

Sobre a Determinação da Curva Taxa-Distorção com Distância de Hamming

Paulo Roberto Rosa Lopes Nunes

Resumo—Neste trabalho a forma de calcular a curva taxa-distorção para o caso binário com distorção de Hamming é estendida para o caso N -ário com a mesma distorção. Não se trata de novo resultado, mas de um estudo de caso bastante interessante que ilustra a metodologia moderna no desenvolvimento e, principalmente, no ensino da Teoria da Informação. É um bom exercício de motivação para ser desenvolvido em sala de aula.

Palavras-Chave—teoria taxa-distorção, teoria da informação, ensino

Abstract—In this work, the usual way to determine the rate-distortion curve for the binary case using hamming distortion is extended to the N -ary case with the same distortion. No new results are presented, but a very interesting case study to illustrate modern methodology to develop and, especially to teach information theory. It's a nice motivation exercise to be developed in class room.

Keywords—rate-distortion theory, information theory, teaching

I. INTRODUÇÃO

Para duas variáveis discretas X e Y , definida uma distorção de símbolo simples $d(i, j)$ entre os valores que X e Y podem assumir, a menor taxa de bits que pode ser atingida quando se representa X por Y admitindo-se uma distorção média máxima D , é dada por

$$R(D) = \min_{\{P[Y=j|X=i]\} \in P_D} I(X; Y), \quad (1)$$

onde P_D é o conjunto de todas as distribuições de probabilidades de transição $\{P[Y=j|X=i]\}$ que produzem uma distorção média menor ou igual a D e $I(X; Y)$ é a informação mútua entre as variáveis X e Y .

A curva taxa-distorção ($R(D)$), como é conhecida a equação (1), em geral não é de fácil determinação. Para o caso particular em que X e Y são variáveis binárias e o critério de distorção é a distância de Hamming, o que resulta em distorção média igual à probabilidade de erro, a forma mais conhecida para determiná-la consiste em propor um valor mínimo para $I(X; Y)$ e mostrar que este mínimo pode ser atingido para todos os valores de probabilidade de erro de interesse (e.g. [1]).

O que se apresenta aqui é a generalização desde procedimento para o caso N -ário, mantendo-se o critério de máxima probabilidade de erro admitida.

Não se chega a nenhum novo resultado, apenas estende-se o caso binário explorando conceitos que podem ajudar estudantes a melhor absorver as ideias e metodologias utilizadas na Teoria da Informação, sendo este o objetivo deste trabalho.

Nas considerações finais, interessantes conclusões sobre a generalização do procedimento para o caso de distribuições com suporte infinito e outras distorções são apresentadas.

II. ESPECIFICAÇÃO DO PROBLEMA

Seja um conjunto de mensagens representadas pela variável aleatória discreta X , que pode assumir N diferentes valores. Sem perda de generalidade pode-se assumir que X assume valores no conjunto

$$\mathcal{X} = \{1, 2, 3 \dots N\}$$

com probabilidades $p_1, p_2 \dots p_N$.

Seja também uma variável aleatória Y , que pode assumir os mesmos valores de X , e uma medida de distorção entre os valores que X e Y podem assumir:

$$d_{ij} = d(X = i, Y = j) = \begin{cases} 0 & i = j \\ 1 & i \neq j \end{cases}$$

A distorção média por símbolo é dada por

$$\bar{d} = E[d_{XY}] = \sum_{i=1}^N \sum_{j=1}^N d_{ij} P[X = i, Y = j] = P[\text{erro}] = p_e.$$

Deseja-se expressar $R(D)$ dada por (1), em função de N e da distribuição de probabilidades de X .

O cálculo de um limiar inferior para esta curva pode ser encontrado em [2] como um caso particular do problema mais geral, utilizando-se o método dos coeficientes de Lagrange. Para o caso particular em que os valores de X são equiprováveis, a determinação da curva $R(D)$ pode ser realizada de forma simples utilizando-se a desigualdade de

Paulo Roberto Rosa Lopes Nunes, Secção de Engenharia Elétrica, Instituto Militar de Engenharia, Rio de Janeiro, Brasil, Email: prnunes@ime.eb.br.

