8 Teoria da Informação: Capacidade do Canal de Transmissão

Comecemos por recordar a introdução ao capítulo 1: o estudo de um sistema de comunicações digitais envolve dois aspectos cruciais:

1. a eficiência da representação da informação gerada pela fonte;
2. a taxa de transmissão à qual é possível enviar a informação com fiabilidade através de um canal ruidoso.

A teoria da informação estabelece os limites fundamentais associados às questões acima referidas. A saber:

I. número mínimo de unidades de informação binária (bit) por símbolo necessário para representar completamente a fonte;
II. o valor máximo da taxa de transmissão que garante fiabilidade da comunicação através de um canal ruidoso.

No capítulo 1 fizemos o estudo do primeiro problema no contexto do modelo de fonte discreta sem memória. Aqui iremos abordar a segunda questão estabelecendo o resultado referido em II usando o modelo de canal discreto sem memória que iremos introduzir na secção seguinte.

8.1 Canais Discretos sem Memória

Um canal discreto é um modelo estatístico que a um alfabeto de entrada representado por uma variável aleatória discreta X faz corresponder, de acordo com uma qualquer lei de transição, uma variável aleatória discreta Y. A variável aleatória discreta X modela o alfabeto de entrada do canal, e Y modela o alfabeto de saída. O modelo descrito está ilustrado na Figura 8.1. Note-se que, em geral, a cardinalidade do alfabeto de saída pode ser diferente da do alfabeto de entrada.

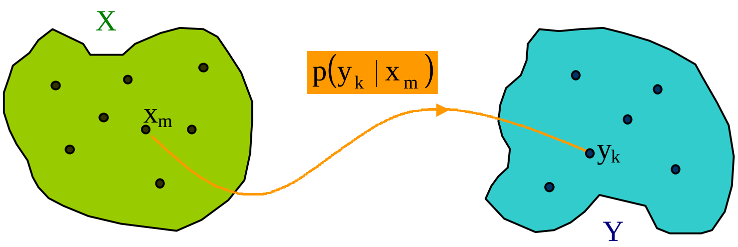

Figura 8.1: Modelo de um canal discreto

Def. 8.1- Canal discreto sem memória: um canal discreto sem memória é determinado pelo mecanismo estatístico que descreve o transporte de informação entre a fonte e o destinatário e que se define pelo conjunto de probabilidades condicionais

$$p(y_k | x_m),\ m = 1,2,\ldots,M;\ k = 1,2,\ldots,K. \quad (8.1)$$

Note-se que cada símbolo de saída só depende de um símbolo de entrada e não de uma sequência. Daí a designação de canal sem memória. Em geral,

$$\forall m = 1,2,\ldots,M: \ p(y_k \mid x_m) \neq 0, \ k \neq m,$$

ou seja, o processo de transmissão está sujeito a erros. A ocorrência de erros de transmissão decorre do facto de o canal ser ruidoso. Assim, à emissão do símbolo particular x_m pode corresponder a recepção de qualquer um dos símbolos do alfabeto de saída e, portanto,

$$\sum_{k=1}^{K} p(y_k \mid x_m) = 1, \ m = 1,2,\ldots,M. \tag{8.2}$$

A medida de fiabilidade da transmissão através de um canal discreto é dada pela probabilidade média de erro por símbolo. Seja

$$p(x_m) = P(X = x_m), \ m = 1,2,\ldots,M, \tag{8.3}$$

a probabilidade de o símbolo x_m ser transmitido. Naturalmente, ocorre um erro de transmissão se o símbolo recebido for um qualquer y_k, $k \neq m$, isto é,

$$P_e = \sum_{\substack{k=1 \\ k \neq m}}^{K} P(Y = y_k). \tag{8.4}$$

Sendo conhecidas as distribuições de probabilidade apriori (8.3) e de transição (8.1), então a partir da distribuição conjunta

$$p(y_k, x_m) = p(y_k \mid x_m) p(x_m), \ m = 1,2,\ldots,M; k = 1,2,\ldots,K, \tag{8.5}$$

podemos calcular a distribuição de probabilidade marginal da saída

$$p(y_k) = P(Y = y_k) = \sum_{m=1}^{M} p(y_k, x_m). \tag{8.6}$$

Usando (8.5) e (8.6) em (8.4), obtém-se

$$P_e = \sum_{\substack{k=1 \\ k \neq m}}^{K} \sum_{m=1}^{M} p(y_k \mid x_m) p(x_m). \tag{8.7}$$

Verificamos assim que, no caso de um canal discreto sem memória, a probabilidade média de erro de transmissão por símbolo é completamente determinada pela distribuição apriori e pelas probabilidades de transição (8.3) e (8.1), respectivamente.

