Princípios Básicos de Teoria da Informação

➢ Entropia

- Até que limite é possível comprimir um conjunto de dados?
- Codificação por Entropia.

- Entropia – Uma Possível Medida de Informação
- Taxa de Informação
- Codificação por Entropia
- Códigos Univocamente Decodificáveis
- Códigos Instantâneos (Códigos Prefixos)
- Procedimento geral para testar se um código é UD
- Teorema da Codificação de Fonte – *Noiseless Coding Theorem*
- Eficiência de um Código por Entropia
- Códigos Ótimos – Códigos de Huffman
- Método para Construção de Códigos Ótimos

➢ Capacidade do Canal – Teorema Fundamental de Shannon

- Qual a maior taxa de transmissão de informação possível em um canal de transmissão para que não ocorram erros?
- Códigos Corretores de Erro.

Entropia – Uma Possível Medida de Informação

- A observação da ocorrência de um evento do espaço amostral de uma variável aleatória nos dá informação.
- Eventos raros contém mais informação do que eventos comuns.

  "O sol nasceu hoje pela manhã"

  "Porto Alegre foi atingida por um terremoto hoje pela manhã"
- A entropia (proposta em 1928 por Hartley) é uma medida logarítmica de informação que reflete este raciocínio intuitivo.

- Por exemplo, se estivermos registrando o valor das amostras na saída do quantizador de um codificador que 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 (que são mensagens de $N = \log_2 M$).
- A saída 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 Auto-Informação $h(m_k)$ implícita na ocorrência de uma mensagem m_k com probabilidade de ocorrência p_k , é definida por

$$h(m_k) = -\log_2(p_k) \ [bits] \tag{6.1}$$

A partir da equação (6.1) pode-seconcluir que:

- Como $0 \le p_k \le 1$, $h(m_k)$ 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 monotonicamente crescente com u, a Auto-Informação $h(m_k) = -\log_2(p_k)$ de uma mensagem rara é maior do que a de uma mensagem comum.

- 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

$$H(X) = E\{h(m_k)\} = E\{-\log_2(p_k)\} = -\sum_{k=0}^{M-1} p_k \log_2(p_k) \ [bits] \tag{6.2}$$

onde o $E\{\cdot\}$ é o operador estatístico que retorna o valor esperado do argumento [Carlson].

Note em (6.2) 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.

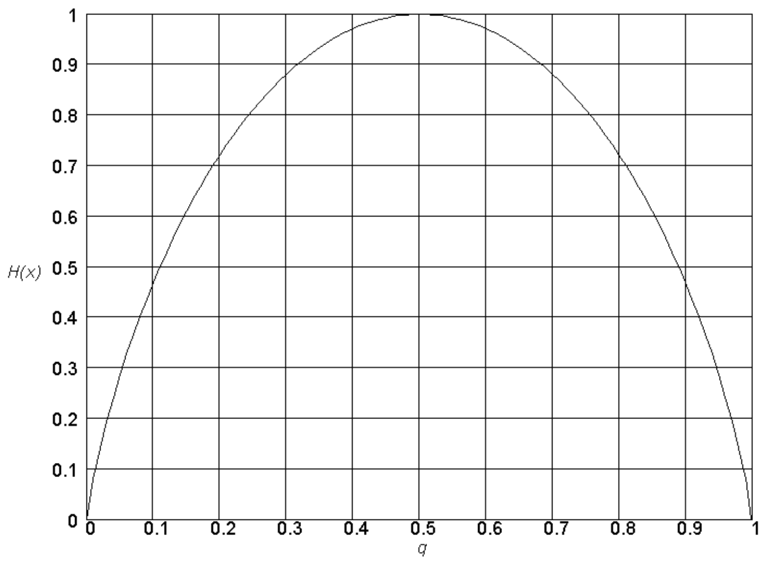
Exemplo 1:

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 (6.2) temos

$$H(X) = -q \log_2 q - (1-q)\log_2(1-q) \ [bits/mensagem] \tag{6.3}$$

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



Entropia de X em função de q.

Note na figura 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.

> 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). [Ash]

Taxa de Informação

Seja uma fonte de informação A aplicada à entrada de um codificador.

Suponhamos que estamos registrando a saída X do quantizador e calculando a entropia $H(X)$.

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

$$R = rH \;\; [bits/s] \tag{6.4}$$

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

Exemplo 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 intervalo de amostragem de $m(t)$ é $T_s = 1/2f_M = 50\mu s$.

