Capítulo 8

Teoria da Informação

A teoria dos sinais e dos sistemas de comunicação, constituindo embora uma ferramenta válida de análise, não aborda o processo fundamental das comunicações que é o da produção e transferência de *informação*. Reconhecendo a necessidade de uma teoria própria para este processo, Claude Shannon, matemático dos Laboratórios Bell que ali trabalhara em sistemas de comunicação durante a segunda grande guerra, desenvolveu uma "Teoria Matemática das Comunicações" que publicou pela primeira vez em 1948.

Segundo Shannon, a questão principal da engenharia de comunicações era a seguinte: Dada uma fonte de informação qualquer, como é que devem ser representadas as mensagens por ela emitidas, de modo a poderem ser transmitidas fiavelmente através de um canal de comunicação, dadas as inerentes limitações físicas deste?

Ao atacar esta questão Shannon concentrou-se na informação da mensagem em si, e não nos sinais utilizados para a transmitir. Esta abordagem deu origem ao que é hoje conhecido por *Teoria da Informação*, que se tem desenvolvido como disciplina híbrida da matemática e da engenharia.

A Teoria da Informação estuda quatro problemas fundamentais:

1. A *medida* da informação produzida por uma fonte;

2. A *codificação da fonte* destinada a representar a informação produzida pela fonte com o menor número possível de símbolos;

3. A *capacidade* do canal de comunicação para transmitir informação,

ou seja, o máximo da quantidade de informação por unidade de tempo que é possível transmitir num canal;

4. A *codificação do canal* como mecanismo para melhor utilizar a capacidade do canal que normalmente se designa por codificação para control de erros.

Nesta teoria o termo *canal* é utilizado em sentido lato, designando o sistema (ou a *via*) de comunicação completo entre a fonte e o destino da informação. O termo *codificação* é utilizado igualmente no seu sentido mais genérico de representação de mensagens quer por formas de onda contínuas quer por formas de onda discretas ou outros símbolos. Os quatro problemas atrás referidos conjugam-se naquele que constitui o teorema fundamental da Teoria da Informação, a saber:

> Dado um canal de comunicação e uma fonte cujo débito de informação não excede a capacidade do canal, existe um código tal que a informação pode ser transmitida através do canal com uma frequência de erros arbitrariamente pequena, apesar da presença de ruido.

O aspecto surpreendente deste teorema é a promessa que faz de transmissão de informação *sem erros* através de um canal ruidoso, o que se torna possível através de operações de *codificação*.

O processo de codificação envolve, em geral, duas operações distintas de codificação/descodificação, representadas em diagrama na figura 8.1. O *codificador* e o *descodificador do canal* executam a tarefa de *control de erros*, detectando-os e eventualmente corrigindo-os através de mecanismos que se estudarão no próximo capítulo e pelos quais os efeitos do ruído podem ser reduzidos ou eliminados. Assim, podemos desde já considerar que o conjunto dos blocos da figura 8.1 contidos no bloco tracejado, equivale a um *canal sem ruído*. A teoria da informação vai um passo mais além afirmando que a codificação óptima do canal dá origem a um canal equivalente sem ruído com uma capacidade de transmissão de informação bem definida. Desde que o débito de informação da fonte não ultrapasse a capacidade do canal, o *codificador da fonte* representa a informação através de símbolos – normalmente binários – tentando fazê-lo com o menor número possível desses símbolos, operação esta que se designa por compressão da fonte. O *descodificador da fonte* executa a operação inversa retirando dos símbolos recebidos a informação neles contida entregando-a ao destino. Pode dizer-se então que o par *codificador/descodificador da fonte* têm o papel de adaptar a fonte ao canal equivalente sem ruído. Começaremos por considerar o caso da informação representada sob a

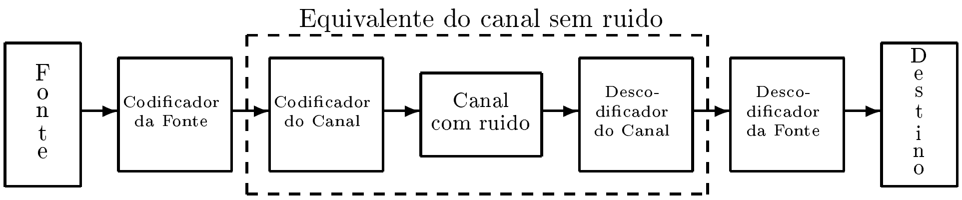

Figura 8.1: Sistema de comunicação com codificação da fonte e do canal

forma digital ou *discreta*. Trataremos a medida da informação, a codificação da fonte, a transmissão da informação e a capacidade do canal discreto. Estes conceitos serão depois extendidos ao caso mais realista da transmissão de informação em canais *contínuos* onde as mensagens tomam a forma de sinais variáveis no tempo, concluindo com a *lei de Hartley-Shannon* que define a capacidade do canal em função da largura de banda e da razão sinal-ruído, servindo de termo de comparação para a avaliação do desempenho de sistemas de comunicação.

Convém recordar o que se disse no início relativamente ao conceito de *canal* em Teoria da Informação: é a *via* de comunicação simplex completa a partir da saída do codificador da fonte e até à entrada do respectivo descodificador no destino, isto é, o bloco a tracejado na figura 8.1.

8.1 Medida da Informação

O ponto de partida da teoria da informação é a medida da *informação*, termo técnico que não deve ser confundido com *dados*, nem com *conhecimento* ou *significado*, conceitos cuja definição e medida são ainda objecto de debate.

No contexto das comunicações, informação não é mais do que o produto, o bem ou o objecto imaterial útil produzido por uma fonte que tem de ser transferido para um utilizador num destino. Se a informação está préviamente disponível no destino a transferência será zero.

Suponhamos, por exêmplo, uma pessoa que planeia as suas férias de verão para o Algarve e ouve uma das seguintes previsões do tempo para essa altura:

No Algarve o sol vai nascer
No Algarve vai chover
No Algarve vai nevar

A primeira mensagem não contém nenhuma informação pois é certo que o sol nasce todos os dias em toda a parte. A previsão de chuva porém contém informação não disponível anteriormente. A terceira previsão fornece ainda mais informação dado que nevar no Algarve é um acontecimento raro e porventura inesperado.

É senso comum que quanto menos provável fôr uma determinada mensagem maior a quantidade de informação nela contida. Pode pois concluir-se que a medida da informação deve relacionar-se com o grau de *incerteza* do destinatário quanto à mensagem que vai receber. Posto de outra forma, a informação mede a *liberdade de escolha* exercida pela fonte ao seleccionar uma mensagem dentro do conjunto universo das possíveis mensagens.

8.1.1 Informação própria

Torna-se evidente, portanto, que a medida da informação deve ser uma função da *probabilidade* de ocorrência da mensagem.

Seja $f()$ essa função e x_i uma mensagem arbitrária tal que a probabilidade do acontecimento x_i ser seleccionado para transmissão é $P(x_i) = P_i$. A quantidade de informação, I_i, associada à ocorrência de x_i, designada de *informação própria* ou *auto-informação* de x_i, será então $I_i = f(P_i)$.

A função $f()$ deve possuir as seguintes propriedades:

(i) $f(P_i) \geq 0 \quad \text{para} \quad 0 \leq P_i \leq 1$

(ii) $\lim_{P_i \to 1} f(P_i) = 0$

(iii) $f(P_i) > f(P_j) \quad \text{para} \quad P_i < P_j$

Estas propriedades fundamentam-se, ainda, no senso-comum. A primeira indica que a informação nunca é negativa, a segunda que ela é nula se o acontecimento fôr certo e a terceira indica que a informação aumenta com a incerteza.

Existem muitas funções que possuem estas três propriedades. Contudo, considere-se o caso em que a fonte produz duas mensagens sucessivas e independentes, x_i e x_j, com probabilidade conjunta $P(x_i x_j) = P_{ij} = P_i P_j$. Pretende-se que a informação total seja igual à soma da informação das mensagens individuais o que exige a seguinte propriedade adicional:

(iv) $f(P_i P_j) = f(P_i) + f(P_j)$

A *única* função que satisfaz estas quatro propriedades é a função logarítmica negativa, $-\log_b()$. A base b que se utilizar para o logaritmo define a *unidade* de medida da informação. A convenção adoptada na teoria da informação é tomar $b = 2$ e designar a unidade correspondente por *bit*.

Definição 8.1 — Bit como unidade de medida de informação
O bit é a quantidade de informação necessária para escolher uma entre duas alternativas igualmente prováveis ou, a quantidade de informação contida numa mensagem emitida por uma fonte capaz de emitir apenas duas mensagens distintas e equiprováveis.

Portanto, e por definição, a quantidade de informação, ou informação própria, I_i numa mensagem x_i é dada por:

$$I_i \stackrel{def}{=} \log_2 \frac{1}{P_i} \quad bits \tag{8.1}$$

Se a fonte possuir um alfabeto de apenas duas mensagens, x_1 e x_2, e se $P(x_1) = P(x_2) = 1/2$, tem-se $I_1 = I_2 = \log_2 2 = 1$ *bit*.

Deve ter-se cuidado em distinguir *bits* de informação dos *dígitos binários* utilizados para a representar ou codificar – especialmente porque um dígito binário pode transportar mais ou menos do que um bit de informação.

De modo a evitar a confusão, o dígito binário é muitas vezes designado por *binit* e, no contexto da Teoria da Informação, o termo *bit* designa a unidade de medida de informação e não o dígito binário.

Por simplicidade utiliza-se também o termo *símbolo* em alternativa ao termo *mensagem*.