Exemplo 8.1: Canal binário simétrico. Consideremos o caso do canal binário simétrico de grande interesse teórico e importância prática. Neste caso $K = M = 2$, o alfabeto de entrada é $x_0 = 0$ e $x_1 = 1$, e o de saída é $y_0 = 0$ e $y_1 = 1$. O canal é designado como simétrico porque a probabilidade p de receber 1 supondo ter sido transmitido 0 é igual à probabilidade de receber 0 supondo ter sido transmitido 1, como se ilustra no diagrama da Figura 8.2. Independentemente da distribuição apriori, $P_e = p$.

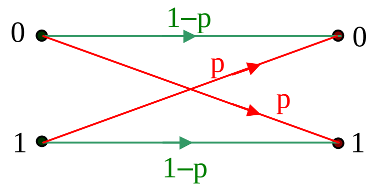

Figura 8.2: Canal binário simétrico

8.2 Informação Mútua e Informação Condicional

O conceito de entropia, introduzido no capítulo 1, pode ser estendido a alfabetos conjuntos.

Def. 8.2- Entropia conjunta. Sejam X e Y os alfabetos de entrada e de saída de um canal discreto sem memória, cuja distribuição de probabilidade conjunta é dada por (8.5). Então, a entropia conjunta dos alfabetos X e Y é

$$H(X,Y) = -\sum_{m=1}^{M}\sum_{k=1}^{K} p(x_m, y_k)\log_2 p(x_m, y_k). \tag{8.8}$$

Recordemos que $[-\log_2 p(x_m)]$ mede a incerteza inicial (apriori) associada à transmissão do símbolo x_m. Por outro lado, $[-\log_2 p(x_m \mid y_k)]$ mede a incerteza final sobre a transmissão do símbolo x_m após ter sido recebido o símbolo y_k. O ganho de informação sobre o símbolo de entrada x_m após observação do símbolo de saída y_k é então dado pela diferença entre a incerteza inicial e a incerteza final, ou seja,

$$I(x_m, y_k) = \log_2 p(x_m \mid y_k) - \log_2 p(x_m). \tag{8.9}$$

A relação anterior pode escrever-se na forma

$$I(x_m, y_k) = \log_2 \frac{p(x_m \mid y_k)}{p(x_m)}; \tag{8.10}$$

como é sabido,

$$\frac{p(x_m \mid y_k)}{p(x_m)} = \frac{p(y_k \mid x_m)}{p(y_k)} \tag{8.11}$$

e, portanto,

$$I(y_k, x_m) = \log_2 \frac{p(y_k \mid x_m)}{p(y_k)} \tag{8.12}$$

e

$$I(x_m, y_k) = I(y_k, x_m). \tag{8.13}$$

É devido a esta relação de simetria que a quantidade $I(x_m, y_k)$ (ou $I(y_k, x_m)$) é designada por informação mútua entre os acontecimentos $X = x_m$ e $Y = y_k$. A informação mútua média entre os alfabetos X e Y será então dada por

$$I(X,Y) = \sum_{m=1}^{M}\sum_{k=1}^{K} I(x_m, y_k)p(x_m, y_k). \tag{8.14}$$

Usando (8.10) e a identidade (equivalente a (8.11))

$$\frac{p(x_m \mid y_k)}{p(x_m)} = \frac{p(x_m, y_k)}{p(x_m)p(y_k)}$$

em (8.14), obtemos

$$I(X,Y) = \sum_{m=1}^{M}\sum_{k=1}^{K} p(x_m, y_k)\log_2\left[\frac{p(x_m, y_k)}{p(x_m)p(y_k)}\right]. \tag{8.15}$$

Para interpretar o conceito de informação mútua média, comecemos por recordar que $[-\log_2 p(x_m \mid y_k)]$ determina a incerteza final sobre a transmissão do símbolo x_m, uma vez observado o símbolo de saída y_k, isto é, representa a informação própria de x_m condicionada pela observação de y_k:

$$I(x_m \mid y_k) = -\log_2 p(x_m \mid y_k). \tag{8.16}$$

Def. 8.3- Entropia condicional. A entropia condicional mede, em termos médios, o acréscimo de informação sobre o símbolo de entrada do canal ganho pela observação do símbolo de saída:

$$H(X \mid Y) = -\sum_{m=1}^{M}\sum_{k=1}^{K} p(x_m, y_k)\log_2 p(x_m \mid y_k). \tag{8.17}$$

