TE903 – Comunicação Digital

Introdução à Teoria de Informação e Codificação de Fonte

Evelio M. G. Fernández

21 de outubro de 2019

Notes

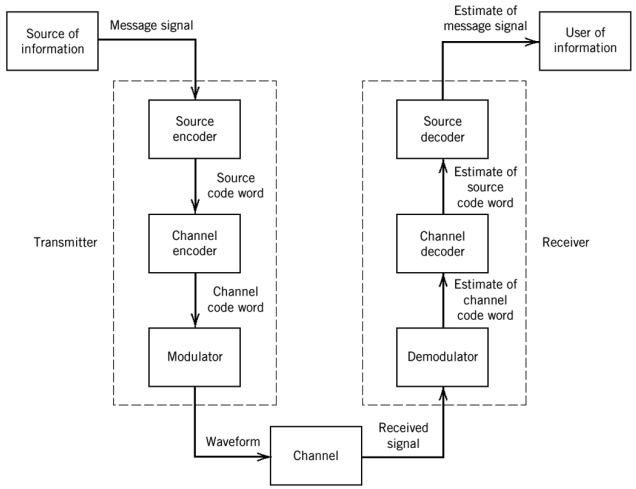
Introdução à Teoria de Informação

Em 1948, Claude Shannon publicou o trabalho "*A Mathematical Theory of Communications*". A partir do conceito de comunicações de Shannon, podem ser identificadas três partes:

1. **Codificação de Fonte:** Shannon mostrou que em princípio sempre é possível transmitir a informação gerada por uma fonte a uma taxa igual à sua entropia.
2. **Codificação de Canal:** Shannon descobriu um parâmetro calculável que chamou de Capacidade de Canal e provou que, para um determinado canal, comunicação livre de erros é possível desde que a taxa de transmissão não seja maior que a capacidade do canal.
3. **Teoria Taxa-Distorção:** A ser utilizada em compressão com perdas.

Notes

Sistema de Comunicação Digital



Notes

Compressão de Dados

- Arte ou ciência de representar informação de uma forma compacta. Essas representações são criadas identificando e utilizando estruturas que existem nos dados para eliminar redundância.

- Dados:
  - Caracteres num arquivo de texto;
  - Números que representam amostras de sinais de áudio, voz, imagens, etc.

Notes

Algoritmos de Compressão

1. **MODELAGEM** – Extrair informação sobre a redundância da fonte e expressar essa redundância na forma de um modelo;
2. **CODIFICAÇÃO** – Uma descrição do modelo e uma descrição de como os dados diferem do modelo são codificados possivelmente utilizando símbolos binários.

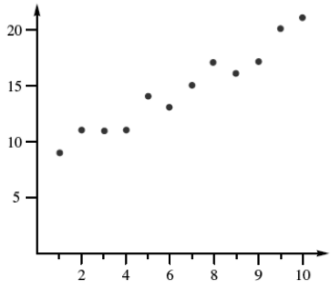
   Diferença: dados – modelo = resíduo

Notes

Exemplo 1

9	11	11	11	14	13	15	17	16	17	20	21

$\hat{x}_n = n+8 \qquad n = 1, 2, \ldots$

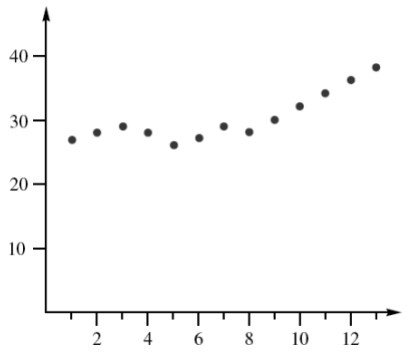


Resíduo: $e_n = x_n - \hat{x}_n \in \{-1,0,1\} \Rightarrow$ 2 bits para representá-lo

Notes

Exemplo 2

27	28	29	28	26	27	29	28	30	32	34	36	38



27	1	1	−1	−2	1	2	−1	2	2	2	2	2

Resíduo: $e_n = x_n - x_{n-1}$

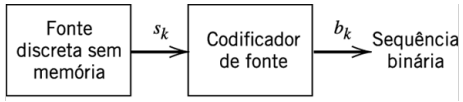
Notes

Medidas de Desempenho

1. Taxa de Compressão
   - Ex: 4:1 ou 75 %
2. Fidelidade
   - Distorção (*Rate-Distortion Theory*)

Notes

Exemplo 3



Símbolo (s_k)	Prob. (p_k)	I	II	III	IV
A	1/2	00	0	0	0
B	1/4	01	11	10	01
C	1/8	10	00	110	011
D	1/8	11	01	1110	0111