Fano. O que se apresenta aqui é uma solução para qualquer distribuição de probabilidades de X .

III. SOLUÇÃO DO PROBLEMA

A informação mútua entre duas variáveis aleatórias X e Y é dada por

$$I(X;Y) = H(X) - H(X|Y)$$

Minimizar $I(X;Y)$, portanto, equivale a encontrar um máximo para $H(X|Y)$, uma vez que $H(X)$ independe das probabilidades condicionais $P[Y = j|X = i]$.

Antes de procurar um mínimo para $I(X;Y)$ como especificado em (1) é interessante maximizar $H(X|Y)$ com a restrição "*d exatamente igual a D*".

Para fazer isto, pode-se procurar um valor de entropia que com certeza é maior ou igual a $H(X|Y)$ quando $p_e = D$ e então mostrar que este valor de entropia pode ser atingido.

Quando $p_e = D$, a entropia de X dado Y não pode ser maior que a entropia de uma variável N -ária que indique qual o valor que X assumiu referenciado ao valor que Y assumiu, desde que a entropia desta variável N -ária seja máxima dada a restrição de probabilidade de erro. Pode-se então pensar em uma variável aleatória N -ária que assuma o valor zero se não houver erro e os valores $1, 2, \dots, N-1$ quando ocorrer erro. O valor que esta variável assume indica quanto deve se somar a Y , módulo N , para se obter X .

De forma precisa, pode-se definir esta variável N -ária como sendo a função invertível de X , na presença de Y :

$$f(X) = (X - Y) \bmod N.$$

Pode-se escrever:

$$\begin{aligned} H(X|Y) &= H((X - Y) \bmod N | Y) \leq H((X - Y) \bmod N) \\ &\leq H\left(1 - p_e, \frac{p_e}{N-1}, \dots, \frac{p_e}{N-1}\right) = H_{max} \\ &= H(p_e) + p_e \log_2(N-1) \end{aligned} \quad (2)$$

Nas expressões acima, a primeira igualdade decorre do teorema do processamento da informação, a primeira desigualdade é consequência do fato de condicionamento reduzir a entropia e a segunda desigualdade ocorre porque a distribuição em questão é a de máxima entropia quando um valor de probabilidade, no caso $1 - p_e$, é fixado. A expressão para H_{max} é facilmente determinada.

Como

$$H(X|Y) = \sum_{i=1}^N q_i H(X|Y = i),$$

onde $q_i = \text{Prob}[Y = i]$, para que $H(X|Y)$ seja igual a H_{max} é suficiente que:

$$H(X|Y = i) = H_{max} \quad i = 1, 2, \dots, N,$$

pois $q_i > 0$ e $\sum q_i = 1$.

A. Quando H_{max} pode ser atingido

Pelo Teorema de Bayes sabe-se que

$$P[X = i|Y = j]P[Y = j] = P[Y = j|X = i]P[X = i]$$

Dado então um conjunto de probabilidades de transição $X \rightarrow Y$ e uma distribuição de probabilidades para X , pode-se sempre determinar a distribuição de probabilidade de Y pelo teorema da probabilidade total, e as probabilidades de transição $Y \rightarrow X$ pelo Teorema de Bayes.

Porém, se forem dadas as probabilidades de transição de $Y \rightarrow X$ e a distribuição de X , não se pode garantir que existe uma distribuição de probabilidades de Y que gere a dada distribuição de probabilidades de X , mas, se existir, pode-se determinar $H(X|Y)$ dadas as probabilidades de transição $Y \rightarrow X$ e a distribuição de X .

Considerando-se as probabilidades de transição de Y para X , para que $H(X|Y)$ seja máxima dada a probabilidade de erro p_e é necessário que as probabilidades de transição sejam como na figura 1. Nessa figura, no lado esquerdo estão representados os valores que a variável aleatória Y pode assumir e no lado direito os valores que a variável aleatória X pode assumir. Cada seta que conecta um valor de Y , seja j , com um valor de X , seja i , representa a probabilidade de transição $P[X = i|Y = j]$. Apenas algumas transições estão representadas.

