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 Pos-Graduação em Matemática da Universidade Federal de Santa Catarina

Prof. Italo José Dejter, Ph.D.
Coordenador

Banca Examinadora:

Prof. Inder Jeet Taneja, Ph.D.
Orientador

Prof. Rajamani Doraiswami, Ph.D.

Prof. Italo José Dejter, Ph.D.

CAPACIDADE-GAMA GENERALIZADA DE UM CANAL DISCRETO SEM MEMÓRIA

MARIA APARECIDA PION ABUABARA

MARÇO, 1981.

Aos meus pais, Oscar e Mercedes.

AGRADECIMENTOS

Ao Professor Inder Jeet Taneja, orientador deste trabalho, pela segurança, dedicação, incentivo, bem como pelos caminhos novos que nos mostrou.

À Universidade Federal de Santa Catarina, pelos meios fornecidos para a realização deste trabalho.

RESUMO

Apresentamos no Capítulo 1, o conceito de um sistema de comunicação e definimos a Entropia de Shannon, bem como a Entropia de Renyi, a Entropia de Daróczy e a Entropia Gama Generalizada.

Esta última entropia, tem um grande número de interessantes propriedades algébricas e analíticas similares às da entropia de Shannon, aqui discutidas, no Capítulo 2.

A capacidade-gama de um canal discreto sem memória é definida por meio a entropia-gama, no Capítulo 3. O teorema de maximização é provado; e, para canais simétricos calculamos a capacidade-gama. Exemplos são dados para a computação da capacidade-gama, de onde a capacidade de Shannon pode ser derivada fazendo o limite $\gamma \to 1$.

ÍNDICE

ABSTRACT

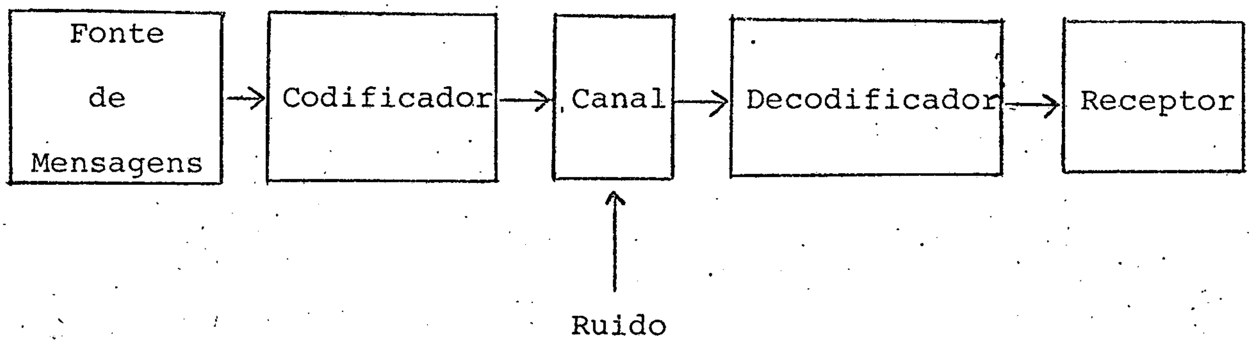
In this work, we study the algebraic and analytic properties of generalized gammma-entropy, which is similar to Shannon's entropy. As over central theorem, we present necessary and suficient conditions for an imput probability vector to achieve gamma-capacity on a DMC. For symmetric DMC, we calculate the gamma-capacity, from which the Shannon's capacity can be derived as the limiting case as $\gamma \to 1$.

CAPÍTULO 1

INTRODUÇÃO

§ 1.- Sistemas de Comunicação

A teoria da Comunicação trata originalmente com sistemas para transmitir informação ou dados de um ponto para outro. Um diagrama para visualizar o comportamento de tais sistemas é dado na figura abaixo



A *fonte de mensagens* é uma componente do sistema capaz de reproduzir mensagens. Poderá representar uma pessoa ou máquina. O *codificador* é o responsável pela mudança da forma das mensagens, isto é, transforma a linguagem da fonte para a linguagem do canal, mantendo inalterado o seu conteúdo. O *canal* é o meio através do qual a mensagem é propagada. Poderá representar, por exemplo, uma linha telefônica, um rádio transmisor de alta frequência. O canal está usualmente sujeito a vários tipos de perturbações ruidosas, as quais em uma linha telefônica, por exemplo, poderá ser na forma de cruzamento de outras linhas, ruído termal. O *decodificador* funciona de modo contrário ao do codificador, pois, após receber a mensagem que foi transmitida pelo canal, deve ser capaz de decifrá-la, de modo que a mesma seja inteligível pelo receptor. E, receptor é o ponto de destino da mensagem.

C. E. Shannon (1948) desenvolveu uma teoria matemática, chamada teoria da informação tratando com os aspectos fundamentais dos sistemas de comunicação. As características eminentes desta teoria são: primeira, uma grande ênfase sobre a teoria de probabilidade e, segunda, um interêsse especial com o codificador e o decodificador, ambos em termos de seus papéis funcionais e em termos da existência (ou não existência) de codificadores e decodificadores que atingem um nível de desempenho dado. Passados 32 anos, a teoria da informção tem sido feita mais precisa, extendida, e conduzida ao ponto onde é aplicada nos sistemas práticos de comunicação.

Como em qualquer teoria matamática, a teoria da informação trata somente com modelos matemáticos e não com fontes físicas e canais físicos. Pode-se pensar, portanto, que o caminho apropriado para começar o desenvolvimento da teoria seria com uma discussão de como construir modelos matemáticos apropriados para fontes e canais físicos. Isto, contudo, não é o caminho que as teorias são construídas, primeiramente porque a realidade física é raramente capaz de ser precisamente modelada por modelos tratáveis matemáticamente.

§ 2.- Entropia de Shannon

Seja $X = \{1,2,\ldots,m\}$ uma variável aleatoria discreta com distribuição de probabilidade $P = (p_1,p_2,\ldots,p_m)$, onde $p_i = p(i)$, $i=1,2,\ldots,m$.

O conjunto de todos os vetores probabilidades m-dimensional é denotado por Δ_m, isto é,

$$\Delta_m = \{P = (p_1,p_2,\dots,p_m);\ p_i \geqslant 0,\ \sum_{i=1}^{m} p_i = 1\}$$

A entropia de Shannon é definida por

$$H(X) = H(p_1,p_2,\dots,p_m) = -\sum_{i=1}^{m} p_i \log_2 p_i \ ,$$

onde $X = (p_1,p_2,\dots,p_m) \in \Delta_m$.

Correspondentemente, para uma variável aleatória bidimensional (X,Y) com distribuição de probabilidade conjunta $p(i,j)$, $i=1,2,\dots,m$; $j=1,2,\dots,n$, podemos definir a entropia conjunta por

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

DEFINIÇÕES :

1.1.- A probabilidade do evento $X = i$ no primeiro experimento indiferente ao segundo experimento é definido por

$$p(i) = \sum_{j=1}^{n} p(i,j)\ ,\quad i=1,2,\dots,m$$

1.2.- A probabilidade do evento $Y = j$ no segundo experimento indiferente ao primeiro experimento é definido por

$$q(j) = \sum_{i=1}^{m} p(i,j)\ ,\quad j=1,2,\dots,n$$

1.3.- Probabilidade condicional. A probabilidade de ocorrer o evento $X = i$ do primeiro experimento dado que o evento $Y = j$ do segundo experimento ocorrer é definida por

$$r(i/j) = \frac{p(i,j)}{q(j)}\ ,\quad i=1,2,\dots,m;\ j=1,2,\dots,n$$

1.4.- Probabilidade condicional. A probabilidade de ocorrer o evento $Y = j$ do segundo experimento dado que o evento $X = i$ do primeiro experimento ocorrer é definido por

$$r(j/i) = \frac{p(i,j)}{p(i)}, \quad i=1,2,\ldots,m; \; j=1,2,\ldots,n.$$

Observação.- Se X e Y são independentes, então

$r(i/j) = p(i)$, o que implica que $p(i) = \dfrac{p(i,j)}{q(j)}$, isto é,

$p(i,j) = p(i).q(j)$.

A incerteza condicional de Y dado que $X = i$ é dada por

$$H(Y/X=i) = - \sum_{j=1}^{n} r(j/i)\log_2 r(j/i), \quad i=1,2,\ldots,m.$$

Logo, a incerteza condicional de Y dado X é dada como a incerteza média de $H(Y/X=i)$ com pesos $p(i)$, $i=1,2,\ldots,m$, isto é,

$$H(Y/X) = \frac{\sum_{i=1}^{m} p(i)H(Y/X=i)}{\sum_{i=1}^{m} p(i)} =$$

$$= \sum_{i=1}^{m} p(i)H(Y/X=i) =$$

$$= -\sum_{i=1}^{m} \sum_{j=1}^{n} p(i)r(j/i)\log_2 r(j/i) =$$

$$= -\sum_{i=1}^{m} \sum_{j=1}^{n} p(i,j)\log_2 r(j/i).$$

Similarmente, temos que

$$H(X/Y) = -\sum_{i=1}^{m} \sum_{j=1}^{n} p(i,j) \log_2 r(i/j)$$

Através das definições acima, podemos obter os seguintes resultados:

i) $H(X,Y) = H(X) + H(Y/X) = H(Y) + H(X/Y)$

ii) $H(X,Y) \leq H(X) + H(Y)$

iii) $H(Y/X) \leq H(Y)$,

com a igualdade em (ii) e (iii) se, e somente se, X e Y são independentes.

