

UNIVERSIDADE FEDERAL DE SANTA CATARINA

MEDIDAS DE INFORMAÇÃO E TEOREMAS DE CODIFICAÇÃO

SEM RUÍDO

Rosimary Pereira

Outubro - 1982

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. Inder Jeet Taneja, Ph.D.
Coordenador

BANCA EXAMINADORA:



Prof. Gur Dial, Ph.D.
Orientador



Prof. Inder Jeet Taneja, Ph.D.



Prof. Carlos Alberto de Bragança Pereira, Dr.

A

meus pais

AGRADECIMENTOS

Ao professor Gur Dial pela sua orientação, estímulo e compreensão durante a elaboração deste trabalho.

Aos meus familiares que de alguma forma contribuíram para a realização deste trabalho.

Aos meus colegas da EAM/SC e do Curso de Pós-Graduação pelo carinho, incentivo e colaboração.

Estendo meus agradecimentos à Universidade Federal de Santa Catarina.

RESUMO

Na literatura da teoria de informação, existem diferentes medidas de informação tais como entropia de Shannon, entropias do tipo β , de ordem α , tipo (α, β) , ordem (α, β) e de ordem α e tipo β .

Neste trabalho apresentaremos generalizações da desigualdade de Shannon para essas entropias e aplicações destas nos teoremas de codificação.

ABSTRACT

In the literature of information theory, exists various measures of information such as Shannon's entropy, entropy of type β , order α , type (α, β) , order (α, β) and of order α and type β .

In this work we will present generalizations of Shannon's inequality for these entropies and their applications to coding theorems.

Í N D I C E

CAPÍTULO 1 - INTRODUÇÃO	1
1.1 - Sistemas de Comunicação	1
1.2 - Medida de Informação	2
1.3 - Entropia de Shannon	5
1.4 - Generalizações da entropia de Shannon ...	6
1.5 - Propriedades da entropia de Shannon e de suas generalizações	9
1.6 - Medida de Imprecisão	14
1.7 - Generalizações da Medida de Imprecisão ..	14
CAPÍTULO 2 - ENTROPIAS DE ORDEM α , TIPO β e TEOREMAS DE CO- DIFICAÇÃO	17
2.1 - Desigualdade de Shannon	17
2.2 - Codificação sem Ruído	19
2.3 - Entropia de Tipo β	27
2.4 - Entropia de Ordem α	38
CAPÍTULO 3 - ENTROPIAS DE ORDEM (α, β) TIPO (α, β) , E DE ORDEM α E TIPO β E TEOREMAS DE CODIFICAÇÃO	44
3.1 - Entropia de Ordem (α, β)	44
3.2 - Entropia de Tipo (α, β)	51
3.3 - Entropia de Ordem α e Tipo β	59
BIBLIOGRAFIA	67

INTRODUÇÃO

Neste trabalho, trataremos do problema da codificação sem ruído.

O objetivo principal desta codificação é minimizar o custo médio para transmissão de informações.

No capítulo 1, apresentaremos a entropia de Shannon dando sua definição e algumas propriedades.

No capítulo 2, mostraremos uma relação envolvendo duas distribuições de probabilidades discretas, chamada desigualdade de Shannon. A importância desta relação é conhecida no teorema da codificação. Após essa demonstração, generalizaremos essa desigualdade para as entropias de Renyi e Daróczy.

O capítulo 3, consta de generalizações dessa desigualdade de Shannon com aplicações nos teoremas de codificação para as entropias de ordem (α, β) , tipo (α, β) , de ordem α e tipo β .

CAPÍTULO 1

INTRODUÇÃO

1.1 - Sistemas de Comunicação

Teoria de informação é um novo ramo da Matemática com grande aplicação nos sistemas de comunicação. Ela está relacionada com três conceitos básicos: a medida de informação; a capacidade de um canal de comunicação para transmitir informações; e uma codificação como meio capaz de utilizar os canais com suas capacidades máximas.

Shannon^[16] em 1948 desenvolveu esta teoria com objetivo de tratar os problemas matemáticos ligados com a transmissão de informações sobre canais de comunicação.

Além de Shannon, Nyquist^[14] e Hartley^[10] também deram contribuições pioneiras neste ramo.

Um sistema de comunicação é um conjunto de processos que possibilita a transmissão de um ponto denominado fonte para um outro ponto qualquer. E tem uma forma esquemática como a representada na figura 1.1 abaixo.

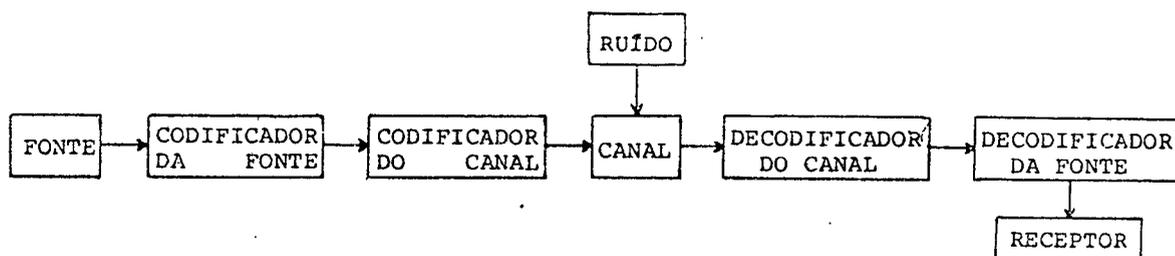


Figura 1.1

Todos os sistemas de comunicação possuem a mesma função que é transmitir informações.

A palavra informação será usada num sentido especial que não deve ser confundida com seu uso habitual. No momento nós identificamos informação com o custo ou o tempo que gastamos para transmitir uma mensagem da fonte para o receptor. Ao definirmos informação, consideraremos não somente a mensagem gerada ou transmitida, mas também o conjunto de todas mensagens das quais escolheremos uma.

Intuitivamente, a quantidade de informação será uma medida do tempo ou do custo gasto para transmitir mensagens.

1.2 - Medida de Informação

Suponhamos que temos dois aparelhos x_1 , x_2 que devem marcar quente e frio. Para isso usaremos dígitos binários da seguinte maneira.

$$x_i \text{ quente} = 1$$

$$x_i \text{ frio} = 0$$

onde $p_r \{x_i = 0\} = p_r \{x_i = 1\} = \frac{1}{2}$

Montaremos um esquema que nos permite avaliar o grau de incerteza que envolve a captação de uma mensagem emitida pela fonte $\{x_1, x_2\}$.

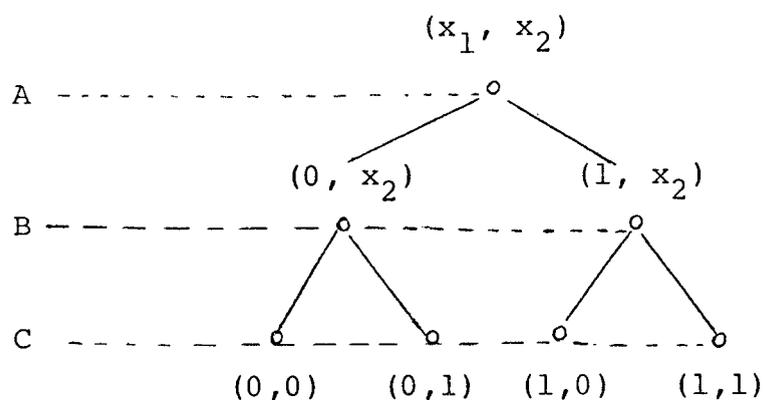


Figura 1.2

Analisaremos agora a figura 1.2. Ela apresenta três níveis A,B,C de sustentação dos nós.

No nível C, cada nó domina apenas uma mensagem. Não havendo incerteza. A informação de qualquer um deles se define como um bit.

No nível B, cada nó nos dá duas mensagens distintas dado que a segunda componente é desconhecida, implicando que a incerteza das mensagens é um.

No nível A, o único nó nos dá quatro mensagens distintas, dado que temos duas componentes desconhecidas, implicando que a incerteza associada com cada mensagem é igual a dois.

Concluimos que quanto maior a incerteza sobre a mensagem, maior será a quantidade de informação transmitida. Através da tabela seguinte estabeleceremos uma primeira medida de informação.

Nível	Incerteza	Quantidade de mensagem
A	2	$4 = 2^2$
B	1	$2 = 2^1$
C	0	$1 = 2^0$

Tabela 1.1

É fácil verificar que a quantidade de mensagem enviada por um nó é sempre uma potência de dois, cujo expoente é a incerteza da mesma.

Tiramos a seguinte conclusão

$$2^I = m$$

onde m = quantidade de mensagem

I = incerteza

Aplicando logaritmo na base 2 em ambos os lados, temos:

$$I = \log_2 m \quad (1.1)$$

No início falávamos em mensagens equiprováveis, isto é, para n mensagens, temos que a probabilidade de ocorrência de cada mensagem é $P = \frac{1}{m}$

Logo podemos escrever (1.1) como:

$$I = \log_2 \frac{1}{P} \quad \text{ou} \quad I = -\log_2 P \quad (1.2)$$

Portanto (1.2) nos mostra que a quantidade de informação contida numa mensagem transmitida por uma fonte é dada pelo

simétrico do logaritmo da probabilidade de ocorrência da mensagem.

Também se pode mostrar para mensagens não equiprováveis, o que significa na não igualdade dos valores da informação transmitida por cada uma das mesmas. Surge então a necessidade de calcularmos um valor médio de informação, definido como uma ponderação das informações providas por cada mensagem, pela probabilidade respectiva, representada por H .

Então, para uma variável aleatória $X = (x_1, x_2, \dots, x_n)$ onde cada x_i tem probabilidade p_i , $i = 1, \dots, n$, e informação $-\log_2 p_i$ estabeleceu-se que "a informação média é a média ponderada das informações das mensagens possíveis de serem geradas pela fonte" dada por:

$$H(P) = - \sum_{i=1}^n p_i \log_2 p_i \quad (1.3)$$

Observação - Neste trabalho, caso não venha especificado a base do logaritmo, esta será igual a dois.

1.3 - Entropia de Shannon

Suponhamos uma variável aleatória $X = (x_1, \dots, x_n)$ onde cada elemento x_i , $i = 1, \dots, n$, tem representação através de sua probabilidade de ocorrência p_i .

Denotando o conjunto de todas as distribuições de probabilidades completas por:

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

definimos a entropia como uma função $H : \Delta_n \rightarrow \mathbb{R}$.

Definição 1 -
$$H(P) = - \sum_{i=1}^n p_i \log p_i \quad \text{para } P \in \Delta_n.$$

$H(P)$ denominada entropia de Shannon é a medida de informação da da em (1.3) e enunciada pela primeira vez por Claude Shannon em 1948 [16].

É uma função contínua, aditiva e possuinte de outras propriedades relacionadas posteriormente.

1.4 - Generalizações da Entropia de Shannon

A primeira generalização da entropia de Shannon foi dada por Renyi [15] em 1961 e é conhecida como entropia de ordem α ou entropia de Renyi.

Definição 2 -
$$H_\alpha(P) = \frac{1}{1-\alpha} \log \left(\sum_{i=1}^n p_i^\alpha \right) \quad \text{para } \alpha \neq 1, \alpha > 0 \text{ e } P \in \Delta_n$$
 (1.4)

onde α é um parâmetro.

Observação:
$$\lim_{\alpha \rightarrow 1} H_\alpha(P) = H(P)$$

Havrda e Charvát [11] em 1967 e Daróczy [7] em 1970 introduziram a entropia do tipo β ou de grau β da seguinte maneira:

Definição 3 -
$$H^\beta(P) = (2^{1-\beta} - 1)^{-1} \cdot \left(\sum_{i=1}^n p_i^\beta - 1 \right) \quad \text{para } \beta \neq 1, \beta > 0.$$
 (1.5)

onde β é um parâmetro e $P \in \Delta_n$.

Observação - $\lim_{\beta \rightarrow 1} H^\beta(P) = H(P)$.

Aczél^[1] em 1963 introduziu a entropia de ordem (α, β) que é uma generalização da entropia de Renyi envolvendo dois parâmetros. Esta é definida por:

$$\text{Definição 4} - H_{\alpha, \beta}(P) = (\beta - \alpha)^{-1} \cdot \log \frac{\sum_{i=1}^n p_i^\alpha}{\sum_{i=1}^n p_i^\beta} \text{ para } \alpha, \beta > 0, \alpha \neq \beta \quad (1.6)$$

onde $P \in \Delta_n$, α e β são parâmetros.

Observações:

- 1) $\lim_{\alpha \rightarrow 1} H_{\alpha, \beta}(P) = H_\beta(P)$.
- 2) $\lim_{\beta \rightarrow 1} \lim_{\alpha \rightarrow 1} H_{\alpha, \beta}(P) = H(P)$

Podemos escrever esta entropia de outra maneira, como segue:

$$\begin{aligned} H_{\alpha, \beta}(P) &= (\beta - \alpha)^{-1} \cdot (\log \sum_{i=1}^n p_i^\alpha - \log \sum_{i=1}^n p_i^\beta) \\ &= (\beta - \alpha)^{-1} \cdot \left(\frac{1-\alpha}{1-\alpha} \cdot \log \sum_{i=1}^n p_i^\alpha - \frac{1-\beta}{1-\beta} \log \sum_{i=1}^n p_i^\beta \right) \\ H_{\alpha, \beta}(P) &= \frac{1-\alpha}{\beta-\alpha} \cdot H_\alpha(P) + \frac{\beta-1}{\beta-\alpha} \cdot H_\beta(P) \end{aligned} \quad (1.7)$$

onde $H_\alpha(P)$ e $H_\beta(P)$ são as entropias de Renyi.

Sharma e Taneja^[19] introduziram a entropia do tipo (α, β) como segue:

Definição 5 - $H^{\alpha, \beta}(P) = (2^{1-\alpha} 2^{1-\beta})^{-1} \cdot \left(\sum_{i=1}^n p_i^\alpha - \sum_{i=1}^n p_i^\beta \right)$ (1.8)

para $\alpha, \beta > 0$, $\alpha \neq \beta$, $P \in \Delta_n$, onde α, β são dois parâmetros.

Observações:

- 1) $\lim_{\alpha \rightarrow 1} H^{\alpha, \beta}(P) = H^\beta(P)$.
- 2) $\lim_{\beta \rightarrow 1} \lim_{\alpha \rightarrow 1} H^{\alpha, \beta}(P) = H(P)$.

Podemos escrever esta entropia de outra maneira, como segue:

$$\begin{aligned}
 H^{\alpha, \beta}(P) &= (2^{1-\alpha} 2^{1-\beta})^{-1} \cdot \left(\sum_{i=1}^n p_i^\alpha - 1 + 1 - \sum_{i=1}^n p_i^\beta \right) \\
 &= (2^{1-\alpha} 2^{1-\beta})^{-1} \cdot \left[(2^{1-\alpha} - 1) \cdot \frac{\left(\sum_{i=1}^n p_i^\alpha - 1 \right)}{(2^{1-\alpha} - 1)} + \right. \\
 &\quad \left. + (2^{1-\beta} - 1) \cdot \frac{\left(1 - \sum_{i=1}^n p_i^\beta \right)}{(2^{1-\beta} - 1)} \right] \\
 H^{\alpha, \beta}(P) &= \frac{A_\alpha}{A_\alpha - A_\beta} \cdot H^\alpha(P) + \frac{A_\beta}{A_\beta - A_\alpha} \cdot H^\beta(P) \quad (1.9)
 \end{aligned}$$

onde $A_\alpha = 2^{1-\alpha} - 1$

$A_\beta = 2^{1-\beta} - 1$

e $H^\alpha(P)$, $H^\beta(P)$ são as entropias de Daróczy.