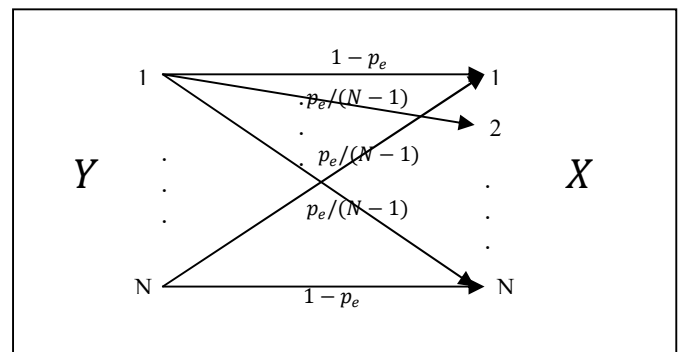


Fig 1- Probabilidades de transição de Y para X

Seja Y com distribuição de probabilidade $\{q_1, q_2, \dots, q_N\}$ e X com distribuição de probabilidade $\{p_1, p_2, \dots, p_N\}$. Pode-se escrever, observando a figura 1, que:

$$p_i = q_i(1 - p_e) + \sum_{j \neq i} \frac{q_j p_e}{N-1} = q_i(1 - p_e) + \frac{p_e(1 - q_i)}{N-1} \quad (3)$$

ou

$$q_i = \frac{p_i(N-1) - p_e}{N(1 - p_e) - 1} = \frac{p_i N - p_i - p_e}{N - N p_e - 1} \quad (4)$$

Procura-se uma distribuição de Y que gere a distribuição estabelecida para X através do canal representado na fig.1. Seja

$$I_{min} = H(X) - H(p_e) - p_e \log_2(N-1)$$

Se existir tal distribuição, pode-se determinar a informação mútua mínima entre X e Y :

$$I(Y;X)_{\min} = I_{\min},$$

se não existir, têm-se apenas um limitante inferior para a informação mútua entre X e Y , podendo-se apenas dizer que

$$I(Y;X)_{\min} > I_{\min}$$

É fácil verificar que o somatório dos q_i 's é unitário. É preciso determinar então para quais valores de p_e todos os valores de q_i serão maiores que zero, o que garante que serão também menores que 1.

O denominador de (4) é igual a zero quando $(1 - p_e) = 1/N$ ou $p_e = 1 - 1/N = (N - 1)/N$. Neste ponto específico o numerador de (4) é nulo apenas para o caso em que $p_i = 1/N$.

Para $p_e < (N - 1)/N$, o denominador de (4) é positivo, e para que q_i seja positivo é necessário que

$$(N - 1)p_i > p_e$$

Já se $p_e > (N - 1)/N$, o numerador de (4) precisa ser negativo para que exista q_i positivo, ou seja,

$$(N - 1)p_i < p_e$$

Considerando-se que $p_i > 0$ para todo i , conclui-se que:

$$H(X|Y) = H_{\max} \quad \text{se } p_e < (N - 1) \min_i(p_i)$$

e

$$H(X|Y) = H_{\max} \quad \text{se } (N - 1) \max_i(p_i) < p_e < 1$$

Se $(N - 1) \min_i(p_i) \leq p_e \leq \min[(N - 1) \max_i(p_i), 1]$, pode-se apenas afirmar que

$$H(X|Y) < H_{\max}$$

Não é difícil verificar que quando $p_e = (N - 1)/N$, $H_{\max}$ atinge o valor máximo $\log_2 N$, que é também o valor máximo que $H(X)$ pode alcançar. Como $\min_i(p_i) \leq \frac{1}{N} \leq \max_i(p_i)$, o ponto onde $H_{\max}$ é máximo está no intervalo onde se tem apenas um limitante para $H(X|Y)$, podendo-se então escrever que se $(N - 1) \min_i(p_i) \leq p_e \leq \min[(N - 1) \max_i(p_i), 1]$

$$H(X|Y) < H_{\max} \leq \log_2 N$$

sem esquecer que

$$H(X|Y) < H(X).$$

Conclui-se então que $I(X;Y)_{\min}$, para valores de p_e menores que $(N - 1) \min_i(p_i)$ e maiores que $(N - 1) \max_i(p_i)$, se este último valor for menor que 1, é dada por:

$$I(X;Y)_{\min} = H(X) - H(p_e) - p_e \log_2(N - 1)$$

A relação entre a curva R - D dada por (1) e $I(X;Y)_{\min}$ é de fácil determinação, visto que a curva R - D não pode ser crescente e $I(X;Y)_{\min}$ claramente é crescente se $p_e > (N - 1) \max_i(p_i)$. Portanto, para a determinação da curva R - D interessa apenas o primeiro intervalo, onde $I(X;Y)_{\min}$ é decrescente, podendo-se dizer que

$$R(D) = H(X) - H(D) - D \log_2(N - 1)$$

se

$$0 \leq D < (N - 1) \min_i(p_i)$$

Admitindo-se uma probabilidade de erro maior ou igual a $1 - \max_i(p_i)$, atinge-se a taxa zero fazendo Y igual ao valor de i com maior probabilidade de ocorrência, podendo-se afirmar que

$$R(D) = 0$$

se

$$D \geq 1 - \max_i(p_i)$$

No intervalo em que apenas um limitante inferior pode ser especificado para $I(X;Y)_{\min}$, pode-se afirmar, levando-se em conta que a informação mútua não pode ser negativa, que:

$$R(D) > \max[0, H(X) - H(D) - D \log_2(N - 1)]$$

se

$$(N - 1) \min_i(p_i) \leq D < 1 - \max_i(p_i)$$

Como neste intervalo $I_{\min}$ é um limitante inferior para $R(D)$ e como $R(D) = 0$ no limite superior do intervalo, conclui-se que $I_{\min}$ atinge o eixo horizontal neste intervalo, podendo-se apenas dizer que $R(D) > 0$ entre este ponto e o limite superior do intervalo.

B. Particularização para o caso binário

Neste caso $N = 2$ e $p_2 = 1 - p_1$. Chega-se a:

$$R(D) = \begin{cases} H(p_1) - H(D) & 0 < D < \min_i(p_i) \\ 0 & 1 - \max_i(p_i) = \min_i(p_i) \leq D \leq 1 \end{cases}$$

Esta curva está representada na figura 2 para $p_2 = p_1 = \frac{1}{2}$.

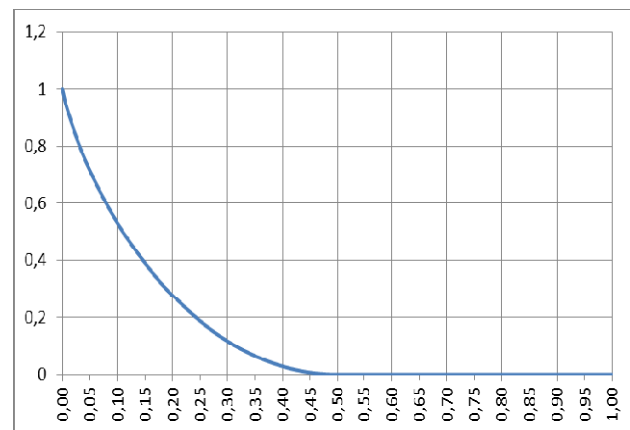


Fig 2 – Curva $R(D)$ para o caso binário equiprovável

C. Particularização para o caso $p_i = 1/N$

Neste caso, $I(Y; X)_{\min} = I_{\min}$ para todos os valores de p_e , e:

$$R(D) = \begin{cases} \log_2(N) - H(D) - D \log_2(N-1) & 0 \leq D < \frac{N-1}{N} \\ 0 & \frac{N-1}{N} \leq D \leq 1 \end{cases}$$

Na figura 3 esta curva está exemplificada para $N = 8$.

