MEDIDAS GENERALIZADAS DE INFORMAÇÃO E SUAS DECOMPOSIÇÕES ESTATÍSTICAS

ORIENTADOR: INDER JEET TANEJA

MARIA DA GRAÇA OLIVEIRA DUARTE

OUTUBRO DE 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.

PROF. ÍTALO JOSÉ DEJTER, Ph.D.
Coordenador

BANCA EXAMINADORA:

PROF. INDER JEET TANEJA, Ph.D.
Orientador

PROF. ÍTALO JOSÉ DEJTER, Ph.D.

Gurdial

PROF. GUR DIAL, Ph.D.

R E S U M O

Apresentamos neste trabalho alguns exemplos de aplicações da Informação Prevista e generalizações das Medidas de Informação Aperfeiçoada para N revisões. Mostramos também, casos particulares e propriedades das generalizações obtidas.

ABSTRACT

In this, work, we present some examples of applications of Information Prediction and give characterization of the generalization on the measure of Information Improvement due to N revisions.

Also, we present some particulars cases and properties of the generalized measure.

Í N D I C E

INTRODUÇÃO

Partindo de idéias básicas e medidas anteriormente usadas estendemos estas idéias e medidas para casos mais gerais proporcionando dessa maneira aplicações mais amplas em várias áreas de estudos.

No Capítulo 1, apresentamos idéias gerais da Teoria da Informação e as medidas usadas na realização de nosso trabalho.

Serão mostrados no Capítulo 2, alguns exemplos de aplicações da Informação Prevista.

No Capítulo 3, fazemos generalizações do trabalho apresentado por Sharma e Mittal [13], usando uma Medida de Informação Aperfeiçoada para N revisões, apresentada por Taneja e Arora [17]. Apresentamos ainda, casos particulares e propriedades das generalizações obtidas.

C A P Í T U L O 1

1. INTRODUÇÃO

Em 1948, Shannon [14] define pela primeira vez, quantitativamente através de um esboço matemático a natureza estatística de fonte, canal e saída de uma comunicação modelo.

A importância e várias aplicações deste novo campo de estudos deve-se a suas formulações abstratas que em muitas situações são consideradas convenientes.

Trabalhos em diferentes áreas demonstram interesse sobre o assunto desenvolvido e o adaptam conforme as suas necessidades. Como a maior parte de suas aplicações baseiam-se nas medidas da Teoria da Informação, estão sendo encontradas novas medidas e novas generalizações.

Nos últimos anos, a bibliografia em Teoria da Informação aumentou substancialmente e foram descobertas muitas aplicações em várias áreas de estudos como Psicologia, Economia, Computação, Ecologia, etc..

Na revisão e reformulação dos conceitos básicos são indispensáveis novas condições e novas generalizações. Esta é a finalidade do trabalho que apresentamos.

No decorrer deste nos restringimos a aspectos da Teoria da Informação e os associamos a nossa pesquisa.

1.1. Entropia de Shannon e seus Casos Bidimensionais

O estudo da Teoria da Informação está fundamentalmente baseado na Entropia de Shannon.

Seja X uma variável aleatória assumindo um número finito de n valores distintos: $x_1, x_2, \ldots, x_n$ com probabilidades $P = (p_1, p_2, \ldots, p_n)$, $p_i \geq 0$, $\sum_{i=1}^{n} p_i = 1$; então a entropia de de Shannon [14] e sua distribuição de probabilidade é dada por:

$$H(p_1, p_2, \ldots, p_n) = - \sum_{i=1}^{n} p_i \log p_i \tag{1.1}$$

onde a base do logaritmo é em geral arbitrária (>1). Contudo no caso discreto é comum usar a base 2. Igualmente consideramos que $0.\log o = 0$.

Denotamos por δ_n o conjunto de todas probabilidades de distribuição

$P = (p_1, p_2, \ldots, p_n)$, $p_i \geq 0$, $\sum_{i=1}^{n} p_i = 1$, isto é:

$$\delta_n = \{P = (p_1, p_2, \ldots, p_n) : p_i \geq 0, \quad \sum_{i=1}^{n} p_i = 1\}, \tag{1.2}$$

Maiores detalhes sobre estudos desta medida podem ser encontrados em livros de Teoria da Informação de Ash [3], Galloger [6], ou Feinstein [5], Guiasu [7].

Visto que os problemas de comunicação requerem análise da mensagem emitida por um canal e recebida por outro canal, a idéia da entropia (medida), precisa ser desenvolvida para o caso bidimensional.

Sejam $X = (x_1,x_2,...,x_n)$ e $Y = (y_1,y_2,...,y_n)$ duas variáveis aleatórias com probabilidade conjunta:

$$p(x_i, y_j) = P\{X = x_i \quad e \quad Y = y_j\} \tag{1.3}$$

A probabilidade condicional de $Y = y_j$ quando $X = x_i$ será denotada por $p(y_j/x_i)$ e as probabilidades individuais de $X = x_i$ e $Y = y_j$ são dadas respectivamente por $p(x_i)$ e $p(y_j)$ então:

$$p(x_i,y_j) = p(x_i)p(y_j/x_i) = p(y_j)p(x_i/y_j),$$

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

$$p(y_j) = \sum_{j=1}^{m} p(x_i,y_j),$$

onde $\sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i,y_j) = 1, \quad \sum_{j=1}^{m} p(y_j/x_i) = 1,$

$$\sum_{i=1}^{n} p(x_i/y_j) = 1, \text{ etc..}$$

Apresentamos cinco entropias que Shannon associou a um sistema bidimensional (X,Y)

1) Entropia conjunta $H(X,Y)$, dada por:

$$H(X,Y) = - \sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i,y_j) \log p(x_i,y_j), \tag{1.4}$$

2) Entropia marginal de X dada por:

$$H(X) = - \sum_{i=1}^{n} p(x_i) \log p(x_i) \tag{1.5}$$

3) Entropia marginal de Y dada por:

$$H(Y) = - \sum_{j=1}^{m} p(y_j) \log p(y_j), \tag{1.6}$$

4) Entropia condicional de Y dado $X=x_i$ é:

$$H(Y/X=x_i) = - \sum_{j=1}^{m} p(y_j/x_i) \log p(y_j/x_i), \tag{1.7}$$

e a entropia condicional de Y dado X é dada por:

$$H(Y/X) = - \sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i,y_j) \log p(y_j/x_i), \tag{1.8}$$

5) Entropia condicional de X dado Y é dada por:

$$H(X,Y) = - \sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i,y_j) \log p(x_i,y_j), \tag{1.9}$$

De algumas relações básicas e desigualdades entre estas medidas associadas ao sistema bidimensional citamos as seguintes:

$$H(X/Y) = H(X) + H(Y/X) = H(Y) + H(X/Y), \tag{1.10}$$

$$H(X,Y) \leqslant H(X) + H(Y), \qquad (1.11)$$

$$H(X) \geqslant H(X/Y) \qquad (1.12)$$

com igualdade em (1.11) e (1.12) se e somente se X e Y são independentes.

Shannon [14] provou a unicidade da medida (1.1) partindo de um conjunto de postulados. Esta medida foi subsequentemente caracterizada de várias maneiras conforme os vários autores. (Ver: Aczél [1], Mothai e Rothie [12]).

Através de alterações feitas aos axiomas e mais outros acrescentados tem sido obtidas novas generalizações dessa medida.

Considerando que a entropia satisfaz:

i) a aditividade

$$H(P*Q) = H(P) + H(Q) \qquad (1.13)$$

onde $P = (p_1, p_2, \ldots, p_n)$; $Q = (q_1, q_2, \ldots, q_m)$

$$\sum_{i=1}^{n} p_i \leqslant 1; \qquad \sum_{j=1}^{m} q_j \leqslant 1;$$

$$P*Q = (p_1 q_1, \ldots, p_1 q_m, p_2 q_1, \ldots, p_2 q_m, \ldots, p_n q_1, \ldots, p_n q_m).$$

ii) Propriedade do valor médio:

$$H(P \vee Q) = \phi^{-1}\left\{\frac{W(P)\,\phi[H(P)]+W(Q)\,\phi[H(Q)]}{W(P)+W(Q)}\right\} \tag{1.14}$$

onde $W(P) = \sum_{i=1}^{n} p_i$; $W(Q) = \sum_{j=1}^{m} q_j$;

$W(P) + W(Q) \leqslant 1$

ϕ é uma função contínua estritamente monótona. Considerando $P = (p_1,p_2,\ldots,p_n)$, $\sum_{i=1}^{n} p_i \leqslant 1$, Rényi apresentou uma generalização da Entropia de Shannon como:

$$H_1(P) = -\sum_{i=1}^{n} p_i \log p_i / \sum_{i=1}^{n} p_i \; ; \tag{1.15}$$

$$H_\alpha(P) = \frac{1}{1-\alpha} \log \left(\sum_{i=1}^{n} p_i^\alpha / \sum_{i=1}^{n} p_i\right), \; \alpha \neq 1, \; \alpha > 0, \tag{1.16}$$

Esta entropia (1.16) é chamada entropia de ordem α, e é uma generalização de (1.1).

Havrda - Charvát [8] e mais tarde Daróczy [4] obtiveram a entropia de grau β que também é uma generalização da Entropia de Shannon:

$$H^\beta(P) = \sum_{i=1}^{n} (p_i^\beta - 1)/(2^{1-\beta}-1) \; ; \tag{1.17}$$

$$\beta \neq 1, \; \beta > 0, \; \sum_{i=1}^{n} p_i = 1$$

Esta entropia de grau β foi obtida através das seguintes propriedades:

$$H^{\beta}(P*Q) = H^{\beta}(P) + H^{\beta}(Q) + (2^{1-\beta}-1)\, H^{\beta}(P)\, H^{\beta}(Q), \tag{1.18}$$

e $H^{\beta}(P) = \sum_{i=1}^{n} f(x_i)$, onde

$$f(p_i) = (p_i^{\beta} - p_i)(2^{1-\beta}-1)^{-1}, \tag{1.19}$$

A entropia de grau α e ordem β é dada por:

$$H_{\alpha}^{\beta}(P) = (2^{1-\beta}-1)^{-1}\left[\left(\sum_{i=1}^{n} p_i^{\alpha}\right)^{\frac{\beta-1}{\alpha-1}}\right], \tag{1.20}$$

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

Maiores detalhes sobre estas medidas foi apresentado por Taneja [18], e um estudo conjunto da entropia de Shannon e entropia de grau β também foi apresentada por Taneja [15], [18].

1.2. Informação Própria de Evento Único

Caracterizamos a informação própria com a ocorrência de um evento único com probabilidade p, $0 < p \leqslant 1$, sob a não aditividade. Estas noções são básicas no estudo que apresentamos.

Denotamos por $I(p)$, a informação própria de evento único com probabilidade p, sob os seguintes postulados:

Postulado 1:

$I(p)$ é uma função contínua de p em $(0,1]$

Postulado 2:

$$I(pq) = I(p) + I(q) + \lambda I(p)\, I(q),$$

$$\lambda \neq 0$$

Postulado 3:

$$I(\tfrac{1}{2}) = 1$$

Teorema 1: A informação própria de evento único com probabilidade p, $0 < p \leqslant 1$, sob os postulados acima é dada por:

$$I(p) = (p^{\beta-1}-1)\,(2^{1-\beta}-1)^{-1}, \quad \beta \neq 1, \tag{1.21}$$

Dada uma distribuição $P = (p_1, p_2, \ldots, p_n)$, $\sum_{i=1}^{n} p_i \leqslant 1$, é possível encontrar as informações próprias não aditivas de cada evento individual. A média aritmética das informações próprias dos eventos individuais é definida como a entropia não aditiva de uma distribuição, ou seja:

$$I(p_1,p_2,\ldots,p_n) = \sum_{i=1}^{n} p_i\ I(p_i) / \sum_{i=1}^{n} p_i, \tag{1.22}$$

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

$$= \sum_{i=1}^{n} (p_i^{\beta} - 1)/(2^{1-\beta} - 1) \sum_{i=1}^{n} p_i, \quad \beta \neq 1,$$

que é a entropia de grau β da distribuição P.

A entropia, em geral, pode ser considerada como uma média generalizada da informação própria com pesos como funções das probabilidades correspondentes, isto é:

$$I(p_1,p_2,\ldots,p_n) = \phi^{-1} \left\{ \frac{\sum_{i=1}^{n} f(p_i)\ \phi[I(p_i)]}{\sum_{i=1}^{n} f(p_i)} \right\}, \tag{1.23}$$

onde ϕ é uma função estritamente monótona. Temos que (1.23) deve satisfazer a condição:

$$I(P*Q) = I(P) + I(Q) + (2^{1-\beta} - 1)\ I(P)\ I(Q), \tag{1.24}$$

Tomando $f(x) = x$, temos o seguinte teorema (ref. Sharma e Mittal [13]).