Taxa de compressão:

$$\bar{L} = \sum_{k=0}^{K-1} l_k \cdot p_k \quad \Longrightarrow$$

- $\bar{L}_{\text{I}} = 2$ bits/símbolo
- $\bar{L}_{\text{II}} = 1{,}25$ bits/símbolo
- $\bar{L}_{\text{III}} = \bar{L}_{\text{IV}} = 1{,}875$ bits/símbolo

Fidelidade: Decodificação da sequência: $\ldots 0011 \ldots$ pelo 'código' II?

Notes

Códigos Prefixos

- Nenhuma palavra código é prefixo de qualquer outra palavra-código;
- Todo código prefixo é instantâneo (o final das palavras-código é bem definido);
- Um código prefixo é sempre U.D. (a recíproca não é sempre verdadeira);
- Existe um código prefixo binário se e somente se

$$\sum_{k=0}^{K-1} 2^{l_k} \leq 1 \rightarrow \text{Desigualdade de Kraft-McMillan.}$$

- Com relação aos códigos da tabela:
  - I $\rightarrow 4 \times 2^{-2} = 1$
  - II $\rightarrow 2^{-1} + 3 \times 2^{-2} = 1{,}25 > 1$
  - III e IV $\rightarrow 2^{-1} + 2^{-2} + 2^{-3} + 2^{-4} = 0{,}9375 < 1$

Notes

Códigos Prefixos

Dado um conjunto de comprimentos de palavras código que satisfaz a desigualdade de Kraft-McMillan, SEMPRE será possível encontrar um código prefixo com esses comprimentos para as suas palavras-código. O comprimento médio das palavras do código estará limitado pela entropia da fonte de informação.

Notes

Medida da Informação

- **Modelo da fonte:** $S \triangleq$ V.A. discreta com Prob$(S = s_k) = p_k$, $k = 0,1,\ldots,K-1$
- **Informação obtida com a revelação do evento:** $S = s_k \rightarrow I(s_k) = -\log_2 p_k$ bits
- **Propriedades de** $I(\cdot)$
  - Se $p_k = 1 \Rightarrow I(s_k) = 0$
  - $0 \leq p_k \leq 1 \Rightarrow I(s_k) \geq 0$
  - $p_k < p_i \Rightarrow I(s_k) > I(s_i)$
  - Se s_k e s_i são estatisticamente independentes $\Rightarrow I(s_k s_i) = I(s_k) + I(s_i)$

Notes

Medida da Informação – Entropia

Entropia da Fonte, $H(S)$

Valor médio de $I(s_k)$ sobre o alfabeto S. É uma medida do número médio de símbolos necessários para codificar a fonte.

$$H(S) \triangleq E[I(s_k)]$$
$$= -\sum_{k=0}^{K-1} p_k \log_2 p_k \text{ bits/símbolo da fonte}$$

Propriedades de $H(S)$

- $H(S) \geq 0$;
- $H(S) = 0$ se e somente se $p_k = 1$ para algum k;
- $H(S) \leq \log_2 K$ com igualdade se e somente se $p_k = \frac{1}{k}$ para todo k.

Exemplo: Entropia da fonte da tabela?

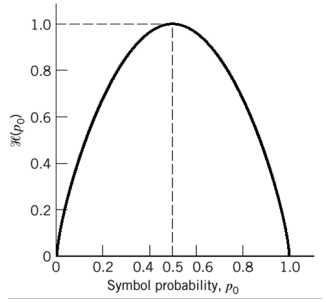
$\rightarrow H(S) = -0{,}5 \log_2 0{,}5 - 0{,}25 \log_2 0{,}25 - 2 \times 0{,}125 \log_2 0{,}125 = 1{,}75$ bits/símbolo

Notes

Entropia de uma Fonte Binária sem Memória

Fonte binária com Prob$(S = 0) = p$ e Prob$(S = 1) = 1 - p$

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



Notes

Teorema da Codificação de Fonte

- Dada uma fonte discreta sem memória com entropia $H(S)$, o comprimento médio $\overline{L}$ de um código U.D. para a codificação desta fonte é limitado por:

$$\overline{L} \geq H(S)$$

com igualdade se e somente se:

$p_k = r^{-l_k}, \quad r = 0,1,\ldots,K-1 \Rightarrow$ códigos absolutamente ótimos

- Códigos (quase) absolutamente ótimos:

$$\Rightarrow -\log_r p_k \leq l_k \leq -\log_r p_k + 1$$

$$\Rightarrow \frac{H(S)}{\log_2 r} \leq \overline{L} \leq \frac{H(S)}{\log_2 r} + 1$$

Notes

Extensão de uma Fonte Discreta sem Memória

- Fonte extendida $\rightarrow$ símbolos tratados em blocos de n símbolos da fonte original;
- Alfabeto (S^n) com K^n blocos (mensagens) distintos de comprimento n.

Exemplo: Fonte binária, $S = \{0,1\}$ com Prob$(S = 0) = 0,25$ e Prob$(S = 1) = 0,75$

$$\Rightarrow H(S) = -0,25 \log_2 0,25 - 0,75 \log_2 0,75 = 0,81 \text{ bits/símbolo}$$

Considerar agora mensagens ou palavras de comprimento $n = 3$

Mensagem (s_k)	Probabilidade
000	1/64
001	3/64
010	3/64
011	9/64
100	3/64
101	9/64
110	9/64
111	27/64

- $H(S^3) = 2,45$ bits/mensagem;
- Notar que $H(S^3) = 3H(S)$;
- Em geral, $H(S^n) = nH(S)$.

Notes

Extensão de uma Fonte Discreta sem Memória

Sejam:

- $\overline{L} \rightarrow$ comprimento médio em caracteres do código por símbolo da fonte S;
- $\overline{L}_n \rightarrow$ comprimento médio em caracteres do código por símbolo da fonte extendida S^n;
- $\frac{\overline{L}_n}{n} \rightarrow$ comprimento médio em caracteres do código por símbolo da fonte S;

$$\Longrightarrow \frac{H(S)}{\log_2 r} \leq \frac{\overline{L}_n}{n} < \frac{H(S)}{\log_2 r} + \frac{1}{n}$$

$$\lim_{n\to\infty} \frac{\overline{L}_n}{n} = \frac{H(S)}{\log_2 r}$$

Exercício: Seja uma fonte com $S = \{s_0, s_1\}$ e probabilidades $\{3/4, 1/4\}$. Seja a fonte extendida de ordem $n = 2$. Para codificá-las considere os códigos $\{0,1\}$ e $\{0,10,110,111\}$, respectivamente. Determine $\overline{L}$, $\overline{L}_n$ e $\frac{\overline{L}_n}{n}$.

Notes

Códigos de Huffmann Binários

1. Ordenar em uma coluna os símbolos do mais provável ao menos provável;
2. Associar '0' e '1' aos dois símbolos menos prováveis e combiná-los (soma das probabilidades individuais);
3. Repetir 1 e 2 até a última coluna que terá apenas dois símbolos; associa-se '0' e '1'.

Notes

Exercício

Obter um código um código ótimo binário pelo método de Huffmann para a seguinte fonte discreta de informação:

Símbolo	Prob.
s_0	0,4
s_1	0,2
s_2	0,2
s_3	0,1
s_4	0,1

a) Verifique se é código prefixo;
b) Determine $\overline{L}$;
c) Satisfaz o teorema da codificação de fonte?
d) É código absolutamente ótimo ou quase absolutamente ótimo?

Notes

Códigos Ótimos *r*-ários

- **Método de Huffmann:** aplica-se o método com o seguinte artifício:
- Adicionam-se ao alfabeto original símbolos fictícios com probabilidade zero de ocorrência, até o número de símbolos assim gerado ser congruente a $1 \bmod (r-1)$;
- Aplica-se o método de Huffmann agrupando-se r símbolos de cada vez. O código gerado é um código r-ário ótimo para o alfabeto original.

Notes

Exercício

Obter um código ótimo ternário pelo método de Huffmann para a seguinte fonte discreta de informação:

Símbolo	Prob.
s_0	1/3
s_1	1/6
s_2	1/6
s_3	1/9
s_4	1/9
s_5	1/9

a) Verifique se é código prefixo;
b) Determine $\overline{L}$;
c) Satisfaz o teorema da codificação de fonte?
d) É código absolutamente ótimo ou quase absolutamente ótimo?

