Capítulo II

Princípios Básicos de Teoria da Informação

A área da ciência denominada Teoria da Informação responde, no contexto de Comunicações Digitais, duas questões fundamentais:

1- Até que limite é possível comprimir um conjunto de dados? (antecipação da resposta: até o valor da <u>Entropia</u> da informação expressa pelo conjunto).

2- Qual a maior taxa de transmissão de informação possível em um canal de transmissão para que não ocorram erros? (resposta: Capacidade do Canal – Teorema Fundamental de Shannon, discutido no Capítulo I).

No entanto, não só a Comunicações Digitais serve a Teoria da Informação. Contribuições fundamentais tem sido feitas, entre outras áreas, em física estatística (termodinâmica), em ciência da computação (complexidade de algoritmos), em inferência estatística (*Occam's Razor: "The simplest explanation is best"*) e em probabilidade e estatística (taxas de erro para estimação ótima).

2.1 Entropia – Uma Possível Medida de Informação

O conceito de informação é muito amplo para que possa ser capturado completamente por uma simples definição. No entanto, se estivermos tratando de uma variável aleatória discreta *X*, com espaço de amostras definido pelo conjunto $\Omega_X = \{m_k\} = \{m_0, m_1, \cdots, m_{M-1}\}$ de M eventos m_k com probabilidade de ocorrência p_k, $k = 0,1,\cdots,M-1$, é possível definir uma quantidade chamada de Entropia a qual apresenta muitas propriedades que concordam com noção intuitiva do que uma medida de informação deve expressar.

Intuitivamente, a observação da ocorrência de um evento do espaço amostral de uma variável aleatória nos dá <u>informação</u>. Eventos raros contém mais <u>informação</u> do que eventos comuns.

Por exemplo, nós inferimos e aprendemos muito pouco se alguém disser "O sol nasceu hoje pela manhã" ou disser "O sol nasceu à leste hoje pela manhã". Mas inferimos e aprendemos consideravelmente mais se alguém disser "São Paulo foi atingido por um terremoto hoje pela manhã" ou "São Paulo foi atingido por um furacão hoje pela manhã".

Em 1928 Hartley propôs uma medida logarítmica de medida de informação que reflete o raciocínio intuitivo do parágrafo anterior. Vamos supor que estamos registrando o valor das amostras na saída *X* do quantizador do Codificador de Fonte e que o quantizador apresente *M* níveis de quantização. Após o registro de um número suficiente de amostras é feito um estudo estatístico da probabilidade de ocorrência de cada uma das *M* possíveis amostras (ou mensagens de $N = \log_2 M$ bits sob o ponto de vista do código compressor), conforme discutido na Seção 1.2. A saída *X* do quantizador pode ser considerada uma variável aleatória discreta *X*, com espaço de amostras definido pelo conjunto

$\Omega_X = \{m_k\} = \{m_0, m_1, \cdots, m_{M-1}\}$ de M mensagens m_k com probabilidade de ocorrência p_k, $k = 0,1,\cdots,M-1$. Segundo Hartley, a <u>Auto-Informação</u> $h(m_k)$ implícita na ocorrência de uma mensagem m_k com probabilidade de ocorrência p_k, é definida por

$$h(m_k) = -\log_2(p_k) \ [bits] \tag{2.1}$$

Analisemos (2.1): Visto que $0 \le p_k \le 1$, $h(m_k)$ é sempre um número positivo. $h(m_k)$ é medida em $[bits]$ devido à função logarítmica em base 2. Como $\log_2(u)$ é uma função monotonamente crescente com u, a Auto-Informação $h(m_k) = -\log_2(p_k)$ de uma mensagem rara é maior que a de uma mensagem comum. Ainda, a derivada $\frac{d}{dp_k}\left(-\log_2(p_k)\right) = \frac{-1}{p_k \ln(2)}$ é uma medida da sensibilidade de $h(m_k)$ à variação da probabilidade p_k de ocorrência da mensagem m_k. Observe, portanto, que a Auto-Informação varia pouco entre mensagens m_k comuns mas apresenta alta variação entre mensagens raras. Finalmente, observe que as propriedades de (2.1) concordam com noção intuitiva de informação discutida em parágrafo anteriores.

