

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 em Matemática da Universidade Federal de Santa Catarina.

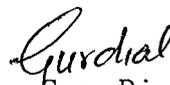


Prof. Dr. Inder Jeet Taneja  
Coordenador

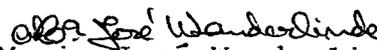
Banca Examinadora:



Prof. Dr. Inder Jeet Taneja  
Orientador



Prof. Dr. Gur Dial



Prof<sup>a</sup> M.Sc. Maria José Wanderlinde

UM MÉTODO ITERATIVO PARA CALCULAR A CAPACIDADE RELATIVA

PONDERADA

ORIENTADOR: PROF: DR. INDER JEET TANEJA

MARIANO MOREIRA

Março-1983

AGRADECIMENTOS

Ao Prof. Dr. Inder Jeet Taneja, pela orientação segura, criteriosa, dedicada e amiga demonstrada na realização desta pesquisa.

Aos colegas, parentes e amigos que direta ou in diretamente me apoiaram na realização da mesma.

Aos meus pais, que me deram a vida.

À minha esposa Valéria e aos meus filhos e filhas: Mariano Jr. , Charlles, Micheline e Caroline, pelas horas que a eles deixei de dedicar para a realização de este trabalho.

Estendo meus agradecimentos à Universidade Federal de Santa Catarina em particular ao Departamento de Matemática que forneceu os meios necessários para que este trabalho se concretizasse.

R E S U M O

Neste trabalho, propusemo-nos a elaborar um método iterativo para calcular a capacidade relativa ponderada de um canal discreto sem memória, isto é, a máxima velocidade de transmissão da informação por custo, com pesos.

O algoritmo iterativo tem por base o trabalho apresentado por Jimbo e Kunisawa [9] no qual introduzimos a velocidade de transmissão da informação com ponderação.

Mostramos também que a convergência do método é monótona não decrescente.

A B S T R A C T

In this work, we propose to elaborate an iteration method for calculating the relative weighted capacity of a discrete memoryless channel, i.e., the maximum of information transmission rate per cost, with weights.

The iterative algorithm is based on the work of Jimbo-Kunisawa [9] in which we introduce information transmission rate with weights.

We also show that the convergence of the procedure is monotone non decreasing.



## INTRODUÇÃO

No Capítulo I, apresentamos os conceitos de Entropias Ponderada, Ponderada Conjunta, Poderada Condicional, definimos informação comunicada pelo canal, com pesos, damos interpretações para o comprimento médio ponderado Ref. Longo<sup>[17]</sup> e finalmente é dada a codificação de letras únicas.

No Capítulo II, apresentamos um algoritmo para calcular a capacidade do canal usando Entropia Ponderada FLEMMING, Diva, M.<sup>[6]</sup>, onde definimos informação comunicada pelo canal, damos uma expressão da capacidade ponderada de um canal discreto, mostramos um Procedimento Iterativo e alguns teoremas que dão limites superior e inferior da capacidade.

No Capítulo III, apresentamos um Método de Iteração para calcular a Capacidade Relativa Ponderada de um canal discreto sem memória, damos a definição de Capacidade Relativa Ponderada, o conceito de Discriminação Ponderada e provamos um Teorema de Monotonicidade e Convergência do Procedimento de Iteração.

## CAPÍTULO I

### ENTROPIA PONDERADA E TEOREMA DE CODIFICAÇÃO

#### 1.1 - INTRODUÇÃO

A tarefa de um elo de comunicação é de reproduzir a saída normal de uma fonte de informação de tal modo que ela possa ser distinguida de alguma outra saída (considere somente fontes discretas).

Portanto, a descrição de uma fonte de informação reduz-se a uma lista de suas possíveis saídas (mais suas propriedades estatísticas), sem nenhuma atenção aos seus significados, ou seus efeitos, a algum observador colocado no final do elo de comunicação. Este ponto de vista corresponde para W. WEAVER o "Nível A" do problema da comunicação relativa à pergunta "como, precisamente, podem os símbolos de comunicação ser transmitidos?" (cf. SHANNON e WEAVER, [14, p. 24]).

Mesmo se os símbolos da fonte são recebidos de um modo não ambíguo, nada implica que o significado que eles transmitem ao observador se relacione com o significado que a fonte desejava transmitir. ("Nível B" de WEAVER). Além disso, mesmo que o significado seja corretamente transmitido, ele pode ter um efeito na conduta do observador que é bastante diferente daquele que a fonte esperava. Este aspecto adicional do problema da comunicação (correspondendo ao "Nível C" de WEAVER) é talvez menos difícil de lidar do que o aspecto semântico.

BELLIS e GUIAŞU<sup>[3]</sup> tentaram formalizar a interrelação entre os aspectos técnico (Nível A) e Pragmático (Nível C)

do problema da comunicação, no caso das fontes de informação discretas sem memória. Eles enriquecem a descrição usual da fonte de informação (isto é, um alfabeto finito e uma distribuição de probabilidades finita) dotando cada letra da fonte com uma medida paramétrica adicional, por exemplo, o proveito ou vantagem que o usuário consegue quando o evento correspondente a esta letra ocorre.

Deve ser ressaltado que o parâmetro adicional deve ter um caráter subjetivo, visto que ele se refere aos objetivos e finalidades do observador e não ao caráter objetivo como o possuído, por exemplo pelas fontes estatísticas. BELLIS e GUIAŞU<sup>[3]</sup> chamaram o parâmetro adicional de utilidade e consideraram o seguinte modelo para um experimento aleatório finito (ou fonte de informação) A:

$$A = \begin{bmatrix} x_1 & x_2 & \dots & x_n \\ p_1 & p_2 & \dots & p_n \\ w_1 & w_2 & \dots & w_n \end{bmatrix} = \begin{bmatrix} X \\ P \\ W \end{bmatrix} \quad (1.1)$$

onde X é o alfabeto, P a distribuição de probabilidades (d.p.) e  $W = (w_1, w_2, \dots, w_n)$  é a distribuição de utilidades (d.  $\mu$ ) ou Pesos. Os  $w_i$  são números reais não negativos. Eles também ressaltaram que a função (aqui e no que segue, logaritmos são tomados na base 2)

$$k(X) = k(P, W) = - \sum_{i=1}^n p_i w_i \log p_i, \quad (1.2)$$

Tem algumas propriedades interessantes e pode ser tomada como uma medida satisfatória para a quantidade média de informação valiosa ou proveitosa proporcionada por uma letra da fonte.

Esta Entropia Ponderada (1.2) foi estudada extensivamente por GUIAŞU (1977) [7].

A Entropia Ponderada foi definida tendo-se em vista as necessidades surgidas em vários campos importantes, por exemplo, na física, onde é necessário levar em consideração além das probabilidades, algumas características qualitativas dos eventos.

A fórmula apresentada para a entropia é uma medida de incerteza ou informação comunicada por um experimento probabilístico dependendo das probabilidades dos eventos e também de pesos qualitativos (objetivos ou subjetivos) dos possíveis eventos.

#### Definição 1.1 - Entropia Ponderada Conjunta

Sendo X e Y duas variáveis aleatórias com distribuição de probabilidades conjunta  $p(i,j)$ ,  $i = 1,2,\dots,n$ ,  $j = 1,2,\dots,m$ ,  $p(i,j) \geq 0$  e  $\sum_{i=1}^n \sum_{j=1}^m p(i,j) = 1$

Seja  $W(i,j) \geq 0$ ,  $i = 1,2,\dots,n$ ,  $j = 1,2,\dots,m$  um peso qualitativo associado ao evento elementar de probabilidade  $p(i,j)$ . A entropia Ponderada Conjunta é definida por

$$K(X,Y) = - \sum_{i=1}^n \sum_{j=1}^m W(i,j) p(i,j) \log p(i,j) \quad (1.3)$$

X e Y podem ser interpretados, também, como entrada e saída para um canal:

#### Definição 1.2 - Entropia Ponderada Condicional

Serão interpretadas como Entropias Ponderadas Condicionais as seguintes quantidades:

$$K(X/Y) = - \sum_{i=1}^n \sum_{j=1}^m W(i,j) p(i,j) \log p(i/j) \quad (1.4)$$

$$K(Y/X) = - \sum_{i=1}^n \sum_{j=1}^m W(i,j) p(i,j) \log p(j/i) \quad (1.5)$$

Onde  $p(i/j)$  e  $p(j/i)$  são probabilidades condicionais de  $X = x_i$  dado  $Y = y_j$  e  $Y = y_j$  dado  $X = x_i$  respectivamente.

Nota-se facilmente que se os pesos de todos os eventos são iguais, isto é,  $W(i,j) = 1$ ,  $i = 1, 2, \dots, n$  e  $j = 1, 2, \dots, m$ , então as quantidades (1.4) e (1.5) podem ser interpretadas, respectivamente, como

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

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

introduziremos as seguintes quantidades:

$$K(X) = - \sum_{i=1}^n \sum_{j=1}^m W(i,j) p(i,j) \log p_i \quad (1.6)$$

$$K(Y) = - \sum_{i=1}^n \sum_{j=1}^m W(i,j) p(i,j) \log q_j \quad (1.7)$$

onde  $q_j$  representa a probabilidade de  $Y = y_j$  e também pode ser escrita como

$$q_j = \sum_{i=1}^n p_i p(j/i) \quad (1.8)$$

Podemos interpretar a quantidade (1.6) como sendo a Entropia Ponderada de uma variável aleatória  $X$  tendo eventos elementares  $x_i$  ( $i = 1, 2, \dots, n$ ) e distribuição de probabilidades  $p_i$  ( $i = 1, 2, \dots, n$ ) se o peso de qualquer evento elementar  $x_i$  da variável aleatória  $X$  é justamente o principal valor dos pesos  $W(i, j)$  ( $j = 1, 2, \dots, m$ ) com respeito à distribuição de probabilidades  $p(j/i)$  ( $j = 1, 2, \dots, m$ ), a saber

$$\begin{aligned} K(X) &= - \sum_{i=1}^n \sum_{j=1}^m W(i, j) p(j/i) p_i \log p_i \\ &= - \sum_{i=1}^n W_i p_i \log p_i \end{aligned} \quad (1.9)$$

Onde 
$$W_i = \sum_{j=1}^m W(i, j) p(j/i) \quad (1.10)$$

Da mesma forma temos de (1.7)

$$\begin{aligned} K(Y) &= - \sum_{j=1}^m \sum_{i=1}^n W(i, j) p(i/j) q_j \log q_j \\ &= - \sum_{j=1}^m W_j q_j \log q_j \end{aligned} \quad (1.11)$$

Onde 
$$W_j = \sum_{i=1}^n W(i, j) p(i/j) \quad (1.12)$$

e  $q_j$  foi definido em (1.8).

