1 LIMITES FUNDAMENTAIS NA TEORIA DA INFORMAÇÃO

O trabalho de Shannon sobre a teoria da informação, publicado em 1948, estabelece a base teórica para que sistemas de comunicações sejam eficientes e confiáveis.

Os principais tópicos abordados nessas notas de aulas são:

- A *Entropia* como uma medida básica de informação.
- O Teorema da *Codificação de Fonte* e algoritmos de compactação de dados.
- *Informação Mútua* e sua relação com a *capacidade do canal de informação* para a transmissão de informação.
- Teorema da *Codificação de Canal* como base para comunicações confiáveis.
- Teorema da *Capacidade de Informação* como a base para a solução de compromisso entre a largura de faixa do canal e a relação sinal-ruído.

1.1. INTRODUÇÃO

No contexto das comunicações, a teoria da informação fornece uma modelagem matemática que permite responder duas questões fundamentais:

1. Qual é a complexidade irredutível abaixo da qual um sinal não pode ser comprimido?
2. Qual é a máxima taxa de transmissão para uma comunicação confiável em um canal ruidoso?

As respostas para as duas questões acima definem bem as duas principais vertentes da teoria da informação: *codificação de fonte e codificação de canal*.

1.2. INCERTEZA, INFORMAÇÃO E ENTROPIA

Suponha que um *experimento probabilístico* envolva a observação da saída de uma fonte de eventos discretos em unidades de intervalo de tempo. Estes eventos podem ser modelados como variáveis discretas aleatórias (s_k) que fazem parte de um conjunto ou *alfabeto* (**S**). Assim, $\mathbf{S} = \{s_0, s_1, \ldots, s_{K-1}\}$ com probabilidades $P(\mathbf{S} = s_k) = p_k$, para $k = 0, 1, \ldots, K-1$, que satisfaz a igualdade

$$\sum_{k=0}^{K-1} p_k = 1 .$$

Além disso, admita que os símbolos emitidos pela fonte sejam estatisticamente independentes. Uma fonte com tais propriedades é definida como *fonte discreta sem memória.*

O termo *sem memória* tem o sentido de que o símbolo emitido a qualquer tempo é independente de uma escolha prévia.

A quantidade de informação produzida pela fonte está associada à *incerteza ou surpresa*. Se não há surpresa não há informação. A quantidade de informação, $I(s_k)$, obtida por um evento, s_k, é definida como:

$$I(s_k) = \log_2\left(\frac{1}{p_k}\right) \tag{1.1}$$

NOTA DO PROFESSOR

Conforme mencionado no livro texto, a base do logaritmo é arbitrária. Entretanto, uma vez que os sistemas de comunicações digitais operam com base binária, a base do logaritmo é 2. Como consequência, cada símbolo de uma fonte discreta sem memória apresenta, neste caso, uma quantidade de informação em número de bits, que depende exclusivamente da probabilidade de ocorrência de cada símbolo.

A quantidade de informação apresenta as seguintes propriedades:

1. $I(s_k) = 0$ *para* $p_k = 1$
2. $I(s_k) \geq 0$ *para* $0 \leq p_k \leq 1$
3. $I(s_k) > I(s_l)$ *então* $p_k < p_l$
4. $I(s_k s_l) = I(s_k) + I(s_l)$

A quantidade de informação produzida por uma fonte durante um intervalo de tempo arbitrário depende do conjunto de símbolos emitidos pela fonte. A média da quantidade de informação $I(s_k)$ de uma fonte de *alfabeto* (**S**) é dada por

$$H(\mathbf{S}) = E[I(s_k)]$$
$$H(\mathbf{S}) = \sum_{k=0}^{K-1} p_k \log_2\left(\frac{1}{p_k}\right). \tag{1.2}$$

$H(\mathbf{S})$ que é definida como *entropia*. A entropia determina a quantidade média de informação por símbolo (evento) da fonte.

❑ PROPRIEDADES DA ENTROPIA

Limitantes:

$$0 \le H(\mathbf{S}) \le \log_2 K \tag{1.3}$$

Consequências:

1. $H(\mathbf{S}) = 0$, se e somente se a probabilidade $p_k = 1$ para um dado valor de k e todas as outras probabilidades são iguais a zero. Neste caso não há incerteza.
2. $H(\mathbf{S}) = \log_2 K$, se e somente se $p_k = 1/K$, ou seja, todos os símbolos são equiprováveis. Neste caso a incerteza é máxima.

Nos canais de comunicação digital, uma fonte de grande interesse é a fonte binária sem memória. Para essa fonte é interessante o entendimento do comportamento da entropia em função da probabilidade de ocorrência dos eventos "0" e "1". Esse comportamento é mostrado no Exemplo 1.1.

EXEMPLO 1.1

Considere uma fonte discreta sem memória que emite os símbolos $s_0 = 0$ e $s_1 = 1$, com probabilidades p_0 e p_1, respectivamente. A entropia desta fonte é

$$H(\mathbf{S}) = \sum_{k=0}^{K-1} p_k \log_2\left(\frac{1}{p_k}\right) = p_0 \log_2\left(\frac{1}{p_0}\right) + p_1 \log_2\left(\frac{1}{p_1}\right)$$

Mas $p_1 = 1 - p_0$.

$$H(\mathbf{S}) = p_0 \log_2\left(\frac{1}{p_0}\right) + (1 - p_0)\log_2\left(\frac{1}{1 - p_0}\right)$$

Consequentemente, a entropia da fonte torna-se

$$\mathrm{H}(p_0) = -p_0 \log_2 p_0 - (1 - p_0)\log_2(1 - p_0) \tag{1.4}$$

Onde, $\mathbf{H}(p_0)$ é chamada de *função entropia* e

1. Quando $p_0 = 0$, $H(\mathbf{S}) = 0$.
2. Quando $p_1 = 1$, $H(\mathbf{S}) = 0$.
3. A entropia é máxima quando $p_1 = p_0 = ½$.

A figura apresentada a seguir apresenta a função entropia em função de p_0.

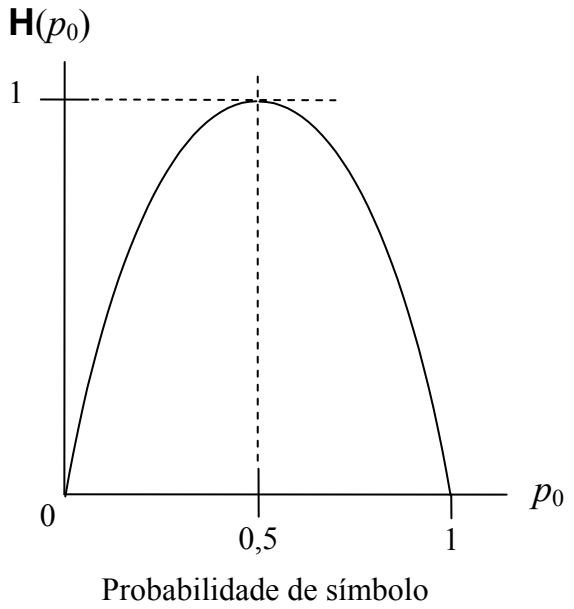


Figura 1.1 - Função entropia de uma fonte discreta sem memória.

1.3. TEOREMA DA CODIFICAÇÃO DE FONTE

A codificação de fonte é o processo pelo qual os dados de uma fonte discreta são representados de forma a permitir uma transmissão *eficiente*. Probabilidades diferentes para a ocorrência dos símbolos de uma fonte discreta podem ser exploradas para uma codificação de fonte com um *código de comprimento variável*. Entretanto, códigos de comprimento variável exigem o conhecimento prévio da estatística da fonte. Um exemplo clássico de código de comprimento variável é o Código Morse, que faz uso da estatística da fonte. Na construção do código, Morse observou que na língua inglesa a letra *e* é a letra que aparece com maior frequência e no Código Morse ela é representada por um "·". Por outro lado, a letra *q* é a letra que ocorre com menos frequência e é representada pela sequência "- · - -".

No caso específico dos sistemas de comunicações digitais é conveniente que dois requisitos funcionais sejam satisfeitos:

1. As palavras códigos devem estar na forma binária.
2. As palavras códigos devem ser inequivocamente decodificáveis.

O comprimento médio das palavras códigos de uma codificação de fonte é determinado por

$$\overline{L} = \sum_{k=0}^{K-1} p_k l_k \tag{1.5}$$

onde l_k é o número de bits da palavra código correspondente ao símbolo k.

PRIMEIRO TEOREMA DE SHANNON

TEOREMA DA CODIFICAÇÃO DE FONTE

Dada uma fonte discreta sem memória de entropia $H(\mathbf{S})$, o comprimento médio das palavras códigos para um esquema de codificação livre de distorção é limitado como

$$\overline{L} \geq H(\mathbf{S}). \tag{1.6}$$

De acordo com o Primeiro Teorema de Shannon a entropia representa um limite fundamental do número médio de bits por símbolo necessários para representar uma fonte discreta sem memória. Em outras palavras, um esquema de codificação de fonte pode ser feito de modo que o comprimento médio das palavras códigos seja tão pequeno quanto à entropia, mas nunca menor do que ela. Assim, eficiência de codificação pode ser definida pela relação

$$\eta = \frac{H(\mathbf{S})}{\overline{L}}. \tag{1.7}$$

Uma vez que o Teorema da Codificação de Fonte admite que $\overline{L} = H(\mathbf{S})$, então o maior valor possível para a eficiência de codificação de fonte é igual a unidade.

1.4. COMPACTAÇÃO DE DADOS

Para uma transmissão eficiente as informações redundantes devem ser removidas do sinal que será transmitido. Neste texto os termos compactação e compressão apresentam significados distintos. Aqui, o termo *compactação* está associado a um processo onde não há perda de informação, enquanto o termo *compressão* é usado nos processos em que se admite perda de informação. Em outras palavras, a *compactação* o mesmo que uma *compressão sem perdas*. Na compactação de dados, o processo de codificação de fonte é limitado, necessariamente, pela entropia da fonte.

❑ CÓDIGOS PREFIXOS

Um código prefixo é um código de fonte que apresenta decodificação inequívoca, ou seja, sua decodificação não resulta em ambiguidade. Um código prefixo é aquele que nenhuma palavra código é prefixo para qualquer outra palavra código.

DEFINIÇÃO: Um prefixo, com i bits, é a parte inicial de qualquer palavra código com n bits, sendo $i \leq n$.

Para ilustrar o conceito, observe a tabela a seguir.

Tabela 1.1 - Exemplos de códigos de fonte.

Símbolo	Probabilidade	Código I	Código II	Código III
S_0	0,5	0	0	0
S_1	0,25	1	10	01
S_2	0,125	00	110	011
S_3	0,125	11	111	0111

A partir do apresentado na Tabela 1.1 verifica-se que:

1. O Código I não é um código prefixo.
2. O Código II é um código prefixo.
3. O Código III não é um código prefixo.

A decodificação de um código prefixo pode ser feita de acordo com uma árvore de decisão. A árvore de decisão pode ser construída de acordo com as seguintes regras:

1. A partir de um estado inicial criam-se dois ramos. Para um ramo é atribuído o bit "0" e para o outro um bit "1".
2. Se houver uma palavra código com apenas um bit, o ramo correspondente à palavra código deve ser terminado com o símbolo correspondente.
3. O(s) ramo(s) não terminado em símbolo deve ser bifurcado em novos dois ramos. Para um ramo é atribuído o bit "0" e para o outro um bit "1".
4. Se houver uma palavra código com dois bits, o ramo correspondente à palavra código deve ser terminado com o símbolo correspondente.
5. O(s) ramo(s) não terminado em símbolo deve ser bifurcado em novos dois ramos. Para um ramo é atribuído o bit "0" e para o outro um bit "1".
6. Este procedimento se repete até que todos os ramos terminem em símbolos.