Notes

Fonte com Alfabeto Pequeno

Símbolo	Código
a_1	0
a_2	11
a_3	10

$\mathcal{P}(a_1) = 0{,}95$
$\mathcal{P}(a_2) = 0{,}02$
$\mathcal{P}(a_3) = 0{,}03$

- $\overline{L} = 1{,}05$ bits/símbolo
- $H(A) = 0{,}335$ bits/símbolo
- Redundância $= 0{,}715$ bits/símbolo (213% da entropia)
- São necessários duas vezes mais bits do que o prometido pela entropia!

Notes

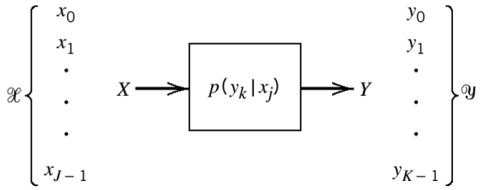
Segunda Extensão da Fonte

Símbolo	Prob.	Código
a_1a_1	0,9025	0
a_1a_2	0,0190	111
a_1a_3	0,0285	100
a_2a_1	0,0190	1101
a_2a_2	0,0004	110011
a_2a_3	0,0006	110001
a_3a_1	0,0285	101
a_3a_2	0,0006	110010
a_3a_3	0,0009	110000

- $\overline{L_2} = 1{,}222$ bits/símbolo
- $\overline{L_2}/2 = 0{,}611$ bits/símbolo (ainda 72% acima da entropia!)
- $\overline{L_n}/n \to H(A) \Rightarrow$ extensão de ordem $n = 8 \Rightarrow$ fonte com 6561 símbolos!
- Huffman: precisa criar todas as palavras-código!
- $\Rightarrow$ Codificação artimética, códigos baseados em dicionários, etc.

Notes

Canal Discreto sem Memória



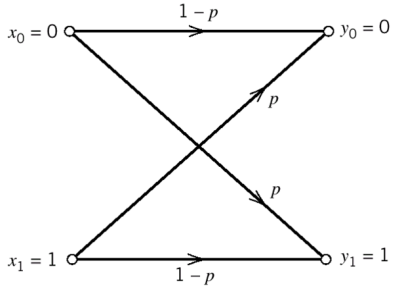
Notes

Matriz de Canal ou Matriz de Transição

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

Notes

Canal Binário Simétrico



Notes

Entropia Condicional

- $H(X) \rightarrow$ medida da incerteza *a priori* sobre X;

Incerteza sobre X após a observação de Y?

$\rightarrow$ Entropia condicional de X dado que $Y = y_k$:

$$H(X|Y = y_k) = \sum_{j=0}^{J-1} p(x_j|y_k) \log \frac{1}{p(x_j|y_k)}$$

$H(X|Y = y_k) \rightarrow$ é uma V.A. que toma valores $H(X|Y = y_0)$, $H(X|Y = y_1)$,..., $H(X|Y = y_{K-1})$, com probabilidades $p(y_0)$, $p(y_1)$,..., $p(y_{K-1})$.

$\Rightarrow$ A entropia condicional de X dado Y é o valor esperado de $H(X|Y = y_k)$ sobre o alfabeto de Y:

$$\begin{aligned} H(X|Y) &= \sum_{k=0}^{K-1} H(X|Y = y_k)p(y_k) \\ &= \sum_{k=0}^{K-1} \sum_{j=0}^{J-1} p(x_j|y_k)p(y_k) \log \frac{1}{p(x_j|y_k)} \end{aligned}$$

$H(X|Y) \rightarrow$ é a incerteza que ainda existe sobre X após a observação da saída do canal.

Notes

Informação Mútua

$\Rightarrow H(X) - H(X|Y) \rightarrow$ incerteza sobre a entrada 'dissipada' ou resolvida observando-se a saída do canal.

$\Rightarrow$ **Informação Mútua:** $I(X,Y) = H(X) - H(X|Y)$

- Quantidade de informação adquirida a respeito de X pela observação de Y;
- Informação processada pelo canal.