São válidas as igualdades

$$i) \quad K(X, Y) = K(X) + K(Y/X) \quad (1.13)$$

$$ii) \quad K(X, Y) = K(Y) + K(X/Y) \quad (1.14)$$

Definição 1.3 - Informação Comunicada pelo Canal

Definimos a informação comunicada com pesos pelo canal sobre X a partir de Y como sendo

$$B(X/Y) = K(X) - (K(X/Y)) \quad (1.15)$$

Utilizando (1.2) e (1.4) podemos escrever:

$$B(X/Y) = \sum_{i=1}^n \sum_{j=1}^m W(i,j) p(i,j) \log \frac{p(i/j)}{P_i} \quad (1.16)$$

$$= \sum_{i=1}^n \sum_{j=1}^m W(i,j) p(i,j) \log \frac{p(j/i)}{P_j} \quad (1.16')$$

Que também é chamada informação média mútua e satisfaz as propriedades:

$$i) \quad B(X/Y) \geq 0 \quad (\text{Não negativa}) \quad (1.17)$$

$$ii) \quad B(X/Y) = B(Y/X) \quad (\text{Simetria}) \quad (1.18)$$

(Ref. GUIAŞU (1977) [7]).

Definição 1.4 - Capacidade Ponderada do Canal

A capacidade ponderada do canal discreto é definida por:

$$C_w = \max_{P \in \bar{P}} B(X/Y) \quad (1.19)$$

onde

$$\bar{P} = \{P = (p_1, p_2, \dots, p_n), p_i \geq 0, i = 1, 2, \dots, n, \sum_{i=1}^n p_i = 1\}.$$

## 1.2 - Teorema de Codificação

Mais recentemente GUIAŞU e PICARD<sup>[8]</sup> consideraram o problema de codificar as letras de saída por A em (1.1), por meio de um código de prefixo de letra única, cujas palavras códigos  $V_1, V_2, \dots, V_n$  tem comprimentos  $l_1, l_2, \dots, l_n$  satisfazendo a igualdade

$$\sum_{i=1}^n D^{-l_i} = 1 \quad (1.20)$$

D sendo o tamanho do alfabeto código. Eles definiram a seguinte quantidade:

$$L_w = \frac{\sum_{i=1}^n l_i w_i p_i}{\sum_{i=1}^n w_i p_i} \quad (1.21)$$

Que eles chamaram de comprimento médio ponderado do código e obtiveram um limite inferior para (1.21).

GUIAŞU e PICARD não motivaram adequadamente a introdução de  $L_w$  mas percebe-se que  $L_w$  pode ser de alguma importância na tentativa de juntar o "Nível A" e o "Nível C" do problema da comunicação.

Nas secções subsequentes ressaltaremos alguns problemas nos quais uma quantidade como o comprimento médio ponderado  $L_w$  em (1.21) é interessante e nos quais tal quantidade pode ser minimizada. O comportamento de  $L_w$  quando as seqüências da fonte se tornam cada vez maiores, é então investigado.

### 1.2.1 - Interpretação para o Comprimento Médio Ponderado.

Como um primeiro exemplo de uma situação na qual um Comprimento Médio Ponderado do tipo mostrado em (1.21) surge, considere o seguinte problema de transmissão. Uma fonte de informação A gera as letras  $\{x_1, x_2 \dots x_n\}$  com probabilidades  $\{p_1, p_2 \dots p_n\}$  e estas letras devem ser codificadas por meio de um código de prefixo com palavras código de comprimento  $\{l_1, l_2 \dots l_n\}$ . Desde que as diferentes letras são de diferentes importâncias para o receptor, suas palavras código devem ser tratadas diferentemente (por exemplo, elas devem ser transmitidas através do canal de comunicação sob diferentes condições de razão sinal-ruído, ou elas devem ir através de diferentes canais, ou elas devem ser gravadas usando diferentes materiais e/ou memórias). Isto permite um custo diferente por letra para cada palavra código, digamos  $W_i$ , para a palavra código de comprimento  $l_i$  da letra  $x_i$ , portanto, o custo  $C_i$  da letra  $x_i$  a ser transmitida é proporcional ao produto  $l_i w_i$ , digamos:

$$C_i = K l_i W_i \quad (1.22)$$

Para facilitarmos usaremos  $K = 1$ .

Se tentarmos minimizar o custo médio  $C_M = \sum_{i=1}^n C_i p_i$ ,

Somos levados a minimizar a quantidade

$$C_M = \sum_{i=1}^n l_i w_i p_i \quad (1.23)$$

ou, equivalentemente, a quantidade  $L_w$  que aparece em (1.21).

Considere agora o seguinte problema de armazenagem-

transmissão. As letras de saída  $\{x_1, x_2 \dots x_n\}$  em A devem ser novamente transmitidas usando um código de prefixo com palavras códigos de comprimento  $\{l_1, l_2 \dots l_n\}$ . Antes da transmissão, entretanto, cada letra deve permanecer por um certo tempo em um dispositivo de armazenagem (memória) e esse tempo é proporcional ao comprimento  $l_i$  da correspondente palavra código: somente quando a palavra código foi transmitida completamente, deve a letra sair do armazém (memória). Cada letra ocupa uma certa parte, digamos  $W_i$  da memória e podemos supor que o custo total de armazenagem da letra  $x_i$  por  $l_i$  segundos é proporcional a  $l_i w_i$ . Portanto, minimizar o custo médio novamente acarreta minimizar a quantidade (1.23).

Pode-se pensar na seguinte interpretação prática da situação:  $x_1, x_2 \dots x_n$  são  $n$  diferentes tipos ou recursos que chegam com probabilidades  $p_1, p_2 \dots p_n$  à memória onde permanecem  $l_1 \cdot l_2 \dots l_n$  unidades de tempo,  $l_i$  sendo o tempo requerido para examinar  $x_i$  e torná-lo pronto para deixar a memória. Se os elementos do tipo  $x_i$  ocupam uma parte  $w_i$  da memória, então a definição de custo dado em (1.22) é bastante natural. Observe que não existe, em princípio, um elo entre  $w_i$  e  $l_i$ . O primeiro depende do espaço ocupado por  $x_i$ , enquanto que o segundo pode depender de outros parâmetros (por exemplo: complexidade, fragilidade, etc).

É claro que neste exemplo o Parâmetro Adicional  $w_i$  relativo à  $i$ -ésima letra não representa uma utilidade de  $x_i$  relativa a algum observador, mesmo que assim seja em alguns outros casos. Portanto em algumas situações o nome utilidade para o parâmetro adicional é inadequado.

Examinemos mais um caso no qual o comprimento médio ponderado pode ser de interesse. Considere o seguinte jogo:

com probabilidades  $p_i$ , M. Júnior escolhe um elemento  $x_i$  de um conjunto  $X = \{x_1, x_2, \dots, x_n\}$ . CHARLLES tem que encontrar  $x_i$  perguntando ao M. Júnior perguntas binárias da forma "O elemento desconhecido pertence a  $X_j$ ?" onde  $X_j$  é algum sub-conjunto de  $X$ . Após  $l_i$  perguntas CHARLLES é capaz de identificar unicamente  $x_i$  (cada um dos sub-conjuntos  $X_2, X_3, \dots, X_{l_i}$  pode ser escolhido pelo CHARLLES de acordo com a(s) resposta(s) já dada(s) pelo M. Júnior à(s) pergunta(s) binária(s) anterior(es). Suponha que quando CHARLLES tiver encontrado  $x_i$  ele recebe uma quantia em dinheiro  $M_i$  ( $i = 1, 2, \dots, n$ ) do M. Júnior que em recompensa recebe uma quantia em dinheiro  $l_i w_i$  do CHARLLES (ambos  $M_i$  e  $w_i$  bem como  $p_i$  são quantidades fixas para todo  $i$ . Em certos casos é razoável supor que  $w_i$  é alguma função crescente de  $M_i$ ). Portanto, CHARLLES pode usar o grau de liberdade possuído por sua estratégia para minimizar sua perda média  $\sum_{i=1}^n p_i l_i w_i$ , enquanto que M. Júnior não tem nenhuma flexibilidade.

Exemplos práticos de tal situação são bastante comuns por exemplo, em diagnósticos médicos, e em geral, naqueles problemas de procura onde o custo para realizar os testes necessários para encontrar o objeto desconhecido (doença, elemento químico, etc) depende do elemento e é independente de sua frequência relativa. Muitas das atividades humanas de natureza utilitária ou de natureza exploratória se parecem ao jogo descrito acima.

Agora esqueçamos as interpretações particulares do problema e o formulemos mais formalmente na linguagem da teoria da informação, que tem a vantagem de fornecer uma bem conhecida faixa de referência. Suponha uma fonte de informação como A em (1.1), letras de saída  $x_i \in X$  com probabilidades  $p_i \in P$ . Cada letra  $x_i$  é, além disso, caracterizada por um parâme-

tro adicional  $W_i$ , e o custo  $C_i$  de transmissão de  $x_i$  através do canal sem ruído alimentado por A é proporcional ao produto  $l_i W_i$ , onde  $l_i$  é o comprimento da palavra código associado com  $x_i$ . Então o par codificador-decodificador é solicitado a minimizar o custo médio de transmissão ou, equivalente, o comprimento médio ponderado  $L_w$ .