É frequente ter de se converter logaritmos de uma base b qualquer, por exêmplo a natural ou a decimal, para a base 2. É fácil verificar que

$$\log_2 \alpha = \frac{\log_b \alpha}{\log_b 2} = \frac{\ln \alpha}{\ln 2} = \frac{\log_{10} \alpha}{\log_{10} 2} \tag{8.2}$$

8.1.2 Entropia

Consideremos agora uma fonte de informação que emite uma sequência de símbolos (mensagens) seleccionados de um alfabeto de m símbolos distintos. Seja $X = \{x_1, x_2, \cdots, x_m\}$ esse conjunto alfabeto. Podemos tratar cada símbolo x_i como uma mensagem que ocorre com probabilidade P_i e transporta a auto-informação I_i. Os valores das probabilidades devem obviamente satisfazer a igualdade $\sum_{i=1}^{m} P_i = 1$.

Suponhamos que os sucessivos símbolos são estatísticamente independentes e são produzidos pela fonte a um débito médio de r_s símbolos por segundo. Suponhamos ainda que a fonte é estacionária, o que significa que as probabilidades não variam com o tempo. Estas propriedades definem o que se designa por *fonte discreta sem memória*.

A quantidade de informação produzida pela fonte durante um intervalo de símbolo arbitrário é uma variável aleatória discreta que toma valores $I_1, I_2, \cdots, I_m$. A informação média por símbolo é então dada pela média estatística,

$$\mathcal{H}(X) \stackrel{def}{=} \sum_{i=1}^{m} P_i\, I_i = \sum_{i=1}^{m} P_i \log_2 \frac{1}{P_i} \quad \textit{bits/símbolo} \tag{8.3}$$

grandeza que é chamada *entropia* da fonte. A equação 8.3 pode ser interpretada como representando a quantidade média de informação obtida quando uma variável aleatória X toma um valor.

O valor de $\mathcal{H}(X)$ para uma dada fonte depende das probabilidades dos símbolos da fonte, P_i, e da cardinalidade, m, do alfabeto estando limitado por

$$0 \leq \mathcal{H}(X) \leq \log_2 m \tag{8.4}$$

O limite inferior corresponde à inexistência de incerteza o que acontece quando um dos símbolos tem probabilidade $P_j = 1$ e todos os restantes $P_i = 0$ $(i \neq j)$ ou seja, quando a fonte emite sempre o mesmo símbolo. Portanto, a prova do limite inferior na relação 8.4 obtem-se facilmente bastando notar que $\alpha \log_2(1/\alpha) \to 0$ quando $\alpha \to 0$. O limite superior corresponde à máxima incerteza que ocorre quando $\forall i : P_i = 1/m$, isto é, quando todos os símbolos são igualmente prováveis. A prova deste limite superior pode ser obtida resolvendo o problema 8.6.

8.1.3 Débito de informação

A equação 8.3 também pode ser interpretada da seguinte maneira: quando uma fonte emite uma sequência de $n \gg 1$ símbolos, a informação total a produzida é aproximadamente igual a $n\,\mathcal{H}(X)$ bits. Dado que a fonte produz r_s símbolos por segundo, a duração desta sequência é de n/r_s seg. A informação deve pois ser transferida a um débito médio $n\,\mathcal{H}(X)/(n/r_s) = r_s\,\mathcal{H}(X)$ bits por segundo. O *débito médio de informação* de uma fonte é então definido por

$$\mathcal{R} \stackrel{def}{=} r_s\,\mathcal{H}(X) \quad bits/seg \tag{8.5}$$

uma grandeza crítica em sistemas de transmissão. O teorema fundamental da teoria da informação (devido a Shannon) diz que a informação produzida por qualquer fonte discreta sem memória pode ser codificada em dígitos binários e transmitida através de um canal sem ruído a um ritmo binário r_b, com $r_b \geq \mathcal{R}$.

Exêmplo 8.1 – Fonte binária

Para ilustrar a variação de $\mathcal{H}(X)$ entre aqueles extremos consideremos o caso especial mas importante da fonte binária ($m = 2$) com

$$P_1 = p \qquad\qquad P_2 = 1 - p$$

Substituindo estes valores das probabilidades na equação 8.3 obtem-se a entropia da fonte binária

$$\mathcal{H}(X) = \Omega(p) \stackrel{def}{=} p\,\log_2 \frac{1}{p} + (1-p)\,\log_2 \frac{1}{1-p} \tag{8.6}$$

Designamos esta entropia pela função $\Omega()$ pois vamos utilizá-la mais adiante. O gráfico de $\Omega(p)$ na figura 8.2 mostra um máximo centrado em $p = 1 - p = 1/2$ onde se tem $\mathcal{H}(X) = \log_2 2 = 1$ bit/símb. $\mathcal{H}(X)$ decresce monótonamente para zero quando $p \to 1$ ou $p \to 0$.

Exêmplo 8.2 – Fonte quaternária

Suponhamos uma fonte que emite r_s = 2000 simb/seg seleccionados de um alfabeto de quatro símbolos ($m = 4$) com probabilidade dos símbolos e auto-informação dados na tabela 8.1.

A equação 8.3 fornece o valor da entropia desta fonte

$$\mathcal{H}(X) = \frac{1}{2}\times 1 + \frac{1}{4}\times 2 + \frac{1}{8}\times 3 + \frac{1}{8}\times 3 = 1.75 \;\; \text{bits/símb}$$

Tabela 8.1:

x_i	P_i	I_i
A	1/2	1
B	1/4	2
C	1/8	3
D	1/8	3

que é ligeiramente inferior ao valor máximo $\log_2 = m$. *O débito de informação é*

$$\mathcal{R} \ = \ 2000 \times 1.75 \ = \ 3500 \ \ \text{bits/seg}$$

e através de codificação apropriada, deve ser possível transmitir a informação da fonte a um ritmo binário $r_b \ \geq \ 3500$ dig bin/seg *(resolver o problema 8.8).*

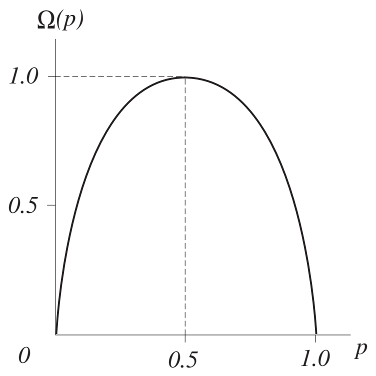

Figura 8.2: Função de entropia binária

8.2 Codificação da fonte discreta sem memória

Quando uma fonte discreta sem memória produz m símbolos[1] equiprováveis de tal forma que $\mathcal{R} = r_s \log_2 m$, todos os símbolos transportam a mesma quantidade de informação e pode obter-se uma transmissão com

[1]tenha-se em atenção que, neste contexto, símbolo é sinónimo de mensagem

bom rendimento utilizando codificação de linha multi-nível com m níveis distintos (codificação M-ária) e com um ritmo na linha[2] r_c igual ao débito de símbolos r_s. Mas quando os símbolos possuem probabilidades de ocorrência diferentes e se tem

$$\mathcal{R} \;=\; r_s\,\mathcal{H}(X) \;<\; r_s \log_2 m \tag{8.7}$$

torna-se necessário *codificar a fonte* em função da quantidade de informação variável por símbolo.

Nesta secção aborda-se um método para efectuar esta codificação considerando-se apenas a codificação binária. Também se podem obter resultados equivalentes para o caso da codificação não-binária desde que se adopte uma base apropriada para os logaritmos.

O codificador binário da figura 8.3 converte os símbolos da fonte em *palavras de código* compostas por dígitos binários produzidos a um ritmo de r_b *dígitos binários/seg.* A saída do codificador pode ser encarada como uma fonte binária com entropia

$$\mathcal{H}(X) \;=\; \Omega(p) \;=\; p\,\log_2 \frac{1}{p} + (1-p)\,\log_2 \frac{1}{1-p} \qquad \textit{bits/dig-bin} \tag{8.8}$$

e com débito de informação

$$\mathcal{R} \;=\; r_b\,\Omega(p) \;\leq\; r_b \log_2 2 \;=\; r_b \qquad \textit{bits/s} \tag{8.9}$$

em que p é a probabilidade de um dos dois símbolos binários. É óbvio que a operação de codificação não pode gerar informação adicional àquela que é produzida pela fonte, nem mesmo perder informação desde que o código seja unívocamente decifrável. Assim, igualando os débitos de informação

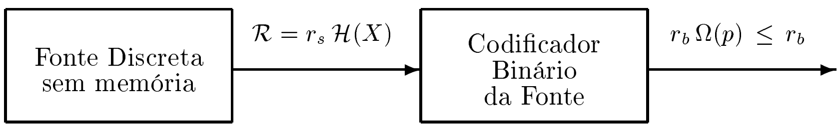

Figura 8.3: Fonte, Codificador da Fonte e débitos de informação

à entrada e à saida do codificador tem-se $r_s\,\mathcal{H}(X) = r_b\,\Omega(p) \leq r_b$, ou seja,

[2] r_c designa o ritmo de símbolos do código no canal de transmissão. Quando esses símbolos são dígitos binários o respectivo ritmo costuma designar-se por r_b.

$r_b/r_s \geq \mathcal{H}(X)$, em que a quantidade $\overline{N} = \frac{r_b}{r_s}$ é um parâmetro importante designado por *comprimento médio do código*. Fisicamente, $\overline{N}$ corresponde ao número médio de dígitos binários por símbolo da fonte. Se fôr N_i o comprimento da palavra de código correspondente ao símbolo i, $\overline{N}$ é a média estatística

$$\overline{N} = \sum_{i=1}^{m} P_i N_i \tag{8.10}$$

8.2.1 Códigos óptimos

Considere-se uma fonte cujos m símbolos do alfabeto ocorrem com probabilidades $P_1, P_2, \cdots, P_m$ e estão codificados em binário com comprimentos dos códigos $N_1, N_2, \cdots, N_m$. Considere-se ainda que os símbolos estão ordenados por ordem decrescente de probabilidade ($P_1 \geq P_2 \geq \cdots \geq P_m$). Pode mostrar-se que um código óptimo deve satisfazer os seguintes requisitos:

(i) Não existem dois símbolos codificados com a mesma combinação de dígitos binários.