Analogamente,

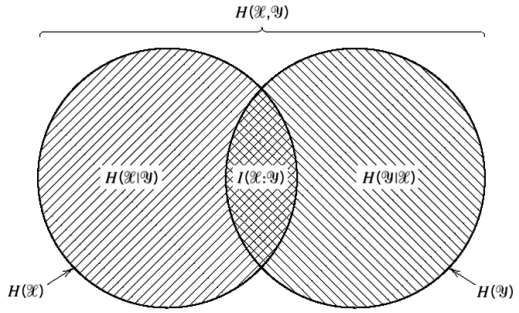
$$I(Y,X) = H(Y) - H(Y|X)$$

- Quantidade de informação adquirida a respeito de Y pelo envio de X;
- Medida da incerteza a respeito da saída do canal que é resolvida enviando-se a entrada do canal.

Notes

Propriedades da Informação Mútua

1. $I(X,Y) = I(Y,X)$;
2. $I(X,Y) \geq 0$. A igualdade $I(X,Y) = 0$ ou $H(X) = H(X|Y)$ acontece se e somente se X e Y são V.A.s independentes;
3. $I(X,Y) = H(X) + H(Y) - H(X,Y)$.



Notes

Exercício

Considere um canal DMC representado pela matriz de transição a seguir, onde os símbolos de entrada são equiprováveis:

$$\mathbf{P} = \begin{bmatrix} \frac{10}{16} & \frac{2}{16} & \frac{4}{16} \\ \frac{5}{16} & \frac{6}{16} & \frac{5}{16} \\ \frac{6}{16} & \frac{1}{16} & \frac{9}{16} \end{bmatrix}$$

Determine:

a) A quantidade de informação recebida por símbolo;
b) A quantidade de informação transportada pelo canal;
c) A probabilidade do símbolo x_2 ter sido transmitido dado que y_1 foi recebido.

Notes

Capacidade do Canal DMC

→ Para um dado DMC, em que situação a informação processada pelo canal será máxima?

$$
\begin{aligned}
I(X,Y) &= H(Y) - H(Y|X) \\
&= \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} p(x_j,y_k) \log \left(\frac{p(y_k|x_j)}{p(y_k)} \right) \\
&= I(Y,X)
\end{aligned}
$$

Capacidade do Canal DMC

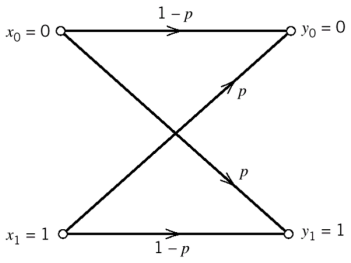
Valor máximo da informação mútua por cada utilização do canal, relativamente a todas as possíveis distribuições de probabilidades *a priori*, isto é,

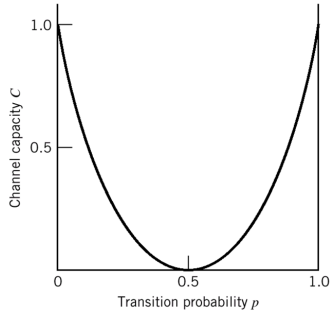
$$C = \max_{\{p(x_j)\}} I(X,Y)$$

C: bits/uso do canal ou bits/intervalo de sinalização

Notes

Exemplo: Capacidade do Canal BSC





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

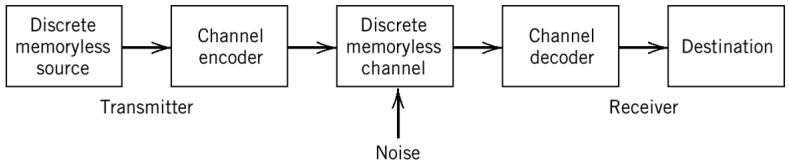
Notes

Capacidade de Canal

- A capacidade de canal não é somente uma propriedade de um canal físico particular;
- Um canal não significa apenas o meio físico de propagação das mensagens, mas também:
  - A especificação do tipo de sinais (binário, r-ário, ortogonal, etc);
  - O tipo de receptor usado (determinante da probabilidade de erro do sistema);
- Todas estas informações estão incluídas na matriz de transição do canal. Esta matriz especifica completamente o canal.