### 1.2.2 - Codificação de Letras Únicas.

Consideremos primeiramente o caso da codificação do prefixo de uma letra única para a fonte de informação.

Escrevendo  $L_w$  como segue:

$$L_w = \sum_{i=1}^n a_i l_i \quad (1.24)$$

Onde:

$$a_i = \frac{p_i w_i}{\sum_{j=1}^k p_j w_j} \quad (1.25)$$

é uma distribuição de probabilidades auxiliar  $\mathcal{B}$  no alfabeto fonte, torna-se claro que  $L_w$  é realmente um comprimento médio de palavra código com respeito, entretanto, à distribuição de probabilidades  $\mathcal{B}$  e não à distribuição de probabilidades original P. Como uma consequência, temos o seguinte Lema 1, que de certo modo generaliza o resultado de GUIAŞU e PICARD [8] e é provado de um modo mais direto.

Lema 1:

Se os comprimentos de palavras códigos  $l_1, l_2, \dots, l_n$  satisfazem a desigualdade de Kraft

$$\sum_{i=1}^n D^{-l_i} \leq 1 \quad (1.26)$$

Então a quantidade  $L_w$  é limitada inferiormente por

$$L_w \geq \frac{K(P, W) - \overline{w \log w} + \bar{w} \log \bar{w}}{\bar{w} \log D} \quad (1.27)$$

onde  $K(P, W)$  é definido por (1.2) e a barra significa o valor médio com relação à Distribuição de Probabilidades  $P$ .

Demonstração:

A equação (1.27) é uma consequência imediata de desigualdade clássica

$$L_w \geq K(\mathcal{D}) / \log D$$

Onde

$$K(\mathcal{D}) = - \sum_{i=1}^n a_i \log a_i$$

Note que: Desde que  $x \log x$  é uma função convexa de  $W$ , a seguinte desigualdade vale:

$$\overline{w \log w} \geq \bar{w} \log \bar{w}$$

E portanto  $K(P, W)$  parece não ser tão básico quanto seria desejável a este respeito.

O seguinte Lema 2, novamente emprestado da Teoria Clássica, dá uma visão da precisão do limite inferior em (1.27).

Lema 2:

Escolhido apropriadamente os comprimentos  $l_1, l_2, \dots, l_n$  no código do Lema 1,  $L_w$  satisfaz a seguinte desigualdade:

$$L_w < \frac{K(P, W) - \overline{w} \log \overline{w} + \overline{w} \log \overline{w} + 1}{\overline{w} \log D} \quad (1.28)$$

Demonstração:

Escolha os valores de  $l_i$  ( $1 \leq i \leq n$ ) de modo a satisfazer:

$$D^{-l_i} \leq a_i = \frac{W_i P_i}{\overline{W}} < D^{-l_i + 1} \quad (1.29)$$

E portanto segue-se (1.28) imediatamente. (Observe que a desigualdade na esquerda de (1.29) implica que a desigualdade de Kraft (1.26) é satisfeita e portanto o código correspondente pode ser escolhido para ser um código prefixo).

## CAPÍTULO II

ALGORÍTMO PARA CALCULAR A CAPACIDADE DO CANAL  
USANDO ENTROPIA PONDERADA

Neste capítulo apresentaremos um algoritmo para calcular a capacidade do canal arbitrário discreto, utilizando a Entropia Ponderada definida no Cap. I (1.2). O procedimento iterativo é similar ao desenvolvido por Arimoto em 1972 (Ref. Arimoto (1972)<sup>[1]</sup>) é generalizado o conceito da Entropia Ponderada Condicional e mostrada a concavidade da função  $B(X/Y)$  bem como é apresentado em algoritmo que tem a propriedade de convergência monotônica para a capacidade; e algumas desigualdades que dão limite superior e inferior para a capacidade. Neste capítulo usaremos Logarítmos Naturais.

Para um estudo mais detalhado ver FLEMMING, Diva M.<sup>[6]</sup>

### 2.1 - Informação Comunicada pelo Canal

Seja um canal discreto com  $n$  símbolos de entrada e  $m$  símbolos de saída e com uma matriz canal  $m \times n$   $M = \{p(j/i)\}$ ,  $i = 1, 2, \dots, n$ ;  $j = 1, 2, \dots, m$ , onde  $P(j/i) \geq 0$  e  $\sum_{i=1}^n p(j/i) = 1$ . Sabemos que para cada  $i$  ( $i = 1, 2, \dots, n$ ) existe pelo menos um  $j$  ( $j = 1, 2, \dots, m$ ) tal que  $p(j/i) > 0$

$$\rho^k = \{P = (p_1, p_2, \dots, p_k) : p_i > 0 ; \sum_{i=1}^k p_i = 1\}$$

$$\bar{\rho}^k = \{P = (p_1, p_2, \dots, p_k) : p_i \geq 0 ; \sum_{i=1}^k p_i = 1\}$$

Então um vetor probabilidade n dimensional será  $p\{(j/. ) = \{p(j/1), p(j/2) \dots p(j/n)\}$  indicado por  $\bar{\rho}^n$ .

A informação comunicada pelo Canal M é definida por:

$$B(X/Y, P) = (K(X, P) - K(X/Y, P)) \quad (2.1)$$

Onde  $P = (p_1, p_2, \dots, p_n) \in \bar{\rho}^n$  é um vetor probabilidade de entrada e  $K(X, P)$ ;  $K(X/Y, P)$  são a Entropia Ponderada e Entropia Ponderada Condicional em (1.2) e (1.4) respectivamente, ou reescrevendo

$$K(X, P) = - \sum_{i=1}^n w_i p_i \log p_i \quad (2.2)$$

$$K(X/Y, P) = - \sum_{i=1}^n \sum_{j=1}^m w(i, j) p(i, j) \log p(i/j)$$

$$= - \sum_{i=1}^n \sum_{j=1}^m w(i, j) p(i, j) \log \frac{p(j/i)p_i}{\sum_{k=1}^m p(j/k)p_k}$$

$$= - \sum_{i=1}^n \sum_{j=1}^m w(i, j) p(j/i)p_i \log \frac{p(j/i)p_i}{\sum_{k=1}^m p(j/k)p_k}$$

(2.3)

A capacidade ponderada do canal discreto é dada por

$$C_w = \max_{P \in \bar{\rho}^n} B(X/Y, P) \quad (2.4)$$

Por outro lado podemos generalizar o conceito de entropia ponderada

rada condicional.

Se introduzirmos a matriz  $\phi, n \times m$  tal que:

$$\phi = \{\phi(i/j)\}, \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, m, \quad \phi(i/j) \geq 0$$

e

$$\sum_{i=1}^n \phi(i/j) = 1 \quad (2.5)$$

A entropia condicional generalizada ponderada é definida como:

$$\begin{aligned} J(X/Y, P, \phi) &= \\ &= - \sum_{i=1}^n \sum_{j=1}^m w(i,j) p(j/i) p_i \log \phi(i/j) \end{aligned} \quad (2.6)$$

Então, se  $\phi(i/j)$  é definida por:

$$\phi(i/j) = \frac{p(j/i) p_i}{\sum_{k=1}^m p(j/k) p_k} = p^*(i/j) \quad (2.7)$$

Então (2.6) torna-se igual a (2.3).

Vale a desigualdade:

$$J(X/Y, P, \phi) \geq J(X/Y, P, S) \quad (2.8)$$

onde  $S$  é a matriz da qual a entrada é a  $(i, j)$ ésima  $p^*(i/j)$  como definida em (2.7). Por esse fato obtém-se uma outra caracterização da capacidade como:

$$C_w = \max_{P \in \mathcal{P}^n} \max_{\phi \in \Phi} [K(X, P) - J(X/Y, P, \phi)] \quad (2.9)$$

onde  $\Phi$  significa o conjunto de todas as matrizes satisfazendo (2.5).

Teorema 2.1:

A informação comunicada pelo canal,  $B(X/Y, P)$  é uma função convexa para baixo, das probabilidades de entrada.

Especificamente se  $a_1, a_2, \dots, a_r$  são números não negativos tais que  $\sum_{k=1}^r a_k = 1$  e definirmos uma distribuição de entrada  $p_0(i) = \sum_{k=1}^r a_k p_k(i)$  e uma distribuição de saída

$$q_0(j) = \sum_{k=1}^r a_k q_k(j).$$

Então  $B_0(X/Y, P)$  correspondente a  $p_0(i)$  satisfaz

$$B_0(X/Y, P) \geq \sum_{k=1}^r a_k B_k(X/Y, P) \quad \text{onde}$$

$B_k(X/Y, P)$  é a informação comunicada pelo canal quando a distribuição de entrada é  $p_k(i)$ ,  $i = 1, 2, \dots, n$  e  $k = 1, 2, \dots, r$ .

Demonstração:

$$\text{Seja } \Delta B = B_0(X/Y, P) - \sum_{k=1}^r a_k B_k(X/Y, P) \quad (2.10)$$

Pela simetria da informação comunicada, podemos escrever:

$$\Delta B = B_0(Y/X, P) - \sum_{k=1}^r a_k B_k(Y/X, P) \quad (2.11)$$

ou ainda:

$$\begin{aligned} \Delta B &= K_0(Y, P) - K_0(Y/X, P) - \sum_{k=1}^r a_k [K_k(Y, P) - K_k(Y/X, P)] \\ &= K_0(Y, P) - \sum_{k=1}^r a_k K_k(Y, P) - [K_0(Y/X, P) - \sum_{k=1}^r a_k K_k(Y/X, P)] \end{aligned} \quad (2.12)$$

Agora:

$$\begin{aligned}
 K_O(Y/X,P) &= - \sum_{i=1}^n \sum_{j=1}^m w(i,j) p_O(i,j) \log p_O(j/i) \\
 &= - \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^r w(i,j) a_k p_k(i,j) \log p_O(j/i) \\
 &= \sum_{k=1}^r a_k \left[ - \sum_{i=1}^n \sum_{j=1}^m w(i,j) p_k(i,j) \log p_O(j/i) \right] \quad (2.13)
 \end{aligned}$$

Como  $p_O(j/i)$  são probabilidades do canal  $\{p(j/i)\}; i = 1, 2, \dots, n;$   
 $j = 1, 2, \dots, m$

$$\begin{aligned}
 p_O(j/i) &= \sum_{k=1}^r a_k p_k(j/i) \\
 &= a_1 p_1(j/i) + a_2 p_2(j/i) + \dots + a_r p_r(j/i) \\
 &= a_1 p(j/i) + a_2 p(j/i) + \dots + a_r p(j/i) \\
 &= (a_1 + a_2 + \dots + a_r) p(j/i) \text{ usando } \sum_{k=1}^r a_k = 1
 \end{aligned}$$

Temos  $p_O(j/i) = p(j/i)$ .