(ii) Nenhum código de símbolo é prefixo de nenhum outro.

(iii) Os símbolos com maior probabilidade possuem códigos mais curtos, isto é, $N_1 \leq N_2 \leq \cdots \leq N_m$.

(iv) Os códigos dos dois símbolos menos prováveis têm o mesmo comprimento ($N_m = N_{m-1}$) e diferem apenas no dígito final.

O primeiro requisito garante que cada símbolo é univocamente decifrável de modo a garantir que não há perda de informação. Pela equação 8.1, que define a medida da quantidade de informação própria I_i do símbolo x_i em *bits*, verifica-se que, idealmente, o símbolo x_i deveria ser codificado com uma palavra de código com $N_i = -\log_2 P_i$ *dígitos binários* que seria o código mais eficiente. Contudo, na maior parte dos casos, o valor de $-\log_2 P_i$ não é inteiro e o melhor que se consegue é codificar x_i com um número de dígitos que é o inteiro imediatamente superior pelo que, na melhor das hipóteses ter-se-á para N_i

$$-\log_2 P_i \leq N_i < -log_2 P_i + 1 \tag{8.11}$$

Tomando a desigualdade do lado esquerdo pode obter-se sucessivamente

$$\begin{aligned} N_i &\geq -\log_2 P_i \\ 2^{-N_i} &\leq P_i \\ \sum_{i=1}^{m} 2^{-N_i} &\leq 1 \end{aligned}$$

pelo que se pode verificar que este primeiro requisito impõe uma condição, embora indirecta, aos valores que N_i pode tomar, isto é, uma condição necessária e suficiente para que um código binário seja univocamente decifrável é que os comprimentos dos códigos N_i obedeçam à *desigualdade de Kraft*

$$\mathrm{Kr} = \sum_{i=1}^{m} 2^{-N_i} \leq 1 \tag{8.12}$$

O segundo requisito, característico dos chamados códigos *instantâneos*, embora não estritamente necessário, assegura que as sequências de dígitos binários do código podem ser descodificadas passo a passo, isto é, percorridas sequencialmente da esquerda para a direita descodificam ao primeiro ajuste (*match*) com um padrão de dígitos que seja o código de um símbolo, ou seja, os códigos são as folhas da árvore de código.

O terceiro requisito é óbvio pois, se não fôr observado, basta trocar os códigos dos dois símbolos nessas condições para se obter um comprimento médio menor (eq 8.10).

Para demonstrar o quarto requisito, suponhamos que $N_m > N_{m-1}$. Visto que o código para o símbolo x_{m-1} não pode ser prefixo do código de x_m, os primeiros N_{m-1} dígitos deste constituem um código único e portanto os restantes dígitos podem ser retirados. Se ambos os códigos destes dois símbolos diferirem nalguma posição excepto na última podem-se retirar os últimos dígitos de cada um deles até essa posição para se obter melhores códigos.

Multiplicando os termos das desigualdades 8.11 por P_i e somando para todo o alfabeto, e tendo em consideração a definição da entropia dada pela equação 8.3 obtém-se sucessivamente

$$-P_i \log_2 P_i \leq P_i N_i < -\log_2 P_i + P_i$$

$$-\sum_{i=1}^{m} P_i \log_2 P_i \leq \sum_{i=1}^{m} P_i N_i < -\sum_{i=1}^{m} \log_2 P_i + \sum_{i=1}^{m} P_i$$

$$\mathcal{H}(X) \leq \overline{N} < \mathcal{H}(X) + 1 \tag{8.13}$$

donde se verifica que o código óptimo tem um comprimento médio em *dígitos binários* por símbolo que é superior, em menos de uma unidade, ao valor da entropia em *bits* por símbolo do conjunto original de símbolos. Este facto conduz a uma outra interpretação para a entropia de uma variável aleatória X como sendo o menor número médio de dígitos binários necessários para representar um valor de X.

8.2.2 Rendimento da codificação e compressão

O rendimento de uma codificação mede-se pela aproximação conseguida entre os valores numéricos do comprimento médio do código e da entropia da fonte ou, consequentemente, pela aproximação entre o ritmo binário à saída do codificador e o ritmo de entropia da fonte, e expressa-se pela relação

$$\rho = \frac{\mathcal{H}(X)}{\overline{N}} = \frac{\mathcal{R}}{r_b} \leq 1 \tag{8.14}$$

O processo de codificação mais simples gera *códigos de comprimento fixo* com todas as palavras de código possuindo o mesmo comprimento $N_i = N_f$. A desigualdade de Kraft dá $\mathrm{Kr} = m2^{-N_f} \leq 1$ e portanto, para ser decifrável, o código terá de ter comprimento (fixo) $N_f \geq \log_2 m$ e o seu rendimento será $\frac{\mathcal{H}(X)}{N_f} \leq \frac{\mathcal{H}(X)}{\log_2 m}$. Quando $\mathcal{H}(X) < \log_2 m$ só se poderão obter melhores rendimentos quando se utilizarem *códigos de comprimento variável*.

A compressão, c, obtida numa determinada codificação da fonte mede-se pela *redução* percentual do número médio, $\overline{N}$, de dígitos binários por símbolo relativamente à codificação da mesma fonte se esta fosse feita com o código de comprimento fixo mínimo $N_f = \min\limits_{n \in \mathcal{Z}_+} n \geq \log_2 m$, isto é

$$c = \frac{N_f - \overline{N}}{N_f} \times 100\ \%$$

Exêmplo 8.3 – **Codificação da fonte quaternária**
Na fonte do exêmplo 8.2 os quatro símbolos A, B, C e D correspondem às formas de onda obtidas por amostragem e quantização a quatro níveis de um sinal analógico (PAM quantizado). Determinar um código binário

para a fonte e comparar o ritmo binário obtido com o débito de informação da fonte.

Os resultados obtidos no exêmplo 8.2 foram $\mathcal{H}(X) = 1.75$ *bits/símbolo e* $\mathcal{R} = 3500$ *bits/s. Consideremos, na tabela 8.2, vários potenciais códigos binários para esta fonte: O código I tem comprimento fixo com* $\overline{N} =$

Tabela 8.2: Códigos binários para a fonte quaternária

x_i	P_i	Código I	Código II	Código III	Código IV
A	1/2	00	0	0	0
B	1/4	01	1	01	10
C	1/8	10	10	011	110
D	1/8	11	11	0111	111
$\overline{N}$		2.0	1.25	1.875	1.75
Kr		1.0	1.5	0.9375	1.0

$\log_2 m = 2$ *dig bin/simb o que, comparado com* $\mathcal{H}(X) = 1.75$ *bits/simb, dá um rendimento* $\rho = \mathcal{H}(X)/\overline{N} = \frac{1.75}{2} \approx 88\%$ *que não é muito baixo.*

Aplicando a média estatística ao código II obtem-se

$$\overline{N} = \frac{1}{2} + \frac{1}{4} + 2 \times \frac{1}{8} + 2 \times \frac{1}{8} = 1.25 \; < \; \mathcal{H}(X)$$

e o número de Kraft dá

$$Kr = 2^{-1} + 2^{-1} + 2^{-2} + 2^{-2} = 1.5 \; > \; 1$$

mas o resultado $\overline{N} < \mathcal{H}(X)$ *não tem significado porque* $Kr > 1$ *o que quer dizer que o código II não é univocamente decifrável. Por exêmplo, a sequência de código 10011 poderia ser descodificada como BAABB, CABB, CAD, etc. Este código destrói realmente a informação da fonte e portanto não pode ser adoptado.*

O código III, conhecido por código de vírgula, *tem* $Kr < 1$ *assegurando a sua decifragem. O início de cada palavra de código é marcada pelo dígito 0, chamado* a vírgula. *As vírgulas acabam por ser dígitos redundantes e conduzem a* $\overline{N} = 1.875 > \mathcal{H}(X)$ *o que no entanto já é melhor do que o código de comprimento fixo (código I).*

O código IV é um código em árvore que possui a propriedade de nenhuma palavra de código aparecer como prefixo de nenhuma outra. Assim, por *exêmplo, a sequência 110010111 representa sem ambiguidade a sequência de mensagens CABD. Este código é* óptimo *relativamente à fonte em questão dado que* $\overline{N} = 1.75 = \mathcal{H}(X)$ *com* $Kr = 1$.

8.2.3 Códigos de Huffman

Existem vários algoritmos para calcular o código para uma dada fonte. Mencionaremos um dos mais conhecidos e simples de construir e que conduz a códigos decifráveis de comprimento variável com bom rendimento. São os códigos de Huffman também designados de códigos de Shannon-Fano. O algoritmo conduz a um código em árvore e é o seguinte:

(1) Ordenar os símbolos por ordem decrescente de probabilidade;

(2) Dividir o conjunto assim ordenado em dois subconjuntos tais que a soma das probabilidades em cada um deles seja o mais aproximadamente possível igual a metade da soma das probabilidades no conjunto anterior. Manter a ordenação.

(3) O dígito seguinte do código binário dos símbolos do primeiro dos sub-conjuntos é o **0** e o dos do outro é o **1**;

(4) Se os sub-conjuntos contêm um só elemento, a codificação terminou para esses sub-conjuntos;

(5) Repetir para cada um dos restantes sub-conjuntos (passo 2.)

Exêmplo 8.4 – Codificação de Huffman
Aplicar o algoritmo de Huffman para codificar a fonte com $m = 8$ *cuja estatística é a indicada na tabela 8.3.*

Tabela 8.3: Estatística da fonte

x_i	A	B	C	D	E	F	G	H
P_i	0.50	0.15	0.15	0.08	0.08	0.02	0.01	0.01

Os símbolos são ordenados por ordem decrescente de probabilidade e são efectuados os passos de codificação indicados na tabela 8.4.