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

$$H(X) = -\frac{1}{8}\log_2\left(\frac{1}{8}\right) - \frac{3}{8}\log_2\left(\frac{3}{8}\right) - \frac{3}{8}\log_2\left(\frac{3}{8}\right) - \frac{1}{8}\log_2\left(\frac{1}{8}\right) = 1.8 \;\; [bits/mensagem] \tag{6.5}$$

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

Assim,

$$R = rH = 2f_M\,[mensagens/s] \times H[bits/mensagem] = 36000\,[bits/s] \tag{6.6}$$

Codificação por Entropia

- Considerando que o quantizador de um codificador apresente M níveis de quantização e codifique o sinal $m(t)$ quantizado com seqüências de $N = \log_2 M$ bits.
- O código para compressão de dados considera cada uma das M possíveis seqüências de N bits como uma mensagem de N bits e associa a cada uma delas uma palavra-código cujo nº de bits depende da probabilidade de ocorrência da mensagem.
- **Palavras-código com menos bits são atribuídas a mensagens com maior probabilidade de ocorrência, e palavras-código com mais bits são atribuídas a mensagens com menor probabilidade de ocorrência.**
- Este critério é crucial para a eficiência da compressão. Um código que segue este critério faz com que mensagens que ocorrem freqüentemente necessitem de menos bits para serem transmitidas e, portanto, o efeito global é o de permitir que mais informação possa ser transmitida no mesmo intervalo de tempo.
- Quando um sistema digital é projetado, é feito um estudo estatístico da probabilidade de ocorrência de cada uma das M possíveis mensagens para que o código compressor possa ser especificado. O conjunto de M valores obtidos, cuja soma forçosamente tende para 1.0, é uma boa aproximação das probabilidades de ocorrência de cada uma das M possíveis mensagens.
- Códigos para compressão com base no princípio $probabilidade \uparrow \;\Rightarrow\; bits \downarrow$ são denominados de processos para Codificação por Entropia.

O veterano Código Morse, utilizado para enviar informação por telegrafia desde a I Guerra Mundial, é um exemplo histórico desta classe de códigos.

Cada letra do alfabeto A – Z é uma mensagem do Código Morse.

O conjunto de caracteres utilizado para compor as palavras-código do Código Morse é o conjunto $\{" \cdot " , " - "\}$.

A cada mensagem é atribuída uma seqüência de "pontos" e/ou "traços" representados em telegrafia por tons audíveis curtos e/ou longos.

O mapeamento $mensagem \rightarrow palavra-código$ do Código Morse é tal que letras mais prováveis na escrita inglesa são associadas a palavras-código curtas e letras menos prováveis são associadas a palavras-código longas.

A letra "E", por exemplo, é a letra mais freqüente na escrita em inglês e é representada por um único $" \cdot "$.

- **A Entropia é uma medida do conteúdo de informação associado a uma variável aleatória discreta *X*, com espaço de amostras definido pelo conjunto $\mathbf{\Omega} = \{x_i\} = \{x_0, x_1, \cdots, x_{M-1}\}$ de *M* eventos x_i com probabilidade de ocorrência p_i, $i = 0,1,\cdots, M-1$.**
- **Quando *X* é a saída de uma fonte de informação discreta, a entropia $H(X)$ da fonte representa a quantidade média de informação emitida pela fonte.**
- Podemos considerar um código para compressão por entropia como um operador $\mathbf{\theta}\{\}$, tal que $\mathbf{S} = \mathbf{\theta}\{\mathbf{\Omega}\}$, onde

  $\mathbf{\Omega} = \{x_i\} = \{x_0, x_1, \cdots, x_{M-1}\}$ é o conjunto de M possíveis **mensagens** a serem codificadas e

  $\mathbf{S} = \{s_i\} = \{s_0, s_1, \cdots, s_{M-1}\}$ é o conjunto de M possíveis **palavras-código** ou **símbolos** resultantes da codificação.
- O operador $\mathbf{\theta}\{\cdot\}$ efetua um mapeamento unívoco entre cada mensagem e respectiva palavra-código, tal que:
  - mensagens com maior probabilidade de ocorrência são mapeadas em palavras-código de menor tamanho (no caso de um código binário, “tamanho” refere-se ao número de bits), e
  - mensagens com menor probabilidade de ocorrência são mapeadas em palavras-código de maior tamanho.
- O **conjunto de caracteres do código** ou **alfabeto do código** é o conjunto $\mathbf{A} = \{a_0, a_1, \cdots, a_{D-1}\}$ composto por D elementos, de cuja composição são formadas cada palavra-código.
- As palavras-código formadas do alfabeto $\mathbf{A}$, as quais constituem o **conjunto imagem** do mapeamento $\mathbf{\theta}\{\}$, são assumidas serem distintas entre si, caso contrário $\mathbf{\theta}\{\cdot\}$ não seria unívoco.

Exemplo 3:
Seja o alfabeto $\mathbf{A} = \{a_0, a_1, a_2\}$ e o conjunto de mensagens $\mathbf{\Omega} = \{x_0, x_1, x_2, x_3\}$.
Um possível código $\mathbf{\theta}\{\}$ seria conforme tabela ao lado.