Teorema 2: A entropia simétrica e contínua dada em (1.23) por $f(x) = x$, de uma distribuição de probabilidade $P = (p_1,p_2,\ldots,p_n)$, $\sum_{i=1}^{n} p_i = 1$, satisfazendo (1.24), pode ser expressa unicamente pelas seguintes formas:

$$I(p_1,p_2,\ldots,p_n;\ 1,\beta) = \left[\exp(\beta-1)\sum_{i=1}^{n} p_i \log p_i - 1\right](2^{1-\beta}-1)^{-1}$$

$$\beta \neq 1,\ \beta > 0, \qquad (1.25)$$

$$I(p_1,p_2,\ldots,p_n;\ \alpha,\beta)$$

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

$$\alpha\neq 1,\ \beta\neq 1, \alpha>0, \beta>0, \qquad (1.26)$$

Casos Particulares

1) $\lim_{\beta\to 1} I(p_1,p_2,\ldots,p_n;\ 1,\beta) = -\sum_{i=1}^{n} p_i \log p_i,$ (1.27)

é a entropia de Shannon.

2) $\lim_{\beta\to 1} I(p_1,p_2,\ldots,p_n;\ \alpha,\beta) = \frac{1}{1-\alpha}\log \sum_{i=1}^{n} p_i^{\alpha},\ \alpha\neq 1,\ \alpha>0;$ (1.28)

é a entropia de Rényi de ordem α.

3) $\lim_{\alpha\to 1}\left[\lim_{\beta\to 1} I(p_1,p_2,\ldots,p_n;\ \alpha,\beta)\right] = -\sum_{i=1}^{n} p_i \log p_i$ (1.29)

é a entropia de Shannon.

1.3. Informação de Kulback e Informação de Kerridge

A entropia de Shannon associa uma medida de informação com uma distribuição simples de probabilidade. Duas medidas semelhantes introduzidas por Kulback-Leibler [10] e Kerridge [9] associam uma medida com um par de distribuições de probabilidades de uma incerteza variável.

Estas medidas são mais gerais do que a entropia de Shannon.

i) Informação de Kulback

Seja X uma variável aleatória. Tomando um número finito de valores $x_1, x_2, \ldots, x_n$ com probabilidades $Q=(q_1, q_2, \ldots, q_n) \in \delta_n$, de um experimento E. Seja a frequência relativa definida como $P = (p_1, p_2, \ldots, p_n) \in \delta_n$, então a informação de Kulback de um experimento é dada por:

$$I(P;Q) = \sum_{i=1}^{n} p_i \log_2 \left(\frac{p_i}{q_i} \right), \tag{1.30}$$

Supondo que, se algum q_i for zero então o correspondente p_i também será zero. Fazendo $0 \log_2 (\frac{0}{0}) = 0 \log_2 0 - 0 \log_2 0 = 0$

No caso do experimento não relacionar as probabilidades com o resultado mas averiguar qual o experimento ocorrido, então a informação em uma observação é:

$$- \sum_{i=1}^{n} q_i \log_2 q_i \tag{1.31}$$

o que é a entropia de Shannon. Isto mostra que a medida de Kulback é uma generalização da entropia de Shannon.

Um estudo mais detalhado dessa medida com suas aplicações em estatística encontra-se em Kulback [11].

Algumas aplicações dessa medida em Economia foi apresentada por Theil [19].

ii) Informação de Kerridge

Suponhamos um dado experimento no qual as probabilidades dos n eventos distintos $x_1, x_2, \ldots, x_n$ são $Q=(q_1, q_2, \ldots$ $\ldots, q_n) \in \delta_n$, então suas reais probabilidades são $P = (p_1, p_2, \ldots$ $\ldots, p_n) \in \delta_n$, logo a transmissão pode ser menos precisa em dois casos:

1) a transmissão pode ser indefinida por falta de informação;

2) a transmissão pode ser incorreta porque a informação é incorreta.

Kerridge [9] introduziu uma medida que considera estes dois aspectos. Sua medida é dada por:

$$H(P;Q) = - \sum_{i=1}^{n} p_i \log_2 q_i \tag{1.32}$$

Considerando que, se qualquer q_i for zero, então o correspondente p_i também será zero, e é adotada a convenção $0 \log_2 0 = 0$. Nessa entropia temos:

$$H(P;Q) = - \sum_{i=1}^{n} p_i \log_2 p_i + \sum_{i=1}^{n} p_i \log_2 \left(\frac{p_i}{q_i} \right)$$

$$= H(P) + E(P;Q) \quad , \qquad (1.33)$$

Kerridge chamou $E(P;Q)$ de erro condicional.

Obviamente se $P=Q$, isto é, quando $p_i = q_i$ para $i=1,2,\ldots,n$ então $E(P;Q) = 0$ e a medida passa a ser a entropia de Shannon. Desta forma a medida de Kerridge abrange os dois aspectos e portanto é uma generalização da entropia de Shannon.

Kerridge [9] caracterizou (1.32) considerando os seguintes postulados:

K_1:$H(P;Q)$ é uma função de p_i e q_i onde $P=(p_1, p_2,\ldots,p_n) \varepsilon \delta_n$ e $Q = (q_1,q_2,\ldots,q_n) \varepsilon \delta_n$.

K_2: Quando n prováveis alternativas são idênticas a medida é uma função monótona decrescente de n.

K_3: Se uma transmissão é interrompida em um número de transmissões subsidiárias, a medida da transmissão original é a soma das medidas das transmissões subsidiárias, isto é:

$$H(p_1,p_2,p_3;q_1,q_2,q_3) = H(p_1,1-p_1;q_1,1-q_1)$$

$$+ (1-p_1) H \left(\frac{p_2}{1-p_1} , \frac{p_3}{1-p_1} ; \frac{q_2}{1-q_1} , \frac{q_3}{1-q_1} \right)$$

onde $p_1+p_2+p_3 = q_1+q_2+q_3 = 1$

$$K_4: H(p_1,p_2,p_3;q_1,q_2,q_3)=H(p_1,p_2+p_3;q_1q_2)$$

o que significa que se as alternativas das quais a soma das probabilidades são associadas, então a medida de uma transmissão é inalterada.

O postulado K_4 é o único que aplica medidas conceitualmente diferentes da entropia de Shannon.

1.4. Informação Aperfeiçoada Estudada por Theil

O conceito teórico de informação como foi abordado em várias medidas, por exemplo: Entropia de Shannon, Informação de Kulback e Informação de Kerridge tem encontrado vasto campo de aplicações nas mais variadas ciências.

Seja $P=(p_1,p_2,\ldots,p_n) \in \delta_n$ a distribuição de probabilidade posterior de um conjunto de n elementos num experimento cuja distribuição de probabilidade prevista anteriormente é $Q=(q_1,q_2,\ldots,q_n) \in \delta_n$, então sabemos que:

$$I(P;Q) = \sum_{i=1}^{n} p_i \log_2 \left(\frac{p_i}{q_i} \right) \tag{1.34}$$

é a informação de Kulback. Theil define a Informação Aperfeiçoada como:

$$I(P;Q;R) = \sum_{i=1}^{n} p_i \log_2 \left(\frac{r_i}{q_i} \right) \tag{1.35}$$

onde $R=(r_1,r_2,\ldots,r_n) \in \delta_n$ é a distribuição da probabilidade original revisada $Q=(q_1,q_2,\ldots,q_n) \in \delta_n$ na distribuição $P=(p_1,p_2,\ldots,p_n) \in \delta_n$ realizada após algum experimento E. No caso de $r_i=q_i$ para cada i, então $I(P;Q;R)$ é zero, mostrando que não é a Informação Aperfeiçoada. Isto ocorre sempre que as probabilidades previstas revisadas forem iguais.

$I(P;Q;R)$ dada em (1.35) inclue como casos particulares a Entropia de Shannon, a Informação de Kulback e a Informação de Kerridge.

Taneja e Arora [17] estudaram uma medida de Informação Aperfeiçoada devida a N revisões, dada por:

$$I(P;Q;R_1;R_2;\ldots;R_N) = -\sum_{i=1}^{n} p_i \log \left[\frac{\prod_{K=1}^{N} r_{Ki}}{q_i^N} \right]^{1/N} , \qquad (1.36)$$

onde $R_K = (r_{K1}, r_{K2}, \ldots, r_{Kn})$, $K = 1,2,\ldots,N$, são as N distribuições de probabilidades previstas e revisadas de um conjunto de n eventos $\xi_1,\xi_2,\ldots,\xi_n$, sendo que a distribuição de probabilidade prevista original é $Q=(q_1,q_2,\ldots,q_n)$ e as revisões são feitas tendo como base a distribuição de probabilidade $P=(p_1,p_2,\ldots,p_n)$.

A medida acima é uma generalização da Informação Aperfeiçoada de Theil [19].

$$I(P;Q;R) = -\sum_{i=1}^{n} p_i \log \left(\frac{r_i}{q_i} \right) \qquad (1.37)$$

A medida (1.36) reduz-se a (1.37) quando todos R_i são iguais, isto é: $R_1=R_2=....=R_N=R$. Esta medida foi apresentada pois se revisarmos o experimento duas, três,....N vezes, temos respectivamente N distribuições de probabilidades previstas e revisadas $R_1,R_2,...,R_N$ cuja distribuição de probabilidade original é Q e as revisões são feitas baseando-se na distribuição P. Definimos então a informação aperfeiçoada devido a N revisões, como:

$$I(P;Q;R_1;...;R_N) = \frac{I(P;Q;R_1)+I(P;Q;R_2)+...+I(P;Q;R_N)}{N}$$

que é idêntica a (1.36).

Taneja e Arora [17] caracterizaram a medida (1.36) e a generalizou segundo uma equação funcional, obtendo:

$$I(P;Q;R_1;...;R_N) = \frac{(2^{-\gamma}-1)^{-1}}{N} \left\{ \sum_{i=1}^{n} p_i \left[\frac{\prod_{k=1}^{N} r_{ki}}{q_i^N} \right]^{\gamma} -1 \right\}, \qquad (1.38)$$

A medida (1.38) reduz-se a (1.36) se for calculado o limite de (1.38) quando $\gamma \to 0$.

No capítulo 3 apresentamos medidas ainda mais gerais do que (1.38).

C A P Í T U L O   2

2. FORMAS GENERALIZADAS DA MENSAGEM E SUA INFORMAÇÃO PREVISTA

Seja o evento E com probabilidade p, qualquer que seja a natureza do evento. É estabelecido que E sempre ocorre; portanto p=1. Vamos supor que uma mensagem não nos dá a certeza de que E ocorre. A probabilidade p de E ocorrer é substituída, sendo que a nova probabilidade de E é q, um número entre zero e um. É uma generalização do caso que estabelece a certeza de que E ocorre. O nosso problema é medir a quantidade de informação oriunda da mensagem que troca as probabilidades de E, de p para q.

2.1. Probabilidades e Informação Anterior e Posterior

Chamaremos de probabilidade anterior à probabilidade original p de E, e chamaremos de probabilidade posterior ao valor que é estabelecido pela mensagem. O problema é medir a capacidade da informação da mensagem que transforma p em q, (anterior em posterior).

O ponto inicial é a probabilidade anterior p. O ponto final é a certeza que E ocorre. Entre estes dois pontos duas alternativas são consideradas:

1) Uma mensagem é recebida e transforma p em q. Esta mensagem é seguida de outra mensagem que transforma q em 1 (certeza).

2) p é transformada diretamente em 1 (certeza), sem passos intermediários.

Como a probabilidade inicial p e a probabilidade final (certeza), são as mesmas nas duas alternativas anteriores, a informação total também será igual.

A informação provida pela segunda alternativa é $h(p) = -\log p$. A informação total proporcionada pela primeira alternativa é igual a informação da mensagem que transforma p em q mais h(q). Ou seja, a informação da mensagem que transforma p (anterior) em q (posterior) é igual a:

$$h(p) - h(q) = \log \frac{q}{p} \tag{2.1}$$

A informação será zero quando $p=q$. O valor da informação é $-\log p$ para $q=1$, incluindo assim o caso especial da mensagem que estabelece a certeza de E ocorrer. Será um valor negativo quando $q < p$, isto é, quando supõem-se que E ocorre no final, mas a mensagem estabelece que E torna-se menos provável do que no início $(q < p)$.

2.2. Informação Prevista

A informação definida em (2.1) é usada quando supõem-se que E ocorre no final. Se E não ocorre no final podemos proceder de maneira análoga, substituindo p por 1-p e q por

1-q em (2.1).