Para o Código II, a árvore de decisão está apresentada na Figura 1.2. Note que sua decodificação é inequívoca e *instantânea*.

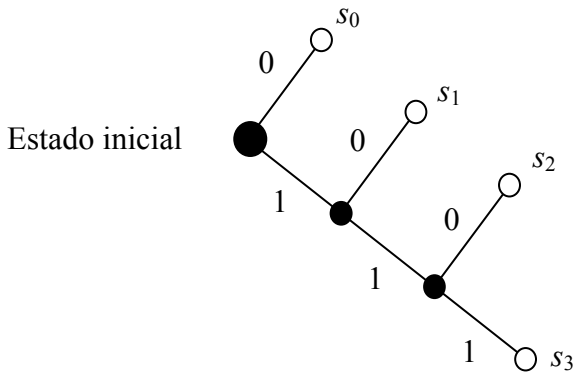


Figura 1.2 - Árvore de decodificação para o Código II.

A sequência 1011111000... é inequivocamente decodificável e corresponde correspondente à sequência de símbolos $s_1\, s_3\, s_2\, s_0\, s_0$.

Uma condição necessária, mas não suficiente, para que um código possa ser um código prefixo é atender a desigualdade de *Kraft-McMillan,* definida como:

$$\sum_{k=0}^{K-1} 2^{-l_k} \leq 1 \tag{1.8}$$

Assim verifica-se que:

- Para o Código I:

$$\sum_{k=0}^{K-1} 2^{-l_k} = 2^{-1} + 2^{-1} + 2^{-2} + 2^{-2} = 1{,}5$$

- Para o Código II:

$$\sum_{k=0}^{K-1} 2^{-l_k} = 2^{-1} + 2^{-2} + 2^{-3} + 2^{-3} = 1$$

- Para o Código III:

$$\sum_{k=0}^{K-1} 2^{-l_k} = 2^{-1} + 2^{-2} + 2^{-3} + 2^{-4} = 0{,}9375$$

Observações:

1. O Código I viola a desigualdade de *Kraft-McMillan.*
2. A desigualdade de *Kraft-McMillan* foi satisfeita pelos Códigos II e III, entretanto, apenas o Código II é um código prefixo.
3. O Código III é inequivocamente decodificável apesar de não ser um código prefixo, ou seja, um código inequivocamente decodificável não precisa ser, necessariamente um código prefixo, entretanto, sua decodificação não é *instantânea*.

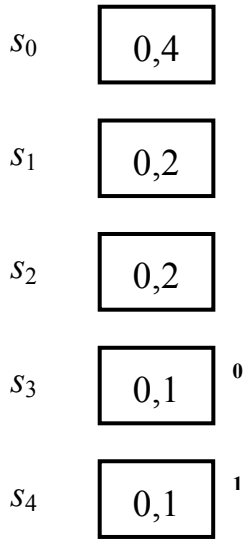
❑ ALGORITMO DE HUFFMAN

O procedimento de Huffman para a criação de um código prefixo consiste de um pequeno conjunto de regras. Essas regras são apresentadas no exemplo apresentado a seguir.

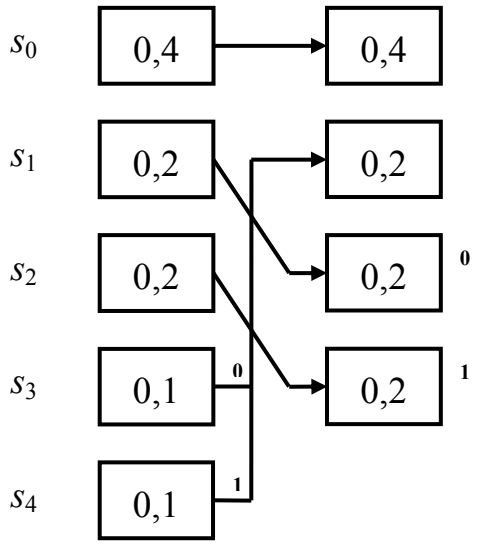
EXEMPLO 1.2

Suponha uma fonte discreta sem memória $\mathbf{S} = \{s_0, s_1, s_2, s_3, s_4\}$ com probabilidades $P(\mathbf{S} = s_k) = \{0{,}4;\ 0{,}2;\ 0{,}2;\ 0{,}1;\ 0{,}1\}$. Um código prefixo pode ser obtido de acordo com o conjunto de regras apresentadas a seguir (algoritmo de Huffman).

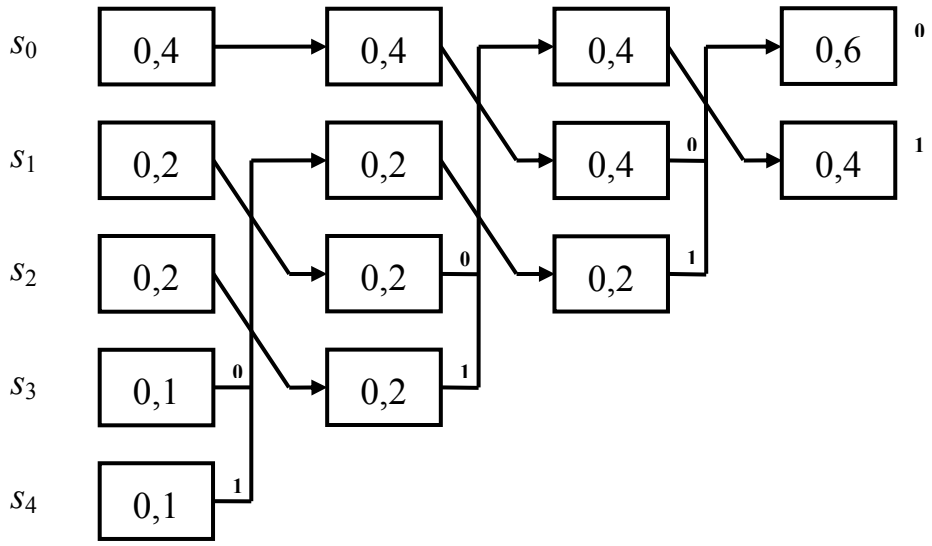
1. Os símbolos da fonte são listados em ordem decrescente de probabilidade. Aos dois símbolos de menor probabilidade, localizados no pé da coluna, são atribuídos os 0 e 1.



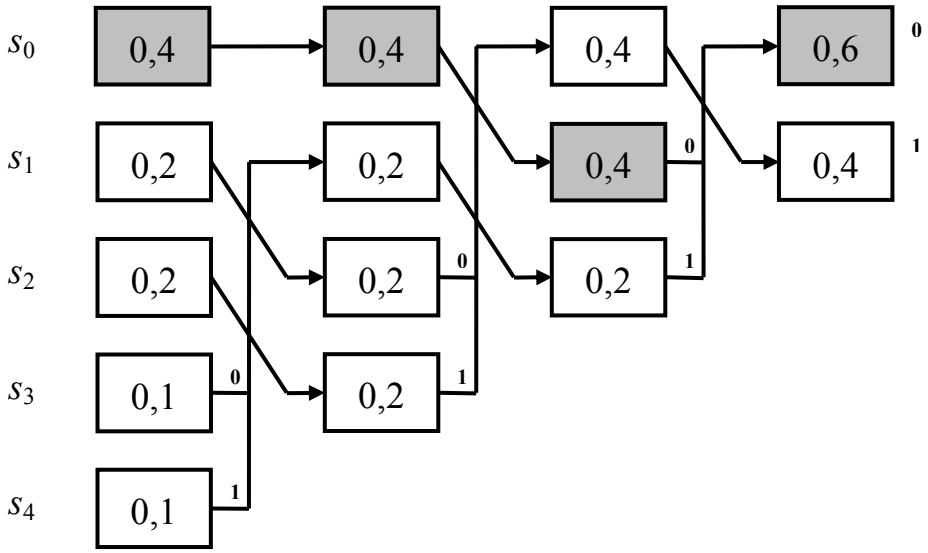
2. Os dois símbolos de menor probabilidade são agrupados em um único símbolo, cuja probabilidade é a soma das probabilidades. As probabilidades são novamente listadas em ordem decrescente de probabilidade em uma segunda coluna. Por um motivo que será explicado posteriormente, na reordenação dos símbolos por ordem decrescente de probabilidade o novo símbolo, resultado do agrupamento, foi colocado na posição mais alta possível, sem violar a ordem decrescente de probabilidade.



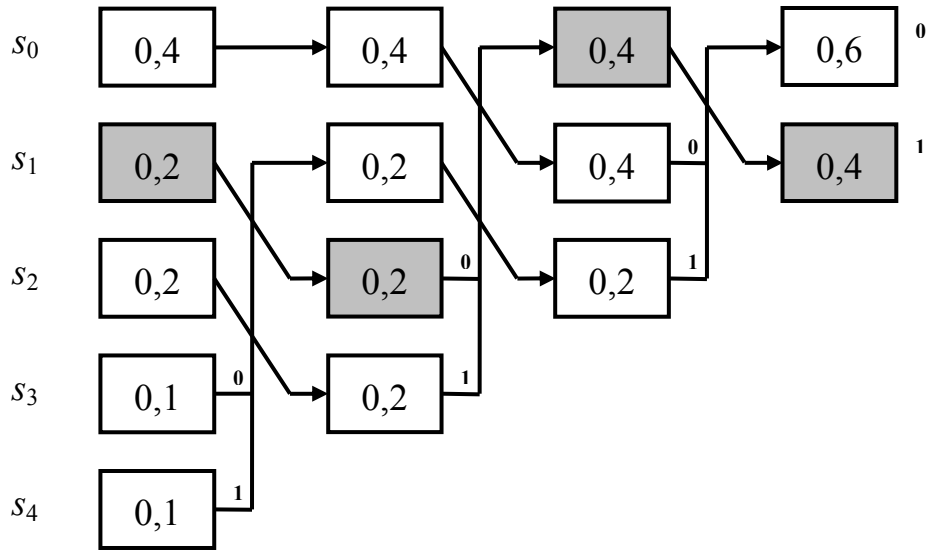
3. O procedimento é repetido até que restem apenas dois símbolos, aos quais são atribuídos novamente os bits 0 e 1.