Então (2.13) fica sendo escrita como:

$$K_O(Y/X,P) = \sum_{k=1}^r a_k \left[ - \sum_{i=1}^n \sum_{j=1}^m w(i,j) p_k(i,j) \log p(j/i) \right]$$

$$\text{ou } K_O(Y/X,P) = \sum_{k=1}^r a_k K_k(Y/X,P)$$

$$\text{Portanto } K_O(Y/X,P) - \sum_{k=1}^r a_k K_k(Y/X,P) = 0 \quad (2.14)$$

Usando (2.14) e (2.12), temos:

$$\begin{aligned}
 \Delta B &= K_0(Y,P) - \sum_{k=1}^r a_k K_k(Y,P) \\
 &= - \sum_{j=1}^m w_j q_0(j) \log q_0(j) - \sum_{k=1}^r a_k \left( - \sum_{j=1}^m w_j q_k(j) \log q_k(j) \right) \\
 &= - \sum_{j=1}^m \sum_{k=1}^r w_j a_k q_k(j) \log q_0(j) + \sum_{j=1}^m \sum_{k=1}^r w_j a_k q_k(j) \log q_k(j) \\
 &= \sum_{k=1}^r a_k \left[ - \sum_{j=1}^m w_j q_k(j) \log q_0(j) + \sum_{j=1}^m w_j q_k(j) \log q_k(j) \right] \quad (2.15)
 \end{aligned}$$

Agora:

$$\sum_{j=1}^m q_0(j) = \sum_{j=1}^m \sum_{k=1}^r a_k q_k(j) = \sum_{k=1}^r a_k \sum_{j=1}^m q_k(j) = 1$$

Portanto:  $\sum_{j=1}^m q_0(j) = 1$  e  $\sum_{j=1}^m q_k(j) = 1$  para  $k = 1, 2, \dots, r$

Podemos então aplicar a desigualdade de SHANNON (Ref. Ash(1965)) [2]

$$- \sum_{j=1}^m q_k(j) \log q_k(j) \leq - \sum_{j=1}^m q_k(j) \log q_0(j) \quad (2.16)$$

$$k = 1, 2, \dots, r$$

ou levando-se em consideração que  $w_j \geq 0$  podemos de (2.16) escrever:

$$- \sum_{j=1}^m w_j q_k(j) \log q_0(j) + \sum_{j=1}^m w_j q_k(j) \log q_k(j) \geq 0 \quad (2.17)$$

Portanto, usando (2.17) em (2.15) concluímos que  $\Delta B \geq 0$ , ou seja:

$$B_0(X/Y, P) - \sum_{k=1}^r a_k B_k(X/Y, P) \geq 0 \quad \text{e isto implica que}$$

$B(X/Y, P)$  é uma função convexa para baixo de  $\bar{\mathcal{P}}^n$ .

Teorema 2.2:

O vetor probabilidade  $p^0 = (p_1^0, p_2^0, \dots, p_n^0) \in \bar{\mathcal{P}}^n$  maximiza  $B(X/Y, P)$  se e somente se:

$$\sum_{j=1}^m w(i, j) p(j/i) \log \frac{p(j/i)}{\sum_{k=1}^n p(j/k) p_k} \begin{cases} = C_w & \text{se } p_i^0 > 0 \\ < C_w & \text{se } p_i^0 = 0 \\ & i = 1, 2, \dots, n \end{cases} \quad (2.18)$$

2.2 - Procedimento Iterativo e Convergência

Baseado na caracterização (2.9) da capacidade propomos um procedimento para calcular a capacidade com os seguintes passos:

i) Inicialmente, escolhemos um vetor probabilidade arbitrário  $p^1 \in \mathcal{P}^n$  (na prática, uma distribuição uniforme de probabilidades  $p_i^1 = \frac{1}{n}$  para todo  $i = 1, 2, \dots, n$  é geralmente conveniente. Então as seguintes etapas são iterativas para  $t=1, 2, \dots$

ii) Maximize  $K(X, P^t) - J(X/Y, P^t, \phi)$  com respeito a  $\phi \in \Phi$  uma vez fixado  $P^t$ . De acordo com (2.8), a maximização de  $\phi$  é:

$$\phi^t(i/j) = \frac{p(j/i)p_i^t}{\sum_{k=1}^n p(j/k)p_k^t} \quad (2.19)$$

$$\begin{aligned} e \quad C(t,t) &= \max_{\phi \in \Phi} [K(X,P^t) - J(X/Y,P^t,\phi)] \\ &= K(X,P^t) - J(X/Y,P^t,\phi^t) \end{aligned} \quad (2.20)$$

iii) Maximize  $K(X,P) - J(X/Y,P,\phi^t)$  em relação a  $P \in \bar{\mathcal{P}}^n$ , Fixando  $\phi^t$ , este vetor probabilidade denotado por  $P^{t+1}$  é dado por:

$$p_i^{t+1} = \text{Exp} \left[ \frac{-\gamma}{w_i} - 1 + \frac{s_i^t}{w_i} \right] \quad (2.21)$$

$$\text{onde} \quad s_i^t = \sum_{j=1}^m w(j/i) p(j/i) \log \phi^t(i/j) \quad (2.22)$$

e  $\gamma$  é solução da equação

$$\sum_{k=1}^n \text{Exp} \left( \frac{-\gamma + s_k^t}{w_k} \right) = e$$

Lema 2.1:

Para cada  $\phi \in \Phi$  fixo

$$\begin{aligned} \max_{P \in \bar{\mathcal{P}}^n} [K(X,P) - J(X/Y,P,\phi)] &= K(X,P^*) - J(X/Y,P^*,\phi) = \\ &= \frac{1}{e} \sum_{i=1}^n [(\gamma + w_i) \text{Exp} \left( \frac{-\gamma + s_i^t}{w_i} \right)] \leq C_w \end{aligned} \quad (2.23)$$

$$\text{onde} \quad s_i^t = \sum_{j=1}^m w(i,j) p(j/i) \log \phi(i/j) \quad (2.24)$$

e  $\gamma$  é solução da equação:

$$\sum_{i=1}^n \text{Exp} \left( \frac{-\gamma + s_i}{w_i} \right) = e \quad (2.25)$$

com  $P^* = (p_1^*, p_2^* \dots p_n^*) \in \bar{\mathcal{P}}^n$  dado por

$$p_i^* = \text{Exp} \left[ \frac{-\gamma}{w_i} - 1 + \frac{s_i}{w_i} \right], \quad i = 1, 2, \dots, n.$$

Teorema 2.3:

Seja  $p^1 \in \mathcal{P}^n$ . Então os valores  $C(t, t)$  definidos em (2.20) convergem monotonicamente para baixo da capacidade  $C_w$  quando  $t \rightarrow \infty$ .

Demonstração:

Da etapa iii), seja o vetor probabilidade definido em (2.21)

$$p_i^{t+1} = \text{Exp} \left[ \frac{-\gamma}{w_i} - 1 + \frac{s_i^t}{w_i} \right].$$

Usando (2.19) e (2.22) podemos reescrever

$$p_i^{t+1} = p_i^t \frac{1}{e} \text{Exp} \left\{ \frac{-\gamma + \sum_{j=1}^n w(i, j) p(j/i) \log \frac{p(j/i)}{\sum_{k=1}^n p(j/k) p_k^t}}{w_i} \right\} \quad (2.26)$$

Seja também

$$C(t+1, t) = \max_{P \in \bar{\mathcal{P}}^n} [K(X, P) - J(X/Y, P, \phi^t)]$$

ou por (2.23)

$$C(t+1, t) = K(X, P^{t+1}) - J(X/Y, P^{t+1}, \phi^t) \quad (2.27)$$

Então, a seguir, pelo Lema 1, temos

$$C(t+1, t) = \frac{1}{e} \sum_{i=1}^n [(\gamma + w_i) \text{Exp} \left( \frac{-\gamma + s_i^t}{w_i} \right)] \quad (2.28)$$

ou ainda

$$C(t+1, t) = \frac{\gamma}{e} \sum_{i=1}^n \text{Exp} \left( \frac{-\gamma + s_i^t}{w_i} \right) + \frac{1}{e} \sum_{i=1}^n w_i \text{Exp} \left( \frac{-\gamma + s_i^t}{w_i} \right)$$

Usando (2.25) e (2.21), temos

$$C(t+1, t) = \gamma + \sum_{i=1}^n w_i p_i^{t+1} \quad (2.29)$$

Além disso, pelas definições de  $C(t, t)$  e  $C(t+1, t)$  segue que:

$$C(1, 1) \leq C(2, 1) \leq C(2, 2) \leq \dots \leq C(t, t) \leq C(t+1, t) \leq \dots \leq C_w \quad (2.30)$$