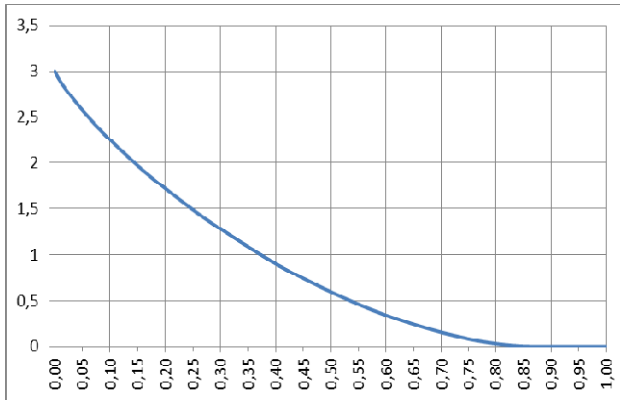


Fig 3- Curva $R(D)$ para o caso $N = 8$, idênticas probabilidades.

D. Caso geral

Como já visto:

$$R(D) = \begin{cases} H(X) - H(D) - D \log_2(N-1) & 0 \leq D < (N-1) \min_i(p_i) \\ 0 & 1 - \max_i(p_i) \leq D \leq 1 \end{cases}$$

$$R(D) > \max[0, H(X) - H(D) - D \log_2(N-1)]$$

$$\text{se } (N-1) \min_i(p_i) \leq D < 1 - \max_i(p_i)$$

Esta curva está exemplificada na figura 4, para $N = 5$ e a seguinte distribuição de probabilidades de X :

$$\{0,4 \ 0,3 \ 0,125 \ 0,1 \ 0,075\}$$

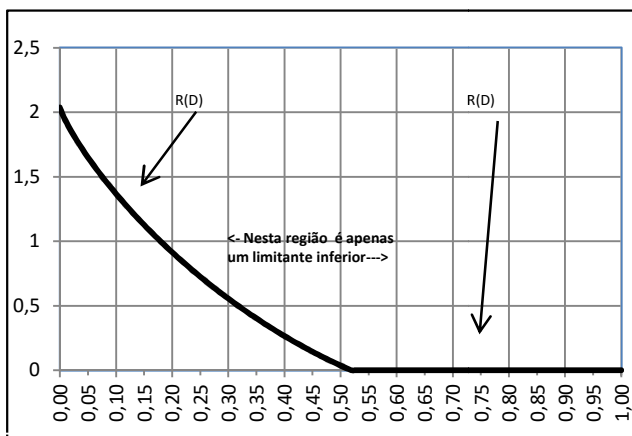


Fig 4- Caso geral. $N = 5$ (vide texto)

IV. CONSIDERAÇÕES FINAIS

Cabem alguns comentários sobre os três casos considerados:

O caso binário e o caso uniforme permitem a especificação da curva $R(D)$ para todos os valores de D . No primeiro caso, apesar do valor de $H_{\max}$ não ser, em geral, atingível para todos os valores de D , na região onde tal não ocorre a solução $R(D) = 0$ é clara. Já no caso equiprovável, $H_{\max}$, e consequentemente $I_{\min}$, é atingível para todos os valores de D , sendo a curva bem determinada para qualquer valor de N . Somente para distribuições uniformes $H_{\max}$ é atingível para qualquer valor de probabilidade de erro.

É interessante considerar o caso em que N é muito grande e a distribuição de probabilidades é uniforme. Nesta situação pode-se assumir que $\log_2(N) \gg H(D)$ e que $N-1 \cong N$, o que permite escrever que, quando $\log_2(N) \gg 1$:

$$R(D) \approx (1-D) \log_2 N \quad 0 \leq D < 1. \quad (5)$$

Esta expressão é, na verdade, um limitante superior trivial para a curva $R(D)$ de qualquer distribuição N -ária uniforme, e, como se pode ver, assintoticamente atingível.

Um aspecto interessante da expressão (5) é que uma representação com o valor mínimo de taxa especificado por esta curva é facilmente obtida se uma sequência de M realizações independentes e igualmente prováveis de X for simplesmente truncada em $(1-D)M$ elementos e armazenada. Os elementos restantes não armazenados são substituídos por estimativas. A distorção média entre as duas sequências é, no máximo, igual a D , e o número médio de bits necessário para armazenar cada elemento da representação, ou seja, a entropia média por elemento da sequência original, é igual a $\frac{(1-D)M \log_2 N}{M}$ bits. Supondo que N é suficientemente grande para que (5) valha, pode-se dizer que a curva $R(D)$ foi alcançada.

