SOBRE O PROBLEMA DE CODIFICAÇÃO COM CUSTO MÍNIMO

ORIENTADOR: PROF. DR. INDER JEET TANEJA

ANTÔNIO JOÃO DA SILVA

Maio - 1981

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. Italo José Dejter
Coordenador

Banca Examinadora:

Prof. Dr. Inder Jeet Taneja
Orientador

Prof. Dr. Gur Dial

Prof. Dr. Teófilo Abuabara Saad

Aos meus pais,

A Neusa, ao Tony e a Eliane.

AGRADECIMENTOS

Ao Professor Dr. Inder Jeet Taneja, Orientador deste trabalho, pelo incentivo dado e segurança demonstrada na realização desta pesquisa.

Estendo meus agradecimentos a todos que me apoiaram e a Universidade Federal de Santa Catarina.

RESUMO

Analisando o teorema de Abu-Bokr El Sayed, que fala sobre o problema de codificação com custo mínimo e que diz $\check{C}(t) \geq 1-M$, estabelecemos o seguinte melhoramento:

Teorema:

$$\frac{1}{2} C_1(t) = \frac{1}{2} \frac{\left(\sum_{i=1}^{m} p_i^{1/t+1}\right)^{t+1} - 1}{D^t - 1} \geq 1-M,$$

para $0 \neq t > -1$, onde $M = \max\{p_1, p_2, \ldots, p_m\}$.

ABSTRACT

Analyzing the theorem of Abu-Bokr El Sayed, which treats the problem of codification with minimum cost, and which says $\check{C}(t) \geq 1-M$, we establish the following improvement:

Theorem:

$$\frac{1}{2} C_1(t) = \frac{1}{2} \frac{\left(\sum_{i=1}^{m} p_i^{1/t+1}\right)^{t+1} - 1}{D^t - 1} \geq 1-M,$$

for $0 \neq t > -1$, where $M = \max\{p_1, p_2, \ldots, p_m\}$.

ÍNDICE

INTRODUÇÃO

No Capítulo I, apresentamos o conceito e diagrama de um Sistema de Comunicação, bem como definimos e provamos as Entropias de Shannon, Rényi, Daróczy, Gama-Generalizada, de Ordem α e Grau β e também as relações entre elas.

No Capítulo II, definimos Codificação sem ruído, palavra código, média ordinária e média exponencial do comprimento das palavras códigos e média quase-aritmética dos comprimentos das palavras códigos. Ainda apresentamos limites inferiores da função custo.

No Capítulo III, fizemos a análise completa sobre o trabalho de Abu-Bakr El Sayed, a respeito dos problemas de codificação com custo mínimo, bem como três contra exemplos.

No Capítulo IV, provamos um teorema sobre codificação com custo mínimo, porém, mais forte que o provado no Capítulo III.

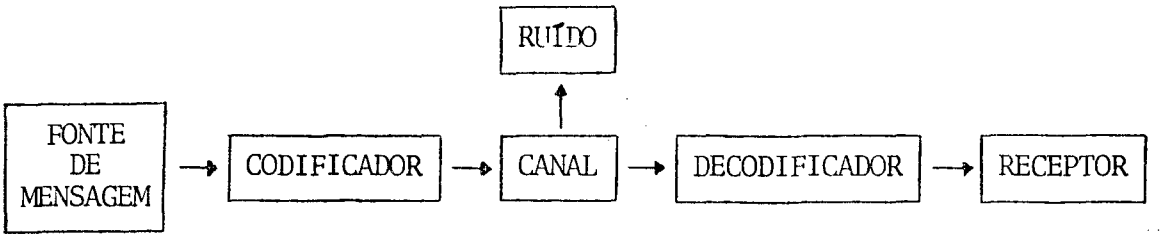
CAPÍTULO I

INTRODUÇÃO

1.1 - SISTEMAS DE COMUNICAÇÃO

C. E. Shannon (1948) desenvolveu uma teoria matemática, que tratava dos aspectos fundamentais dos sistemas de comunicação, denominada teoria da informação. Esta teoria trabalha muito com probabilidades e tem um objetivo principal que é minimizar os custos de transmissão; que para tal, utiliza de forma apropriada, os codificadores e os decodificadores; procurando de uma forma funcional atingir um desempenho ótimo, quando aplicada em sistemas de comunicação.

Para termos uma idéia geral de como funciona o sistema de comunicação para transmitir informação de um ponto para outro, damos um diagrama que visualiza o comportamento de tais sistemas, que é:



Sabemos, que toda teoria matemática procura somente os modelos matemáticos ou as expressões matemáticas para resolver os problemas práticos existentes, assim também é, a teoria da informação. Para construir esta teoria foi muito discutido, qual seria o caminho mais apropriado que melhor se adapta aos problemas de comunicação.

1.2 - ENTROPIA DE SHANNON

Consideremos $X = \{x_1, x_2, \ldots, x_m\}$ uma variável aleatória discreta com distribuição probabilística $P = (p_1, p_2, \ldots, p_m)$, onde $p_i \equiv p(x_i)$; $i = 1, 2, \ldots, m$; e $\sum_{i=1}^{m} p_i = 1$.

Denotamos o conjunto de todas as distribuições probabilísticas por Δ_m, isto é,

$$\Delta_m = \{P = (p_1, p_2, \ldots, p_m);\ p_i \geq 0;\ \sum_{i=1}^{m} p_i = 1\}. \quad (1.1)$$

A entropia de Shannon é definida por

$$H(X) = H(p_1, p_2, \ldots, p_m) = -\sum_{i=1}^{m} p_i \log_2 p_i; \quad (1.2)$$

onde $(p_1, p_2, \ldots, p_m) \in \Delta_m$.

Correspondente, para um experimento bidimensional (X,Y) com distribuição probabilística conjunta $p(x_i, y_j) = (p(i,j)$, com $i = 1, 2, \ldots, m;\ j = 1, 2, \ldots, n$; definimos a entropia conjunta de X e Y por

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

A incerteza condicional de Y, sendo $X = x_i$ é definida por

$$H\left(\frac{Y}{X=x_i}\right) = -\sum_{j=1}^{n} r\left(\frac{j}{i}\right) \log_2 r\left(\frac{j}{i}\right); \quad (1.4)$$

onde $r(\frac{j}{i})$ é a probabilidade condicional de $Y=y_j$ $(j = 1, 2, \ldots, n)$, dado que $X=x_i$, $(i = 1, 2, \ldots, m)$.

Portanto, a incerteza condicional de Y dado X, é definida como a incerteza média de $H(\frac{Y}{X=x_i})$ com pesos $p(x_i)$, $(i = 1, 2, \ldots, m)$, é dada por

$$H\left(\frac{Y}{X}\right) = \sum_{i=1}^{m} \sum_{j=1}^{m} p_i\ r\left(\frac{j}{i}\right) \log_2 r\left(\frac{j}{i}\right). \quad (1.5)$$

Com isto, temos os seguintes resultados:

$$H(X,Y) = H(X) + H\left(\frac{Y}{X}\right) = H(Y) + H\left(\frac{X}{Y}\right) \quad (1.6)$$

$$H(X,Y) \leq H(X) + H(Y) \quad (1.7)$$

$$H\left(\frac{Y}{X}\right) \leq H(Y), \quad (1.8)$$

com a igualdade em (1.7) e (1.8) se, e somente se, X e Y forem independentes.

1.3 - GENERALIZAÇÕES DA ENTROPIA DE SHANNON

1.3.1 - Entropia de Rényi

Em 1961, Rényi fez uma generalização da entropia de Shannon e definiu entropia de ordem α, dada por

$$H_\alpha(X) = \frac{1}{1-\alpha} \log_2 \left(\sum_{i=1}^{m} p_i^\alpha \right), \quad \alpha>0, \ \alpha\neq 1. \quad (1.9)$$

Correspondentemente, ao experimento conjunto de X e Y com distribuição de probabilidade conjunta $p(i,j)$, $(i = 1, 2, \ldots, m)$, $(j = 1, 2, \ldots, n)$, podemos definir a entropia conjunta de ordem α por

$$H_\alpha(X,Y) = \frac{1}{1-\alpha} \log_2 \sum_{i=1}^{m} \sum_{j=1}^{n} p^\alpha(i,j) \quad (1.10)$$

A entropia condicional de ordem α de X dado Y, é definida por

$$H_\alpha\left(\frac{X}{Y}\right) = \frac{\alpha}{1-\alpha} \log_2 \left[\sum_{j=1}^{n} \left(\sum_{i=1}^{m} p_i^\alpha \, r\left(\frac{j}{i}\right)^\alpha \right)^{\frac{1}{\alpha}} \right], \quad \alpha>0, \ \alpha\neq 1. \quad (1.11)$$

Podemos verificar os seguintes resultados:

$$H_\alpha(X,Y) \leq H_\alpha(X) + H_\alpha(Y) \quad (1.12)$$

$$H_\alpha\left(\frac{X}{Y}\right) \leq H_\alpha(X); \quad (1.13)$$

com a igualdade em (1.12) e (1.13) se, e somente se, X e Y forem independentes.