O código resultante possui comprimento médio $\overline{N} = 2.18$ *dig bin e um rendimento* $\rho = \frac{2.15}{2.18} \approx 99\%$ *(resolver o problema 8.11).*

Tabela 8.4: Codificação da fonte

x_i	P_i	Passos de codificação						Código
		1	2	3	4	5	6	
A	0.50	0						0
B	0.15	1	0	0				100
C	0.15	1	0	1				101
D	0.08	1	1	0				110
E	0.08	1	1	1	0			1110
F	0.02	1	1	1	1	0		11110
G	0.01	1	1	1	1	1	0	111110
H	0.01	1	1	1	1	1	1	111111
$\mathcal{H}(X) = 2.15$								$\overline{N} = 2.18$

8.2.4 Codificação por Blocos

Exêmplo 8.5 – Codificação por blocos

Suponhamos uma fonte sem memória que emite símbolos de um alfabeto X com apenas dois símbolos, $X = \{A, B\}$, ocorrendo com probabilidades $P_A = 0.8$ e $P_B = 0.2$. A entropia desta fonte é $\mathcal{H}(X) = 0.8 \times \log_2(1.25) + 0.2 \times \log_2(5) = 0.722$ bits/símbolo mas o melhor que se consegue para a sua codificação é utilizar um dígito binário por cada símbolo, ou seja, $A = 0$ e $B = 1$ cujo comprimento médio é $\overline{N} = 1$ dígito binário/símbolo$_X$. O rendimento desta codificação é de 72.2% e a compressão é de 0 %. Estes valores podem ser melhorados se a codificação se fizer por blocos de símbolos da fonte. Por exêmplo, se se codificarem dois símbolos de cada vez, pode-se considerar que se trata de um novo alfabeto Y com quatro símbolos $Y = \{AA, AB, BA, BB\}$ com probabilidades $P_{AA} = 0.64$, $P_{AB} = 0.14$, $P_{BA} = 0.14$ e $P_{BB} = 0.04$, em que a probabilidade conjunta é $P_{ij} = P_i \cdot P_j$ pelo facto de se tratar de uma fonte sem memória e portanto os símbolos serem estatisticamente independentes. A tabela 8.5 mostra o código de Huffman para Y que possui um comprimento médio $\overline{N}_2 = 1.560$ dígitos binários/símbolo$_Y$, ou seja $\overline{N} = \frac{\overline{N}_2}{2} = 0.780$ dígitos binários/símbolo$_X$. O rendimento e a compressão obtidos com esta codificação por blocos de $K = 2$ símbolos são $\rho = \frac{\mathcal{H}(X)}{\overline{N}} = \frac{0.722}{0.780} = 0.926$ e $c = \frac{N_f - \overline{N}}{N_f} \times 100 = \frac{1-0.780}{1} = 22$ %, bastante superiores aos obtidos com a

Tabela 8.5:

y_i	P_{y_i}	Código
AA	0.64	0
AB	0.16	11
BA	0.16	100
BB	0.04	101
$\overline{N}_2$		1.56

codificação imediata ($K = 1$).

Codificando a fonte em blocos de $K = 3$ símbolos de cada vez, isto é, através de um alfabeto com oito símbolos obtém-se ainda melhores resultados. O alfabeto é $Z = \{AAA, AAB, ABA, ABB, BAA, BAB, BBA, BBB\}$ verificando-se, nestas condições, $\overline{N}_3 = 2.184$ dígitos binários/símbolo$_Z$ pelo que $\overline{N} = \overline{N}_3/3 = 0.728$, $\rho = \frac{0.722}{0.728} = 0.992$ e $c = \frac{1-0.728}{1} = 27.2$ %.

O ponto até que este processo deve ser levado deve ser avaliado face aos valores máximos teóricos que é possível obter e que ocorrem quando $\overline{N} = \mathcal{H}(X)$ sendo, neste caso, $\rho = \frac{0.722}{0.722} = 1$ e $c_{max} = \frac{1-0.722}{1} = 27.8$ %.

A codificação por blocos conduz tendencialmente a um código óptimo, isto é, com $K \to \infty$ tem-se $\overline{N}_K \to \mathcal{H}(X)$, $\rho \to 1$ e $c \to c_{max}$. De facto, para a codificação por blocos, a desigualdade 8.13 escreve-se

$$\mathcal{H}(X) \leq \overline{N}_K < \mathcal{H}(X) + 1$$

donde, dividindo por K e tendo em atenção que a entropia da fonte não se altera com a codificação, se obtém

$$\mathcal{H}(X) \leq \frac{\overline{N}_K}{K} < \mathcal{H}(X) + \frac{1}{K}$$

ou, visto que $\overline{N} = \frac{\overline{N}_K}{K}$,

$$\mathcal{H}(X) \leq \overline{N} < \mathcal{H}(X) + \frac{1}{K}$$

Podemos agora enunciar um dos teoremas fundamentais da Teoria da Informação embora não procedamos à sua demonstração geral:

Teorema 8.1 – Teorema de Shannon da Codificação da Fonte
Toda a fonte de informação caracterizada por um valor da entropia $\mathcal{H}(X)$ *bits/símbolo, pode ser codificada em binário de tal forma que o comprimento médio do código,* $\overline{N}$*, é limitado por*

$$\mathcal{H}(X) \;\leq\; \overline{N} \;\leq\; \mathcal{H}(X) + \epsilon \tag{8.15}$$

em que ϵ *é uma quantidade positiva tão pequena quanto se desejar.*

O código óptimo é aquele em que $\epsilon = 0$, ou seja, $\overline{N} = \mathcal{H}(X)$. Na prática, nem sempre se consegue obter um código nestas condições sendo satisfatório um código sub-óptimo com $\overline{N} > \mathcal{H}(X)$ desde que este possua um bom rendimento. Na codificação por blocos está-se a fazer $\epsilon = \frac{1}{K}$.

8.3 Fontes com memória

Até aqui considerámos fontes discretas sem memória cujos símbolos sucessivos são estatisticamente independentes. No entanto a maior parte das fontes de informação possui *memória*, na medida em que a probabilidade de emissão de um determinado símbolo depende dos que foram emitidos anteriormente. A linguagem escrita que se rege pelas regras da gramática constitui um exêmplo de uma fonte com memória. Verifica-se, por exêmplo, que a letra U que eventualmente ocorre com uma probabilidade $P(U) \approx 0.2$, em termos de frequência relativa, ocorre com probabilidade igual a 1 sempre que a letra anterior fôr Q. A probabilidade condicional desta situação é indicada por $P(U|Q) \approx 1$, isto é, praticamente não existe na forma escrita da língua portuguesa nenhum caso em que a letra Q não venha seguida da letra U.

Esta influência pode estender-se a um ou mais símbolos anteriores. É quase certo, por exêmplo, que a seguir à frase *PODE MOSTRAR-SE* aparece a palavra *QUE*, ou seja, $P(\text{QUE}|\text{PODE MOSTRAR-SE}) \approx 1$.

O efeito da *memória* é o de reduzir a incerteza dando origem a um valor para a entropia mais baixo do que aquele que seria obtido se se utilizasse a estatística absoluta da fonte em lugar da condicional.

Suponhamos que uma fonte possui *memória de primeira ordem*, isto é, a fonte só se *lembra* do símbolo precedente. Para formular uma expressão da entropia, consideremos a *probabilidade condicional* $P(x_i|x_j)$ do símbolo x_i ser escolhido depois do símbolo x_j. Substituindo $P_i = P(x_i)$ por $P(x_i|x_j)$

na equação 8.3 obtem-se a *entropia condicional* relativamente ao símbolo x_j

$$\mathcal{H}(X|x_j) \stackrel{def}{=} \sum_{i=1}^{m} P(x_i|x_j) \log_2 \frac{1}{P(x_i|x_j)} \tag{8.16}$$

que representa a informação média por símbolo dado que o símbolo anterior foi um determinado x_j. Fazendo a média estatística para todos os possíveis símbolos imediatamente anteriores ter-se-á

$$\mathcal{H}(X) = \sum_{j=1}^{m} P(x_j)\, \mathcal{H}(X|x_j) \tag{8.17}$$

expressão que fornece a entropia *real* da fonte com memória de primeira ordem e que pode ser escrita sob a forma desenvolvida

$$\mathcal{H}(X) = \sum_{i=1}^{m}\sum_{j=1}^{m} P(x_j)P(x_i|x_j) \log_2 \frac{1}{P(x_i|x_j)} \tag{8.18}$$

$$\mathcal{H}(X) = \sum_{i=1}^{m}\sum_{j=1}^{m} P(x_i\, x_j) \log_2 \frac{1}{P(x_i|x_j)} \tag{8.19}$$

As relações 8.18 e 8.19 anteriores são equivalentes onde $P(x_i\, x_j) = P(x_j) \cdot P(x_i|x_j)$ representa a *probabilidade conjunta* dos símbolos x_i e x_j.

Pode-se deduzir uma expressão equivalente para o caso geral de uma fonte com memória de ordem n. A notação será mais complexa pois x_j tem de ser substituido pelo *estado* da fonte em termos dos n símbolos anteriormente emitidos e haverá m^n possíveis estados a considerar. A fonte pode então ser modelada por um autómato de estados finitos.