A média da Auto-Informação das M mensagens m_k do conjunto $\Omega_X = \{m_0, m_1, \cdots, m_{M-1}\}$ é denominada de <u>Entropia</u> da variável aleatória X ($\equiv$ Entropia do conjunto Ω_X de mensagens). Assim, a Entropia $H(X)$ da variável aleatória X cujo espaço de amostras é o conjunto Ω_X de M mensagens é dada por

$$H(X) = E\{h(m_k)\} = E\{-\log_2(p_k)\} = -\sum_{k=0}^{M-1} p_k \log_2(p_k) \ [bits] \tag{2.2}$$

onde o $E\{\cdot\}$ é o operador estatístico que retorna o valor esperado do argumento [Carlson]. Note em (2.2) que, se as M mensagens apresentam probabilidade de ocorrência iguais (mensagens equiprováveis), então $p_k = 1/M$ para $k = 0,1,\cdots,M-1$ e $H(X) = -\frac{1}{M}\sum_{k=0}^{M-1}\log_2\left(\frac{1}{M}\right) = \log_2(M) \ [bits]$.

Além da interpretação da Entropia $H(X)$ como sendo a informação média implícita no conjunto de mensagens $\Omega_X = \{m_0, m_1, \cdots, m_{M-1}\}$, conjunto que é o espaço de amostras da variável aleatória X, são válidas também as seguintes interpretações:

1- A informação média obtida como resultado da observação de uma realização da variável aleatória X (realização = ocorrência de uma mensagem m_k na saída do quantizador). Nesta interpretação, $H(X)$ é melhor quantificada na unidade $[bits/realização]$, ou no caso da saída do quantizador, em $[bits/mensagem]$, ou mais genericamente, em $[bits/símbolo]$.

2- A incerteza média sobre X antes de ocorrer uma observação.

3- A incerteza média sobre X removida após ocorrer uma observação.

Exemplo 2.1: Seja um sistema para transmissão digital que utilize no Codificador de Fonte um conjunto $\Omega_X = \{m_0, m_1\}$ com $M = 2$ possíveis mensagens (ou $M = 2$ níveis de quantização sob o ponto de vista do quantizador). Seja q a probabilidade de que a saída X do quantizador assuma o valor m_0, isto é, $q = P(X = m_0)$. Determine o gráfico da entropia de X em função de q.

Solução: Se $q = P(X = m_0)$, então $P(X = m_1) = 1 - q$. De (2.2) temos

$$H(X) = -q \log_2 q - (1-q)\log_2(1-q) \;\; [bits/mensagem] \tag{2.3}$$

A Figura 2.1 mostra o gráfico $H(X) \times q$.

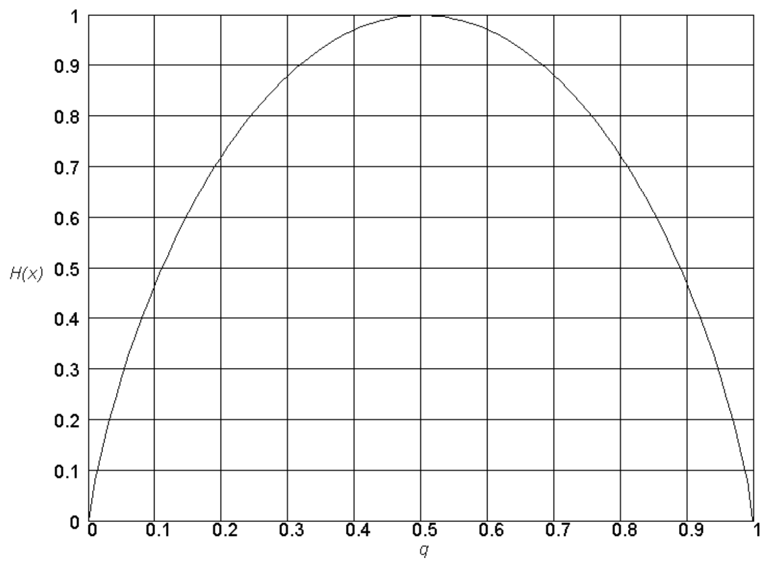

Figura 2.1: Entropia de X em função de q.