Seja \mathbb{R} o conjunto dos números reais.

Arimoto^[3] definiu uma medida de informação, que é conhecida como medida R-norm, da seguinte maneira.

Definição 6 - $H_R(P) = \frac{R}{R-1} \cdot [(\sum_{i=1}^n p_i^R)^{1/R} - 1]$ para $R \in \mathbb{R}, R \neq 1$

(1.10)

e $R > 0$.

Observação:

$$\lim_{R \rightarrow 1} H_R(P) = H(P).$$

Esta medida R-norm foi recentemente estudada também por Boekee e Van der Lubbe [22].

Sharma e Mittal [18] introduziram a entropia de ordem α e tipo β , que é definida da seguinte maneira:

Definição 7 - $H_\alpha^\beta(P) = (2^{1-\beta} - 1)^{-1} \cdot [(\sum_{i=1}^n p_i^\alpha)^{\frac{\beta-1}{\alpha-1}} - 1]$ para

(1.11)

$\alpha, \beta > 0, \alpha \neq \beta$ onde α, β são parâmetros e $P \in \Delta_n$.

Observações:

- 1) $\lim_{\beta \rightarrow 1} H_\alpha^\beta(P) = H_\alpha(P)$.
- 2) Quando $\alpha = \beta \neq 1$, $H_\alpha^\beta(P) = H^\beta(P)$
- 3) $\lim_{\alpha \rightarrow 1} \lim_{\beta \rightarrow 1} H_\alpha^\beta(P) = H(P)$.

1.5 - Propriedades da Entropia de Shannon e de suas generalizações.

Escreveremos as propriedades de maneira unificada pois a maioria delas é semelhante para as entropias apresentadas. Qua se todas as propriedades são válidas para $n = 2, 3, \dots$

Sejam $P = (p_1, \dots, p_n)$ e $Q = (q_1, \dots, q_n) \in \Delta_n$ duas distribuições de probabilidades finitas.

Especificaremos as várias propriedades para F_n , onde F_n representa a função de $\Delta_n \rightarrow \mathbb{R}$.

p_1 - Simetria:

Para toda permutação arbitrária $\{a_1, \dots, a_n\}$ de $\{1, 2, \dots, n\}$,

$$F_n(p_1, \dots, p_n) = F(p_{a_1}, \dots, p_{a_n})$$

p_2 - Recursividade:

$$F_n(p_1, \dots, p_n) = F_{n-1}(p_1+p_2, \dots, p_n) + \Delta(p_1, p_2) \cdot F_2\left(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2}\right)$$

Para todo $p_1 + p_2 > 0$ e todo $n = 3, 4, \dots$

onde $\Delta = f(p_1, p_2)$; $0 \leq p_1 < 1$, $0 \leq p_2 < 1$ e $p_1 + p_2 \leq 1$ e $f: \Delta_n \rightarrow \mathbb{R}$.

p_3 - Expansibilidade ou zero-indiferente:

$$F_n(p_1, \dots, p_n) = F_{n+1}(p_1, \dots, p_n, 0).$$

p_4 - Aditividade:

$$F_{mn}(p_1 q_1, \dots, p_1 q_m, \dots, p_n q_1, \dots, p_n q_m) = F_n(p_1, \dots, p_n) + F_m(q_1, \dots, q_m)$$

Para $P \in \Delta_n$, $Q \in \Delta_m$.

p_5 - Aditividade - forte

$$F_{mn}(p_{11}, \dots, p_{1m}, \dots, p_{n1}, \dots, p_{nm}) = F_m(p_1, \dots, p_n) + \\ + \sum_{i=1}^n p_i F_n(p_{ii}/p_i, \dots, p_{ni}/p_i)$$

Para $p_{ik} \geq 0$, $i = 1, \dots, n$, $\sum_{i=1}^n p_i = 1$ e $p_i = \sum_{k=1}^n p_{ik} > 0$.

p_6 - Sub-aditividade:

$$F_{mn}(p_{11}, \dots, p_{1m}, \dots, p_{m1}, \dots, p_{mn}) \leq F_m(\sum_{i=1}^n p_{ii}, \dots, \sum_{i=1}^n p_{mi}) + \\ + F_n(\sum_{j=1}^m p_{ji}, \dots, \sum_{j=1}^m p_{jn})$$

Para todos $(p_{1i}, \dots, p_{mi}) \in \Delta_m$,

$$(p_{j1}, \dots, p_{jn}) \in \Delta_n.$$

p_7 - Não-negatividade:

$$F_n(p_1, \dots, p_n) \geq 0,$$

Com a igualdade se e somente se um $p_i = 1$ e os demais são zero.

p_8 - Desigualdade:

$$F_n(p_1, \dots, p_n) \leq F_n(1/n, \dots, 1/n),$$

com a igualdade se e somente se $p_i = \frac{1}{n}$ para todo $i = 1, \dots, n$.

p_9 - Não-aditividade:

$$F_{mn}(p_1q_1, \dots, p_1q_m, \dots, p_nq_m) = F_n(p_1, \dots, p_n) + \\ + F_m(q_1, \dots, q_m) + c \cdot F_n(p_1, \dots, p_n) \cdot F_m(q_1, \dots, q_m)$$

Para todo $P \in \Delta_n$, $Q \in \Delta_m$ e c uma constante que não depende de P e Q .

p_{10} - Não-aditividade forte:

$$F_{mn}(p_1p_{11}, \dots, p_1p_{1n}, \dots, p_m p_{m1}, \dots, p_m p_{mn}) = \\ = F_m(p_1, \dots, p_m) + \sum_{i=1}^n \Delta(p_i) F_n(q_1, \dots, q_n)$$

Para $p_{ij} \geq 0$; $\sum_{j=1}^n p_{ij} = 1$ para todo $i = 1, \dots, m$.

$$p_j \geq 0; \quad \sum_{j=1}^n p_j = 1.$$

p_{11} - Representatividade:

$$\text{Soma arbitrária: } F_n(p_1, \dots, p_n) = \sum_{i=1}^n f(p_i)$$

onde $f: [0, 1] \rightarrow \mathbb{R}$.

p_{12} - Continuidade:

$F_n(p_1, \dots, p_n)$ é uma função contínua de suas variáveis.

p_{13} - Monotocidade:

$F_n(1/n, \dots, 1/n)$ é uma função monótona crescente de n .

P_{14} - Integrabilidade a Lebesgue:

$F_2(p, 1-p)$ é Lebesgue integrável em $[0, 1]$.

P_{15} - Mensurabilidade:

$F_2(p, 1-p)$ é uma função mensurável de $p \in [0, 1]$.

P_{16} - Normalidade:

$$F_2\left(\frac{1}{2}, \frac{1}{2}\right) = 1.$$

A entropia de Shannon tem as seguintes propriedades:

P_1, P_2 com $\Delta(p_1, p_2) = p_1 + p_2$; de P_3 a P_8, P_{10} com $\Delta(p_k) = p_k$;

P_{11} com $f(p_i) = -p_i \log p_i$; de P_{12} a P_{16} ,

$$P_{17} \Rightarrow H(p_1, \dots, p_{m-1}, p_m q_1, p_m q_2, \dots, p_m q_{n-m+1}) = H_m(p_1, \dots, p_m) + p_m H_{n-m+1}(q_1, \dots, q_{n-m+1})$$

$$\text{onde } \sum_{i=1}^m p_i = 1 \text{ e } \sum_{i=1}^{n-m+1} q_i = 1.$$

A entropia de ordem α satisfaz as seguintes propriedades: $P_1, P_3, P_4, P_7, P_8, P_{12}, P_{13}$ e P_{16} .

A entropia não aditiva do tipo β satisfaz as seguintes propriedades: $P_1; P_2$ com $\Delta(p_1, p_2) = (p_1 + p_2)^\beta$; $P_3; P_7; P_8; P_9$ com $c = (2^{1-\beta} - 1)$; P_{10} com $\Delta(p_{12}) = p_k^\beta$; $P_{12}; P_{13}$ e P_{16} .

A entropia do tipo (α, β) satisfaz as seguintes propriedades: $P_1, P_3, P_7, P_{11}, P_{12}$ e P_{16} .

A entropia de ordem (α, β) satisfaz: $P_1, P_3, P_5, P_7, P_{12}$ e P_{16} .

A entropia de ordem α e tipo β satisfaz: $p_1, p_3, p_7, p_8, p_{12}$ e p_{16} .

1.6 - Medida de Imprecisão

Seja X uma variável aleatória e sejam P, Q duas distribuições de probabilidades associadas com esta variável X . Kerridge^[12] introduziu a seguinte medida, denominada medida de imprecisão.

$$\text{Definição 8 - } H_S(P, Q) = - \sum_{i=1}^n p_i \log q_i \quad \text{para } P, Q \in \Delta_n. \quad (1.12)$$

Quando $P = Q \Rightarrow H_S(P, P) = H(P)$.

1.7 - Generalizações da Medida de Imprecisão

Definição 9 - Medida de imprecisão de ordem α [21].

$$H_\alpha(P, Q) = \frac{1}{1-\alpha} \cdot \log \left(\frac{\sum_{i=1}^n p_i^\alpha}{\sum_{i=1}^n p_i^\alpha q_i^{1-\alpha}} \right), \quad \alpha > 0, \quad \alpha \neq 1. \quad (1.13)$$

Observações:

1) Quando $P = Q \Rightarrow H_\alpha(P, P) = H(P)$.

2) $\lim_{\alpha \rightarrow 1} H_\alpha(P, Q) = H_S(P, Q)$.

Definição 10 - Medida de imprecisão do tipo β [21].

$$H^\beta(P, Q) = \frac{1}{2^{1-\beta} - 1} \cdot \left(\frac{\sum_{i=1}^n p_i^\beta}{\sum_{i=1}^n p_i^\beta q_i^{1-\beta}} - 1 \right) \text{ para } \beta > 0 ; \beta \neq 1 \quad (1.14)$$

Observações:

- 1) Quando $P = Q \Rightarrow H^\beta(P, P) = H^\beta(P)$.
- 2) $\lim_{\beta \rightarrow 1} H^\beta(P, Q) = H(P, Q)$.

Definição 11 - Medida de imprecisão de ordem (α, β) .

$$H_{\alpha, \beta}(P, Q) = \frac{1}{\beta - \alpha} \cdot \log \left(\frac{\sum_{i=1}^n p_i^\alpha \sum_{i=1}^n p_i^\beta q_i^{1-\beta}}{\sum_{i=1}^n p_i^\alpha q_i^{1-\alpha} \cdot \sum_{i=1}^n p_i^\beta} \right) \text{ para } \alpha, \beta > 0 ; \alpha \neq \beta. \quad (1.15)$$

Observações:

- 1) $\lim_{\alpha \rightarrow 1} H_{\alpha, \beta}(P, Q) = H_\beta(P, Q)$.
- 2) $\lim_{\alpha \rightarrow 1} \lim_{\beta \rightarrow 1} H_{\alpha, \beta}(P, Q) = H_S(P, Q)$.

Definição 12 - Medida de imprecisão do tipo (α, β) .

$$H^{\alpha, \beta}(P, Q) = (2^{1-\alpha} - 2^{1-\beta})^{-1} \cdot \left(\frac{\sum_{i=1}^n p_i^\alpha}{\sum_{i=1}^n p_i^\alpha q_i^{1-\alpha}} - \frac{\sum_{i=1}^n p_i^\beta}{\sum_{i=1}^n p_i^\beta q_i^{1-\beta}} \right) \text{ para } \alpha, \beta > 0 ; \alpha \neq \beta. \quad (1.16)$$

Observações:

- 1) $\lim_{\alpha \rightarrow 1} H^{\alpha, \beta}(P, Q) = H^\beta(P, Q)$

$$2) \lim_{\beta \rightarrow 1} \lim_{\alpha \rightarrow 1} H^{\alpha, \beta}(P, Q) = H_S(P, Q).$$

No próximo capítulo mostraremos relações que envolvem as entropias definidas em (1.3), (1.4) e (1.5) com suas respectivas medidas de imprecisão.

CAPÍTULO 2

ENTROPIA DE ORDEM α E DE TIPO β e TEOREMAS
DE CODIFICAÇÃO

Neste capítulo vamos mostrar a desigualdade de Shannon e suas generalizações no caso da entropia de Renyi e entropia do tipo β . Como aplicações, provaremos os teoremas de codificação.

2.1 - Desigualdade de Shannon

No seguinte teorema, a desigualdade de Shannon será provada.

Teorema 1^[4] - Sejam $P, Q \in \Delta_n$, então temos:

$$H(P) = - \sum_{i=1}^n p_i \log p_i \leq - \sum_{i=1}^n p_i \log q_i = H(P, Q), \quad (2.1)$$

Observação: $\log x = \log e \cdot \log_e x$

Prova:

Sabemos que logaritmo é uma função convexa, $\log_e x$ encontra-se sempre abaixo de sua tangente.

Considerando a tangente para $x = 1$, obtemos $\log_e x \leq x - 1$ com a igualdade se e somente se $x = 1$.

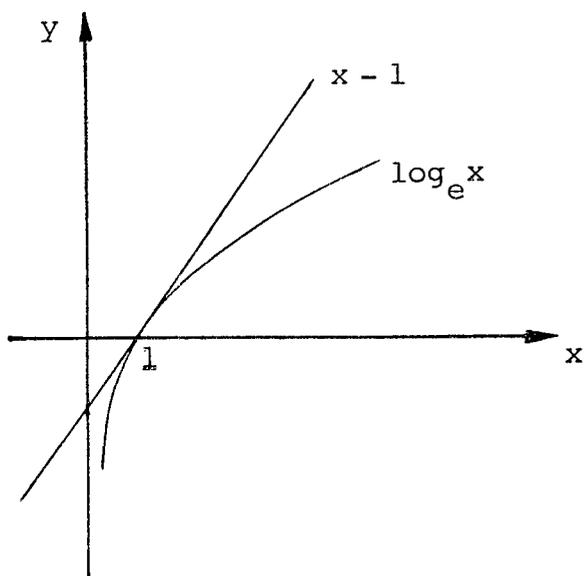


Figura 1.3

Seja $x = q_i/p_i$

então $\log_e q_i/p_i \leq q_i/p_i - 1$

ou $\log_e q_i - \log_e p_i \leq q_i/p_i - 1$

$p_i \log_e q_i - p_i \log_e p_i \leq q_i - p_i$

Somando sob $(i = 1, \dots, n)$ temos:

$$\sum_{i=1}^n p_i \log_e q_i - \sum_{i=1}^n p_i \log_e p_i \leq 0$$

$$- \sum_{i=1}^n p_i \log_e p_i \leq - \sum_{i=1}^n p_i \log_e q_i$$

$$- \sum_{i=1}^n p_i \log p_i \leq - \sum_{i=1}^n p_i \log q_i$$

Com a igualdade se e somente se $x = q_i/p_i = 1$.

