta}-1)^{-1}\left\{\left[\sum_{j=1}^{m} p_o^\beta(y_j) - \sum_{k=1}^{r} a_k \sum_{j=1}^{m} p_k^\beta(y_j)\right]\right.$$

$$- \sum_{i=1}^{n} p_o^\beta(x_i)\left(\sum_{j=1}^{m} p_o^\beta(y_j/x_i) - 1\right)$$

$$\left. - \left[\sum_{k=1}^{r} a_k \sum_{i=1}^{n} p_k^\beta(x_i)\left(\sum_{j=1}^{m} p_k^\beta(y_j/x_i) - 1\right)\right]\right\}$$

Como o canal tem as probabilidades condicionais $p(y_j/x_i)$, $i = 1,2,\ldots,n$ e $j = 1,2,\ldots,m$ fixadas, temos

$$p_o(y_j/x_i) = \sum_{k=1}^{r} a_k\, p_k(y_j/x_i)$$

$$= a_1\, p_1(y_j/x_i) + a_2\, p_2(y_j/x_i) + \ldots + a_r\, p_r(y_j/x_i)$$

$$= a_1 \, p(y_j/x_i) + a_2 \, p(y_j/x_i) + \ldots + a_r \, p(y_j/x_i)$$

$$= p(y_j/x_i) \, (a_1 + a_2 + \ldots + a_r) = p(y_j/x_i)$$

Portanto

$$p_o(y_j/x_i) = p(y_j/x_i) = p_k(y_j/x_i) \quad , \text{ para } k = 1,2,\ldots,r \tag{2.25}$$

Logo

$$\Delta I^{\beta} = (2^{1-\beta} - 1)^{-1} \left\{ \left[\sum_{j=1}^{m} p_o^{\beta}(y_j) - \sum_{k=1}^{r} a_k \sum_{j=1}^{m} p_k^{\beta}(y_j) \right] - \left[\sum_{i=1}^{n} p_o^{\beta}(x_i) - \sum_{k=1}^{r} a_k \sum_{i=1}^{n} p_k^{\beta}(x_i) \right] \cdot \left[\sum_{j=1}^{m} p^{\beta}(y_j/x_i) - 1 \right] \right\} \tag{2.26}$$

Sabemos que

$$p^{\beta}(y_j/x_i) \leqslant p(y_j/x_i) \quad , \; \beta \geqslant 1 \; , \; j = 1,2,\ldots,m \tag{2.27}$$

$$\sum_{j=1}^{m} p^{\beta}(y_j/x_i) \leqslant \sum_{j=1}^{m} p(y_j/x_i)$$

Portanto,

$$\sum_{j=1}^{m} p^{\beta}(y_j/x_i) - 1 \leqslant 0 \quad , \quad \beta \geqslant 1 \tag{2.28}$$

Como

$$\sum_{j=1}^{m} p_o^{\beta}(y_j) \geqslant \sum_{k=1}^{r} a_k \sum_{j=1}^{m} p_k^{\beta}(y_j) \quad , \; \beta \geqslant 1 \tag{2.29}$$

(Ref. Gallager, [1968], pág. 523)

e como

$$\sum_{i=1}^{n} p_o^\beta(x_i) \geq \sum_{k=1}^{r} a_k \sum_{i=1}^{n} p_k^\beta(x_i) \ , \ \beta \geq 1 \tag{2.30}$$

(Ref. Gallager, [1968], pág. 523)

(2.28) com (2.29) e (2.30) dá $\Delta I^\beta \geq 0$, isto é,

$$I_o^\beta(X/Y) \geq \sum_{k=1}^{r} a_k \ I_k^\beta(X/Y) \ , \quad \beta \geq 1 \ .$$

Logo, $I^\beta(X/Y)$ é uma função convexa para baixo de $P = (p(x_1),\ldots,p(x_n))$, para $\beta \geq 1$.