Então temos:

$$h(1-p) - h(1-q) = \log \frac{1-q}{1-p} . \qquad (2.2)$$

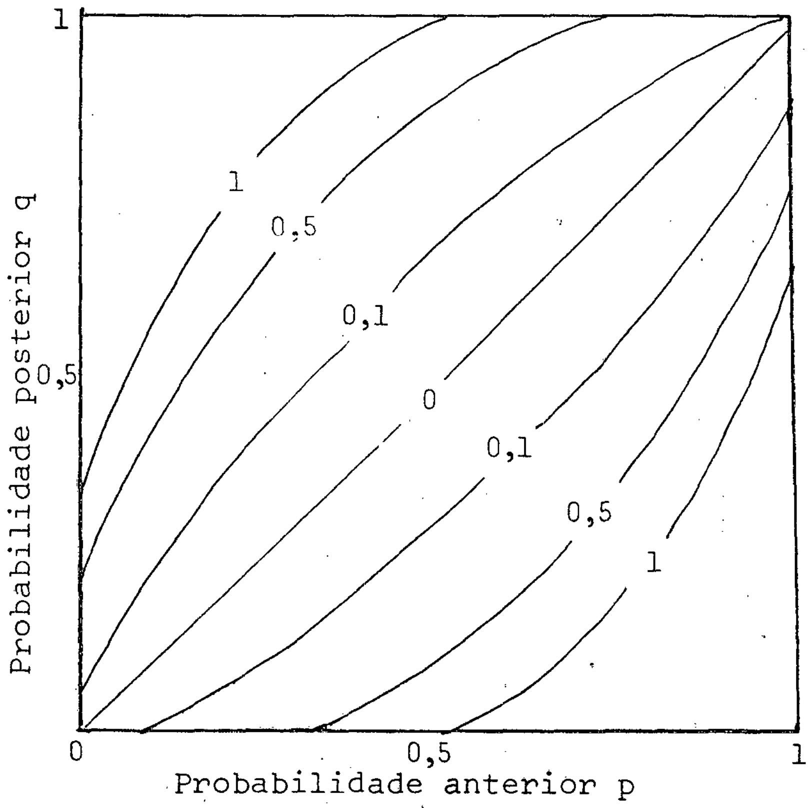
A informação recebida é (2.1) ou (2.2) e não as distinguimos até saber se E ocorre ou não. A informação recebida será (2.1) quando E ocorre e a probabilidade é q, e a informação é (2.2) quando E não ocorre e a probabilidade considerada é 1-q.

A informação prevista é obtida pela mensagem que transforma a probabilidade anterior (p,1-p) na probabilidade posterior (q,1-q) e é dada por:

$$I = q \log \frac{q}{p} + (1-q) \log \frac{1-q}{1-p} , \qquad (2.3)$$

I será igual a zero quando p=q. Será provado mais tarde que é possível sempre que $p \neq q$

Figura 2.1



A figura 2.1 apresenta o gráfico da probabilidade posterior (q,1-q) sendo dada a probabilidade anterior (p,1-p).

2.2.1. Extensão para Vários Eventos

A aproximação pode ser estendida para um sistema completo de n eventos com suas probabilidades anteriores e probabilidades posteriores expostas a seguir:

Eventos	E_1	$E_2 \ldots\ldots E_n$
Probabilidades anteriores:	p_1	$p_2 \ldots\ldots p_n$
Probabilidades posteriores:	q_1	$q_2 \ldots\ldots q_n$

Se transformarmos os E_i ocorridos, a informação é provida pela mensagem que transforma os p_i em q_i, e (2.1) é igual a log (q_i/p_i). Como a mensagem estabelece que a probabilidade de E_i é q_i, a informação prevista é:

$$I(Q:P) = \sum_{i=1}^{n} q_i \log \frac{q_i}{p_i} \tag{2.4}$$

A função (2.4) é nula quando $p_i=q_i$, para $i=1,2,\ldots,n$. Será positiva quando $p_i \neq q_i$ para cada i. Se um dos eventos for probabilidade posterior única, a função (2.4) torna-se $-\log p_i$ pela definição de informação. No caso de equiprobabilidades anteriores, $p_i = \frac{1}{n}$, para cada i, temos:

$$I(Q:P) = \sum_{i=1}^{n} q_i \log \frac{q_i}{1/n} = \log n - \sum_{i=1}^{n} q_i \log \frac{1}{q_i} , \tag{2.5}$$

Sabemos que log n é a entropia do caso equiprobabilidade e que é também o valor da entropia máxima. Logo o terceiro membro é a diferença entre as entropias anteriores e posteriores.

2.22. Prova da Não Negatividade da Informação Prevista

Para analisar o sinal da informação prevista (2.4) introduzimos:

$$a_i = \frac{p_i - q_i}{q_i} , \quad i=1,2,\ldots,n. \tag{2.6}$$

que é a diferença entre as correspondentes probabilidades anterior e posterior.

$$\sum_{i=1}^{n} q_i\, a_i = \sum_{i=1}^{n} (p_i - q_i) = 1 - 1 = 0 \qquad (2.7)$$

Substituindo os valores de a_i e p_i em (2.4) temos:

$$I(Q:P) = - \sum_{i=1}^{n} q_i \log \frac{p_i}{q_i}$$

$$I(Q:P) = - \sum_{i=1}^{n} q_i \log (1 + a_i) , \qquad (2.8)$$

Aplicando (2.7), obtemos:

$$I(Q:P) = \sum_{i=1}^{n} q_i \left[a_i - \log (1 + a_i) \right]$$

$$= \sum_{i=1}^{n} q_i\, f(a_i) ; \qquad (2.9)$$

onde $f(x) = x - \log (1 + x)$ (2.10)

Derivando $f(x)$, temos:

$$f'(x) = 1 - \frac{1}{1+x} = \frac{x}{1+x} > 0 \quad \text{se} \quad x > 0$$

$$< 0 \quad \text{se} \quad -1 \leqslant x \leqslant 0, \qquad (2.11)$$

Visto que $f(0)=0$, concluimos da derivada positiva que $f(x) > 0$ para $x > 0$; e da derivada negativa, $f(x) > 0$ para $x < 0$. Portanto f é estritamente positiva, exceto quando assume o valor zero.

Então de (2.9), $I(Q:P) > 0$, exceto quando todos os a_i são nulos.

Observe que a informação prevista (2.4) não tem limite superior finito. Isto pode ser verificado particularizando $q_i > p_i = 0$, para algum i. A probabilidade anterior particularizada no caso de E_1 ser zero, faz a informação aumentar porque a probabilidade foi aumentada segundo um fator infinito.

2.2.3. Decomposição da Informação Prevista

Sejam os eventos $E_1, E_2, \ldots, E_n$ e suas probabilidades $p_1, p_2, \ldots, p_n$. Esses eventos são combinados em G conjuntos de eventos $S_1, S_2, \ldots; S_G$ de tal modo que cada E_i tem valores em S_G, onde $g=1,2,\ldots,G$. Definimos probabilidade anterior e posterior por:

$$R_g = \sum_{i \varepsilon S_g} p_i \quad , \quad T_g = \sum_{i \varepsilon S_g} q_i \quad , \quad g = 1,2,\ldots,G \tag{2.13}$$

A informação prevista (2.4) após associações dos eventos é dada por:

$$I_0(Q:P) = \sum_{g=1}^{G} T_g \log \frac{T_g}{R_g} \tag{2.14}$$

Considere:

$$\sum_{i=1}^{n} q_i \log \frac{q_i}{p_i} = \sum_{g=1}^{G} T_g \sum_{i\varepsilon S_g} \frac{q_i}{T_g} \left(\log \frac{T_g}{R_g} + \log \frac{q_i/T_g}{p_i/R_g}\right)$$

$$= \sum_{g=1}^{G} T_g \log \frac{T_g}{R_g} + \sum_{g=1}^{G} T_g \sum_{i\varepsilon S_g} \frac{q_i}{T_g} \log \frac{q_i/T_g}{p_i/R_g}$$

ou

$$I(Q:P) = I_0(Q:P) + \sum_{g=1}^{G} T_g\, I_g(Q:P) \quad , \qquad (2.15)$$

onde:

$$I_g(Q:P) = \sum_{i\varepsilon S_g} \frac{q_i}{T_g} \log \frac{q_i/T_g}{p_i/R_g} \;, \; g=1,\ldots,G \qquad (2.16)$$

Como p_i/R_g e q_i/T_g, $i \varepsilon S_g$, são probabilidades condicionais de E_i dados S_g, $I_g(Q:P)$ é a informação prevista de suas mensagens sob a condição que um dos eventos de S_g ocorrerá no final. Temos $I_g(Q:P) = 0$ se e somente se a mensagem troca as probabilidades de todos os eventos de S_g nas mesmas proporções ($q_i/T_g = p_i/R_g$ ou $q_i/p_i = T_g/R_g$ para cada $i \varepsilon S_g$). Esta decomposição é feita em duas etapas. A primeira mensagem estabelece que as probabilidades anteriores $R_1,R_2,\ldots,R_G$ são modificadas para $T_1,T_2,\ldots,T_G$, assim a informação prevista é $I_0(Q:P)$. Então, sob a condição de que um dos eventos de S_G ocorrerá, considerar a mensagem subsequente que estabelece as modificações das probabilidades condicionais anterior p_i/R_g para q_i/T_g, $i \varepsilon S_g$. A informação prevista será $I_g(Q:P)$. Como a probabilidade de S_g é T_g; pela primeira mensagem, a informação total é igual ao lado direito de (2.15), e sua equação estabelece que o total é igual a informação prevista da mensagem que fixa o modo como as probabilidades de todos os n eventos são alterados.

2.3. Resultados Adicionais na Avaliação do Aproveitamento de Alunos

Apresentamos a seguir uma aplicação da entropia (2.3) na avaliação do aproveitamento dos alunos que cursaram as disciplinas oferecidas pelo Departamento de Matemática da Universidade Federal de Santa Catarina no período compreendido entre 1976 a 1980.

O Departamento de Matemática da U.F.S.C., oferece várias disciplinas aos alunos matriculados nos diversos cursos oferecidos por essa Universidade.

Foram escolhidas todas as disciplinas ininterruptamente durante os anos de 1976 a 1980. Obtivemos um total de 21 disciplinas e as agrupamos em 10 conjuntos de disciplinas afins, sendo:

M_1 - Elementos de Cálculo Diferencial e Integral, Cálculo Diferencial e Integral I, Cálculo Diferencial e Integral II, Cálculo Diferencial e Integral III e Cálculo Diferencial e Integral IV.

M_2 - Matemática Básica I e Matemática Básica II.

M_3 - Matemática Superior I, Matemática Superior II e Matemática Superior III.

M_4 - Matemática Comercial e Financeira.

M_5 - Fundamentos de Matemática I e Fundamentos de Matemática II.

M_6 - Álgebra Linear e Geometria Analítica I e Álgebra Linear e Geometria Analítica II.

M_7 - Introdução à Álgebra e Álgebra I.

M_8 - Análise Matemática I e Análise Matemática II.

M_9 - Funções de uma Variável Complexa.

M_{10} - Métodos de Matemática Aplicada.

Foram computadas as aprovações anuais dessas disciplinas no referido período e os dados foram usados nesse trabalho.

2.3.1. Entropia da Avaliação dos Alunos do Departamento de Matemática

Sejam:

p_j - a percentagem de alunos aprovados na j-ésima disciplina oferecida pelo Departamento de Matemática;

w_j - todos os alunos de uma disciplina;

$W_r = \sum_j w_j (j \in R_r)$ - todos os alunos de um conjunto de disciplina;

D_r - todas as disciplinas oferecidas pelo Departamento, (as que foram consideradas).

A proporção de alunos matriculados que conseguiram aprovação em cada conjunto de disciplinas é:

$$P_r = \sum_{j \varepsilon D_r} \frac{w_j}{W_r} \; p_j \qquad r = 1,2,\ldots,10 \tag{2.17}$$

A proporção de alunos reprovados de D_r é $1-P_r$. Portanto a entropia da avaliação dos alunos por grupo de disciplina é:

$$K_r = P_r \log \frac{1}{P_r} + (1-P_r) \log \frac{1}{1-P_r} \tag{2.18}$$

onde K representa a entropia da avaliação dos alunos de cada grupo de disciplina.

A tabela 2.1 apresenta a entropia da avaliação dos alunos dos 10 conjuntos de disciplinas consideradas nos anos de 1976 a 1980. São expressos em percentagem os valores da tabela.

A entropia média de avaliação dos alunos nos conjuntos de disciplinas é:

$$\overline{K} = \sum_{r=1}^{10} W_r \, K_r \; ; \tag{2.19}$$

Essa entropia média de avaliação ocorre de 1977 a 1979, tendo um decréscimo em 1980 em relação ao ano anterior.

TABELA 2.1

DISCI- PLINAS	ENTROPIA DE AVALIAÇÃO				
	1976	1977	1978	1979	1980
M_1	0,779	0,815	0,816	0,827	0,81
M_2	0,829	0,835	0,777	0,846	0,821
M_3	0,829	0,831	0,755	0,816	0,708
M_4	0,611	0,826	0,461	0,551	0,443
M_5	0,839	0,835	0,846	0,443	0,799
M_6	0,529	0,732	0,722	0,801	0,827
M_7	0,805	0,737	0,799	0,828	0,776
M_8	0,763	0,846	0,797	0,799	0,625
M_9	0,752	0,642	0,815	0,845	0,752
M_{10}	0,298	0,22	0,569	0,546	0,697
MÉDIA	3390,33	4010,15	4525,9	6982,8	5545,4

Comparando os dados da tabela 2.1 com a percentagem de alunos aprovados e reprovados observamos que K_r será máximo nos grupos onde aprovações e reprovações se equipararem em número K_r será menor nos grupos onde prevalece o número de aprovados ou de reprovados.