Rescrevendo (8.15)

$$\begin{aligned}
I(X,Y) &= \sum_{m=1}^{M}\sum_{k=1}^{K} p(x_m, y_k)[\log_2 p(x_m \mid y_k) - \log_2 p(x_m)] \\
&= \sum_{m=1}^{M}\sum_{k=1}^{K} p(x_m, y_k)\log_2 p(x_m \mid y_k) - \sum_{m=1}^{M}\log_2 p(x_m)\sum_{k=1}^{K} p(x_m, y_k) \\
&= \sum_{m=1}^{M}\sum_{k=1}^{K} p(x_m, y_k)\log_2 p(x_m \mid y_k) - \sum_{m=1}^{M}\log_2 p(x_m)p(x_m),
\end{aligned}$$

e comparando com (8.17), verificamos que a primeira parcela do lado direito é exactamente $-H(X \mid Y)$ enquanto que a segunda é a entropia do alfabeto de entrada. Temos então

$$\begin{aligned}
I(X,Y) &= H(X) - H(X \mid Y) \\
&= H(Y) - H(Y \mid X) = I(Y,X).
\end{aligned} \tag{8.18}$$

A informação mútua média mede, portanto, a incerteza média existente apriori sobre o símbolo de entrada do canal que é resolvida (também em média) pela observação da saída do canal. A relação anterior pode ainda escrever-se noutra forma. Com efeito, a expressão (8.8) da entropia conjunta pode rescrever-se como

$$\mathrm{H}(X,Y)=-\sum_{m=1}^{M}\sum_{k=1}^{K}p(x_m,y_k)\log_2\left[\frac{p(x_m,y_k)}{p(x_m)p(y_k)}p(x_m)p(y_k)\right],$$

ou ainda como

$$\begin{aligned}\mathrm{H}(X,Y)=&-\sum_{m=1}^{M}\sum_{k=1}^{K}p(x_m,y_k)\log_2\left[\frac{p(x_m,y_k)}{p(x_m)p(y_k)}\right]\\&-\sum_{k=1}^{K}\underbrace{\left(\sum_{m=1}^{M}p(x_m,y_k)\right)}_{p(y_k)}\log_2 p(y_k)\\&-\sum_{m=1}^{M}\underbrace{\left(\sum_{k=1}^{K}p(x_m,y_k)\right)}_{p(x_m)}\log_2 p(x_m);\end{aligned}$$

Identificando as parcelas uma a uma, concluímos que

$$\mathrm{I}(X,Y)=\mathrm{H}(X)+\mathrm{H}(Y)-\mathrm{H}(X,Y). \tag{8.19}$$

A Figura 8.3 ilustra de modo simbólico as relações entre as diversas quantidades introduzidas e que caracterizam a fonte, o canal, e a respectiva saída.

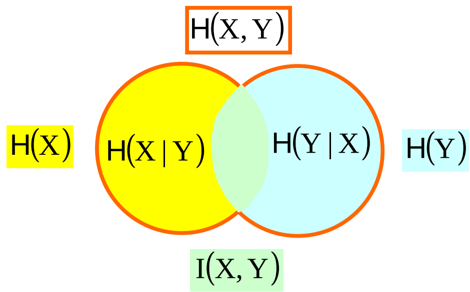

Figura 8.3: Relações entre as quantidades que caracterizam a fonte, o canal, e a saída do canal

8.3 Capacidade de um Canal Discreto sem Memória

Já anteriormente se fez notar que um canal discreto sem memória é caracterizado pelas probabilidades de transição definidas em (8.1). No entanto, a informação mútua média entre os alfabetos de entrada (fonte) e de saída do canal depende também da distribuição apriori definida em (8.3). Sendo o canal independente da fonte, e sendo a informação mútua média o ganho de informação sobre a entrada do canal após observação da correspondente saída, é natural pensar-se em optimizar a eficiência de utilização do canal, maximizando o ganho de informação atrás referido. Uma vez que as probabilidades de transição caracterizam o canal e estão fora do contolo de quem pretende desenhar o sistema de comunicações, o processo de optimização referido terá de assentar sobre a distribuição de probabilidade apriori.

Def. 8.4- Capacidade do canal discreto sem memória. A capacidade de um canal discreto sem memória é o máximo da informação mútua média, por cada utilização do canal, e relativamente a todas as possíveis distribuições de probabilidade apriori, isto é,

$$C = \max_{p(x_m),\, m=1,2,\ldots,M} I(X,Y). \tag{8.20}$$

A capacidade de canal mede-se em unidades binárias de informação (bit) por intervalo de sinalização (tempo de transmissão de um símbolo ou duração de cada utilização do canal).