§ 3 .- Informação Comunicada Pelo Canal

A "informação comunicada pelo canal sobre X a partir de Y" é definida por

$$I(X/Y) = H(X) - H(X/Y) =$$

$$= -\sum_{i=1}^{m} p(i) \log_2 p(i) + \sum_{i=1}^{m} \sum_{j=1}^{n} p(i,j) \log_2 r(i/j)$$

$$= -\sum_{i=1}^{m} \sum_{j=1}^{n} p(i,j) \log_2 p(i) + \sum_{i=1}^{m} \sum_{j=1}^{n} p(i,j) \log_2 r(i/j) =$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{n} p(i,j) \log_2 \frac{r(j/i)}{p(i)} =$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{n} p(i)r(j/i)\log_2 \frac{r(j/i)}{q(j)}.$$

Esta medida de informação é também chamada de Informação Mútua e satisfaz as seguintes propriedades:

i) Não negativa: $I(X/Y) \geqslant 0$.

ii) Simétrica: $I(X/Y) = I(Y/X)$.

iii) Concavidade: $I(X/Y)$ es uma função côncava sobre Δ_m.

§ 4.- Canais Discretos Sem Memória

Definição 1.5.- Um canal discreto sem memória com alfabeto de entrada $X = \{1,2,...,m\}$ e alfabeto de saida $Y = \{1,2,...,n\}$, com distribuição de entrada $(p(1),p(2),...,p(m))$ e distribuição de saida $(q(1),q(2),...,q(n))$ é caracterizado pela matriz mxn,

$$\left[r(j/i)\right] = \begin{bmatrix} r(1/1) & r(2/1) & \cdots & r(n/1) \\ r(1/2) & r(2/2) & \cdots & r(n/2) \\ \vdots & \vdots & & \vdots \\ r(1/m) & r(2/m) & \cdots & r(n/m) \end{bmatrix}.$$

onde $r(j/i)$, $i=1,2,...,m$; $j=1,2,...,n$ representa a probabilidade condicional da palavra código recebida $Y = j$ enquanto $X = i$ é transmitida, com $\sum_{j=1}^{n} r(j/i) = 1$, para $i=1,2,...,m$. A matriz $\left[r(j/i)\right]$ é chamada matriz-canal.

Definição 1.6.- A capacidade C de um canal discreto sem memória é definido como o valor máximo de $I(X/Y)$, onde a maximização é tomada sobre todos os $P \varepsilon \Delta_m$.

Desde que $I(X/Y)$ é uma função côncava sobre Δ_m, aplicando o teorema de Kuhn-Fucker, pode-se demonstrar o seguinte teorema, Gallager (1968) |6|.

Teorema 1.1.- Condições necessárias e suficientes para um vetor probabilidade de entrada $P = (p_1, p_2, \ldots, p_m)$ atingir a capacidade C de um canal discreto sem memória são que

$$\sum_{j=1}^{n} r(j/i) \log_2 \frac{r(j/i)}{\sum_{k=1}^{m} p_k(j/k)} = C, \text{ para todo } i \text{ com } p_i > 0$$

$$\leq C, \text{ para todo } i \text{ com } p_i = 0 .$$

§ 5.- Entropia de Rényi

Em 1961, Rényi |9| propôs uma generalização da entropia de Shannon. Propriedades conhecidas da entropia de Rényi e sua caracterização tem sido estudadas recentemente por Ben-Bassat e Raviv (1978), |3|.

A entropia de ordem α (ou entropia de Rényi) é definida por

$$_{\alpha}H(X) = \frac{1}{1-\alpha} \log_e \left(\sum_{i=1}^{m} p_i^{\alpha} \right), \quad \alpha > 0, \ \alpha \neq 1.$$

Quando $\alpha \to 1$, $_{\alpha}H$ torna-se a entropia de Shannon, a menos de uma constante. A entropia de ordem α pode ser reescrita na forma

$$_{\alpha}H(X) = \frac{\alpha}{1-\alpha} \log_e \left(\sum_{i=1}^{m} p_i^{\alpha} \right)^{1/\alpha}, \quad \alpha > 0, \ \alpha \neq 1.$$

A entropia condicional de ordem α de X dado Y é então definida como

$$_{\alpha}H(X/Y) = \frac{\alpha}{1-\alpha}\log_e \sum_{j=1}^{n} q_j\left(\sum_{i=1}^{m} r(i/j)^{\alpha}\right)^{1/\alpha} =$$

$$= \frac{\alpha}{1-\alpha}\log_e \sum_{j=1}^{n}\left(\sum_{i=1}^{m} p_i^{\alpha} r(j/i)^{\alpha}\right)^{1/\alpha},$$

onde $\alpha > 0$, $\alpha \neq 1$. Logo, a informação comunicada pelo canal de ordem α sobre X a partir de Y é dada por

$$_{\alpha}I(X/Y) = {_{\alpha}H(X)} - {_{\alpha}H(X/Y)} =$$

$$= \frac{-\alpha}{1-\alpha}\log_e \frac{\sum_{j=1}^{n}\left(\sum_{i=1}^{m} p_i^{\alpha} r(j/i)^{\alpha}\right)^{1/\alpha}}{\left(\sum_{i=1}^{m} p_i^{\alpha}\right)^{1/\alpha}},$$

onde $\alpha > 0$, $\alpha \neq 1$. E a capacidade de ordem α, de um canal discreto sem memória é definida como sendo o máximo de $_{\alpha}I(X/Y)$, tomado sobre todas as distribuições de probabilidade de entrada $P = (p_1, p_2, \ldots, p_m)$.

§ 6.- Entropia de Daróczy

Daróczy (1970) |5| introduziu o conceito de funções informação de grau β, e por meio dessas funções definiu as entropias de grau β por

$$H^{\beta}(X) = (2^{1-\beta} - 1)^{-1} \left(\sum_{i=1}^{m} p_i^{\beta} - 1\right), \quad \beta > 0, \ \beta \neq 1.$$

A entropia de Shannon $H(X) = -\sum_{i=1}^{m} p_i \log_2 p_i$ é o limite da função $H^{\beta}(X)$, quando $\beta \to 1$.

Esta entropia tem um grande número de interessantes propriedades algébricas e analíticas similares às da entropia de Shannon. Ver Daróczy (1970), |5|.

A entropia condicional de grau β de X dado Y é dada por

$$H_{\beta}(X/Y) = (2^{1-\beta} - 1)^{-1} \sum_{i=1}^{m} p_i^{\beta} \left(\sum_{j=1}^{n} r(j/i)^{\beta} - 1\right),$$

onde $\beta > 0$, $\beta \neq 1$. E, então, a informação comunicada pelo canal de grau β sobre X a partir de Y é definida por

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

Finalmente, definimos a capacidade de grau β de um canal discreto sem memória caracterizado pela matriz $R = [r(j/i)]$, $i=1,2,\ldots,m$; $j=1,2,\ldots,n$, como sendo

$$C_{\beta} = \max_{P \in \Delta_m} I^{\beta}(X/Y).$$

Ademais, estas entropias de ordem α e de grau β satisfazem muitas outras propriedades. Ver Taneja (1979), |11|.

A generalização do teorema 1.1 para as entropias de ordem α e de grau β foi feita por Guerra (1980), |7|.

§ 7.- Entropia-Gama Generalizada

A entropia-γ para uma distribuição de probabilidade $(p_1, p_2, \ldots, p_m)$ é definida por

$$H_\gamma(p_1, p_2, \ldots, p_m) = \frac{1}{\gamma - 1}\left[\left(\sum_{i=1}^{m} p_i^{1/\gamma}\right)^{\gamma} - 1\right], \quad \gamma > 0, \; \gamma \neq 1.$$

No caso limite $\gamma \to 1$, a entropia-γ reduz-se à entropia de Shannon, a menos de uma constante.

Estas entropias possuem interessantes propriedades algébricas e analíticas similares às da entropia de Shannon, as quais serão estudadas no Capítulo 2. Também, no Capítulo 2, a entropia-γ condicional é definida e suas propriedades analizadas

Agora veremos uma maneira diferente de definir entropias-gamas generalizadas. Ver também |1| ou |9|.

<u>Definição 1.7</u>.- Seja $f(x)$ uma função escalar com valores reais, definida e não-negativa em $(0,1]$, com derivadas contínuas em $(0,1]$ e tal que $f(1) = 0$. Definimos funções entropia generalizadas como

$$H_f(p_1, p_2, \ldots, p_m) = \inf \sum_{i=1}^{m} p_i f(\tilde{p}_i) \; ,$$

onde a operação inf é tomada sobre todas as distribuições de probabilidade $\tilde{P} = (\tilde{p}_1, \tilde{p}_2, \ldots, \tilde{p}_m) \in \Delta_m$.

Consideremos as seguintes funções escalares

$$f^{\gamma}(x) = \frac{1 - x^{1-\gamma}}{1 - \gamma}, \quad \gamma \neq 1, \ \gamma > 0,$$

$$f^{1}(x) = -\log_e x = \lim_{\gamma \to 1} f^{\gamma}(x).$$

Então, segue de um cálculo simples que

$$H_{\gamma}(P) = \inf_{\tilde{P}} \sum_{i=1}^{m} p_i f^{\gamma}(\tilde{p}_i) =$$

$$= \begin{cases} 1 - \max_i p_i, & \text{para } \gamma = 0 \\ \dfrac{1 - \left(\sum_{i=1}^{m} p_i^{1/\gamma}\right)^{\gamma}}{1 - \gamma}, & \text{para } \gamma \neq 1 \text{ e } \gamma > 0 \\ \sum_{i=1}^{m} -p_i \log_e p_i, & \text{para } \gamma = 1 \end{cases}$$

§ 8.- Relação Entre a Entropia-Gama e a Entropia de Ordem α.

A entropia-gama $H_{\gamma}(P)$ relaciona-se com a entropia de Rényi (ou de ordem α), definida no § 5 como

$$_{\alpha}H(P) = \frac{1}{1-\alpha} \log_e \left(\sum_{i=1}^{m} p_i^{\alpha} \right).$$

Tomando $\gamma = \alpha^{-1}$ e observando a desigualdade $\log_e x \leqslant x - 1$, temos as seguintes relações:

$$H_{\gamma}(P) \leqslant {}_{\alpha}H(P) \leqslant H(P), \text{ para } 0 \leqslant \gamma = \alpha^{-1} < 1,$$

$$H_{\gamma}(P) = {}_{\alpha}H(P) = H(P), \text{ para } \gamma = \alpha = 1,$$

$$H_{\gamma}(P) \geqslant {}_{\alpha}H(P) \geqslant H(P), \text{ para } \gamma = \alpha^{-1} > 1,$$

onde $H(P) = \sum_{i=1}^{m} -p_i \log_e p_i$.