As propriedades da entropia de Rényi e suas características foram estudadas mais recentemente por Ben-Bassat e Raviv (1978).

1.3.2 - Entropia de Daróczy

Em 1970, Daróczy, introduziu o conceito de funções da informação de grau β, e por meio dessas funções definiu as entropias de grau β por

$$H^\beta(X) = (2^{1-\beta}-1)^{-1} \cdot \left(\sum_{i=1}^{m} p_i^\beta - 1 \right), \quad \beta>0, \ \beta\neq 1. \quad (1.14)$$

É fácil verificar que

$$\lim_{\beta \to 1} H^\beta(X) = H^1(X) = -\sum_{i=1}^{m} p_i \log p_i,$$

a qual é a entropia de Shannon.

Esta entropia de grau β também foi estudada por Havrda e Charvát (1967).

A entropia de Daróczy tem muitas propriedades similares as da entropia de Shannon.

Correspondente ao experimento conjunto de X e Y com distribuição de probabilidade conjunta $p(i,j)$, $(i = 1, 2, \ldots, m)$, $(j = 1, 2, \ldots, n)$, definimos a entropia conjunta de grau β por

$$H^{\beta}(X,Y) = (2^{1-\beta}-1)^{-1}.\left[\sum_{i=1}^{m}\sum_{j=1}^{n} p^{\beta}(i,j) - 1\right], \quad \beta>0,\ \beta\neq 1. \quad (1.15)$$

A entropia condicional de grau β de X, dado Y, é definida por

$$H^{\beta}\left(\frac{X}{Y}\right) = (2^{1-\beta}-1)^{-1}.\sum_{i=1}^{m} p_i^{\beta}\left(\sum_{j=1}^{n} r\left(\frac{j}{i}\right) - 1\right), \quad \beta>0,\ \beta\neq 1. \quad (1.16)$$

Estas entropias, tem as seguintes propriedades:

$$H^{\beta}(X,Y) = H^{\beta}\left(\frac{Y}{X}\right) + H^{\beta}(X) = H^{\beta}\left(\frac{X}{Y}\right) + H^{\beta}(Y); \quad (1.17)$$

$$H^{\beta}(X,Y) = H^{\beta}(X) + H^{\beta}(Y) + (2^{1-\beta}-1)^{-1}.H^{\beta}(X).H^{\beta}(Y); \quad (1.18)$$

$$H^{\beta}\left(\frac{Y}{X}\right) = H^{\beta}(Y) + (2^{1-\beta}-1).H^{\beta}(X).H^{\beta}(Y); \quad (1.19)$$

$$H^{\beta}\left(\frac{X}{Y}\right) \leq H^{\beta}(X); \quad (1.20)$$

com a igualdade em (1.20), se, e somente se, X e Y forem independentes; a igualdade (1.18) é válida para X e Y independentes.

1.3.3 - Entropia Gama Generalizada

A entropia - γ para uma distribuição de probabilidade $(p_1, p_2, \ldots, p_m)$, (foi estudada por Arimoto (1971)), é definida por

$${}_{\gamma}H(p_1, p_2, \ldots, p_m) = \frac{1}{2^{\gamma-1}-1}\left[\left(\sum_{i=1}^{m} p_i^{1/\gamma}\right)^{\gamma}-1\right],$$

$\gamma>0,\ \gamma\neq 1;$ (1.21)

Quando $\gamma\to 1$, a entropia - γ reduz-se a entropia de Shannon.

Esta entropia também tem interessantes propriedades algébricas e analíticas similares as da entropia de Shannon.

Correspondente ao experimento do conjunto X e Y com distribuição de probabilidade conjunta p(i,j), (i = 1, 2,..., m), (j = 1, 2,..., n), definimos entropia - γ conjunta por

$$ {}_{\gamma}H(X,Y) = \frac{1}{2^{\gamma-1}-1}\left[\left(\sum_{j=1}^{n}\sum_{i=1}^{m} p(i,j)^{1/\gamma}\right)^{\gamma} - 1\right]. \qquad (1.22) $$

A entropia - γ condicional de X dado Y, é definida por

$$ {}_{\gamma}H\left(\frac{X}{Y}\right) = \frac{1}{2^{\gamma-1}-1}\left[\sum_{j=1}^{n}\left(\sum_{i=1}^{m} p_i^{1/\gamma} r\left(\frac{j}{i}\right)^{1/\gamma}\right)^{\gamma} - 1\right]. \qquad (1.23) $$

Temos ainda as seguintes relações, da entropia - γ :

a) se X e Y são independentes, então:

$$ {}_{\gamma}H(X,Y) = {}_{\gamma}H(X) + {}_{\gamma}H(Y) + (2^{\gamma-1}-1).{}_{\gamma}H(X).{}_{\gamma}H(Y). \qquad (1.24) $$

$$ {}_{\gamma}H\left(\frac{X}{Y}\right) = {}_{\gamma}H(X); \qquad (1.25) $$

ambas para $\gamma>0$, $\gamma\neq1$;

b) $$ {}_{\gamma}H(X,Y) \geq {}_{\gamma}H(Y) + {}_{\gamma}H\left(\frac{X}{Y}\right), \text{ se } \gamma>1; \qquad (1.26) $$

$$ {}_{\gamma}H(X,Y) \leq {}_{\gamma}H(Y) + {}_{\gamma}H\left(\frac{X}{Y}\right), \text{ se } 0<\gamma<1; \qquad (1.27) $$

com a igualdade em (1.26) e (1.27) se, e somente se, X e Y forem independentes.

1.4 - RELAÇÕES ENTRE AS ENTROPIAS

1.4.1 - Entre Entropia Gama e Entropia de Ordem α

A entropia gama ${}_{\gamma}H(P)$ relaciona-se com a entropia de Rényi (ou de ordem α), definida em (1.9), como

$$ H_{\alpha}(P) = \frac{1}{1-\alpha}\log_2\left(\sum_{i=1}^{m} p_i^{\alpha}\right); $$

fazendo $\gamma=\alpha^{-1}$ e observando a desigualdade $\log_2 x \leq x-1$; desta forma

temos as seguintes relações:

$$_{\gamma}H(P) \leqslant H_{\alpha}(P) \leqslant H(P), \quad p/\ 0 \leqslant \gamma \leqslant \alpha^{-1} < 1;$$

$$_{\gamma}H(P) = H_{\alpha}(P) = H(P), \quad p/\ \gamma = \alpha = 1;$$

$$_{\ }H(P) \geqslant H_{\alpha}(P) \geqslant H(P), \quad p/\ \gamma = \alpha^{-1} > 1;$$

onde $H(P) = -\sum_{i=1}^{m} p_i \log_2 p_i$.

1.4.2 - Entre Entropia de Grau α e ordem α

Podemos verificar que

$$H^{\alpha}(X) = (2^{1-\alpha}-1)^{-1}\{\exp_2\left[(1-\alpha)\ H_{\alpha}(X)\right]-1\}, \quad (1.28)$$

onde $H^{\alpha}(X)$ representa entropia de grau α e $H_{\alpha}(X)$ representa entropia de ordem α.

1.5 - ENTROPIA DE ORDEM α E GRAU β

Em 1975, Sharma e Mittal fizeram uma generalização da entropia de Shannon e definiram uma entropia de ordem α e grau β, por

$$H_{\alpha}^{\beta}(X) = (2^{1-\beta}-1)^{-1}\left[\left(\sum_{i=1}^{m} p_i^{\alpha}\right)^{\beta-1/\alpha-1} - 1\right]; \quad (1.29)$$

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

Casos particulares:

(a) $\lim\limits_{\beta \to 1} H_{\alpha}^{\beta}(X) = \frac{1}{1-\alpha} \log_2\left(\sum_{i=1}^{m} p_i^{\alpha}\right)$;

que é a entropia de ordem α.

(b) Quando $\alpha = \beta \neq 1$, temos

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

que é a entropia de grau β.

(c) Quando $\gamma = \frac{1}{\alpha} = \frac{1}{2-\beta}$,

temos

$$_{\gamma}H(X) = (2^{\gamma-1}-1)^{-1}.\left[\left(\sum_{i=1}^{m} p_i^{1/\gamma}\right)^{\gamma} - 1\right], \quad \gamma\neq 1,\ \gamma>0;$$

que é a entropia estudada por Arimoto (1971).

(d) Quando $\alpha\to 1$ e $\beta\to 1$, temos

$$\lim_{\alpha\to 1}\left[\lim_{\beta\to 1} H_{\alpha}^{\beta}(X)\right] = -\sum_{i=1}^{m} p_i \log_2 p_i,$$

a qual é a entropia de Shannon.