Usaremos esta desigualdade nos teoremas seguintes.

2.2 - Codificação sem ruído

Codificação sem ruído significa que para transmitir uma mensagem usaremos um sistema como dado na figura 1.1 com um canal sem ruído.

Usaremos este canal porque nos permite uma transmissão correta da entrada para a saída.

Tentamos maximizar o número de mensagens que pode ser transmitido através do canal num dado tempo.

Tomamos $X = \{x_1, \dots, x_n\}$ como um conjunto de mensagens e $A = \{a_1, \dots, a_D\}$ um conjunto com D símbolos chamado de código alfabeto.

Se os símbolos x_i forem comunicados propriamente, então cada x_i , $i = 1, \dots, n$, deve ser representado por uma seqüência de símbolos do código A . Sendo cada seqüência chamada de palavra código w_i de comprimento N_i .

x	probabilidade	palavra código	comprimento
x_1	P_1	w_1	N_1
x_2	P_2	w_2	N_2
\vdots	\vdots	\vdots	\vdots
x_n	P_n	w_n	N_n

Comunicar perfeitamente significa transmitir uma mensagem dada no tempo mínimo possível.

Se a taxa de transmissão através do canal é fixa, então para essa transmissão ser eficiente implica que as palavras códigos deverão ter o menor comprimento possível. Logo, o objetivo da codificação sem ruído é minimizar o comprimento médio do código dado por:

$$L = \sum_{i=1}^n p_i N_i \quad (2.2)$$

Observação: Ao longo deste trabalho, usaremos códigos binários, isto é, $A = \{a_1, a_2\}$ caso não venha especificado.

O problema da decifrababilidade única

Exemplo 1: Sejam $X = \{x_1, x_2, x_3, x_4\}$ e $A = \{0, 1\}$

$x_1 \rightarrow 0$

$x_2 \rightarrow 010$

$x_3 \rightarrow 01$

$x_4 \rightarrow 10$

Mas podemos decifrar 010 como x_2, x_3x_1, x_1x_4 .

Um código é dito decifrável unicamente se toda seqüência finita de símbolos do código corresponde no máximo uma mensagem.

Para não ocorrer o problema dado no exemplo 1, deve-se codificar de modo que uma palavra não seja prefixo da outra. Se um código tem esta propriedade então é chamado de código instantâneo.

Exemplo 2: Sejam $X = \{x_1, \dots, x_4\}$ e $A = \{0, 1\}$

$x_1 \rightarrow 0$

$x_2 \rightarrow 100$

$x_3 \rightarrow 110$

$x_4 \rightarrow 101$

Podemos decifrar 010100100110 somente como $x_1x_4x_1x_1x_2x_3$

Todo código instantâneo é decifrável unicamente. A recíproca é falsa.

Teorema 2 - Um código instantâneo com comprimentos das palavras N_i existe se e somente se valer a desigualdade de Kraft dada por $\sum_{i=1}^n D^{-N_i} \leq 1$.

Prova: Veja Aczél^[2].

Exemplo 3: Sejam $X = \{x_1, \dots, x_9\}$, $A = \{0, 1, 2\}$, $D = 3$ com comprimento das palavras 1, 2, 2, 2, 2, 2, 3, 3, 3.

$$\sum_{i=1}^9 D^{-N_i} = \sum_{i=1}^9 3^{-N_i} = 3^{-1} + 5 \cdot 3^{-2} + 3 \cdot 3^{-3} = 1$$

$\sum_{i=1}^9 D^{-N_i} = 1 \rightarrow$ existe um código instantâneo com estes comprimentos das palavras.

Então temos:

$x_1 \rightarrow 0$	$x_6 \rightarrow 21$
$x_2 \rightarrow 10$	$x_7 \rightarrow 220$
$x_3 \rightarrow 11$	$x_8 \rightarrow 221$
$x_4 \rightarrow 12$	$x_9 \rightarrow 222$
$x_5 \rightarrow 20$	

Temos visto que a condição $\sum_{i=1}^n D^{-N_i} \leq 1$ é necessária e suficiente para existência de um código instantâneo com comprimento das palavras N_1, \dots, N_n . No seguinte teorema será enunciado um resultado forte, que a mesma condição é necessária e suficiente para existência de um código decifrável unicamente. A parte sufi-

ciente é imediata porque todo código instantâneo é decifrável unicamente.

Teorema 3 - Se um código é decifrável unicamente então $\sum_{i=1}^n D^{-N_i} \leq 1$,
onde D é a dimensão do código.

Como iremos trabalhar com códigos binários, podemos dizer que se um código é decifrável unicamente então:

$$\sum_{i=1}^n 2^{-N_i} \leq 1 \quad (2.3)$$

Prova: Veja Aczél [2]

A seguir daremos uma cota inferior de L .

Teorema 4 - Teorema de codificação sem ruído [4].

Se $L = \sum_{i=1}^n p_i N_i$ é o comprimento médio das palavras códigos de um código decifrável unicamente, então:

$$L \geq H(P) \quad (2.4)$$

com a igualdade se e somente se $p_i = D^{-N_i}$, $i = 1, \dots, n$

Prova:

Usando a desigualdade de Shannon para $q_i = \frac{2^{-N_i}}{\sum_{i=1}^n 2^{-N_i}}$,

temos:

$$\begin{aligned} - \sum_{i=1}^n p_i \log p_i &\leq - \sum_{i=1}^n p_i \log \frac{2^{-N_i}}{\sum_{i=1}^n 2^{-N_i}} = \\ &= \sum_{i=1}^n p_i \log \sum_{i=1}^n 2^{-N_i} - \sum_{i=1}^n p_i \log 2^{-N_i} = \end{aligned}$$

$$\begin{aligned}
 &= \sum_{i=1}^n p_i N_i \log 2 + \log \sum_{i=1}^n 2^{-N_i} \\
 &= L + \log \sum_{i=1}^n 2^{-N_i} \tag{2.5}
 \end{aligned}$$

com a igualdade se e somente se $p_i = \frac{2^{-N_i}}{\sum_{i=1}^n 2^{-N_i}}$.

por (2.3) segue que $\log \sum_{i=1}^n 2^{-N_i} \leq \log 1 = 0$ (2.6)

$$\Rightarrow - \sum_{i=1}^n p_i \log p_i \leq L.$$

Logo, $L \geq H(P)$.

Nos resta mostrar que $p_i = 2^{-N_i}$, $i = 1, 2, \dots, n$.

de (2.5) temos que $H(P) \leq L + \log \sum_{i=1}^n 2^{-N_i}$

$$H(P) = L \Leftrightarrow \log \sum_{i=1}^n 2^{-N_i} \geq 0$$

Por (2.6) temos $\log \sum_{i=1}^n 2^{-N_i} = 0$

$$\text{Logo, } \sum_{i=1}^n 2^{-N_i} = 1 = \sum_{i=1}^n p_i \Rightarrow 2^{-N_i} = p_i.$$

Definição 1 - Um código que é limitado inferiormente ($L \geq H(P)$) é denominado código absolutamente ótimo.

Em geral não podemos construir um código absolutamente ótimo para um dado conjunto de probabilidade p_1, \dots, p_n , pois se escolhermos N_i satisfazendo $p_i = D^{-N_i} = -\log p_i$ não pode ser

um inteiro positivo sempre.

Temos uma melhor possibilidade com o teorema seguinte.

Teorema 5 - Dado uma variável aleatória $X = (x_1, \dots, x_n)$ com incerteza $H(X)$, existe um código instantâneo de dimensão 2 cujo comprimento médio do código, satisfaz:

$$H(X) \leq L < H(X) + 1.$$

Prova:

Sejam os intervalos reais $\delta_i = [-\log p_i, -\log p_i + 1)$.
(2.7)

Em cada δ_i , existe exatamente um inteiro positivo N_i tal que:

$$0 < -\log p_i \leq N_i < -\log p_i + 1, \quad i = 1, \dots, n. \quad (2.8)$$

Temos uma sucessão de números inteiros positivos $\{N_1, N_2, \dots, N_n\}$. É possível construir um código instantâneo com comprimentos das palavras códigos N_1, \dots, N_n pois $N_i \geq -\log p_i$ por (2.7).

$$-N_i \leq \log p_i$$

$$\log 2^{-N_i} \leq \log p_i \Rightarrow 2^{-N_i} \leq p_i$$

$$\text{Logo } \sum_{i=1}^n 2^{-N_i} \leq \sum_{i=1}^n p_i = 1.$$

Portanto existe um código instantâneo com comprimento das palavras códigos N_1, \dots, N_n .

De (2.8) segue que:

$$-\log p_i \leq N_i < -\log p_i + 1.$$

$$-\sum_{i=1}^n p_i \log p_i \leq \sum_{i=1}^n p_i N_i < -\sum_{i=1}^n \log p_i + 1.$$

Logo, $H(X) \leq L < H(X) + 1$.

Podemos sempre aproximar a cota inferior como desejamos se permitimos o uso do código em blocos. Pois em geral, este código decresce o comprimento médio das palavras códigos por valor de X .

Ao invés de relacionarmos uma palavra código para cada símbolo X_i , tomaremos uma série de s observações independentes de X e relacionaremos uma palavra código para um grupo de s símbolos. Ou seja, construímos um código para um vetor aleatório $Y = (X_1, \dots, X_s)$ onde cada X_i , $i = 1, \dots, s$ são independentes e cada X_i tem a mesma distribuição de probabilidades de X .

Exemplo 4:

a) Seja $Y = (x_1, x_2)$, $A = \{0, 1\}$

X	prob.	palavra	comp.
x_1	3/4	0	1
x_2	1/4	1	1

$$L = \frac{3}{4} \cdot 1 + \frac{1}{4} \cdot 1 = 1$$

b) $Y = (x_1, x_2)$, $A = \{0, 1\}$

Y	prob.	palavra	comp.
$x_1 x_1$	9/16	0	1
$x_1 x_2$	3/16	10	2
$x_2 x_1$	3/16	110	3
$x_2 x_2$	1/16	111	3

$$L = \frac{9}{16} \cdot 1 + \frac{3}{16} \cdot 2 + \frac{3}{16} \cdot 3 + \frac{1}{16} \cdot 3 = \frac{27}{16}$$

$$\therefore L = \frac{27}{16} \text{ para dois valores de } X.$$

$$\therefore L = \frac{27}{32} \text{ para um valor de } X.$$

Usando esses argumentos mostramos o seguinte teorema^[4].

Teorema 6 - Para cada inteiro positivo s , existe um código instantâneo de X^s tal que L_s é o comprimento médio, então temos:

$$\lim_{s \rightarrow \infty} \frac{L_s}{s} = H(P).$$

Prova:

Suponhamos $Y = (X_1, \dots, X_s)$ como dado anteriormente, e que L_s é o comprimento médio das palavras códigos para Y , então:

$$H(Y) \leq L_s < H(Y) + 1. \quad (2.9)$$

$$\begin{aligned} \text{Como } H(Y) &= H(X_1, \dots, X_s) \\ &= H(X_1) + \dots + H(X_s) \quad \text{pois } x_i \text{ são independentes.} \\ &= H(X) + \dots + H(X) \\ &= s H(X) \end{aligned}$$

Então, (2.9) pode ser reescrita da seguinte maneira:

$$s H(X) \leq L_s < s H(X) + 1$$

$$H(X) \leq \frac{L_s}{s} < H(X) + \frac{1}{s}$$

onde $\frac{L_s}{s}$ é o comprimento médio do código para cada valor de X .

Logo, $\lim_{S \rightarrow \infty} \frac{L_S}{S} = H(P)$.

Este teorema nos diz que cifrando longas seqüências de entrada é possível fazer o comprimento médio das palavras códigos para cada símbolo de entrada tão próximo de $H(P)$ quanto se queira.

2.3 - Entropia do tipo β .

2.3.1 - Generalização da desigualdade de Shannon.

Nath e Mittal^[13] em 1973 fizeram uma generalização da desigualdade de Shannon e definiram uma outra medida de comprimento.

Antes, necessitamos de algumas definições.

Definição 2 - Seja $P = (p_1, \dots, p_n) \in \Delta_n$. A distribuição de probabilidade $Q = (q_1, \dots, q_n) \in \Delta_n$ é dita uma distribuição de potência β derivável de P se existe um número real β tal que $q_i = p_i^{(\beta)}$,

$$\text{onde } p_i^{(\beta)} = \frac{p_i^\beta}{W_\beta(P)}, \quad i = 1, \dots, n \quad \text{e} \quad W_\beta(P) = \sum_{i=1}^n p_i^\beta. \quad (2.10)$$

Quando (2.10) é válida, escreveremos $Q = P^{(\beta)}$.

Definição 3 - Sejam $P, Q \in \Delta_n$.

Uma distribuição de probabilidades $R = (r_1, r_2, \dots, r_n) \in \Delta_n$ é dita uma distribuição de potência (β, γ) derivável de P, Q se existem números reais β e γ tal que

$r_i = r_i(\beta, \gamma)$ onde

$$r_i(\beta, \gamma) = \frac{p_i^\beta q_i^\gamma}{\sum_{i=1}^n p_i^\beta q_i^\gamma} \quad \text{para } i = 1, 2, \dots, n. \quad (2.11)$$

Quando (2.11) é válida, escreveremos

$$R = R_{(\beta, \gamma)}(P, Q)$$

$$P^{(x)} = R_{(x, 0)}(P, Q), \quad Q^{(y)} = R_{(0, y)}(P, Q)$$

$$P^{(x+y)} = R_{(x, y)}(P, P).$$

Lema 1 - Para duas distribuições de probabilidades $P, Q \in \Delta_n$, segue que:

$$H^\beta(P) \leq H^\beta(\tilde{R}_\beta, Q); \quad \beta > 0, \quad \beta \neq 1 \quad (2.12)$$

onde $\tilde{R}_\beta = R_{\beta, 1-\beta}(P, Q)$.

Prova:

$$\text{Tome } F_r(P \| Q) = \sum_{i=1}^n p_i^r q_i^{1-r}$$

$$W_r(P) = \sum_{i=1}^n p_i^r$$

(i) Seja $0 < \beta < 1$,

Usando a desigualdade de Hölder $\sum_{i=1}^n p_i a_i^r \leq (\sum_{i=1}^n p_i a_i)^r$

para $r < 1$, onde p_i, a_i são números não negativos sobre um conjunto finito de $i, 1 \leq i \leq n$ e $\sum_{i=1}^n p_i = 1$.

Fazendo $r = \beta$, $a_i = p_i/q_i$ temos que $F_\beta(P \| Q) \leq 1$. (2.13)

A igualdade em (2.13) ocorre se e somente se $p_i = q_i$, $i = 1, \dots, n$.

Segue que: $1 - \frac{W_\beta(P)}{F_\beta(P \parallel Q)} \leq 1 - W_\beta(P)$

Como $(1 - 2^{1-\beta})^{-1} < 0$, temos:

$$(1 - 2^{1-\beta})^{-1} \cdot \left(1 - \frac{W_\beta(P)}{F_\beta(P \parallel Q)}\right) \geq (1 - 2^{1-\beta})^{-1} \cdot (1 - W_\beta(P))$$

o que demonstra o lema para $0 < \beta < 1$.

(ii) Para $1 < \beta < \infty$ a prova segue similarmente.