CAPÍTULO 2

Propriedades da Entropia-Gama Generalizada

A entropia-γ, introduzida no § 7, Cap. 1, é uma função real definida sobre Δ_m ($m=2,3,\ldots$), isto é,

$$H_\gamma : \Delta_m \longrightarrow \mathbb{R}, \quad m=2,3,\ldots,$$

onde R é o conjunto dos números reais. Nos teoremas seguintes resumimos as propriedades da entropia-γ.

Teorema 2.1.- As entropias $H_\gamma : \Delta_m \longrightarrow \mathbb{R}$, $m=2,3,\ldots$; $\gamma > 0$, $\gamma \neq 1$, satisfazem as seguintes propriedades:

1ª) Não-Negativa: $H_\gamma(p_1,p_2,\ldots,p_m) \geqslant 0$;

2ª) Simétrica: $H_\gamma(p_1,p_2,\ldots,p_m)$ é uma função simétrica de suas variáveis;

3ª) Decisiva: $H_\gamma(0,1) = H_\gamma(1,0) = 0$;

4ª) Expansível: $H_\gamma(p_1,p_2,\ldots,p_m) = H_\gamma(p_1,p_2,\ldots,p_m,0)$;

5ª) Contínua: $H_\gamma(p_1,p_2,\ldots,p_m)$ é uma função contínua.

Estas propriedades são consequências imediatas da definição de entropia-γ.

Teorema 2.2.- Se $P = (p_1,p_2,\ldots,p_m)$ é uma distribuição de probabilidade, então

(i) $\lim_{\gamma\to 1} H_\gamma(P) = -\sum_{i=1}^{m} p_i \log_e p_i$, e

(ii) $\lim_{\gamma\to 1} H_\gamma(P) = 1 - \max_{1<i<m} p_i$.

<u>Prova</u>:

$$\text{(i)}\ \lim_{\gamma\to 1} H_\gamma(P) = \lim_{\gamma\to 1} \frac{\left(\sum_{i=1}^{m} p_i^{1/\gamma}\right)^{\gamma} - 1}{\gamma - 1} =$$

$$= \lim_{\gamma\to 1} \frac{\frac{d}{d\gamma}\left(\sum_{i=1}^{m} p_i^{1/\gamma}\right)^{\gamma} - 1}{\frac{d}{d\gamma}(\gamma - 1)} =$$

$$= \lim_{\gamma\to 1} \left(\sum_{i=1}^{m} p_i^{1/\gamma}\right)^{\gamma} \left[\log_e\left(\sum_{i=1}^{m} p_i^{1/\gamma}\right) - \frac{1}{\gamma}\,\frac{\sum_{i=1}^{m} p_i^{1/\gamma}\log_e p_i}{\left(\sum_{i=1}^{m} p_i^{1/\gamma}\right)}\right] =$$

$$= -\sum_{i=1}^{m} p_i \log_e p_i .$$

(ii) Seja $p_\ell \geq p_i$, $i=1,2,\ldots,m$, e suponhamos que p_i coincida com p_ℓ para q valores de i. Logo,

$$\lim_{\gamma\to 0} \frac{\left(\sum_{i=1}^{m} p_i^{1/\gamma}\right)^{\gamma} - 1}{\gamma - 1} = \lim_{\gamma\to 0} \frac{p_\ell\left(\sum_{i=1}^{m} \left(\frac{p_i}{p_\ell}\right)^{1/\gamma}\right)^{\gamma} - 1}{\gamma - 1} = \frac{p_\ell - 1}{-1} =$$

$$= 1 - p_\ell = 1 - \max_{1\leqslant i\leqslant m} p_i .$$

<u>Teorema 2.3</u>.- Para todo $(p_1,p_2,\ldots,p_m)\in\Delta_m$ $(m=2,3,\ldots)$, temos que

$$0 \leqslant H_\gamma(p_1,p_2,\ldots,p_m) \leqslant H_\gamma\left(\frac{1}{m},\frac{1}{m},\ldots,\frac{1}{m}\right) = \frac{1}{\gamma - 1}\,(m^{\gamma-1} - 1),$$

isto é, a entropia-γ é máxima quando todas as probabilidades são iguais.

<u>Prova</u>: Desde que a função $\omega(x) = x^{1/\gamma}$, $x \geqslant 0$, é côncava se $\gamma > 1$, temos que

$$\sum_{i=1}^{m} \frac{1}{m}\omega(p_i) \leqslant \omega\left(\sum_{i=1}^{m} \frac{1}{m}\, p_i\right),$$

isto é,

$$\sum_{i=1}^{m} \frac{1}{m}\, p_i^{1/\gamma} \leqslant \left(\sum_{i=1}^{m} \frac{1}{m}\, p_i\right)^{1/\gamma} = \left(\frac{1}{m}\right)^{1/\gamma},$$

já que $\sum_{i=1}^{m} p_i = 1$. Portanto, temos que

$$\sum_{i=1}^{m} p_i^{1/\gamma} \leqslant m\left(\frac{1}{m}\right)^{1/\gamma}$$

donde temos que

$$\left(\sum_{i=1}^{m} p_i^{1/\gamma}\right)^{\gamma} \leqslant \left[m\left(\frac{1}{m}\right)^{1/\gamma}\right]^{\gamma} = m^{\gamma-1} ,$$

logo,

$$\left(\sum_{i=1}^{m} p_i^{1/\gamma}\right)^{\gamma} - 1 \leqslant m^{\gamma-1} - 1,$$

o que implica que

$$\frac{1}{\gamma - 1}\left[\left(\sum_{i=1} p_i^{1/\gamma}\right)^{\gamma} - 1\right] \leqslant \frac{1}{\gamma - 1}(m^{\gamma-1} - 1),$$

ou seja,

$$H_{\gamma}(p_1,p_2,\ldots,p_m) \leqslant H_{\gamma}\left(\frac{1}{m},\frac{1}{m},\ldots,\frac{1}{m}\right) = \frac{1}{\gamma - 1}(m^{\gamma-1} - 1).$$

Analogamente, se $0 < \gamma < 1$, temos o resultado, desde que $\omega(x) = x^{1/\gamma}$, $x \geqslant 0$, é convexa e $\dfrac{1}{\gamma - 1} < 0$, provando o teorema.

<u>Observação 1</u>.- É interessante observar que a função

$$\phi_{\gamma}(m) = H_{\gamma}\left(\frac{1}{m},\frac{1}{m},\ldots,\frac{1}{m}\right),$$

é monotônica crescente, isto é, $\phi_\gamma(m) \leqslant \phi_\gamma(m+1)$. Isto é uma consequência dos teorema 2.1 e 2.3:

$$\phi_\gamma(m) = H_\gamma(\frac{1}{m},\frac{1}{m},\ldots,\frac{1}{m}) =$$

$$= H_\gamma(\frac{1}{m},\frac{1}{m},\ldots,\frac{1}{m},0) \leqslant$$

$$\leqslant H_\gamma(\frac{1}{m+1}, \frac{1}{m+1}, \ldots, \frac{1}{m+1}) =$$

$$= \phi_\gamma(m+1).$$

Observação 2.- Se $0 < \gamma < 1$, então, pela monotonicidade de ϕ_γ, temos que: Para todo $(p_1,p_2,\ldots,p_m) \varepsilon \Delta_m$ $(m=2,3,\ldots)$,

$$H_\gamma(p_1,p_2,\ldots,p_m) \leqslant \phi_\gamma(m) \leqslant$$

$$\leqslant \lim_{m\to\infty} \frac{1}{\gamma-1}(m^{\gamma-1} - 1) =$$

$$= \frac{1}{\gamma-1}(-1) =$$

$$= \frac{1}{\gamma-1}.$$

No caso $\gamma > 1$, esta afirmação não é verdadeira, pois $\lim_{m\to\infty}\phi_\gamma(m) = +\infty$.

Teorema 2.4.- A função $H_\gamma(p_1,p_2,\ldots,p_m)$, $\gamma > 0$, $\gamma \neq 1$, é côncava.

Prova: A função $F(P) = F(p_1,p_2,\ldots,p_m) = (\sum_{i=1}^{m} p_i^{1/\gamma})^\gamma$ é côncava para $\gamma > 1$ e convexa para $0 < \gamma < 1$. De fato, se $0 \leqslant \lambda \leqslant 1$ e,

$P = (p_1,p_2,\ldots,p_m)$ e $Q = (q_1,q_2,\ldots,q_m)$ são duas distribuições de probabilidade, então

$$F(\lambda P+(1-\lambda)Q) \geq [\sum_{i=1}^{m}(\lambda p_i+(1-\lambda)q_i)^{1/\gamma}]^{\gamma} \geq$$

$$\geq [\sum_{i=1}^{m}(\lambda p_i)^{1/\gamma}]^{\gamma} + [\sum_{i=1}^{m}(1-\lambda)^{1/\gamma}q_i^{1/\gamma}]^{\gamma} =$$

$$= \lambda(\sum_{i=1}^{m}p_i^{1/\gamma})^{\gamma} + (1-\lambda)(\sum_{i=1}^{m}q_i^{1/\gamma})^{\gamma} =$$

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

se $\gamma > 1$, onde a desigualdade acima é obtida usando a desigualdade de Minkowski |6|.

Se $0 < \gamma < 1$, a desigualdade é contrária, provando assim a afirmação.