Seja  $P^0 = (p_1^0, p_2^0, \dots, p_n^0) \in \bar{P}^n$  um dos vetores probabilidades de entrada que maximizam a informação comunicada pelo canal, isto é:

$$B(X/Y, P^0) = C_w \quad (2.31)$$

Então, usando as notações

$$q_j^0 = \sum_{i=1}^n p(j/i) p_i^0 \quad (2.32)$$

$$q_j^t = \sum_{i=1}^n p(j/i) p_i^t \quad (2.33)$$

e levando em consideração (2.29) e o teorema 2.2 temos:

$$\begin{aligned} & \sum_{k=1}^n w_k p_k^0 \log \frac{p_k^{t+1}}{p_k^t} = \\ & = \sum_{k=1}^n w_k p_k^0 \log \left\{ \frac{p_k^t \frac{1}{e} \text{Exp} \left( \frac{-\gamma + \sum_{j=1}^m w(k,j) p(j/k) \log \frac{p(j/k)}{q_j^t}}{w_k} \right)}{p_k^t} \right\} \end{aligned}$$

Por (2.26)

$$\begin{aligned} & = - \sum_{k=1}^n w_k p_k^0 + \sum_{k=1}^n w_k p_k^0 \left\{ \frac{-\gamma + \sum_{j=1}^m w(k,j) p(j/k) \log \frac{p(j/k)}{q_j^t}}{w_k} \right\} \\ & = - \sum_{k=1}^n w_k p_k^0 - \gamma + \sum_{k=1}^n \sum_{j=1}^m w(k,j) p(j/k) p_k^0 \log \frac{p(j/k)}{q_j^t} \cdot \frac{q_j^0}{q_j^0} \\ & = - \sum_{k=1}^n w_k p_k^0 + \gamma + \sum_{k=1}^n \sum_{j=1}^m w(k,j) p(j/k) p_k^0 \log \frac{p(j/k)}{q_j^0} + \\ & \quad + \sum_{k=1}^n \sum_{j=1}^m w(k,j) p(j/k) p_k^0 \log \frac{q_j^0}{q_j^t} \\ & = - \sum_{k=1}^n w_k p_k^0 + \sum_{k=1}^n w_k p_k^{t+1} - C(t+1, t) + C_w + \\ & \quad + \sum_{k=1}^n \sum_{j=1}^m w(k,j) p(j/k) p_k^0 \log \frac{q_j^0}{q_j^t} \quad (2.34) \end{aligned}$$

Usando (2.29) e (2.31), Sabemos que  $\sum_{k=1}^n w_k p_k^{t+1}$  é não negativo. Considerando a concavidade da função  $B(X/Y, P)$  e que por (2.30)

$$p_i^t \leq p_i^{t+1} \quad \text{podemos afirmar que}$$

$$p_i^t \leq p_i^0 \quad \text{para qualquer } t = 1, 2, \dots$$

Logo:

$$p(j/i)p_i^t \leq p(j/i)p_i^0 \quad (p(j/i) \geq 0) \quad e$$

$$\sum_{i=1}^n p(j/i)p_i^t \leq \sum_{i=1}^n p(j/i)p_i^0 \quad \text{ou } q_j^t \leq q_j^0$$

isto implica que  $\frac{q_j^0}{q_j^t} \geq 1$  e  $\log \frac{q_j^0}{q_j^t} \geq 0$

Portanto:  $\sum_{k=1}^n \sum_{j=1}^m w(k,j) p(j/k)p_k^0 \log \frac{q_j^0}{q_j^t}$  é não negativo.

Conseqüentemente (2.34) fica:

$$\sum_{k=1}^n w_k p_k^0 \log \frac{p_k^{t+1}}{p_k^t} \geq - \sum_{k=1}^n w_k p_k^0 - C(t+1, t) + C_w \quad (2.35)$$

ou ainda:

$$\sum_{k=1}^n w_k p_k^0 \log \frac{e^{p_k^{t+1}}}{p_k^t} \geq C_w - C(t+1, t) \quad (2.36)$$

Obviamente, esta desigualdade vale para cada  $t = 1, 2, \dots$ , fazendo o somatório de (2.36) para  $t = 1, 2, \dots, N$  temos

$$\sum_{t=1}^N [C_w - C(t+1, t)] \leq \sum_{t=1}^N \sum_{k=1}^n w_k p_k^0 \log \frac{e^{p_k^{t+1}}}{p_k^t}$$

$$\begin{aligned}
&= \sum_{k=1}^n w_k p_k^0 + \sum_{k=1}^n w_k p_k^0 \log p_k^2 - \sum_{k=1}^n w_k p_k^0 \log p_k^1 + \sum_{k=1}^n w_k p_k^0 \log p_k^3 \\
&- \sum_{k=1}^n w_k p_k^0 \log p_k^2 + \dots + \sum_{k=1}^n w_k p_k^0 \log p_k^{N+1} - \\
&- \sum_{k=1}^n w_k p_k^0 \log p_k^N
\end{aligned}$$

Fazendo as simplificações temos:

$$\sum_{t=1}^N [C_w - C(t+1, t)] < \sum_{k=1}^n w_k p_k^0 + \sum_{k=1}^n w_k p_k^0 \log \frac{p_k^{N+1}}{p_k^1} \quad (2.37)$$

$$\begin{aligned}
\sum_{t=1}^N [C_w - C(t+1, t)] &< \sum_{k=1}^n w_k p_k^0 \log \frac{e^{p_k^{N+1}}}{p_k^1} \\
&< \sum_{k=1}^n w_k p_k^0 \log \frac{e^{p_k^0}}{p_k^1} \quad (2.38)
\end{aligned}$$

Para cada inteiro  $N \geq 1$ .

Notar que o lado direito de (2.38) é finito e constante para  $p^1 \in \mathcal{P}^n$ . Desde que o valor  $C_w - C(t+1, t)$  é não negativo e não crescente com acréscimo  $t$ , temos claramente que:

$$\lim_{t \rightarrow \infty} C(t+1, t) = \lim_{t \rightarrow \infty} C(t, t) = C_w \quad (2.39)$$

Corolário 2.1:

$$\lim_{N \rightarrow \infty} q_j^N = q_j^0 \quad \text{onde} \quad q_j = \sum_{i=1}^n p(j/i) p_i \quad (2.40)$$

### 2.3 - Limites Superior e Inferior da Capacidade

Vamos mostrar, agora, algumas propriedades de  $C_w$ . Em particular, daremos fórmulas aproximadas que dão os limites superior e inferior da capacidade.

Seja:  $C_{w, \phi} = \max_{P \in \mathcal{P}^n} [K(X, P) - J(X/Y, P, \phi)]$

$$= \frac{1}{e} \sum_{i=1}^n [(\gamma + w_i) \text{Exp} \left( \frac{-\gamma + s_i}{w_i} \right)] \leq C_w \quad (2.41)$$

(pelo Lema 2.1)

onde  $s_i = \sum_{j=1}^m w(i, j) p(j/i) \log \phi(i/j)$  e  $\gamma$  é solução da equa-

$$\text{ção } \sum_{i=1}^n \text{Exp} \left( \frac{-\gamma + s_i}{w_i} \right) = e$$

Pelo teorema 2.2.

$$\max_{\phi \in \Phi} C_{w, \phi} = C_w \quad (2.42)$$

#### Teorema 2.4:

Sejam  $M_1$  e  $M_2$  matrizes  $m \times n$  canais,  $\alpha$  um número arbitrário tal que  $0 \leq \alpha \leq 1$  e  $\phi$  uma  $n \times m$  matriz arbitrária, então a seguinte desigualdade vale:

$$C(\alpha M_1 + (1 - \alpha) M_2, \phi) \leq \alpha C(M_1, \phi) + (1 - \alpha) C(M_2, \phi) \quad (2.43)$$

#### Corolário 2.2:

Com as condições do teorema 2.4, podemos escrever que:

$$C(\alpha M_1 + (1 - \alpha)M_2) \leq \alpha C(M_1) + (1 - \alpha) C(M_2). \quad (2.44)$$

Teorema 2.5:

$$C_w \geq \frac{1}{n} \log n \left( \sum_{i=1}^n w_i \right) - \frac{\sum_{i=1}^n w_i - \Delta}{n} \log (n-1) \\ - K \left( \frac{\sum_{i=1}^n w_i}{n}, \frac{\sum_{i=1}^n w_i}{n}, \frac{\Delta}{\sum_{i=1}^n w_i}, 1 - \frac{\Delta}{\sum_{i=1}^n w_i} \right), \quad (2.45)$$

onde  $\Delta = w_{i_j}$  e  $i_j$  significa um dos inteiros arbitrários escolhidos de 1 a n correspondentes a cada j e

$$K \left( \frac{\sum_{i=1}^n w_i}{n}, \frac{\sum_{i=1}^n w_i}{n}, \frac{\Delta}{\sum_{i=1}^n w_i}, 1 - \frac{\Delta}{\sum_{i=1}^n w_i} \right) \text{ é a entropia ponderada}$$

já definida e dada por

$$K(X) = K(w_1, w_2, p_1, p_2) = - \sum_{i=1}^2 w_i p_i \log p_i .$$

Teorema 2.6:

$$C_w \geq \frac{1}{e} \sum_{i=1}^n [(\gamma + w_i) \text{Exp} \left\{ \frac{-\gamma + \sum_{j=1}^m (w(i,j) p(j/i) \log \frac{p(j/i)}{\sum_{k=1}^n p(j/k)})}{w_i} \right\} ,$$

onde  $\gamma$  foi definida em (2.22).