2.3.2 - Pseudo-generalização da desigualdade de Shannon

Definiremos outra medida para duas distribuições de pro babilidades $P, Q \in \Delta_n$ como segue:

Definição 4 - $H^{1\beta}(P, Q) = (1 - 2^{1-\beta})^{-1} \cdot \left[1 - \left(\sum_{i=1}^n p_i q_i^{\frac{\beta-1}{\beta}}\right)\right]; \beta > 0; \beta \neq 1$.

(2.14)

Esta medida não deve ser interpretada como uma medida de imprecisão desde que $H^{1\beta}(P, P) \neq H^\beta(P)$.

Mas $H^{1\beta}(P, Q)$ é uma generalização da medida de imprecisão $H_S(P, Q)$ como definida em (1.12).

No teorema seguinte, daremos uma relação entre (2.14) e $H^\beta(P)$, que é conhecida como pseudo-generalização da desigualdade de Shannon dado que $H^{1\beta}(P, Q)$ não é uma medida de imprecisão no sentido usual.

Teorema 5 - Se $P, Q \in \Delta_n$, segue que: [21]

$$(1-2^{1-\beta})^{-1} \cdot (1 - \sum_{i=1}^n p_i^\beta) \leq (1-2^{1-\beta})^{-1} \cdot [1 - (\sum_{i=1}^n p_i q_i^{\frac{\beta-1}{\beta}})^\beta] ;$$

$$\beta > 0 , \beta \neq 1 , \quad (2.15)$$

com a igualdade se e somente se $q_i = \frac{p_i^\beta}{\sum_{i=1}^n p_i^\beta}$; $i = 1, \dots, n$.

Prova:

Usando a desigualdade de Hölder, $\sum_{i=1}^n a_i b_i \geq (\sum_{i=1}^n a_i^r)^{1/r} \cdot (\sum_{i=1}^n b_i^s)^{1/s}$ para a_i, b_i números não negativos definidos sobre um conjunto finito de i ; $1 \leq i \leq n$.

Fazendo $a_i = p_i^{\frac{\beta}{\beta-1}} \cdot q_i$

$$b_i = p_i^{\frac{-\beta}{\beta-1}}$$

$$r = \frac{\beta-1}{\beta} \text{ e } s = 1 - \beta$$

temos: $(\sum_{i=1}^n p_i q_i^{\frac{\beta-1}{\beta}})^{\frac{\beta}{\beta-1}} \cdot (\sum_{i=1}^n p_i^\beta)^{\frac{1}{1-\beta}} \leq 1$ para $\beta > 0$; $\beta \neq 1$.

(2.16)

Consideremos dois casos:

1) Para $\beta > 1$,

$$(2.16) \text{ reduz-se para } \sum_{i=1}^n p_i^\beta \geq (\sum_{i=1}^n p_i q_i^{\frac{\beta-1}{\beta}})^\beta \quad (2.17)$$

Como $(1 - 2^{1-\beta})^{-1} > 0$; (2.17) reduz-se para (2.15).

2) Para $0 < \beta < 1$,

$$(2.16) \text{ reduz-se para } \sum_{i=1}^n p_i^\beta \leq \left(\sum_{i=1}^n p_i q_i^{\frac{\beta-1}{\beta}} \right)^\beta \quad (2.18)$$

Como $(1 - 2^{1-\beta})^{-1} < 0$; (2.18) reduz-se para (2.15).

Logo, temos provado o teorema para $\beta > 0$; $\beta \neq 1$.

2.3.3 - Aplicação no teorema de codificação.

Sabemos por Einstein^[8] (1958) que sempre existe um código decifrável unicamente com comprimentos N_1, \dots, N_n se e somente se $\sum_{i=1}^n D^{-N_i} \leq 1$.

Em geral escolhemos códigos que possui o comprimento mínimo.

$$L = \sum_{i=1}^n p_i N_i.$$

Apresentamos a seguir três medidas de comprimento do tipo β .

Definição 5 - $L_\beta = (1 - 2^{1-\beta})^{-1} \cdot \left[1 - \frac{1}{\sum_{i=1}^n p_i^{(\beta)} 2^{N_i(\beta-1)}} \right]$ para $\beta \neq 1$; $\beta > 0$,

$$(2.19)$$

$$\text{onde } p_i^{(\beta)} = \frac{p_i^\beta}{\sum_{i=1}^n p_i^\beta} \quad \text{e}$$

$$\lim_{\beta \rightarrow 1} L_\beta = L.$$

Teorema 6 - Se N_1, \dots, N_n denotam os comprimentos de um código decifrável unicamente satisfazendo (2.3), então:^[13]

$$L_\beta \geq H^\beta(P), \quad \beta \neq 1 ; \beta > 0 \quad (2.20)$$

Prova:

$$\text{Escolhemos } Q \in \Delta_n \text{ tal que } q_i = \frac{2^{-N_i}}{\sum_{i=1}^n 2^{-N_i}}, \quad i = 1, \dots, n. \quad (2.21)$$

em (2.12) segue que:

$$(1 - 2^{1-\beta})^{-1} \cdot \left(1 - \sum_{i=1}^n p_i^\beta\right) \leq (1 - 2^{1-\beta})^{-1} \left[1 - \frac{(\sum_{i=1}^n p_i^\beta) \cdot (\sum_{i=1}^n 2^{-N_i})^{1-\beta}}{\sum_{i=1}^n p_i^\beta 2^{N_i(\beta-1)}}\right]$$

Usando (2.3) temos:

$$H^\beta(P) \leq (1 - 2^{1-\beta})^{-1} \cdot \left[1 - \frac{\sum_{i=1}^n p_i^\beta}{\sum_{i=1}^n p_i^\beta 2^{N_i(\beta-1)}}\right] = L_\beta.$$

Portanto, demonstramos o teorema para $0 < \beta < 1$, $1 < \beta < \infty$.

A igualdade em (2.20) é válida se e somente se $p_i = \frac{2^{-N_i}}{\sum_{i=1}^n 2^{-N_i}}$

isto implica em $\log \frac{1}{p_i} = N_i + \log \left(\sum_{i=1}^n 2^{-N_i}\right)$.

Portanto, é sempre possível ter uma palavra código satisfazendo:

$$\log \frac{1}{p_i} \leq N_i < \log \frac{1}{p_i} + 1 \quad (2.22)$$

$$\frac{1}{p_i} \leq 2^{N_i} < \frac{2}{p_i} \quad (2.23)$$

Por (2.20), $H^\beta(P)$ é a cota inferior de L_β .

É possível encontrar também uma cota superior de L_β .

Então, temos o seguinte teorema.

Teorema 7 - Para um código decifrável unicamente satisfazendo (2.3) segue: [13]

$$(2^\beta - 2)L_\beta < 2^\beta - \sum_{i=1}^n p_i^\beta \quad \text{para } \beta > 0; \beta \neq 1. \quad (2.24)$$

Prova:

De (2.23) segue que $2^{N_i} \cdot p_i < 2$

$$2^{(\beta-1)N_i} \cdot p_i^\beta < 2^\beta \cdot 2^{-N_i}$$

Somando ambos os lados com respeito a i e usando (2.3) temos:

$$\sum_{i=1}^n p_i^\beta 2^{(\beta-1)N_i} < 2^\beta$$

$$\text{Então, } 1 - \frac{\sum_{i=1}^n p_i^\beta}{\sum_{i=1}^n p_i^\beta 2^{(\beta-1)N_i}} < 1 - \frac{\sum_{i=1}^n p_i^\beta}{2^\beta}$$

Multiplicando ambos os lados por $(1 - 2^{1-\beta})^{-1}$ encontramos

$$2^\beta L_\beta < (1 - 2^{1-\beta})^{-1} \cdot (2^\beta - \sum_{i=1}^n p_i^\beta)$$

$$2^\beta (1 - 2^{1-\beta}) \cdot L_\beta < 2^\beta - \sum_{i=1}^n p_i^\beta. \quad (2.25)$$

A cota superior em (2.25) é maior do que a unidade.

Quando $\beta \rightarrow \infty \Rightarrow L_\beta \rightarrow 1$.

Como $H^\beta(P)$ não é aditiva, o teorema de codificação sem ruído não segue no caminho usual.

$$\text{Definição 6 - } L_\beta = (1 - 2^{1-\beta})^{-1} \cdot \left[1 - \frac{\sum_{i=1}^n p_i^\beta \left(\sum_{i=1}^n 2^{-N_i} \right)^{1-\beta}}{\sum_{i=1}^n p_i^\beta 2^{(\beta-1)N_i}} \right], \beta > 0; \beta \neq 1 \quad (2.26)$$

A seguir mostraremos uma relação entre esta medida e $H^\beta(P)$, [21].

Teorema 8 - Se N_i ; $i = 1, 2, \dots, n$ são os comprimentos das palavras códigos satisfazendo (2.3) então:

$$H^\beta(P) \leq L_\beta < 2^{-\beta} \cdot H^\beta(P) + \frac{1 - 2^{-\beta}}{1 - 2^{1-\beta}} \quad (2.27)$$

O limite superior de L_β é válido somente para $\beta > 1$.

Prova:

Substituindo $q_i = \frac{2^{-N_i}}{\sum_{i=1}^n 2^{-N_i}}$ em (2.12) segue que

$$H^\beta(P) \leq L_\beta.$$

A igualdade em (2.27) acontece se e somente se $p_i = \frac{2^{-N_i}}{\sum_{i=1}^n 2^{-N_i}}$. (2.28)

Mostraremos agora, a cota superior de L_β em (2.27).

Escolhemos N_i , $i = 1, \dots, n$ satisfazendo (2.28).

Então, (2.28) pode ser reescrito como:

$$-\log p_i = \log 2^{N_i} + \log \sum_{i=1}^n 2^{-N_i} \quad (2.29)$$

Como $\sum_{i=1}^n 2^{-N_i} \leq 1 \Rightarrow N_i \geq -\log p_i$.

Por outro lado, como $N_i \geq 1$ e usando (2.29), segue que:

$$N_i < -\log p_i + 1.$$

Combinando estes resultados, temos que:

$$-\log p_i \leq N_i < -\log p_i + 1.$$

O qual é equivalente a
$$\frac{1}{p_i} \leq 2^{-N_i} < \frac{2}{p_i} \quad (2.30)$$

Assim, é sempre possível ter um código satisfazendo (2.30).

De (2.30) segue que para $\beta > 1 \Rightarrow p_i^\beta 2^{(\beta-1)N_i} \leq 2^\beta \cdot 2^{-N_i}$.

Somando ambos os lados com respeito a i , temos:

$$\sum_{i=1}^n p_i^\beta 2^{(\beta-1)N_i} \leq \sum_{i=1}^n 2^\beta \cdot 2^{-N_i} \leq 2^\beta \quad (2.31)$$

Como $\beta > 1$, $(\sum_{i=1}^n 2^{-N_i})^{1-\beta} \geq 1$ e assim segue que:

$$\frac{\sum_{i=1}^n p_i^\beta (\sum_{i=1}^n 2^{-N_i})^{1-\beta}}{\sum_{i=1}^n p_i^\beta 2^{(\beta-1)N_i}} \geq \frac{\sum_{i=1}^n p_i^\beta}{\sum_{i=1}^n p_i^\beta 2^{(\beta-1)N_i}} > \frac{\sum_{i=1}^n p_i^\beta}{2^\beta}$$

$$\text{Como } (1 - 2^{1-\beta})^{-1} < 0 \Rightarrow L_\beta < (1 - 2^{1-\beta})^{-1} \cdot (1 - \frac{\sum_{i=1}^n p_i^\beta}{2^\beta}) \quad (2.32)$$

Usando a definição de $H^\beta(P)$, o lado direito de (2.32) pode ser reescrito como uma cota superior de L_β , como temos em (2.27).

Observações:

1) Para $\beta \geq 1$, a cota superior de L_β em (2.27) é maior do que a unidade.

Contudo, quando $\beta \rightarrow \infty \Rightarrow L_\beta \rightarrow 1$.

2) Em (2.19) definimos uma medida de comprimento que é independente do valor $(\sum_{i=1}^n 2^{-N_i})$. Vamos representá-la por L_β^* .

Desde que $(\sum_{i=1}^n 2^{-N_i})^{1-\beta} > 1$ para $\beta > 1$ e

$(\sum_{i=1}^n 2^{-N_i})^{1-\beta} < 1$ para $0 < \beta < 1$. Pode-se mostrar que $L_\beta \leq L^*$ usando (2.27), então segue que: $H^\beta(P) \leq L_\beta^*$, onde a igualdade aparece se e somente se $p_i = 2^{-N_i}$, $i = 1, \dots, n$.

Escolhendo $\beta > 1$, e supondo que todo N_i satisfaz (2.30) e usando (2.32) com $(1 - 2^{1-\beta})^{-1} > 0$ resulta em

$$L_\beta \leq L_\beta^* < 2^{-\beta} \cdot H^\beta(P) + \frac{2^{-\beta}}{1 - 2^{1-\beta}}.$$

Comparando este resultado com (2.27) segue que: L_β e L_β^* têm as mesmas cotas inferiores e superiores.

Então, $\lim_{\beta \rightarrow 1} L_\beta^* = L$.

Assim, temos uma medida de comprimento, que é uma generalização do comprimento médio L , independente de $(\sum_{i=1}^n 2^{-N_i})$.

Definição 7 - $L_\beta^1 = (1 - 2^{1-\beta})^{-1} \cdot [1 - (\sum_{i=1}^n p_i 2^{N_i \frac{1-\beta}{\beta}})]^\beta$; $\beta > 0$; $\beta \neq 1$.

(2.34)

No teorema seguinte, provaremos uma relação entre esta medida de comprimento e $H^\beta(P)$, [21].

Teorema 9 - Se N_i , $i = 1, \dots, n$, são os comprimentos das palavras códigos satisfazendo (2.3) então:

$$H^\beta(P) \leq L_\beta^1 < 2^{1-\beta} \cdot H^\beta(P) + 1, \quad \beta > 1 \quad (0 < \beta < 1) \quad (2.35)$$

Prova:

$$\text{Substituindo } q_i = \frac{2^{-N_i}}{\sum_{i=1}^n 2^{-N_i}}, \quad i = 1, \dots, n, \quad \text{em (2.15)}$$

e usando a desigualdade dada em (2.3) encontramos a primeira parte de (2.35).

A seguir, encontraremos a cota superior de L_β^1 .

$$\text{Temos } L_\beta^1 = H^\beta(P) \quad \text{se e somente se } 2^{-N_i} = \frac{p_i^\beta}{\sum_{i=1}^n p_i^\beta} \quad \text{que é equivalente a } N_i = -\log p_i^\beta + \log \left(\sum_{i=1}^n p_i^\beta \right), \quad i = 1, \dots, n.$$

Escolhendo N_i , tal que:

$$-\log \left(\frac{p_i^\beta}{\sum_{i=1}^n p_i^\beta} \right) \leq N_i \quad -\log \left(\frac{p_i^\beta}{\sum_{i=1}^n p_i^\beta} \right) + 1 \quad (2.36)$$

$$\text{segue que:} \quad 2^{-N_i} > \frac{p_i^\beta}{\sum_{i=1}^n p_i^\beta}. \quad (2.37)$$

$$\text{Como } \beta > 1, \quad (2.37) \quad \left[\sum_{i=1}^n p_i^\beta 2^{-N_i \left(\frac{\beta-1}{\beta} \right)} \right]^\beta > \sum_{i=1}^n p_i^\beta 2^{1-\beta} \quad (2.38)$$

$$\text{e finalmente encontramos } L_\beta^1 < \frac{1 - \sum_{i=1}^n p_i^\beta \cdot 2^{1-\beta}}{1 - 2^{1-\beta}} \quad (2.39)$$

$$\text{onde } \frac{1 - \sum_{i=1}^n p_i^\beta \cdot 2^{1-\beta}}{1 - 2^{1-\beta}} \quad \text{é equivalente a } 2^{1-\beta} \cdot H^\beta(P) + 1.$$

Portanto, demonstramos o teorema.