Proposição 2.2:

$I^\beta(X/Y)$ é uma função continuamente diferenciável nas variáveis $p(x_1),\ldots,p(x_n)$. (Ref. Daróczy, [1970]).

Apresentaremos agora o segundo teorema de nosso trabalho que nos fornece a capacidade de grau β de um canal discreto sem memória.

Teorema 2.4 - (**Capacidade de Grau β**)

Um conjunto de condições necessárias e suficientes sob o vetor de probabilidade $P = (p(x_1),\ldots,p(x_n))$ para alcançar a capacidade C^β, de um canal discreto sem memória com entrada X e saída Y e matriz canal $[p(y_j/x_i)]$, $i = 1,2,\ldots,n$ e $j = 1,2,\ldots,m$ é

$$(2^{1-\beta}-1)^{-1}\left[p^{\beta-1}(x_i) - 1 - \sum_{j=1}^{m} p^{\beta-1}(x_i)\, p^{\beta}(y_j/x_i) + \sum_{j=1}^{m} p^{\beta-1}(y_j)\, p(y_j/x_i)\right] \begin{cases} = C^{\beta} & \text{se } p(x_i) > 0 \\ \leqslant C^{\beta} & \text{se } p(x_i) = 0 \end{cases}$$

$$\text{para} \quad i = 1,2,\ldots,n \qquad \beta > 1 \qquad (2.31)$$

Prova:

A função que vamos maximizar é da forma

$$I^{\beta}(X/Y) = H^{\beta}(X) - H^{\beta}(X/Y)$$

$$= (2^{1-\beta}-1)^{-1}\left[\sum_{i=1}^{n} p^{\beta}(x_i) - 1\right] - \sum_{j=1}^{m} p^{\beta}(y_j)\, H^{\beta}(X/Y = y_j),$$

$$\beta \neq 1 \;,\; \beta > 0 \qquad (2.32)$$

$$= (2^{1-\beta}-1)^{-1}\left\{\left[\sum_{i=1}^{n} p^{\beta}(x_i) - 1\right] - \sum_{j=1}^{m} p^{\beta}(y_j)\left[\sum_{i=1}^{n} p^{\beta}(x_i/y_j) - 1\right]\right\}$$

$$= (2^{1-\beta}-1)^{-1}\left\{\sum_{i=1}^{n} p^{\beta}(x_i) - 1 - \sum_{i=1}^{n}\sum_{j=1}^{m} p^{\beta}(y_j)\, p^{\beta}(x_i/y_j) + \sum_{j=1}^{m} p^{\beta}(y_j)\right\}$$

Usando (1.6) e (1.7) temos que

$$I^{\beta}(X/Y) = (2^{1-\beta}-1)^{-1}\left\{\sum_{i=1}^{n} p^{\beta}(x_i) - 1 - \sum_{i=1}^{n}\sum_{j=1}^{m} p^{\beta}(x_i)\, p^{\beta}(y_j/x_i) + \sum_{j=1}^{m} p^{\beta}(y_j)\right\}$$

$$\beta \neq 1 \;,\; \beta > 0 \qquad (2.33)$$

A expressão (2.33) é definida para todos os valores reais não negativos das variáveis $p(x_1),\dots,p(x_n)$. A expressão (2.33) é a informação comunicada de grau β, por um único canal, somente para $0 \leq p(x_i) \leq 1$, $i = 1,2,\dots,n$, $\sum_{i=1}^{n} p(x_i) = 1$ e $I^{\beta}(X/Y)$ é uma função continuamente diferenciável nas variáveis $p(x_1),\dots,p(x_n)$. As quantidades $p(y_j/x_i)$ são fixadas para um dado canal. Vamos tentar maximizar (2.33) sujeito a condição $\sum_{i=1}^{n} p(x_i) = 1$. Usando o método dos multiplicadores de Lagrange, queremos maximizar

$$I^{\beta}(X/Y) + \lambda \sum_{i=1}^{n} p(x_i) \quad \text{tal que}$$

$$\frac{\partial}{\partial\, p(x_i)} \left[I^{\beta}(X/Y) + \lambda \sum_{i=1}^{n} p(x_i) \right] = 0 \qquad (2.34)$$