Exêmplo 8.6 – Codificação por blocos da fonte com memória
Consideremos a fonte binária do exêmplo 8.5 e suponhamos que, para além das probabilidades dos símbolos isoladamente $P_A = 0.8$ *e* $P_B = 0.2$ *se verifica também que a probabilidade de emissão de um símbolo depende do símbolo que foi emitido imediatamente antes desse. Suponhamos que essa dependência se reflete nas probabilidades condicionais* $P_{A|A} = 0.9$ *e* $P_{A|B} = 0.4$*, isto é, que se verifica que a probabilidade do símbolo* A *ocorrer a seguir a um símbolo* A *é de* 0.9 *e a probabilidade do símbolo* B *ser emitido a seguir a um símbolo* A *é de* 0.4*. Naturalmente, as probabilidades complementares são respectivamente* $P_{B|A} = 1 - P_{A|A} = 0.1$ *e* $P_{B|B} = 1 - P_{A|B} = 0.6$*. Esta fonte, que possui memória de primeira ordem, pode*

ser modelada pelo diagrama de transições da figura 8.4 em que P_A *e* P_B *são as probabilidades associadas ao estado inicial e* P'_A *e* P'_B *as probabilidades*

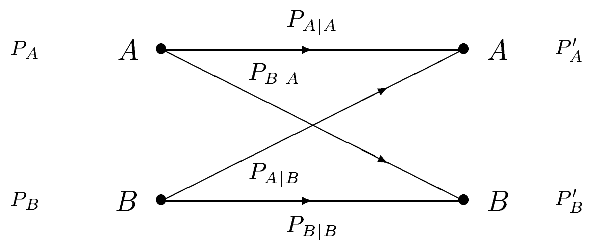

Figura 8.4: Modelo da fonte binária com memória

associadas ao estado seguinte em que já foi emitido um símbolo. Estas probabilidades relacionam-se pelas equações de transição de estado

$$\begin{aligned} P'_A &= P_A \cdot P_{A|A} \; + \; P_B \cdot P_{A|B} \\ P'_B &= P_A \cdot P_{B|A} \; + \; P_B \cdot P_{B|B} \end{aligned}$$

que podem ser escritas sob a forma matricial

$$[P'_A \; P'_B] = [P_A \; P_B] \cdot \begin{bmatrix} P_{A|A} & P_{B|A} \\ P_{A|B} & P_{B|B} \end{bmatrix} \tag{8.20}$$

em que a matriz do lado direito representa a matriz de transição de estado.

No caso deste exêmplo, visto que as probabilidades dos símbolos considerados isoladamante são as mesmas, ter-se-á $P'_A = P_A = 0.8$ *e* $P'_B = P_B = 0.2$ *pelo que a sua matriz de transição*

$$\begin{bmatrix} 0.9 & 0.1 \\ 0.4 & 0.6 \end{bmatrix}$$

deve ser consistente com a equação de transição de estado 8.20. Nestas condições, a entropia real da fonte deve ser inferior à encontrada nas condicões do exêmplo 8.5. De acordo com a equação 8.18 tem-se então

$$\begin{aligned} \mathcal{H}(X) &= P_A P_{A|A} \log_2 \frac{1}{P_{A|A}} \; + \; P_A P_{B|A} \log_2 \frac{1}{P_{B|A}} \; + \\ & \quad + P_B P_{A|B} \log_2 \frac{1}{P_{A|B}} \; + \; P_B P_{B|B} \log_2 \frac{1}{P_{B|B}} \end{aligned}$$

$$\begin{aligned}\mathcal{H}(X) &= 0.8 \times 0.9 \log_2 \frac{1}{0.9} + 0.8 \times 0.1 \log_2 \frac{1}{0.1} + \\ &\quad + 0.2 \times 0.4 \log_2 \frac{1}{0.4} + 0.2 \times 0.6 \log_2 \frac{1}{0.6} \\ \mathcal{H}(X) &= 0.569 \; \textit{bits/símbolo}\end{aligned}$$

O valor obtido para a entropia neste caso é de 0.569 *bits/símbolo, inferior ao valor de* 0.722 *bits/símbolo obtido anteriormente, se os símbolos fossem independentes. A codificação de Huffman desta fonte por blocos de* $K = 2$ *símbolos fornece o código da tabela 8.6, tendo agora que se considerar que* $P_{AA} = P_A P A|A = 0.72$, $P_{AB} = P_A P_{B|A} = 0.08$, $P_{BA} = P_B P_{A|B} = 0.08$ *e* $P_{BB} = P_B P_{B|B} = 0.12$

Tabela 8.6:

y_i	P_{y_i}	Código
AA	0.72	0
BB	0.12	11
AB	0.08	100
BA	0.08	101
$\overline{N}_2$		1.44

que é semelhante ao da tabela 8.5 mas o comprimento médio é agora também menor, isto é, $\overline{N} = \overline{N}_2/2 = 0.72$ *dígitos binários/símbolo*$_X$ *sendo a compressão conseguida de* $c = \frac{N_f - \overline{N}}{N_f} \times 100 = \frac{1-0.720}{1} = 28$ *%. Devido à memória, e consequentemente a uma menor entropia, esta fonte pode ser bastante mais comprimida que na situação anterior sendo o limite máximo teórico da compressão* $c_{\max} = \frac{1-0.569}{1} = 43.1$ *%. O rendimento do código para blocos de* $K = 2$ *é, neste caso,* $\rho = \frac{\mathcal{H}(X)}{\overline{N}} = \frac{0.0.569}{0.720} = 0.79$ *bastante inferior ao encontrado anteriormente que era de* 0.926 *pelo que se deve considerar codificar esta fonte com blocos maiores (resolver o problema 14).*

Uma fonte com memória é dita *redundante* quando as probabilidades condicionais reduzem significativamente $\mathcal{H}(X)$ em relação ao limite superior $\log_2 m$. A redundância num texto em língua Portuguesa pode ser tão elevada como 50%, o que significaria que metade das letras ou palavras num texto longo não são essênciais à sua compreensão e portanto à entrega da informação completa ao destinatário. Vários factores contribuem para que tal seja possivel como seja a dedução por contexto, isto é, dve sr capz d lr sta frse msmo fltndo lgmas ltrs.

Se a incerteza é reduzida por efeito de memória então a *previsibilidade* aumenta. A codificação de fontes com memória pode portanto ser baseada em métodos de previsão. O estudo destes métodos está fora do âmbito deste curso pelo que somente fazemos menção à sua existência.

Concluimos esta secção notando que a codificação da fonte é um proceso que *retira a redundância* produzida pela fonte, processo que também se designa por *compressão da fonte*. A redundância, especialmente na linguagem escrita, constitui um mecanismo natural de correcção de erros. Uma vez retirada essa redundância, resultando uma frase *comprimida*, como a que se exêmplificou atrás, a perda ou alteração de uma letra poderá torná-la parcial ou completamente incompreensível.

8.4 Transmissão de informação: o canal

Iremos estudar de seguida a transmissão de informação em canais de comunicação sob o ponto de vista da teoria da informação. Começaremos por considerar o caso em que tanto a fonte como o canal de transmissão são discretos, isto é, as mensagens são representadas e transmitidas sob a forma de símbolos de um determinado alfabeto finito. Passaremos depois ao caso mais realista da fonte e do canal contínuos onde as mensagens são produzidas e transmitidas sob a forma de funções contínuas do tempo.

8.4.1 Canais discretos

Informação mútua

Considere-se o sistema de transmissão de informação representado na figura 8.5. A fonte discreta selecciona símbolos do seu alfabeto X para transmissão através do canal. Por seu lado, o ruído e outras formas de perturbação da transmissão podem transformar os símbolos transmitidos noutros do alfabeto Y do destino. Pretende-se medir a informação transmitida. Os vários tipos de probabilidades de símbolos que estarão em causa são os seguintes:

Figura 8.5: Sistema discreto de transmissão de informação

$P(x_i)$	probabilidade da fonte seleccionar o símbolo x_i para transmissão;
$P(y_j)$	probabilidade de ser recebido no destino o símbolo y_i;
$P(x_i\,y_j)$	probabilidade conjunta de ser transmitido o símbolo x_i e recebido o símbolo y_j;
$P(x_i\|y_j)$	probabilidade condicional de ter sido transmitido o símbolo x_i dado ter-se recebido o símbolo y_j;
$P(y_j\|x_i)$	probabilidade condicional de se receber o símbolo y_j dado ter-se transmitido o símbolo x_i.

Consideramos o canal *invariante no tempo* e *sem memória*, portanto as probabilidades condicionais são independentes do tempo e dos símbolos precedentes. Em especial, os $P(y_j|x_i)$ representam as *probabilidades de transição directas* do canal, indicadas na figura 8.6 a qual ilustra o modelo de um canal m-ário com ruído.

Se o sistema é suposto entregar $Y = y_k$ quando $X = x_k$, então as *probabilidades de erro* dos símbolos são dadas por $P(y_j|x_i)$ para $j \neq i$. Por outro lado, ao receber-se um determinado símbolo y_j, a incerteza quanto àquele que lhe deu origem ter sido o símbolo x_i, depende da probabilidade condicional $P(x_i|y_j)$, a que corresponde uma *perda* de informação que designaremos por $I(x_i|y_j)$ medida em *bits.*

A quantidade de informação transferida (recebida no destino) quando é transmitido o símbolo x_i e recebido o símbolo y_j é chamada de *informação mútua* e dada então por

$$I(x_i, y_j) \;=\; I(x_i) \;-\; I(x_i|y_j) \tag{8.21}$$

em que $I(x_i) = I_i$ é a auto-informação do símbolo x_i produzido na fonte. Em função das probabilidades tem-se

$$I(x_i, y_j) \;=\; \log_2 \frac{1}{P(x_i)} - \log_2 \frac{1}{P(x_i|y_j)} \tag{8.22}$$

Determinemos agora os valores médios da informação que, por consistência com a notação que se adoptou, vamos designar por *entropia*, $\mathcal{H}$.

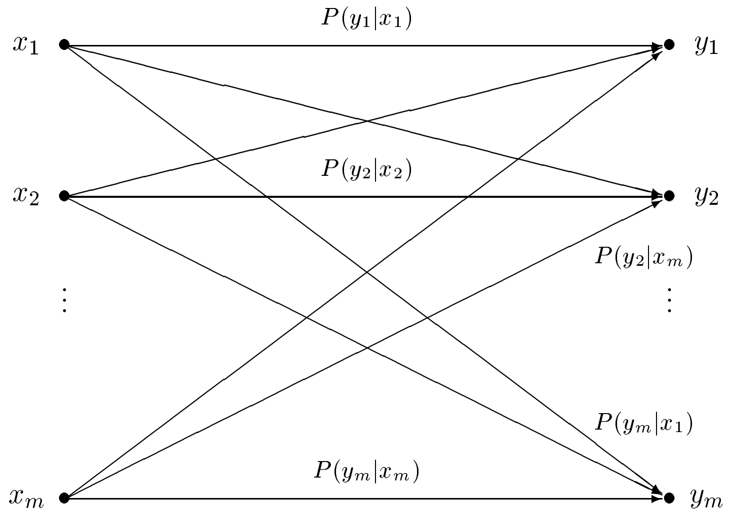

Figura 8.6: Probabilidades de transição para um canal discreto com ruido

Estarão em causa várias entropias.