Exemplo 8.2: Canal binário simétrico. Consideremos o canal binário simétrico já introduzido no Exemplo 8.1, Figura 8.2, e completamente especificado pelo parâmetro p, probabilidade de erro. Seja $p_0 = \Pr(X = x_0)$; obviamente $p_1 = \Pr(X = x_1) = 1 - p_0$. Usando estes dados em (8.18) e maximizando em ordem a p_0, conclui-se que

$$C = I(X,Y)\big|_{p_0=p_1=1/2}, \tag{8.21}$$

isto é, a informação mútua média é máxima quando os símbolos de entrada são equiprováveis e a capacidade deste canal vale

$$C = 1 + p\log_2 p + (1-p)\log_2(1-p). \tag{8.22}$$

Note-se que (8.22) se pode escrever como

$$C = 1 - H(p),$$

onde

$$H(p) = -p\log_2 p - (1-p)\log_2(1-p)$$

é a função (1.15) representada na Figura 1.1 do capítulo 1. Recordando essa figura, podemos concluir que

1. quando não existe ruído de canal, $p = 0$ e a capacidade atinge o seu valor máximo de um bit por cada intervalo de transmissão;
2. quando o canal é ruidoso, $p \neq 0$ e a capacidade atinge o valor mínimo de 0 bit por intervalo de transmissão quando $p = 1/2$.

8.4 Teorema da Codificação de Canal

O teorema da codificação de canal é um dos resultados mais importantes derivados por Shannon no contexto da Teoria da Informação. Este teorema estabelece um limite fundamental sobre o valor da taxa média de transmissão fiável de informação através de um canal ruidoso.

Teorema da Codificação de Canal
Considere-se uma fonte discreta sem memória com alfabeto X, entropia $H(X)$, e que gera um símbolo em cada T_s segundos. Consideremos também um canal discreto sem memória com capacidade C e que é usado uma vez em cada T_c segundos. Então, se

$$\frac{H(X)}{T_s} \leq \frac{C}{T_c} \tag{8.23}$$

existe um código segundo o qual a saída da fonte pode ser transmitida através do canal e reconstruída com probabilidade de erro arbitrariamente pequena. Ao contrário, se

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

não é possível garantir a reconstrução fiável da informação transmitida.

A demonstração deste teorema que não faremos aqui não é construtiva[1]. Isto é, apenas se garante a existência do código não o definindo. Embora o problema da codificação de canal venha a ser discutido mais tarde em maior detalhe, faz sentido introduzir as ideias fundamentais.

O objectico da codificação de canal é aumentar a resistência do sistema de comunicações digital face aos efeitos do ruído de canal. No caso particular dos códigos de bloco, a cada bloco de k bits da sequência binária gerada pela fonte faz-se corresponder um bloco de n bits (palavra de código) com $n > k$. Este processo de codificação deve ser concebido de modo que a descodificação tenha solução única. Note-se que do universo de 2^n blocos binários de comprimento n apenas 2^k são palavras de código (as que correspondem numa relação de um para um aos blocos binários de comprimento k gerados pela fonte). Este esquema está representado simbolicamente na Figura 8.4. No caso de a transmissão se efectivar sem erros, o processo de descodificação conduz ao bloco de comprimento k que havia sido gerado pela fonte. Quando ocorrem erros de transmissão, a palavra de comprimento n recebida pode não ser uma palavra de código e o erro é detectado e/ou corrigido. Daqui resulta naturalmente uma probabilidade de erro por bit inferior à que existiria se não fosse usada a codificação de canal. A codificação de canal pode ser encarada como o processo dual da codificação de fonte. Enquanto neste caso se elimina redundância para melhorar a eficiência, na codificação de canal controla-se a redundância introduzida para melhorar a fiabilidade.

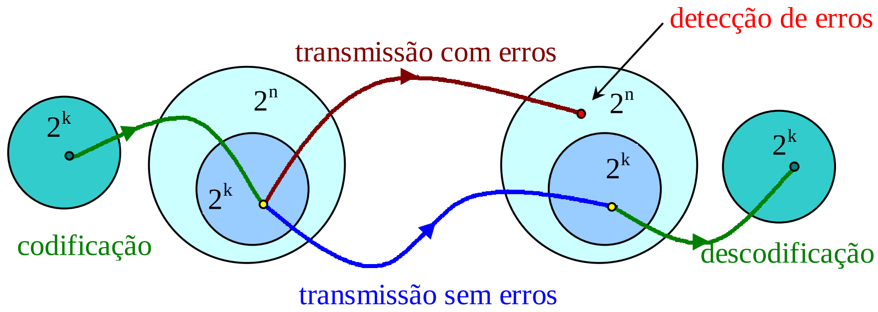

Figura 8.4: Codificação de canal

[1] Referência indicada no fim do capítulo 1.