A entropia de ordem α e grau β satisfaz as seguintes propriedades:

$$H_{\alpha}^{\beta}(X) + H_{\alpha}^{\beta}(\tfrac{Y}{X}) + (2^{1-\beta}-1)\,.\,H_{\alpha}^{\beta}(X)\,.\,H_{\alpha}^{\beta}(\tfrac{Y}{X}) =$$

$$= (2^{1-\beta}-1)^{-1}\left[\exp_2(H_{\alpha}(X) + H_{\alpha}(\tfrac{Y}{X}) - 1\right] \quad (1.30)$$

$$H_{\alpha}^{\beta}(\tfrac{X}{Y}) \leq H_{\alpha}^{\beta}(X) \quad (1.31)$$

com a igualdade em (1.31) se, e somente se, X e Y forem independentes. Onde $H_{\alpha}^{\beta}(\frac{X}{Y})$ e $H_{\alpha}^{\beta}(\frac{Y}{X})$ representam a entropia condicional de X dado Y e entropia condicional de Y dado X, respectivamente, e dada por

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

com $\beta\neq 1$, $\alpha\neq 1$, $\beta>0$ e $\alpha>0$.

Ademais, estas entropias generalizadas satisfazem muitas outras propriedades, ver Taneja (1979).

CAPÍTULO II

CANAIS SEM RUÍDO

2.1 - CODIFICAÇÃO SEM RUÍDO

No sistema clássico apresentado no Capítulo I, o canal é sem ruído, se ele permite transmissão perfeita da entrada à saída, isto é, não requer o problema da correção de erro. (Isto significa que, queremos somente maximizar o número de mensagens que pode ser mandada pelo canal em um tempo dado (fixo)).

Seja $X = \{x_1, x_2, \ldots, x_m\}$ um conjunto finito de mensagens, e, seja $P = \{p_1, \ldots, p_m\}$ uma distribuição de probabilidades associada com X, tal que a probabilidade de x_i é p_i, $i = 1, 2, \ldots, m$ e $\sum_{i=1}^{m} p_i = 1$, com $p_i \geq 0$ e $(i = 1, 2, \ldots, m)$.

Cada símbolo x_i, associado com uma seqüência finita do alfabeto código $A = \{0, 1, 2, \ldots, D-1\}$, onde $D > 1$, (D é a dimensão ou tamanho do alfabeto código), é chamado palavra código e a coleção de todas as palavras códigos é chamado código.

Se um código tem a propriedade que nenhuma palavra é prefixo da outra, então, o código é chamado código instantâneo.

Cada código instantâneo é decifrável unicamente. A recíproca é falsa, pois, existem códigos que são decifráveis unicamente mas, não são instantâneos.

Suponhamos que representamos as mensagens de X por palavras códigos; isto é; por seqüências finitas dos elementos do conjunto $\{0, 1, 2, \ldots, D-1\}$, onde $D > 1$, então, podemos mostrar que existe um código instantâneo/decifrável unicamente (ref. Feinstein (1958)), que representa x_i $(i = 1, 2, \ldots, m)$ pela palavra código de comprimento (número dos elementos) n_i $(i = 1, 2, \ldots, m)$ se, e somente se, o conjunto dos comprimentos das palavras códigos, inteiros positivos $N = \{n_1, n_2, \ldots, n_m\}$ satisfaz a desigualdade de Kraft

$$\sum_{i=1}^{m} D^{-n_i} \leq 1. \quad (2.1)$$

Uma palavra código associada com x_i de comprimento n_i para

qualquer i = 1, 2,..., m, com probabilidade p_i; então, escolhemos códigos nos quais $\bar{n} = \sum_{i=1}^{m} p_i n_i$ (comprimento médio) é mínimo, sendo este o motivo da codificação sem ruído.

Podemos provar que, se H(X) representa a incerteza (entropia de Shannon) associada com X, então existe um código instantâneo de dimensão D cujo comprimento médio das palavras códigos ($\bar{n}$) satisfaz

$$\frac{H(X)}{\log D} \leq \bar{n} < \frac{H(X)}{\log D} + 1 \qquad (2.2)$$

Agora, para cada inteiro positivo s, existe um código instantâneo X^s tal que se $\bar{n}_s$ é o comprimento médio, então

$$\lim_{s\to\infty} \frac{\bar{n}_s}{s} = \frac{H(X)}{\log D} \qquad (2.3)$$

Isto quer dizer que, cifrando suficientemente longas seqüências de entradas é possível fazer o comprimento médio de palavras códigos, para cada símbolo de entrada tão próximo de H(X) quanto se queira.

Este é o teorema de codificação sem ruído.

Para demonstração de (2.2) e (2.3) acima, referência Ash (1965).

Os resultados (2.2) e (2.3), são também, estendidos para entropia de ordem α (definida no Capítulo I), por Campbell (1965); e para entropia de ordem α e grau β provado por Gupta (1975).

Usando a entropia ponderada de ordem α, as desigualdades (2.2) e (2.3) similares, também foram estudadas por Gurdial e Pessoa (1977).

Agora seja $\phi: [1, \infty[\to \mathbb{R}$ uma função contínua estritamente crescente, tem uma inversa ϕ^{-1} que também é contínua e estritamente crescente. Isto define uma média quase aritmética do comprimento da palavra código

$$L(P, N, \phi) = \phi^{-1}\left[\sum_{i=1}^{m} p_i \, \phi(n_i)\right], \qquad (2.4)$$

para todo o N satisfazendo (2.1). A razão de chamar L um comprimento médio é que, para N = {n, n,..., n}; isto é; quando todas as palavras códigos são de iguais comprimentos n, então L (P, N, ϕ) = n. Entretanto se $\phi(x) = \phi_0(x) = x$ e $x \in [1, \infty[$;

então

$$L(P, N, \phi) = \sum_{i=1}^{m} p_i n_i, \quad (2.5)$$

é a média ordinária ou aritmética do comprimento da palavra código. Campbell, (1965, 1966), também introduziu a média exponencial do comprimento da palavra código, para as quais $\phi(x) = \phi_t(x) = D^{tx}$ com $x \in [1, \infty[$; $t \neq 0$; por

$$L(P, N, \phi_t) = \frac{1}{t} \log_D \sum_{i=1}^{m} p_i D^{tn_i} \quad (2.6)$$

É fácil ver que

$$\lim_{t \to 0} L(P, N, \phi_t) = L(P, N, \phi_0).$$

É bem conhecido (Reza,61; Campbell, 65, 66; Aczel, 74), que para todo P e N satisfazendo (2.1),

$$L(P, N, \phi_0) = \sum_{i=1}^{m} p_i n_i \geq - \sum_{i=1}^{m} p_i \log_D p_i, \quad (2.7)$$

e, para $t > -1$, $t \neq 0$,

$$L(P, N, \phi_t) = \frac{1}{t} \log_D \sum_{i=1}^{m} p_i D^{tn_i} \geq \frac{t+1}{t} \log_D \sum_{i=1}^{m} p_i^{\frac{1}{t+1}} \quad (2.8)$$

O lado da mão direita de (2.7) é a entropia de Shannon enquanto que o lado da mão direita de (2.8) é a entropia de Rényi (de ordem $\frac{1}{t+1}$).

Uma vantagem de admitir comprimentos das palavras códigos não-inteiros (Campbell, 66) é que os limites inferiores para os lados da mão direita de (2.7) e (2.8) são atualmente atingidos. Porém, se eventualmente restringirmos nossa resolução para comprimentos inteiros das palavras códigos, é fácil provar que (Reza,61; Campbell,65; Aczel,74),

$$L(P, N^*, \phi_0) = \sum_{i=1}^{m} p_i n_i^* < - \sum_{i=1}^{m} p_i \log_D p_i + 1 \quad (2.9)$$

se

$$- \log_D p_i \leq n_i^* < - \log_D p_i + 1 \ (i = 1, \ldots, m), \quad (2.10)$$

e para todo $t \neq -1$, $t \neq 0$,

$$L(P, N^*, \phi_t) = \frac{1}{t} \log_D \sum_{i=1}^{m} p_i D^{tn_i^*} < \frac{t+1}{t} \log_D \sum_{i=1}^{m} p_i^{\frac{1}{t+1}} + 1 \quad (2.11)$$

se

$$- \log_D \left(p_i^{\frac{1}{t+1}} \Big/ \sum_{i=1}^{m} p_i^{\frac{1}{t+1}}\right) \leq n_i^* < - \log_D \left(p_i^{\frac{1}{t+1}} \Big/ \sum_{i=1}^{m} p_i^{\frac{1}{t+1}}\right) + 1$$

$(i = 1, 2, \ldots, m)$. (2.12)

Podemos obter estas pela transitividade de (2.5) e (2.6).

Quando t=-1, é fácil mostrar que

$$\lim_{t \to -1} \left(\frac{t+1}{t} \log_D \sum_{i=1}^{m} p_i^{\frac{1}{t+1}}\right) = - \log_D \max(p_1, \ldots, p_m). \quad (2.13)$$

(Portanto o lado da mão direita de (2.13) é a entropia de Rényi de ordem α).