Assim, o valor médio da informação recebida num símbolo y_j em relação a todos os possíveis símbolos produzidos pela fonte será dado pela média estatística

$$\begin{aligned} I(X, y_j) &= \sum_{i=1}^{m} P(x_i) \log_2 \frac{1}{P(x_i)} - \sum_{i=1}^{m} P(x_i|y_j) \log_2 \frac{1}{P(x_i|y_j)} \\ &= \mathcal{H}(X) - \mathcal{H}(X|y_j) \end{aligned} \tag{8.23}$$

A informação média recebida por símbolo, considerados todos os possíveis símbolos do alfabeto do destino é dada novamente pela média estatística

$$\begin{aligned} I(X,Y) &= \mathcal{H}(X) - \sum_{j=1}^{m} P(y_j)\, \mathcal{H}(X|y_j) \\ &= \mathcal{H}(X) - \sum_{j=1}^{m}\sum_{i=1}^{m} P(y_j)\, P(x_i|y_j) \log_2 \frac{1}{P(x_i|y_j)} \\ &= \mathcal{H}(X) - \underbrace{\sum_{j=1}^{m}\sum_{i=1}^{m} P(x_i\, y_j) \log_2 \frac{1}{P(x_i|y_j)}}_{\mathcal{H}(X|Y)} \end{aligned} \tag{8.24}$$

A informação média recebida não é mais do que a informação mútua média, ou seja a *entropia mútua* $\mathcal{H}(X,Y) \equiv I(X,Y)$

$$\mathcal{H}(X,Y) = \mathcal{H}(X) - \mathcal{H}(X|Y) \tag{8.25}$$

Se agora se tomar em consideração que $P(x_i)\,P(y_j|x_i) = P(y_j)\,P(x_i|y_j)$ e que portanto $I(X,Y) = I(Y,X)$, pode trocar-se X e Y na equação 8.25 obtendo-se

$$\mathcal{H}(X,Y) = \mathcal{H}(Y) - \mathcal{H}(Y|X) \tag{8.26}$$

tendo-se introduzido três novas quantidades com significado físico bem definido

$$\mathcal{H}(X|Y) = \sum_{j=1}^{m}\sum_{i=1}^{m} P(x_i\,y_j)\,\log_2 \frac{1}{P(x_i|y_j)} \tag{8.27}$$

$$\mathcal{H}(Y|X) = \sum_{i=1}^{m}\sum_{j=1}^{m} P(x_i\,y_j)\,\log_2 \frac{1}{P(y_j|x_i)} \tag{8.28}$$

$$\mathcal{H}(Y) = \sum_{j=1}^{m} P(y_j)\,\log_2 \frac{1}{P(y_j)} \tag{8.29}$$

em que

$\mathcal{H}(X\|Y)$	é a equivocação;
$\mathcal{H}(Y\|X)$	é a entropia do ruído;
$\mathcal{H}(Y)$	é a entropia do destino.

A equação 8.25 diz que a transferência de informação média por símbolo é igual à entropia da fonte menos a equivocação. A equivocação representa a informação perdida no canal com ruído. A equação 8.26 diz que a informação transferida é também igual à *entropia no destino* $\mathcal{H}(Y)$ menos a *entropia do ruído* $\mathcal{H}(Y|X)$ introduzida pelo canal. A interpretação de $\mathcal{H}(Y|X)$ como entropia do ruído segue-se à observação anterior segundo a qual o conjunto das probabilidades de transição directas $P(y_j|x_i)$ inclui as probabilidades de erro dos símbolos.

A entropia mútua pode então obter-se de duas maneiras: ou pela equação 8.25 ou pela equação 8.26.

Exêmplo 8.7 – Canal binário simétrico

Consideremos o canal binário simétrico *(BSC*[3]*) cujo modelo está representado na figura 8.7. A fonte produz dois símbolos com probabilidades*

$$P(x_1) = p \qquad P(x_2) = 1 - p$$

[3]BSC – sigla para a denominação inglesa de *Binary Simmetric Channel*

e no destino há dois possíveis símbolos com probabilidades de transição directa

$$
\begin{aligned}
P(y_1|x_2) &= P(y_2|x_1) = \alpha \\
P(y_1|x_1) &= P(y_2|x_2) = 1 - \alpha
\end{aligned}
$$

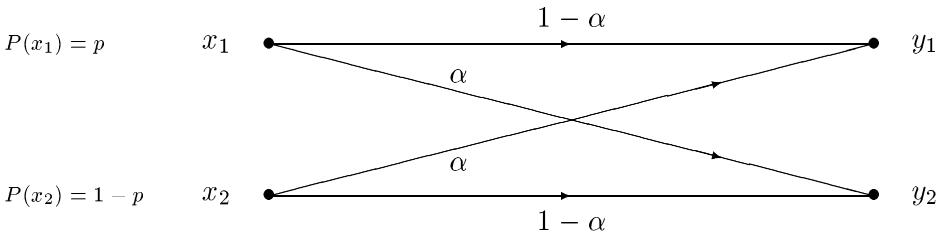

Figura 8.7: Modelo do canal binário simétrico

Este modelo representa qualquer sistema de transmissão binário no qual os erros são estatísticamente independentes e as probabilidades de erro são as mesmas para ambos os símbolos e portanto a probabilidade média de erro por símbolo é dada por

$$P_e = P(x_1)\,P(y_2|x_1) + P(x_2)\,P(y_1|x_2) = p\alpha + (1-p)\alpha = \alpha$$

Dado conhecerem-se as probabilidades de transição directas, pode utilizar-se a equação $\mathcal{H}(X,Y) = \mathcal{H}(Y) - \mathcal{H}(Y|X)$ para calcular a informação média mútua, ou entropia mútua em termos de p e α. Esta será a informação média transferida para o destino.

A entropia no destino $\mathcal{H}(Y)$ pode ser obtida tratando a saida do canal como uma fonte binária com probabilidades de símbolos $P(y_1)$ e $P(y_2) = 1 - P(y_1)$, isto é

$$\mathcal{H}(Y) = \Omega[P(y_1)]$$

em que $\Omega()$ é a função de entropia binária definida pela equação 8.8 e onde

$$P(y_1) = P(x_1)\,P(y_1|x_1) + P(x_2)\,P(y_1|x_2) = \alpha + p - 2\alpha p$$

O cálculo da entropia do ruído $H(Y|X)$ faz-se a partir da equação 8.28 que dá

$$\mathcal{H}(Y|X) \;=\; \Omega(\alpha)$$

resultado que é independente de p devido à simetria do canal.

Pode-se agora escrever uma expressão para a entropia mútua em função das expressões obtidas

$$\mathcal{H}(X,Y) \;=\; \Omega(\alpha + p - 2\alpha p) \;-\; \Omega(\alpha) \tag{8.30}$$

o que mostra que a transferência de informação através de um canal binário simétrico depende tanto da probabilidade de erro α como da probabilidade da fonte p. Se o ruído fôr pequeno, então $\alpha \ll 1$ e $\mathcal{H}(X,Y) \approx \Omega(p) = \mathcal{H}(X)$. Se o ruído fôr muito grande, então $\alpha = \frac{1}{2}$ e $\mathcal{H}(X,Y) = 0$ (resolver os problemas 8.15 e 8.16).

Capacidade do canal discreto

Cada canal possui normalmente alfabetos de origem e destino fixos e probabilidades de transição directas também fixas dado o sistema de transmissão completo, quando já em exploração, ser constituido por equipamento bem determinado resultante do respectivo projecto e funcionar com parâmetros fixos tais como o tipo de modulação, a potência do sinal, o meio físico de transmissão utilizado, etc. Deste modo, as únicas quantidades variáveis em $\mathcal{H}(X,Y)$ são as probabilidades da *fonte* $P(x_i)$. Consequentemente, para que a *transferência de informação* seja *máxima* é necessária uma determinada estatística para a fonte que poderá ser obtida, por exêmplo, através da codificação da fonte. Seja C_c o valor máximo de $\mathcal{H}(X,Y)$ assim obtido, isto é,

$$C_c \;\stackrel{def}{=}\; \max_{P(x_i)} \mathcal{H}(X,Y) \quad \text{bits/símbolo} \tag{8.31}$$

C_c representa a máxima quantidade de informação transferida por símbolo do canal e que é designada por *capacidade do canal*. É mais frequente, porém, medir a capacidade de um canal em termos do débito de informação. Assim, se fôr r_c o ritmo máximo de símbolos permitido no canal, então a sua capacidade é dada por

$$C \;=\; r_c\, C_c \quad \text{bits/seg} \tag{8.32}$$

que representa o ritmo máximo de transferência de informação. A importância do conceito de capacidade do canal torna-se mais evidente no contexto do *teorema fundamental da codificação do canal*, devido a Shannon, que se referiu no início deste capítulo e cujo enunciado é o seguinte:

Teorema 8.2 – **Teorema de Shannon da Codificação do Canal**
Se um canal possui capacidade C e uma fonte produz informação a um débito $R \leq C$, então existe um método de codificação para a transmissão pelo canal tal que a saída da fonte pode ser transmitida por esse canal com uma frequência de erros arbitráriamente pequena. Pelo contrário, se $R > C$, então não é possível transmitir a informação sem erros.