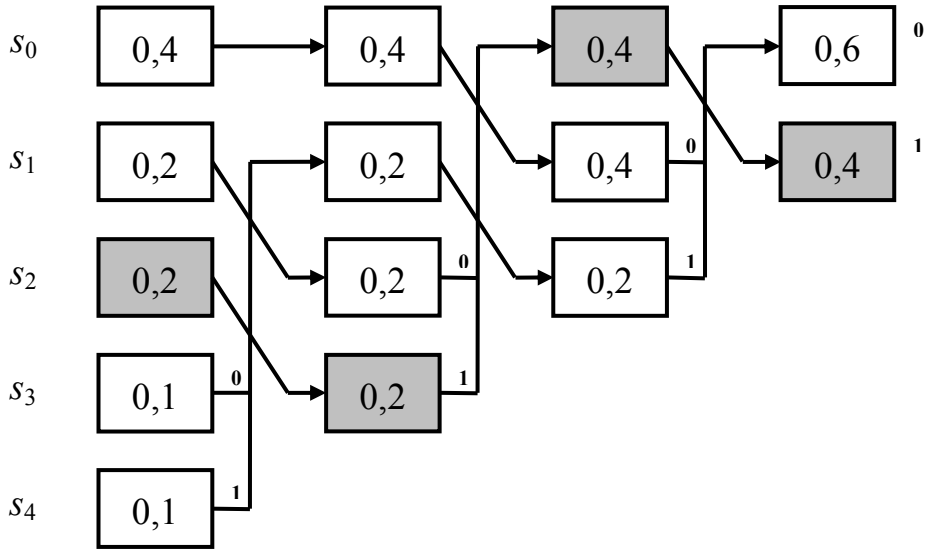
4. As palavras códigos binárias são obtidas fazendo o caminho inverso, a partir da última coluna, em direção a primeira coluna, anotando-se os bits encontrados em cada percurso, até cada símbolo listado na primeira coluna, conforme mostrado nas figuras a seguir.



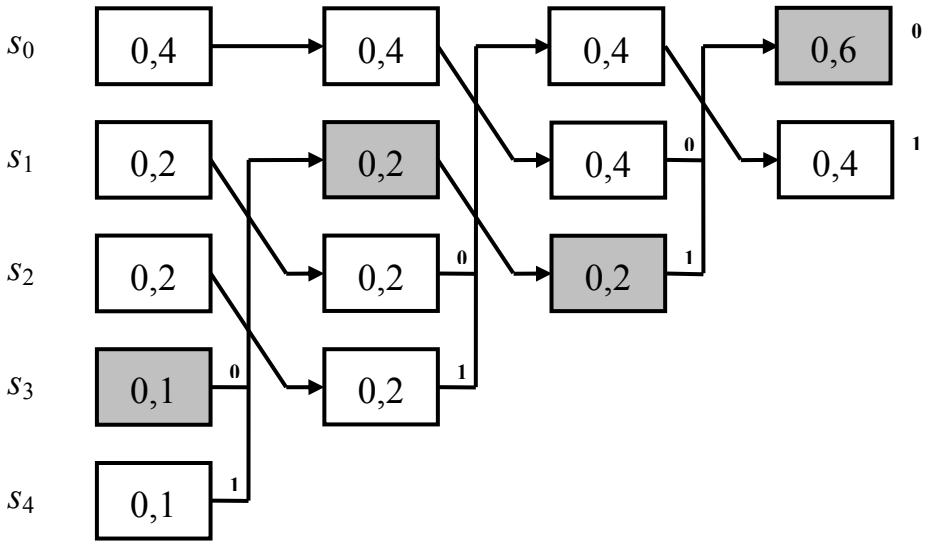
$s_0 = 00$



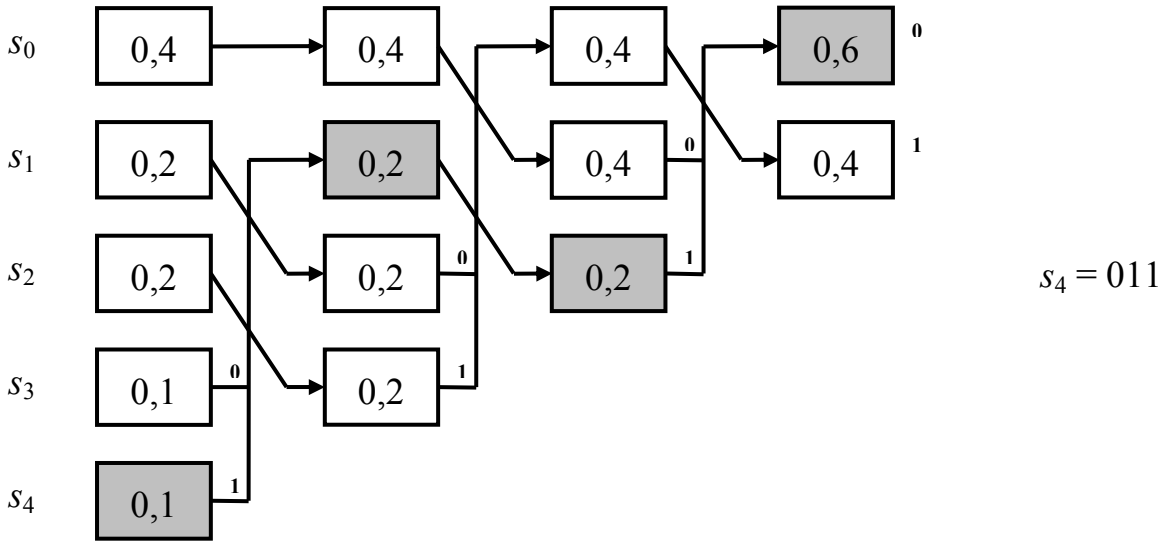
$s_1 = 10$



$s_2 = 11$



$s_3 = 010$



O resultado da codificação está apresentado na Tabela 1.2.

Tabela 1.2 - Palavras códigos obtidas no Exemplo 1.2.

s_0	00
s_1	10
s_2	11
s_3	010
s_4	011

O comprimento médio é das palavras códigos obtidas neste exemplo é:

$$\overline{L} = \sum_{k=0}^{K-1} p_k l_k = 0{,}4(2) + 0{,}2(2) + 0{,}2(2) + 0{,}1(3) + 0{,}1(3)$$

$$\overline{L} = 2{,}2 \text{ bits/símbolo}$$

Sendo a entropia determinada por:

$$H(\mathrm{S}) = \sum_{k=0}^{K-1} p_k \log_2\left(\frac{1}{p_k}\right) = 0{,}4\log_2\left(\frac{1}{0{,}4}\right) + 0{,}2\log_2\left(\frac{1}{0{,}2}\right) + 0{,}2\log_2\left(\frac{1}{0{,}2}\right) + 0{,}1\log_2\left(\frac{1}{0{,}1}\right) + 0{,}1\log_2\left(\frac{1}{0{,}1}\right)$$

$$H(\mathrm{S}) = 2{,}12193 \text{ bits/símbolo}$$

A eficiência obtida nesta codificação é:

$$\eta = \frac{H(\mathbf{S})}{\bar{L}} = \frac{2{,}12193}{2{,}2} \quad \Rightarrow \quad \eta = 0{,}9645$$

A árvore de decisão para o código do Exemplo 1.2 está mostrada na Figura 1.3.

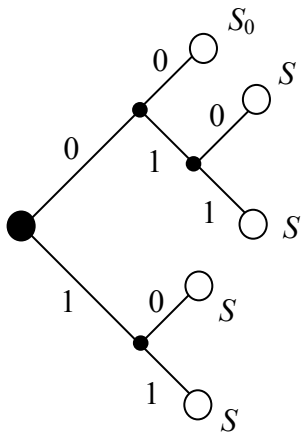


Figura 1.3 - Árvore de decodificação para o código do Exemplo 1.2.

* * *

O processo de codificação de Huffman não é único. Pelo menos duas variações no processo de codificação são possíveis.

1. No agrupamento dos símbolos, o rótulo 0 e 1 para cada símbolo é arbitrário, poderia ser 1 e 0. Entretanto, a mesma convenção deve ser mantida do início ao fim.
2. A colocação de símbolos combinados com mesma probabilidade que outro(s), na posição mais alta da coluna, também é arbitrário, poderia não ser na posição mais alta.

Qualquer que seja a forma arbitrada para a codificação, o comprimento médio do código sempre será o mesmo, entretanto, a colocação de símbolos combinados com mesma probabilidade que outro(s), na posição mais alta da coluna, resulta em uma menor variância do comprimento das palavras códigos.

A variância do comprimento das palavras códigos, σ^2, é determinada por

$$\sigma^2 = \sum_{k=0}^{K-1} p_k \left(l_k - \bar{L}\right)^2 . \tag{1.9}$$

NOTA DO PROFESSOR

Suponha a sequência 00010110111000. Verifica-se através da árvore de decisão apresentada no Exemplo 1.3 que a sequência é inequivocamente decodificável como sendo S_0; S_3; S_2; S_4; S_1 e S_0.

Conclusão: Este exemplo mostra que o algoritmo de Huffman produz um código prefixo com comprimento médio próximo da entropia e inequivocamente decodificável.

Conforme mostrado pelo Primeiro Teorema de Shannon é possível obter um código cujo comprimento médio das palavras códigos seja igual a entropia, ou seja,

$$\overline{L} = H(\mathsf{S}). \tag{1.10}$$

Isso ocorre quando a probabilidade de um evento é igual ao inverso da base numérica da codificação elevado ao comprimento da palavra código correspondente a este evento. Em outras palavras,

$$p_k = 2^{-l_k}. \tag{1.11}$$

Entretanto, isso é uma casualidade uma vez que é necessário que a correspondência estabelecida por (1.11) ocorra para todas as probabilidades e suas respectivas palavras códigos. Para esta condição particular, diz-se que o código prefixo está *casado* com a fonte. Note que o Código II é um código que atende esta condição, ou seja, é um código prefixo casado com a fonte.

Quando um código não é casado com a fonte é possível obter um comprimento médio das palavras códigos tão próximo da entropia quanto desejado por meio de um código prefixo estendido. Para isso é necessário estender a fonte original **S** para uma fonte $\mathbf{S}^n$, ou seja, o comprimento médio das palavras códigos aproxima-se da entropia na medida em que o valor de n é aumentado na fonte estendida. Como o número de símbolos de uma fonte estendida $\mathbf{S}^n$ é K^n, a diminuição do comprimento médio do código tem como preço um aumento da complexidade de decodificação.

❑ EXTENSÃO DE UMA FONTE DISCRETA SEM MEMÓRIA

Considere a fonte discreta sem memória $\mathbf{S} = \{s_0, s_1, \ldots, s_{K-1}\}$ com um número de símbolos $K = 3$ e com probabilidades $P(\mathbf{S} = s_k) = p_k$, para $k = 0, 1, \ldots, K-1$, logo

$$\mathbf{S} = \{s_0, s_1, s_2\}$$

$$P(\mathbf{S} = s_k) = \{p_0; p_1; p_2\}.$$

Considere os blocos formados conforme apresentado na Tabela 1.3.

Tabela 1.3 - Extensão da fonte **S** em $\mathbf{S}^2$.