Consequentemente, a função

$$\tilde{H}(p_1,p_2,\ldots,p_m) = \frac{1}{\gamma-1}[(\sum_{i=1}^{m}p_i^{1/\gamma})^{\gamma} - 1],$$

é côncava, para $\gamma > 0$, $\gamma \neq 1$.

Seja $X = \{1,2,\ldots,m\}$ uma variável aleatória discreta. Podemos escrever a entropia-γ para a variável aleatória X como

$$H_{\gamma}(X) = H_{\gamma}(p_1,p_2,\ldots,p_m),$$

onde $p_i = p(i)$, $i=1,2,\ldots,m$.

Correspondentemente, para uma variável aleatória discreta bidemensional (X,Y), temos a entropia-γ conjunta

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

onde p(i,j), i=1,2,...,m; j=1,2,...,n, são as distribuições de probabilidade conjunta.

A <u>entropia-γ condicional</u> de uma variável aleatória X com respeito a Y = j, é definida por

$$H_\gamma(X/Y=j) = \frac{1}{\gamma-1}\left[\left(\sum_{i=1}^{m} r(i/j)^{1/\gamma}\right)^\gamma - 1\right],\ j=1,2,...,n,$$

onde r(i/j), i=1,2,...,m; j=1,2,...,n são as probabilidades condicionais. E, correspondentemente, a <u>entropia-γ condicional</u> de uma variável aleatória X com respeito a Y é dada por

$$H_\gamma(X/Y) = \sum_{j=1}^{n} q(j) H_\gamma(X/Y=j),$$

onde q(j) é a probabilidade quando Y = j, j=1,2,...,n. Assim,

$$H_\gamma(X/Y) = \frac{1}{\gamma-1}\left\{\sum_{j=1}^{n} q(j)\left[\left(\sum_{i=1}^{m} r(i/j)^{1/\gamma}\right)^\gamma - 1\right]\right\} =$$

$$= \frac{1}{\gamma-1}\left\{\sum_{j=1}^{n} q(j)\left(\sum_{i=1}^{m} r(i/j)^{1/\gamma}\right)^\gamma - \sum_{j=1}^{n} q(j)\right\} =$$

$$= \frac{1}{\gamma-1}\{\sum_{j=1}^{n}(\sum_{i=1}^{m} q(j)^{1/\gamma} r(i/j)^{1/\gamma})^{\gamma} - 1\} =$$

$$= \frac{1}{\gamma-1}\{\sum_{j=1}^{n}(\sum_{i=1}^{m} p_i^{1/\gamma} r(j/i)^{1/\gamma})^{\gamma} - 1\} ,$$

devido ao fato de $r(i/j)q(j) = r(j/i)p(i) = p(i,j)$.

Teorema 2.5.- Se X e Y são variáveis aleatórias discretas independentes, então

$$H_\gamma(X,Y) = H_\gamma(X) + H_\gamma(Y) + (\gamma-1)H_\gamma(X).H_\gamma(Y).$$

Prova: Desde que X e Y são variáveis aleatórias discretas independentes, então temos que

$$H_\gamma(X/Y) = \frac{1}{\gamma-1}[(\sum_{i=1}^{m}\sum_{j=1}^{n} p(i,j)^{1/\gamma})^{\gamma} - 1] =$$

$$= \frac{1}{\gamma-1}[(\sum_{i=1}^{m}\sum_{j=1}^{n} p(i)^{1/\gamma} q(j)^{1/\gamma})^{\gamma} - 1] =$$

$$= \frac{1}{\gamma-1}[(\sum_{i=1}^{m} p(i)^{1/\gamma})^{\gamma}(\sum_{j=1}^{n} q(j)^{1/\gamma})^{\gamma} - 1] =$$

$$= \frac{1}{\gamma-1}[\{(\sum_{i=1}^{m} p(i)^{1/\gamma})^{\gamma}-1\}\{\sum_{j=1}^{n} q(j)^{1/\gamma})^{\gamma}-1\}+(\sum_{i=1}^{m} p(i)^{1/\gamma})^{\gamma}+(\sum_{j=1}^{n} q(j)^{1/\gamma})^{\gamma}-2$$

$$= \frac{1}{\gamma-1}\Big[\{(\sum_{i=1}^{m} p(i)^{1/\gamma})^{\gamma} - 1\}\{(\sum_{j=1}^{n} q(j)^{1/\gamma})^{\gamma} - 1\} +$$

$$+ \{(\sum_{i=1}^{m} p(i)^{1/\gamma})^{\gamma} - 1\} + \{(\sum_{j=1}^{n} q(j)^{1/\gamma})^{\gamma} - 1\}\Big] =$$

$$= H_{\gamma}(X) + H_{\gamma}(Y) + (\gamma-1)H_{\gamma}(X)H_{\gamma}(Y),$$

o que prova o teorema.

No caso limite $\gamma\to 1$, temos a conhecida propriedade de aditividade da entropia de Shannon.

<u>Teorema 2.6</u>.- Se X e Y são variáveis aleatórias discretas independentes, então

$$H_{\gamma}(X/Y) = H_{\gamma}(X).$$

<u>Prova</u>:

$$H_{\gamma}(X/Y=j) = \frac{1}{\gamma-1}\Big[(\sum_{i=1}^{m} r(i/j)^{1/\gamma})^{\gamma} - 1\Big] =$$

$$= \frac{1}{\gamma-1}\Big[(\sum_{i=1}^{m} p(i)^{1/\gamma})^{\gamma} - 1\Big] =$$

$$= H_{\gamma}(X).$$

Portanto,

$$H_{\gamma}(X/Y) = \sum_{j=1}^{n} q(j)H_{\gamma}(X/Y=j) = \sum_{j=1}^{n} q(j)H_{\gamma}(X) = H_{\gamma}(X).$$

<u>Lema 2.1</u>.- Se x_1, x_2, y_1, y_2 são números reais tais que $0 < x_i < y_i$, $i=1,2$, então

$$(y_1 + y_2)^\gamma - (y_1^\gamma + y_2^\gamma) \geq (x_1 + x_2)^\gamma - (x_1^\gamma + x_2^\gamma),$$

se $\gamma > 1$, e

$$(y_1 + y_2)^\gamma - (y_1^\gamma + y_2^\gamma) \leq (x_1 + x_2)^\gamma - (x_1^\gamma + x_2^\gamma),$$

se $0 < \gamma < 1$.

<u>Prova</u>: Suponhamos $\gamma > 1$. Seja

$$f(x) = (x + y_2)^\gamma - x^\gamma \text{ , para } x \geq 0.$$

Então,

$$f'(x) = \gamma(x + y_2)^{\gamma-1} - \gamma x^{\gamma-1} = \gamma\left[(x + y_2)^{\gamma-1} - x^{\gamma-1}\right] \geq 0,$$

o que implica que f é crescente. Assim, $f(y_1) \geq f(x_1)$, isto é, $(y_1 + y_2)^\gamma - y_1^\gamma \geq (x_1 + y_2)^\gamma - x_1^\gamma$, o que implica que

$$(y_1 + y_2)^\gamma - (y_1^\gamma + y_2^\gamma) \geq (x_1 + y_2)^\gamma - (x_1^\gamma + y_2^\gamma) \tag{1}$$

Similarmente, seja

$$g(x) = (x_1 + x)^\gamma - x^\gamma, \text{ para } x \geq 0.$$

Então,

$$g'(x) = \gamma\left[(x_1 + x)^{\gamma-1} - x^{\gamma-1}\right] \geq 0,$$

o que implica que g é crescente. Assim, $g(y_2) \geq g(x_2)$, ou seja

$(x_1 + y_2)^\gamma - y_2^\gamma \geq (x_1 + x_2)^\gamma - x_2^\gamma$, o que implica que

$$(x_1 + y_2)^\gamma - (x_1^\gamma + y_2^\gamma) \geq (x_1 + x_2)^\gamma - (x_1^\gamma + x_2^\gamma). \quad (2)$$

Das desigualdades (1) e (2), obtemos que

$$(y_1 + y_2)^\gamma - (y_1^\gamma + y_2^\gamma) \geq (x_1 + x_2)^\gamma - (x_1^\gamma + x_2^\gamma).$$

Se $0 < \gamma < 1$, a demonstração é análoga, com as desigualdades contrárias. Isso prova o lema.

Usando o método da indução matemática, obtemos as seguintes duas desigualdades:

$$\left(\sum_{j=1}^{n} y_j\right)^\gamma - \sum_{j=1}^{n} y_j^\gamma \geq \left(\sum_{j=1}^{n} x_j\right)^\gamma - \sum_{j=1}^{n} x_j^\gamma ,$$

se $\gamma > 1$, e

$$\left(\sum_{j=1}^{n} y_j\right)^\gamma - \sum_{j=1}^{n} y_j^\gamma \leq \left(\sum_{j=1}^{n} x_j\right)^\gamma - \sum_{j=1}^{n} x_j^\gamma ,$$

se $0 < x_j < y_j$, $j=1,2,\ldots,n$.

<u>Teorema 2.7</u>.- Se X e Y são variáveis aleatórias discretas então

$$H_\gamma(X,Y) \geq H_\gamma(Y) + H_\gamma(X/Y) ,$$

se $\gamma > 1$, e

$$H_\gamma(X,Y) \leq H_\gamma(Y) + H_\gamma(X/Y),$$

se $0 < \gamma < 1$.