2.3.2. Uma Medida de Aproveitamento

Pode ser mostrado que a entropia de avaliação do grupo de disciplinas K_r nunca pode ser menor do que a entropia média $\bar{\bar{H}}_r$ das disciplinas de D_r.

Considere a diferença K_r-$\bar{H}_r$ usando a definição de $\overline{H}_r$:

$$\overline{H}_r = \sum_{j\varepsilon D_r} \frac{w_j}{W_r} H_j = \sum_{j\varepsilon D_r} \frac{w_j}{W_r} \left[p_j \log \frac{1}{p_j} + (1-p_j) \log \frac{1}{1-p_j}\right]$$

$$K_r - \overline{H}_r = P_r \log \frac{1}{P_r} + (1-P_r) \log \frac{1}{1-P_r}$$

$$- \sum_{j\varepsilon D_r} \frac{w_j}{W_r} \left[p_j \log \frac{1}{p_j} + (1-p_j) \log \frac{1}{1-p_j}\right]$$

usando (2.17) temos:

$$K_r - \overline{H}_r = \sum_{j\varepsilon D_r} \frac{w_j}{W_r} p_j \left(\log \frac{1}{P_r} - \log \frac{1}{p_j}\right)$$

$$+ \sum_{j\varepsilon D_r} \frac{w_j}{W_r} (1-p_j) \left(\log \frac{1}{1-P_r} - \log \frac{1}{1-p_j}\right)$$

Simplicando temos:

$$K_r - \overline{H}_r = \sum_{j\varepsilon D_r} \frac{w_j}{W_r} \left[p_j \log \frac{p_j}{P_r} + (1-p_j) \log \frac{1-p_j}{1-P_r}\right] ; \qquad (2.18)$$

Considerando em particular a expressão entre os colchetes.

$$I_j = p_j \log \frac{p_j}{P_r} + (1-p_j) \log \frac{1-p_j}{1-P_r} , \qquad j \ \varepsilon \ D_r ; \qquad (2.19)$$

Comparando com (2.3), temos que I_j é a informação prevista da mensagem que transforma a probabilidade anterior (P_r, $1-P_r$), na probabilidade posterior (p_j, $I-p_j$). Temos que $I_j = 0$ se e somente se as duas composições são idênticas, e $I_j>0$, se são diferentes. Se todos os alunos de uma disciplina são aprovados ($p_j=1$), então $I_j = -\log P_r$. Se todos os alunos são reprovados temos que $I_j = -\log(1-P_r)$. Assim, quando num conjunto de disciplina predomina a aprovação, a presença de uma disciplina onde todos os alunos são aprovados resulta um baixo valor $I_j=-\log P_r$, mas se em uma disciplina todos os alunos são reprovados aumenta o valor $I_j = -\log(1-P_r)$.

Em (2.18) os valores são todos positivos, assim $K_r-\bar{H}_r$ é positivo exceto nas disciplinas de D_r que tem a mesma composição, ou seja, $K_r-\bar{H}_r = 0$. Portanto a entropia média $\bar{H}_r$ das disciplinas de D_r é no máximo igual a entropia de avaliação K_r do conjunto de disciplinas. Temos que $\bar{H}_r < K_r$ considerando as diferenças entre as composições de avaliação dessas disciplinas. Quando H_r é menor que o máximo permitido segundo a composição de D_r, $(K_r-\bar{H}_r)$ é uma medida de avaliação.

A entropia $K_r-\bar{H}_r$ é mostrada na tabela 2.2. Os valores são geralmente maiores para os que tem os maiores K_r. A entropia média de avaliação dos alunos do Departamento de Matemática é:

$$\sum_{r=1}^{10} W_r\,(K_r - \bar{H}_r) = \bar{K} - \bar{H} \quad . \tag{2.20}$$

TABELA 2.2

DISCI-PLINAS	100 $(K_r - \bar{H}_r)$				
	1976	1977	1978	1979	1980
M_1	4,8	3,7	3,9	1,8	4,3
M_2	4,8	0,1	0,2	12,7	0,2
M_3	10,4	2,7	1,7	1,1	3,1
M_4	0,0	0,0	0,0	0,0	0,0
M_5	1,6	3,1	6,2	0,5	3,3
M_6	2,6	3,0	5,8	0,3	0,3
M_7	5,8	0,1	0,1	0,1	0,1
M_8	1,9	4,7	1,7	0,1	0,1
M_9	0,0	0,0	0,0	0,0	0,0
M_{10}	0,0	0,0	0,0	0,0	0,0
MÉDIA	211,61	131,72	190,23	963,16	179,76

Na tabela 2.2 verificamos que os menores valores constatam maior homogeneidade quanto ao nível do curso e dos alunos no grupo de disciplinas. A Entropia Média de Avaliação dos alunos do Departamento está apresentada na última linha. Ela decresce em 1977, crescendo em 1978 e 1979. Ela é máxima em 1979 e decresce novamente em 1980.

2.3.3. O Aproveitamento como Função da Entropia de Avaliação dos Grupos de Disciplinas

A tabela 2.3 apresenta K_r-$\bar{H}_r$ como uma percentagem do correspondente K_r de cada conjunto de disciplinas. Os dados variam de um conjunto a outro, mas apresentamos nas duas últimas linhas uma média dos cinco grupos com maior entropia de avaliação K_r e outra média dos cinco grupos com menor entropia de avaliação K_r. Os dados são obtidos das razões médias $\frac{K_r - \overline{H}_r}{K_r}$ em percentagem.

A tabela mostra que em média a redução da entropia média de avaliação das disciplinas atinge 2,1% nos grupos com menor K_r e atinge 6,6% nos grupos com maior entropia de avaliação.

TABELA 2.3

$$\frac{100\ (K_r - H_r)}{K_r}$$

DISCIPLINAS	1976	1977	1978	1979	1980
M_1	6,0	4,53	4,77	2,17	5,30
M_2	5,79	0,12	0,25	15,01	0,26
M_3	12,54	3,24	2,25	1,34	4,37
M_4	0,0	0,0	0,0	0,0	0,0
M_5	1,9	3,71	7,32	1,12	4,13
M_6	4,9	4,09	8,03	0,37	0,36
M_7	7,2	0,13	0,125	0,12	0,128
M_8	2,49	5,55	2,13	0,12	0,16
M_9	0,0	0,0	0,0	0,0	0,0
M_{10}	0,0	0,0	0,0	0,0	0,0
Grupos com:					
Menor K_r	1,478	1,75	2,108	0,325	0,9
Maior K_r	6,686	2,52	2,871	3,73	2,03

2.4. O Índice não Similar

Apresentamos o índice não similar pelo seu frequente uso em sociologia. No exemplo apresentado anteriormente a aplicação desse índice resultaria numa comparação entre o número de alunos aprovados em cada disciplina e o número de alunos aprovados em cada grupo de disciplina.

Na nossa notação temos que:

$w_j\ p_j$ - é o número de alunos aprovados na i-ésima disciplina;

$W_r\ P_r$ - é o número de alunos aprovados no r-ésimo grupo;

$\dfrac{w_j\ p_j}{W_r\ P_r}$ - é a relação entre os alunos aprovados na j-ésima disciplina e os alunos aprovados no r-ésimo conjunto de disciplinas. Analogamente, para os alunos reprovados a relação é $\dfrac{w_j\ (1-p_j)}{W_r\ (1-P_r)}$, assim o índice não similar de D_r é dado por:

$$d_r = \frac{1}{2} \sum_{j \varepsilon D_r} \left| \frac{w_j\ p_j}{W_r\ P_r} - \frac{w_j\ (1-p_j)}{W_r\ (1-P_r)} \right| \tag{2.21}$$

$$= \frac{1}{2} \sum_{j \varepsilon D_r} \frac{w_j}{W_r} \left| \frac{p_j}{W_r} - \frac{1 - p_j}{1 - P_r} \right|$$

Sendo:

$$\left|\frac{p_j}{P_r} - \frac{1-p_j}{1-P_r}\right| = \frac{|p_j(1-P_r) - (1-p_j)P_r|}{P_r(1-P_r)} = \frac{|p_j-P_r|}{P_r(1-P_r)} ,$$

podemos simplificar (2.21), obtendo

$$d_r = \frac{\sum_{j\varepsilon D_r} \frac{w_j}{W_r} |p_j-P_r|}{2\ P_r(1-P_r)} \tag{2.22}$$

Se todas as disciplinas de D_r tem a mesma composição $d_r = 0$ (o valor mínimo).

Se numa disciplina todos os alunos são reprovados ou todos são aprovados teremos $d_r = 1$ (o valor máximo).

Chamaremos de D_{r1} para o conjunto de disciplinas onde todos os alunos são aprovados e D_{r2} para o conjunto de disciplinas onde todos os alunos são reprovados, assim $p_j=1$ para $j\varepsilon D_{r1}$ e $p_j=0$ para $j\varepsilon D_{r2}$. Usando (2.22) temos:

$$d_r = \frac{(1-P_r) \sum_{j\varepsilon D_{r1}} \frac{w_j}{W_r} + P_r \sum_{j\varepsilon D_{r2}} \frac{w_j}{W_r}}{2\ P_r(1 - P_r)} \tag{2.23}$$

Mas $\sum_j (\frac{w_j}{W_r})$ é igual a P_r para $j\varepsilon D_{r1}$ e $(1 - P_r)$ para $j\varepsilon D_{r2}$. Logo:

$$d_r = \frac{(1-P_r)P_r + P_r(1-P_r)}{2\ P_r(1-P_r)}$$

Portanto: $d_r = 1$

2.4.1. Aplicação Prática do Índice não Similar

Seja P a proporção de alunos aprovados em todas as disciplinas do Departamento de Matemática consideradas anteriormente.

$$P = \sum_{j=1}^{10} w_j\ p_j \qquad (2.24)$$

Considerando as aprovações e reprovações médias nos 10 conjuntos de disciplinas como foi definido em (2.20) e (2.18) temos:

$$\overline{K}-\overline{H} = \sum_{r=1}^{10} \sum_{j \varepsilon D_r} w_j \left[p_j \log \frac{p_j}{P_r} + (1-p_j) \log \frac{1-p_j}{1-P_r} \right], \qquad (2.25)$$

A expressão entre os colchetes pode ser da forma seguinte:

$$p_j \log \frac{p_j}{P} + (1-p_j) \log \frac{1-p_j}{1-P} - \left[p_j \log \frac{P_r}{P} + (1-p_j) \cdot \log \frac{1-P_r}{1-P} \right],$$

então (2.25) será escrita assim:

$$\overline{K}-\overline{H} = \sum_{j=1}^{N} w_j \left[p_j \log \frac{p_j}{P} + (1-p_j) \log \frac{1-p_j}{1-P} \right]$$

$$- \sum_{r=1}^{10} W_r \left[P_r \log \frac{P_r}{P} + (1-P_r) \log \frac{1-P_r}{1-P} \right] \qquad (2.26)$$

Isto mostra que o aproveitamento médio nos conjuntos de disciplinas é igual à diferença de dois outros valores obtidos, sendo um deles a composição da avaliação de todas as N disciplinas do Departamento, e o outro valor é a composição da avaliação dos 10 conjuntos de disciplinas (em relação às disciplinas do Departamento).

Agora aplicando o índice não similar

$$\frac{\sum_{j=1}^{N} w_j \left| p_j - P \right|}{2\,P(1-P)} \quad \text{e} \quad \frac{\sum_{r=1}^{10} W_r \left| P_r - P \right|}{2\,P(1-P)} \qquad (2.27)$$

em (2.26), obtemos:

$$\sum_{r=1}^{10} W_r\, d_r = \sum_{r=1}^{10} \frac{\sum_{j \varepsilon D_r} w_j \left| p_j - P_r \right|}{2\,P_r(1-P_r)} \,, \qquad (2.28)$$

Comparando (2.27) e (2.28), concluimos que não existe uma relação que expresse uma delas em função das outras duas expressões, o que é uma desvantagem no uso do índice não similar. A diferença absoluta é o maior obstáculo para decomposições simples como (2.26).

O índice não similar é limitado a duas alternativas, (alunos aprovados e alunos reprovados conforme o exemplo dado). Considerando que a entropia (2.18) pode ser estendida a um grande número de alternativas, o uso do que a aplicação do índice não similar.

2.5. A Precisão na Decomposição Prevista

Uma aplicação da informação (2.4) é no caso de uma decomposição ser prevista uma ou mais vezes com a finalidade de obter maior precisão. A decomposição prevista é representada por um conjunto de probabilidades anteriores e a decomposição observada é representada por um conjunto de probabilidades posteriores.

2.5.1. Descrição de um Exemplo de Consumo sendo Observada a Intenção

Uma pesquisa foi feita com o objetivo de saber da pretenção das pessoas em obter um determinado objeto num certo período de tempo. As respostas podem ser: afirmativas, negativas ou incertas.

Após um ano da primeira pesquisa faz-se uma nova pesquisa para verificar se as mesmas pessoas realizaram a sua intenção expressa na pesquisa anterior.

As pesquisas revelam dois tipos de comportamento. Os dados obtidos indicam que as pessoas apesar de pretenderem adquirir o objeto não o compraram, e vice-versa.