No caso geral, o valor $(N-1) \min_i(p_i)$ é importante na medida em que demarca a região onde a curva $R(D)$ é determinada. Uma tendência natural é tentar estender os resultados obtidos para o caso de distribuições com entropia finita sobre infinitos valores, ou seja, para o caso em que $N \rightarrow \infty$, $p_i > 0 \forall i$ e $H(X) \leq K < \infty$. Pode-se mostrar que neste caso $(N-1) \min_i(p_i)$ tende a zero quando $N \rightarrow \infty$, obtendo-se um limitante inferior trivial (zero) para a curva $R(D)$.

Para distribuições discretas de probabilidade $\{p_i\}$, $p_i > 0$, $i \in \{-\infty, \dots, -1, 0, 1, \dots, +\infty\}$, pode-se estender o procedimento considerando outras distorções, como módulo do erro ou erro quadrático. Se a distorção for função de

$(X - Y)$, identificada a distribuição de máxima entropia de $(X - Y)$ dada a distorção média admissível, a expressão similar à equação (3) pode ser expressa como uma equação de convolução discreta: $p_i = q_i * r_i$, onde r_i é o valor de $P[(X - Y) = i]$ na situação de máxima entropia. A identificação desta convolução está ilustrada na figura 5. Observe que (3) pode ser vista como uma equação de convolução discreta cíclica. Técnicas de transformada podem, então, ser aplicadas para a determinação da curva $R-D$ de distribuições discretas com suporte infinito com distorções que sejam função de $(X-Y)$.

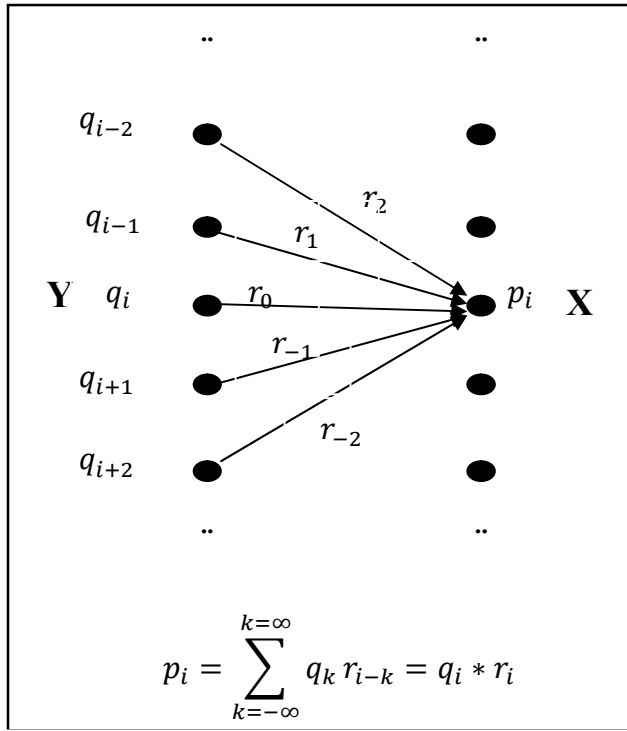


Fig 5 - $\{p_i\}$ pode ser determinada por uma convolução

Estendendo o conceito de curva $R-D$ para variáveis contínuas com suporte infinito, o procedimento leva a uma equação integral de convolução: $p_X(x) = q_Y(x) * r_{X-Y}(x)$. Assumindo que $p_X(x)$ é uma densidade Gaussiana com variância σ^2 e que a distorção é o erro quadrático, e considerando que a densidade de máxima entropia neste caso é uma Gaussiana com variância D , pode-se determinar facilmente a curva $R - D$ para o intervalo $0 \leq D \leq \sigma^2$; utilizando-se a Transformada de Fourier.

VI. REFERÊNCIAS

- [1] Cover, T.M., Thomas, J.A. - Elements of Information Theory (2nd ed.), Wiley-Interscience, 2006
- [2] Gallager, R.G. - Information Theory and Reliable Communication, John Wiley and Sons, 1968