Blocos	s_0s_0	s_0s_1	s_0s_2	s_1s_0	s_1s_1	s_1s_2	s_2s_0	s_2s_1	s_2s_2
Símbolos	σ_0	σ_1	σ_2	σ_3	σ_4	σ_5	σ_6	σ_7	σ_8
Probabilidades	$p_0 . p_0$	$p_0 . p_1$	$p_0 . p_2$	$p_1 . p_0$	$p_1 . p_1$	$p_1 . p_2$	$p_2 . p_0$	$p_2 . p_1$	$p_2 . p_2$

Então,

$$\mathbf{S}^2 = \{\sigma_0, \sigma_1, \ldots, \sigma_8\}$$

é chamada de ***fonte estendida*** que tem K^n símbolos, onde n é o número de símbolos da fonte discreta que, combinados, compõem cada novo símbolo da fonte estendida, ou seja,

$$K^n = 3^2 = 9 \text{ símbolos}$$

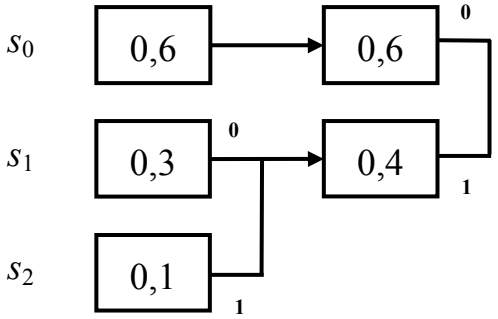
A entropia da fonte estendida é

$$H(\mathbf{S}^n) = nH(\mathbf{S}) \tag{1.12}$$

EXEMPLO 1.3

Suponha uma fonte discreta sem memória **S** = $\{s_0, s_1, s_2\}$ com probabilidades $P(\mathbf{S} = s_k)$ = {0,6; 0,3; 0,1}. Admita que essa fonte deva ser codificada com um código prefixo com eficiência igual ou maior do que 0,95.

Codificação da fonte *S*



$\boldsymbol{s}_0 = 0$
$\boldsymbol{s}_1 = 10$
$\boldsymbol{s}_2 = 11$

$$\overline{L} = \sum_{k=0}^{K-1} p_k l_k = 0{,}6(1) + 0{,}3(2) + 0{,}1(2)$$

$$\overline{L} = 1{,}4 \text{ bits/símbolo}$$

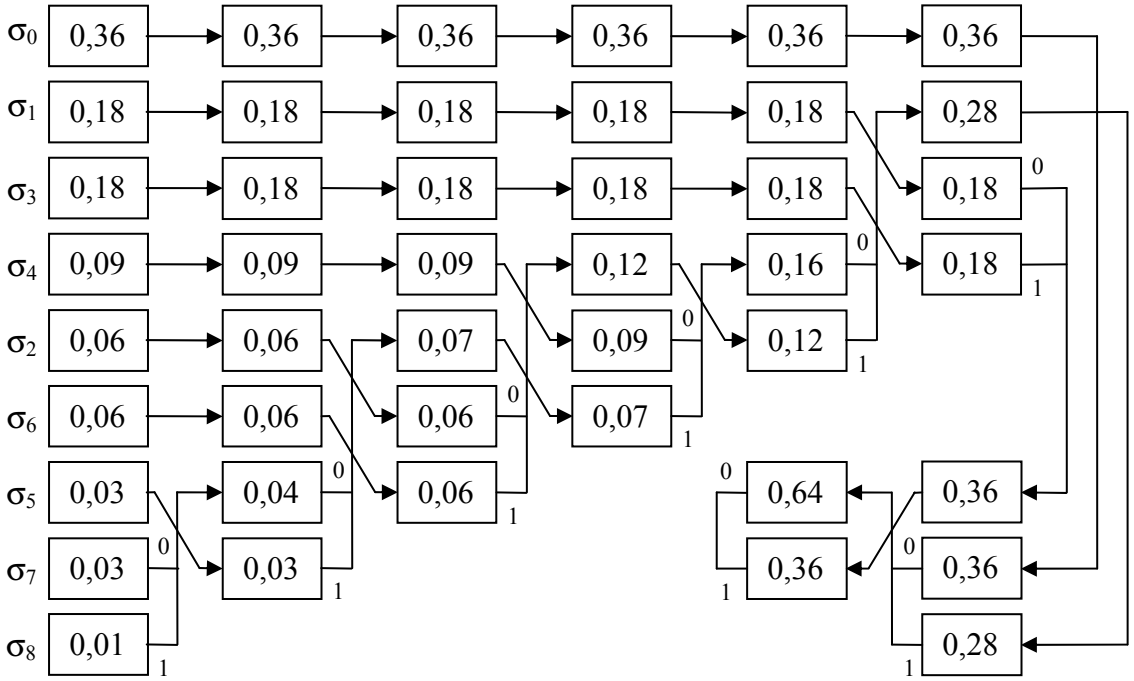
$$H(\mathbf{S}) = \sum_{k=0}^{K-1} p_k \log_2\left(\frac{1}{p_k}\right) = 0{,}6\log_2\left(\frac{1}{0{,}6}\right) + 0{,}3\log_2\left(\frac{1}{0{,}3}\right) + 0{,}1\log_2\left(\frac{1}{0{,}1}\right)$$

$$H(\mathbf{S}) \cong 1{,}295 \text{ bit/símbolo}$$

$$\eta = \frac{H(\mathbf{S})}{\overline{L}} = \frac{1{,}295}{1{,}4} \quad \Rightarrow \quad \eta = 0{,}925$$ A eficiência está abaixo da especificação.

Codificação da fonte S^2

Blocos	s_0s_0	s_0s_1	s_0s_2	s_1s_0	s_1s_1	s_1s_2	s_2s_0	s_2s_1	s_2s_2
Símbolos	σ_0	σ_1	σ_2	σ_3	σ_4	σ_5	σ_6	σ_7	σ_8
Probabilidades	0,36	0,18	0,06	0,18	0,09	0,03	0,06	0,03	0,01



$\sigma_0 = 00$	$\sigma_3 = 11$	$\sigma_6 = 0111$
$\sigma_1 = 10$	$\sigma_4 = 0100$	$\sigma_7 = 010100$
$\sigma_2 = 0110$	$\sigma_5 = 01011$	$\sigma_8 = 010101$

$$\overline{L} = \sum_{k=0}^{K-1} p_k l_k = 0{,}36\times 2 + 0{,}18\times 2 + 0{,}06\times 4 + 0{,}18\times 2 + 0{,}09\times 4 + 0{,}03\times 5 + 0{,}06\times 4 + 0{,}03\times 6 + 0{,}01\times 6$$

$\overline{L} = 2{,}67$ bit/símbolo

$$H(\mathbf{S}^2)=\sum_{k=0}^{K-1} p_k \log_2\left(\frac{1}{p_k}\right)=0{,}36\log_2\left(\frac{1}{0{,}36}\right)+0{,}18\log_2\left(\frac{1}{0{,}18}\right)+0{,}18\log_2\left(\frac{1}{0{,}18}\right)+0{,}09\log_2\left(\frac{1}{0{,}09}\right)$$
$$+0{,}06\log_2\left(\frac{1}{0{,}06}\right)+0{,}06\log_2\left(\frac{1}{0{,}06}\right)+0{,}03\log_2\left(\frac{1}{0{,}03}\right)+$$
$$+0{,}03\log_2\left(\frac{1}{0{,}03}\right)+0{,}01\log_2\left(\frac{1}{0{,}01}\right)$$

$H(\mathbf{S}^2)=2{,}59$ bit/símbolo

Ou, simplesmente,

$$H(\mathbf{S}^2)=2H(\mathbf{S})=2\times 1{,}295=2{,}59 \text{ bit/símbolo}$$

$$\eta=\frac{H(\mathbf{S}^2)}{\overline{L}}=\frac{2{,}59}{2{,}67} \quad \Rightarrow \quad \eta=0{,}97 \qquad \text{A eficiência atende à especificação.}$$

* * *

- **CODIFICAÇÃO DE LEMPEL-ZIV**

Conforme já comentado, a construção de um código prefixo, como os códigos obtidos por meio do algoritmo de Huffman, pressupõe o conhecimento prévio da estatística da fonte. Isso representa uma desvantagem em algumas aplicações, como por exemplo, na compactação de texto. A estatística da fonte para um texto pode mudar de acordo com língua utilizada, ou seja, a estatística da fonte para a língua inglesa é diferente da estatística da fonte da língua portuguesa. Além disso, dentro de uma mesma língua, a estatística da fonte pode variar de acordo com a natureza do texto. Para aplicações tais como compactações de textos existem códigos que se adaptam à estatística da fonte. O algoritmo de Lempel-Ziv é adaptativo e não necessita do conhecimento prévio da estatística da fonte. Para ilustrar este algoritmo considere a seguinte sequência binária:

000101110010100101. . .

O mecanismo de codificação pode ser entendido de acordo com a sequência apresentada a seguir.

1. Inicialmente os bits 0 e 1 são previamente armazenados nas posições numéricas 0 e 1, conforme apresentado a seguir. Tais posições numéricas compõem o *livro de códigos*.

Posição numérica	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**0**
Subsequências	0	1							
Blocos codificados	-	-							

2. Em seguida, a subsequência mais curta, ainda não armazenada em nenhuma posição binária, é armazenada na posição 2 pelo codificador. Esta subsequência é **00.** O processo de codificação consiste em transmitir a posição de memória onde se encontra a parte da subsequência **00** já armazenada anteriormente, na forma binária, e o que é *novidade* simplesmente é repetido. Ou seja, da subsequência **00**, o primeiro **0** se encontra na posição numérica **0** (**000** em binário) e o segundo **0** simplesmente é repetido. Veja a seguir.

Posição numérica	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**0**
Subsequências	0	1	00						
Blocos codificados	-	-	000 0						

3. Novamente a próxima subsequência mais curta ainda não armazenada em nenhuma posição binária, que é a subsequência **01**, é armazenada na posição 3. A posição de memória onde se encontra a parte da subsequência **01** já armazenada é a posição 0 (**000** em binário) e a parte *novidade* **1** é repetida. Logo, o bloco codificado agora é **0001**, conforme mostrado a seguir.

Posição numérica	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**0**
Subsequências	0	1	00	01					
Blocos codificados	-	-	0000	0001					

4. Este procedimento é repetido conforme apresentado na sequência a seguir.

Posição numérica	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**0**
Subsequências	0	1	00	01	011				
Blocos codificados	-	-	0000	0001	0111				

Posição numérica	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**0**
Subsequências	0	1	00	01	011	10			
Blocos codificados	-	-	0000	0001	0111	0010			

Posição numérica	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**0**
Subsequências	0	1	00	01	011	10	010		
Blocos codificados	-	-	0000	0001	0111	0010	0110		

Posição numérica	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**0**
Subsequências	0	1	00	01	011	10	010	100	
Blocos codificados	-	-	0000	0001	0111	0010	0110	1010	

Posição numérica	**-**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**0**
Subsequências	-	1	00	01	011	10	010	100	101
Blocos codificados	-	-	0000	0001	0111	0010	0110	1010	1011

Para a sequência não codificada 00 01 011 10 010 100 101..., obteve-se a sequência codificada 0000 0001 0111 0010 0110 1010 1011... .

O processo de decodificação é iniciado a partir de três premissas:

1) O codificador e o decodificador possuem o mesmo número de memórias no livro de códigos e, portanto, o decodificador conhece o tamanho do bloco codificado.
2) O decodificador também inicia o processo de decodificação com os bits 0 e 1 armazenados nas posições numéricas 0 e 1.
3) O decodificador compõe o seu próprio livro de código usando as memórias na mesma sequência que o codificador.

A partir das premissas acima é possível verificar como o decodificador compõe o seu próprio livro de código e como a decodificação é feita observando-se exclusivamente os blocos codificados recebidos. Esta questão está colocada como exercício proposto no final do capítulo.