<u>Prova:</u> Caso I: $\gamma > 1$. Temos que

$$\sum_{i=1}^{m} r(j/i)^{1/\gamma} p(i)^{1/\gamma} \geqslant \left(\sum_{i=1}^{m} r(j/i)p(i)\right)^{1/\gamma} =$$

$$= \left(\sum_{i=1}^{m} p(i,j)\right)^{1/\gamma} =$$

$$= q(j)^{1/\gamma},$$

para $j=1,2,\ldots,n$. Usando o lema 2.1, obtemos que

$$\left(\sum_{j=1}^{n} \sum_{i=1}^{m} r(j/i)^{1/\gamma} p(i)\right)^{\gamma} - \sum_{j=1}^{n} \left(\sum_{i=1}^{m} r(j/i)^{1/\gamma} p(i)^{1/\gamma}\right)^{\gamma} \geqslant$$

$$\geqslant \left(\sum_{j=1}^{n} q(j)^{1/\gamma}\right)^{\gamma} - \sum_{j=1}^{n} q(j) ,$$

donde temos que

$$\sum_{j=1}^{n} \left(\sum_{i=1}^{m} r(j/i)^{1/\gamma} p(i)^{1/\gamma}\right)^{\gamma} - 1 + \left(\sum_{j=1}^{n} q(j)^{1/\gamma}\right)^{\gamma} \leqslant$$

$$\leqslant \left(\sum_{j=1}^{n} \sum_{i=1}^{m} p(i,j)^{1/\gamma}\right)^{\gamma},$$

logo

$$\sum_{j=1}^{n} \left(\sum_{i=1}^{m} r(j/i)^{1/\gamma} p(i)^{1/\gamma}\right)^{\gamma} - 1 + \left(\sum_{j=1}^{n} q(j)^{1/\gamma}\right)^{\gamma} - 1 \leqslant$$

$$\leqslant \left(\sum_{j=1}^{n} \sum_{i=1}^{m} p(i,j)^{1/\gamma}\right)^{\gamma} - 1 ,$$

e, portanto

$$\frac{1}{\gamma-1}\Big[\sum_{j=1}^{n}\Big(\sum_{i=1}^{m} r(j/i)p(i)^{1/\gamma}\Big)^{\gamma} - 1\Big] + \frac{1}{\gamma-1}\Big[\Big(\sum_{j=1}^{n} q(j)^{1/\gamma}\Big)^{\gamma} - 1\Big] \leqslant$$

$$\leqslant \frac{1}{\gamma-1}\Big[\Big(\sum_{j=1}^{n}\sum_{i=1}^{m} p(i,j)^{1/\gamma}\Big)^{\gamma} - 1\Big] ,$$

isto é,

$$H_\gamma(X/Y) + H_\gamma(Y) \leqslant H_\gamma(X,Y).$$

<u>Caso II</u> : $0 < \gamma < 1$. Neste caso, temos que

$$\sum_{i=1}^{m} r(j/i)^{1/\gamma}p(i)^{1/\gamma} \leqslant \Big(\sum_{i=1}^{m} r(j/i)p(i)\Big)^{1/\gamma} =$$

$$= \Big(\sum_{i=1}^{m} p(i,j)\Big)^{1/\gamma} =$$

$$= q(j)^{1/\gamma},$$

para $j=1,2,\ldots,n$. Novamente, usando o lema 2.1, obtemos que

$$\Big(\sum_{j=1}^{n} q(j)^{1/\gamma}\Big)^{\gamma} - \sum_{j=1}^{n} q(j) \leqslant$$

$$\leqslant \Big(\sum_{j=1}^{n}\sum_{i=1}^{m} r(j/i)^{1/\gamma}p(i)^{1/\gamma}\Big)^{\gamma} - \sum_{j=1}^{n}\Big(\sum_{i=1}^{m} r(j/i)^{1/\gamma}p(i)^{1/\gamma}\Big)^{\gamma},$$

donde temos que

$$\Big(\sum_{j=1}^{n} q(j)^{1/\gamma}\Big)^{\gamma} - 1 + \sum_{j=1}^{n}\Big(\sum_{i=1}^{m} r(j/i)^{1/\gamma}p(i)^{1/\gamma}\Big)^{\gamma} \leqslant$$

$$\leqslant \left(\sum_{j=1}^{n} \sum_{i=1}^{m} p(i,j)^{1/\gamma} \right),$$

logo,

$$\left(\sum_{j=1}^{n} q(j)^{1/\gamma} \right)^{\gamma} - 1 + \sum_{j=1}^{n} \left(\sum_{i=1}^{m} r(j/i)^{1/\gamma} \right)^{\gamma} - 1 \leqslant$$

$$\leqslant \left(\sum_{j=1}^{n} \sum_{i=1}^{m} p(i,j)^{1/\gamma} \right)^{\gamma} - 1,$$

donde, já que $\frac{1}{\gamma-1} < 1$, segue que

$$H_\gamma(Y) + H_\gamma(X/Y) \geqslant H_\gamma(X,Y),$$

o que prova o teorema.

Como vimos no § 2, Cap. 1, temos a igualdade para a entropia de Shannon.

CAPÍTULO 3

Capacidade-γ Generalizada de um Canal Discreto sem Memória.

Neste capítulo, definimos a capacidade-γ ($\gamma > 1$) de um canal discreto sem memória por meio da entropia-γ. Introduzimos também, o conceito de funções pseudocôncavas e, verificamos que estas funções são facilmente maximizadas em uma região convexa. Então, provaremos o teorema que dá condições necessárias e suficientes para um canal discreto sem memória atingir a capacidade-γ. Para os canais discretos sem memória simétricos, calculamos à capacidade-γ. Finalmente, exemplos são dados para a computação da capacidade-γ. Devemos observar também que, sob condições mais fracas provamos o teorema 4.41 de |6|.

Um canal discreto sem memória com alfabeto de entrada $X = \{1,2,\ldots,m\}$ e alfabeto de saida $Y = \{1,2,\ldots,n\}$ é caracterizado pela matriz mxn $R = [r(j/i)]$, $j=1,2,\ldots,n$; $i=1,2,\ldots,m$, com $r(j/i) \geq 0$ e $\sum_{j=1}^{n} r(j/i) = 1$, $i=1,2,\ldots,m$. Aqui, $r(j/i)$ representa a probabilidade condicional do j-ésimo símbolo de saida, se o i-ésimo símbolo de entrada foi transmitido.

Consideramos uma distribuição de probabilidade de entrada arbitrária $(p_1,p_2,\ldots,p_m)\in\Delta_m$ sobre o alfabeto de entrada, que induz a distribuição $(q_1,q_2,\ldots,q_n)\in\Delta_n$ sobre o alfabeto de saida dada por

$$q_j = \sum_{i=1}^{m} p_i r(j/i)$$

A velocidade de transmição ou informação comunicada pelo canal é definida por

$$I_\gamma(X/Y) = H_\gamma(X) - H_\gamma(X/Y), \ \gamma > 0, \ \gamma \neq 1.$$

Definimos então, a capacidade-γ de um canal discreto sem memória caracterizado por a matriz R, como a quantidade

$$C_\gamma = \max_{(p_1,p_2,\ldots,p_m)\in\Delta_m} \left[H_\gamma(X) - H_\gamma(X/Y)\right],$$

onde $\gamma > 0$, $\gamma \neq 1$, e as distribuições de probabilidade de X e Y são dadas por $(p_1,p_2,\ldots,p_m)$ e $(q_1,q_2,\ldots,q_n)$, $q_j = \sum_{i=1}^{m} p_i r(j/i)$, $j=1,2,\ldots,n$, respectivamente. E, então $I_\gamma(X/Y)$ pode ser escrita como

$$I_\gamma(X/Y) = \frac{1}{\gamma-1}\left[\left(\sum_{i=1}^{m} p_i^{1/\gamma}\right)^\gamma - \sum_{j=1}^{n}\left(\sum_{i=1}^{m} p_i^{1/\gamma} r(j/i)^{1/\gamma}\right)\right] =$$

$$= \frac{1}{\gamma-1}\left(\sum_{i=1} p_i^{1/\gamma}\right)^\gamma \left[1 - \sum_{j=1}^{n} \frac{\left(\sum_{i=1}^{m} p_i^{1/\gamma} r(j/i)^{1/\gamma}\right)^\gamma}{\left(\sum_{i=1}^{m} p_i^{1/\gamma}\right)^\gamma}\right]$$

$$= \frac{1}{\gamma-1}\left(\sum_{i=1}^{m} p_i^{1/\gamma}\right)^\gamma \left[1 - \sum_{j=1}^{n}\left(\sum_{i=1}^{m} \frac{p_i^{1/\gamma}}{\sum_{i=1}^{m} p_i^{1/\gamma}} r(j/i)^{1/\gamma}\right)^\gamma\right].$$

Seja

$$p_i' = \frac{p_i^{1/\gamma}}{\sum_{i=1}^{m} p_i^{1/\gamma}}, \quad i=1,2,\ldots m.$$

Então, $\sum_{i=1}^{m} p_i' = 1$. Agora, como $p_i'^{\gamma} = \dfrac{p_i}{\left(\sum_{i=1}^{m} p_i^{1/\gamma}\right)^{\gamma}}$, então temos que

$$\sum_{i=1}^{m} p_i'^{\gamma} = \frac{1}{\left(\sum_{i=1}^{m} p_i^{1/\gamma}\right)^{\gamma}} .$$

Logo,

$$I_{\gamma}(X/Y) = \frac{1}{\gamma-1} \frac{1}{\left(\sum_{i=1}^{m} p_i'^{\gamma}\right)} \left[1 - \sum_{j=1}^{n} \left(\sum_{i=1}^{m} p_i' \, r(j/i)^{1/\gamma}\right)^{\gamma}\right],$$

onde $\gamma > 0$, $\gamma \neq 1$ e $p_i' = \dfrac{p_i^{1/\gamma}}{\sum_{i=1}^{m} p_i^{1/\gamma}}$, $i=1,2,\ldots,m$.

<u>Lema 3.1</u>.- A função

$$F(\gamma,p') = \sum_{j=1}^{n} \left(\sum_{i=1}^{m} p_i' \, r(j/i)^{1/\gamma}\right)^{\gamma},$$

é convexa em Δ_m, para $\gamma > 1$.