Teorema 2.7:

$$C_w \geq \frac{2}{n} \log n \left( \sum_{i=1}^n w_i \right) -$$

$$- \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m w(i,j) p(j/i) \log \sum_{k=1}^n p(j/k) - K(X,Y) \quad (2.46)$$

onde  $K(X,Y)$  é a entropia ponderada conjunta definida em (1.3)

com  $p(i,j) > 0$ ,  $\sum_{i=1}^n \sum_{j=1}^m p(i,j) = 1$ ,  $w(i,j) \geq 0$ ,  $i = 1, 2, \dots, n$ ,  
 $j = 1, 2, \dots, m$ ,

ou

$$K(X,Y) = - \sum_{i=1}^n \sum_{j=1}^m w(i,j) p(i,j) \log p(i,j)$$

Teorema 2.8:

$$C_w \leq w_i \log n + \max_{p_i} \left[ \sum_{j=1}^m w(i,j) p(j/i) \log \frac{p(j/i)}{\sum_{k=1}^n p(j/k)} \right]$$

(2.47)

## CAPÍTULO III

UM MÉTODO DE ITERAÇÃO PARA CALCULAR A  
CAPACIDADE RELATIVA PONDERADA3.1 - Introdução

A capacidade de um canal arbitrário, discreto, sem memória e conhecida como a máxima velocidade de transmissão da informação.

MUROGA (1953)<sup>[11]</sup>, CHENG (1974)<sup>[5]</sup>, TAKANO (1975)<sup>[15]</sup> e outros, estudaram métodos para avaliar a capacidade, enquanto ARIMOTO (1972)<sup>[1]</sup> e BLAHUT (1972)<sup>[4]</sup> independentemente apresentaram o mesmo método de iteração para calculá-la.

É um método simples e sistemático e aparenta ser importante por considerar o caso onde não podemos descuidar a incidência do fator custo para a transmissão dos símbolos de entrada.

Em tais casos, seria melhor dar atenção ao máximo da velocidade de transmissão da informação por custo que é a capacidade relativa, a qual foi introduzida por REZA (1961)<sup>[12]</sup>.

O método de iteração que apresentaremos neste capítulo destina-se a calcular a capacidade relativa ponderada de um canal discreto sem memória, isto é: a máxima velocidade de transmissão da informação (com pesos) por custo.

O algoritmo tem por base o trabalho apresentado por JIMBO e KUNISAWA<sup>[9]</sup> no qual introduzimos a velocidade de transmissão da informação com ponderação que foi apresentada por FLEMMING, Diva M.<sup>[6]</sup>. Mostraremos também que a con

vergência do método é monótona não decrescente.

A seguir será apresentado um método de iteração para calcular a capacidade relativa ponderada. Este método é simples e bem tratável e contém o método de Jimbo e Kunisawa como um caso especial.

### 3.2 - Capacidade Relativa Ponderada

Um canal discreto, sem memória, com  $n$  símbolos de entrada e  $m$  símbolos de saída é caracterizado por uma matriz de transição  $n \times m$ :

$$M = [p(j/i)] \quad (i = 1, 2, \dots, n ; j = 1, 2, \dots, m) \quad \text{onde}$$

$$p(j/i) \geq 0 \quad \text{e} \quad \sum_{j=1}^m p(j/i) = 1.$$

Cada entrada  $p(j/i)$  representa a probabilidade condicional para receber o  $j$ -ésimo símbolo de saída se o  $i$ -ésimo símbolo de entrada foi transmitido. Supomos que a matriz de transição  $M$  não contenha colunas nulas, isto é: não existam símbolos de saída que nunca foram recebidos. Vamos denotar por  $\mathcal{P}^k$  e  $\bar{\mathcal{P}}^k$  dois tipos de vetores probabilidades de entrada como definidos no Cap. II.

Vamos definir a velocidade de transmissão da informação (ou informação comunicada pelo canal)  $B_w(P)$  para o canal  $M$  como:

$$B_w(P) = \sum_{i=1}^n \sum_{j=1}^m w(i,j) p(j/i) p_i \log \frac{p(j/i)}{q_j} \quad (3.1)$$

$$\text{com } q_j = \sum_{i=1}^n p_i p(j/i), \quad P \in \mathcal{P}^k$$

onde  $B_w(P) = B(X/Y)$  (Ver Ref. (1.16'))

Por conveniência definimos  $0 \log \frac{0}{0} = 0 \log 0 = 0$

Para facilitar a notação, seja:

$$D_{i,w}(P) = \sum_{j=1}^m w(i,j) p(j/i) \log \frac{p(j/i)}{q_j}, \quad (3.2)$$

onde o vetor  $Q = (q_1, q_2, \dots, q_m)$  é o vetor probabilidade de saída correspondente ao vetor probabilidade de entrada  $P = (p_1, p_2, \dots, p_n)$ . Usando esta notação podemos representar a velocidade de transmissão da informação como:

$$B_w(P) = \sum_{i=1}^n p_i D_{i,w}(P) \quad (3.3)$$

A discriminação ponderada de dois vetores probabilidades

$R = (r_1, r_2, \dots, r_t)$  e  $P = (p_1, p_2, \dots, p_t)$  é

$$D_w(R,P) = \sum_{i=1}^t w_i r_i \log \frac{r_i}{p_i} \quad (3.4)$$

onde os  $w_i$  são escolhidos convenientemente de forma que  $D_w(R,P)$  é não negativa e convexa (pode ser provado através de linha similar à discriminação de KULLBACK) Ref. KULLBACK<sup>[10]</sup>.

Então  $D_{i,w}(P)$  é, também, uma discriminação ponderada de  $p(j/i)$  e  $Q$ .

Agora, introduziremos um fator custo  $l_i > 0$  ao  $i$ -ésimo símbolo de entrada para algum  $i$ . Então uma função custo  $C_M(P)$  está definida por:

$$C_M(P) = \sum_{i=1}^n p_i l_i w_i \quad \text{para } P \in \bar{P}^k \quad (3.5)$$

Evidentemente,  $C_M(P) > 0$  para algum  $P \in \bar{\rho}^k$ .

Definimos a capacidade relativa ponderada  $C_{W,M}$  por:

$$C_{W,M} = \max_{P \in \bar{\rho}} \frac{B_W(P)}{C_M(P)}, \quad (3.6)$$

isto é, a capacidade relativa ponderada está interpretada como o máximo da velocidade de transmissão da informação por custo.

No caso quando cada custo  $l_i = 1$  ( $i = 1, 2, \dots, n$ ) e cada peso  $w_i = 1$  ( $i = 1, 2, \dots, n$ ), o custo pensado  $C_M(P) = 1$ . Assim a capacidade relativa ponderada coincide com a capacidade de SHANNON [13].

Daremos um método de iteração para calcular a capacidade relativa ponderada  $C_{W,M}$  diretamente. O procedimento é o seguinte:

### 3.3 - Procedimento de Iteração e Convergência

- i) Seja  $b = \min_i l_i w_i$
- ii) Inicialmente, escolhemos um vetor probabilidade arbitrário  $P^0 \in \bar{\rho}^k$ .
- iii) Após o  $k$ -ésimo vetor probabilidade  $P^k$  ter sido obtido, usando-o, construir o  $(k+1)$ -ésimo vetor probabilidade  $P^{k+1}$  como segue:

$$\tilde{p}_i^{k+1} = p_i^k \text{Exp} [b D_{i,W}(P^k) / l_i w_i] \quad i = 1, 2, \dots, n$$

$$T^k = \sum_{i=1}^n \tilde{p}_i^{k+1}$$

$$p_i^{k+1} = \tilde{p}_i^{k+1} / T^k \quad i = 1, 2, \dots, n.$$

então podemos mostrar o seguinte teorema:

Teorema:

Para a seqüência dos vetores probabilidades  $\{P^k\}$   $k = 0, 1, 2, \dots$  construída de acordo com o procedimento de iteração dado acima, a seqüência:  $\{B_W(P^k)/C_M(P^k)\}$  é monótona não decrescente e converge para a capacidade relativa ponderada  $C_{W,M}$ .

A demonstração deste teorema está baseada, principalmente, em duas etapas. A primeira sobre a monotonicidade do procedimento e a segunda sobre a convergência.

Primeiramente, mostraremos a monotonicidade do procedimento. Denotaremos por  $P$  e  $R$  o  $k$ -ésimo e o  $(k+1)$ -ésimo vetores probabilidades do procedimento de iteração respectivamente.

Sejam:

$$x_{i,w} = x_{i,w}(P) = D_{i,w}(P) / l_i^{w_i} \quad (3.7)$$

$$f_{i,w} = \text{Exp} [b x_{i,w}] \quad i = 1, 2, \dots, n \quad (3.8)$$

$$T = \sum_{i=1}^n p_i f_{i,w} \quad (3.9)$$

Obviamente temos  $r_i = p_i f_{i,w} / T$ . Definimos o vetor probabilidade  $\mu_w = (\mu_{1,w}, \mu_{2,w}, \dots, \mu_{n,w})$  por

$$\mu_{i,w} = \frac{p_i l_i^{w_i}}{C_M(P)} \quad (3.10)$$

O lema a seguir garante a monotonicidade do procedi-

mento.