Mensagem	Palavra-código s_i associada a x_i por $s_i = \mathbf{\theta}\{x_i\}$
x_0	$a_0 a_1$
x_1	$a_0 a_1 a_2$
x_2	a_0
x_3	a_1

Exemplo 4: Seja o alfabeto $\mathbf{A} = \{a_0, a_1, a_2\}$ e o conjunto de mensagens $\mathbf{\Omega} = \{x_0, x_1, x_2, x_3\} = \{00, 01, 10, 11\}$ resultante da codificação da saída de um Quantizador com 4 níveis de quantização. Um possível código $\boldsymbol{\theta}\{\}$ seria

Mensagem	Seqüência	Palavra-código s_i associada a x_i por $s_i = \boldsymbol{\theta}\{x_i\}$
x_0	00	$a_0 a_1$
x_1	01	$a_0 a_1 a_2$
x_2	10	a_0
x_3	11	a_1

Obs: As palavras-código usualmente originam-se de um alfabeto binário $\mathbf{A} = \{0,1\}$.

Exemplo 5: Seja o alfabeto $\mathbf{A} = \{0,1\}$ e o conjunto de mensagens $\mathbf{\Omega} = \{x_0, x_1, x_2, x_3\} = \{00, 01, 10, 11\}$. Um possível código $\boldsymbol{\theta}\{\}$ seria

Mensagem	Seqüência	Palavra-código s_i associada a x_i por $s_i = \boldsymbol{\theta}\{x_i\}$
x_0	00	0
x_1	01	010
x_2	10	01
x_3	11	10

Obs: O **tamanho** ℓ_i de uma palavra-código ou símbolo s_i é definido pelo número de caracteres do alfabeto $\mathbf{A}$ utilizado na sua construção.

Exemplo 6: Seja o código binário ($A = \{0,1\}$) do Exemplo 5. O tamanho ℓ_i de cada palavra-código ou símbolo s_i é

Mensagem	Seqüência	Símbolo s_i associado a x_i por $s_i = \boldsymbol{\theta}\{x_i\}$	ℓ_i
x_0	00	0	1
x_1	01	010	3
x_2	10	01	2
x_3	11	10	2

O objetivo da **Codificação por Entropia** é encontrar um código $\boldsymbol{\theta}\{\cdot\}$ que **minimize o tamanho médio** $\overline{L}$ **dos símbolos emitidos pela fonte**, a partir do conjunto de M possíveis símbolos $\mathbf{S}=\{s_i\}=\{s_0,s_1,\cdots,s_{M-1}\}$, sendo $\overline{L}$ dado por

$$\overline{L}=\sum_{i=0}^{M-1} p_i \ell_i \tag{6.7}$$

onde p_i é a probabilidade de ocorrência da mensagem x_i, e ℓ_i é o tamanho do símbolo s_i associado à mensagem x_i através do código $\boldsymbol{\theta}\{\cdot\}$.

- A Codificação por Entropia assume que a fonte é **sem memória**.
- Uma fonte é considerada sem memória quando as mensagens emitidas pela fonte são estatisticamente independentes, i.e., a ocorrência de uma determinada mensagem x_i não afeta a probabilidade de ocorrência da mensagem x_j, com $i,j=0,1,\cdots,M-1$.
- Esta condição é necessária pois, caso contrário, a função $\overline{L}=f(p_i,\ell_i)$ a ser minimizada, dada por (6.7), dependeria do desenrolar temporal da seqüência de mensagens emitidas pela fonte, o que resultaria em um código $\boldsymbol{\theta}\{\cdot\}$ variável no tempo.
- Embora poucas fontes físicas sigam exatamente o modelo de uma fonte sem memória, códigos $\boldsymbol{\theta}\{\cdot\}$ constantes no tempo (resultantes da suposição de independência estatística) são amplamente utilizados como códigos compressores, mesmo quando a dependência estatística da fonte resulta na impossibilidade de minimização de $\overline{L}$ durante a totalidade do tempo de codificação.

Exemplo 7:

Seja um sistema para transmissão digital que utilize no Codificador de Fonte um conjunto $\mathbf{\Omega} = \{x_0, x_1, x_2, x_3\} = \{00, 01, 10, 11\}$ 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 $P(X = x_0) = 1/2$, $P(X = x_1) = 1/4$ e $P(X = x_2) = P(X = x_3) = 1/8$.

O código compressor $\boldsymbol{\theta}\{\cdot\}$ é conforme tabela abaixo

Mensagem	Seqüência	Palavra-código s_i associada a x_i por $s_i = \boldsymbol{\theta}\{x_i\}$
x_0	00	0
x_1	01	10
x_2	10	110
x_3	11	111

Determine:

- a entropia da saída do quantizador $H(X)$ e
- o comprimento médio $\overline{L}(\boldsymbol{\theta})$ do código $\boldsymbol{\theta}\{\cdot\}$.