Desta mesma forma, considerando o limite $t \to -1$ em (2.6) obtemos

$$L(P, N, \phi_{-1}) = - \log_D \sum_{i=1}^{m} p_i D^{-n_i} \geq - \log_D \max(p_1, \ldots, p_m).$$

Mais geral ainda, Campbell recentemente provou (ref. Aczel e Daróczy (1975); pg 156; sec. 5.4,) que para todo $t \leq -1$

$$L(P, N, \phi_t) = \frac{1}{t} \log_D \sum_{i=1}^{m} p_i D^{tn_i} \geq \frac{1}{t} \log_D \max(p_1, \ldots, p_m), \quad (2.14)$$

enquanto que (outra vez $t \leq -1$)

$$L(P, N^*, \phi_t) = \frac{1}{t} \log_D \sum_{i=1}^{m} p_i D^{tn_i^*} < \frac{1}{t} \log_D \max(p_1, \ldots, p_m) + 1, \quad (2.15)$$

se $n_{i_0}^* = 1$, $n_i^* \geq \log_D \frac{D-1}{D(m-1)}$ $(i \neq i_0)$ onde

$p_{i_0} = \max(p_1, \ldots, p_m)$ (2.16)

(Todos estes $(n_1^*, \ldots, n_m^*)$ também satisfazem (2.1).

Sobre os lados da mão direita de (2.9), (2.11) e (2.15), +1 podem ser recolocados por $\Sigma > 0$, arbitrariamente pequeno, se decodificamos seqüências de mensagens independentes, conectivamente.

O mínimo ou limite inferior das propriedades (2.7) (2.8) e (2.14) dão interesse a seguinte interpretação de média quase aritmética dos comprimentos das palavras códigos, (conforme Campbell, 66). A função ϕ em (2.4) pode ser entendida como função custo, $\phi(n)$ sendo o custo de usar uma palavra código de comprimento n. É razoável supor que ϕ é (estritamente) crescente sobre o conjunto de inteiros positivos e então pode sempre ser estendida a uma fun

ção estritamente crescente e contínua sobre $[-1, \infty[$. Isto é conveniente porque ϕ^{-1} pode ser aplicado sobre mais que um conjunto enumerável.

Agora o "custo médio" de codificação de mensagens $X = \{x_1, x_2, \ldots, x_m\}$ da (distribuição probabilística $P = \{p_1, \ldots, p_m\}$) por uma distribuição $N = \{n_1, n_2, \ldots, n_m\}$ de comprimentos das palavras códigos é

$$C = \sum_{i=1}^{m} p_i \phi(n_i).$$

Um problema de codificação de algum interesse é minimizar o custo C por uma escolha apropriada da distribuição N, sujeita a condição (2.1). Visto que $L(P, N, \phi) = \phi^{-1}(C)$ e ϕ^{-1} é (contínua) estritamente crescente, um problema equivalente é minimizar o comprimento médio da palavra código, $L(P, N, \phi)$.

Existem constantes multiplicativas e aditivas contidas nas funções custo dadas por

(i) $\phi(x) = ax + b$ $(a>0)$ para todo $x \in [1, \infty[$

(ii) $\phi(x) = aD^{tx} + b$ $(a.t>0)$ para todo $x \in [1, \infty[$
(ref. Aczél (1974)).

(Elas não influenciam nos comprimentos médios das palavras códigos (2.5) e (2.6)). Calculando os custos médios, pode ser oportuno normalizá-los. Uma possível normalização fixaria custo unitário para decodificar uma palavra código de comprimento 1 e custo zero no caso de uma palavra código de comprimento 0. Então, no entanto, temos

$$\tilde{\phi_0}(n) = n \quad (n = 0, 1, 2, \ldots) \qquad (2.17)$$

mas no lugar de ϕ_t, temos

$$\tilde{\phi_t}(n) = \frac{D^{tn} - 1}{D^t - 1} \quad (t \neq 0,\ n = 0, 1, 2, \ldots) \qquad (2.18)$$

(uma das vantagens é que $\tilde{\phi_0} = \lim_{t \to 0} \tilde{\phi_t}$, enquanto que $\phi_0 \neq \lim_{t \to 0} \phi_t$).

As desigualdades (2.7), (2.8) e (2.14) mostram que os custos médios não podem ser menores que

$$-\sum_{i=1}^{m} p_i \log_D p_i \quad (0 \log_2 0 = 0)\ p/t = 0 \qquad (2.19)$$

$$\frac{\left(\sum_{i=1}^{m} p_i^{1/t+1}\right)^{t+1} - 1}{D^t - 1} \quad p/t \neq 0,\ t > -1, \qquad (2.20)$$

$$\frac{1 - \max(p_1, p_2, \ldots, p_m)}{1 - D^t}, \quad p\ t \leq -1, \qquad (2.21)$$

sempre que as funções custos são $\phi_t^{\sim}$, dadas por $\phi_0^{\sim}(x) = x$ e $\phi_t^{\sim}(x) = \frac{D^{tx} - 1}{D^t - 1}$ $p/t \neq 0$ e $x \in [1, \infty[$.

As inequações (2.10), (2.12) e (2.16) mostram com que N obtemos a proximidade aos limites inferiores (2.19), (2.20) e (2.21) dos custos médios, respectivamente.

CAPÍTULO III

SOBRE O PROBLEMA DE CODIFICAÇÃO COM CUSTO MÍNIMO - I

Aczél (1974), provou que, a média aritmética do comprimento da palavra código $\bar{L}_0$ e a média exponencial do comprimento das palavras códigos $\bar{L}_t$ são somente aditiva, e a média quase aritmética dos comprimentos das palavras códigos. Mais adiante, ele também provou que, sob as condições de aditividade e quase aritmeticidade da média dos comprimentos da palavra código e da normalização dos custos médios, os custos médios de decodificação das mensagens $X = \{x_1, x_2, \ldots, x_m\}$ de distribuição probabilística $P = \{p_1, p_2, \ldots, p_m\}$, tinham limites inferiores dados por

$$\check{C}(t) = C_0 = -\sum_{i=1}^{m} p_i \log_D p_i, \quad p/t=0; \qquad (3.1)$$

$$= C_1(t) = \frac{\left(\sum_{i=1}^{m} p_i^{1/t+1}\right)^{t+1} - 1}{D^t - 1}, \quad p/t>-1,\ t\neq 0; \qquad (3.2)$$

$$= C_2(t) = \frac{1 - M}{1 - D^t}, \quad p/t\leq -1, \qquad (3.3)$$

(onde $M = \max(p_1, \ldots, p_m)$.

(NOTA: O limite para $t\leq -1$ foi provado por Campbell (ref. Aczel e Daróczy (1975); pag. 156; sec. 5.4).

Em seguida, daremos um futuro limite inferior, que é independente de t, para os limites inferiores dos custos médios de codificação acima. (ref. Abu-Bakr El-Sayed (1979))

TEOREMA 3.1

Seja a média quase aritmética dos comprimentos das palavras códigos aditiva, e, seja o custo médio de codificação do conjunto X de mensagens normalizado. Se

I - as mensagens são eqüiprováveis e $D\leq m$, ou

II - uma codificação binária ($D=2$) e $M\geq\frac{1}{2}$, então, os custos

médios não podem ser menores que 1 - M, isto é, $\breve{C}(t) \geq 1 - M$, para todo $t \in \mathbb{R}$.

Prova:

Mostraremos primeiro que

$$\lim_{t \to 0} C_1(t) = C_0,$$

ou seja,

$$\lim_{t \to 0} \frac{(\sum_{i=1}^{m} p_i^{1/t+1})^{t+1} - 1}{D^t - 1} = - \sum_{i=1}^{m} p_i \log_D p_i.$$

Seja $\psi(t) = (\sum_{i=1}^{m} p_i^{1/t+1})^{t+1}$, implica que,

$$\ell n\ \psi(t) = (t+1)\ \ell n\ (\sum_{i=1}^{m} p_i^{1/t+1}).$$

Aplicando derivada nos dois lados, temos:

$$(\ell n\ \psi(t))' = \left[(t+1).\ell n(\sum_{i=1}^{m} p_i^{1/t+1})\right]',$$

isto é,

$$\frac{\psi'(t)}{\psi(t)} = (t+1).\frac{(\sum_{i=1}^{m} p_i^{1/t+1})'}{\sum_{i=1}^{m} p_i^{1/t+1}} + 1.\ell n\ (\sum_{i=1}^{m} p_i^{1/t+1}),$$

isto é,

$$\frac{\psi'(t)}{\psi(t)} = \frac{t+1}{\sum_{i=1}^{m} p_i^{1/t+1}} \cdot \sum_{i=1}^{m} (p_i^{1/t+1} \cdot \frac{-\ell n\ p_i}{(t+1)^2}) + \ell n\ (\sum_{i=1}^{m} p_i^{1/t+1}),$$

e multiplicando por $\psi(t)$, vem:

$$\psi'(t) = \psi(t)\left[\frac{1}{\sum_{i=1}^{m} p_i^{1/t+1}} \cdot \sum_{i=1}^{m} (p_i^{1/t+1} \cdot \frac{-\ell n\ p_i}{t+1}) + \ell n\ (\sum_{i=1}^{m} p_i^{1/t+1})\right],$$

que substituindo $\psi(t)$ e simplificando vem:

$$\psi'(t) = (\sum_{i=1}^{m} p_i^{1/t+1})^{t+1} \cdot \left[-\frac{1}{t+1} \cdot \frac{\sum_{i=1}^{m} (p_i^{1/t+1} \cdot \ell n\ p_i)}{\sum_{i=1}^{m} p_i^{1/t+1}} + \right.$$

$$+ \ell n \left(\sum_{i=1}^{m} p_i^{1/t+1} \right) \Bigg],$$

e observamos que

$$\lim_{t \to 0} C_1(t) = \lim_{t \to 0} \frac{\left(\sum_{i=1}^{m} p_i^{1/t+1} \right)^{t+1} - 1}{D^t - 1}$$