Lema:

$$\text{Temos: } \frac{B_w(P)}{C_M(P)} < \frac{\log \theta}{b} < \frac{B_w(R)}{C_M(R)}$$

$$\text{onde } \theta = \sum_{i=1}^n \mu_{i,w} f_{i,w}$$

Demonstração:

Para mostrar que  $\frac{B_w(P)}{C_M(P)} < \frac{\log \theta}{b}$  basta aplicar a desigualdade de JENSEN e temos:

$$\begin{aligned} \frac{\log \theta}{b} &= \frac{\log \sum_{i=1}^n \mu_{i,w} f_{i,w}}{b} > \frac{1}{b} \sum_{i=1}^n \mu_{i,w} \log f_{i,w} = \\ &= \frac{1}{b} \sum_{i=1}^n \frac{p_i l_i w_i}{C_M(P)} \log \text{Exp} [b x_{i,w}] = \frac{1}{b} \sum_{i=1}^n \frac{p_i l_i w_i}{C_M(P)} \cdot b x_{i,w} \\ &= \sum_{i=1}^n \frac{p_i l_i w_i}{C_M(P)} \cdot \frac{D_{i,w}(P)}{l_i w_i} = \frac{\sum_{i=1}^n p_i D_{i,w}(P)}{C_M(P)} = \frac{B_w(P)}{C_M(P)} \end{aligned}$$

Por (3.10), (3.8), (3.7) e (3.3).

$$\text{O que resulta em: } \frac{B_w(P)}{C_M(P)} < \frac{\log \theta}{b}$$

Para mostrar que  $\frac{\log \theta}{b} < \frac{B_w(R)}{C_M(R)}$  é suficiente mostrar

que:

$$G = b B_w(R) - C_M(R) \log \theta > 0 \quad (3.11)$$

Seja  $S = (s_1, s_2, \dots, s_m)$  um vetor probabilidade de

saída correspondente ao vetor probabilidade de entrada

$R = (r_1, r_2, \dots, r_n)$  isto é:

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

Da convexidade da função  $f(z) = z \log z$  ( $z \geq 0$ ) podemos mostrar a seguinte desigualdade (similar ao mostrado por Kullback), Ref. KULLBACK [10].

$$D_w(R, P) \geq D_w(S, Q) \quad (3.12)$$

Então temos:

$$B_w(R) = \sum_{i=1}^n \sum_{j=1}^m w(i, j) r_i p(j/i) \log \frac{p(j/i)}{s_j} \cdot \frac{q_j}{q_j}$$

$$B_w(R) = \sum_{i=1}^n \sum_{j=1}^m w(i, j) r_i p(j/i) \log \frac{p(j/i)}{q_j} -$$

$$- \sum_{i=1}^n \sum_{j=1}^m w(i, j) r_i p(j/i) \log \frac{s_j}{q_j}$$

$$B_w(R) = \sum_{i=1}^n r_i D_{i,w}(P) -$$

$$- \sum_{i=1}^n \sum_{j=1}^m w(i, j) r_i p(j/i) \log \frac{s_j}{q_j} \quad \text{por (3.2)}$$

Mas  $r_i p(j/i) = s_j p(i/j)$

$$B_w(R) = \sum_{i=1}^n r_i D_{i,w}(P) -$$

$$- \sum_{i=1}^n \sum_{j=1}^m w(i, j) s_j p(i/j) \log \frac{s_j}{q_j}$$

$$B_w(R) = \sum_{i=1}^n r_i D_{i,w}(P) - \sum_{j=1}^m w_j s_j \log \frac{s_j}{q_j} \quad \text{por (1.12)}$$

$$B_w(R) = \sum_{i=1}^n r_i D_{i,w}(P) - D_w(S,Q) \quad \text{por (3.4)}$$

$$B_w(R) \geq \sum_{i=1}^n r_i D_{i,w}(P) - D_w(R,P) \quad \text{por (3.12)}$$

$$B_w(R) \geq \sum_{i=1}^n r_i x_{i,w} l_{i,w} - D_w(R,P) \quad \text{por (3.7)}$$

$$B_w(R) \geq \sum_{i=1}^n r_i l_{i,w} \frac{\log f_{i,w}}{b} - D_w(R,P) \quad \text{por (3.8)}$$

Então (3.11) fica:

$$G \geq b \left[ \frac{1}{b} \sum_{i=1}^n r_i l_{i,w} \log f_{i,w} - D_w(R,P) \right] - C_M(R) \log \theta \geq 0$$

Mas 
$$\theta = \sum_{i=1}^n \mu_{i,w} f_{i,w} = \sum_{i=1}^n \frac{p_i l_{i,w}}{C_M(P)} \cdot \frac{T r_i}{P_i} \quad \text{por (3.10)}$$

e temos 
$$\theta = \frac{T \sum_{i=1}^n r_i l_{i,w}}{C_M(P)} = \frac{T C_M(R)}{C_M(P)} \quad \text{por (3.5)}$$

Então:

$$\begin{aligned} G &\geq b \left[ \frac{1}{b} \sum_{i=1}^n r_i l_{i,w} \log f_{i,w} - D_w(R,P) \right] - C_M(R) \log \frac{T C_M(R)}{C_M(P)} \\ &= \sum_{i=1}^n r_i l_{i,w} \log \frac{f_{i,w}}{b} - b D_w(R,P) - C_M(R) \log \frac{C_M(R)}{C_M(P)} - \\ &\quad - C_M(R) \log T \end{aligned}$$

$$= \sum_{i=1}^n r_i l_i w_i \log \frac{f_{i,W}}{T} - b D_W(R,P) - C_M(R) \log \frac{C_M(R)}{C_M(P)} \geq 0$$

Portanto: usando (3.4) e como  $r_i = p_i f_{i,W}/T$

Temos:

$$G \geq \sum_{i=1}^n r_i l_i w_i \log \frac{r_i}{p_i} - b \sum_{i=1}^n w_i r_i \log \frac{r_i}{p_i} - C_M(R) \log \frac{C_M(R)}{C_M(P)} \geq 0$$

$$G \geq \sum_{i=1}^n r_i w_i (l_i - b) \log \frac{r_i}{p_i} - C_M(R) \log \frac{C_M(R)}{C_M(P)} \geq 0$$

$$G \geq \sum_{i=1}^n r_i w_i (l_i - b) \log \frac{\sum_{i=1}^n r_i w_i (l_i - b)}{\sum_{i=1}^n p_i w_i (l_i - b)} - C_M(R) \log \frac{C_M(R)}{C_M(P)} \geq 0$$

(por convexidade da discriminação de KULLBACK).

O que resulta:

$$G \geq [C_M(R) - c] \log \frac{[C_M(R) - c]}{[C_M(P) - d]} - C_M(R) \log \frac{C_M(R)}{C_M(P)}$$

onde  $c = \sum_{i=1}^n r_i w_i b$  e  $d = \sum_{i=1}^n p_i w_i b$ .

Usando o resultado de [9], podemos afirmar que  $G \geq 0$ , desde que  $C_M(R) \geq c > 0$  e  $C_M(P) \geq d > 0$ , a qual é válida, obviamente.

Agora mostraremos a convergência do procedimento de iteração

Demonstração do Teorema:

Sejam:

$$x_{i,w}^k = D_{i,w}(P^k) / l_{i,w_i} \quad (3.13)$$

$$f_{i,w}^k = \text{Exp} [b x_{i,w}^k] \quad (3.14)$$

$$\mu_{i,w}^k = \frac{p_{i,w_i}^k}{C_M(P^k)} \quad (i = 1, 2, \dots, n, \quad k = 1, 2, \dots) \quad (3.15)$$

Também sejam:

$$T^k = \sum_{i=1}^n p_i^k f_{i,w}^k \quad (3.16)$$

$$\theta^k = \sum_{i=1}^n \mu_{i,w}^k f_{i,w}^k \quad (3.17)$$

Então pela definição:

$$\begin{aligned} p_i^{k+1} &= \tilde{p}_i^{k+1} / T^k = p_i^k \text{Exp} [b D_{i,w}(P^k) / l_{i,w_i}] / T^k \\ &= p_i^k \text{Exp} [b x_{i,w}^k] / T^k = p_i^k f_{i,w}^k / T^k \end{aligned} \quad (3.18)$$

Do Lema anterior, temos

$$\frac{B_w(P^k)}{C_M(P^k)} \leq \frac{\log \theta^k}{b} \leq \frac{B_w(R^k)}{C_M(R^k)} \leq C_{w,M} \quad k = 0, 1, 2, \dots$$

Então ambas as seqüências  $\{B_w(P^k)/C_M(P^k)\}$  e  $\{\log \theta^k/b\}$  são monótonas, não decrescentes e convergem para o mesmo valor. Denotaremos por  $P^*$  um certo vetor probabilidade que atinge a capa-

cidade relativa ponderada e seja  $\mu_{i,w}^* = p_i^* l_{i,w} / C_M(P^*)$ . Ainda mais, denotamos por  $Q^k$  ( $k = 0, 1, 2, \dots$ ) e  $Q^*$  os vetores probabilidades de saída correspondentes aos vetores probabilidades de entrada  $P^k$  ( $k = 0, 1, 2, \dots$ ) e  $P^*$  respectivamente. Também denotamos

$$D(\mu_w^*, \mu_w^k) = \sum_{i=1}^n \mu_{i,w}^* \log \frac{\mu_{i,w}^*}{\mu_{i,w}^k}$$

Considere: 
$$\frac{\mu_{i,w}^{k+1}}{\mu_{i,w}^k} = \frac{f_{i,w}^k}{\theta^k} . \quad (3.19)$$

Temos:

$$\begin{aligned} D(\mu_w^*, \mu_w^k) - D(\mu_w^*, \mu_w^{k+1}) &= \\ &= \sum_{i=1}^n \mu_{i,w}^* \log \frac{\mu_{i,w}^*}{\mu_{i,w}^k} - \sum_{i=1}^n \mu_{i,w}^* \log \frac{\mu_{i,w}^*}{\mu_{i,w}^{k+1}} \\ &= \sum_{i=1}^n \mu_{i,w}^* \log \mu_{i,w}^* - \sum_{i=1}^n \mu_{i,w}^* \log \mu_{i,w}^k - \\ &\quad - \sum_{i=1}^n \mu_{i,w}^* \log \mu_{i,w}^* + \sum_{i=1}^n \mu_{i,w}^* \log \mu_{i,w}^{k+1} \\ &= \sum_{i=1}^n \mu_{i,w}^* \log \frac{\mu_{i,w}^{k+1}}{\mu_{i,w}^k} = \sum_{i=1}^n \mu_{i,w}^* \log \frac{f_{i,w}^k}{\theta^k} \\ &= \sum_{i=1}^n \mu_{i,w}^* \log f_{i,w}^k - \sum_{i=1}^n \mu_{i,w}^* \log \theta^k \\ &= \sum_{i=1}^n \mu_{i,w}^* \log f_{i,w}^k - \log \theta^k \end{aligned}$$

$$= \sum_{i=1}^n \mu_{i,w}^* \log [\text{Exp} (b x_{i,w}^k)] - \log \theta^k \quad (\text{Por } 3.14)$$