Solução:

Mensagem	p_i	Símbolo s_i associado a x_i por $s_i = \boldsymbol{\theta}\{x_i\}$	ℓ_i
x_0	$1/2$	0	1
x_1	$1/4$	10	2
x_2	$1/8$	110	3
x_3	$1/8$	111	3

$$H(X) = -\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right) - \frac{1}{8}\log_2\left(\frac{1}{8}\right) - \frac{1}{8}\log_2\left(\frac{1}{8}\right) = 1.75 \ \ [bits/mensagem] \qquad (6.8)$$

$$\overline{L}(\boldsymbol{\theta}) = \frac{1}{2}\times 1 + \frac{1}{4}\times 2 + \frac{1}{8}\times 3 + \frac{1}{8}\times 3 = 1.75 \ \ [bits/símbolo] \qquad (6.9)$$

Exemplo 8: Seja o código compressor $\boldsymbol{\theta}\{\cdot\}$ conforme definido abaixo:

Mensagem	p_i	Símbolo s_i associado a x_i por $s_i = \boldsymbol{\theta}\{x_i\}$
x_0	$1/3$	0
x_1	$1/3$	10
x_2	$1/3$	11

Determine a entropia $H(X)$ da fonte e o comprimento médio $\overline{L}(\boldsymbol{\theta})$ do código $\boldsymbol{\theta}\{\cdot\}$.

Solução:

$$H(X) = -\frac{1}{3}\log_2\left(\frac{1}{3}\right) - \frac{1}{3}\log_2\left(\frac{1}{3}\right) - \frac{1}{3}\log_2\left(\frac{1}{3}\right) = 1.58 \ \ [bits/mensagen] \tag{6.10}$$

$$\overline{L}(\boldsymbol{\theta}) = \frac{1}{3}\times 1 + \frac{1}{3}\times 2 + \frac{1}{3}\times 2 = 1.67 \ \ [bits/símbolo] \tag{6.11}$$

Códigos Univocamente Decodificáveis

- Um código que pretenda ser útil deve pertencer à classe de códigos Univocamente Decodificáveis, caso contrário é impossível efetuar a decodificação sem que ocorra ambigüidade.

- Um código é Univocamente Decodificável (UD) quando qualquer seqüência de caracteres do alfabeto $\mathbf{A}$ passível de ser formada a partir da <u>justaposição</u> de um número qualquer de símbolos quaisquer pertencentes a $\mathbf{S}=\{s_i\}=\{s_0, s_1, \cdots, s_{M-1}\}$ puder ser associada, ao ser decodificada, a uma única mensagem em $\mathbf{\Omega}=\{x_i\}=\{x_0, x_1, \cdots, x_{M-1}\}$.

- **<u>Conceito de justaposição</u>**: A justaposição de N símbolos (ou palavras-código) $s_i, s_{i+1}, \cdots, s_{i+N-1}$ é a seqüência α formada pela transmissão do símbolo s_i seguido da transmissão do símbolo s_{i+1}, e assim sucessivamente até a transmissão do símbolo s_{i+N-1}, cuja representação é $\alpha = s_i s_{i+1} \cdots s_{i+N-1}$.

Exemplo 9: Verifique se o código $\boldsymbol{\theta}\{\}$ abaixo é UD.

Mensagem	Palavra-código s_i associada a x_i por $s_i = \boldsymbol{\theta}\{x_i\}$
x_0	0
x_1	010
x_2	01
x_3	10

Solução:

A seqüência 010 poderia corresponder a qualquer uma das seguintes seqüências de mensagens x_1, $x_2 x_0$ ou $x_0 x_3$.

- Portanto $\boldsymbol{\theta}\{\}$ não é UD.

Códigos Instantâneos (=Códigos Prefixos)

- No Exemplo 9 a ambigüidade do código $\theta\{\}$ talvez pudesse ser resolvida se aguardássemos a recepção de bits adicionais, mas tal tempo de espera é indesejável, dada à sempre existente busca por velocidade de decodificação.

- Uma maneira de assegurar que um código seja UD **e** que nenhum tempo de espera seja necessário para a correta decodificação é utilizar códigos denominados Prefixos ou Instantâneos (a denominação "Instantâneo" decorre da não necessidade de aguardar a recepção de bits adicionais para que se resolva ambigüidades).

- **Nota**: Todos os códigos instantâneos são UD, mas nem todos os códigos UD são instantâneos. Ou seja, o conjunto dos códigos instantâneos é um sub-conjunto do conjunto dos códigos UD.

- Um código instantâneo ou prefixo pode ser decodificado sem referência a palavras-código futuras porque o final de uma palavra-código é imediatamente reconhecido no decodificador.