Aplicando L'Hospital, temos:

$$\lim_{t \to 0} C_1(t) = \frac{\left[\left(\sum_{i=1}^{m} p_i^{1/t+1} \right)^{t+1} - 1 \right]'}{\left[D^t - 1 \right]'} = \frac{\left[\left(\sum_{i=1}^{m} p_i^{1/t+1} \right)^{t+1} \right]'}{D^t \ell n \, D},$$

mas, como $\left(\sum_{i=1}^{m} p_i^{1/t+1} \right)^{t+1} = \psi(t)$, vem:

$$\lim_{t \to 0} C_1(t) = \lim_{t \to 0} \frac{\psi'(t)}{D^t \ell n D} = \frac{\psi'(0)}{\ell n \, D},$$

pois,

$$\psi'(0) = \sum_{i=1}^{m} p_i \left[\frac{-1}{\sum_{i=1}^{m} p_i} \cdot \sum_{i=1}^{m} (p_i \ell n \, p_i) + \ell n \left(\sum_{i=1}^{m} p_i \right) \right],$$

isto é,

$$\psi'(0) = 1 . \left[-1 \cdot \sum_{i=1}^{m} (p_i \ell n \, p_i) + \ell n \, 1 \right],$$

isto é,

$$\psi'(0) = - \sum_{i=1}^{m} (p_i \ell n \, p_i),$$

logo,

$$\lim_{t \to 0} C_1(t) = \frac{1}{\ell n D} \left[- \sum_{i=1}^{m} (p_i \ell n \, p_i) \right] =$$

$$= \frac{1}{\ell n D} \left[- \sum_{i=1}^{m} p_i \frac{\log_D p_i}{\log_D e} \right] = - \sum_{i=1}^{m} p_i \log_D p_i .$$

Também é claro que:

$\frac{1 - M}{1 - D^t}$ é crescente com t,

e

$$\lim_{t \to -\infty} \frac{1 - M}{1 - D^t} = 1 - M,$$

pois,

$$\lim_{t \to -\infty} \frac{1 - M}{1 - D^t} = \frac{1 - M}{1 - D^{-\infty}} = \frac{1 - M}{1 - 0} = 1 - M.$$

Portanto, $\frac{1 - M}{1 - D^t} \geq 1 - M$.

a) Sejam as mensagens eqüiprováveis, isto é,

$p_1 = p_2 = \ldots = p_m = \frac{1}{m}$.

Portanto, $M = \frac{1}{m}$, e como temos m.p's, então:

$$C_1(t) = \frac{\left[m(\frac{1}{m})^{1/t+1}\right]^{t+1} - 1}{D^t - 1} = \frac{m^t - 1}{D^t - 1},$$

logo,

$$C_1(t) = \frac{m^t - 1}{D^t - 1}.$$

Se $D = m$, então $C_1(t) = 1$.

Se $D < m$, então C_1 será crescente com t. Isto, pode ser mostrado considerando a derivada $C_1'(t)$.

$$C_1'(t) = \frac{(D^t-1) \cdot m^t \cdot \ell n\ m - (m^t-1) \cdot D^t\ \ell n\ D}{(D^t-1)^2};$$

ou seja

$$C_1'(t) > 0 \Longleftrightarrow (D^t-1)\ m^t\ \ell n\ m > (m^t-1)\ D^t \cdot \ell n\ D$$

$$\Longleftrightarrow \frac{D^t-1}{D^t \ell nD} > \frac{m^t-1}{m^t \ell nm}.$$

Esta última inequação estará verificada se provarmos que a função

$$f(x) = \frac{x^t - 1}{x^t\ \ell n\ x} \quad \text{é decrescente;} \quad (\text{onde } D < m)$$

façamos

$$f(x) = \frac{x^{-t}(x^t-1)}{\ell n\ x} = \frac{1-x^{-t}}{\ell n\ x},$$

logo,

$$f(x) = \frac{1-x^{-t}}{\ell n\ x},$$

que calculando f'(x), vem:

$$f'(x) = \frac{\ell n\ x\ (1-x^{-t})' - (1-x^{-t}) \cdot (\ell n\ x)'}{(\ell n\ x)^2} =$$

$$= \frac{1}{(\ell n\, x)^2} \cdot \left[t\, \ell n\, x.\ x^{-t-1} - \frac{1}{x}\,(1-x^{-t})\right]$$

$$= \frac{1}{(\ell n\, x)^2} \left[\ell n\, x^t \cdot x^{-t-1} - \frac{1}{x}\,(1-x^{-t})\right],$$

logo

$$f'(x) = \frac{1}{(\ell n\, x)^2} \left[x^{-t-1} \cdot \ell n\, x^t - \frac{1}{x}\,(1-x^{-t})\right].$$

Porém, $\ell n\, x^t \leq x^t - 1$. (A igualdade é verdadeira somente se $x^t = 1$, e no nosso caso $D^t \neq 1$ e $m^t \neq 1$, e portanto, vamos considerar somente a estrita inequação).

Substituindo, $\ell n\, x^t$ por $x^t - 1$, temos que o lado esquerdo fica menor que o lado direito na derivada acima, pois, $\ell n\, x^t = x^t - 1$ é o máximo de $\ell n\, x^t$.

Portanto, temos

$$f'(x) < \frac{1}{(\ell n\, x)^2} \cdot \left[x^{-t-1}(x^t-1) - x^{-1}(1-x^{-t})\right] =$$

$$= \frac{1}{(\ell n\, x)^2} \left[x^{-1} - x^{-t-1} - x^{-1} + x^{-t-1}\right] = \frac{0}{(\ell n\, x)^2} = 0,$$

logo

$$f'(x) < 0.$$

Assim sendo, f é decrescente e, por conseqüência C_1 é crescente com t. Então, segue que $C_1 > 1-M$, porque

$$\lim_{t\to\infty} \frac{1-M}{1-D^t} = 1-M$$

Assim sendo, ambos C_1 e C_2 são crescente e

$$\lim_{t\to -1} C_1(t) = \frac{m^{-1}-1}{D^{-1}-1} = \frac{1-M}{1-D^{-1}} = C_2(-1).$$

Por isso, por mensagens eqüiprováveis, a função $\check{C}$ é limitada inferiormente pela quantidade 1-M.

b) A seguir, consideremos o caso de mensagens com probabilidades que podem ser diferentes. Assumimos primeiro que m=2 e por conseqüência D=2.

Seja $P = \{p_1, p_2\} = \{p, 1-p\}$ e seja $p \geq 1-p$, isto é, $p \geq \frac{1}{2}$. Em outras palavras, $M = \max\{p_1, p_2\} = p \geq \frac{1}{2}$.

A inequação

$$C_1(t) = \frac{\left(\sum_{i=1}^{m} p_i^{1/t+1}\right)^{t+1} - 1}{D^t - 1} \geq 1-M \quad (3.4)$$

é a que queremos provar, a qual no nosso caso significa (substituindo $p_1 = p$, $p_2 = 1-p$, $D = 2$ e $M = p$):

$$\frac{\left[p^{1/t+1} + (1-p)^{1/t+1}\right]^{t+1} - 1}{2^t - 1} \geq 1-p \quad (3.5)$$

isto é,

$$\left[p^{1/t+1} + (1-p)^{1/t+1}\right]^{t+1} - 1 \geq (1-p)(2^t-1) \text{ se } t>0$$

$$\leq (1-p)(2^t-1) \text{ se } -1<t<0;$$

isto é,

$$\left[p^{1/t+1} + (1-p)^{1/t+1}\right]^{t+1} \geq p+2^t(1-p) \text{ se } t>0,$$

$$\leq p+2^t(1-p) \text{ se } -1<t<0.$$

Dividindo ambos os lados por $p(\frac{1}{2} \leq p \leq 1)$ e tomando $\frac{1-p}{p} = r$, $0 \leq r \leq 1$, $(p \geq 1-p)$, temos para provar que

$$(1+r^{1/t+1})^{t+1} \geq 1+2^t.r \text{ se } t>0, \leq 1+2^t.r \text{ se } -1<t<0;$$

o que é verdade pois,

$$\frac{\left[p^{1/t+1} + (1-p)^{1/t+1}\right]^{t+1}}{p} \geq \frac{p + 2^t(1-p)}{p}, \text{ se } t>0$$

$$\leq \frac{p + 2^t(1-p)}{p} \text{ se } -1<t<0;$$

isto é,

$$\left[\frac{p^{1/t+1}}{p^{1/t+1}} + \frac{(1-p)^{1/t+1}}{p^{1/t+1}}\right]^{t+1} \geq 1 + 2^t\frac{(1-p)}{p} \text{ se } t>0,$$

$$\leq 1 + 2^t\frac{(1-p)}{p} \text{ se } -1<t<0;$$

logo,

$$\left[1+r^{1/t+1}\right]^{t+1} \geq 1+2^t.r \text{ se } t>0, \leq 1+2^t.r \text{ se } -1<t<0.$$

Seja $B(r) = (1+r^{1/t+1})^{t+1}$ e $R(r) = 1+2^t.r$.