Para $0 < \beta < 1$, a prova da cota superior de L_β^1 segue similarmente.

Observações:

1) A cota superior de L_{β}^1 dado em (2.35) é maior do que a unidade.

Contudo, quando $\beta \rightarrow \infty$, encontramos $L_{\beta}^1 \rightarrow 1$.

2) Quando $\beta \rightarrow 1$.

$$\lim_{\beta \rightarrow 1} L_{\beta}^1 = L.$$

Então, (2.35) pode ser reescrita como:

$$H(P) \leq L < H(P) + 1$$

que é também conhecido com o resultado dado por Shannon, veja Aczél (1975).

2.4 - Entropia de Ordem α

Sabemos de (1.4) que esta entropia foi definida por Renyi (1961) do seguinte modo:

$$H_{\alpha}(P) = (1 - \alpha)^{-1} \cdot \log \sum_{i=1}^n p_i^{\alpha}, \quad \alpha > 0 \text{ e } \alpha \neq 1.$$

2.4.1 - Generalização da desigualdade de Shannon.

Teorema 10 - Se $P, Q \in \Delta_n$ são distribuições de probabilidades completas, então temos a seguinte desigualdade, [21]:

$$(1 - \alpha)^{-1} \cdot \log \sum_{i=1}^n p_i^{\alpha} \leq (1 - \alpha)^{-1} \cdot \log \left(\frac{\sum_{i=1}^n p_i^{\alpha}}{\sum_{i=1}^n p_i^{\alpha} q_i^{1-\alpha}} \right) \text{ para } \alpha > 0; \alpha \neq 1 \quad (2.42)$$

onde a igualdade aparece se e somente se $p_i = q_i$, $i = 1, \dots, n$.

Prova:

A prova segue imediatamente da desigualdade de Hölder

$$\left(\sum_{i=1}^n p_i a_i \right)^r \leq \sum_{i=1}^n p_i a_i^r \quad \text{para } r > 1,$$

$$\left(\sum_{i=1}^n p_i a_i \right)^r \geq \sum_{i=1}^n p_i a_i^r \quad \text{para } r < 1;$$

onde p_i, a_i são números não negativos e r um número positivo.

Fazendo $r = \alpha$

$$p_i = q_i$$

$$a_i = p_i/q_i.$$

Consideremos dois casos:

1) Se $\alpha > 1$,

$$\sum_{i=1}^n p_i^\alpha q_i^{1-\alpha} = \sum_{i=1}^n (p_i/q_i)^\alpha \cdot q_i \geq \left(\sum_{i=1}^n p_i/q_i \cdot q_i \right)^\alpha = \left(\sum_{i=1}^n p_i \right)^\alpha = 1$$

$$\sum_{i=1}^n p_i^\alpha \geq \frac{\sum_{i=1}^n p_i^\alpha}{\sum_{i=1}^n p_i^\alpha q_i^{1-\alpha}}$$

Tomando logaritmo em ambos os lados e com $(1 - \alpha)^{-1} < 0$, segue que:

$$(1 - \alpha)^{-1} \cdot \log \sum_{i=1}^n p_i^\alpha \leq (1 - \alpha)^{-1} \cdot \log \left(\frac{\sum_{i=1}^n p_i^\alpha}{\sum_{i=1}^n p_i^\alpha q_i^{1-\alpha}} \right).$$

2) Se $0 < \alpha < 1$,

$$\sum_{i=1}^n p_i^\alpha \leq \frac{\sum_{i=1}^n p_i^\alpha}{\sum_{i=1}^n p_i^\alpha q_i^{1-\alpha}}$$

Tomando logaritmo em ambos os lados e multiplicando por $(1 - \alpha)^{-1} > 0$ temos (2.42).

Portanto, demonstramos o teorema para $\alpha > 0$, $\alpha \neq 1$.

2.4.2 - Pseudo-generalização da desigualdade de Shannon.

Definição 8 - $H_\alpha^1(P, Q) = \frac{\alpha}{1-\alpha} \log \left(\sum_{i=1}^n p_i q_i^{\frac{\alpha-1}{\alpha}} \right)$ para

$$\alpha \neq 1, \alpha > 0, P \in \Delta_n \quad (2.43)$$

Semelhante a $H^{1\beta}(P, Q)$ notamos que $H_\alpha^1(P, Q)$ não é uma medida de imprecisão no sentido usual.

Usando esta definição e $H_\alpha(P)$ temos o seguinte teorema.

Teorema 11 - Se $P, Q \in \Delta_n$ segue que:

$$(1 - \alpha)^{-1} \cdot \log \sum_{i=1}^n p_i^\alpha \leq \frac{\alpha}{1-\alpha} \cdot \log \left(\sum_{i=1}^n p_i q_i^{\frac{\alpha-1}{\alpha}} \right), \alpha \neq 1, \alpha > 0 \quad (2.44)$$

Prova:

A prova segue direto de (2.17) e (2.18) substituindo β por α , tomando logaritmo em ambos os lados e multiplicando-os por $(1 - \alpha)^{-1}$.

2.4.3 - Aplicação no teorema de codificação.

$$\text{Definição 9} - L_\alpha = (1-\alpha)^{-1} \cdot \log \left[\frac{\sum_{i=1}^n p_i^\alpha \left(\sum_{i=1}^n 2^{-N_i} \right)^{1-\alpha}}{\sum_{i=1}^n p_i^\alpha 2^{(\alpha-1)N_i}} \right], \quad \text{para}$$

$$\alpha \neq 1, \alpha > 0. \quad (2.45)$$

O teorema seguinte mostra uma relação entre esta medida e $H_\alpha(P)$.

Teorema 12 - Se N_i , $i = 1, \dots, n$ são os comprimentos das palavras códigos satisfazendo (2.3), então [21]:

$$H_\alpha(P) \leq L_\alpha < H_\alpha(P) + \frac{\alpha}{\alpha-1} \quad (2.46)$$

Prova:

$$\text{Substituindo } q_i = \frac{2^{-N_i}}{\sum_{i=1}^n 2^{-N_i}}, \quad i = 1, 2, \dots, n \text{ em} \quad (2.44)$$

temos:

$$H_\alpha(P) \leq (1-\alpha)^{-1} \cdot \log \left(\frac{\sum_{i=1}^n p_i^\alpha \left(\sum_{i=1}^n 2^{-N_i} \right)^{1-\alpha}}{\sum_{i=1}^n p_i^\alpha 2^{(\alpha-1)N_i}} \right) = L_\alpha \quad (2.47)$$

A igualdade em (2.47) acontece se e somente se $p_i = \frac{2^{-N_i}}{\sum_{i=1}^n 2^{-N_i}}$,

$i = 1, \dots, n$.

Para encontrarmos a cota superior de L_α , escolhemos N_i satisfazendo (2.30) então (2.32) aparece.

Tomando logaritmo em ambos os lados e multiplicando-os por

$(1-\alpha)^{-1} < 0$ se $\alpha > 1$, encontramos a cota superior de L_α dada

em (2.46).

Observações:

- 1) Para $\alpha > 1$, a cota superior de L_α é maior do que a unidade.
- 2) A medida de comprimento independente do valor $(\sum_{i=1}^n 2^{-N_i})$ que é considerada uma generalização de L , pode ser definida por

$$L_\alpha^* = (1 - \alpha)^{-1} \cdot \log \left[\frac{\sum_{i=1}^n p_i^\alpha}{\sum_{i=1}^n p_i^\alpha 2^{(\alpha-1)N_i}} \right] \text{ para } \alpha > 0; \alpha \neq 1.$$

Analogamente ao fato que L_β^* tem a mesma cota superior e inferior como L_β em (2.27), podemos também mostrar que L_α^* tem a mesma cota inferior e superior como L_α em (2.46).

Definição 10 - $L_\alpha^1 = \frac{\alpha}{1-\alpha} \log \left[\sum_{i=1}^n p_i 2^{-N_i \left(\frac{\alpha-1}{\alpha}\right)} \right], \alpha > 0; \alpha \neq 1$ (2.48)

L_α^1 é o comprimento médio dado por Campbell (1965, 1966), [5,6]. O teorema seguinte também foi dado por Campbell.

Teorema 13 - Se $N_i, i = 1, \dots, n$ são os comprimentos das palavras códigos, então:

$$H_\alpha(P) \leq L_\alpha^1 < H_\alpha(P) + 1, \alpha > 0; \alpha \neq 1 \quad (2.49)$$

Prova:

Segue facilmente substituindo $q_i = \frac{2^{-N_i}}{\sum_{i=1}^n 2^{-N_i}}$,

$i = 1, \dots, n$ em (2.44) e usando a desigualdade dada em (2.3).

A igualdade em (2.49) aparece se e somente se

$$2^{-N_i} = \frac{p_i^\alpha}{\sum_{i=1}^n p_i^\alpha} \text{ e isto equivale a } N_i = -\log p_i^\alpha + \log \sum_{i=1}^n p_i^\alpha.$$

A cota superior de L_α^1 em (2.49) pode ser encontrada usando (2.37) com β no lugar de β e seguindo os mesmos passos feitos para a entropia do tipo β , tomando logaritmo em ambos os lados e multiplicando-os por $(1 - \alpha)^{-1}$.

Observações:

- 1) A cota superior de L_α^1 é maior do que a unidade para $D = 2$.
- 2) $\lim_{\alpha \rightarrow 1} L_\alpha^1 = L$.

Quando $\alpha \rightarrow 1$, temos que (2.49) reduz-se para $H(P) \leq L < H(P) + 1$ que é o resultado de Shannon.

Acabamos de apresentar aplicações das generalizações da desigualdade de Shannon para as entropias de ordem α e do tipo β . Usaremos este capítulo e algumas definições dadas no primeiro capítulo para generalizar a desigualdade de Shannon para outras entropias definidas anteriormente, com aplicações que serão dadas no próximo capítulo.

CAPÍTULO 3

ENTROPIAS DE ORDEM (α, β) , TIPO (α, β) , DE ORDEM α
E TIPO β E TEOREMAS DE CODIFICAÇÃO

Este capítulo consta de generalizações da desigualdade de Shannon para as entropias de ordem (α, β) , tipo (α, β) , de ordem α e tipo β .

Daremos também aplicações destas nos teoremas de codificação como fizemos para as entropias de Renyi e de tipo β .

3.1 - Entropia de ordem (α, β)

Como vimos no primeiro capítulo, esta medida foi introduzida por Aczél (1963). Por (1.7) podemos escrevê-la como

$$H_{\alpha, \beta}(P) = \frac{1-\alpha}{\beta-\alpha} \cdot H_{\alpha}(P) + \frac{\beta-1}{\beta-\alpha} H_{\beta}(P) \quad \text{onde } H_{\alpha}(P), H_{\beta}(P)$$

são as entropias de Renyi para $\alpha, \beta > 0$; $\alpha \neq \beta$.

A seguir faremos uma generalização da desigualdade de Shannon dada em (2.1) para esta entropia.

3.1.1 - Generalização da desigualdade de Shannon.

Podemos escrever a medida de imprecisão de ordem (α, β) definida em (1.15) da seguinte maneira:

$$H_{\alpha,\beta}(P,Q) = \frac{1-\alpha}{\beta-\alpha} \cdot H_{\alpha}(P,Q) + \frac{1-\beta}{\alpha-\beta} \cdot H_{\beta}(P,Q) \quad \text{para } \alpha, \beta > 0; \alpha \neq \beta \quad (3.1)$$

onde $H_{\alpha}(P,Q)$ e $H_{\beta}(P,Q)$ são as medidas de imprecisão de ordem α e β definidas em (1.13).

Então, temos o seguinte teorema.

Teorema 1 - Para duas distribuições de probabilidades $P, Q \in \Delta_n$, temos:

$$H_{\alpha,\beta}(P) \leq H_{\alpha,\beta}(P,Q) \quad , \quad 0 < \alpha < 1 \leq \beta \quad (0 < \beta < 1 \leq \alpha) \quad (3.2)$$

Prova:

Provaremos (3.2) somente para o caso $(0 < \alpha < 1 \leq \beta)$ porque o outro caso segue similarmente por simetria.

Se $\beta = 1$, então (3.2) reduz-se para (2.42) que é conhecido.

Consideremos o caso $0 < \alpha < 1 < \beta$.

Desde que (2.42) é válido para todo $\alpha > 0; \alpha \neq 1$, temos:

$$H_{\alpha}(P) \leq H_{\alpha}(P,Q) \quad \text{para } 0 < \alpha < 1 \quad (3.3)$$

e

$$H_{\beta}(P) \leq H_{\beta}(P,Q) \quad \text{para } \beta > 1 \quad (3.4)$$

Multiplicando (3.3) por $\frac{1-\alpha}{\beta-\alpha}$ (> 0) e (3.4) por $\frac{1-\beta}{\alpha-\beta}$ (> 0) e somando as equações resultantes, temos:

$$\frac{1-\alpha}{\beta-\alpha} H_{\alpha}(P) + \frac{1-\beta}{\alpha-\beta} H_{\beta}(P) \leq \frac{1-\alpha}{\beta-\alpha} H_{\alpha}(P,Q) + \frac{1-\beta}{\alpha-\beta} H_{\beta}(P,Q)$$

Logo, por (1.7) e (3.1), $H_{\alpha,\beta}(P) \leq H_{\alpha,\beta}(P,Q)$

3.1.2 - Pseudo-generalização da desigualdade de Shannon.

Definimos a medida $H_{\alpha, \beta}^1(P, Q)$ como:

$$H_{\alpha, \beta}^1(P, Q) = \frac{\alpha}{\beta - \alpha} \log \left(\sum_{i=1}^n p_i q_i^{\frac{\alpha-1}{\alpha}} \right) - \frac{\beta}{\beta - \alpha} \log \left(\sum_{i=1}^n p_i q_i^{\frac{\beta-1}{\beta}} \right) \quad (3.5)$$

para $\alpha, \beta > 0$; $\alpha \neq \beta$, $P, Q \in \Delta_n$.

Podemos escrevê-lo como:

$$H_{\alpha', \beta}^1(P, Q) = \frac{1-\alpha}{\beta-\alpha} H_{\alpha}^1(P, Q) + \frac{1-\beta}{\alpha-\beta} H_{\beta}^1(P, Q) \quad (3.6)$$

Notamos que $H_{\alpha, \beta}^1(P, P) \neq H_{\alpha, \beta}(P)$.

Contudo, $H_{\alpha, \beta}^1(P, Q)$ é uma generalização de $H_S(P, Q)$ dada em (1.12).

$$\lim_{\alpha \rightarrow 1} \lim_{\beta \rightarrow 1} H_{\alpha, \beta}^1(P, Q) = H_S(P, Q).$$

No seguinte teorema damos uma relação entre (1.7) e (3.6).