Note na Figura 2.1 que $H(X)$ é máxima quando as mensagens m_0 e m_1 têm a mesma probabilidade de ocorrência, i.e., $q = (1-q) = 0.5$. Na realidade, este comportamento acontece não só para um espaço de amostras Ω_X com apenas $M = 2$ mensagens de probabilidades iguais, mas ocorre também para qualquer quantidade M de mensagens de mesma probabilidade. Demonstra-se [Ash] que

> O valor máximo da entropia da variável aleatória X é $H(X) = \log_2(M)$, valor que ocorre quando as probabilidades de ocorrência dos M elementos do espaço de amostras Ω_X são todas iguais a $1/M$ (i.e., os M elementos de Ω_X são equiprováveis).

2.2 Taxa de Informação

Sejam duas fontes de informação A e B aplicadas à entrada de seus respectivos Codificadores de Fonte. Suponhamos que estamos registrando as saídas X_A e X_B dos quantizadores respectivos a cada fonte e que, com isto, calculemos as entropias $H(X_A)$ e $H(X_B)$ e que obtivemos $H(X_A) = H(X_B)$.

Poderíamos ser levados a crer que o Codificador de Canal que segue o Codificador de Fonte teria o mesmo trabalho de codificação a realizar quer a ele seja aplicado a saída do Codificador de Fonte A quer a ele seja aplicado a saída do Codificador de Fonte B. No entanto, consideremos a situação que a fonte A apresente um espectro mais amplo que o da fonte B. Portanto, a princípio, a freqüência de amostragem da fonte A terá de ser maior que a da fonte B para evitar *aliasing*. Portanto, embora a entropia de X_{A} seja igual à entropia de X_{B}, X_{A} gera mais mensagens por segundo do que X_{B}, e, obviamente, isto gera um diferenciação no trabalho realizado pelo Codificador de Canal.

Para levar em conta o efeito da taxa de amostragem, existe o conceito de Taxa de Informação. Se a fonte é amostrada a uma taxa tal que o quantizador gera $r\,[mensagens/segundo]$ com uma entropia $H\,[bits/mensagem]$ então a Taxa de Informação R é definida como

$$R = rH \ \ [bits/s] \tag{2.4}$$

e é uma medida do número médio de bits que necessita ser transportado por segundo através do sistema.

Exemplo 2.2: Seja um sistema para transmissão digital que utilize no Codificador de Fonte um conjunto $\Omega_X = \{m_0, m_1, m_2, m_3\}$ com $M = 4$ possíveis mensagens (ou $M = 4$ níveis de quantização sob o ponto de vista do quantizador). As amostras na saída X do quantizador são tais que a ocorrência de uma não altera a probabilidade de ocorrência da outra (i.e., as mensagens são estatisticamente independentes). As probabilidades são $P(X = m_0) = P(X = m_3) = 1/8$ e $P(X = m_1) = P(X = m_2) = 3/8$. O sinal $m(t)$ a ser quantizado possui um espectro cuja freqüência máxima é $f_M = 10\,\text{KHz}$ e é amostrado a uma taxa igual à Freqüência de Nyquist. Determine a Taxa de Informação gerada pelo sinal $m(t)$ na saída X do quantizador.

Solução: A informação média gerada por $m(t)$ em X é

$$H(X) = -\frac{1}{8}\log_2\left(\frac{1}{8}\right) - \frac{3}{8}\log_2\left(\frac{3}{8}\right) - \frac{3}{8}\log_2\left(\frac{3}{8}\right) - \frac{1}{8}\log_2\left(\frac{1}{8}\right) = 1.8 \ \ [bits/mensagem] \tag{2.5}$$

O intervalo de amostragem de $m(t)$ é $T_s = 1/2f_M = 50\mu s$. Portanto são geradas $r = 1/T_s = 20000 \ \ [mensagens/segundo]$**.** Assim,

$$R = rH = 2f_M\,[mensagens/segundo] \times H[bits/mensagem] = 36000\,[bits/s] \tag{2.6}$$

2.X Referências Bibliográficas

[Carlson] A. B. Carlson, *Communication Systems*, McGraw-Hill, 1965.

[Ash] R. Ash, *Information Theory*, Interscience - John Wiley & Sons, 1967.