<u>Prova</u>: Sejam P' e Q' em Δ_m e $0 \leqslant \lambda \leqslant 1$. Então,

$$F(\gamma,\lambda P'+(1-\lambda)Q') = \sum_{j=1}^{n} \left(\sum_{i=1}^{m} (\lambda p_i' + (1-\lambda) q_i') r(j/i)^{1/\gamma}\right)^{\gamma} =$$

$$= \sum_{j=1}^{n} \left[\lambda \sum_{i=1}^{m} p_i' r(j/i)^{1/\gamma} + (1-\lambda) \sum_{i=1}^{m} q_i' r(j/i)^{1/\gamma}\right]^{\gamma} \leqslant$$

$$\leq \sum_{j=1}^{n} |\lambda(\sum_{i=1}^{m} p_i' r(j/i)^{1/\gamma})^\gamma + (1-\lambda)(\sum_{i=1}^{m} q_i' r(j/i)^{1/\gamma})^\gamma| =$$

$$= \lambda \sum_{j=1}^{n} (\sum_{i=1}^{m} p_i' r(j/i)^{1/\gamma})^\gamma + (1-\lambda)(\sum_{i=1}^{m} q_i' r(j/i)^{1/\gamma})^\gamma =$$

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

onde a desigualdade é obtida do fato da função $\omega(x) = x^{1/\gamma}$, $x \geq 0$, se: convexa com respeito a x, para $\gamma > 1$.

Lema 3.2.- A função

$$E(\gamma,P') = 1 - \sum_{j=1}^{n} (\sum_{i=1}^{m} p_i' r(j/i)^{1/\gamma})^\gamma,$$

é não-negativa e côncava em Δ_m, para $\gamma > 1$.

Prova: Da desigualdade de Minkowski (ver Gallager, 1968, |6|), para $\gamma > 1$, temos que

$$\sum_{j=1}^{n} (\sum_{i=1}^{m} p_i' r(j/i)^{1/\gamma})^\gamma \leq [\sum_{i=1}^{m} p_i' (\sum_{j=1}^{n} r(j/i)^{1/\gamma}]^\gamma = (\sum_{i=1}^{m} p_i')^\gamma = 1,$$

donde

$$0 \leq 1 - \sum_{j=1}^{n} (\sum_{i=1}^{m} p_i' r(j/i)^{1/\gamma})^\gamma = E(\gamma,P'),$$

isto é, $E(\gamma,P')$ é não-negativa.

A concavidade é uma consequencia imediata do lema 3.1. Isto prova o lema.

Definição 3.1.- Uma função real f definida sobre um conju to convexo $A \subset \mathbb{R}^m$ é pseudocôncava sobre A se para todo $x \in A$, todo $y \in A$

$$\nabla f(x)(y-x) \leq 0 \quad \text{implicar} \quad f(y) \leq f(x).$$

Concavidade implica pseudoconcavidade, mas existem funções pseudo-côncavas que não são côncavas.

É fácil ver (|8|) que se g é uma função real definida sobre um conjunto convexo $A \subset \mathbb{R}^m$, g positiva e côncava sobre A, e se h é uma função real definida também sobre A, h não-negativa e côncava sobre A, então a função g.h é pseudocôncava sobre A.

Teorema 3.1.- A informação comunicada

$$I_\gamma(X/Y) = \frac{1}{\gamma-1}\left(\sum_{i=1}^{m} p_i^{1/\gamma}\right)^\gamma \left[1 - \sum_{j=1}^{n}\left(\sum_{i=1}^{m} \frac{p_i^{1/\gamma}}{\sum_{i=1}^{m} p_i^{1/\gamma}} r(j/i)^{1/\gamma}\right)^\gamma\right],$$

onde $\gamma > 1$, é uma função pseudocôncava, não-negativa sobre Δ_m.

Prova: A função $\left(\sum_{i=1}^{m} p_i^{1/\gamma}\right)^\gamma$ é côncava e positiva para $\gamma > 1$ (teorema 2. Cap. 2). Do lema 3.2, a função $E(\gamma,P') = 1 - \sum_{j=1}^{n}\left(\sum_{i=1}^{m} p_i' r(j/i)^{1/\gamma}\right)^\gamma$, $\gamma > 1$, é não-negativa e côncava sobre Δ_m. Logo, a função $1 - \sum_{j=1}^{n}\left(\sum_{i=1}^{m} \frac{p_i^{1/\gamma}}{\sum_{i=1}^{m} p_i^{1/\gamma}} r(j/i)^{1/\gamma}\right)^\gamma$, $\gamma > 1$ é não-negativa e côncava sobre Δ_m. Assim, como consequência da observação acima, a função $I_\gamma(X/Y)$ é pseudocôncava, não-negativa sobre Δ_m.

O teorema seguinte se encontra demonstrado em Gallager (1968), |6|, 68), sob a condição mais forte de concavidade da função f sobre Δ_m.

Teorema 3.2.- Seja f(P) uma função pseudocôncava de $P=(p_1,p_2,\ldots,p_m)$ no conjunto Δ_m. Suponhamos que as derivadas parciais $\frac{\partial f(P)}{\partial p_i}$ são definidas e contínuas sobre Δ_m com a possível exceção que $\lim_{p_i \to 0} \frac{\partial f(P)}{\partial p_i}$ pode ser $+\infty$. Então,

$$\frac{\partial f(P)}{\partial p_i} = \lambda, \text{ para todo } i \text{ tal que } p_i > 0, \tag{1}$$

$$\frac{\partial f(P)}{\partial p_i} \leqslant \lambda, \text{ para todo } i \text{ tal que } p_i = 0, \tag{2}$$

para algūm nūmero λ, são condicões necessārias e suficientes sobre um vetor probabilidade P maximizar f sobre Δ_m.

<u>Prova</u>: (Suficiência).- Suponhamos que as condições

$$\frac{\partial f(P)}{\partial p_i} = \lambda, \text{ para todo } i \text{ tal que } p_i > 0,$$

$$\frac{\partial f(P)}{\partial p_i} \leqslant \lambda, \text{ para todo } i \text{ tal que } p_i = 0$$

são satisfeitas, para algūm nūmero λ e algūm vetor de probabilidade P. Mostraremos que para qualquer outro vetor probabilidade Q, $f(Q) - f(P) \leqslant 0$, estabelecendo que P maximiza f.

Temos que

$$\nabla f(P)(Q - P) = \sum_{i=1}^{m} \frac{\partial f(P)}{\partial p_i}(q_i - p_i).$$

As hipōteses (1) e (2) excetuam o caso $\dfrac{\partial f(P)}{\partial p_i} = +\infty$. Observe que se $p_i > 0$, então

$\dfrac{\partial f(P)}{\partial p_i}(q_i - p_i) = \lambda(q_i - p_i)$. Tambēm, se $p_i = 0$ temos que $q_i - p_i > 0$ e então

$\frac{\partial f(P)}{\partial p_i}(q_i - p_i) \leqslant \lambda(q_i - p_i)$. Portanto, segue-se que

$$\sum_{i=1}^{m} \frac{\partial f(P)}{\partial p_i}(q_i - p_i) \leqslant \lambda \sum_{i=1}^{m}(q_i - p_i) = \lambda\left(\sum_{i=1}^{m} q_i - \sum_{i=1}^{m} p_i\right) = 0.$$

Assim, $\nabla f(P)(Q - P) \leqslant 0$, o que implica que $f(Q) \leqslant f(P)$, já que f é pseudocôncava.

(Necessidade): Suponhamos que P maximiza f. Então, temos que

$$f(\theta Q + (1-\theta)P) - f(P) \leqslant 0,$$

para qualquer vetor probabilidade Q e qualquer $0 < \theta < 1$. Dividindo esta desigualdade por θ e tomando o limite quando $\theta \to 0$, temos que

$$\left.\frac{\partial f(\theta Q + (1-\theta)P)}{\partial \theta}\right|_{\theta=0} \leqslant 0 ,$$

ou seja,

$$\sum_{i=1}^{m} \frac{\partial f(P)}{\partial p_i}(q_i - p_i) \leqslant 0. \qquad (3)$$

Pelo menos uma componente de P é estritamente positiva, e suponhamos, por simplicidade de notação, que $p_1 > 0$. Seja I_i um vetor unitário co[illegible] 1 na i-ésima componente e 0 nas demais, e escolhemos $Q=P+\varepsilon I_i-\varepsilon I_1$. Desde que $p_1 > 0$, Q é um vetor probabilidade para $0 \leqslant \varepsilon \leqslant p_1$. Substituindo este Q em (3), temos que

$$\varepsilon\frac{\partial f(P)}{\partial p_i} - \varepsilon\frac{\partial f(P)}{\partial p_1} \leqslant 0.$$

isto é,

$$\frac{\partial f(P)}{\partial p_i} \leqslant \frac{\partial f(P)}{\partial p_1}. \tag{4}$$

Se $p_i > 0$, ε pode ser também escolhido negativo, e então a desigualdade (4) fica invertida, dando

$$\frac{\partial f(P)}{\partial p_i} = \frac{\partial f(P)}{\partial p_1}, \qquad p_i > 0 \tag{5}$$

Finalmente, escolhendo $\lambda = \dfrac{\partial f(P)}{\partial p_1}$, (4) e (5) tornan-se equivalente a

$$\frac{\partial f(P)}{\partial p_i} = \lambda, \quad \text{se } p_i > 0,$$

$$\frac{\partial f(P)}{\partial p_i} \leqslant \lambda, \quad \text{se } p_i = 0.$$

Para completar a prova, consideremos um P tal que $\dfrac{\partial f(P)}{\partial p_i} = +\infty$ para algúm i e mostremos que um tal P não pode maximizar f. Suponhamos, por simplicidade de notação, que $p_1 > 0$. Temos que

$$\frac{f(P + \varepsilon I_i - \varepsilon I_1) - f(P)}{\varepsilon} = \frac{f(P + \varepsilon I_i - \varepsilon I_1) - f(P + \varepsilon I_i)}{\varepsilon} + \frac{f(P+\varepsilon I_i) - f(P)}{\varepsilon}.$$

Quando $\varepsilon \to 0$, o primeiro têrmo do lado direito da igualdade acima fica limitado pela continuidade de $\frac{\partial f}{\partial p_1}$. O segundo têrmo tende ao infinito, de modo que o lado esquerdo é positivo, para ε suficientemente pequeno. Logo,

$$f(P + \varepsilon I_i - \varepsilon I_1) \geq f(P),$$

para ε suficientemente pequeno. Isto mostra que P não maximiza f, completando a prova do teorema.