Portanto

$$\frac{\partial}{\partial\, p(x_i)} I^{\beta}(X/Y) = (2^{1-\beta} - 1)^{-1} \left\{ \beta\, p^{\beta-1}(x_i) - \sum_{j=1}^{m} \beta\, p^{\beta-1}(x_i)\, p^{\beta}(y_j/x_i) \right.$$

$$\left. + \sum_{j=1}^{m} \beta \left[\sum_{i=1}^{n} p(x_i)\, p(y_j/x_i) \right]^{\beta-1} .\; p(y_j/x_i) \right\} \qquad (2.35)$$

Então

$$\beta(2^{1-\beta} - 1)^{-1} \left\{ p^{\beta-1}(x_i) - \sum_{j=1}^{m} p^{\beta-1}(x_i)\, p^{\beta}(y_j/x_i) + \sum_{j=1}^{m} p^{\beta-1}(y_j)\, p(y_j/x_i) \right\}$$

$$+ \lambda = 0 \qquad (2.36)$$

Logo

$$\lambda = - \beta(2^{1-\beta} - 1)^{-1} \left\{ p^{\beta-1}(x_i) - \sum_{j=1}^{m} p^{\beta-1}(x_i)\, p^{\beta}(y_j/x_i) + \sum_{j=1}^{m} p^{\beta-1}(y_j) . p(y_j/x_i) \right\}$$

para $i = 1, 2, \ldots, n$

Pelo Lema 2.2, como $I^{\beta}(X/Y)$ é uma função convexa para baixo de $P = (p(x_1), \ldots, p(x_n))$, para $\beta \geq 1$ e as derivadas parciais de $I^{\beta}(X/Y)$ são contínuas, então as condições necessárias e suficientes em $P = (p(x_1), \ldots, p(x_n))$ para maximizar $I^{\beta}(X/Y)$ são

$$\frac{\partial}{\partial\, p(x_i)}\, I^{\beta}(X/Y) \begin{cases} = \lambda & \text{se} \quad p(x_i) > 0 \\ \leqslant \lambda & \text{se} \quad p(x_i) = 0 \end{cases} \tag{2.37}$$

para $i = 1,2,\ldots,n$

Escolhendo agora

$$C^{\beta} = \lambda\, \beta - (2^{1-\beta} - 1)^{-1}$$

temos

$$(2^{1-\beta}-1)^{-1}\left[p^{\beta-1}(x_i) - 1 - \sum_{j=1}^{m} p^{\beta-1}(x_i)\, p^{\beta}(y_j/x_i) + \sum_{j=1}^{m} p^{\beta-1}(y_j)\, p(y_j/x_i)\right] \begin{cases} = C^{\beta} & \text{se } p(x_i) > 0 \\ \leqslant C^{\beta} & \text{se } p(x_i) = 0 \end{cases} \tag{2.38}$$

para $i = 1,2,\ldots,n$

$\beta > 1$

Vamos mostrar, agora, que C^{β} é a capacidade do canal.

Multiplicando ambos os lados de (2.38) por $p(x_i)$ e tomando a soma sob todos i em que $p(x_i) > 0$, temos

$$(2^{1-\beta}-1)^{-1}\left[\sum_{i=1}^{n} p^{\beta}(x_i) - \sum_{i=1}^{n} p(x_i) - \sum_{j=1}^{m}\sum_{i=1}^{n} p^{\beta}(x_i)\, p^{\beta}(y_j/x_i) + \sum_{i=1}^{n} p(x_i) \sum_{j=1}^{m} p^{\beta-1}(y_j)\, p(y_j/x_i)\right] = \sum_{i=1}^{n} p(x_i)\, C^{\beta}$$

Isto implica que

$$I^{\beta}(X/Y) = C^{\beta} \quad , \quad \text{isto é} \; ,$$

$$\max_{\substack{p(x_i)\\ i=1,2,\ldots,n}} \quad I^{\beta}(X/Y) = C^{\beta} \quad . \tag{2.39}$$

Portanto, C^{β} é a capacidade do canal de grau β, $\beta > 1$, maximizando $I^{\beta}(X/Y)$ sobre o vetor probabilidade $(p(x_1),\ldots,p(x_n))$.