- **Um código é chamado Instantâneo se nenhuma palavra-código é prefixo de nenhuma outra palavra-código pertencente ao código.**

Conceito de prefixo:

Sejam as seqüências α_a, α_b e α_c, formadas pela justaposição de, respectivamente, N_a, N_b e N_c palavras-código s_i pertencentes ao código $\theta\{\}$, sendo $N_a = N_b + N_c$ um número qualquer de palavras-código.

Dizemos que α_b é prefixo de α_a, se α_a puder ser representada por $\alpha_b\alpha_c$, para alguma seqüência α_c denominada sufixo.

Exemplo 10: Verifique se o código $\theta\{\}$ abaixo é Instantâneo.

Mensagem	Palavra-código s_i associada a x_i por $s_i = \theta\{x_i\}$
x_0	10
x_1	00
x_2	11
x_3	110

Solução:

Como 11 é prefixo de 110 , $\theta\{\}$ não é Instantâneo.

Não podemos afirmar que não seja UD pelo fato de não ser Instantâneo.

Procedimento geral para testar se um código é UD

Seja um código $\mathbf{\theta}\{\cdot\}$ com alfabeto $\mathbf{A}=\{a_0,a_1,\cdots,a_{D-1}\}$ e conjunto imagem $\mathbf{S}=\{s_i\}=\{s_0,s_1,\cdots,s_{M-1}\}$. Para testar se $\mathbf{\theta}\{\cdot\}$ é UD, constrói-se a seqüência de conjuntos $\mathbf{S}_0,\mathbf{S}_1,\cdots$ da seguinte maneira:

1. $\mathbf{S}_0$ é o próprio conjunto imagem $\mathbf{S}=\{s_i\}=\{s_0,s_1,\cdots,s_{M-1}\}$.

2. Para definir $\mathbf{S}_1$, forma-se a partir de $\mathbf{S}_0$ o conjunto $\mathbf{P}$ de todos os pares $s_i s_j$ de palavras-código, $s_i \neq s_j$, possíveis de serem formados por justaposição de duas palavras-código distintas pertencentes ao conjunto $\mathbf{S}_0$:

	s_0	s_1	$\cdots$	s_{M-1}
s_0	-	$s_0\ s_1$	$\cdots$	$s_0\ s_{M-1}$
s_1	$s_1\ s_0$	-	$\cdots$	$s_1\ s_{M-1}$
$\vdots$	$\vdots$	$\vdots$	-	$\vdots$
s_{M-1}	$s_{M-1}\ s_0$	$s_{M-1}\ s_1$	$\cdots$	-

Formação do conjunto $\mathbf{P}=\{s_0s_1,s_0s_2,\cdots,s_{M-1}s_{M-2}\}$ de M^2-M elementos.

3. Se a palavra-código $s_i\in\mathbf{S}_0$ é prefixo da palavra-código $s_j\in\mathbf{S}_0$, i.e. $s_j=s_i\sigma$, então o sufixo σ é um elemento do conjunto $\mathbf{S}_1$, i.e. $\sigma\in\mathbf{S}_1$.

Executa-se a verificação $s_j \overset{?}{=} s_i\sigma$ para todos os elementos de $\mathbf{P}$ até que todos os sufixos sejam atribuídos ao conjunto $\mathbf{S}_1=\{\alpha_0,\alpha_1,\cdots\}$, onde cada seqüência α_k de caracteres de $\mathbf{A}$ é um sufixo originado pelo resultado positivo do teste $s_j \overset{?}{=} s_i\sigma$.

4. Para definir $\mathbf{S}_n$, $n > 1$, compara-se $\mathbf{S}_0$ e $\mathbf{S}_{n-1}$ de modo bidirecional:

I) Se uma palavra-código $s_i \in \mathbf{S}_0$ é prefixo de uma seqüência $\alpha_j \in \mathbf{S}_{n-1}$ tal que $\alpha_j = s_i\sigma$, então o sufixo $\sigma \in \mathbf{S}_n$.

II) Se uma seqüência $\alpha'_j \in \mathbf{S}_{n-1}$ é prefixo de uma palavra-código $s'_i \in \mathbf{S}_0$ tal que $s'_i = \alpha'_j\sigma'$, então o sufixo $\sigma' \in \mathbf{S}_n$.

5. Define-se tantos conjuntos $\mathbf{S}_n$ até um valor de n tal que $\mathbf{S}_n = \{ö\}$ ou até um valor de n tal que $\mathbf{S}_n = \mathbf{S}_{n-1}$.

6. O código $\boldsymbol{\theta}\{\}$ é UD se e somente se **nenhum** dos conjuntos da seqüência de conjuntos $\mathbf{S}_1, \mathbf{S}_2, \cdots$ contenha uma palavra-código que pertença ao conjunto $\mathbf{S}_0$.

Exemplo 11: Verifique se o código $\boldsymbol{\theta}\{\}$ abaixo com alfabeto $\mathbf{A} = \{a, b, c, d, e\}$ é UD.