2.5.2. Informação Relativa de uma Decomposição Prevista

O objetivo da observação é prever como muitas pessoas realmente adquirem um dado objeto. Seja p a proporção prevista de objetos adquiridos e seja 1-p a proporção prevista de objetos não adquiridos. Considere q para a correspondente proporção observada de compras e 1-q para não compras e seja a informação (2.3).

$$I_2(Q:P) = q \log \frac{q}{p} + (1-q) \log \frac{1-q}{1-p}$$

A função (2.3) pode ser interpretada como a informação relativa da decomposição prevista (p,1-p). Consideramos a proporção prevista como a probabilidade anterior e a proporção observada como a probabilidade posterior, então (q,1-q) é a probabilidade posterior da probabilidade anterior (p,1-p). Quando as proporções previstas e observadas são $p_1,p_2,\ldots,p_n$ e $q_1,q_2,\ldots,q_n$ respectivamente, (2.3) pode ser estendido para:

$$I(Q:P) = \sum_{i=1}^{n} q_i \log \frac{q_i}{p_i} \qquad (2.29)$$

2.5.3. Exemplo Prático de Informação Relativa de uma Decomposição

A tabela 2.4 foi obtida através de pesquisas feitas trimestralmente durante determinado tempo. Foram feitas 19 observações e foram averiguadas as intenções das pessoas em adquirir um carro nos próximos 6 meses.

TABELA 2.4

INTENÇÃO	COMPORTAMENTO		
	Não Compraram	Compraram	Total
Não Comprariam	0,849	0,068	0,917
Comprariam	0,051	0,032	0,083
Total	0,900	0,100	1

As observações levam a 19 valores informação relativa, um para cada observação.

Verificamos pela Tabela 2.4 que aproximadamente 40% dos que pretendiam comprar um carro o fizeram, e 7,5% dos que não tencionavam comprar também adquiriram um carro. Se estas percentagens (40% e 7,5%) são constantes obtemos uma previsão da taxa de consumo. Escrever x_t para os que pretendiam comprar e $1-x_t$ para os que não pretendiam. A correspondente taxa é igual a:

$$0,4x_t + 0,075(1-x_t) = 0,075 + 0,325x_t \quad , \tag{2.30}$$

A condição na prática não é fácil de aplicar mas pode ser usada como uma aproximação para a taxa de consumo.

A informação média relativa desse método é mostrada na 2ª linha da tabela 2.5. O valor mostrado na primeira linha dessa tabela é a média aritmética das 19 informações relativas.

TABELA 2.5

MÉTODO PREVISTO	19 OBSERVAÇÕES	15 ÚLTIMAS OBSERVAÇÕES
Intenção inicial (x_t)	0,2341	
Método (2.30)	0,0764	0,0652
Taxa Média de consumo	0,2868	
após um trimestre (2.32)		0,0475
Após dois trimestres (2.33)		0,0371
Após três trimestres (2.34)		0,0243

É interessante o aperfeiçoamento do método "intenção inicial". Ele é inteiramente possível ainda que as previsões assim modificadas sejam desprezíveis.

Se a informação média relativa correspondente ao método taxa média de consumo for menor ou igual a do procedimento (2.30), o uso desse último método torna-se inadequado. A terceira linha da tabela 2.5 mostra que esse não é o caso. Observe, também que o método "intenção inicial" deu um resultado só um pouco melhor do que o da taxa média de consumo.

2.6. Informação Aperfeiçoada por uma Revisão Previsão

Supondo que a decomposição $(q_1, q_2, \ldots, q_n)$ é prevista para $(p_1, p_2, \ldots, p_n)$ e que em alguma etapa mais tarde a previsão é revisada para um novo conjunto de proporções $p_1', p_2', \ldots, p_n$. A informação relativa original é I(Q:P) como foi definida em (2.29) e o novo valor é I(Q:P'), sendo substituido os p_i em (2.29) por p_i'. Se a nova previsão é mais exata do que as anteriores, I(Q:P') deve ser menor do qüe I(Q:P). Portanto a diferença:

$$I(Q:P)-P(Q:P') = \sum_{i=1}^{n} q_i \log \frac{q_i}{p_i} - \sum_{i=1}^{n} q_i \log \frac{q_i}{P_i'}$$

$$= \sum_{i=1}^{n} q_i \log \frac{p_i'}{p_i} \qquad (2.31)$$

A diferença acima é chamada de Informação Aperfeiçoada da Revisão Prevista e é nula quando $p_i = p_i'$ considerados para cada i, mas esta é uma condição suficiente e não uma condição necessária. O valor máximo de uma informação aperfeiçoada I(Q:P) é obtido de uma revisão prevista exata. O aperfeiçoamento pode ser negativo, o que indica que a revisão não aperfeiçoou mas sim deteriorizou a informação.

Voltando agora ao exemplo anterior de carros comprados temos que a t-ésima observação nos dá a taxa prevista x_t e para o mesmo período a taxa de consumo observada é y_t originada da (t+4)-ésima observação.

Um trimestre após ao t-ésimo exame temos y_{t-3} como sendo a taxa de consumo que será utilizada para aperfeiçoar (2.30). Usamos pesos proporcionais.

$$\frac{1}{4}\, y_{t-3} + \frac{3}{4}\,(0{,}075 + 0{,}326x_t) \tag{2.32}$$

onde $\frac{1}{4}$ deve-se ao fato das observações serem feitas trimestralmente durante um ano. Após dois trimestres usamos y_{t-2} em vez de y_{t-3}. Como y_{t-2} e y_{t-3} são dois trimestres aumentamos o peso para $\frac{1}{2}$.

$$\frac{1}{2}\, y_{t-2} + \frac{1}{2}\,(0{,}075 + 0{,}325x_t) \tag{2.33}$$

Após três trimestres usamos y_{t-1} o que nos dá um peso igual a $\frac{3}{4}$.

$$\frac{3}{4}\, y_{t-1} + \frac{1}{4}\,(0{,}075 + 0{,}325x_t) \tag{2.34}$$

As informações médias relativas aos procedimentos (2.32), (2.33), (2.34) são mostradas na última coluna da tabela 2.5.

Os resultados indicam que as sucessivas taxas de consumo obtidas y_{t-3}, y_{t-2} e y_{t-1} são realmente informações referentes a y_t.

2.6.1. Decomposição da Informação Relativa e a Informação Aperfeiçoada

A decomposição (2.15) é válida também quando a informação prevista I(Q:P) é interpretada como uma informação relativa prevista de n cotas gastas. A decomposição é então uma combinação de produtos em G grupos de produtos e I(Q:P) é a informação relativa. $I_g(Q:P)$ é a medida relativa nos g-ésimos grupos, (g=1,2,.. ...,G). Temos que $I_g(Q:P)=0$ se e somente se o gasto previsto de cada produto do g-ésimo grupo é considerado como uma parte do gasto total do grupo.

Observe que o peso de $I_g(Q:P)$ na decomposição é Q_g, a cota observada gasta do grupo.

A informação aperfeiçoada (2.31) pode ser usada similarmente.

$$I_0(Q:P) - I_0(Q:P') = \sum_{g=1}^{G} T_g \log \frac{R'_g}{R_g} , \qquad (2.35)$$

onde $R'_g = \sum_{i\epsilon S_g} p'_i$, e no g-ésimo grupo é:

$$I_g(Q:P) - I_g(Q:P') = \sum_{i\epsilon S_g} \frac{q_i}{T_g} \log \frac{p'_i/R'_g}{p_i/R_g} \qquad g=1,2,\ldots,G \qquad (2.36)$$

Pode-se verificar que somando a (2.35) uma média proporcional das G expressões (2.36) com pesos $Q_1,Q_2,\ldots,Q_g$ obtemos I(Q:P) - I(Q:P') que é a informação relativa definida em (2.31).

Um exemplo de cotas gastas foi dado por Theil em 1971. Neste exemplo é apresentada a cota gasta de quatro grupos de produtos em 17 observações feitas antes da 2ª guerra mundial e 14 após a guerra.

No exemplo citado acima, Theil [19] mostra que $I_0(Q:P)$, a informação relativa entre os grupos alimentos e os grupos não alimentos é sempre menor do que $I(Q:P)$, a informação média relativa de demanda.

C A P Í T U L O 3

MEDIDAS GENERALIZADAS DA INFORMAÇÃO COM N REVISÕES

Neste capítulo apresentamos generalizações do trabalho apresentado por Sharma e Mittal [13] usando a medida de Informação Aperfeiçoada para N revisões (1.36) obtida por Taneja e Arora [17].

Inicialmente generalizamos os teoremas 1 e 2 do capítulo 1. A seguir apresentamos casos particulares e propriedades das generalizações obtidas.

3.1. Informação Própria não Aditiva

Seja a ocorrência de um evento x, associado a duas probabilidades p e q respectivamente. Caracterizamos a seguir $I(p,q,r_1,\ldots,r_N)$, a medida de informação própria sob a não-aditividade.

Teorema 1: Se uma função $I(p,q,r_1,\ldots,r_N)$ satisfaz os seguintes axiomas:

A_1: $I(p,q,r_1,\ldots,r_N)$ é contínua em $(0,1]$

A_2: $I(p_1p_2,q_1q_2,r_{11}r_{21},r_{12}r_{22},\ldots,r_{1N}r_{2N})$

$$=I(p_1,q_1,r_{11},r_{12},\ldots,r_{1N})+I(p_2,q_2,r_{21},r_{22},\ldots,r_{2N})$$

$$+\lambda I(p_1,q_1,r_{11},r_{12},\ldots,r_{1N})I(p_2,q_2,r_{21},r_{22},\ldots,r_{2N}), \qquad (3.2)$$

para todos os $p_1,p_2,q_1,q_2,r_{11},r_{12},\ldots,r_{1N}r_{2N}\ \varepsilon\ (0,1]$

com $\lambda \neq 0$

$$A_3=(1,1,\ldots,1,\frac{1}{2})=I(1,1,\ldots,\frac{1}{2},1)=\ldots=I(1,1,\frac{1}{2},\ldots,1)=\frac{1}{N}, \qquad (3.3)$$

$$A_4:I(1,\frac{1}{2},\frac{1}{2},\ldots,\frac{1}{2})=0 \qquad (3.4)$$

$$A_5:I(\frac{1}{2},1,\frac{1}{2},1,\ldots,1)=\frac{1}{N} \qquad (3.5)$$

então:

$$I(p,q,r_1,r_2,\ldots,r_N) = \frac{(2^{-\gamma}-1)^{-1}}{N}\left[\left(\frac{\prod_{k=1}^{N} r_k}{q_N}\right)^{\gamma}-1\right],$$

$$\gamma \neq 0, \qquad (3.6)$$

Prova:

Seja $I(p,q,r_1,r_2,\ldots,r_N)$

$= f(p,q,r_1,r_2,\ldots,r_N)$, então de A_2, temos:

$$f(p_1p_2,q_1q_2,r_{11}r_{21},r_{12}r_{22},\dots,r_{1N}r_{2N}) =$$

$$f(p_1,q_1,r_{11},r_{12},\dots,r_{1N})+f(p_2,q_2,r_{21},r_{22},\dots,r_{2N})$$

$$+\lambda f(p_1,q_1,r_{11},r_{12},\dots,r_{1N})f(p_2,q_2,r_{21},r_{22},\dots,r_{2N}) \qquad (3.7)$$

com $\lambda \neq 0$

ou

$$1+\lambda f(p_1p_2,q_1q_2,r_{11}r_{21},r_{12}r_{22},\dots,r_{1N}r_{2N})$$

$$=[1+\lambda f(p_1,q_1,r_{11},r_{12},\dots,r_{1N}][1+\lambda f(p_2,q_2,r_{21},r_{22},\dots,r_{2N})] \;,$$

sendo $\lambda \neq 0$

ou seja:

$$F(p_1p_2,q_1q_2,r_{11}r_{21},r_{12}r_{22},\dots,r_{1N}r_{2N})$$

$$=F(p_1,q_1,r_{11},r_{12},\dots,r_{1N})F(p_2,q_2,r_{21},r_{22},\dots,r_{2N}) \qquad (3.8)$$

onde

$$F(p_1,q_1,r_{11},r_{12},\dots,r_{1N})=1+\lambda f(p_1,q_1,r_{11},r_{12},\dots,r_{1N})$$

As soluções gerais contínuas da equação funcional (3.8) são dadas por:

$$F(p,q,r_1,r_2,\dots,r_N)=p^{\alpha}q^{\beta},r_1^{\gamma_1},\dots,r_N^{\gamma_N} \;, \qquad (3.9)$$

onde $\alpha > 0$, β, $\gamma_1, \dots, \gamma_N$ são constantes arbitrárias.