NOTA DO PROFESSOR

Em uma primeira análise, podemos chegar à conclusão que este algoritmo não compacta nada, pelo contrário. Essa é uma idéia falsa uma vez que a sequência utilizada é muito curta para o algoritmo *reconhecer* padrões de repetição. Conforme mencionado no livro texto, na prática, são utilizados blocos com 12 bits, o que resulta em 4096 posições no *livro código*. Além disso, o algoritmo de Lempel-Ziv reconhece redundância entre caracteres, o que não ocorre com o algoritmo de Huffman. Particularmente eficiente para compactação de textos, o algoritmo de Lempel-Ziv pode atingir uma compactação de 43% contra 55% do algoritmo de Huffman, para textos em inglês.

1.5. CANAIS DISCRETOS SEM MEMÓRIA

Um canal discreto sem memória é um modelo estatístico com uma entrada X e uma saída Y, que é uma versão *ruidosa* de X. Tanto os símbolos que entram em X quanto os que saem de Y são variáveis aleatórias. Em outras palavras, é um canal que, em cada unidade de tempo, aceita um símbolo de entrada x_j, pertencente a um alfabeto $\boldsymbol{X}$, e em resposta, emite na saída um símbolo y_k, de um alfabeto $\boldsymbol{Y}$. O termo *discreto* significa que os dois alfabetos possuem tamanhos finitos e o termo *sem memória* significa que o símbolo corrente presente na saída depende apenas do símbolo corrente presente na entrada e de nenhum outro anterior.

Considere a representação mostrada na figura a seguir.

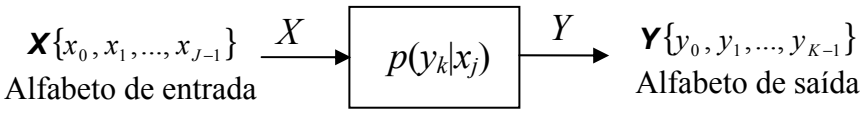


Figura 1.4 - Modelamento do canal discreto sem memória.

Na Figura 1.4 a designação $p(y_k|x_j)$ representa o conjunto das probabilidades de transição

$$p(y_k \mid x_j) = P(Y = y_k \mid X = x_j) \qquad \text{para todos } j \text{ e } k.$$

Note que os alfabetos $\boldsymbol{X}$ e $\boldsymbol{Y}$ não necessariamente têm o mesmo tamanho. Tanto pode ocorrer $\boldsymbol{X \geq Y}$ como $\boldsymbol{X \leq Y}$.

Uma forma conveniente de descrever as várias possibilidades de transição de um canal discreto sem memória é na forma de uma matriz, conforme apresentado a seguir.

$$\mathbf{P} = \begin{bmatrix} p(y_0 \mid x_0) & p(y_1 \mid x_0) & \cdots & p(y_{K-1} \mid x_0) \\ p(y_0 \mid x_1) & p(y_1 \mid x_1) & \cdots & p(y_{K-1} \mid x_1) \\ \vdots & \vdots & & \vdots \\ p(y_0 \mid x_{J-1}) & p(y_1 \mid x_{J-1}) & \cdots & p(y_{K-1} \mid x_{J-1}) \end{bmatrix} \tag{1.13}$$

A matriz P, de dimensões $J \times K$, é chamada de *matriz de canal* ou *matriz de transição*. Note que cada linha de **P** corresponde a um símbolo fixo de entrada e cada coluna a um símbolo fixo de saída. Note também que uma propriedade fundamental de **P** é que a soma de todos os elementos de uma linha da matriz é igual a 1, isto é:

$$\sum_{k=0}^{K-1} p(y_k \mid x_j) = 1 \qquad \text{para todos os } j. \tag{1.14}$$

A *distribuição de probabilidade conjunta* das variáveis aleatórias x_j e y_k é dada por:

$$p(x_j, y_k) = p(y_k \mid x_j)p(x_j) \tag{1.15}$$

A *distribuição de probabilidade marginal* da variável aleatória y_k é obtida fazendo:

$$p(y_k) = \sum_{j=0}^{J-1} p(y_k \mid x_j)p(x_j) \qquad \text{para } k = 0, 1, \ldots, K-1 \tag{1.16}$$

EXEMPLO 1.4

Um canal de grande interesse teórico e prático é o *canal binário simétrico*. Este canal é um caso especial do *canal discreto sem memória* quando $J = K = 2$. O alfabeto de entrada do canal possui dois símbolos ($x_0 = 0$, $x_1 = 1$) e o de saída também dois símbolos ($y_0 = 0$, $y_1 = 1$). O canal é simétrico porque a probabilidade de se receber **1** quando um **0** é transmitido é a mesma de se receber **0** quando um **1** é transmitido. Esta probabilidade condicional é representada por p. O diagrama da probabilidade de transição de um canal simétrico binário é mostrado na figura a seguir.

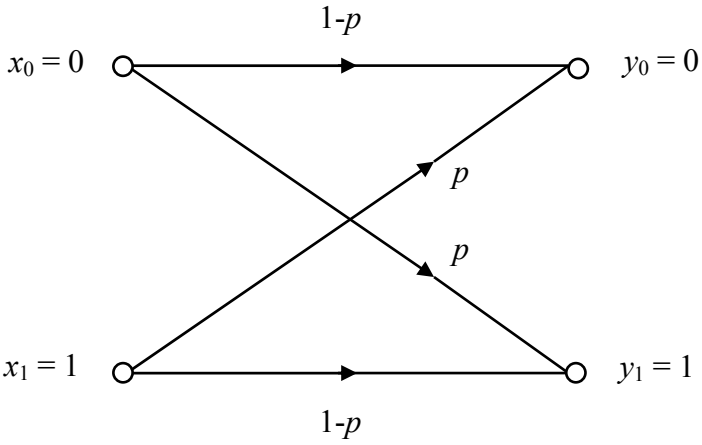


Figura 1.5 - Canal binário simétrico sem memória.

Os termos que compõem a matriz de probabilidade são:

$$p_{10} = P(y = 1 \mid x = 0)$$
$$p_{01} = P(y = 0 \mid x = 1)$$
$$p_{10} = p_{01} = p$$

* * *

1.6. INFORMAÇÃO MÚTUA

Dado que conhecemos a saída do canal (obtida de um alfabeto $\boldsymbol{Y}$) que é uma versão ruidosa da entrada do canal (obtida de um alfabeto $\boldsymbol{X}$), e que a entropia $H(\boldsymbol{X})$ é uma medida da incerteza a priori sobre X, como podemos medir a incerteza sobre X, depois da observação de Y?

A resposta é a *entropia condicional* de X selecionada de um alfabeto $\boldsymbol{X}$, dado que $Y = y_k$, representada pela notação $H(\boldsymbol{X}|Y = y_k)$. Especificamente:

$$H(\boldsymbol{X} \mid Y = y_k) = \sum_{j=0}^{J-1} p(x_j \mid y_k) \log_2 \left[\frac{1}{p(x_j \mid y_k)} \right]. \tag{1.17}$$

Que é uma variável aleatória e assume valores $H(\boldsymbol{X}|Y = y_0)$, ... , $H(\boldsymbol{X}|Y = y_{K-1})$ com probabilidades $p(y_0)$, ... , $p(y_{K-1})$, respectivamente.

A média da entropia $H(\boldsymbol{X}|Y = y_k)$, sobre o alfabeto de saída $\boldsymbol{Y}$, representada pela notação $H(\boldsymbol{X}|\boldsymbol{Y})$, é dada por:

$$
\begin{aligned}
H(\boldsymbol{X}\,|\,\boldsymbol{Y}) &= \sum_{k=0}^{K-1} H(\boldsymbol{X}\,|\,Y=y_k)\,p(y_k) \\
&= \sum_{k=0}^{K-1}\sum_{j=0}^{J-1} p(x_j\,|\,y_k)\,p(y_k)\log_2\left[\frac{1}{p(x_j\,|\,y_k)}\right] \\
&= \sum_{k=0}^{K-1}\sum_{j=0}^{J-1} p(x_j,y_k)\log_2\left[\frac{1}{p(x_j\,|\,y_k)}\right]
\end{aligned}
\tag{1.18}
$$

Assim, $H(\boldsymbol{X}|\boldsymbol{Y})$ que é chamada simplesmente de *entropia condicional*, representa a quantidade de incerteza restante sobre a entrada do canal depois da saída do canal ter sido observada. Mas $H(\boldsymbol{X})$ representa a incerteza sobre a entrada do canal antes da observação da saída. Consequentemente, a diferença $H(\boldsymbol{X})$ - $H(\boldsymbol{X}|\boldsymbol{Y})$ deve representar a incerteza sobre a entrada do canal que é resolvida pela observação de sua saída. Essa diferença é chamada de *informação mútua* do canal, representada por $I(\boldsymbol{X};\,\boldsymbol{Y})$.

Logo,

$$
\begin{aligned}
I(\boldsymbol{X};\boldsymbol{Y}) &= H(\boldsymbol{X}) - H(\boldsymbol{X}|\boldsymbol{Y}) \\
I(\boldsymbol{Y};\boldsymbol{X}) &= H(\boldsymbol{Y}) - H(\boldsymbol{Y}|\boldsymbol{X})
\end{aligned}
\tag{1.19}
$$

Onde $H(\boldsymbol{Y})$ é a entropia da saída do canal e $H(\boldsymbol{Y}|\boldsymbol{X})$ é a entropia da saída do canal, dada uma entrada de canal. Fazendo as substituições adequadas e algumas manipulações, obtém-se:

$$
I(\boldsymbol{X};\boldsymbol{Y}) = I(\boldsymbol{Y};\boldsymbol{X}) = \sum_{k=0}^{K-1}\sum_{j=0}^{J-1} p(x_j\,|\,y_k)\,p(y_k)\log_2\left[\frac{p(y_k\,|\,x_j)}{p(y_k)}\right] \tag{1.20}
$$

- **PROPRIEDADES DA INFORMAÇÃO MÚTUA**

1. A informação mútua do canal é simétrica, isto é

$$
I(\boldsymbol{X};\boldsymbol{Y}) = I(\boldsymbol{Y};\boldsymbol{X}) \tag{1.21}
$$

2. A informação mútua é sempre não negativa

$$
I(\boldsymbol{X};\boldsymbol{Y}) \geq 0
$$

3. A informação mútua de um canal está relacionada com a entropia conjunta da entrada do canal e da saída do canal, por

$$
I(\boldsymbol{X};\boldsymbol{Y}) = H(\boldsymbol{X}) + H(\boldsymbol{Y}) - H(\boldsymbol{Y},\boldsymbol{X}) \tag{1.22}
$$

1.7. CAPACIDADE DE CANAL

A *Capacidade de canal*, de um canal discreto sem memória, é a *informação mútua máxima*, $I(\boldsymbol{X}; \boldsymbol{Y})$, em qualquer intervalo de sinalização, onde a maximização deve levar em consideração todas as possíveis distribuições de probabilidade da entrada $\{p(x_j)\}$ em $\boldsymbol{X}$.

$$C = \max_{\{p(x_j)\}} I(\boldsymbol{X};\boldsymbol{Y}) \geq 0 \tag{1.23}$$

A capacidade de canal C é medida em *bits por uso do canal* ou *bits por transmissão*.

Note que a capacidade de canal C depende apenas das probabilidades de transições $p(y_k|x_j)$, que define o canal. O cálculo de C envolve a maximização da informação mútua $I(\boldsymbol{X}|\boldsymbol{Y})$ sobre a variável J [i.e., as probabilidades de entrada $p(x_0)$, ... , $p(x_{j-1})$]. Em geral, o problema variacional da determinação da capacidade de canal C é uma tarefa desafiadora.