Teorema 2 - Se P, Q são duas distribuições de probabilidades discretas, então:

$$H_{\alpha, \beta}(P) \leq H_{\alpha, \beta}^1(P, Q) \quad ; \quad 0 < \alpha < 1 \leq \beta \quad (0 < \beta < 1 \leq \alpha) \quad (3.7)$$

onde a igualdade aparece se e somente se $\frac{p_i^\alpha}{\sum_{i=1}^n p_i^\alpha} = \frac{p_i^\beta}{\sum_{i=1}^n p_i^\beta}$.

Prova:

Vamos provar (3.7) somente para $0 < \alpha < 1 \leq \beta$ porque o outro caso segue por simetria.

Se $\beta = 1$, (3.7) reduz-se para (2.44).

Consideremos o caso $0 < \alpha < 1 < \beta$.

Desde que (2.44) é válido para $\alpha > 0$, $\alpha \neq 1$, temos:

$$H_{\alpha}(P) \leq H_{\alpha}^1(P, Q) \quad \text{para } 0 < \alpha < 1 \quad (3.8)$$

e

$$H_{\beta}(P) \leq H_{\beta}^1(P, Q) \quad \text{para } \beta > 1 \quad (3.9)$$

Multiplicando (3.8) por $\frac{1-\alpha}{\beta-\alpha}$ (> 0) e (3.9) por $\frac{1-\beta}{\alpha-\beta}$ (> 0) e somando as equações resultantes, temos:

$$\frac{1-\alpha}{\beta-\alpha} H_{\alpha}(P) + \frac{1-\beta}{\alpha-\beta} H_{\beta}(P) \leq \frac{1-\alpha}{\beta-\alpha} H_{\alpha}^1(P, Q) + \frac{1-\beta}{\alpha-\beta} H_{\beta}^1(P, Q) .$$

Logo, por (1.7) e (3.6) segue que $H_{\alpha, \beta}(P) \leq H_{\alpha, \beta}^1(P, Q)$

Desde que $H_{\alpha, \beta}^1(P, Q)$ não é uma medida de imprecisão no sentido usual, chamaremos (3.7) como uma pseudo-generalização da desigualdade de Shannon.

3.1.3 - Aplicação no teorema da codificação.

Discutiremos a seguir, a aplicação dos teoremas 1 e 2 no teorema de codificação.

Usaremos um conjunto finito de n -símbolos de entrada com distribuição de probabilidades $P = (p_1, \dots, p_n) \in \Delta_n$ que será codificado em termos de n símbolos de um alfabeto com tamanho 2.

Sejam N_1, N_2, \dots, N_n os comprimentos das palavras códigos, sabemos que se existe um código decifrável unicamente então

temos a desigualdade dada em (2.3).

Para o comprimento médio L foi provado que $L \geq H(P)$. Definiremos outras medidas de comprimento e mostraremos esta relação para a entropia de ordem (α, β) .

$$\text{Definição 1} - L_{(\alpha, \beta)} = \frac{1}{\beta - \alpha} \left[\frac{\sum_{i=1}^n p_i^\alpha \left(\sum_{i=1}^n 2^{-N_i} \right)^{1-\alpha} \cdot \sum_{i=1}^n p_i^\beta \left(\sum_{i=1}^n 2^{-N_i} \right)^{\beta-1}}{\sum_{i=1}^n p_i^\alpha \left(\sum_{i=1}^n 2^{-N_i} \right)^{\alpha-1} \cdot \sum_{i=1}^n p_i^\beta \left(\sum_{i=1}^n 2^{-N_i} \right)^{1-\beta}} \right] \quad (3.10)$$

para $\alpha, \beta > 0$; $\alpha \neq \beta$.

Verificamos que $\lim_{\alpha \rightarrow 1} \lim_{\beta \rightarrow 1} L_{(\alpha, \beta)} = L$.

A seguir mostraremos uma relação entre esta medida de comprimento e a entropia de ordem (α, β) .

Teorema 3 - Se N_1, \dots, N_n denotam os comprimentos das palavras de um código decifrável unicamente, então:

$$H_{\alpha, \beta}(P) \leq L_{(\alpha, \beta)} \quad \text{para} \quad (0 < \alpha < 1 \leq \beta); \quad 0 < \beta < 1 \leq \alpha. \quad (3.11)$$

Prova:

Provaremos somente para $0 < \alpha < 1 < \beta$, pois o outro caso segue similarmente.

Se $\beta = 1$, (3.11) reduz-se para (2.47).

Consideremos o caso $0 < \alpha < 1 < \beta$.

$$\text{Seja } Q = (q_1, \dots, q_n) \in \Delta_n \text{ em (3.2) com } q_i = \frac{2^{-N_i}}{\sum_{i=1}^n 2^{-N_i}} \quad (3.12)$$

segue que: $H_{\alpha, \beta}(P) \leq L_{(\alpha, \beta)}$

A igualdade em (3.11) ocorre se e somente se $p_i = \frac{2^{-N_i}}{\sum_{i=1}^n 2^{-N_i}}$

$$-\log p_i = N_i + \log \sum_{i=1}^n 2^{-N_i}.$$

Portanto, é sempre possível ter um código satisfazendo

$$-\log p_i \leq N_i < -\log p_i + 1$$

$$\text{que equivale a } \frac{1}{p_i} \leq 2^{N_i} < \frac{2}{p_i} \quad (3.13)$$

A relação dada em (3.11) mostra que $H_{\alpha, \beta}(P)$ é a maior cota inferior de $L_{(\alpha, \beta)}$.

$$\text{Definição 2} - L_{(\alpha, \beta)}^1 = \frac{\alpha}{\beta - \alpha} \log \left[\sum_{i=1}^n p_i 2^{\left(\frac{1-\alpha}{\alpha}\right) N_i} \right] - \frac{\beta}{\beta - \alpha} \log \left[\sum_{i=1}^n p_i 2^{\left(\frac{1-\beta}{\beta}\right) N_i} \right] \quad (3.14)$$

para $\alpha, \beta > 0$; $\alpha \neq \beta$.

O teorema seguinte mostra uma relação entre $L_{(\alpha, \beta)}^1$ e $H_{(\alpha, \beta)}(P)$.

Teorema 4 - Se N_i , $i = 1, \dots, n$ são os comprimentos das palavras códigos satisfazendo (2.3) temos:

$$H_{\alpha, \beta}(P) \leq L_{(\alpha, \beta)}^1 \quad \text{para } 0 < \alpha \leq 1 < \beta \quad (0 < \beta \leq 1 < \alpha) \quad (3.15)$$

Prova:

Provaremos para $0 < \alpha \leq 1 < \beta$, pois o outro caso segue similarmente.

Se $\alpha = 1$ (3.15) reduz-se para (2.49).

Consideremos o caso $0 < \alpha < 1 < \beta$.

Colocando $q_i = \frac{2^{-N_i}}{\sum_{i=1}^n 2^{-N_i}}$, $i = 1, \dots, n$ em (3.7) segue que:

$$H_{\alpha, \beta}(P) \leq \frac{\alpha}{\beta - \alpha} \log \left(\frac{\sum_{i=1}^n p_i (2^{-N_i})^{\frac{\alpha-1}{\alpha}}}{(\sum_{i=1}^n 2^{-N_i})^{\frac{\alpha-1}{\alpha}}} \right) - \frac{\beta}{\beta - \alpha} \log \left(\frac{\sum_{i=1}^n p_i (2^{-N_i})^{\frac{\beta-1}{\beta}}}{(\sum_{i=1}^n 2^{-N_i})^{\frac{\beta-1}{\beta}}} \right),$$

Por (2.3) segue que $(\sum_{i=1}^n 2^{-N_i})^{\frac{\alpha-1}{\alpha}} \leq 1$ e $(\sum_{i=1}^n 2^{-N_i})^{\frac{\beta-1}{\beta}} \leq 1$

Logo, $H_{\alpha, \beta}(P) \leq L_{(\alpha, \beta)}^1$.

A seguir daremos uma cota superior de $L_{(\alpha, \beta)}^1$.

Teorema 5 - Para um código decifrável unicamente, satisfazendo (2.3) segue:

$$L_{(\alpha, \beta)}^1 < H_{\alpha, \beta}(P) + 1 \quad \text{para } 0 < \alpha < 1 \leq \beta \quad (0 < \beta < 1 \leq \alpha) \quad (3.16)$$

Prova:

Podemos escrever $L_{(\alpha, \beta)}^1$ definida em (3.14) como:

$$L_{(\alpha, \beta)}^1 = \frac{1-\alpha}{\beta-\alpha} L_{\alpha}^1 + \frac{\beta-1}{\beta-\alpha} L_{\beta}^1 \quad (3.17)$$

mostraremos para $0 < \alpha < 1 \leq \beta$ pois o outro caso segue similarmente.

Se $\beta = 1$, (3.16) reduz-se a (2.49).

Sabemos que (2.49) vale para $\alpha > 0$, $\alpha \neq 1$, então:

$$L_{\alpha}^1 \leq H_{\alpha}(P) + 1 \quad \text{para } 0 < \alpha < 1 < \beta \quad (3.18)$$

e

$$L_{\beta}^1 \leq H_{\beta}(P) + 1 \quad \text{para } \beta > 1 \quad (3.19)$$

multiplicando (3.18) por $\frac{1-\alpha}{\beta-\alpha}$ (> 0) e (3.19) por $\frac{\beta-1}{\beta-\alpha}$ (> 0) e somando as equações resultantes segue que:

$$\begin{aligned} \frac{1-\alpha}{\beta-\alpha} \cdot L_{\alpha}^1 + \frac{\beta-1}{\beta-\alpha} L_{\beta}^1 &< \frac{1-\alpha}{\beta-\alpha} H_{\alpha}(P) + \frac{1-\alpha}{\beta-\alpha} + \\ + \frac{\beta-1}{\beta-\alpha} H_{\beta}(P) + \frac{\beta-1}{\beta-\alpha} &= \frac{1-\alpha}{\beta-\alpha} H_{\alpha}(P) + \frac{\beta-1}{\beta-\alpha} H_{\beta}(P) + 1 \end{aligned}$$

Logo, usando (1.7) e (3.17) demonstramos o teorema.

3.2 - Entropia do tipo (α, β)

Definida por Shannon e Taneja^[19] como:

$$H^{\alpha, \beta}(P) = \frac{\sum_{i=1}^n (p_i^{\alpha} - p_i^{\beta})}{2^{1-\alpha} - 2^{1-\beta}} \quad \text{para } \alpha, \beta > 0 ; \alpha \neq \beta \text{ como dada em (1.8)}$$

3.2.1 - Generalização da desigualdade de Shannon.

Do segundo capítulo sabemos que uma distribuição de generalidades $Q \in \Delta_n$ é dita distribuição de potência β derivável de P se existe um número real β tal que $q_i = p_i^{(\beta)}$.

Seja $F_r(P || Q) = \sum_{i=1}^n p_i^r q_i^{1-r}$ onde r é um parâmetro.

Teorema 6 - Para duas distribuições de probabilidades $P, Q \in \Delta_n$, temos:

$$H^{\alpha, \beta}(P) \leq (2^{1-\alpha} - 2^{1-\beta})^{-1} \left[\frac{W_{\alpha}(P)}{F_{\alpha}(P \parallel Q)} - \frac{W_{\beta}(P)}{F_{\beta}(P \parallel Q)} \right], \quad 0 < \alpha \leq 1 < \beta, \quad (3.20)$$

$$0 < \beta \leq 1 < \alpha.$$

Prova:

(i) Seja $0 < \alpha \leq 1 < \beta$.

$$\begin{aligned} \text{Usando a desigualdade de Hölder } \left(\sum_{i=1}^n p_i a_i \right)^r &\leq \sum_{i=1}^n p_i a_i^r, \quad r > 1, \\ &\geq \sum_{i=1}^n p_i a_i^r, \quad r < 1, \end{aligned}$$

onde a_i, p_i são números não-negativos.

Fazendo $a_i = \frac{p_i}{q_i}$ segue que:

$$\begin{aligned} F_{\alpha}(P \parallel Q) &= \sum_{i=1}^n p_i^{\alpha} q_i^{1-\alpha} = \sum_{i=1}^n (p_i/q_i)^{\alpha} \cdot q_i \leq \\ &\leq \left(\sum_{i=1}^n p_i/q_i \cdot q_i \right)^{\alpha} = \left(\sum_{i=1}^n p_i \right)^{\alpha} = 1 \end{aligned}$$

para $0 < \alpha \leq 1$,

$$F_{\beta}(P \parallel Q) \geq 1 \quad \text{para } \beta > 1 \quad (3.21)$$

$$\text{De (3.21) segue que } \frac{W_{\alpha}(P)}{F_{\alpha}(P \parallel Q)} - \frac{W_{\beta}(P)}{F_{\beta}(P \parallel Q)} \geq W_{\alpha}(P) - W_{\beta}(P) \quad (3.22)$$

Como $(2^{1-\alpha} - 2^{1-\beta})^{-1} > 0$, temos:

$$H^{\alpha, \beta}(P) \leq (2^{1-\alpha} - 2^{1-\beta})^{-1} \cdot \left[\frac{W_{\alpha}(P)}{F_{\alpha}(P \parallel Q)} - \frac{W_{\beta}(P)}{F_{\beta}(P \parallel Q)} \right]$$

(ii) Para $0 < \beta \leq 1 < \alpha$ a prova segue similarmente.

3.2.2 - Pseudo-generalização da desigualdade de Shannon.

De (1.9) podemos escrever a entropia do tipo (α, β) como

$$H^{\alpha, \beta}(P) = \frac{A_\alpha}{A_\alpha - A_\beta} H^\alpha(P) + \frac{A_\beta}{A_\beta - A_\alpha} H^\beta(P) \text{ onde } A_\alpha = 2^{1-\alpha} - 1 \text{ e } A_\beta = 2^{1-\beta} - 1$$

e $H^\alpha(P)$, $H^\beta(P)$ são entropias do tipo α e β .

Definimos a medida $H^{1(\alpha, \beta)}(P, Q)$ por:

$$H^{1(\alpha, \beta)}(P, Q) = \frac{1}{A_\alpha - A_\beta} \cdot \left[\left(\sum_{i=1}^n p_i q_i^{\frac{\alpha-1}{\alpha}} \right)^\alpha - \left(\sum_{i=1}^n p_i q_i^{\frac{\beta-1}{\beta}} \right)^\beta \right] \quad (3.23)$$

Verificamos que $\lim_{\beta \rightarrow 1} \lim_{\alpha \rightarrow 1} H^{1(\alpha, \beta)}(P, Q) = H_S(P, Q)$.

E podemos escrevê-la como:

$$H^{1(\alpha, \beta)}(P, Q) = \frac{A_\alpha}{A_\alpha - A_\beta} \cdot H^{1\alpha}(P, Q) + \frac{A_\beta}{A_\beta - A_\alpha} \cdot H^{1\beta}(P, Q) \quad (3.24)$$

onde $H^{1\alpha}(P, Q)$ e $H^{1\beta}(P, Q)$ são definidas em (2.14).

Relacionando a entropia do tipo (α, β) e (3.24) temos o seguinte teorema.