De notar a extrema importância deste teorema dado que ele *promete* a possibilidade de transmissão isenta de erros bastando para isso que $R \leq C$.

A sua demonstração geral está fora do âmbito deste curso pelo que terminaremos esta secção discutindo o princípio geral da codificação do canal. Este assunto será abordado em mais detalhe no capítulo que se segue.

Codificação para o canal binário simétrico

Consideremos o sistema de codificação representado na figura 8.8. O Canal Binário Simétrico (BSC) possui uma capacidade $C_c = 1 - \Omega(\alpha)$ com $\alpha < 1/2$ (ver Problema 8.17) e um ritmo de símbolos r_c simb do canal/s. A fonte, com entropia $\mathcal{H}(X)$, produz informação a um débito de $\mathcal{R}$ bits/s. De modo a maximizar a informação mútua, um codificador binário da

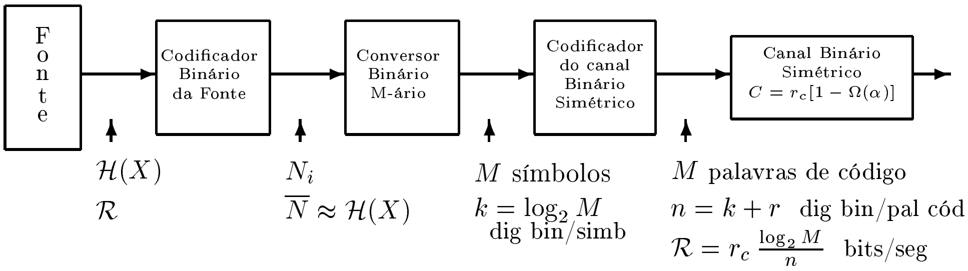

Figura 8.8: Codificação para um canal binário simétrico (BSC)

fonte óptimo opera sobre os símbolos da fonte e produz dígitos binários equiprováveis. A codificação do canal para control de erros é efectuada em dois estágios. No primeiro, os dígitos binários do código da fonte (os dígitos de informação) são agrupados em blocos de k dígitos produzindo assim $M = 2^k$ símbolos distintos (símbolos M-ários representados por $\log_2 M = k$ dígitos binários). No segundo, o codificador do canal binário simétrico, representa cada símbolo M-ário numa *palavra de* *código do canal*, com n dígitos binários por adição de r dígitos binários redundantes[4] $(n = k + r)$.

A informação média por dígito binário do canal é $\frac{\log_2 M}{n}$ bits/dig bin, sendo estes dígitos gerados ao ritmo de símbolos do canal, r_c. Podemos, pois, expressar o débito de informação da fonte por

$$\mathcal{R} \;=\; r_c \, \frac{\log_2 M}{n} \quad \text{bits/s} \tag{8.33}$$

em que M e n são os parâmetros do codificador do canal. O teorema de Shannon exige que $R \leq C = r_c \, C_c$ o que equivale a $\frac{\log_2 M}{n} \leq C_c$. Os parâmetros M e n devem portanto estar relacionados por

$$M \;=\; 2^{n\,(C_c - \epsilon)} \tag{8.34}$$

com $0 \leq \epsilon < C_c$. Pode mostrar-se que ϵ pode ser arbitráriamente pequeno e portanto $R \to C$ e que uma codificação do canal apropriada permite recuperar os símbolos M-ários no destino com uma probabilidade de erro arbitráriamente pequena, desde que o comprimento n da palavra de código seja *muito grande*. De facto, ter-se-á eventualmente que fazer $n \to \infty$ para garantir uma transmissão isenta de erro. A codificação ideal do canal binário simétrico resulta assim num *tempo de atraso infinito*, pelo que qualquer sistema prático de codificação, devendo ter necessáriamente um tempo de atraso limitado e um comprimento de palavra finito, ficará sempre aquém do desempenho ideal.

Por outro lado, Shannon não propôs qualquer algoritmo explícito para a determinação das palavras de código.

8.4.2 Canais contínuos

Começaremos esta secção pela definição da medida de informação produzida por uma fonte que emite mensagens sob a forma de sinais contínuos. Seguidamente, baseando-nos em hipóteses razoáveis àcerca da transmissão de sinais contínuos, chegaremos a uma expressão para a capacidade do canal contínuo em função da largura de banda e da razão de potências sinal-ruído, resultado que é conhecido por *lei de Hartley-Shannon*. Esta lei, que define o sistema de comunicação ideal, serve de padrão para a avaliação comparativa do desempenho de diferentes sistemas de comunicação e de termo de referência quando se projectam esses sistemas.

[4] redundantes no sentido de não transportarem informação da fonte

Informação contínua

Uma fonte de informação contínua produz um *sinal* variável no tempo $x(t)$. Consideremos os possíveis sinais como um conjunto de formas de onda geradas por algum processo aleatório que supomos ser ergódico[5]. Suponhamos ainda que o processo tem largura de banda limitada, B_x, o que significa, pelo teorema da amostragem, que $x(t)$ fica completamente caracterizado pelos *valores das amostras* obtidas periódicamente a uma frequência de amostragem não inferior a $2\,B_x$ ($f_a \geq 2\,B_x$). Assim, em qualquer instante de amostragem, o conjunto dos possíveis valores de amplitude das amostras constitui uma *variável aleatória contínua* X_x descrita por uma função de densidade de probabilidade $p_X(x)$.

A quantidade média de informação por amostra de $x(t)$ é medida pela função de entropia

$$\mathcal{H}(X_x) \stackrel{def}{=} \int_{-\infty}^{\infty} p_X(x)\, \log_2 \frac{1}{p_X(x)}\, dx \tag{8.35}$$

equação que é semelhante à que define a entropia de uma fonte discreta (equação 8.3), onde se substituiram o sinal de somatório pelo de integral e as probabilidades P_i pela *f.d.p.* $p_X(x)$. Contudo, a equação 8.35 representa uma medida *relativa* da informação e não uma medida absoluta. A entropia absoluta de uma fonte contínua é sempre infinita o que é razoável de admitir devido ao facto de X_x ser uma variável aleatória contínua e portanto poder tomar um número incontável de valores.

A entropia relativa, sendo finita, é uma medida útil de informação de fontes contínuas se se evitarem comparações com diferentes sinais de referência.

A questão que se coloca no caso das fontes contínuas é semelhante à que se colocou no caso das fontes discretas: Qual a função de densidade de probabilidade $p_X(x)$ que maximiza o valor da entropia $\mathcal{H}(X_x)$ para uma dada fonte?

O resultado que então se obteve (secção 8.1.2) — expresso na relação 8.4 e demonstrado no problema 6 — foi que $\mathcal{H}(X)$ era máxima quando os símbolos eram estatísticamente independentes, tendo-se concluido que esse máximo era obtido com $P_i = 1/m$.

Podemos então afirmar que se a variável aleatória contínua X_x toma um grande número de valores e estes são estatisticamente independentes e de

[5] o que permite tomar as médias temporais pelas médias de conjunto

valor quadrático médio limitado — os valores das amplitudes das amostras do sinal $x(t)$ — então, pelo *teorema do limite central*, a função de densidade de probabilidade $p_X(x)$ dessa variável é Gaussiana de média nula e é a que maximiza $\mathcal{H}(X_x)$

$$p_X(x) \;=\; \frac{1}{\sqrt{2\pi\sigma_x^2}}\, e^{-\frac{x^2}{2\sigma_x^2}} \tag{8.36}$$

O valor quadrático médio de X_x não é mais do que a potência média, $S = \sigma_x^2$, do sinal $x(t)$

$$S \;=\; \overline{x^2} \;=\; \int_{-\infty}^{+\infty} x^2\, p_X(x)\, dx \tag{8.37}$$

que tem de ser finita.

O valor do máximo da entropia obtem-se de seguida substituindo o valor de $p_X(x)$ na equação 8.35, dando

$$\mathcal{H}(X_x)_{max} \;=\; \frac{1}{2}\,\log_2(2\pi e S) \tag{8.38}$$

Capacidade do canal contínuo

A transferência de informação num canal contínuo toma a forma de transmissão de um sinal. A fonte emite um sinal $x(t)$ o qual, depois de corrompido pelo ruído de transmissão, chega ao destino sob a forma de um outro sinal $y(t)$. A *informação mútua média*, ou *entropia mútua*, é definida por analogia com o caso discreto (ver equação 8.26)

$$\mathcal{H}(X_x, Y_y) \;=\; \mathcal{H}(Y_y) \;-\; \mathcal{H}(Y_y|X_x) \tag{8.39}$$

em que nos interessa tomar esta forma dado que, tal como no caso discreto, normalmente conhecemos explícitamente a $f.d.p$ de transição directa $p_Y(y|x)$ e não a inversa $p_X(x|y)$.

Dado que as amplitudes do ruído, $n(t)$, são estatísticamente independentes e ele se manifesta aditivamente ao sinal, isto é,

$$y(t) \;=\; x(t) \;+\; n(t) \tag{8.40}$$

a entropia do ruído é definida por uma expressão equivalente à da equação 8.35

$$\mathcal{H}(Y_y|X_x) \;\stackrel{def}{=}\; \int_{-\infty}^{\infty} p_N(n)\, \log_2 \frac{1}{p_N(n)}\, dn \tag{8.41}$$

em que $p_N(n)$ é a função de densidade de probabilidade do ruído que será também Gaussiano

$$p_N(n) \;=\; \frac{1}{\sqrt{2\pi\sigma_n^2}}\; e^{-\frac{n^2}{2\sigma_n^2}} \tag{8.42}$$

O valor quadrático médio do sinal de ruído $n(t)$ representa a sua potência média, $N = \sigma_n^2$.