Mensagem	Palavra-código s_i associada a x_i por $s_i = \boldsymbol{\theta}\{x_i\}$
x_0	a
x_1	c
x_2	ad
x_3	abb
x_4	bad
x_5	deb
x_6	$bbcde$

Solução:

$\mathbf{S}_0$	$\mathbf{S}_1$	$\mathbf{S}_2$	$\mathbf{S}_3$	$\mathbf{S}_4$	$\mathbf{S}_5$	$\mathbf{S}_6$	$\mathbf{S}_7$	$\mathbf{S}_8$
a	d	eb	de	b	ad	d	eb	$\{ö\}$
c	bb	cde			$bcde$			
ad								
abb								
bad								
deb								
$bbcde$								

- Visto que $ad \in \mathbf{S}_5$ e $ad \in \mathbf{S}_0$, logo $\boldsymbol{\theta}\{\}$ não é UD.
- Note que poderíamos ter encerrado o procedimento ao obter $\mathbf{S}_5$, quando, então, já temos elementos suficientes para decidir que $\boldsymbol{\theta}\{\}$ não é UD.

Exemplo 12: Verifique se o códigos $\theta_{I}\{\}$, $\theta_{II}\{\}$ e $\theta_{III}\{\}$ são UD e/ou Instantâneos.

Mensagem	$s_i = \theta_{I}\{x_i\}$	$s_i = \theta_{II}\{x_i\}$	$s_i = \theta_{III}\{x_i\}$
x_0	1	0	0
x_1	00	10	01
x_2	01	110	011
x_3	10	111	111

Solução:

- Verificando $\theta_{I}\{\}$:

S_0	S_1	S_2
1	0	0
00		1
01		
10		

- $\theta_{I}\{\}$ não é Instantâneo (1 é prefixo de 10), mas pode ser UD.
- Visto que $1 \in S_2$ e $1 \in S_0$, $\theta_{I}\{\}$ não é UD.

- Verificando $\theta_{II}\{\}$:

S_0	S_1
0	$\{\emptyset\}$
10	
110	
111	

- $\theta_{II}\{\}$ é Instantâneo (nenhuma palavra-código é prefixo de nenhuma outra) então $\theta_{II}\{\}$ é UD.
- Apenas a título de ilustração vamos verificar se $\theta_{II}\{\}$ é UD: Como nenhum dos conjuntos da seqüência de conjuntos $S_1, S_2, \cdots$ contém uma palavra-código que pertença ao conjunto S_0, $\theta_{II}\{\}$ é UD .

- Verificando $\theta_{III}\{\}$:

S_0	S_1	S_2
0	1	11
01	11	1
011		
111		

- $\theta_{III}\{\}$ não é Instantâneo (0 é prefixo de 01, por exemplo), mas pode ser UD.
- Aplicando o procedimento para verificação de código UD: Como nenhum dos conjuntos da seqüência de conjuntos $S_1, S_2, \cdots$ contém uma palavra-código que pertença ao conjunto S_0, $\theta_{III}\{\}$ é UD .

Teorema da Codificação de Fonte
(*Noiseless Coding Theorem*)

"Seja uma variável aleatória discreta X, com espaço de amostras definido pelo conjunto $\mathbf{\Omega} = \{x_i\} = \{x_0, x_1, \cdots, x_{M-1}\}$ de M eventos estatisticamente independentes x_i com probabilidade de ocorrência p_i, $i = 0, 1, \cdots, M-1$.

Então é possível construir um código Instantâneo $\mathbf{\theta}\{\cdot\}$ com um conjunto de palavras-código $\mathbf{S} = \{s_i\} = \{s_0, s_1, \cdots, s_{M-1}\}$ formadas a partir do alfabeto $\mathbf{A} = \{a_0, a_1, \cdots, a_{D-1}\}$, tal que o conjunto $\mathbf{L} = \{\ell_i\} = \{\ell_0, \ell_1, \cdots, \ell_{M-1}\}$ dos tamanhos das palavras-código respectivas em $\mathbf{S}$ satisfaça a desigualdade

$$\frac{H(X)}{\log_2 D} \leq \overline{L} < \frac{H(X)}{\log_2 D} + 1 \qquad (6.12)$$

onde:

$H(X)$ é a Entropia X da fonte e

$\overline{L}$ é o tamanho médio das palavras-códigos, dado por $\overline{L} = \sum_{i=0}^{M-1} p_i \ell_i$".

[Ash][Cover]

O Teorema da Codificação de Fonte (TCF) garante a viabilidade teórica de implementação de códigos instantâneos D-ários, cujo tamanho médio dos símbolos pode ser reduzido a um valor tão pequeno quanto o valor da Entropia $H(X)$ da fonte, ou, se impossível, pelo menos a um valor menor que $H(X)+1$.