EXEMPLO 1.5

Considere o canal simétrico binário apresentado no Exemplo 1.4.

A entropia $H(X)$ é maximizada quando os símbolos do alfabeto de entrada são equiprováveis, i.e., $p(x_0) = p(x_1) = ½$, onde $x_0 = 0$ e $x_1 = 1$. A informação mútua, $I(\boldsymbol{X}; \boldsymbol{Y})$, é similarmente maximizada, pois

$$C = I(\boldsymbol{X};\boldsymbol{Y})\big|_{\mathrm{p(x0)=p(x1)=½}}$$
$$p(y_0 \mid x_1) = p(y_1 \mid x_0) = p$$
$$p(y_0 \mid x_0) = p(y_1 \mid x_1) = 1 - p$$

Conforme já apresentado,

$$I(\boldsymbol{X};\boldsymbol{Y}) = I(\boldsymbol{Y};\boldsymbol{X}) = \sum_{k=0}^{K-1} \sum_{j=0}^{J-1} p(x_j \mid y_k)\, p(y_k) \log_2 \left[\frac{p(y_k \mid x_j)}{p(y_k)} \right]$$

Logo,

$$C = 1 + p \log_2 p + (1-p) \log_2 (1-p)$$

Usando a função entropia definida anteriormente (1.4), repetida abaixo por conveniência, i.e.,

$$\mathsf{H}(p) = -p \log_2 p - (1-p) \log_2 (1-p)$$

chega-se finalmente a

$$C = 1 - \mathsf{H}(p)\,. \tag{1.24}$$

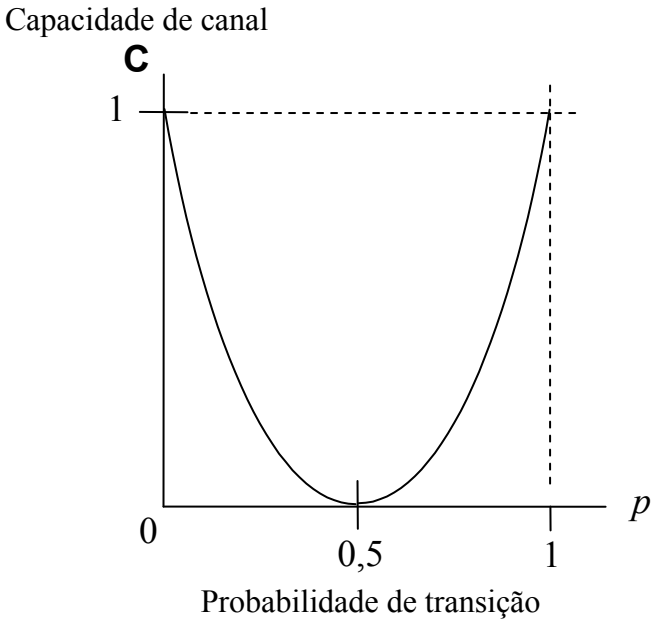


Figura 1.6 - Capacidade de canal de um canal binário discreto sem memória em função da probabilidade de transição.

O resultado apresentado na Figura 1.6 permite concluir que:

1. Quando o canal é livre de ruído, ou seja $p = 0$, a capacidade de canal C assume o seu valor máximo de um bit por uso do canal.
2. Quando a probabilidade condicional de erro é $p = ½$ devido ao ruído, a capacidade de canal assume seu menor valor que é zero, enquanto a entropia assume seu maior valor que é 1 (máxima incerteza). Neste caso o canal é dito *não utilizável*.

1.8. TEOREMA DA CAPACIDADE DE CANAL

A presença de ruído nos canais de comunicações digitais é inevitável e ruído provoca erro. Em canais muito ruidosos a probabilidade de erro pode chegar a 10^{-1} o que significa que em 10 bits transmitidos, 9 são recebidos corretamente. Este nível de confiabilidade é inaceitável para a maior parte das aplicações. A probabilidade de erro aceitável depende da aplicação. Entretanto, uma probabilidade de erro de 10^{-6} ou menor é, frequentemente, um requisito necessário para a maioria das aplicações. Para a obtenção de altos níveis de desempenho, o uso de códigos corretores de erro, ou *codificação de canal* é inevitável.

O objetivo da codificação de canal é aumentar a robustez de um sistema de comunicação digital na presença de ruído e, basicamente, este processo consiste da inserção de redundâncias na sequência binária por um codificador de canal, antes da transmissão, conforme apresentado na Figura 1.7. Na recepção, um decodificador de canal verifica a sequência recebida e com o auxílio da redundância introduzida, detecta ou mesmo corrige automaticamente alguns padrões de erro.

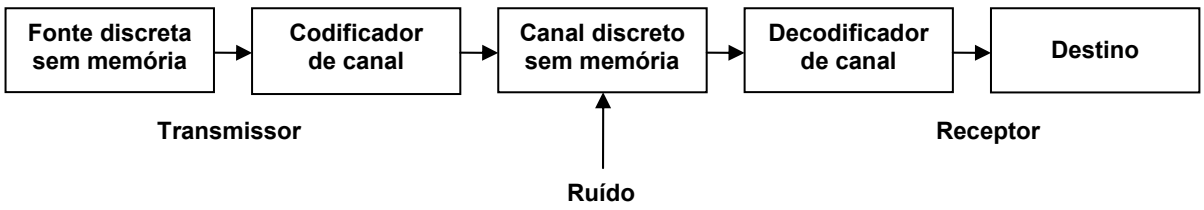


Figura 1.7 - Diagrama em blocos simplificado de um sistema de comunicação.

A fidelidade entre a sequência decodificada no receptor e a sequência originalmente transmitida influencia na probabilidade de erro final e este resultado depende do código utilizado e de algumas características do canal. Como regra geral, quanto maior é a capacidade de correção de erro do código, maior deve ser a quantidade de redundância introduzida e maior é a complexidade de decodificação. Neste ponto uma questão de extrema importância é pode ser colocada da seguinte forma:

Será que existe um esquema de codificação de canal eficiente o suficiente para de tornar a probabilidade de erro de bit arbitrariamente baixa sem que a quantidade de redundância introduzida seja demasiadamente alta?

Segundo Shannon, a resposta é um categórico **sim**, demonstrado por meio do *Teorema da Codificação de Canal*.

Até aqui a variável *tempo* não foi considerada na discussão sobre a capacidade de canal. Admitindo que uma fonte emita símbolos a cada T_s segundos e que a entropia da fonte, $H(\mathbf{S})$, é a medida de bits por símbolo transmitido, então $H(\mathbf{S})/T_s$ têm a dimensão de *bits por segundo*. Por outro lado, C é a capacidade de canal em *bits por uso do canal*. Admitindo ainda que o canal possa ser utilizado durante T_c segundos, então a capacidade de canal dividida pelo tempo de utilização tem a dimensão de *bits por segundo*, que representa a máxima taxa de transferência de informação pelo canal.

SEGUNDO TEOREMA DE SHANNON

TEOREMA DA CODIFICAÇÃO DE CANAL

Suponha uma fonte discreta sem memória com alfabeto $\mathbf{S}$ e entropia $H(\mathbf{S})$ bits por símbolo de fonte que produz um símbolo a cada T_s segundos. Seja um canal discreto sem memória que tem uma capacidade de C bits por uso do canal e é usado a cada T_c segundos. Então se

$$\frac{H(\mathsf{S})}{T_s} \leq \frac{C}{T_c} \tag{1.25}$$

deve existir um esquema de codificação de tal forma que a saída da fonte pode ser transmitida pelo canal com uma taxa de erros arbitrariamente baixa.

O parâmetro C/T_c é chamado de taxa crítica. Quando a igualdade é obtida diz-se que o sistema está transmitindo na taxa crítica. Inversamente, se

$$\frac{H(\mathsf{S})}{T_s} > \frac{C}{T_c}$$

não é possível transmitir informação pelo canal com taxa arbitrariamente baixa.

O teorema da codificação de canal é considerado o mais importante resultado da teoria da informação. Ele determina a capacidade de canal C como um *limite fundamental* sobre a taxa na qual uma transmissão pode ser realizada, livre de erros, em um canal discreto sem memória. Entretanto duas observações são importantes:

1. O teorema não mostra como construir bons códigos. Ele deve ser interpretado no sentido de uma *prova de existência*, i. e., desde que a limitação imposta pelo teorema seja satisfeita, então a existência do código é possível.
2. O teorema não apresenta um resultado preciso para a probabilidade de erro depois da decodificação de canal. Ele indica que a probabilidade de erro tende para zero conforme o comprimento do código aumenta, mais uma vez, desde que a limitação imposta pelo teorema seja satisfeita.

❑ APLICAÇÃO DO TEOREMA DA CODIFICAÇÃO DE CANAL EM CANAIS SIMÉTRICOS BINÁRIOS

Considere uma fonte discreta sem memória que emite símbolos 0's e 1's, com probabilidades iguais, a cada T_s segundos. Considere também que a entropia da fonte é igual a um bit por símbolo de fonte e, portanto, a fonte emite informação a uma taxa de $1/T_s$ bits por segundo. Considere ainda a existência de um codificador de canal cujo código possui uma taxa de codificação r, sendo r definido como:

$$r = \frac{k}{n},$$

onde k é o número de bits de informação e n é o número de bits da sequência ou do bloco codificado. Admita que o codificador de canal produza um símbolo a cada T_c segundos e, consequentemente, a taxa de transmissão de símbolos é $1/T_c$ símbolos por segundo. Desta forma, o codificador de canal ocupa um canal simétrico binário a cada T_c segundos. Portanto, a capacidade do canal por unidade de tempo é C/T_c bits por segundo. De acordo com o teorema da codificação de canal, se

$$\frac{1}{T_s} \leq \frac{C}{T_c} \tag{1.26}$$

a probabilidade de erro pode ser arbitrariamente baixa se usado um esquema de codificação de canal adequado.

Como

$$r = \frac{T_c}{T_s}$$

então

$$r \leq C .$$

Ou seja, para $r \leq C$, deve existir um código capaz de permitir uma transmissão com taxa de erro tão baixa quanto se queira. Para que este conceito fique claro, veja o Exemplo 1.

Para que o significado do teorema da codificação de canal, aplicada a um canal binário simétrico fique clara, veja o exemplo a seguir.

EXEMPLO 1.6

Seja um canal binário simétrico cuja entrada é equiprovável e a probabilidade de transição do canal (probabilidade de erro) é igual a 0,1. Qual deve ser o maior número de bits de informação para cada bloco de 7 bits de forma ser possível, teoricamente, a obtenção de uma taxa de erro arbitrariamente baixa ?

SOLUÇÃO:

Do Exemplo 1.5, tem-se

$$C = 1 - \mathsf{H}(p)$$

$$\mathsf{H}(p) = -p\log_2 p - (1-p)\log_2(1-p)$$

Como $p = 0{,}1$

$$C = 1 - \left(-0{,}1\frac{\log 0{,}1}{\log 2} - 0{,}9\frac{\log 0{,}9}{\log 2} \right)$$

$$C = 0{,}531$$

Conforme apresentado

$$r = \frac{k}{n} \le C$$

$$k \le nC$$

Como $n = 7$

$$k = \lfloor nC \rfloor = \lfloor 7 \times 0{,}531 \rfloor = \lfloor 3{,}72 \rfloor$$

$$k = 3$$

* * *

1.9. ENTROPIA DIFERENCIAL E INFORMAÇÃO MÚTUA PARA CONJUNTOS CONTÍNUOS

Até o momento as fontes e os canais considerados foram de conjuntos de variáveis que são discretas em amplitude. Alguns dos conceitos apresentados serão estendidos para conjunto contínuos de forma a permitir estabelecer outro limite fundamental da teoria da informação que é o Terceiro Teorema de Shannon ou *Teorema da Capacidade de Informação*.