Como:

$$F(p,q,r_1,r_2,\dots,r_N) = 1 + \lambda f(p,q,r_1,r_2,\dots,r_N)$$

temos que:

$$1 + \lambda f(p,q,r_1,r_2,\dots,r_N) = p^{\alpha} q^{\beta} r_1^{\gamma_1} \dots r_N^{\gamma_N}$$

então:

$$f(p,q,r_1,r_2,\dots,r_N) = \frac{p^{\alpha} q^{\beta} r_1^{\gamma_1} \dots r_N^{\gamma_N} - 1}{\lambda}, \qquad (3.10)$$

mas por A_3

$$\frac{1}{N} = I(1,1,\dots,1,\frac{1}{2}) = \frac{1^{\alpha}\, 1^{\beta}\, 1^{\gamma_1} \dots (\frac{1}{2})^{\gamma_N} - 1}{\lambda}$$

$$\frac{\lambda}{N} = (\frac{1}{2})^{\gamma_N} - 1, \quad \text{ou seja}$$

$$\lambda = N(2^{-\gamma_N} - 1).$$

Logo:

$$I(p,q,r_1,r_2,\dots,r_N) = \frac{p^{\alpha} q^{\beta} r_1^{\gamma_1} \dots r_N^{\gamma_N} - 1}{N(2^{-\gamma_N} - 1)}$$

Ainda por A_3

$$\frac{1}{N} = I(1,1,1,\ldots,\frac{1}{2},1) = \frac{1\ 1\ \ldots\ (\frac{1}{2})^{\gamma_{N-1}} - 1}{N(2^{-\gamma_N} - 1)}$$

$$2^{-\gamma_N} - 1 = 2^{-\gamma_{N-1}} - 1$$

logo $\gamma_N = \gamma_{N-1} = \gamma$

Portanto usando A_3 sucessivamente obtemos:

$$\gamma_1 = \gamma_2 = \ldots\ldots = \gamma_N = \gamma$$

Usando A_4 temos:

$$0 = I(1,\frac{1}{2},\frac{1}{2},\ldots,\frac{1}{2}) = \frac{1\ (\frac{1}{2})^{\beta}\ (\frac{1}{2})^{\gamma_1}\ \ldots\ (\frac{1}{2})^{\gamma_N} - 1}{N(2^{-\gamma_N} - 1)},$$

Sendo:

$$\gamma_1 = \gamma_2 = \ldots = \gamma_N = \gamma$$

$$2^{-\beta}\ 2^{-\gamma}\ \ldots\ 2^{-\gamma} - 1 = 0, \text{ logo } \beta = -N\gamma$$

Por A_5 temos:

$$\frac{1}{N} = I(\frac{1}{2},1,\frac{1}{2},1,\ldots,1) = \frac{(\frac{1}{2})^{\alpha}\ 1^{\beta}\ (\frac{1}{2})^{\gamma_1}\ 1^{\gamma_2}\ \ldots\ 1^{\gamma_N} - 1}{N(2^{-\gamma_N} - 1)}$$

Como $\gamma_1 = \gamma_2 = \ldots = \gamma_N = \gamma$, temos:

$$\frac{1}{N} = \frac{2^{-\alpha}\, 2^{-\gamma} - 1}{N(2^{-\gamma} - 1)} \implies \alpha = 0$$

Portanto:

$$I(p,q,r_1,r_2,\ldots,r_N) = \frac{(2^{-\gamma}-1)^{-1}}{N}\left[\left(\frac{\prod_{K=1}^{N} r_k}{q^N}\right)^{\gamma} - 1\right], \quad \gamma \neq 0$$

3.2. Medidas Generalizadas de Informação

Consideremos as medidas não aditivas como uma média generalizada da informação própria $I(P;Q;R_1:\ldots;R_N)$, com pesos como funções das probabilidades, isto é:

$$I(P;Q;R_1;\ldots;R_N) = \phi^{-1}\left\{\frac{\sum_{i=1}^{n} f(p_i)\phi[I(p_i,q_i,r_{i1},\ldots,r_{iN})]}{\sum_{i=1}^{n} f(p_i)}\right\} \qquad (3.11)$$

onde ϕ é uma função contínua estritamente monótona. A função ϕ em (3.11) deve satisfazer a condição:

$$I(PU;QV;R_1S_1;\ldots;R_NS_N)$$

$$= I(P;Q;R_1;\ldots;R_N)+I(U;V;S_1;\ldots;S_N)$$

$$+ N(2^{-\gamma}-1)I(P;Q;R_1;\ldots;R_N)I(U;V;S_1;\ldots;S_N) \qquad (3.12)$$

onde: $U = \{u\}$, $V = \{v\}$, $S_1 = \{s_1\}, \ldots, S_N = \{s_N\}$

Considerando $f(x)=x$, caracterizamos o seguinte teorema:

Teorema 2:

A informação generalizada contínua $I(P,Q,R_1,\ldots,R_N)$ dada em (3.11), para $f(x)=x$, com distribuição de probabilidade $P,Q,R_1,\ldots,R_N$ sendo $P=(p_1,p_2,\ldots,p_n)$, $Q=(q_1,q_2,\ldots,q_n)$, $R_1=(r_{11},r_{12},\ldots,r_{1n}),\ldots,R_N=(r_{N1},r_{N2},\ldots,r_{Nn})$ e

$$\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} q_i = \sum_{i=1}^{n} r_{i1} = \ldots. = \sum_{i=1}^{n} r_{iN} = 1 \quad \text{com} \quad p_i \geqslant 0,$$

$q_i \geqslant 0$; $r_{i1} \geqslant 0, \ldots, r_{iN} \geqslant 0$, $i=1,2,\ldots,n$, satisfazendo a não aditividade. (3.12) pode ser expressa somente por uma das seguintes formas:

$$I(P;Q;R_1;\ldots;R_N;1,\gamma) = \frac{2^{\gamma \sum\limits_{i=1}^{n} p_i \log\left(\frac{\prod\limits_{k=1}^{N} r_{k_i}}{q_i^N}\right)} - 1}{N(2^{-\gamma} - 1)} \qquad (3.13)$$

$$I(P;Q;R_1;\ldots;R_N;\alpha,\gamma) = \frac{\left[\sum\limits_{i=1}^{n} p_i \left(\frac{\prod\limits_{k=1}^{N} r_{k_i}}{q_i^N}\right)^{\alpha-1}\right]^{\frac{\gamma}{\alpha-1}} - 1}{N(2^{-\gamma}-1)}$$

$$\gamma \neq 0,\ \alpha \neq 1,\ \alpha,\ \gamma > 0, \qquad (3.14)$$

Prova:

Temos de (3.12) que:

$I(Pu;Qv;R_1 s_1;\ldots;R_N s_N)$

$=I(P;Q;R_1;\ldots;R_N)+I(u;v;s_1;\ldots;s_N)$

$+CI(P;Q;R_1;\ldots;R_N)I(u;v;s_1;\ldots;s_N)$

onde $C=N(2^{-\gamma}-1)$

Pelo teorema 1, temos que:

$$I(p;q;r_1;\ldots;r_N) = \frac{(2^{-\gamma}-1)^{-1}}{N}\left[\left(\frac{\prod_{k=1}^{N} r_k}{q^N}\right)^{\gamma} - 1\right], \qquad (3.15)$$

Usando (3.11) e (3.15) temos:

$$\phi^{-1}\left\{\frac{\sum_{i=1}^{n} p_i\,\phi\left[\left(\frac{\prod_{k=1}^{N} r_{k_i} s_k}{(u_i v)^N}\right)^{\gamma} - 1\right]}{C}\right\}$$

$$= \phi^{-1}\left\{\frac{\sum_{i=1}^{n} p_i\,\phi\left[\left(\frac{\prod_{k=1}^{N} r_{k_i}}{u_i^N}\right)^{\gamma} - 1\right]}{C}\right\} + \frac{\left(\frac{\prod_{k=1}^{N} s_k}{v^N}\right)^{\gamma} - 1}{C}$$

$$+C\phi^{-1}\left\{\frac{\sum_{i=1}^{n} p_i\,\phi\left[\left(\frac{\prod_{k=1}^{N} r_{k_i}}{u_i^N}\right)^{\gamma}-1\right]}{C}\right\}\cdot\left\{\frac{\left(\frac{\prod_{k=1}^{N} s_k}{v^N}\right)^{\gamma}-1}{C}\right\}$$

ou seja:

$$\phi-1\left\{\frac{\sum_{i=1}^{n} p_i\,\phi\left[\left(\frac{\prod_{k=1}^{N} r_{k_i}\,s_k}{(u_i v)^N}\right)^{\gamma}-1\right]}{C}\right\} =$$

$$= \phi-1\left\{\frac{\sum_{i=1}^{n} p_i\,\phi\left[\left(\frac{\prod_{k=1}^{N} r_{k_i}}{u_i^N}\right)^{\gamma}-1\right]}{C}\right\}\cdot\left(\frac{\prod_{k=1}^{N} s_k}{v^N}\right)^{\gamma}$$

$$+\frac{\left(\frac{\prod_{k=1}^{N} s_k}{v^N}\right)^{\gamma}-1}{C}, \qquad (3.16)$$

ou

$$\psi^{-1}_{u,v,s_1,\ldots,s_N}\left\{\frac{\sum_{i=1}^{n} p_i\,\psi\left[\left(\frac{\prod_{k=1}^{N} r_{k_i}}{u_i^N}\right)^{\gamma}-1\right]}{C}\right\}$$

$$= \phi^{-1}\left\{\frac{\sum_{i=1}^{n} p_i\,\phi\left[\left(\frac{\prod_{k=1}^{N} r_{k_i}}{u_i^N}\right)^{\gamma}-1\right]}{C}\right\}, \qquad (3.17)$$

onde

$$\Psi_{u,v,s_1,\ldots,s_N}\left[\frac{\left(\frac{\prod_{k=1}^{N} r_{k_i}}{u_i^N}\right)^{\gamma} - 1}{C}\right] = \phi\left[\frac{\left(\frac{\prod_{k=1}^{N} r_{k_i} s_k}{(u_i v)^N}\right)^{\gamma} - 1}{C}\right], \quad (3.18)$$

Existe uma relação linear entre ϕ e $\Psi_{u,v,s_1,\ldots,s_N}$, (ref. a Hardy, Littlewood, e Pólya [20]),

$$\Psi\left(\frac{x^{\gamma}-1}{C}\right) = A(y)\,\phi\left(\frac{x^{\gamma}-1}{C}\right) + B(y)$$

Sendo:

$$x = \frac{\prod_{k=1}^{N} r_{k_i}}{u_i^N} \quad \text{e} \quad y = \frac{\prod_{k=1}^{N} s_k}{v^N},$$

então:

$$\phi\left(\frac{x^{\gamma}y^{\gamma}-1}{C}\right) = A(y)\,\phi\left(\frac{x^{\gamma}-1}{C}\right) + B(y), \quad (3.19)$$

ou seja:

$$g(xy) = A(y)\,g(x) + B(y) \quad (3.20)$$

onde

$$g(x) = \phi\left(\frac{x^{\gamma}-1}{C} \right) \tag{3.21}$$

ou $G(xy)=A(y)G(x)+G(y)$ (3.22)

onde $G(x)=g(x)-g(1)$ (3.23)

Por simetria temos:

$G(xy)=G(yx)$, logo

$A(y)G(x)+G(y)=A(x)G(y)+G(x)$,

desta forma temos:

$$G(x)\left[A(y)-1\right]=G(y)\left[A(x)-1\right] \tag{3.24}$$

isto é

$$\frac{G(x)}{A(x)-1} = \frac{G(y)}{A(y)-1} = \text{constante}$$

Consideramos dois casos:

1) $A(x) - 1 = 0$

2) $A(x) - 1 \neq 0$

1º Caso: Quando $A(x)-1=0$, implica que $A(x)=1$, então (3.22) reduz-se a

$$G(xy) = G(x) + G(y) \tag{3.25}$$

A solução geral de (3.25) é dada por:

$G(x) = A \log x, \; x > 0$ (ver Aczél [2])

Considerando $g(1)=a$ e usando (3.21) e (3.23) temos:

$$\phi\left(\frac{x^{\gamma}-1}{C}\right) = a + A \log x , \qquad \gamma \neq 0 , \qquad (3.26)$$

como $C = N(2^{-\gamma}-1)$, então

$$\phi\left[\frac{x^{\gamma}-1}{N(2^{-\gamma}-1)}\right] = a + A \log x$$

$$= a + \frac{A}{\gamma} \log x^{\gamma}$$

$$= a + \frac{A}{\gamma} \log \left[1 + \frac{x^{\gamma}-1}{N(2^{-\gamma}-1)} \cdot N(2^{-\gamma}-1)\right]$$

fazendo $x = \dfrac{x^{\gamma}-1}{N(2^{-\gamma}-1)}$, obtemos

$$\phi(x) = a + \frac{A}{\gamma} \log \left[1+xN(2^{-\gamma}-1)\right] \qquad \gamma \neq 0 \qquad (3.27)$$