Eficiência de um Código por Entropia

Uma decorrência do TCF é a definição da Eficiência de Codificação η dada por

$$\eta = \frac{H(X)}{\overline{L}\log_2 D} \tag{6.13}$$

- Um código é **<u>Absolutamente Ótimo</u>** (*matched to the source* – casado com a fonte) quando $\eta = 1.0$, isto é, quando $\dfrac{H(X)}{\log_2 D} = \overline{L}$.
- Um código é **<u>Quase Absolutamente Ótimo</u>**, quando $\dfrac{H(X)}{\log_2 D} \le \overline{L} < \dfrac{H(X)}{\log_2 D} + 1$.

Tomemos como exemplo o código estudado no Exemplo 7, em que:

$H(X) = 1.75$ $[bits/mensagem]$, $\overline{L}(\boldsymbol{\theta}) = 1.75$ $[bits/símbolo]$ e $\log_2 D = \log_2 2 = 1$.

Para este código, $\dfrac{H(X)}{\log_2 D} = \overline{L}$.

Portanto, <u>o código é Absolutamente Ótimo</u>.

Códigos Ótimos – Códigos de Huffman

- Embora o TCF nos garanta que é possível obter códigos instantâneos com $\overline{L}$ tão pequeno quanto a própria Entropia $H(X)$ da fonte, nenhuma informação é dada sobre **como** construir tais códigos.
- A construção de tais códigos baseia-se na minimização de $\overline{L} = \sum_{i=0}^{M-1} p_i \ell_i$.
- Um código instantâneo que mimimize $\overline{L}$ é denominado de **Código Ótimo**.
- Existe um teorema que prova que se um código ótimo $\boldsymbol{\theta}^*\{\cdot\}$ resulta em $\overline{L}^*$, é impossível existir um outro código instantâneo $\boldsymbol{\theta}\{\cdot\}$ com tamanho médio $\overline{L}$ tal que $\overline{L} < \overline{L}^*$ [Ash].

Um Código Ótimo D-ário cujas palavras-código $\mathbf{S} = \{s_i\} = \{s_0, s_1, \cdots, s_{M-1}\}$ são formadas a partir do alfabeto $\mathbf{A} = \{a_0, a_1, \cdots, a_{D-1}\}$ satisfaz as seguintes propriedades (se o código for binário cada dígito D-ário é um bit) [Cover]:

1- Palavras-código com maior probabilidade possuem menor tamanho.
2- As D palavras-código menos prováveis possuem o mesmo tamanho.
3- As D palavras-código menos prováveis diferem somente no último dígito D-ário.

Exemplo 13: Verifique se o código $\boldsymbol{\theta}\{\cdot\}$ abaixo é Ótimo.

Mensagem	p_i	Palavra-código s_i associada a x_i por $s_i = \boldsymbol{\theta}\{x_i\}$
x_0	0.6	0
x_1	0.2	100
x_2	0.1	101
x_3	0.04	1101
x_4	0.06	1110

Solução:

- As propriedades 1 e 2 são satisfeitas.
- A propriedade 3 não é satisfeita: x_3 e x_4 não diferem somente no último bit.
- Portanto, $\boldsymbol{\theta}\{\cdot\}$ não é ótimo.

Método para Construção de Códigos Ótimos

Para a construção de $\theta\{\cdot\}$ efetua-se:
Seja, inicialmente, $k=j=0$.
1. Organizar as probabilidades p_i de alto a baixo em uma coluna em ordem decrescente de valor, denominada Coluna $\boldsymbol{k}$.
2. Somar as D menores probabilidades na Coluna $\boldsymbol{k}$ e transferí-las para a próxima coluna (à direita), denominada Coluna $\boldsymbol{k+1}$, obedecendo a ordem decrescente. As demais probabilidades da Coluna $\boldsymbol{k}$ são transferidas inalteradas para a Coluna $\boldsymbol{k+1}$.
3. Incrementar $\boldsymbol{k}$ de 1 e repetir 1 a 3 até restarem somente D probabilidades na Coluna $\boldsymbol{k+1}$, então denominada Coluna $\boldsymbol{j}$.
4. Na Coluna $\boldsymbol{j}$, atribuir a palavra-código representada pelo caractere a_0 à maior probabilidade, atribuir a palavra-código representada pelo caractere a_1 à segunda maior probabilidade, e assim sucessivamente até atribuir a palavra-código representada pelo caractere a_{D-1} à menor probabilidade.
5. Localizar na Coluna $\boldsymbol{j+1}$, imediatamente à esquerda da Coluna $\boldsymbol{j}$, quais as D probabilidades geradoras que, ao serem somadas, resultaram na probabilidade gerada na Coluna $\boldsymbol{j}$. Atribuir às D probabilidades geradoras na coluna Coluna $\boldsymbol{j+1}$ a palavra-código já atribuída à probabilidade gerada na Coluna $\boldsymbol{j}$. Às probabilidades não-geradoras na Coluna $\boldsymbol{j+1}$ são atribuídas as palavras-código já atribuídas às respectivas probabilidades não-geradas por soma na Coluna $\boldsymbol{j}$.
6. Na Coluna $\boldsymbol{j+1}$, às palavras-códigos já atribuídas em 5 às D probabilidades geradoras, justapor a palavra-código representada pelo caractere a_0 àquela geradora de maior probabilidade, justapor a palavra-código representada pelo caractere a_1 àquela geradora de segunda maior probabilidade, e assim sucessivamente até justapor a palavra-código representada pelo caractere a_{D-1} à palavra-código geradora de menor probabilidade.
7. Incrementar $\boldsymbol{j}$ de 1 e repetir 5 a 7 até que todas as colunas tenham palavras-código associadas às probabilidades nelas contidas.
8. Após a execução de 7, o Código de Huffman estará definido na coluna mais à esquerda.