Considere uma variável aleatória contínua X com função densidade de probabilidade $f_X(x)$. Por analogia com a entropia para variáveis discretas pode-se escrever que

$$h(X) = \int_{-\infty}^{\infty} f_x(x)\log_2\left[\frac{1}{f_x(x)}\right]dx, \tag{1.27}$$

onde $h(X)$ é a entropia diferencial de X. Se n variáveis aleatórias contínuas formam um vetor aleatório contínuo $\mathbf{X}$, então utilizando a mesma analogia, pode-se definir a entropia diferencial de $\mathbf{X}$, como

$$h(\mathbf{X}) = \int_{-\infty}^{\infty} f_x(\mathbf{x})\log_2\left[\frac{1}{f_x(\mathbf{x})}\right]d\mathbf{x}, \tag{1.28}$$

onde $f_X(\mathbf{x})$ é a função densidade de probabilidade conjunta de $\mathbf{X}$.

Considere agora um par de variáveis aleatórias contínuas X e Y. Mais uma vez, por analogia ao que foi definida para variáveis discretas, a informação mútua torna-se

$$I(X;Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{\mathrm{X,Y}}(x,y)\log_2\left[\frac{f_X(x\mid y)}{f_{\mathrm{x}}(x)}\right]dx\,dy \tag{1.29}$$

onde $f_{\mathrm{X,Y}}(x,y)$ é a função densidade de probabilidade conjunta de X e Y, e $f_{\mathrm{X}}(\mathrm{x}|\mathrm{y})$ é a função densidade de probabilidade condicional de X, dado que $Y = y$.

Também por analogia, as seguintes propriedades para a informação mútua são válidas:

1. $I(X;Y) = I(Y;X)$
2. $I(X;Y) \geq 0$
3. $I(X;Y) = h(X) - h(X\mid Y) = h(Y) - h(Y\mid X)$

onde $h(X)$ é a entropia diferencial de X e $h(X|Y)$ é a entropia diferencial condicional de X dado Y, que, ainda por analogia, é definida como

$$h(X\mid Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{\mathrm{X,Y}}(x,y)\log_2\left[\frac{1}{f_{\mathrm{x}}(x\mid y)}\right]dx\,dy\,. \tag{1.30}$$

O parâmetro $h(Y|X)$ é a entropia diferencial condicional de Y dado X e é definido de maneira similar a $h(X|Y)$.

1.10. Teorema da Capacidade de Informação

A formulação do *Teorema da Capacidade de Informação* para canais limitados em faixa e potência utiliza a idéia de informação mútua. Para isso, considere um processo estacionário $X(t)$ com média zero e limitado em uma faixa de B hertz. Seja X_k, k = 1, 2, ... , K variáveis aleatórias contínuas obtidas por amostragem uniforme de $X(t)$ à uma taxa de Nyquist de $2B$ por segundo. Estas amostras são transmitidas em T segundos em um canal ruidoso, também limitado em faixa de B hertz. Então o número de amostras, K, é dado por

$$K = 2BT\ ,$$

onde X_k é uma amostra do sinal transmitido em um canal perturbado com ruído Gaussiano branco aditivo (AWGN) de média zero e densidade espectral de potência $N_0/2$. As amostras do sinal recebido são representadas pelas variáveis aleatórias contínuas Y_k, k = 1, 2, ... , K, estatisticamente independentes, de forma que

$$Y_k = X_k + N_k \qquad k = 1, 2, \ldots, K\ ,$$

onde N_k representa a amostras de ruído Gaussiano com média zero e densidade espectral de potência dado por

$$\sigma^2 = N_0 B \ . \tag{1.31}$$

O canal descrito pelas duas últimas equações é chamado de *canal Gaussiano sem memória*, *discreto no tempo* e é modelado conforme mostrado na figura a seguir.

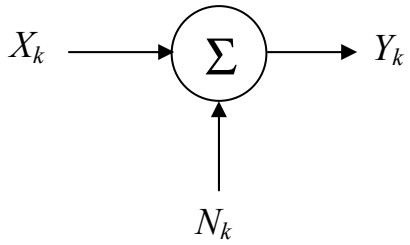


Figura 1.8 - Modelagem do canal gaussiano sem memória discreto no tempo.

Como a transmissão é, tipicamente, limitada em potência, então a potência média transmitida pode ser definida como

$$P = E[X_k^2] \qquad k = 1, 2, \ldots, K$$

A *capacidade de informação* do canal é definida como a máxima informação mútua entre a entrada X_k que satisfaz a limitação de potência apresentada acima, ou seja,

$$C = \max_{f_{X_k}(x)} \{ I(X_k ; Y_k) : E[X_k^2] = P \}, \tag{1.32}$$

onde a maximização é feita com relação a $f_{Xk}(x)$. Por conveniência a informação mútua $I(X_k; Y_k)$ pode ser escrita como

$$I(X_k ; Y_k) = h(Y_k) - h(Y_k \mid X_k)$$

Uma vez que X_k e N_k são variáveis aleatórias independentes e sua soma é igual a Y_k a entropia diferencial condicional de Y_k, dado X_k, é igual a entropia diferencial de N_k, conforme mostrado a seguir.

$$h(N_k) = h(Y_k \mid X_k)$$

Consequentemente,

$$I(X_k ; Y_k) = h(Y_k) - h(N_k) \, .$$

A maximização da informação mútua requer a maximização de $h(Y_k)$. A variância da amostra Y_k é $(P + \sigma^2)$. Logo, Y_k é também uma variável aleatória gaussiana, consequentemente, sua entropia diferencial máxima é

$$h(Y_k) = \frac{1}{2} \log_2 [2\pi e (P + \sigma^2)]$$

e a entropia diferencial de N_k é

$$h(N_k) = \frac{1}{2}\log_2(2\pi e\sigma^2)$$

Substituindo as duas entropias diferenciais apresentadas acima na expressão da *informação mútua*, obtêm-se:

$$C = \frac{1}{2}\log_2\left(1+\frac{P}{\sigma^2}\right) \quad \text{bits por transmissão} \qquad (1.33)$$

Com o canal sendo usado K vezes para a transmissão de K amostras em T segundos, então a capacidade de informação por unidade de tempo é obtida multiplicando-se a equação apresentada acima por K/T. Como $K = 2BT$ e $\sigma^2 = N_0B$, então a capacidade de canal por unidade de tempo torna-se:

$$C = B\log_2\left(1+\frac{P}{N_0B}\right) \quad \text{bits por segundo.} \qquad (1.34)$$

Assim o terceiro e mais famoso teorema de Shannon pode ser enunciado da forma como apresentado a seguir.

TERCEIRO TEOREMA DE SHANNON

TEOREMA DA CAPACIDADE DE INFORMAÇÃO OU TEOREMA DA CAPACIDADE DE CANAL

A capacidade de informação de um canal contínuo com largura de faixa de B hertz, perturbado por ruido Gaussiano branco aditivo de densidade espectral $N_0/2$, é dada por

$$C = B\log_2\left(1+\frac{P}{N_0B}\right),$$

dada em bits por segundo, onde P é a potência média do sinal transmitido.

O Teorema da Capacidade de Informação é um dos mais notáveis resultados da Teoria da Informação onde, em uma única expressão, é destacada a interação entre a largura de faixa do canal e a relação sinal/ruído. Note que a capacidade do canal varia linearmente com a largura de faixa e logaritmicamente com a relação sinal ruído. Desta forma, *é mais fácil aumentar a capacidade do canal aumentando a sua largura de faixa do que aumentando a relação sinal ruído*.

O teorema implica que, para uma dada relação sinal/ruído, é possível transmitir a uma taxa de C bits por segundo com probabilidade de erro arbitrariamente baixa empregando um esquema de codificação de canal suficientemente complexo. Por outro lado, não é possível transmitir a uma taxa maior do que C bits por segundo com qualquer esquema de codificação sem uma probabilidade de erro definida. Assim, o Teorema da Capacidade de Canal define um *limite fundamental* para a taxa de transmissão livre de erros para um canal gaussiano com uma dada relação sinal/ruído e largura de faixa limitada. Para alcançar este limite, no entanto, o sinal deve ter propriedades estatísticas semelhantes a do ruído gaussiano.

Veja o exemplo de aplicação apresentado a seguir.

EXEMPLO 1.7

Seja um canal AWGN limitado em faixa (B = 3,4 kHz) e uma potência de recepção (saída do canal) igual a 1pW. Determine a capacidade do canal, considerando que o mesmo está submetido a uma temperatura equivalente de ruído igual a 290 K. Determine também a relação sinal-ruído em dB.

SOLUÇÃO:

$$C = B\log_2\left(1+\frac{P}{N_0 B}\right) = B\log_2\left(1+\frac{P}{kTB}\right) = B\log_2\left(1+S/N\right)$$

$$S/N = \frac{P}{kTB} = \frac{1\times10^{-12}}{1{,}38\times10^{-23}\times290\times3{,}4\times10^{3}} = 73{,}49\times10^{3}$$

$$C = 3{,}4\times10^{3}\,\frac{\log\left(1+73{,}49\times10^{3}\right)}{\log 2}$$

$$C = 54{,}96 \text{ kbit/s}$$

$$\left(S/N\right)_{dB} = 10\log 73{,}49\times10^{3}$$

$$\left(S/N\right)_{dB} = 48{,}66\ dB$$

CONCLUSÃO:

Deve haver um esquema de codificação de canal, suficientemente complexo, que permita a transmissão a uma taxa de ≅ 55 kbit/s em um canal com 3,4 kHz de largura de faixa para uma relação sinal/ruído de ≅ 49 dB, com uma taxa de erros de bit arbitrariamente baixa.

* * *

1.11. IMPLICAÇÕES DO TEOREMA DA CAPACIDADE DE INFORMAÇÃO

Para analisar as implicações do Teorema da Capacidade de informação, admita a existência de um *sistema ideal* definido como sendo aquele que transmite a uma taxa R_b igual à capacidade de canal C. Desta forma, a potência média do sinal pode ser escrita como

$$P = E_b C\,,$$

onde E_b é a energia transmitida por bit. Assim (1.34) pode ser reescrita como

$$\frac{C}{B} = \log_2\left(1 + \frac{E_b}{N_0}\frac{C}{B}\right).$$

Equivalentemente, a relação entre a energia por bit e a densidade espectral de ruído E_b/N_0 em termos de C/B para o sistema ideal é

$$\frac{E_b}{N_0} = \frac{2^{C/B} - 1}{C/B}.$$

Com a expressão acima é possível traçar a curva que estabelece a relação entre a *eficiência de largura de faixa* R_b/B e a relação E_b/N_0, que é chamado de diagrama da eficiência de largura de faixa. A forma genérica deste diagrama é mostrada na Figura 1.9, onde a curva intitulada "*limite da capacidade*" corresponde ao sistema ideal no qual $R_b = C$.