Fazendo $f(p_i)=p_i$, a expressão (3.11) pode ser escrita dessa forma:

$$\phi\left[I(P,Q,R_1,\ldots,R_N)\right] = \frac{\sum_{i=1}^{n} p_i\,\phi\left[I(p_i,q_i,r_{i1},\ldots,r_{iN})\right]}{\sum_{i=1}^{n} p_i}$$

$$= \sum_{i=1}^{n} p_i\,\phi\left[I(p_i,q_i,r_{i1},\ldots,r_{iN})\right]$$

sendo que $\sum_{i=1}^{n} p_i = 1$

Considerando o valor de $\phi(x)$ dado em (3.27), temos:

$$a + \frac{A}{\gamma}\log\left[1+I(P;Q;R_1,\ldots,R_N)N(2^{-\gamma}-1)\right]$$

$$= \sum_{i=1}^{n} p_i\left\{a + \frac{A}{\gamma}\log\left[1+I(p_i,q_i,r_{i1},\ldots,r_{iN})N(2^{-\gamma}-1)\right]\right\}$$

$$a + \frac{A}{\gamma}\log\left[1+I(P,Q,R_1,\ldots,R_N)N(2^{-\gamma}-1)\right]$$

$$= a + \frac{A}{\gamma}\left\{\sum_{i=1}^{n} p_i\log\left[1+I(p_i,q_i,r_{i1},\ldots,r_{iN})N(2^{-\gamma}-1)\right]\right\}$$

então: $\log\left[1+I(P,Q,R_1,\ldots,R_N)N(2^{-\gamma}-1)\right]$

$$= \sum_{i=1}^{n} p_i\log\left[1+I(p_i,q_i,r_{i1},\ldots,r_{iN})N(2^{-\gamma}-1)\right]$$

Usando o teorema 1, temos:

$$\log\left[1+I(P,Q,R_1,\ldots,R_N)N(2^{-\gamma}-1)\right] = \sum_{i=1}^{n} p_i \log\left[1+\frac{\left(\dfrac{\prod_{k=1}^{N} r_{k_i}}{q_i^N}\right)^{\gamma}-1}{N(2^{-\gamma}-1)}.N(2^{-\gamma}-1)\right]$$

$$\log\left[1+I(P,Q,R_1,\ldots,R_N)N(2^{-\gamma}-1)\right] = \sum_{i=1}^{n} p_i\, \gamma \log\left(\frac{\prod_{k=1}^{N} r_{k_i}}{q_i^N}\right)$$

$$1+I(P,Q,R_1,\ldots,R_N)N(2^{-\gamma}-1) = 2^{\gamma \sum_{i=1}^{n} p_i \log\left(\frac{\prod_{k=1}^{N} r_{k_i}}{q_i^N}\right)}$$

$$I(P,Q,R_1,\ldots,R_N) = \frac{2^{\gamma \sum_{i=1}^{n} p_i \log\left(\frac{\prod_{k=1}^{N} r_{k_i}}{q_i^N}\right)}-1}{N(2^{-\gamma}-1)}, \quad \gamma \neq 0$$

<u>2º Caso</u>: Quando $A(x) - 1 \neq 0$, por (3.24) temos: $\dfrac{G(x)}{A(x)-1} = \dfrac{G(y)}{A(y)-1} = \dfrac{1}{K}$

ou $A(x)-1 = KG(x)$

$A(xy)-1 = KG(xy)$, portanto

$$A(xy) = A(y)\,A(x) \qquad (3.28)$$

A solução geral de (3.28) é dada por

$A(x) = x^{\alpha-1}$ (ref. Aczél [1]), $\alpha \neq 1$, $\alpha > 0$

Se $A(x) = 0$ então $G(x) =$ constante

Logo:

$$G(x) = \frac{x^{\alpha-1} - 1}{K} \,,$$

sendo $G(x) - g(1) = g(x)$ e fazendo $g(1)=a$, temos:

$$g(x) = G(x) + a$$

$$g(x) = a + \frac{x^{\alpha-1}-1}{K}$$

ou $\phi\left(\dfrac{x^{\gamma}-1}{C}\right) = a + \dfrac{x^{\alpha-1}-1}{K}$, por (3.21)

Como $C = N(2^{-\gamma}-1)$, temos:

$$\phi\left(\frac{x^{\gamma}-1}{N(2^{-\gamma}-1)}\right) = a + \frac{x^{\alpha-1}-1}{K}$$

$$= a + \frac{(x^{\gamma})^{\frac{\alpha-1}{\gamma}} - 1}{K}$$

$$= a + \frac{\left[\frac{x^{\gamma}-1}{N(2^{-\gamma}-1)} \cdot N(2^{-\gamma}-1) + 1\right]^{\frac{\alpha-1}{\gamma}} - 1}{K} \qquad (3.29)$$

Trocando x por $\dfrac{x^{\gamma}-1}{N(2^{-\gamma}-1)}$ em (3.29) obtemos:

$$\phi(x) = a + \frac{[N(2^{-\gamma}-1)x+1]^{\frac{\alpha-1}{\gamma}} - 1}{K}, \quad \alpha \neq 1,\ \gamma \neq 0,\ \alpha > 0 \qquad (3.30)$$

Fazendo $f(p_i)=p_i$, usando (3.11) e considerando $\phi(x)$ dado em (3.30), obtemos:

$$a + \frac{[N(2^{-\gamma}-1)I(P,Q,R_1,\ldots,R_N)+1]^{\frac{\alpha-1}{\gamma}} - 1}{K}$$

$$= \sum_{i=1}^{n} p_i \left\{ a + \frac{[N(2^{-\gamma}-1)I(p_i,q_i,r_{i1},\ldots,r_{iN})+1]^{\frac{\alpha-1}{\gamma}} - 1}{K} \right\}$$

$$\left[N(2^{-\gamma}-1)I(P,Q,R_1,\ldots,R_N)+1\right]^{\frac{\alpha-1}{\gamma}}$$

$$= \sum_{i=1}^{n} p_i \left[N(2^{-\gamma}-1)I(p_i,q_i,r_{i1},\ldots,r_{iN})+1\right]^{\frac{\alpha-1}{\gamma}}$$

Usando (3.6) temos:

$$\left[N(2^{-\gamma}-1)I(P,Q,R_1,\ldots,R_N)+1\right]^{\frac{\alpha-1}{\gamma}}$$

$$= \sum_{i=1}^{n} p_i \left[N(2^{-\gamma}-1)\, \frac{\left(\dfrac{\prod_{k=1}^{N} r_{k_i}}{q_i^N}\right)^{\gamma} - 1}{N(2^{-\gamma}-1)} + 1 \right]^{\frac{\alpha-1}{\gamma}}$$

$$\left[N(2^{-\gamma}-1)I(P,Q,R_1,\ldots,R_N)+1\right]^{\frac{\alpha-1}{\gamma}} = \sum_{i=1}^{n} p_i \left[\left(\frac{\prod_{k=1}^{N} r_{k_i}}{q_i^N}\right)^{\gamma}\right]^{\frac{\alpha-1}{\gamma}}$$

$$= \sum_{i=1}^{n} p_i \left(\frac{\prod_{k=1}^{N} r_{k_i}}{q_i^N}\right)^{\alpha-1}$$

$$N(2^{-\gamma}-1)I(P,Q,R_1,\ldots,R_N)+1 = \left[\sum_{i=1}^{n} p_i \left(\frac{\prod_{k=1}^{N} r_{k_i}}{q_i^N}\right)^{\alpha-1}\right]^{\frac{\gamma}{\alpha-1}}$$

$$I(P,Q,R_1,\ldots,R_N) = \frac{\left[\sum_{i=1}^{n} p_i \left(\dfrac{\prod_{k=1}^{N} r_{k_i}}{q_i^N}\right)^{\alpha-1}\right]^{\frac{\gamma}{\alpha-1}} - 1}{N(2^{-\gamma}-1)} .$$

Observações: Foi considerado no teorema anterior que o $\log 0 = 0 \log \frac{0}{0} = 0$ e $0^{\alpha} = 0$ para $\alpha > 0$. Sempre que q_i ou r_{k_i} $(k=1,2,\ldots,N)$ for zero, então o correspondente p_i também é zero para todo $i=1,2,\ldots,n$. A base do logaritmo é sempre 2.

3.3. Casos Particulares

Mostraremos a seguir algumas relações entre a Informação Generalizada $I(P;Q;R_1;\ldots;R_N)$ e outras medidas já conhecidas.

1) $\lim\limits_{\alpha \to 1} I(P;Q;R_1;\ldots;R_1;\ldots;R_N;\alpha,\gamma) = I(P;Q;R_1;\ldots;R_N;1,\gamma)$

Prova:

Em (3.14) temos que:

$$I(P;Q;R_1;\ldots;R_N;\alpha,\gamma) = \frac{\left[\sum\limits_{i=1}^{n} p_i \left(\frac{\prod\limits_{k=1}^{N} r_{k_i}}{q_i^N}\right)^{\alpha-1}\right]^{\frac{\gamma}{\alpha-1}} - 1}{N(2^{-\gamma}-1)}$$

$$N(2^{-\gamma}-1) I(P;Q;R_1;\ldots;R_N;\alpha,\gamma) + 1 = \left[\sum\limits_{i=1}^{n} p_i \left(\frac{\prod\limits_{k=1}^{N} r_{k_i}}{q_i^N}\right)^{\alpha-1}\right]^{\frac{\gamma}{\alpha-1}}$$

$$\log\left[N(2^{-\gamma}-1)I(P;Q;R_1;\ldots;R_N;\alpha,\gamma)+1\right]=\frac{\gamma}{\alpha-1}\log\left[\sum_{i=1}^{n}p_i\left(\frac{\prod_{k=1}^{N}r_{k_i}}{q_i^N}\right)^{\alpha-1}\right]$$

$$\lim_{\alpha\to 1}\left[\log N(2^{-\gamma}-1)I(P;Q;R_1;\ldots;R_N;\alpha,\gamma)+1\right]$$

$$=\gamma\lim_{\alpha\to 1}\frac{\log\left[\sum_{i=1}^{n}p_i\left(\frac{\prod_{k=1}^{n}r_{k_i}}{q_i^N}\right)^{\alpha-1}\right]}{\alpha-1}$$

$$\lim_{\alpha\to 1}\left[\log N(2^{-\gamma}-1)I(P;Q;R_1;\ldots;R_N;\alpha,\gamma)+1\right]$$

$$=\gamma\lim_{\alpha\to 1}\frac{\dfrac{\sum_{i=1}^{n}p_i\left(\frac{\prod_{k=1}^{N}r_{k_i}}{q_i^N}\right)^{\alpha-1}\cdot\ell n\left(\frac{\prod_{k=1}^{n}r_{k_i}}{q_i^N}\right)}{\sum_{i=1}^{n}p_i\left(\frac{\prod_{k=1}^{N}r_{k_i}}{q_i^N}\right)^{\alpha-1}\ell n\ 2}}{1-0}$$

$$\lim_{\alpha\to 1}\left[\log N(2^{-\gamma}-1)I(P;Q;R_1;\ldots;R_N;\alpha,\gamma)+1\right]$$

$$=\gamma\,\frac{\sum_{i=1}^{n}p_i\,\ell n\left(\frac{\prod_{k=1}^{N}r_{k_i}}{q_i^N}\right)}{\sum_{i=1}^{n}p_i\,\ell n\ 2}$$

$$\lim_{\alpha\to 1}\left[\log N(2^{-\gamma}-1)I(P;Q;R_1;\ldots;R_N,\alpha,\gamma)+1\right]$$

$$= \gamma\,\frac{\sum_{i=1}^{n} p_i \log\left(\frac{\prod_{k=1}^{N} r_{k_i}}{q_i^N}\right)}{\sum_{i=1}^{n} p_i}$$

$$\lim_{\alpha\to 1} \log\left[N(2^{-\gamma}-1)I(P;Q;R_1;\ldots;R_N;\alpha,\gamma)+1\right]$$

$$= \gamma \sum_{i=1}^{n} p_i \log\left(\frac{\prod_{k=1}^{N} r_{k_i}}{q_i^N}\right)$$

$$\log \lim_{\alpha\to 1} \left[N(2^{-\gamma}-1)I(P;Q;R_1;\ldots;R_N;\alpha,\gamma)+1\right]$$

$$= \gamma \sum_{i=1}^{n} p_i \log\left(\frac{\prod_{k=1}^{N} r_{k_i}}{q_i^N}\right)$$

$$\lim_{\alpha\to 1}\left[N(2^{-\gamma}-1)I(P;Q;R_1;\ldots;R_N;\alpha,\gamma)+1\right] = 2^{\gamma \sum_{i=1}^{n} p_i \log\left(\frac{\prod_{k=1}^{N} r_{k_i}}{q_i^N}\right)}$$

$$\lim_{\alpha\to 1} I(P;Q;R_1;\ldots;R_N;\alpha,\gamma) = \frac{2^{\gamma \sum_{i=1}^{n} p_i \log\left(\frac{\prod_{k=1}^{N} r_{k_i}}{q_i^N}\right)} - 1}{N(2^{-\gamma}-1)}$$

logo:

$$\lim_{\alpha\to 1} I(P;Q;R_1;\ldots;R_N;\alpha,\gamma) = I(P;Q;R_1;\ldots;R_N;1,\gamma)$$

2) Considerando $\alpha-1 = \gamma$, temos:

$$I(P;Q;R_1;\ldots;R_N) = \frac{\sum_{i=1}^{n} p_i \left(\frac{\prod_{k=1}^{N} r_{k_i}}{q_i^N}\right)^{\gamma} - 1}{N(2^{-\gamma}-1)}$$