Exemplo 8.3: Canal binário simétrico. Consideremos uma fonte binária sem memória que gera bits equiprováveis a uma taxa de um bit em cada T_s segundos, ou seja, $1/T_s$ bps. Neste caso a entropia da fonte é 1, pelo que a taxa de geração de informação é de $1/T_s$ bit/segundo[2]. A sequência fonte é codificada com base num codificador de razão $r = k/n$ e que produz um símbolo em cada T_c segundos. À saída do codificador a taxa de transmissão de dados é $1/T_c$ baud. O canal binário simétrico é assim usado uma vez em cada T_c segundos. Portanto, a capacidade do canal por unidade de tempo é C/T_c bit/segundo, onde C é dada por (8.22). De acordo com o teorema da codificação de canal, se

$$\frac{1}{T_s} \le \frac{C}{T_c}, \qquad (8.24)$$

então pode forçar-se a probabilidade de erro por bit a tomar um valor arbitrariamente pequeno através de uma codificação adequada. Como $r = k/n = T_c/T_s$, a condição (8.24) pode ser modificada para a forma equivalente

$$r \le C. \qquad (8.25)$$

Deste modo, podemos afirmar que existe um código com razão r que, verificando (8.25), garante uma probabilidade de erro por bit tão pequena quanto queiramos.

Suponhamos que o parâmetro que caracteriza o canal binário simétrico vale $p = 10^{-2}$ e que queremos construir um código para o qual $P_e \le 10^{-8}$. Usando aquele valor de p em (8.22), obtemos $C = 0.9192$ bit/segundo. Consideremos agora um código de repetição que para cada bit gerado pela fonte transmite $n = 2m + 1$ bits iguais, ou seja $r = 1/n$. Por exemplo, para $m = 1$, as palavras do código são 000 e 111. Suponhamos ainda que o descodificador funciona por maioria: escolhe 0 se a palavra recebida tiver mais 0's do que 1's e vice-versa. Se for gerado um 0, transmite-se 000, e o descodificador erra se receber 2 1's e um 0 ou 3 1's. Portanto, para o caso geral, a probabilidade de erro por bit vem dada por

$$P_e = \sum_{i=m+1}^{n} \binom{n}{i} p^i (1-p)^{n-i}. \qquad (8.26)$$

Calculando o valor de P_e para diversos valores de $r = 1/n$, podemos construir a Tabela 8.1.

r	1	1/3	1/5	1/7	1/9	1/11
P_e	10^{-2}	3×10^{-4}	10^{-6}	4×10^{-7}	10^{-8}	5×10^{-10}

Tabela 8.1: Probabilidade de erro média nos códigos de repetição

Verificamos assim que P_e decresce à medida que r diminui, e que se atinge o valor máximo especificado de $P_e = 10^{-8}$ quando $r = 1/9$. Repare-se que neste caso transmitimos 9 bits em representação de um único bit da fonte, o que se traduz no facto de $r = 0.1111 << 0.9192 = C$. Ao contrário, o que o teorema da codificação de fonte diz é que basta que $r < C$ (igual no limite). Portanto, os códigos de repetição não são deste ponto de vista os mais eficientes existindo, como veremos mais tarde, métodos de codificação mais adequados.

[2] A distinção entre taxa de geração de dados binários e taxa de geração de informação é muito importante no contexto aqui considerado. Note-se que aqui, ambas têm o mesmo valor pois os símbolos são equiprováveis.

8.5 Entropia Diferencial

O conceito de entropia, introduzido no contexto das variáveis aleatórias discretas, pode ser de algum modo generalizado para o caso das variáveis aleatórias contínuas desde que se tenham em conta alguns detalhes técnicos que discutiremos adiante.

Def. 8.5- Entropia diferencial. Consideremos uma fonte analógica modelada pelo processo $X(t)$. Seja X a variável aleatória contínua que descreve estatisticamente a amplitude $X(t)$ do referido processo em qualquer instante t, e $f_X(x)$ a respectiva densidade de probabilidade. A entropia diferencial de X é

$$h(X) = -\int_{-\infty}^{+\infty} f_X(x)\log_2[f_X(x)]dx\,. \qquad (8.27)$$

O termo entropia diferencial é usado para marcar as diferenças relativamente ao conceito de entropia de fontes discretas. Suponhamos que X tem uma distribuição uniforme

$$f_X(x) = \begin{cases} a^{-1}, & 0 \le x \le a \\ 0, & \text{caso contrário}; \end{cases}$$

substituindo em (8.27), obter-se-ia $h(X) = \log_2 a$, a qual toma valores negativos quando $a < 1$ e torna-se singular quando $a \to \infty$. Recorde-se que a entropia de uma fonte discreta é sempre positiva e limitada. Portanto, ao contrário da entropia de fontes discretas, a entropia diferencial definida para fontes analógicas não pode ser interpretada como uma medida de aleatoriedade ou de incerteza. Suponhamos que a variável aleatória X constitui a forma limite de uma variável aleatória discreta que toma os valores $x_k = k\Delta_x$, $k = \ldots,-1,0,1,\ldots$, e onde $\Delta_x > 0$ é arbitrariamente pequeno. Assim sendo, podemos assumir que