Teorema 7 - Se $P, Q \in \Delta_n$, então:

$$H^{\alpha, \beta}(P) \leq H^{1(\alpha, \beta)}(P, Q), \quad 0 < \alpha < 1 \leq \beta \quad (0 < \beta < 1 \leq \alpha) \quad (3.25)$$

Prova:

(i) $0 < \alpha < 1 \leq \beta$,

Se $\beta = 1$, (3.25) reduz-se para (2.15).

Consideremos o caso $0 < \alpha < 1 < \beta$.

Desde que (2.15) vale para $\alpha > 0$, $\alpha \neq 1$, temos:

$$H^\alpha(P) \leq H^{1\alpha}(P, Q) \quad \text{para } 0 < \alpha < 1 \quad (3.26)$$

e

$$H^\beta(P) \leq H^{1\beta}(P, Q) \quad \text{para } \beta > 1 \quad (3.27)$$

multiplicando (3.26) por $\frac{A_\alpha}{A_\alpha - A_\beta}$ (> 0) e (3.27) por $\frac{A_\beta}{A_\beta - A_\alpha}$ (> 0) e somando as equações resultantes, segue:

$$\frac{A_\alpha}{A_\alpha - A_\beta} H^\alpha(P) + \frac{A_\beta}{A_\beta - A_\alpha} H^\beta(P) \leq \frac{A_\alpha}{A_\alpha - A_\beta} H^{1\alpha}(P, Q) + \frac{A_\beta}{A_\beta - A_\alpha} H^{1\beta}(P, Q) \quad (3.28)$$

Logo, por (1.9) e (3.24) temos que $H^{\alpha, \beta}(P) \leq H^{1(\alpha, \beta)}(P, Q)$.

(ii) Para $0 < \beta < 1 \leq \alpha$, a prova segue similarmente..

3.2.3 - Aplicação no teorema de codificação.

Faremos uma discussão análoga a entropia de ordem (α, β) , isto é, mostraremos uma aplicação dos teoremas 6 e 7 no teorema de codificação usando as mesmas condições.

Para uma distribuição de probabilidade $P \in \Delta_n$ e um código satisfazendo (2.3) definiremos três medidas de comprimento para a entropia do tipo (α, β) .

Definição 3 - $L_{(\alpha, \beta)} = (2^{1-\alpha} - 2^{1-\beta})^{-1} \cdot \left[\frac{1}{\sum_{i=1}^n p_i^{(\alpha)} 2^{(\alpha-1)N_i}} - \frac{1}{\sum_{i=1}^n p_i^{(\beta)} 2^{(\beta-1)N_i}} \right]$

(3.29)

para $\alpha, \beta > 0$; $\alpha \neq \beta$,

onde $p_i^{(\alpha)} = \frac{p_i^\alpha}{W_\alpha(P)}$ e $p_i^{(\beta)} = \frac{p_i^\beta}{W_\beta(P)}$

Verificamos que $\lim_{\alpha \rightarrow 1} \lim_{\beta \rightarrow 1} L_{(\alpha, \beta)} = L$.

Teorema 8 - Se N_1, \dots, N_n denotam os comprimentos de um código de cifrável unicamente então:

$$L_{(\alpha, \beta)} \geq H^{\alpha, \beta}(P) \text{ para } 0 < \alpha \leq 1 < \beta \text{ (} 0 < \beta \leq 1 < \alpha \text{)} \quad (3.30)$$

Prova:

Seja $0 < \alpha \leq 1 < \beta$.

Tomando $Q = (q_1, \dots, q_n) \in \Delta_n$ em (3.20) onde $q_i = \frac{2^{-N_i}}{\sum_{i=1}^n 2^{-N_i}}$ (3.31)

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

Usando (3.31) e (3.20) temos o seguinte:

$$H^{\alpha, \beta}(P) \leq (2^{1-\alpha} - 2^{1-\beta})^{-1} \cdot \left[\frac{W_\alpha(P) \left(\sum_{i=1}^n 2^{-N_i} \right)^{1-\alpha}}{\sum_{i=1}^n p_i^\alpha 2^{(\alpha-1)N_i}} - \frac{W_\beta(P) \left(\sum_{i=1}^n 2^{-N_i} \right)^{1-\beta}}{\sum_{i=1}^n p_i^\beta 2^{(\beta-1)N_i}} \right] \quad (3.32)$$

De (2.3) $\left(\sum_{i=1}^n 2^{-N_i} \right)^{1-\alpha} \leq 1$, $0 < \alpha \leq 1$

$\left(\sum_{i=1}^n 2^{-N_i} \right)^{1-\beta} \geq 1$, $\beta > 1$

Usando estas desigualdades em (3.32) demonstramos o teorema.

Para $0 < \beta \leq 1 < \alpha$, a prova segue similarmente.

A igualdade em (3.30) é válida se e somente se

$$p_i = \frac{2^{-N_i}}{\sum_{i=1}^n 2^{-N_i}}, \quad i = 1, \dots, n.$$

$$\text{Ou seja, } \log \frac{1}{p_i} = N_i + \log \left(\sum_{i=1}^n 2^{-N_i} \right)$$

Portanto, é sempre possível ter um código satisfazendo a relação

$$\log \frac{1}{p_i} \leq N_i < \log \frac{1}{p_i} + 1 \quad (3.33)$$

$$\text{que é equivalente a } \frac{1}{p_i} \leq 2^{N_i} < \frac{2}{p_i} \quad (3.34)$$

No teorema seguinte daremos uma cota superior de $L_{(\alpha, \beta)}$.

Teorema 9 - Para um código decifrável unicamente, satisfazendo

(2.3) temos:

$$(2^{1-\alpha+\beta} - 2) \cdot L_{(\alpha, \beta)} < (2^\beta \sum_{i=1}^n p_i^\alpha - \sum_{i=1}^n p_i^\beta), \quad 0 < \alpha \leq 1 < \beta \quad (3.35)$$

e

$$(2^{1-\beta+\alpha} - 2) \cdot L_{(\alpha, \beta)} < (2^\alpha \sum_{i=1}^n p_i^\beta - \sum_{i=1}^n p_i^\alpha), \quad 0 < \beta \leq 1 < \alpha$$

Prova:

(i) Seja $0 < \alpha \leq 1 < \beta$,

De (3.34) temos $p_i 2^{N_i} < 2$ e $p_i 2^{N_i} \geq 1$,

$$\sum_{i=1}^n p_i^\alpha 2^{(\alpha-1)N_i} \geq 1 \quad \text{para } 0 < \alpha \leq 1,$$

$$\sum_{i=1}^n p_i^\beta 2^{(\beta-1)N_i} < 2^\beta \quad \text{para } \beta > 1,$$

$$\text{segue que: } \frac{\sum_{i=1}^n p_i^\alpha}{\sum_{i=1}^n p_i^\alpha 2^{(\alpha-1)N_i}} - \frac{\sum_{i=1}^n p_i^\beta}{\sum_{i=1}^n p_i^\beta 2^{(\beta-1)N_i}} \leq \sum_{i=1}^n p_i^\alpha - 2^{-\beta} \sum_{i=1}^n p_i^\beta \quad (3.36)$$

multiplicando ambos os lados de (3.36) por $(2^{1-\alpha} - 2^{1-\beta})^{-1} > 0$

encontramos:

$$L_{\alpha, \beta} < (2^{1-\alpha} - 2^{1-\beta})^{-1} \cdot \left(\sum_{i=1}^n p_i^\alpha - \frac{\sum_{i=1}^n p_i^\beta}{2^\beta} \right)$$

$$(2^{\beta+1-\alpha} - 2) L_{(\alpha, \beta)} < 2^\beta \cdot \sum_{i=1}^n p_i^\alpha - \sum_{i=1}^n p_i^\beta$$

isto para o resultado.

(ii) Para $0 < \beta \leq 1 < \alpha$, a prova segue similarmente.

$$\text{Definição 4} - L_{(\alpha, \beta)} = \frac{1}{A_\alpha - A_\beta} \cdot \left[\frac{\sum_{i=1}^n p_i^\alpha \left(\sum_{i=1}^n 2^{-N_i} \right)^{1-\alpha}}{\sum_{i=1}^n p_i^\alpha 2^{(\alpha-1)N_i}} - \frac{\sum_{i=1}^n p_i^\beta \left(\sum_{i=1}^n 2^{-N_i} \right)^{1-\beta}}{\sum_{i=1}^n p_i^\beta 2^{(\beta-1)N_i}} \right] \quad (3.37)$$

para $\alpha, \beta > 0$; $\alpha \neq \beta$.

Verificamos que $\lim_{\alpha \rightarrow 1} \lim_{\beta \rightarrow 1} L_{(\alpha, \beta)} = L$.

Teorema 10 - Se N_1, \dots, N_n denotam os comprimentos de um código decifrável unicamente satisfazendo (2.3), então:

$$H^{\alpha, \beta}(P) \leq L_{(\alpha, \beta)} \quad \text{para } \alpha, \beta > 0; \alpha \neq \beta \quad (3.38)$$

Prova:

(i) $0 < \alpha \leq 1 < \beta$,

Colocar
$$q_i = \frac{2^{-N_i}}{\sum_{i=1}^n 2^{-N_i}}, \quad i = 1, \dots, n \text{ em (3.20)} \quad (3.39)$$

Usando (3.39) e (3.20) temos:

$$H^{\alpha, \beta}(P) \leq (2^{1-\alpha} - 2^{1-\beta})^{-1} \left[\frac{\sum_{i=1}^n p_i^\alpha (\sum_{i=1}^n 2^{-N_i})^{1-\alpha}}{\sum_{i=1}^n p_i^\alpha 2^{-N_i(1-\alpha)}} - \frac{\sum_{i=1}^n p_i^\beta (\sum_{i=1}^n 2^{-N_i})^{1-\beta}}{\sum_{i=1}^n p_i^\beta 2^{-N_i(1-\beta)}} \right] = L^1_{(\alpha, \beta)}$$

(ii) Para $0 < \beta \leq 1 < \alpha$ a prova segue similarmente.

Definição 5 -
$$L^1_{(\alpha, \beta)} = (2^{1-\alpha} - 2^{1-\beta})^{-1} \cdot \left[\left(\sum_{i=1}^n p_i 2^{N_i \frac{1-\alpha}{\alpha}} \right)^\alpha - \left(\sum_{i=1}^n p_i 2^{N_i \frac{1-\beta}{\beta}} \right)^\beta \right]$$

para $\alpha, \beta > 0$; $\alpha \neq \beta$ (3.40)

Teorema 11 - Se N_i , $i = 1, \dots, n$ são os comprimentos das palavras de um código satisfazendo (2.3), então:

$$H^{\alpha, \beta}(P) \leq L^1_{(\alpha, \beta)}, \quad 0 < \alpha < 1 \leq \beta, \quad (0 < \beta < 1 \leq \alpha). \quad (3.41)$$

Prova:

(i) $0 < \alpha < 1 \leq \beta$,

Colocando
$$q_i = \frac{2^{-N_i}}{\sum_{i=1}^n 2^{-N_i}}, \quad i = 1, \dots, n \text{ em (3.25) temos:}$$

$$H^{\alpha, \beta}(P) \leq (2^{1-\alpha} - 2^{1-\beta})^{-1} \cdot \left[\left(\frac{\sum_{i=1}^n p_i (2^{-N_i})^{\frac{\alpha-1}{\alpha}}}{(\sum_{i=1}^n 2^{-N_i})^{\frac{\alpha-1}{\alpha}}} \right) - \left(\frac{\sum_{i=1}^n p_i (2^{-N_i})^{\frac{\beta-1}{\beta}}}{(\sum_{i=1}^n 2^{-N_i})^{\frac{\beta-1}{\beta}}} \right) \right]$$

Usando (2.3) conseguimos $H^{\alpha, \beta}(P) \leq L_{(\alpha, \beta)}^1$.

(ii) Para $0 < \beta < 1 \leq \alpha$ a prova segue similarmente.

3.3 - Entropia de ordem α e tipo β

Esta medida introduzida por Sharma e Mittal [18] é uma medida não-aditiva definida por

$$H_{\alpha}^{\beta}(P) = (2^{1-\beta} - 1)^{-1} \cdot \left[\sum_{i=1}^n (p_i^{\alpha})^{\frac{\beta-1}{\alpha-1}} - 1 \right] \text{ como dado em (1.11).}$$

3.3.1 - Generalização da desigualdade de Shannon.

Definimos $H_{\alpha}^{\beta}(P, Q)$ por:

$$H_{\alpha}^{\beta}(P, Q) = (2^{1-\beta} - 1)^{-1} \cdot \left[\frac{\sum_{i=1}^n p_i^{\alpha} q_i^{\frac{\beta-1}{\alpha-1}}}{\sum_{i=1}^n p_i^{\alpha} q_i^{1-\alpha}} - 1 \right] \text{ para } \alpha, \beta > 0 ; \alpha \neq \beta \quad (3.42)$$

Verificamos que se $p_i = q_i$ para todo $i = 1, \dots, n$, a medida acima reduz-se a entropia de Mittal dada em (1.11).

Definindo $W_r(P) = \sum_{i=1}^n p_i^r$ e $F_r(P || Q) = \sum_{i=1}^n p_i^r q_i^{1-r}$ podemos escrever (3.42) da seguinte forma:

$$H_{\alpha}^{\beta}(P, Q) = (2^{1-\beta} - 1)^{-1} \cdot \left[\frac{W_{\alpha}(P)^{\frac{\beta-1}{\alpha-1}}}{F_{\alpha}(P || Q)} - 1 \right] \text{ para } \alpha \neq \beta ; \alpha, \beta > 0 \quad (3.43)$$

A seguir daremos uma generalização da desigualdade de Shannon para esta entropia.

Teorema 12 - Para duas distribuições $P, Q \in \Delta_n$, temos:

$$H_{\alpha}^{\beta}(P) \leq (2^{1-\beta} - 1)^{-1} \cdot \left[\frac{W_{\alpha}(P)^{\frac{\beta-1}{\alpha-1}}}{F_{\alpha}(P \parallel Q)} - 1 \right], \quad \alpha, \beta > 0; \alpha \neq \beta \quad (3.44)$$

Prova:

$$(i) \quad 0 < \alpha < 1 \quad e \quad 0 < \beta < 1.$$

Usando a desigualdade de Hölder $(\sum_{i=1}^n p_i a_i)^r \leq \sum_{i=1}^n p_i a_i^r$, $r > 1$

$$\geq \sum_{i=1}^n p_i a_i^r, \quad r < 1$$

onde a_i, p_i são números não-negativos.

Fazendo $p_i = q_i$ e $a_i = p_i/q_i$ segue que:

$$F_{\alpha}(P \parallel Q) \leq 1 \quad (3.45)$$

a igualdade em (3.45) acontece se e somente se $p_i = q_i$;
 $i = 1, \dots, n$.

Então,
$$\frac{W_{\alpha}(P)^{\frac{\beta-1}{\alpha-1}}}{F_{\alpha}(P \parallel Q)} - 1 \geq W_{\alpha}(P)^{\frac{\beta-1}{\alpha-1}} - 1$$

Multiplicando ambos os lados por $(2^{1-\beta} - 1)^{-1} > 0$, temos demonstrando o teorema.