2.1) $$\lim_{\gamma\to 0} I(P;Q;R_1;\ldots;R_N) = -\sum_{i=1}^{n} p_i \log\left(\frac{\prod_{k=1}^{n} r_{k_i}}{q_i^N}\right)^{1/N}$$

Prova:

Por (3.6) temos que:

$$\lim_{\gamma\to 0} I(P;Q;R_1;\ldots;R_N) = \lim_{\gamma\to 0} \frac{\sum_{i=1}^{n} p_i \left(\frac{\prod_{k=1}^{N} r_{k_i}}{q_i^N} \right)^{\gamma} - 1}{N(2^{-\gamma}-1)}$$

$$= \lim_{\gamma\to 0} \frac{\sum_{i=1}^{n} p_i \left(\frac{\prod_{k=1}^{N} r_{k_i}}{q_i^N} \right)^{\gamma} \ell n \left(\frac{\prod_{k=1}^{N} r_{k_i}}{q_i^N} \right)}{N(2^{-\gamma}\ell n\ 2\ (-1)-0)}$$

$$= \frac{\sum_{i=1}^{n} p_i\ \ell n \left(\frac{\prod_{k=1}^{N} r_{k_i}}{q_i^N} \right)}{N(-\ell n\ 2)}$$

Concluindo temos que:

$$\lim_{\gamma\to 0} I(P;Q;R_1;\ldots;R_N) = -\sum_{i=1}^{n} p_i \log \left(\frac{\prod_{k=1}^{N} r_{k_i}}{q_i^N} \right)^{1/N}$$

PROPRIEDADES

As medidas (3.13) e (3.14) caracterizadas anteriormente apresentam as seguintes propriedades:

1) Se $R_1 = \dots = R_N = R$, então:

a) $$I(P;Q;R;1,\gamma) = \frac{2^{\gamma \sum_{i=1}^{n} p_i \log \left(\frac{r_i}{q_i}\right)^N} - 1}{N(2^{-\gamma}-1)}$$

$$\lim_{\gamma \to 0} I(P;Q;R;1,\gamma) = - \sum_{i=1}^{n} p_i \log \frac{r_i}{q_i}$$

A medida acima é a Informação Aperfeiçoada de Theil.

Prova:

$$\lim_{\gamma \to 0} I(P;Q;R;1,\gamma) = \lim_{\gamma \to 0} \frac{2^{\gamma \sum_{i=1}^{n} p_i \log \left(\frac{r_i}{q_i}\right)^N} - 1}{N(2^{-\gamma}-1)}$$

$$= \lim_{\gamma \to 0} \frac{2^{\gamma \sum_{i=1}^{n} p_i \log \left(\frac{r_i}{q_i}\right)^N} \ell n\, 2 \sum_{i=1}^{n} p_i \log \left(\frac{r_i}{q_i}\right)^N - 0}{N(-2^{-\gamma}\, n\, 2)}$$

$$= \frac{\ln 2 \sum_{i=1}^{n} p_i \, N \log\left(\frac{r_i}{q_i}\right)}{-N \ln 2}$$

$$= - \sum_{i=1}^{n} p_i \log\left(\frac{r_i}{q_i}\right)$$

b) $$I(P;Q;R;\alpha,\gamma) = \frac{\left[\sum_{i=1}^{n} p_i \left(\frac{r_i}{q_i}\right)^{N(\alpha-1)}\right]^{\frac{\gamma}{\alpha-1}} - 1}{N(2^{-\gamma}-1)}$$

Considerando $\alpha-1 = \gamma$ e calculando o

$\lim\limits_{\gamma \to 0} I(P;Q;R;\gamma)$ obtemos a Informação Aperfeiçoada de Theil.

$$\lim_{\gamma \to 0} (P;Q;R;\gamma) = \lim_{\gamma \to 0} \frac{\sum_{i=1}^{n} p_i \left(\frac{r_i}{q_i}\right)^{N\gamma} - 1}{N(2^{-\gamma}-1)}$$

$$= \lim_{\gamma \to 0} \frac{\sum_{i=1}^{n} p_i \left(\frac{r_i}{q_i}\right)^{N\gamma} . \ln\left(\frac{r_i}{q_i}\right) . N - 0}{N(-2^{-\gamma} \ln 2)}$$

$$= \frac{\sum_{i=1}^{n} p_i \ln\left(\frac{r_i}{q_i}\right)}{- \ln 2}$$

$$= - \sum_{i=1}^{n} p_i \log\left(\frac{r_i}{q_i}\right)$$

2) a) A medida:

$$I(P;Q;R_1;\ldots;R_N,1,\gamma) \begin{matrix} > \\ = \\ < \end{matrix} 0 \quad \Longleftrightarrow$$

$$\sum_{i=1}^{n} p_i \left(\frac{q_i^N}{r_{1i}\, r_{2i} \,\ldots\, r_{Ni}}\right) \begin{matrix} < \\ = \\ > \end{matrix} 1$$

b) A medida:

$$I(P;Q;R_1;\ldots;R_N;\alpha,\gamma) \begin{matrix} > \\ = \\ < \end{matrix} 0 \quad \Longleftrightarrow$$

$$\sum_{i=1}^{n} p_i \left(\frac{q_i^N}{r_{1i}\, r_{2i} \,\ldots\, r_{Ni}}\right)^{1-\alpha} \begin{matrix} > \\ = \\ < \end{matrix} 1$$

3) a) $I(P;U;R_1;\ldots;R_N+Q;V;S_1;\ldots;S_N)$

$$= I(P;U;R_1;\ldots;R_N)+I(Q;V;S_1;\ldots;S_N)$$

$$+N(2^{-\gamma}-1)I(P;U;R_1;\ldots;R_N)I(Q;V;S_1;\ldots;S_N)$$

<u>Prova</u>:

$$I(P;U;R_1;\ldots;R_N)+I(Q;V;S_1;\ldots;S_N)$$

$$+N(2^{-\gamma}-1)I(P;U;R_1;\ldots;R_N)I(Q;V;S_1;\ldots;S_N)$$

$$=I(P;U;R_1;\ldots;R_N)\left[1+N(2^{-\gamma}-1)I(Q;V;S_1;\ldots;S_N)\right]$$

$$+I(Q;V;S_1;\ldots;S_N)$$

$$=\frac{2^{\gamma\sum_{i=1}^{n}p_i\log\left(\frac{\prod_{k=1}^{N}r_{k_i}}{u_i^N}\right)}-1}{N(2^{-\gamma}-1)}\left[1+N(2^{-\gamma}-1)\frac{2^{\gamma\sum_{j=1}^{m}q_j\log\left(\frac{\prod_{k=1}^{N}s_{k_j}}{v_j^N}\right)}-1}{N(2^{-\gamma}-1)}\right]$$

$$+\frac{2^{\gamma\sum_{j=1}^{m}q_j\log\left(\frac{\prod_{k=1}^{N}s_{k_j}}{v_j^N}\right)}-1}{N(2^{-\gamma}-1)}$$

$$=\frac{2^{\gamma\sum_{i=1}^{n}p_i\log\left(\frac{\prod_{k=1}^{N}r_{k_i}}{u_i^N}\right)}\cdot 2^{\gamma\sum_{j=1}^{m}q_j\log\left(\frac{\prod_{k=1}^{N}s_{k_j}}{v_j^N}\right)}-1}{N(2^{-\gamma}-1)}$$

$$=\frac{1}{N(2^{-\gamma}-1)}:2^{\left(\sum_{i=1}^{n}p_i\log\left(\frac{\prod_{k=1}^{N}r_{k_i}}{u_i^N}\right)\right)^{\gamma}+\sum_{j=1}^{m}q_j\log\left(\frac{\prod_{k=1}^{N}s_{k_j}}{v_j^N}\right)^{\gamma}}-1$$

$$=I(P;U;R_1;\ldots;R_N+Q;V;S_1;\ldots;S_N)$$

b) $I(P*Q;U*V;R_1*S_1;\ldots;R_N*S_N)$

$$= I(P;U;R_1;\ldots;R_N)+I(Q;V;S_1;\ldots;S_N)$$

$$+ (2^{-\gamma}-1)N\ I(P;U;R_1;\ldots;R_N)I(Q;V;S_1;\ldots;S_N)$$

$$= I(P;U;R_1,\ldots;R_N)\left[1+N(2^{-\gamma}-1)I(Q;V;S_1;\ldots;S_N)\right]$$

$$+ I(Q;V;S_1;\ldots;S_N)$$

$$= \frac{\left[\sum_{i=1}^{n} p_i\left(\frac{\prod_{k=1}^{N} r_{k_i}}{u_i^N}\right)^{\alpha-1}\right]^{\frac{\gamma}{\alpha-1}} -1}{N(2^{-\gamma}-1)} \cdot \left[\sum_{j=1}^{m} q_j\left(\frac{\prod_{k=1}^{N} s_{k_j}}{v_j^N}\right)^{\alpha-1}\right]^{\frac{\gamma}{\alpha-1}}$$

$$+ \frac{\left[\sum_{j=1}^{m} q_j\left(\frac{\prod_{k=1}^{N} s_{k_j}}{v_j^N}\right)^{\alpha-1}\right]^{\frac{\gamma}{\alpha-1}} -1}{N(2^{-\gamma}-1)}$$

$$= \frac{\left[\sum_{i=1}^{n} p_i\left(\frac{\prod_{k=1}^{N} r_{k_i}}{u_i^N}\right)^{\alpha-1}\right]^{\frac{\gamma}{\alpha-1}} \cdot \left[\sum_{j=1}^{m} q_j\left(\frac{\prod_{k=1}^{N} s_{k_j}}{v_j^N}\right)^{\alpha-1}\right]^{\frac{\gamma}{\alpha-1}} -1}{N(2^{-\gamma}-1)}$$

$$= \frac{\left[\sum_{i=1}^{n} \sum_{j=1}^{m} p_i q_j \left(\frac{\prod_{k=1}^{N} r_{k_i} s_{k_j}}{u_i^N v_j^N} \right)^{\alpha-1} \right]^{\frac{\gamma}{\alpha-1}} - 1}{N(2^{-\gamma}-1)}$$

$$= I(P*Q;U*V;R_1*S_1;\ldots;R_N*S_N)$$

BIBLIOGRAFIA

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

[2] - ACZÉL, J. (1966). Lectures on Functional Equations and Their Applications. Academic Press. New York.

[3] - ASH, R. (1965). Information Theory. Intercience Pub., New York.

[4] - DARÓCZY, Z. (1970). Generalized Information Functions. Information and Control (Vol. 11, página 36-51).

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

[6] - GALLAGER, R.G. (1968). Information Theory and Reliable Communication.- J. Wiley and Sons. Inc. New York

[7] - GUIASU. (1977). Information Theory with Applications. McGraw-Hill. New York. pág. 58-71 e 398-401.

[8] - HAVRDA, J. and F. CHARVÁT. (1967). Quantification Méthod of Classification Processes. Concept of Structural a-entropy -Kybernetika. Vol. 3. pág. 30-35.

[9] - KERRIDGE, D.F. (1961) - Inaccuracy and Inference. J. Royal Statist. Soc. vol. 23, pág. 184-194.

[10] - KULBACK, S. and R. LEIBLER. (1959) - On the Information and Sufficiency. Ann Math. Statist. vol. 22, pág. 72-86.

[11] - KULBACK, S. (1968) - Information Theory and Statistcs, Dover Publ. Inc. New York.

[12] - MATHAI, A.M. e P.N. RATHIE. (1975) - Basic Concepts in Information Theory and Statistics. Wiley Eastern Limited, New Delhi.

[13] - SHARMA, B.D. and D.P. MITTAL. (1975) - New Non-Additive Measure of Entropy for Discrete Probability Distributions. J. Math. Sciences. vol. 10, pág. 28-40.

[14] - SHANNON, C.E. (1948) - A Mathematical Theory of Communication - Bell System Tech. vol. 27, pag. 379-423 e 623-656.

[15] - TANEJA, I.J. (1975) - A Joint Characterization of Shannon's Entropy and Entropy of type β throught a Functional Equation. Journal of Mathematical Sciences. vol. 10, pág. 69-74.

[16] - TANEJA, I.J. (1979) - On the Information Improvement Due to 2 and N Revisions. Information Sciences. vol. 18, pág. 223-233.

|17| - TANEJA, I.J. and P.N. ARORA. (1980) - Characterization of Information Improvement Due to N revision and its Generalization by a Functional Equation. Information Sciences. vol. 20, pág. 127-136.

[18] - TANEJA, I.J. (1979) - Some Contributions to Information Theory. I(A Survey). On Measures of Information - Journal of Comb. Inform. and Syst. Sc. vol. 4, pág. 259-280.

[19] - THEIL, H. (1967) - Economics and Information Theory - North Holland Pub. Co. Amsterdan.

[20] - HARDY, G.H., J.E. LITTLEWOOD and G. PÓLYA. (1952) - Inequalities - Cambridge University Press.