Exemplo 14: Seja uma fonte de informação representada pela variável aleatória discreta X, com espaço de amostras definido pelo conjunto $\mathbf{\Omega}=\{x_i\}=\{x_0,x_1,\cdots,x_{M-1}\}$ de $M=6$ eventos estatisticamente independentes x_i com probabilidade de ocorrência p_i, $i=0,1,\cdots,M-1$, conforme tabela abaixo.

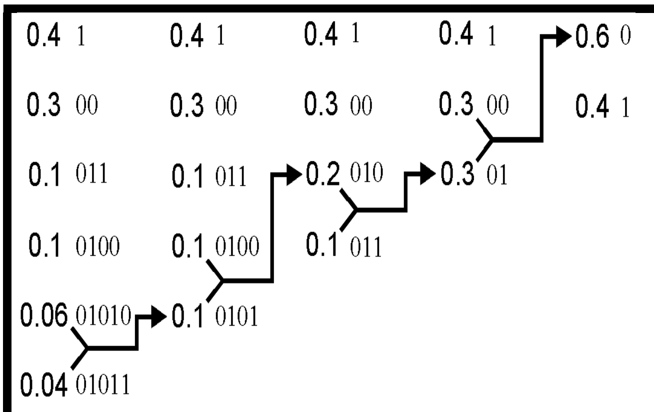
Mensagem	p_i
x_0	0.4
x_1	0.3
x_2	0.1
x_3	0.1
x_4	0.06
x_5	0.04

a) Determine um Código Ótimo $\boldsymbol{\theta}\{\}$ cujo conjunto de palavras-código $\mathbf{S}=\{s_i\}=\{s_0,s_1,\cdots,s_{M-1}\}$ é formado a partir do alfabeto $\mathbf{A}=\{0,1\}$ (código binário).

b) Determine a Eficiência de $\boldsymbol{\theta}\{\cdot\}$.

c) Determine se $\boldsymbol{\theta}\{\cdot\}$ é Absolutamente Ótimo ou Quase Absolutamente Ótimo.

Solução:



O código de Huffman $\boldsymbol{\theta}\{\}$ resultante é mostrado na tabela ao lado.

Mensagem	p_i	Palavra-código s_i (símbolo) associada a x_i por $s_i=\boldsymbol{\theta}\{x_i\}$
x_0	0.4	1
x_1	0.3	00
x_2	0.1	011
x_3	0.1	0100
x_4	0.06	01010
x_5	0.04	01011

$$H(X)=-\sum_{i=0}^{M-1}p_i\log_2(p_i)=2.14\ \ [bits/mensagem]$$

$$\overline{L}(\boldsymbol{\theta})=\sum_{i=0}^{M-1}p_i\ell_i=$$

$$=0.4\times1+0.3\times2+0.1\times3+0.1\times4+0.06\times5+0.04\times5=2.20\ \ [bits/símbolo]$$

$$\eta=\frac{H(X)}{\overline{L}\log_2 D}=\frac{2.14\ \ [bits/mensagem]}{2.20\,[bits/símbolo]}=97.3\%$$

Visto que $\dfrac{H(X)}{\log_2 D}\leq\overline{L}<\dfrac{H(X)}{\log_2 D}+1$, $\boldsymbol{\theta}\{\}$ é Quase Absolutamente Ótimo.

Referências Bibliográficas

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

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

[Proakis] J. G. Proakis, *Digital Communications*, McGraw-Hill, 1995.

[Shannon] C.E. Shannon, "A Mathematical Theory of Communications", *Bell Systems Technical Journal*, vol. 27, pp. 379 –423 (part I) and pp. 623 –656 (part II), 1948.

[Taub] H. Taub and D.L. Schilling, *Principles of Communications Systems*, McGraw-Hill, 1986.