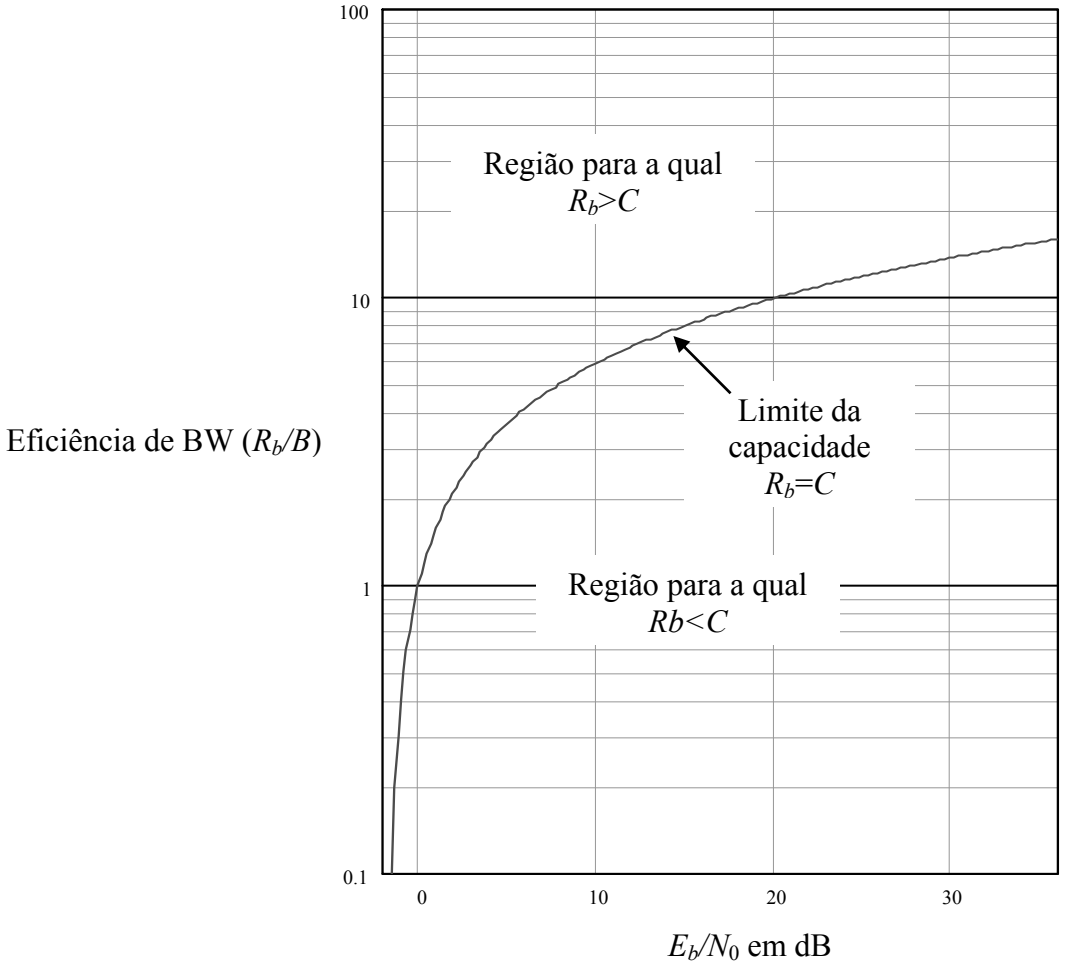


Figura 1.9 - Eficiência × Largura de Faixa para $R_b = C$.

O limite é representado por uma fronteira entre a região em que é possível uma transmissão livre de erros (R_b < C) e a região em que isso não é possível (R_b > C) em um plano da *eficiência de largura de faixa* em função da relação entre a *energia de bit* e a *densidade espectral de ruído*, conforme apresentado a seguir.

1.12. TEORIA DA DISTORÇÃO-TAXA

Conforme apresentado anteriormente, para se obter uma codificação de uma fonte discreta sem memória que permita a transmissão e recuperação da informação sem imperfeições, o Teorema da Codificação de Fonte não pode ser violado. Em outras palavras, isso significa que o comprimento médio do código de fonte não pode ser menor do que a entropia da fonte. Entretanto, em muitas aplicações práticas existem restrições que introduzem imperfeições na codificação, o que resulta em inevitáveis distorções.

Um exemplo disso são as fontes de sinais contínuos, como por exemplo, o sinal de voz. Neste caso, os sistemas de comunicação geralmente impõem restrições quanto ao comprimento das palavras códigos, acarretando na limitação do número de níveis de quantização do sinal. Assim, a aproximação de um sinal contínuo para um número reduzido de níveis discretos produz *erros de quantização severos* cuja consequência é a degradação da qualidade do sinal por meio da introdução, no sinal transmitido, de um *ruído de quantização* com amplitude significativa.

Neste caso, o problema é definido como *codificação de fonte com um critério de fidelidade*. A teoria da distorção-taxa encontra aplicações em dois tipos de situação:

1. Codificação de fonte onde o alfabeto de codificação permitido não pode representar exatamente a informação da fonte, caso no qual é forçosa uma *compressão de dados* com perdas.
2. Transmissão de informação à uma taxa maior do que a capacidade do canal.

Se vista de forma adequada, a teoria da distorção-taxa pode ser vista como uma extensão natural dos teoremas de codificação de Shannon. Para maiores detalhes sobre este assunto veja Seção 9.13, Capítulo 9 do livro Communication System, Simon Haykin.

1.13. COMPRESSÃO DE DADOS

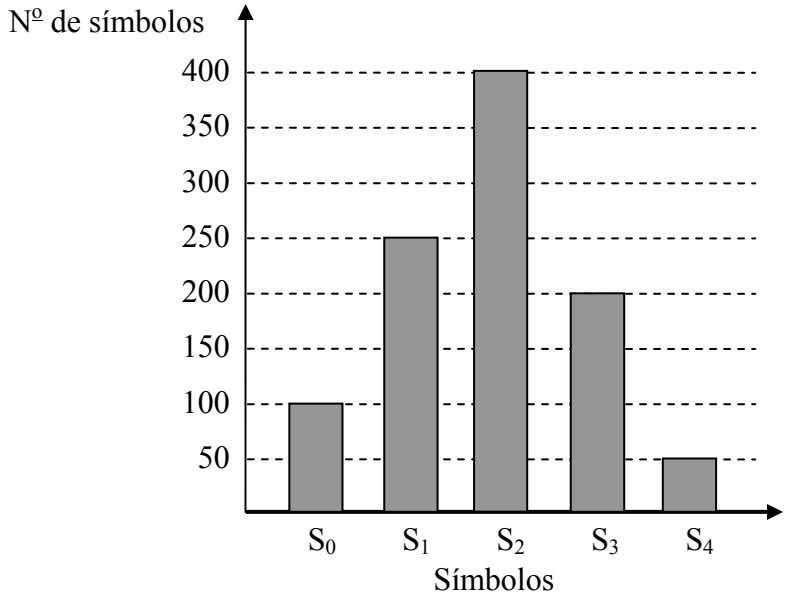
A teoria da distorção-taxa permite considerar a idéia da *compressão* de dados, que envolve uma redução proposital no conteúdo da informação de uma fonte contínua ou discreta. Especificamente, pode-se imaginar um *compressor de dados*, ou um *compressor de sinais*, como um dispositivo que alimenta um código que produz o mínimo de bits necessários para representar a saída da fonte com uma *distorção* aceitável ou permissível. Assim, o compressor de dados retém o conteúdo essencial de informação da saída da fonte descartando finos detalhes de maneira controlada. Desta forma, a compressão de dados é uma operação com perdas no sentido de que a entropia da fonte é reduzida (i.e., há perda de informação) independentemente do tipo de fonte considerada.

No caso de uma fonte discreta, a razão para o uso da compressão de dados é codificar a saída da fonte à uma taxa menor do que a entropia da fonte. Consequentemente, a entropia da fonte é violada, o que significa que a reprodução exata do que a fonte produziu é impossível.

No caso de uma fonte contínua, a entropia da fonte é infinita. Entretanto, na prática é necessário que os níveis de quantização sejam finitos e, consequentemente, o codificador de fonte produzirá codificação à uma taxa finita. Assim, não é possível codificar um sinal contínuo com um número finito de bits sem produzir distorções.

1.14. EXERCÍCIOS

1. O alfabeto de uma fonte discreta sem memória é composto por 5 símbolos. Os símbolos emitidos pela fonte devem ser codificados por meio de um código prefixo a partir do histograma da frequência de emissão apresentado abaixo. O número total de símbolos do histograma foi emitido em 1 ms.



Pede-se:

a) A entropia da fonte.
b) As palavras códigos de um código prefixo para esta fonte.
c) A eficiência de codificação.
d) Determine qual é a taxa de bit na saída do codificador de fonte.

2. O alfabeto de uma fonte discreta sem memória (**S**) possui 3 símbolos. As probabilidades de emissão dos símbolos S_0, S_1 e S_2 são respectivamente 0,55; 0,35 e 0,1. Pede-se:

a) Criar um código prefixo para a fonte **S**.
b) Criar um código prefixo para a fonte estendida $\mathbf{S^2}$.
c) Comparar as eficiências de codificação para os dois códigos e comentar o resultado.

3. Considere um codificador de Lempel-Ziv com 8 posições de memória (de 0 a 7) que parte com o bit **0** armazenado na posição de memória 000 e o bit **1** na posição de memória 001. Codifique a sequência mensagem 11101001100010110100...

4. Considere um decodificador de Lempel-Ziv com 8 posições de memória (de 0 a 7) que parte com o bit **0** armazenado na posição de memória 000 e o bit **1** na posição de memória 001. Reconstrua a sequência mensagem a partir dos blocos codificados 0011 0001 0000 0100 0110 1000 1101 0000 0010...

5. Uma fonte discreta sem memória emite símbolos que são estatisticamente independentes. Admita que esses símbolos sejam codificados com um código prefixo casado com a fonte, resultando nas seguintes palavras códigos: $S_0 = 0001$; $S_1 = 011$; $S_2 = 0011$; $S_3 = 10$; $S_4 = 0010$; $S_5 = 010$; $S_6 = 0000$; $S_7 = 11$. Determine a entropia desta fonte.

6. Um terminal remoto deve enviar símbolos alfanuméricos para um computador à uma taxa de 1 milhão de símbolos por segundo por meio de um canal cuja relação sinal/ruído é igual a 20 dB. Admita que o terminal possua 256 símbolos diferentes e que a transmissão destes símbolos seja equiprovável e estatisticamente independente. Determine a largura de faixa mínima do canal, à luz do teorema da capacidade de informação, de modo que essa transmissão possa ser teoricamente realizável.

7. Considere que o alfabeto de uma fonte possui 64 símbolos com iguais probabilidades de ocorrência. Admita que tais símbolos sejam convertidos em palavras binárias para que sejam transmitidos em um canal AWGN cuja largura de faixa é igual a 10^6 Hz. Determine qual deve ser a menor relação sinal ruído teórica, em dB, para que a transmissão de 10^6 símbolos por segundo possa ser realizada com taxa de erro arbitrariamente baixa.

* * *

1.15. REFERÊNCIA BIBLIOGRÁFICA

[1] HAYKIN, S. - Communications Systems. 4th Ed. John Wiley & Sons. 2000.