A função R é uma linha reta com $R(0) = 1$, $R(1) = 1+2^t$ e inclina-se igual a 2^t.

$$B'(r) = (t+1)(1+r^{1/t+1})^t \cdot \frac{1}{t+1} \cdot r^{\frac{1}{t+1}-1} \cdot 1,$$

isto é,

$$B'(r) = (1+r^{1/t+1})^t \cdot r^{-\frac{t}{t+1}},$$

isto é,

$$B'(r) = r^{-\frac{t}{t+1}} \cdot (1+r^{1/t+1})^t;$$

calculando a derivada segunda, vem:

$$B''(r) = r^{-\frac{t}{t+1}} \cdot t(1+r^{1/t+1})^{t-1} \cdot \frac{1}{t+1} \cdot r^{-\frac{t}{t+1}} +$$

$$+ \left(\frac{-t}{t+1}\right) \cdot r^{\left(\frac{-2t-1}{t+1}\right)} \cdot (1+r^{1/t+1})^t;$$

isto é,

$$B''(r) = \frac{t}{t+1}\left[r^{\frac{-2t}{t+1}} - r^{\frac{-2t-1}{t+1}} - r^{\frac{-2t}{t+1}}\right] \cdot (1+r^{1/t+1})^{t-1};$$

logo,

$$B''(r) = -\frac{t}{t+1} \cdot r^{\frac{-2t-1}{t+1}} (1+r^{1/t+1})^{t-1}.$$

(i) Seja $-1<t<0$:

$B(0) = 1 = R(0)$

$B(1) = 2^{t+1} = 2^t+2^t < 1+2^t = R(1)$

$B'(r)>0$, pois, $0\leq r\leq 1$, isto é, B é crescente.

$B'(0) = 0$, $B''(r)>0$, $B'(1) = 2^t$ $(\frac{1}{2}<2^t<1)$.

Por isto, é claro, conforme figura abaixo, que $B(r)\leq R(r)$, para todo r, $0\leq r\leq 1$.

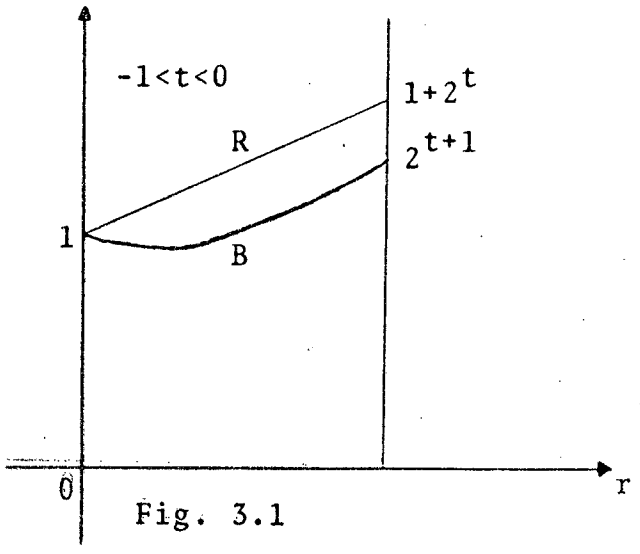


Fig. 3.1

(ii) Para $t>0$:

$B(0) = 1 = R(0)$,

$B(1) = 2^{t+1} > 1+2^t = R(1)$,

$B'(r)>0$, $B'(0) = \infty$, $B'(1) = 2^t$,

$B''(r)<0$.

Neste caso, concluimos conforme figura abaixo, que $B(r) \geq R(r)$, para todo r, $0 \leq r \leq 1$.

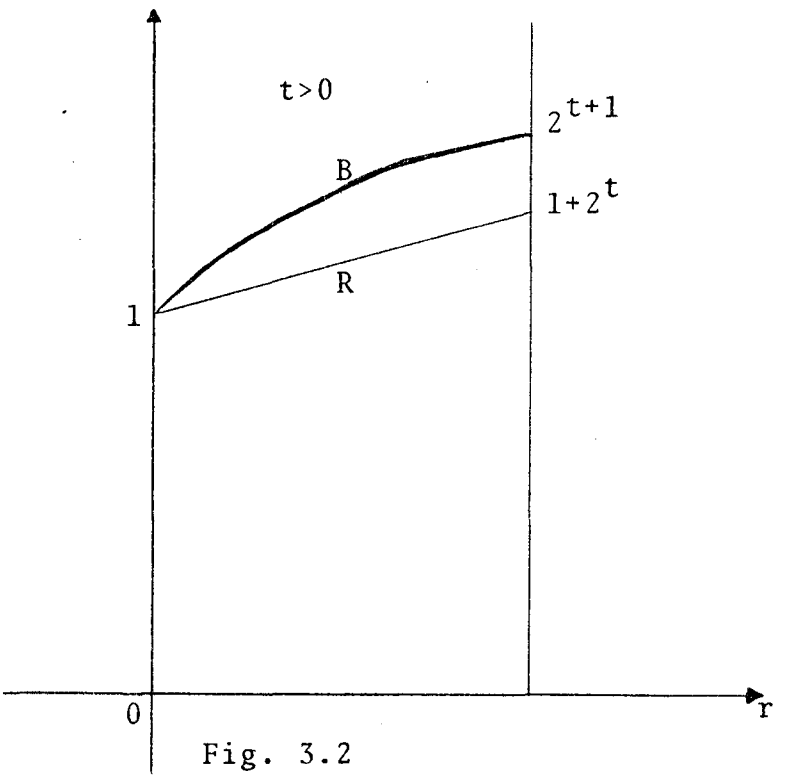


Fig. 3.2

Por isso, para todo $t>-1$, $t\neq 0$, a inequação (3.5) é verdadeira.

Agora, voltamos ao caso de m mensagens ($m\geq 2$). Assumimos que um alfabeto código binário ($D=2$) é usado e que $M \geq \frac{1}{2}$. Visto que a quantidade $\sum_{i=1}^{m} p_i^{1/t+1}$ é simétrica em relação aos p_i, podemos assumir sem perda de generalidade que $p_1 = M$.

Seja $q_1 = p_1 = M$,

$q_2 = p_2 + \ldots + p_m = 1-M$,

isto é, $q_1+q_2 = 1$.

$$\frac{\left(\sum_{i=1}^{m} p_i^{1/t+1}\right)^{t+1} - 1}{2^t - 1} > 1-M.$$

(ii) Seja $t>0$:

Seguindo passos similares, e observando (3.5'), obtemos

$$q_2^{1/t+1} < p_2^{1/t+1} + \ldots + p_m^{1/t+1},$$

e ainda

$$\left(\sum_{k=1}^{2} q_k^{1/t+1}\right)^{t+1} - 1 < \left(\sum_{i=1}^{m} p_i^{1/t+1}\right)^{t+1} - 1.$$

Colocando em ordem contrária e dividindo por 2^t-1, ($2^t-1>0$, pois $t>0$), vem:

$$\frac{\left(\sum_{i=1}^{m} p_i^{1/t+1}\right)^{t+1} - 1}{2^t - 1} > \frac{\left(\sum_{k=1}^{2} q_k^{1/t+1}\right)^{t+1} - 1}{2^t - 1} \geqslant 1-M.$$

Portanto

$$\frac{\left(\sum_{i=1}^{m} p_i^{1/t+1}\right)^{t+1} - 1}{2^t - 1} > 1-M,$$

para todo $t>-1$, $t\neq 0$.

Isto completa a prova.

Observação: Geralmente não é verdadeiro que:

Se $m \geqslant D$ e $M \geqslant \frac{1}{2}$, então $\check{C}(t) \geqslant 1-M$, para todo $t \in \mathbb{R}$. Isto será mostrado exibindo os exemplos contrários abaixo:

Exemplo 1: Neste exemplo o número de mensagens é igual ao número de elementos do conjunto de codificação.

Seja $m=D=3$,

$P = \{0,6;\ 0,0001;\ 0,3999\}$,

$M = 0,6$, então $1-M = 1-0,6 = 0,4$;

como $C_1(t) = \dfrac{\left(\sum_{i=1}^{m} p_i^{1/t+1}\right)^{t+1} - 1}{D^t - 1}$, vem:

$$C_1(3) = \frac{\left(\sum_{i=1}^{3} p_i^{1/3+1}\right)^{3+1} - 1}{3^3 - 1} =$$

$$= \frac{1}{26}\left[\left((0,6)^{1/4} + (0,0001)^{1/4} + (0,3999)^{1/4}\right)^4 - 1\right]$$

$$= 0,3433 < 0,4$$

logo, $C_1(t) < 1-M$.