$$= b \sum_{i=1}^n \mu_{i,w}^* x_{i,w}^k - \log \theta^k$$

$$= b \sum_{i=1}^n \frac{p_{i,w}^*}{C_M(P^*)} \cdot \frac{D_{i,w}(P^k)}{l_i^{w_i}} - \log \theta^k \quad \text{Por (3.15) e (3.13)}$$

$$= \frac{b}{C_M(P^*)} \sum_{i=1}^n p_i^* D_{i,w}(P^k) - \log \theta^k$$

Mas

$$\sum_{i=1}^n p_i^* D_{i,w}(P^k) = \sum_{i=1}^n p_i^* \sum_{j=1}^m w(i,j) p(j/i) \log \frac{p(j/i)}{q_j^k} \cdot \frac{q_j^*}{q_j^*}$$

$$= \sum_{i=1}^n p_i^* \sum_{j=1}^m w(i,j) p(j/i) \log \frac{p(j/i)}{q_j^*} +$$

$$+ \sum_{i=1}^n p_i^* \sum_{j=1}^m w(i,j) p(j/i) \log \frac{q_j^*}{q_j^k}$$

$$= \sum_{i=1}^n p_i^* D_{i,w}(P^*) + \sum_{i=1}^n \sum_{j=1}^m q_j^* p(i/j) w(i,j) \log \frac{q_j^*}{q_j^k} \text{ pois}$$

$$p_i^* p(j/i) = q_j^* p(i/j)$$

$$= \sum_{i=1}^n p_i^* D_{i,w}(P^*) + \sum_{j=1}^m w_j q_j^* \log \frac{q_j^*}{q_j^k} \quad \text{por (1.12)}$$

$$= \sum_{i=1}^n p_i^* D_{i,w}(P^*) + D_w(Q^*, Q^k) \quad \text{por (3.4)}$$

Portanto

$$\begin{aligned}
 D(\mu_w^*, \mu_w^k) - D(\mu_w^*, \mu_w^{k+1}) &= \\
 &= \frac{b}{C_M(P^*)} \left[ \sum_{i=1}^n p_i^* D_{i,w}(P^*) + D_w(Q^*, Q^k) \right] - \log \theta^k \\
 &= b C_{w,M} + \frac{b}{C_M(P^*)} D_w(Q^*, Q^k) - \log \theta^k \\
 &= b C_{w,M} - \log \theta^k + \frac{b}{C_M(P^*)} D_w(Q^*, Q^k) \quad (3.20)
 \end{aligned}$$

Como  $\frac{b}{C_M(P^*)} D_w(Q^*, Q^k)$  é não negativo, temos

$$D(\mu_w^*, \mu_w^k) - D(\mu_w^*, \mu_w^{k+1}) \geq b C_{w,M} - \log \theta^k \quad (\geq 0) \quad \text{ou}$$

$$\frac{1}{b} [D(\mu_w^*, \mu_w^k) - D(\mu_w^*, \mu_w^{k+1})] \geq C_{w,M} - \frac{\log \theta^k}{b} \quad (\geq 0)$$

Adicionando estas desigualdades de  $k = 0$  até  $k = N-1$

$$\sum_{k=0}^{N-1} \left[ C_{w,M} - \frac{\log \theta^k}{b} \right] \leq \sum_{k=0}^{N-1} \frac{1}{b} [D(\mu_w^*, \mu_w^k) - D(\mu_w^*, \mu_w^{k+1})]$$

ou

$$\sum_{k=0}^{N-1} \left[ C_{w,M} - \frac{\log \theta^k}{b} \right] \leq \frac{1}{b} [D(\mu_w^*, \mu_w^0) - D(\mu_w^*, \mu_w^N)]$$

ou

$$0 \leq \sum_{k=0}^{N-1} \left[ C_{w,M} - \frac{\log \theta^k}{b} \right] \leq \frac{1}{b} D(\mu_w^*, \mu_w^0) \quad N = 1, 2, \dots$$

Como o lado direito da desigualdade é finito e independente de  $N$ , a seqüência  $\{\log \theta^k/b\}$  deve convergir para a ca

pacidade relativa ponderada  $C_{w,M}$ . O que completa a demonstração do teorema.

Corolário:

A seqüência de saída  $\{Q^k\}$  correspondente a  $\{P^k\}$  converge.

Demonstração:

Da igualdade (3.20) temos

$$0 \leq \frac{b}{C_M(P^*)} D_w(Q^*, Q^k) \leq D(\mu_w^*, \mu_w^k) - D(\mu_w^*, \mu_w^{k+1})$$

$$k = 0, 1, 2, \dots$$

Adicionando esta desigualdade de  $k = 0$  até  $k = N - 1$  temos:

$$0 \leq \sum_{k=0}^{N-1} \frac{b}{C_M(P^*)} D_w(Q^*, Q^k) \leq \sum_{k=0}^{N-1} [D(\mu_w^*, \mu_w^k) - D(\mu_w^*, \mu_w^{k+1})]$$

$$0 \leq \sum_{k=0}^{N-1} \frac{b}{C_M(P^*)} D_w(Q^*, Q^k) \leq D(\mu_w^*, \mu_w^0) - D(\mu_w^*, \mu_w^N) \quad \text{ou}$$

$$0 \leq \sum_{k=0}^{N-1} \frac{b}{C_M(P^*)} D_w(Q^*, Q^k) \leq D(\mu_w^*, \mu_w^0) \quad \text{ou}$$

$$0 \leq \sum_{k=0}^{N-1} D_w(Q^*, Q^k) \leq \frac{C_M(P^*)}{b} D(\mu_w^*, \mu_w^0)$$

Como  $\frac{C_M(P^*)}{b} D(\mu_w^*, \mu_w^0)$  é finito e independente de  $N$ , temos que

$\sum_{k=0}^{N-1} D(Q^*, Q^k)$  converge, então

$$\lim_{k \rightarrow \infty} D(Q^*, Q^k) = 0 \quad \text{ou} \quad \lim_{k \rightarrow \infty} \sum_{j=1}^m q_j^* \log \frac{q_j^*}{q_j^k} = 0$$

ou

$$\lim_{k \rightarrow \infty} \left[ \sum_{j=1}^m q_j^* \log q_j^* - \sum_{j=1}^m q_j^* \log q_j^k \right] = 0$$

ou

$$\lim_{k \rightarrow \infty} \sum_{j=1}^m q_j^* \log q_j^* - \lim_{k \rightarrow \infty} \sum_{j=1}^m q_j^* \log q_j^k = 0$$

ou

$$\lim_{k \rightarrow \infty} \sum_{j=1}^m q_j^* \log q_j^k = \lim_{k \rightarrow \infty} \sum_{j=1}^m q_j^* \log q_j^*$$

ou

$$\lim_{k \rightarrow \infty} \sum_{j=1}^m q_j^* \log q_j^k = \sum_{j=1}^m q_j^* \log q_j^*$$

ou

$$\sum_{j=1}^m q_j^* \log \lim_{k \rightarrow \infty} q_j^k = \sum_{j=1}^m q_j^* \log q_j^*$$

o que implica em  $\lim_{k \rightarrow \infty} q_j^k = q_j^*$  e portanto a seqüência  $\{Q^k\}$  converge.

É importante notar, aqui, que este processo de convergência é bem melhor e mais fácil do que aquele estudado anteriormente por FLEMMING, Diva M. [6] e TANEJA, I.J. e FLEMMING, Diva M. [16] o qual é um caso particular do aqui apresentado.

REFERÊNCIAS

- [1] ARIMOTO, S. (1972), An Algorithm for Computing the Capacity of Arbitrary Discrete Memoryless Channels, IEEE Trans. Information Theory IT - 18, 14-20.
- [2] ASH, R. (1965), Information Theory - Interscience Publishers, New York.
- [3] BELLIS, M. & GUIASU, S. (1968), A Quantitative-Qualitative Measure of Information in Cybernetic Systems, IEEE Trans. Information Theory, IT - 14, pp. 593-594.
- [4] BLAHUT, R.E. (1972), Computation of Capacity and Rate-Distortion Functions, IEEE Trans. Information Theory IT - 18, 460-473.
- [5] CHENG, M.C. (1974), On the Computation of Capacity of a Discrete Memoryless Channel, Inform. Contr. 24, 292-298.
- [6] DIVA, M. Flemming, Algoritmo para Calcular Capacidade do Canal Usando Entropia Ponderada e sua Generalização - Dissertação de Mestrado/UFSC, 1981.
- [7] GUIASU, S. (1977), Information Theory With Applications, McGraw-Hill, New York, pag. 58-71 e 398-401.
- [8] GUIASU, S. & PICARD, C.F. (1971), Borne Inferieure de La Longueur Utile de Certains Codes, C.R. Acad. Sci. Paris, 273, pp. 248-251.
- [9] JIMBO, M. and KUNISAWA, K. (1979), An Iteration Method for Calculating the Relative Capacity. Inform. Contr. 43, 216-223.
- [10] KULLBACK, S. (1959), "Information Theory and Statistics" Dover, New York.
- [11] MUROGA, S. (1953), On the Capacity of a Discrete Channel, 1, J. Phys. Soc. Japan 8, 484-494.
- [12] REZA, F.M. (1961), "An Introduction to Information Theory", McGraw-Hill, New York.
- [13] SHANNON, C.E. (1948), A Mathematical Theory of Communication, Bell System Tech. J. 27, 379-423, 623-656.

- [14] SHANNON, C.E. & WWAVER, W. (1949), The Mathematical Theory of Communication, The University of Illinois Press , Urbana, 1949.
- [15] TAKANO, S. (1975), On a Method of Calculating the Capacity of a Discrete Memoryless Channel, Inform. Contr. 29, 327-336.
- [16] TANEJA, I.J. & DIVA, M. Flemming, An Iterative Method for Calculating the Weighted Capacity of a Discrete Constant Channel - XIII Colóquio Brasileiro de Matemática. Held Poços de Caldas, MG, July 20-24, 1981.
- [17] LONGO, G. (1976), A Noiseless Coding Theorem for Sources Having Utilities - Siam J. Appl. Math., 739-748.