Notes

Sistema de Comunicação com Codificação de Canal



Notes

Teorema da Codificação de Canal

i. Seja uma fonte discreta sem memória com alfabeto S e entropia $H(S)$ que produz símbolos a cada T_s segundos. Seja um canal DMC com capacidade C que é usado uma vez a cada T_c segundos.
Então, se

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

existe um esquema de codificação para o qual a saída da fonte pode ser transmitida pelo canal e reconstruída com

$$P_e \to \varepsilon, \quad \varepsilon \to 0$$

Notes

Teorema da Codificação de Canal (Cont.)

ii. Pelo contrário, se

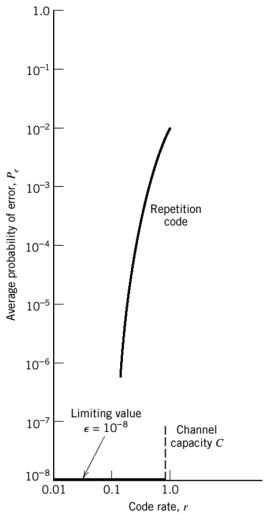
$$\frac{H(S)}{T_s} > \frac{C}{T_c}$$

não é possível o anterior.

$\Rightarrow$ Resultado mais importante da Teoria de Informação.

Notes

Código de Repetição



Notes

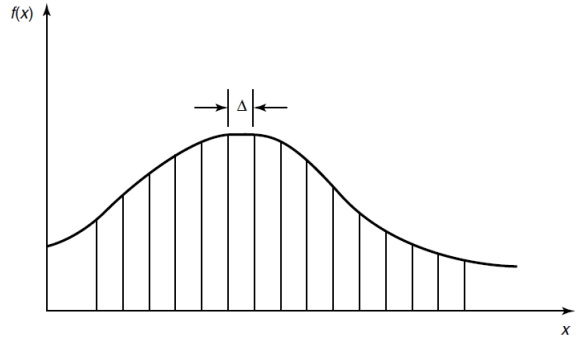
Medida de Informação para Sinais Contínuos

Caso Discreto:

- $X \rightarrow$ V.A. discreta;
- $x \in$ alfabeto finito.

Caso Contínuo:

- $X \rightarrow$ V.A. contínua;
- $x \in$ alfabeto infinito.



Pelo teorema do valor médio, existe um x_i dentro de cada *bin* tal que:

$$f_X(x_i)\Delta = \int_{i\Delta}^{(i+1)\Delta} f_X(x)dx.$$

Notes

Capacidade de Shannon

Capacidade do canal AWGN: $C = B\log_2(1+\gamma)$, onde $\gamma = \frac{P_r}{N_0 B}$

Exemplo (4.1, pag. 101, Goldsmith)
Considere um canal sem fio onde a potência média recebida decai com distância de acordo com

$$P_r(d) = P_t\left(\frac{d_0}{d}\right)^3, \text{ para } d_0 = 10\text{m}$$

Suponha que o canal é AWGN com 30 kHz de largura de banda e densidade espectral de potência de ruído $N_0 = 10^{-9}$ W/Hz. Para uma potência transmitida $P_t = 1$ W, determine a capacidade deste canal para $d = 100$ m e $d = 1$ km.

Notes

Capacidade do Canal sem Fio

$$Y = hX + n$$

- A capacidade do canal sem fio depende da distribuição do fading. Ex: $|h| \rightarrow$ Rayleigh $\Rightarrow |h|^2 \rightarrow$ exponencial;

- $h \rightarrow$ i.i.d. independente do sinal e do ruído;

- Supor que o coeficiente do canal é conhecido apenas no receptor $\Rightarrow$ o receptor conhece a SNR instantânea;

Capacidade (ergódica) de Shannon:

$$C = E[B\log_2(1+\gamma)] = \int_0^\infty B\log_2(1+\gamma)p(\gamma)d\gamma,$$

onde $\gamma = |h|^2\overline{\gamma}$ é a SNR instantânea e $p(\gamma)$ é a sua pdf.

Notes

Capacidade do Canal sem Fio

Exemplo (4.2, pag. 104, Goldsmith)

Considere um canal *flat fading* com ganho de canal h que pode tomar três valores:

$h_1 = 0{,}05$ com probabilidade $p_1 = 0{,}1$;
$h_2 = 0{,}5$ com probabilidade $p_2 = 0{,}5$;
$h_3 = 1$ com probabilidade $p_3 = 0{,}4$.

A potência transmitida é $P_t = 10$ mW, a densidade espectral de potência de ruído é $N_0 = 10^{-9}$ W/Hz e a largura de banda é $B = 30$ kHz. Suponha que o receptor conhece a CSI e o transmissor não. Determine a capacidade (ergódica) de Shannon e comparar com a cacidade do canal AWGN com a mesma SNR média.