Exemplo 2: Neste exemplo, o número de mensagens é maior que o número de alfabetos de codificação, com uma insignificante modificação do anterior. No lugar da probabilidade $p_3 = 0,3999$, temos quatro probabilidades cuja soma é igual ao valor de p_3. Uma destas quatro probabilidades (0,39989997) é muito próxima de p_3 (portanto, dando quase a mesma contribuição $p_3^{1/4}$ para a soma total $\sum_i p_i^{1/4}$) e as outras três são muito pequenas (10^{-8} cada), portanto, crescendo C inadequadamente, mas, mantendo contudo menor do que 0,4(=1-M). Assim seja m=6>D=3,

$P = \{0,6;\ 0,0001;\ 0,39989997;\ 10^{-8};\ 10^{-8};\ 10^{-8}\}$.

$1-M = 0,4$,

$$C_1(3) = \frac{\left(\sum_{i=1}^{6} p_i^{1/4}\right)^4 - 1}{3^3 - 1} = \frac{9,6146}{26} = 0,3698 < 0,4$$

logo $C_1(3) < 1-M = 0,4$.

Exemplo 3: Neste exemplo, usamos um grande número de mensagens e de símbolos de codificação.

Seja, $m = 130$, $D = 50$;

$$P = \{0,5728;\ 0,4144;\ \underbrace{10^{-4};\ 10^{-4};\ \ldots,\ 10^{-4}}_{128 \text{ vezes}}\}$$

Como, $M = 0,5728$,

$1-M = 0,4272$.

$$C_1(3) = \frac{\left[0,5728^{1/4} + 0,4144^{1/4} + (128.10^{-4})^{1/4}\right]^4 - 1}{50^3 - 1}$$

$$< \frac{(0,87 + 0,81 + 12,8)^4}{50^3};$$

logo, $C_1(3) < \frac{(14,48)^4}{50^3} < \frac{15^4}{50^3} = 0,405 < 0,4272 = 1-M.$

De onde, $C_1(3) < 1-M.$

CAPÍTULO IV

SOBRE O PROBLEMA DE CODIFICAÇÃO COM CUSTO MÍNIMO - II

No capítulo III, foi provado que

$$C_1(t) = \frac{\left(\sum_{i=1}^{m} p_i^{1/t+1}\right)^{t+1} - 1}{D^t - 1} \geq 1-M, \text{ para } t>-1 \text{ e } t\neq 0.$$

Agora, vamos provar um resultado mais forte a respeito dos limites inferiores de $C_1(t)$ para $t>-1$, $t\neq 0$. Para tanto vamos enunciar o teorema:

TEOREMA 4.1

Para cada $P = (p_1, p_2, \ldots, p_m) \in \Delta_m$, com $m \geq 2$, $M \geq \frac{1}{2}$ e $D = 2$, temos $\frac{1}{2} C_1(t) \geq 1-M$,

isto é,

$$\frac{1}{2} C_1(t) = \frac{1}{2} C_1(p_1, \ldots, p_m; t) \geq 1-M$$

isto é,

$$\frac{1}{2} C_1(t) = \frac{1}{2} \cdot \frac{1}{2^t-1} \left[\left(\sum_{i=1}^{m} p_i^{1/t+1}\right)^{t+1} - 1 \right] \geq 1-M; \quad (4.1)$$

para $t>-1$, $t\neq 0$, onde $M=\max\{p_1, p_2, \ldots, p_m\}$.

Para demonstrarmos este teorema precisamos o lema seguinte:

Lema 4.1:

A função $C_1(p_1, \ldots, p_m; t)$, com $t>-1$, $t\neq 0$ é côncava para $(p_1, p_2, \ldots, p_m) \in \Delta_m$.

Demonstração:

Consideremos a função

$$F(P) = F(p_1, p_2, \ldots, p_m) = \left(\sum_{i=1}^{m} p_i^{1/t+1}\right)^{t+1}.$$

Vamos provar que F(P) é uma função côncava para $t>-1$ e $t\neq 0$.

Sejam $P = (p_1, p_2, \ldots, p_m) \in \Delta_m$ e $Q = (q_1, q_2, \ldots, q_m) \in \Delta_m$, duas distribuições de probabilidades, então, para $0 \leqslant \lambda \leqslant 1$, temos:

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

$$\geqslant \left[\sum_{i=1}^{m} (\lambda p_i)^{1/t+1}\right]^{t+1} + \left[\sum_{i=1}^{m} (1-\lambda)^{1/t+1} \cdot q_i^{1/t+1}\right]^{t+1}$$

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

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

para $t>0$, onde a desigualdade acima é obtida, usando a desigualdade de Minhowski.

Se $-1<t<0$, a desigualdade é contrária.

Desta forma ficou provada a afirmação do lema.

Consequentemente, a função

$$C_1(p_1, \ldots, p_m; t) = \frac{\left[\left(\sum_{i=1}^{m} p_i^{1/t+1}\right)^{t+1} - 1\right]}{2^t - 1},$$

com $t>-1$, $t\neq 0$ é côncava em Δ_m.

Demonstração do teorema 4.1

Primeiro, vamos provar (4.1) para m=2 e após, estendemos para m>2.

Para o caso m=2, temos

$$\frac{1}{2} C_1(0,1; t) = \frac{1}{2}\left[\frac{(0^{1/t+1} + 1^{1/t+1}) - 1}{2^t - 1}\right] =$$

$$= \frac{1}{2} \cdot \frac{0}{2^t - 1} = 0; \text{ para } t\neq 0;$$

logo, $\frac{1}{2} C_1(0,1;\ t) = 0 \quad (4.2)$

$$\frac{1}{2} C_1(\tfrac{1}{2},\tfrac{1}{2};\ t) = \frac{1}{2}\left[\frac{(2^{-1/t+1} + 2^{-1/t+1})^{t+1} - 1}{2^t - 1}\right] =$$

$$= \frac{1}{2}\left[\frac{(2.2^{-1/t+1})^{t+1} - 1}{2^t - 1}\right] =$$

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

$$= \frac{1}{2}\left[\frac{2^t - 1}{2^t - 1}\right] = \frac{1}{2} \cdot 1 = \frac{1}{2};$$

para $t \neq 0$; logo,

$$\frac{1}{2} C_1(\tfrac{1}{2},\tfrac{1}{2};\ t) = \frac{1}{2} \quad (4.3)$$

$$\frac{1}{2} C_1(1,0;\ t) = \frac{1}{2}\left[\frac{(1^{1/t+1} + 0^{1/t+1})^{t+1} - 1}{2^t - 1}\right] =$$

$$= \frac{1}{2}\left[\frac{(1 - 1)}{2^t - 1}\right] = \frac{1}{2} \cdot \frac{0}{2^t - 1} = 0; \text{ para } t \neq 0,$$

logo,

$$\frac{1}{2} C_1(1,0;\ t) = 0 \quad (4.4)$$

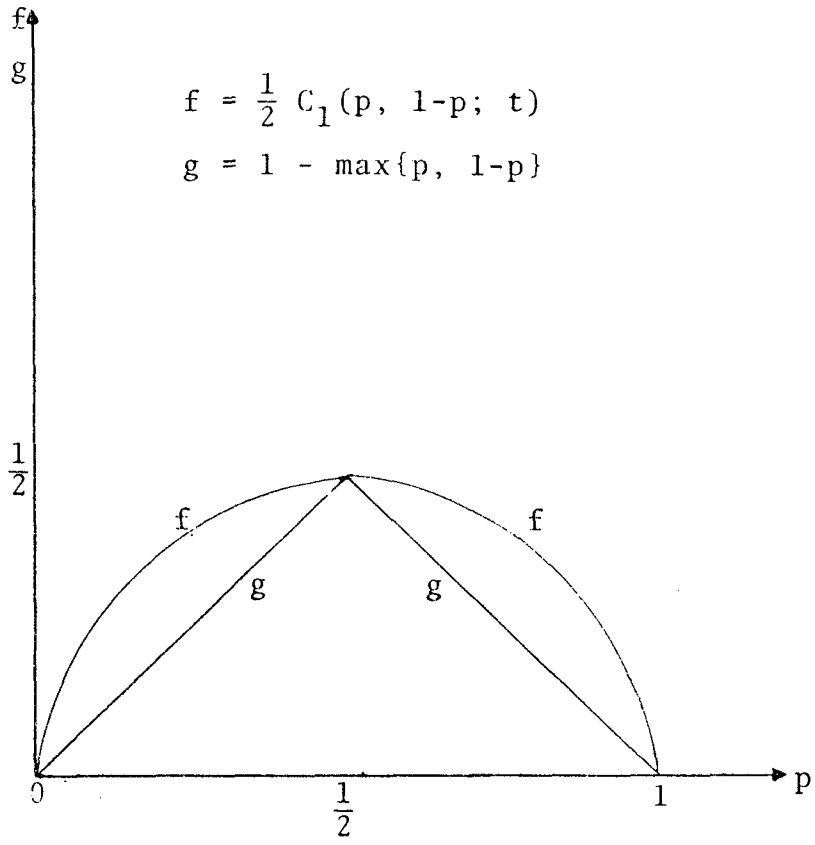
Resumindo temos:

$$\frac{1}{2} C_1(0,1;\ t) = 0$$

$$\frac{1}{2} C_1(\tfrac{1}{2},\tfrac{1}{2};\ t) = \frac{1}{2}$$

$$\frac{1}{2} C_1(1,0;\ t) = 0; \text{ para } t > -1,\ t \neq 0.$$

Conforme figura a seguir vemos que o gráfico de $1-\max(p, 1-p)$, $0 \leq p \leq 1$, consiste de duas linhas retas entre $(0,0)$ e $(\frac{1}{2},\frac{1}{2})$ e entre $(\frac{1}{2},\frac{1}{2})$ e $(1,0)$;