O valor da entropia do ruído obtem-se de igual forma substituindo o valor de $p_N(n)$ na equação 8.41

$$\mathcal{H}(Y_y|X_x) \;=\; \frac{1}{2}\,\log_2(2\pi e N) \tag{8.43}$$

Da equação 8.40 deduz-se que o sinal de destino $y(t)$ terá potência média $\overline{y^2} = S + N$ pelo que $\mathcal{H}(Y_y)$ se obtem da equação 8.38 substituindo S por $S + N$

$$\mathcal{H}(Y_y) \;=\; \frac{1}{2}\,\log_2[2\pi e(S+N)] \tag{8.44}$$

A máxima quantidade de informação transferida por amostra de $y(t)$ dá a *capacidade do canal contínuo.*

$$\begin{aligned}
C_c \;&\stackrel{def}{=}\; \max_{p_X(x)}\; \mathcal{H}(X_x, Y_y) \\
&=\; \max_{p_X(x)}\; \{\mathcal{H}(Y_y) \;-\; \mathcal{H}(Y_y|X_x)\} \\
&=\; \max_{p_X(x)} \left\{\frac{1}{2}\,\log_2[2\pi e(S+N)] \;-\; \frac{1}{2}\,\log_2[2\pi e N]\right\} \qquad (8.45) \\
&=\; \frac{1}{2}\,\log_2\left(1 + \frac{S}{N}\right) \qquad \text{bits/amostra} \qquad (8.46)
\end{aligned}$$

Se o canal de transmissão tem uma *largura de banda* B_T *Hz* então o ritmo máximo de transmissão de amostras é, como se sabe,

$$r_c \;=\; 2\,B_T \tag{8.47}$$

donde se obtem a capacidade do canal contínuo como sendo o máximo do débito médio de transferência de informação por amostra, tal como se definiu em 8.32 com $C = r_c\,C_c$:

$$C \;=\; B_T\,\log_2\left(1 + \frac{S}{N}\right) \qquad \text{bits/seg} \tag{8.48}$$

8.5 Problemas

8.1 – Uma fonte emite uma de quatro mensagens possíveis m_1, m_2, m_3 e m_4 com probabilidades $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$ e $\frac{1}{8}$ respectivamente. Calcular a quantidade de informação de cada mensagem e a informação média por mensagem.

8.2 – Uma carta é tirada de um baralho de cartas de jogo.

a) É informado que a carta que tirou é uma espada. Quanta informação recebeu?

b) Quanta informação recebe se lhe fôr dito que a carta que tirou é um ás?

c) Quanta informação recebe se lhe fôr dito que a carta que tirou é um ás de espadas? Verifique a relação que existe entre este resultado e os obtidos em *a)* e *b)*.

8.3 – Uma fonte emite uma sequência independente de símbolos de um alfabeto consistindo em 5 símbolos A, B, C, D e E com com probabilidades de ocorrência dos símbolos $\frac{1}{4}$, $\frac{1}{8}$, $\frac{1}{8}$, $\frac{3}{16}$ e $\frac{5}{16}$ respectivamente. Determinar a entropia desta fonte.

8.4 – Calcular o débito de entropia de uma fonte telegráfica que emite pontos e traços com probabilidades de ocorrência e tempos de duração do ponto e do traço respectivamente $P_p = \frac{2}{3}$, $\tau_p = 0.2 \quad s$, $P_t = \frac{1}{3}$, $\tau_t = 0.4 \quad s$. Qual o tempo médio de duração de um símbolo desta fonte?

8.5 – Uma fonte de dados tem oito símbolos que são produzidos em blocos de três a um débito de 1000 blocos por segundo. O primeiro símbolo de cada bloco é sempre o mesmo (presumivelmente para sincronismo) e os dois seguintes podem ser preenchidos por quaisquer dos oito símbolos com igual probabilidade. Qual é o débito de entropia, $\mathcal{R}$, da fonte?

8.6 – Mostrar que $\sum_{j=1}^{m} P_j \log_2(1/mP_j) = \mathcal{H}(X) - \log_2 m$ e utilizar este resultado e o facto de que $\ln(\alpha) \leq \alpha - 1$ para provar que $\mathcal{H}(X) \leq \log_2 m$.

8.7 – Uma fonte de informação tem um alfabeto de dimensão m. Um dos símbolos tem probabilidade ϵ enquanto os outros são igualmente prováveis. Determinar $\mathcal{H}(X)$ em termos de m e ϵ.

8.8 – Suponha que uma fonte tem $m = 3$ símbolos com probabilidades $P_1 = p$ e $P_2 = P_3$. Mostre que $\mathcal{H}(X) = \Omega(p) + 1 - p$. Esboce $\mathcal{H}(X)$ em função de p.

8.9 – Uma fonte de dados binária produz símbolos 0 e 1 com $p_0 = \frac{3}{8}$ e $p_1 = \frac{5}{8}$ e a influência entre símbolos em grupos de dois símbolos sucessivos é tal que $p_{1/0} = \frac{3}{4}$ e $p_{0/1} = \frac{1}{16}$. Calcular a entropia real e comparar com a que se obteria se os símbolos fossem independentes entre si.

8.10 – O código Morse internacional utiliza uma sequência de pontos e traços para transmitir letras do alfabeto alfanumérico. O traço é representado por um pulso de três unidades de tempo de duração e o ponto por um pulso de uma unidade de tempo de duração. A probabilidade de ocorrência de um traço é $\frac{1}{3}$ da probabilidade de ocorrência de um ponto.

a) Calcular o conteúdo de informação de um ponto e de um traço.

b) Calcular a informação média do código.

c) Suponha que o traço dura 1 ms que é o mesmo intervalo de tempo que a pausa entre símbolos. Determinar o débito médio de transmissão de informação.

8.11 – Aplicar o algoritmo de Huffman à fonte do exêmplo 8.3. O resultado será idêntico ao código IV. Confirmar que este código possui a propriedade óptima $N_i = I_i$ e que os dígitos binários 0 e 1 são equiprováveis.

8.12 – Uma fonte emite sequências independentes de símbolos de um alfabeto contendo cinco símbolos com probabilidades 0.4, 0.2, 0.2, 0.1 e 0.1.

a) Calcule a entropia da fonte.

b) Defina um codificador para a fonte por blocos de $K = 2$.

8.13 – Determinar um código de comprimento variável para a fonte do problema 8.9

a) Com rendimento não inferior a 60%.

b) Com rendimento não inferior a 80%.

c) Qual a compressão conseguida em cada caso?

8.14 – Determine qual o comprimento mínimo do bloco em que deve codificar a fonte do exêmplo 8.6 por forma a obter um rendimento da codificação de pelo menos 0.93. Determine um código de Huffman para esses blocos e a compressão que assim obtém.

8.15 – Confirmar que $\mathcal{H}(Y|X) = \Omega(\alpha)$ para o canal binário simétrico.

8.16 – Considere um canal com a propriedade de x_i e y_j serem estatísticamente independentes $\forall i, j$. Mostrar que $\mathcal{H}(X|Y) = \mathcal{H}(X)$ e que $\mathcal{H}(X,Y) = 0$. Que significa, físicamente, x_i e y_j serem estatísticamente independentes?

8.17 – Considere o canal binário não-simétrico representado na figura 8.9 em que as probabilidades dos símbolos 0 e 1 chegarem ao destino errados são respectivamente α e β $\quad (\alpha, \beta < 1)$.

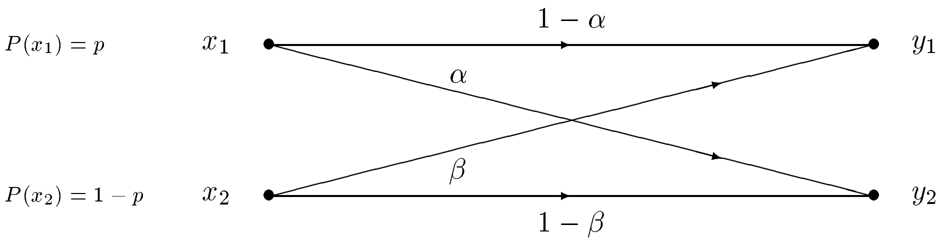

Figura 8.9: Canal binário não-simétrico

a) Determine a entropia na fonte $\mathcal{H}(X)$, a entropia no destino $\mathcal{H}(Y)$, a equivocação $\mathcal{H}(X|Y)$ e a entropia do erro $\mathcal{H}(Y|X)$, quando $p(x=0) = \frac{1}{4}$, $p(x=1) = \frac{3}{4}$, $\alpha = 0.25$ e $\beta = 0.1$.

b) Determine a capacidade do canal para $\alpha = 0.25$ e $\beta = 0.1$.

c) Determine a capacidade do canal binário simétrico $(\alpha = \beta)$.

8.18 – Mostre que a entropia mútua no canal binário não-simétrico do problema 8.17 é dada por

$$\mathcal{H}(X,Y) \;=\; \Omega[\beta + (1-\alpha-\beta)p] \;-\; p\Omega(\alpha) \;-\; (1-p)\Omega(\beta)$$

8.19 – Diz-se que o canal do problema 8.17 é *inútil* se $\beta = 1-\alpha$. Justifique esta afirmação utilizando argumentos intuitivos e depois analíticos.

8.20 – Se a e b forem constantes, mostre que

$$\frac{d}{dp}\Omega(a+bp) \;=\; b\,\log_2 \frac{1-a-bp}{a+bp}$$

e aplique esta relação para confirmar que $\mathcal{H}(X,Y)$ na equação 8.30 é máxima quando $p = 1/2$.

8.21 – O canal do problema 8.17 é chamado um canal Z quando $\beta = 0$. Utilize a relação do problema 8.20 para mostrar que se $\alpha = 1/2$, então $\mathcal{H}(X,Y)$ é máxima quando $p = 2/5$. Depois calcule C_c.

8.22 – Suponha que no canal do problema 8.17 se tem $\alpha = 1/4$ e $\beta = 0$. Utilize a relação do problema 8.20 para calcular C_c.

8.23 – Determine a capacidade do canal ternário simétrico no qual a probabilidade de erro por símbolo é de 2α $(2\alpha < 1)$.

fim do capítulo 8