Observação: Se definimos

$$I^{\beta}(X/Y) = \sum_{i=1}^{n} p(x_i)\ I^{\beta}(X = x_i/Y) \tag{2.40}$$

no teorema 2.4, a expressão (2.31), pode ser escrita por

$$I^{\beta}(X = x_i/Y) \begin{cases} = C^{\beta} & \text{se} \quad p(x_i) > 0 \\ \\ \leqslant C^{\beta} & \text{se} \quad p(x_i) = 0 \end{cases} \tag{2.41}$$

para $i = 1,2,\ldots,n$, $\beta > 1$

onde

$I^{\beta}(X = x_i/Y)$ é a informação comunicada pelo canal de grau β, para entrada x_i a partir da saída Y, é dada por

$$I^{\beta}(X = x_i/Y) = (2^{1-\beta} - 1)^{-1} \left\{ p^{\beta-1}(x_i) - 1 - \sum_{j=1}^{m} p^{\beta-1}(x_i)\ p^{\beta}(y_j/x_i) + \sum_{j=1}^{m} p^{\beta-1}(y_j)\ .\ p(y_j/x_i) \right\} \tag{2.42}$$

para $i = 1,2,\ldots,n$, $\beta > 1$

Caso Particular:

$$\lim_{\beta \to 1} C^{\beta} = C \tag{2.43}$$

a qual é a capacidade do canal, conforme definida na secção 1.2, definição 1.2.5.

BIBLIOGRAFIA

ACZÉL, J. e Z. DARÓCZY (1975): On Measures of Information and Their Characterization - Academic Press, New York.

ARIMOTO, S. (1973): On the Converse to the Coding Theorem for Discrete Memoryless Channels - IEEE Trans. or Inform. Theory, IT - 19(3), pág. 357-359.

ARIMOTO, S. (1975): Information Measures and Capacity or Order for Discrete Memoryless Channels - In the Procedings of the Conference or Information Theory, Kesthely, Hungary, August, 1975.

ASH, R. (1965): Information Theory - Interscience Pub., New York.

DARÓCZY, Z. (1970): Generalized Information Measures - Information and Control, Vol. 16, pág. 33-51.

GALLAGER, R.G. (1965): A Simple Derivation of the Coding Theorem and Some Applications - IEEE Trans. on Inform. Theory, IT-11(1), pág. 3-18.

GALLAGER, R.G. (1968): Information Theory and Reliable Communication - J. Wiley, New York.

JELINEK, F. (1968): Probabilistic Information Theory - McGraw Hill, New York.

RÉNYI, A. (1961): On Measures of Entropy and Information - Vol. I, pág. 547-561, Proceedings of the 4 th Berkeley Symposium Mathematical Statistics and Probability, University of California Press, Berkeley, California.

SHARMA, B.D. e R. AUTAR (1973): An Inversion Theorem and Generalized Entropies for Continuous Distributions - SIAM J. Appl. Math. Vol. 25(2), pág. 125-132.

SHANNON, C.E. (1948): A Mathematical Theory of Communication - Bell Syst. Tech. J. Vol. 27, pág. 379-423, 623-656.

TANEJA, I.J. (1979): Some Contributions to Information Theory - I (A Survey) Lon Measures of Information - Journal of Comb., Inform. and Syst. Sciences, Vol. 4(4), pág. 259-280.