Como $C_1(p_1, \ldots, p_m; t)$ é uma função côncava para $t>-1$, $t\neq 0$, temos:

$$1-M \leq \frac{1}{2} C_1(p_1, p_2; t) = \frac{1}{2} C_1(p, 1-p; t),$$

isto é,

$$1-M \leq \frac{1}{2} C_1(p, 1-p; t), \quad 0\leq p\leq 1. \qquad (4.5)$$

Agora, seja $P\in\Delta_m$, $m>2$ com $M\geq\frac{1}{2}$ e, sem perda de generalidade, assumimos $M=p_1$.

Seja $q_1 = p_1 = M$;

$$q_2 = p_2 + p_3 + \ldots + p_m = 1-M,$$

isto é,

$$q_1 + q_2 = 1.$$

(i) Seja $-1<t<0$:

Visto que,

$$\left(\sum_{k=1}^{n} a_k\right)^w \begin{cases} \leqslant \sum_{k=1}^{n} a_k^w & \text{se } w\leqslant 1 \\ \geqslant \sum_{k=1}^{n} a_k^w & \text{se } w\geqslant 1 \end{cases} \qquad (4.6)$$

(onde $a_k\geqslant 0$, $k = 1, 2, \ldots, n$).

Por conseqüência vem:

$$q_2^{1/t+1} > p_2^{1/t+1} + \ldots + p_m^{1/t+1}, \text{ pois } -1<t<0,$$

logo

$$q_1^{1/t+1} + q_2^{1/t+1} > p_1^{1/t+1} + p_2^{1/t+1} + \ldots + p_m^{1/t+1};$$

Multiplicando os dois lados por $\frac{1}{2}$ e utilizando somatório vem:

$$\frac{1}{2}\left(\sum_{k=1}^{2} q_k^{1/t+1}\right)^{t+1} > \frac{1}{2}\left(\sum_{i=1}^{m} p_i^{1/t+1}\right)^{t+1}$$

Dividindo por $2^t - 1$ e como $2^t - 1<0$, para $-1<t<0$, temos:

$$\frac{1}{2}\,\frac{\left(\sum_{k=1}^{2} q_k^{1/t+1}\right)^{t+1} - 1}{2^t - 1} < \frac{1}{2}\,\frac{\left(\sum_{i=1}^{m} p_i^{1/t+1}\right)^{t+1} - 1}{2^t - 1} \qquad (4.7)$$

Observando que $\sum_{k=1}^{2} q_k = 1$, e aplicando a inequação (4.5), obtemos:

$$\frac{1}{2}\,\frac{\left(\sum_{k=1}^{2} q_k^{1/t+1}\right)^{t+1} - 1}{2^t - 1} = \frac{1}{2}\,\frac{\left(q_1^{1/t+1} + q_2^{1/t+1}\right)^{t+1} - 1}{2^t - 1} =$$

$$= \frac{1}{2}\,\frac{\left(q_1^{1/t+1} + (1-q_1)^{1/t+1}\right)^{t+1} - 1}{2^t - 1} \geqslant$$

$$\geqslant 1-\max\{q_1, q_2\} = 1-q_1 =$$

$$= 1 - p_1 = 1 - M,$$

logo

$$\frac{1}{2}\,\frac{\left(\sum_{k=1}^{2} q_k^{1/t+1}\right)^{t+1} - 1}{2^t - 1} \geqslant 1-M.$$

Combinando esta inequação com a (4.7), obtemos

$$\frac{1}{2}\,\frac{\left(\sum_{i=1}^{m} p_i^{1/t+1}\right)^{t+1} - 1}{2^t - 1} \geq 1-M.$$

(ii) Seja $t>0$:

Seguindo os mesmos passos e observando (4.6), obtemos

$$q_2^{1/t+1} < p_2^{1/t+1} + \ldots + p_m^{1/t+1}.$$

Multiplicando por $\frac{1}{2}$ e somando dos dois lados $q_1^{1/t+1} = p_1^{1/t+1}$, vem

$$\frac{1}{2}\,(q_1^{1/t+1} + q_2^{1/t+1}) < \frac{1}{2}(p_1^{1/t+1} + p_2^{1/t+1} + \ldots + p_m^{1/t+1}),$$

logo,

$$\frac{1}{2}\left[\left(\sum_{k=1}^{2} q_k^{1/t+1}\right)^{t+1} - 1\right] < \frac{1}{2}\left[\left(\sum_{i=1}^{m} p_i^{1/t+1}\right)^{t+1} - 1\right].$$

Colocando em ordem contrária, dividindo por 2^t-1, onde $2^t-1>0$, pois, $t>0$ e utilizando (4.5), vem:

$$\frac{1}{2}\left[\frac{\left(\sum_{i=1}^{m} p_i^{1/t+1}\right)^{t+1} - 1}{2^t - 1}\right] > \frac{1}{2}\left[\frac{\left(\sum_{k=1}^{2} q_k^{1/t+1}\right)^{t+1} - 1}{2^t - 1}\right] \geq 1-M,$$

logo

$$\frac{1}{2}\left[\frac{\left(\sum_{i=1}^{m} p_i^{1/t+1}\right)^{t+1} - 1}{2^t - 1}\right] \geq 1-M;$$

para todo $t>-1$, $t\neq 0$.

Com isto, concluimos a prova do teorema, sendo

$\frac{1}{2}\, C_1(t) \geq 1-M$, para

$t>-1$ e $t\neq 0$.

Observação: $\lim_{t\to 0} C_1(t) = -\sum_{i=1}^{m} p_i \log p_i = C_0$.

O caso para C_0 foi provado em Galleger (1968), isto é, $\frac{1}{2}\, C_0 \geq 1-M$.

REFERÊNCIAS

ACZÉL, J. (1974). Determination of All Additive Quasiarithmetic Mean Codeword Lengths - Z. Wahr. und Vern. Geb., 29, pág. 351 - 360.

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

ARIMOTO, S. (1971). Information - Theoretical Considerations on Estimation Problems - Information and Control, 19, pág. 181 - 190.

ASH, R. B. (1965). Information Theory, Wiley, New York.

BASSAT, M. B. e RAVIV, J. (1978). Rényi's Entropy and Probability of Error - IEEE trans. Inform. Theory, IT - 24, pág. 323 - 331.

CAMPBELL, L. L. (1965). A Coding Theorem and Rényi's Entropy - Information and Control, 8, pág. 423 - 429.

CAMPBELL, L. L. (1966). Definition of entropy by means of a coding problem - Z. Wahr. Vern. Geb., 6, pág. 113 - 118.

DARÓCZY, Z. (1970). Generalized Information Functions - Information and Control, 16, pág. 36 - 51.

EL - SAYED, A. B. (1979). On the Problem of Coding with Minimal Costs - Information and Control, 40, pág. 291 - 300.

FEINSTEIN, A. (1958). Foundations of Information Theory - McGraw Hill, New York.

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

GUPTA, H. C. (1975). Noiseless Coding Theorems for Non-Additive Measures of Entropy and Inacuracy - J. Math. Sci, 10, pág. 86 - 95.

GURDIAL e PESSOA, Fº (1977). On Useful Information of order α - J. Comb. Inf. Syst. Sci, 2, pág. 158 - 162.

HAVRDA, J. e CHARVÁT, F. (1967). Quantification Method of Classification Processos. Concept of Structural a-Entropy-Kybenetika, 3, pág. 30 - 35.

RÉNYI, A. (1961). On Measures of Entropy and Information - Proc. 4th Berkeley Symp. Math. statist. Probability, 1960, 1, 547 - 561. Univ. of California Press, Berkeley, 1961.

REZA, F. M. (1961). An Introduction to Information Theory - McGraw Hill, New York.

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

SHARMA, B. D. e MITTAL, D. P. (1975). New Non - Additive Measures of Entropy for a Discrete Probability Distribution - J. Math. Sci, 10, pág. 28 - 40.

TANEJA, I. J. (1979). Some Contribuitions to Information Theory - I - J. Comb. Inform. Syst. Sci, 4, pág. 253 - 274.