Apresentamos agora, o teorema central do nosso trabalho, que dá condições necessárias e suficientes para um vetor probabilidade atingir a capacidade-γ de um canal discreto sem memória.

Teorema 3.3.- As condições necessárias e suficientes sobre um vetor probabilidade $P=(p_1,p_2,\ldots,p_m)$ atingir a capacidade-γ ($\gamma > 1$) de um canal discreto sem memória com matriz $\dot{R} = [r(j/i)]$, $j=1,2,\ldots,n; i=1,2,\ldots,m$, são que para algum número C_γ,

$$\frac{1}{\gamma-1}\left[\left(\sum_{k=1}^{m} p_k^{1/\gamma}\right)^{\gamma-1} p_i^{\frac{1}{\gamma}-1} - \sum_{j=1}^{n}\left(\sum_{k=1}^{m} p_k^{1/\gamma} r(j/i)^{1/\gamma}\right)^{\gamma-1} p_i^{\frac{1}{\gamma}-1} r(j/i)^{1/\gamma}\right] = C_\gamma \quad (6)$$

para todo i tal que $p_i > 0$, e

$$\frac{1}{\gamma-1}\left[\left(\sum_{k=1}^{m} p_k^{1/\gamma}\right)^{\gamma-1} p_i^{\frac{1}{\gamma}-1} - \sum_{j=1}^{n}\left(\sum_{k=1}^{m} p_k^{1/\gamma} r(j/i)^{1/\gamma}\right)^{\gamma-1} p_i^{\frac{1}{\gamma}-1} r(j/i)^{1/\gamma}\right] \leqslant C_\gamma \quad (7)$$

para todo i tal que $p_i = 0$. Além disso, o número C_γ é a capacidade-γ d canal.

Prova: Desejamos maximizar a função

$$I_\gamma(X/Y) = \frac{1}{\gamma-1}\left[\left(\sum_{i=1}^{m} p_i^{1/\gamma}\right)^{\gamma} - \sum_{j=1}^{n}\left(\sum_{i=1}^{m} p_i^{1/\gamma} r(j/i)^{1/\gamma}\right)^{\gamma}\right], \quad \gamma > 1,$$

sobre Δ_m. Tomando as derivadas parciais, temos que

$$\frac{\partial I_\gamma(X/Y)}{\partial p_i} = \frac{1}{\gamma-1}\Big[\big(\sum_{k=1}^{m} p_k^{1/\gamma}\big)^{\gamma-1} p_i^{\frac{1}{\gamma}-1} - \sum_{j=1}^{n}\big(\sum_{k=1}^{m} p_k^{1/\gamma} r(j/k)^{1/\gamma}\big)^{\gamma-1} p_i^{\frac{1}{\gamma}-1} r(j/i)^{1/\gamma}\Big] \qquad (8)$$

Podemos aplicar o teorema 3.2 de maximização, desde que $I_\gamma(X/Y)$, $\gamma > 1$, é pseudo-côncava em $P\varepsilon\Delta_m$, e suas derivadas parciais satisfazem as condições de continuidade apropriadas. Então, condições necessárias e suficientes sobre P para maximizar $I_\gamma(X/Y)$ são :

$$\frac{\partial I_\gamma(X/Y)}{\partial p_i} = \lambda, \text{ se } p_i > 0,$$

e,

$$\frac{I_\gamma(X/Y)}{\partial p_i} < \lambda, \text{ se } p_i = 0.$$

Usando (8) e tomando $C_\gamma = \lambda$, temos que

$$\frac{1}{\gamma-1}\left[\big(\sum_{k=1}^{m} p_k^{1/\gamma}\big)^{\gamma-1} p_i^{\frac{1}{\gamma}-1} - \sum_{j=1}^{n}\big(\sum_{k=1}^{m} p_k^{1/\gamma} r(j/k)^{1/\gamma}\big)^{\gamma-1} p_i^{\frac{1}{\gamma}-1} r(j/i)^{1/\gamma}\right] \begin{cases} = C_\gamma, & p_i > 0 . \\ \leqslant C_\gamma, & p_i = 0. \end{cases}$$

Agora, multiplicando ambos os lados de (6) por p_i, e somando sobre os i tais que $p_i > 0$, temos o valor máximo de $I_\gamma(X/Y)$ do lado esquerdo e a constante C_γ do lado direito, estabelecendo que C_γ é, de fato, a capacidade do canal.

Definição 3.2.- Um canal discreto sem memória com matriz R é dito simétrico se as colunas podem ser divididas em subconjuntos tal que cada linha dentro de um subconjunto é uma permutação de cada outra linha dentro do mesmo subconjunto e cada coluna (se mais que uma) dentro de um subconjunto é uma permutação de cada outra coluna dentro do mesmo subconjunto. Por exemplo, se R é a matriz

$$\begin{bmatrix} \boxed{\begin{matrix} 0.7 & 0.1 \\ 0.1 & 0.7 \end{matrix}} & \boxed{\begin{matrix} 0.2 \\ 0.2 \end{matrix}} \end{bmatrix} \quad \text{ou} \quad \begin{bmatrix} \boxed{\begin{matrix} 0.7 & 0.2 & 0.1 \\ 0.2 & 0.1 & 0.7 \\ 0.1 & 0.7 & 0.2 \end{matrix}} \end{bmatrix}$$

então o canal é simétrico, desde que cada submatriz satisfaz as propriedades de permutação.

Teorema 3.4.- Para um canal discreto sem memória simétrico com matriz R, a capacidade C_γ é atingida para um vetor probabilidade de entrada P, cujos elementos são

$$p_i = \frac{1}{m},\ i=1,2,\ldots,m.$$

Prova: Sejam as colunas de R divididas em ℓ subconjuntos com n_i colunas, $i=1,2,\ldots,\ell$, no i-ésimo subconjunto. Além disso, denota-se por a_i a soma dos elementos de cada coluna com potência $\frac{1}{\gamma}$ do i-ésimo subconjunto, e por b_i a soma dos elementos de cada linha com potência $\frac{1}{\gamma}$ do i-ésimo subconjunto. Então, precisamos provar que

$$\frac{1}{\gamma-1}\left[\{m(\tfrac{1}{m})^{1/\gamma}\}^{\gamma-1}(\tfrac{1}{m})^{\frac{1}{\gamma}-1} - \sum_{j=1}^{n}\sum_{k=1}^{m}(\tfrac{1}{m})^{1/\gamma} r(j/k)^{1/\gamma})^{\gamma-1}(\tfrac{1}{m})^{\frac{1}{\gamma}-1} r(j/i)^{1/\gamma}\right] = C_\gamma$$

para $i=1,2,\ldots,m$. Mas, o membro esuerdo da equação acima é igual a

$$\frac{1}{\gamma-1}\left[m^{\gamma-1}(\tfrac{1}{m})^{\frac{\gamma-1}{\gamma}}(\tfrac{1}{m})^{\frac{1}{\gamma}-1} - \sum_{j=1}^{n}\{(\tfrac{1}{m})^{1/\gamma}\sum_{k=1}^{m} r(j/k)^{1/\gamma}\}^{\gamma-1}(\tfrac{1}{m})^{\frac{1}{\gamma}-1} r(j/i)^{1/\gamma}\right] =$$

$$= \frac{1}{\gamma-1}\left[m^{\gamma-1} - \sum_{j=1}^{n}(\tfrac{1}{m})^{\frac{\gamma-1}{\gamma}}(\sum_{k=1}^{m} r(j/k)^{1/\gamma})^{\gamma-1}(\tfrac{1}{m})^{\frac{1}{\gamma}-1} r(j/i)^{1/\gamma}\right] =$$

$$= \frac{1}{\gamma-1}\left[m^{\gamma-1} - \sum_{j=1}^{n}(\sum_{k=1}^{m} r(j/k)^{1/\gamma})^{\gamma-1} r(j/i)^{1/\gamma}\right].$$

Escrevendo,

$$\sum_{j=1}^{n} = \sum_{j=1}^{n_1} + \sum_{n_1+1}^{n_2} + \ldots + \sum_{n_1+n_2+\ldots+n_{\ell-1}+1}^{n} ,$$

é facilmente visto que (9) é igual a

$$\frac{1}{\gamma-1}\left[m^{\gamma-1} - \sum_{k=1}^{\ell} a_k{}^{\gamma-1} b_k\right].$$

Portanto,

$$C_\gamma = \frac{1}{\gamma-1}\left[m^{\gamma-1} - \sum_{k=1}^{\ell} a_k{}^{\gamma-1} b_k\right] ,$$

que independe do índice i. Assim, (6) é satisfeita e a capacidade-γ é atingida.