(ii) Para $\alpha, \beta > 1$ ou $0 < \alpha < 1 < \beta$, $0 < \beta < 1 < \alpha$, o teorema segue similarmente.

3.3.2 - Pseudo-generalização da desigualdade de Shannon

Para duas distribuições de probabilidades $P, Q \in \Delta_n$, definimos

$$H_{\alpha}^{1\beta}(P, Q) = (2^{1-\beta} - 1)^{-1} \cdot \left[\left(\sum_{i=1}^n p_i q_i^{\frac{\alpha-1}{\alpha}} \right)^{\alpha \cdot \frac{\beta-1}{\alpha-1}} - 1 \right],$$

$$\alpha, \beta > 0 \quad ; \quad \alpha \neq \beta \quad (3.47)$$

Verificamos que $\lim_{\alpha \rightarrow 1} \lim_{\beta \rightarrow 1} H_{\alpha}^{1\beta}(P, Q) = H_S(P, Q)$.

Teorema 13 - Se $P, Q \in \Delta_n$ então:

$$H_{\alpha}^{\beta}(P) \leq H_{\alpha}^{1\beta}(P, Q) \quad \text{para} \quad \alpha, \beta > 0 \quad ; \quad \alpha \neq \beta \quad (3.48)$$

Prova:

$$(i) \quad 0 < \alpha < 1 < \beta,$$

Usando a desigualdade de Hölder $\sum_{i=1}^n a_i b_i \geq \left(\sum_{i=1}^n a_i^r \right)^{1/r} \cdot \left(\sum_{i=1}^n b_i^s \right)^{1/s}$

com a_i, b_i números não negativos e $0 < r < 1, \quad 0 < s < 1$.

$$\text{Fazendo} \quad a_i = p_i^{\frac{\alpha}{\alpha-1}} \cdot q_i$$

$$b_i = p_i^{\frac{-\alpha}{\alpha-1}}$$

$$r = \frac{\alpha-1}{\alpha} \quad \text{e} \quad s = 1 - \alpha$$

$$\left(\sum_{i=1}^n p_i q_i^{\frac{\alpha-1}{\alpha}} \right)^{\frac{\alpha}{\alpha-1}} \cdot \left(\sum_{i=1}^n p_i^{\alpha} \right)^{\frac{1}{1-\alpha}} \leq 1$$

$$\text{Como} \quad \alpha > 1, \quad \left(\sum_{i=1}^n p_i^{\alpha} \right)^{\frac{1}{\alpha-1}} \geq \left(\sum_{i=1}^n p_i q_i^{\frac{\alpha-1}{\alpha}} \right)^{\frac{\alpha}{\alpha-1}}$$

$$\begin{aligned}
 \text{e como } \beta > 1, \quad & (2^{1-\beta} - 1)^{-1} \cdot \left[\left(\sum_{i=1}^n p_i^\alpha \right)^{\frac{\beta-1}{\alpha-1}} - 1 \right] \leq \\
 & \leq (2^{1-\beta} - 1)^{-1} \cdot \left[\left(\sum_{i=1}^n p_i q_i^{\frac{\alpha-1}{\alpha}} \right)^{\alpha \cdot \frac{\beta-1}{\alpha-1}} - 1 \right]
 \end{aligned}$$

Portanto demonstramos o teorema.

(ii) $0 < \beta < 1 < \alpha$, a prova segue similarmente.

3.3.3 - Aplicação no teorema de codificação

Mostraremos uma aplicação dos teoremas 12 e 13 no teorema de codificação como fizemos anteriormente para outras entropias.

Para um par de distribuições $P, Q \in \Delta_n$, definiremos medidas de comprimento para a entropia $H_\alpha^\beta(P)$.

$$\text{Definição 6 - } L_{(\alpha, \beta)} = (2^{1-\beta} - 1)^{-1} \cdot \left[\frac{\left(\sum_{i=1}^n p_i^\alpha \right)^{\frac{\beta-1}{\alpha-1}}}{\sum_{i=1}^n p_i^\alpha 2^{(\alpha-1)N_i}} - 1 \right] \quad (3.49)$$

para $\alpha, \beta > 0$; $\alpha \neq \beta$.

Verificamos que $\lim_{\alpha \rightarrow 1} \lim_{\beta \rightarrow 1} L_{(\alpha, \beta)} = L$.

Teorema 14 - Se N_1, \dots, N_n denotam os comprimentos de um código decifrável unicamente satisfazendo (2.3), então:

$$H_\alpha^\beta(P) \leq L_{(\alpha, \beta)} \quad , \quad \alpha, \beta > 0 \quad ; \quad \alpha \neq \beta \quad (3.50)$$

Prova:

Escolhendo $Q \in \Delta_n$ tal que $q_i = \frac{2^{-N_i}}{\sum_{i=1}^n 2^{-N_i}}$ em (3.44) segue que:

$$H_{\alpha}^{\beta}(P) \leq (2^{1-\beta} - 1)^{-1} \cdot \left[\frac{\left(\sum_{i=1}^n 2^{-N_i} \right)^{1-\alpha} \cdot \left(\sum_{i=1}^n p_i^{\alpha} \right)^{\frac{\beta-1}{\alpha-1}}}{\sum_{i=1}^n p_i^{\alpha} 2^{-N_i(1-\alpha)}} - 1 \right]$$

Como $\sum_{i=1}^n 2^{-N_i} \leq 1$, então $H_{\alpha}^{\beta}(P) \leq L(\alpha, \beta)$.

A igualdade em (3.50) ocorre se e somente se $p_i = \frac{2^{-N_i}}{\sum_{i=1}^n 2^{-N_i}}$, $i=1, \dots, n$

$$-\log p_i = N_i + \log \left(\sum_{i=1}^n 2^{-N_i} \right)$$

Portanto, é sempre possível ter um código satisfazendo

$$-\log p_i \leq N_i < -\log p_i + 1 \quad (3.51)$$

$$\Rightarrow \frac{1}{p_i} \leq 2^{N_i} < \frac{2}{p_i}$$

Teorema 15 - Para um código decifrável unicamente satisfazendo (2.3) temos:

$$L(\alpha, \beta) < \frac{1 - 2^{-\alpha} \cdot \left(\sum_{i=1}^n p_i^{\alpha} \right)^{\frac{\beta-1}{\alpha-1}}}{1 - 2^{1-\beta}} \quad \text{para } \beta > 1 \quad (3.52)$$

Prova:

De (3.51) segue que $p_i 2^{N_i} < 2$

$$p_i^{\alpha} 2^{(\alpha-1)N_i} < 2^{\alpha} \cdot 2^{-N_i}$$

Somando sobre todos i ($i = 1, \dots, n$) temos:

$$\sum_{i=1}^n p_i^\alpha 2^{(\alpha-1)N_i} < 2^\alpha \cdot \sum_{i=1}^n 2^{-N_i} < 2^\alpha$$

$$\text{Então, } 1 - \frac{\left(\sum_{i=1}^n p_i^\alpha \right)^{\frac{\beta-1}{\alpha-1}}}{\sum_{i=1}^n p_i^\alpha 2^{(\alpha-1)N_i}} < 1 - 2^{-\alpha} \cdot \left(\sum_{i=1}^n p_i^\alpha \right)^{\frac{\beta-1}{\alpha-1}}$$

Multiplicando ambos os lados por $(1 - 2^{1-\beta})^{-1}$, temos demonstrado o teorema.

Caso particular: se $\alpha = \beta > 1$, o resultado acima reduz-se ao resultado de Nath e Mittal.

$$\text{Definição 7 - } L_{(\alpha, \beta)}^1 = (2^{1-\beta} - 1)^{-1} \cdot \left[\left(\sum_{i=1}^n p_i D_i^{\frac{1-\alpha}{\alpha}} \right)^{\alpha \cdot \frac{\beta-1}{\alpha-1}} - 1 \right] \quad (3.53)$$

para $\alpha, \beta > 0$; $\alpha \neq \beta$.

Verificamos que $\lim_{\alpha \rightarrow 1} \lim_{\beta \rightarrow 1} L_{(\alpha, \beta)}^1 = L$.

Teorema 16 - Se N_i , $i = 1, \dots, n$ são os comprimentos das palavras de um código satisfazendo (2.3), então:

$$H_\alpha^\beta(P) \leq L_{(\alpha, \beta)}^1, \quad 0 < \alpha < 1 \leq \beta \quad (0 < \beta < 1 \leq \alpha) \quad (3.54)$$

Prova:

$$(i) \quad 0 < \alpha < 1 \leq \beta,$$

Se $\beta = 1$, (3.54) reduz-se a (2.49).

Consideremos o caso $0 < \alpha < 1 < \beta$.

Colocando $q_i = \frac{2^{-N_i}}{\sum_{i=1}^n 2^{-N_i}}$, $i = 1, \dots, n$ em (3.48) segue que:

$$H_{\alpha}^{\beta}(P) \leq (2^{1-\beta} - 1)^{-1} \cdot \left[\frac{\sum_{i=1}^n p_i 2^{-N_i \left(\frac{\alpha-1}{\alpha}\right)}}{\left(\sum_{i=1}^n 2^{-N_i}\right)^{\frac{\alpha-1}{\alpha}}} \right]^{\alpha \cdot \frac{\beta-1}{\alpha-1}} - 1$$

Como $\sum_{i=1}^n 2^{-N_i} \leq 1$, temos que $H_{\alpha}^{\beta}(P) \leq L_{(\alpha, \beta)}^1$.

(ii) $0 < \beta < 1 \leq \alpha$, a prova segue similarmente.

Teorema 17 - Para um código decifrável unicamente satisfazendo

(2.3), temos:

$$L_{(\alpha, \beta)}^1 < 2^{1-\beta} \cdot H_{\alpha}^{\beta}(P) + 1, \quad \alpha, \beta > 0; \quad \alpha \neq \beta \quad (3.55)$$

Prova:

A igualdade em (3.54) acontece se e somente se

$$2^{-N_i} = \frac{p_i^{\alpha}}{\sum_{i=1}^n p_i^{\alpha}}, \quad i = 1, \dots, n,$$

a qual é equivalente a $N_i = -\log p_i^{\alpha} + \log \left(\sum_{i=1}^n p_i^{\alpha} \right)$ (3.56)

Escolhendo N_i tal que $-\log \left(\frac{p_i^{\alpha}}{\sum_{i=1}^n p_i^{\alpha}} \right) \leq N_i < -\log \left(\frac{p_i^{\alpha}}{\sum_{i=1}^n p_i^{\alpha}} \right) + 1$ (3.57)

e usando esta relação temos $2^{-N_i} \geq \frac{p_i^{\alpha}}{\sum_{i=1}^n p_i^{\alpha} \cdot 2}$ (3.58)

Existem duas possibilidades:

1) Se $\alpha > 1$, (3.58) resulta em $\left[\sum_{i=1}^n p_i 2^{-N_i \left(\frac{1-\alpha}{\alpha}\right)} \right]^{\alpha} \leq \sum_{i=1}^n p_i^{\alpha} 2^{1-\alpha}$ (3.59)

Consideremos dois casos:

a) Se $0 < \beta < 1$,

Elevando ambos os lados de (3.59) a potência $\frac{\beta-1}{\alpha-1}$ temos:

$$\left[\sum_{i=1}^n p_i 2^{-N_i \left(\frac{1-\alpha}{\alpha}\right)} \right]^{\alpha \cdot \frac{\beta-1}{\alpha-1}} \leq \left[\sum_{i=1}^n p_i^\alpha 2^{1-\alpha} \right]^{\frac{\beta-1}{\alpha-1}} \quad (3.60)$$

Multiplicando ambos os lados de (3.60) por $(2^{1-\beta} - 1)^{-1} > 0$ demonstramos o teorema.

b) Se $\beta > 1$, a prova segue similarmente.

2) Se $0 < \alpha < 1$, a prova segue similarmente.

Observações:

- 1) A cota superior de $L_{(\alpha, \beta)}^1$ em (3.55) é maior do que a unidade.
- 2) Se $\beta = \alpha$, (3.55) reduz-se para o resultado provado para entropia do tipo β .

BIBLIOGRAFIA

- 1) ACZÉL, J. & DARÓCZY, Z. (1963) - "Über Verallgemeinerte Quasi Lineare Mittelwerte, die mit Gewichtsfunktionen Gebildet Sind, Publications Mathematicae, 10, 171-190.
- 2) ACZÉL, J. & DARÓCZY, Z. (1975) - On measures of information and their characterization, Academic Press, New York.
- 3) ARIMOTO, S. (1971) - Information theoretical considerations on estimation problems, Inform. Control, 19, 181-194.
- 4) ASH, R.B. (1965) - Information theory, Wiley, New York.
- 5) CAMPBELL, L.L. (1965) - A coding theorem and Renyi's entropy, In: Information control, n° 8, (1965) 423-429.
- 6) CAMPBELL, L.L. (1966) - Definition of Entropy by means of coding problem. In: Z. Wahr. Verw. Geb. n° 6 (1966) 113-118.
- 7) DARÓCZY, Z. (1970) - Generalized information functions, Information and Control, 16, Pags. 36-51.
- 8) FEINSTEIN, A. (1958) - Foundations of information theory, MacGraw-Hill, New York.
- 9) GALLAGER, R.G. (1968) - Information theory and reliable communication, Wiley, New York.
- 10) HARTLEY, R.V.C. (1928) - Transmission of information, BSTJ Pags. 535-563.
- 11) HAVRDA, J. & CHARVAT, F. (1967) - Quantification method of classification processes. The Concept of Structural α -entropy, Kybernetika (Prague) 3, 30-35.
- 12) KERRIDGE, D.F. (1961) - Inaccuracy and inference, J. Roy. Statist. Soc. Ser., n° 23, 184-194.
- 13) NATH, P. & MITTAL, D.P. (1973) - A generalization of Shannon's inequality and its application to coding theory, Information and Control, 23, 439-445.
- 14) NYQUIST, H. (1928) - Certain topics in telegraph. Transmission Theory, AIEE - Trans. on Commun. and Electronics, Vol. 47, 617-644.

- 15) RENYI, A. (1961) - On measures of entropy and information, Proc. 4th, Berkeley Symp. Math. Statist. Probability, n^o 1, 547-561.
- 16) SHANNON, C.E. (1948) - A mathematical theory of communication BSTJ, 27, 378-423, 623-656.
- 17) SHARMA, B.D. & GUPTA, H.C. (1976) - Sub-additive measures of relative information and inaccuracy, Metrika, 23, 155-165.
- 18) SHARMA, B.D. & MITTAL, D.P. (1975) - New non-additive measures of entropy for discrete probability distributions, Journal of Mathematical, Sc. 10, 28-40.
- 19) SHARMA, B.D. & TANEJA, I.J. (1977) - Their generalized additive measures of entropy, EIK. 13, 271-285.
- 20) TANEJA, I.J. (1977) - On axiomatic characterization of entropy of type (α, β) , Aplikace Matematiky, 22(6), 409-417.
- 21) VAN DER LÜBBE, J.C.A. (1979) - On certain coding theorems for the information of order α and of type β . Inf. Theory, Statistical Functions, Random Processes, Trans. 8'th Prague Conference, Vol. C, Prague 1978, 253-266.
- 22) VAN DER LÜBBE, J.C.A. & BOEKEE, D.E. (1978) - Coding theorems for two classes of information measures, Report Number IT-78-06.