$$\Pr(x_k < X \le x_k + \Delta_x) \approx f_X(x_k)\Delta_x$$

e, no limite quando $\Delta_x \to 0$, a entropia da fonte contínua seria

$$\begin{aligned} H(X) &= \lim_{\Delta_x \to 0}\left[-\sum_{k=-\infty}^{+\infty} f_X(x_k)\Delta_x \log_2(f_X(x_k)\Delta_x)\right] \\ &= \lim_{\Delta_x \to 0}\left[-\sum_{k=-\infty}^{+\infty} f_X(x_k)\log_2(f_X(x_k))\Delta_x - \sum_{k=-\infty}^{+\infty} f_X(x_k)\log_2(\Delta_x)\Delta_x\right] \end{aligned}$$

ou, atendendo a (8.27),

$$H(X) = h(X) - \lim_{\Delta_x \to 0} \log_2(\Delta_x)\sum_{k=-\infty}^{+\infty} f_X(x_k)\Delta_x\,. \qquad (8.28)$$

Consideremos a segunda parcela do lado direito de (8.28) e notemos que para valores de Δ_x suficientemente pequenos se tem

$$f_X(x_k)\Delta_x < f_X(x_k)\Delta_x \left|\log_2(\Delta_x)\right|;$$

quando $\Delta_x \to 0$, o termo à direita nesta desiguladade torna-se *infinitamente* maior do que o termo da esquerda, pelo que, embora

$$\lim_{\Delta_x \to 0} \sum_{k=-\infty}^{+\infty} f_X(x_k)\Delta_x = \int_{-\infty}^{+\infty} f_X(x)dx = 1,$$

$$\lim_{\Delta_x \to 0} \log_2 \Delta_x \sum_{k=-\infty}^{+\infty} f_X(x_k)\Delta_x$$

não existe. Daqui se conclui que a entropia (8.28) de uma fonte contínua é *infinitamente grande*. Esta conclusão deriva do facto de ser infinita a quantidade de informação associada ao acontecimento $X = x_0, -\infty < x_0 < +\infty$, quando X é uma variável aleatória contínua. No entanto, e voltando a (8.28), a entropia diferencial pode ser interpretada como a entropia da fonte contínua medida relativamente à referência

$$-\lim_{\Delta_x \to 0} \log_2 \Delta_x \sum_{k=-\infty}^{+\infty} f_X(x_k)\Delta_x .$$

Neste contexto, supondo que X e Y são, respectivamente, a entrada e a saída de um canal sem memória, podemos generalizar o conceito de informação mútua entre X e Y. Em particular,

$$I(X,Y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{XY}(u,v)\log_2\left[\frac{f_{XY}(u,v)}{f_X(u)f_Y(v)}\right]dudv. \qquad (8.29)$$

Entre outros factos, pode mostrar-se que, sendo a entropia diferencia condicional dada por

$$h(X\,|\,Y) = -\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{XY}(u,v)\log_2\left[f_{X|Y}(u\,|\,Y=v)\right]dudv, \qquad (8.30)$$

então

$$\begin{aligned} I(X,Y) &= h(X) - h(X\,|\,Y) \\ &= h(Y) - h(Y\,|\,X) \end{aligned}. \qquad (8.31)$$

Como em (8.31), o termo de referência no cálculo da entropia é o mesmo, a informação mútua pode continuar a ser interpretada tal como no caso de fontes e canais discretos.

8.5.1 Máxima Entropia

Neste parágrafo, vamos resolver um problema cujos resultados serão necessários mais adiante, e cuja formulação se apresenta de seguida.

Calcular a função densidade de probabilidade f_X que maximiza

$$h(X) = -\int_{-\infty}^{+\infty} f_X(x) \log_2 [f_X(x)] dx \tag{8.27}$$

sob as restrições

$$\int_{-\infty}^{+\infty} f_X(u) du = 1 \tag{8.32}$$

e

$$\int_{-\infty}^{+\infty} (u - m_X)^2 f_X(u) du = \sigma^2 \tag{8.33}$$

onde

$$m_X = \int_{-\infty}^{+\infty} u \, f_X(u) du . \tag{8.34}$$

Para resolver este problema, vamos novamente recorrer à técnica dos multiplicadores de Lagrange, começando por determinar os pontos de estacionariedade da Lagrangeana

$$L(f_X) = -\int_{-\infty}^{+\infty} f_X(u) \log_2 [f_X(u)] du + \lambda_1 \left[\int_{-\infty}^{+\infty} f_X(u) du - 1 \right] + \lambda_2 \left[\int_{-\infty}^{+\infty} (u - m_X)^2 f_X(u) du - \sigma^2 \right]. \tag{8.35}$$