Exemplo.- Consideremos o canal discreto sem memória com a seguinte matriz canal $(n+1)x(n+1)$:

$$R = \begin{bmatrix} 1-a & \frac{a}{n} & \ldots, & \frac{a}{n} \\ \frac{a}{n} & 1-a & \ldots & \frac{a}{n} \\ \vdots & & & . \\ \frac{a}{n} & \frac{a}{n} & \ldots & 1-a \end{bmatrix} ,$$

onde $0 \leqslant a \leqslant 1$. Então, R é simétrica e o teorema 3.4 dá a capacidade-γ:

$$C_\gamma = \frac{1}{\gamma-1}[(n+1)^{\gamma-1} - a_1{}^{\gamma-1} b_1] =$$

$$= \frac{1}{\gamma-1}\left[(n+1)^{\gamma-1} - \{(1-a)^{1/\gamma} + n(\tfrac{a}{n})^{1/\gamma}\}^{\gamma-1}\{(1-a)^{1/\gamma} + n(\tfrac{a}{n})^{1/\gamma}\}\right] =$$

$$= \frac{1}{\gamma-1}\left[(n+1)^{\gamma-1} - \{(1-a)^{1/\gamma} + n(\tfrac{a}{n})^{1/\gamma}\}\right].$$

Este resultado é uma generalização de um resultado conhecido para a capacidade de Shannnon.

$$\lim_{\gamma\to 1} \frac{1}{\gamma-1}\left[(n+1)^{\gamma-1} - \{(1-a)^{1/\gamma} + \quad (1-a)^{1/\gamma} + n(\frac{a}{n})^{1/\gamma}\}^{\gamma}\right] =$$

$$= \lim_{\gamma\to 1} \frac{d}{d\gamma}\left[(n+1)^{\gamma-1} - \{(1-a)^{1/\gamma} + n(\frac{a}{n})^{1/\gamma}\}^{\gamma}\right] =$$

$$= \lim_{\gamma\to 1} (n+1)^{\gamma-1}\log_e(n+1) - \lim_{\gamma\to 1} \frac{d}{d\gamma}\{(1-a)^{1/\gamma} + n(\frac{a}{n})^{1/\gamma}\}^{\gamma} =$$

$$= \log_e(n+1) - \lim_{\gamma\to 1} \frac{dy}{d\gamma} ,$$

onde,

$$y = \{(1-a)^{1/\gamma} + n(\frac{a}{n})^{1/\gamma}\}^{\gamma} .$$

Portanto,

$$\log_e y = \gamma\log_e\{(1-a)^{1/\gamma} + n(\frac{n}{a})^{1/\gamma}\},$$

donde temos que

$$\frac{1}{y}\frac{dy}{d\gamma} = \log_e\{(1-a)^{1/\gamma} + n(\frac{a}{n})^{1/\gamma}\} +$$

$$+ \frac{\gamma}{\{(1-a)^{1/\gamma} + n(\frac{a}{n})^{1/\gamma}\}} \circ (\frac{-1}{\gamma^2}) \left[(1-a)^{1/\gamma}\log_e(1-a) + n(\frac{a}{n})^{1/\gamma}\log_e \frac{a}{n}\right] .$$

Portanto,

$$\lim_{\gamma\to 1}\frac{dy}{d\gamma} = \lim_{\gamma\to 1}\{(1-a)^{1/\gamma} + n(\tfrac{a}{n})^{1/\gamma}\}^{\gamma}\left[\log_e[(1-a)^{1/\gamma} + n(\tfrac{a}{n})^{1/\gamma}] - \right.$$

$$\left. - \frac{1}{\gamma}\frac{[(1-a)^{1/\gamma}\log_e(1-a) + n(\frac{a}{n})^{1/\gamma}\log_e \frac{a}{n}]}{|(1-a)^{1/\gamma} + n(\frac{a}{n})^{1/\gamma}|}\right] =$$

$$= -(1-a)\log_e(1-a) - a.\log_e \frac{a}{n}.$$

Portanto, temos que

$$C = \lim_{\gamma\to 1} C_\gamma = \log_e(n+1) + (1-a)\log_e(1-a) + a.\log_e \frac{a}{n} =$$

$$= \log_e(n+1) + (1-a)\log_e(1-a) + a.\log_e a - a.\log_e n.$$

Para um <u>canal binario simētrico</u>, temos com n=1, que

$$C_\gamma = \frac{1}{\gamma-1}\left[2^{\gamma-1} - \{(1-a)^{1/\gamma} + a^{1/\gamma}\}^{\gamma}\right] = \frac{2^{\gamma-1} - 1}{\gamma - 1} - H_\gamma(a,1-a), \quad (10$$

de ónde segue que

$$C = \lim_{\gamma\to 1} C_\gamma = \log_e 2 + (1-a)\log_e(1-a) + a.\log_e a =$$

$$= \log_e 2 - H(a,1-a).$$

<u>Teorema 3.5</u>.- Para um canal discreto sem memōria com matri canal 2xn temos que

$$C_2 = 1 - \sum_{j=1}^{n} r(j/1)^{1/2} r(j/2)^{1/2}.$$

<u>Prova</u>: Tomamos $(p_1,p_2) = (x,1-x)\varepsilon\Delta_2$, $x\varepsilon(0,1)$. Então,

$$H_2(X) - H_2(X/Y) =$$

$$= \{[x^{1/2}+(1-x)^{1/2}]^2-1\}-\{\sum_{j=1}^{n}[x^{1/2}r(j/1)^{1/2}+(1-x)^{1/2}r(j/2)^{1/2}]^2-1\} =$$

$$= x+(1-x)+2x^{1/2}(1-x)^{1/2}-\sum_{j=1}^{n}\left[xr(j/1)+(1-x)r(j/2)+2x^{1/2}(1-x)^{1/2}r(j/1)^{1/2}r(j/2)^{1/2}\right] =$$

$$= 1+2x^{1/2}(1-x)^{1/2}-x\sum_{j=1}^{n}r(j/1)-(1-x)\sum_{j=1}^{n}r(j/2)-2x^{1/2}(1-x)^{1/2}\sum_{j=1}^{n}r(j/1)^{1/2}r(j/2)^{1/2} =$$

$$= 1+2x^{1/2}(1-x)^{1/2}-x-(1-x)-2x^{1/2}(1-x)^{1/2}\sum_{j=1}^{n}r(j/1)^{1/2}r(1/2)^{1/2} =$$

$$= 2x^{1/2}(1-x)^{1/2}[1-\sum_{j=1}^{n}r(j/1)^{1/2}r(j/2)^{1/2}] = T(x).$$

Agora,

$$T'(x) = 2\left[1-\sum_{j=1}^{n}r(j/1)^{1/2}r(j/2)^{1/2}\right]\left[\frac{1}{2}x^{-\frac{1}{2}}(1-x)^{1/2}-\frac{1}{2}(1-x)^{-\frac{1}{2}}x^{1/2}\right] =$$

$$= \left[1-\sum_{j=1}^{n}r(j/1)^{1/2}r(j/2)^{1/2}\right]\left[x^{-\frac{1}{2}}(1-x)^{1/2}-(1-x)^{-\frac{1}{2}}x^{1/2}\right].$$

Assim, $T'(x) = 0$, quando $x = \frac{1}{2}$ com $\sum_{j=1}^{n}r(j/1)^{1/2}r(j/2)^{1/2} \neq 1$ e $0 < x < 1$. Isto implica que a função $T(x)$ (que é pseudocôncava) tem um único valor máximo no ponto $x = \frac{1}{2}$. Portanto, obtemos que

$$C_2 = \max_{(p_1,p_2)\varepsilon\Delta_2} \left[H_2(X) - H_2(X/Y)\right] =$$

$$= \max_{x\varepsilon(0,1)} T(x) =$$

$$= T\left(\frac{1}{2}\right) =$$

$$= 1 - \sum_{j=1}^{n} r(j/1)^{1/2} r(j/2)^{1/2} .$$

Assim, o teorema está provado.

Pará o <u>canal simétrico binario</u> ($r(1/1) = r(2/2) = 1 - a$ e $r(1/2) = r(2/1) = a$), temos que

$$C_2 = 1 - \sum_{j=1}^{2} r(j/1)^{1/2} r(j/2)^{1/2} =$$

$$= 1 - r(1/1)^{1/2} r(1/2)^{1/2} - r(2/1)^{1/2} r(2/2)^{1/2} =$$

$$= 1 - a^{1/2}(1-a)^{1/2} - a^{1/2}(1-a)^{1/2} =$$

$$= 1 - 2a^{1/2}(1-a)^{1/2} =$$

$$= 2 - \left[a^{1/2} + (1-a)^{1/2}\right]^2 =$$

$$= 1 - \frac{1}{2-1}\left[\{a^{1/2} + (1-a)^{1/2}\}^2\right] =$$

$$= 1 - H_2(a,1-a),$$

que coincide com (10) no caso $\gamma = 2$.

REFERÊNCIAS

|1|. Arimoto, S.: Information-Theoretical Considerations on Estimation Problems Information and Control, Vol. 19, Nº 3, October 1971, pag. 181-194.

|2|. Behara, M. and Chawla, J. S.: Generalized Gama-Entropy- Entropy and Ergodic Theory, Selecta Statistica Canadiana, Vol. II, 1974, pag. 15-38

|3|. Ben-Bassat, M. and Raviv, J.: Rényi's Entropy and the Probability of Error, IEEE Transactions on Information Theory, Vol. II-24, Nº 3, May 1978, pag. 324-331.

|4|. Cheng, M. C.: On Capacity and Randon Coding Exponent of DMC, Information and Control, 23, 1973, pag. 288-300.

|5|. Daroczy, Z.:Generalized Information Functions, Information and Control, 16, 1970, pag. 36-51.

|6|. Gallager, R.G.: Information Theory and Reliable Communication, John-Wiley and Sons, Inc. , 1868.

|7|. Guerra, F.: Entropias Generalizadas e Capacidade do canal discreto sem Memória. Tese de Mestrado, UFSC, Novembro 1980.

|8|. Mangasarian, O. L. : Non-linear Programming, Mc.Graw-Hill, Inc., 1969.

|9|. Renyi, A.: On Measures of Entropy and Information, Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability , University of California Press, Berkeley, California, 1961, Vol. I pag. 547-561.

|10|. Shannon, C. E.: A Mathematical Theory of Communication, B.S.T.J. 1948, pag. 379-423, 623-656.

|11|. Taneja, I.J.: Some Contributions to Information Theory and (a survey) on measures of Information, <u>Journal of Comb. and Syst. Sciences</u>, Vol. 4(4), 1979, pag. 259-280.