No entanto, é necessário ter em conta que o problema em causa exige a maximização de (8.35) relativamente a uma função definida sobre o conjunto dos números reais e não relativamente a um parâmetro. Suponhamos que f_X é a função que maximiza (8.27), cumprindo as restrições (8.32) e (8.33). Então, podemos definir

$$\tilde{f}_X(\cdot) = f_X(\cdot) + \varepsilon f_\varepsilon(\cdot) , \tag{8.36}$$

onde $\varepsilon f_\varepsilon(\cdot)$ representa uma perturbação relativamente à função maximizante $f_X(\cdot)$. Sublinhe-se que ε é um parâmetro real arbitrário, tal como $f_\varepsilon(\cdot)$. Os pontos de estacionariedade de (8.35) são aqueles que verificam a condição necessária de existência do máximo

$$\left. \frac{\partial L}{\partial \varepsilon} \right|_{\varepsilon = 0} = 0. \tag{8.37}$$

Se em (8.35), substituirmos $f_X(\cdot)$ por $\tilde{f}_X(\cdot)$ definido em (8.36), e em seguida usarmos (8.37), obteremos

$$\int_{-\infty}^{+\infty} f_\varepsilon(u) \left[-\log_2 e \cdot \ln f_X(u) - \log_2 e + \lambda_1 + \lambda_2 (u - m_X)^2 \right] du = 0 ,$$

o que, atendendo ao facto de $f_\varepsilon(\cdot)$ ser uma função arbitrária, só se verifica se o factor entre $[\ ]$ da integranda for nulo, ou seja

$$\ln f_X(u) = -1 + \frac{\lambda_1}{\log_2 e} + \frac{\lambda_2}{\log_2 e}(u - m_X)^2\ ,\ \forall u \in \mathbb{R}. \qquad (8.38)$$

Usando esta relação, podemos impor as restrições (8.32) e (8.33) e, após resolver o sistema de equações resultante, obtemos os multiplicadores de Lagrange

$$\lambda_1 = \frac{1}{2}\log_2\left(\frac{e}{2\pi\sigma^2}\right)$$
$$\lambda_2 = -\frac{\log_2 e}{2\sigma^2},$$

os quais, uma vez subsituídos em (8.38), permitem obter o resultado final

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x - m_X)^2}{2\sigma^2}}\ ,\ -\infty < x < +\infty.$$

Conclusão: a entropia diferencial da variável aleatória X, cuja variância é σ^2, é máxima se X for gaussiana. Por outras palavras, se X e Y forem variáveis aleatórias contínuas, ambas com a mesma variância, e se X for gaussiana, então

$$h(X) \geq h(Y).$$

8.6 Teorema da Capacidade de Canal

Consideremos o processo $X(t)$, estacionário, de média nula e com largura de banda B. Suponhamos que $X(t)$ é amostrado uniformemente com período T_s. Sejam $X_k = X(kT_s)$ as variáveis aleatórias que modelam estatisticamente as amostras $x(t)$ do processo $X(t)$ nos instantes $t = kT_s$. Admitamos que as amostras x_k são transmitidas através de um canal perturbado por ruído aditivo, branco e gaussiano, com espectro de potência constante e igual a $\eta/2$. Se o canal de transmissão tiver largura de banda B, então a respectiva saída é modelada pela variável aleatória

$$Y_k = X_K + N_k\ ,\ k = 0,1,\ldots,K-1, \qquad (8.39)$$

onde N_k é uma variável aleatória gaussiana, independente de X_k, com média nula e variância

$$\sigma^2 = \eta B. \qquad (8.40)$$

Admitamos ainda que

$$\mathbf{P}_X = E\{X_k^2\},\ k = 0,1,\ldots,K-1. \qquad (8.41)$$

Supondo que a amostragem é feita ao ritmo de Nyquist, então o número K de amostras transmitidas num intervalo de tempo com duração T é

$$K = 2BT\,. \tag{8.42}$$

Por analogia com (8.20), a capacidade do canal gaussiano será

$$C = \max_{f_{X_k}(u)} I(X_k, Y_k), \tag{8.43}$$

onde $I(X_k, Y_k)$ é a informação mútua entre X_k e Y_k, definida em (8.29) e que verifica (8.31), isto é,

$$I(X_k, Y_k) = h(Y_k) - h(Y_k \mid X_k)\,. \tag{8.44}$$

Note-se que, de acordo com (8.39) e (8.40), a variável aleatória $Y_k \mid X_k = x_k$ é gaussiana com média x_k e variância $\sigma^2 = \eta B$. Por outro lado, usando a definição de entropia diferencial (8.27), verifica-se facilmente que a entropia diferencial de uma gaussiana com variância σ^2 não depende do respectivo valor médio e vale $\log_2\left(2\pi e \sigma^2\right)/2$. Portanto, podemos escrever

$$I(X_k, Y_k) = h(Y_k) - h(N_k)\,. \tag{8.45}$$

Como N_k é independente de X_k, então o cálculo da capacidade do canal (8.43) envolve apenas a maximização de $h(Y_k)$, onde Y_k tem variância $\mathbf{P}_X + \eta B$. Este problema foi resolvido no parágrafo 8.5.1: $h(Y_k)$ é máxima com Y_k gaussiana. Temos então

$$h(Y_k) = \frac{1}{2}\log_2\left[2\pi e\left(\mathbf{P}_X + \eta B\right)\right] \tag{8.46}$$

e

$$h(N_k) = \frac{1}{2}\log_2\left[2\pi e \eta B\right]. \tag{8.47}$$

O valor máximo da informação mútua (8.45) resulta de (8.46) e (8.47) pelo que, tendo em conta (8.43), obtém-se a capacidade do canal gaussiano

$$C = \frac{1}{2}\log_2\left(1 + \frac{\mathbf{P}_X}{\eta B}\right) \text{ bit/utilização.}$$

Finalmente, como o canal é usado K vezes em T segundos, usando (8.42), vem

$$C = B\log_2\left(1 + \frac{\mathbf{P}_X}{\eta B}\right) \text{ bit/segundo.} \tag{8.48}$$

De acordo com o teorema da codificação de canal sabemos que, desde que se use o código adequado, é possível transmitir através do canal gaussiano com largura de banda B à taxa máxima C dada por (8.48) com probabilidade de erro arbitrariamente pequena. Para uma largura de banda B fixa, a taxa de transmissão de informação exequível aumenta com a razão $\mathbf{P}_X/\eta$. No entanto, se estes parâmetros se mantiverem fixos, aquela taxa de transmissão aumenta com a largura de banda B aproximando-se de um valor limite quando

$SNR = \mathbf{P}_X/(\eta B)$ se anula. De facto, usando a aproximação $\log_2(1+x) \approx (\log_2 e)^{-1} x$, $x \approx 0$, concluímos que quando $B \to \infty$, $SNR = \mathbf{P}_X/(\eta B) \to 0$, e $C \to (\log_2 e)^{-1} \mathbf{P}_X/\eta$. Portanto, mesmo que a largura de banda do canal cresça indefinidamente, a respectiva capacidade e, portanto, a taxa máxima de transmissão fiável de informação permanece limitada.

8.7 Sistema Ideal de Comunicações

Por hipótese, assumiremos que no sistema ideal a transmissão de informação é feita à taxa máxima, isto é, igual à capacidade C do canal gaussiano. Assim, se E for a energia dispendida na transmissão de cada unidade de informação, podemos dizer que a potência transmitida vale $\mathbf{P}_X = CE$. Assim sendo, (8.48) escreve-se na forma

$$\frac{C}{B} = \log_2\left(1 + \frac{E}{\eta} \cdot \frac{C}{B}\right) \ \mathrm{bit} \cdot \mathrm{s}^{-1} \cdot \mathrm{Hz}^{-1}. \qquad (8.49)$$

Note-se que C/B mede a eficiência espectral máxima do canal, isto é, dá o número máximo de unidades de informação transmitidas por unidade de tempo e por unidade de largura de banda do canal. A Figura 8.5 mostra o andamento da eficiência espectral em função de E/η.

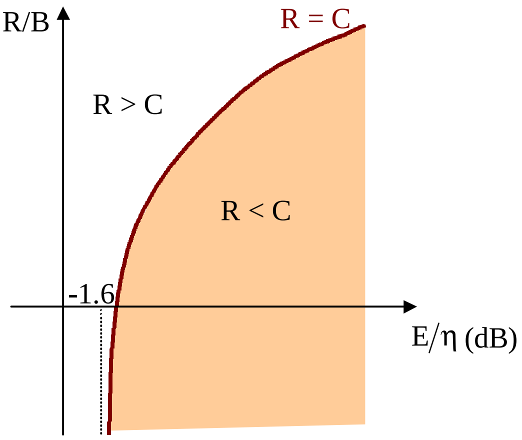

Figura 8.5: Eficiência espectral

A zona sombreada do plano, abaixo da curva $R = C$, constitui a região útil de utilização do canal gaussiano. Para o caso do sistema ideal, o limiar $E/\eta = -1.6\,\mathrm{dB}$ obtém-se invertendo (8.49) com $R = C$. Neste sistema, a probabilidade de erro é $P_e = 0$, se o canal for usado abaixo daquele limiar, isto é, se $E/\eta < -1.6\,\mathrm{dB}$. Ao contrário, se $E/\eta \geq -1.6\,\mathrm{dB}$, então $P_e = 1$.