Em geral,

$$C = E[B\log_2(1+\gamma)] = \int_0^\infty B\log_2(1+\gamma)p(\gamma)d\gamma \leq B\log_2(1+\overline{\gamma}).$$

Notes

Capacidade com Outage

- A capacidade de Shannon define a máxima taxa que pode ser enviada pelo canal com probabilidade de erro assintoticamente pequena no receptor;
- Como o TX não conhece o canal, a taxa de transmissão é constante $\Rightarrow$ estados ruins do canal reduzem a capacidade;
- Quando o canal experimenta um *slow deep fading* a BER não é zero porque o TX não pode adaptar a sua taxa de acordo com a CSI $\Rightarrow$ capacidade é zero!;
- **Capacidade com Outage:** máxima taxa que pode ser transmitida que garanta uma determinada probabilidade de outage, P_{out};
- Premisa: pode ser enviada uma alta taxa pelo canal e ser decodificada exceto quando o canal estiver em *slow deep fading*.

Notes

Capacidade com Outage

- Supor que a SNR na recepção, γ, é constante durante um bloco (longo) de transmissões e depois muda de valor de acordo com a distribuição do fading;
- O TX fixa um valor de SNR mínimo (na recepção), $\gamma_{\min}$, e transmite com uma taxa perto da capacidade para esse valor de SNR,

  $$C = B \log_2(1 + \gamma_{\min}).$$

  A informação será corretamente recebida se $\gamma \geq \gamma_{\min}$;
- A probabilidade de outage será então

  $$P_{\text{out}} = \Pr[\gamma < \gamma_{\min}] = \int_0^{\gamma_{\min}} p(\gamma) d\gamma = F_\Gamma(\gamma_{\min});$$

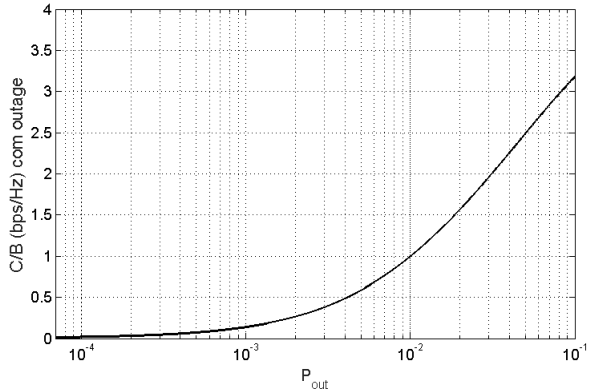
- A capacidade com outage, isto é, a taxa média (ao longo de muitas rajadas de transmissão) que pode ser corretamente recebida nesse cenário será

  $$C_{\text{out}} = (1 - P_{\text{out}}) B \log_2(1 + \gamma_{\min}).$$

Notes

Capacidade vs Outage

O valor de $\gamma_{\min}$ é um parâmetro de projeto escolhido a partir de um valor aceitável de probabilidade de outage. Isto pode ser ilustrado num gráfico da capacidade normalizada $C/B = \log_2(1 + \gamma_{\min})$ em função de P_{out} :



Notes

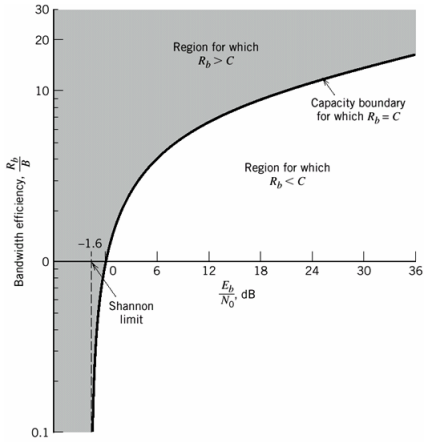
Exemplo: Capacidade vs Outage

Exemplo (4.3, pag. 105, Goldsmith)

Supondo o mesmo canal do exemplo anterior com $B = 30$ kHz e três possíveis valores de SNR recebida: $\gamma_1 = 0{,}8333$ com $p(\gamma_1) = 0{,}1$, $\gamma_2 = 83{,}33$ com $p(\gamma_1) = 0{,}5$ e $\gamma_3 = 333{,}33$ com $p(\gamma_3) = 0{,}4$. Determine a capacidade vs. outage para este canal e encontre o valor médio da taxa corretamente recebida para $P_{\text{out}} < 0{,}1$, $P_{\text{out}} = 0{,}1$ e $P_{\text{out}} = 0{,}6$.

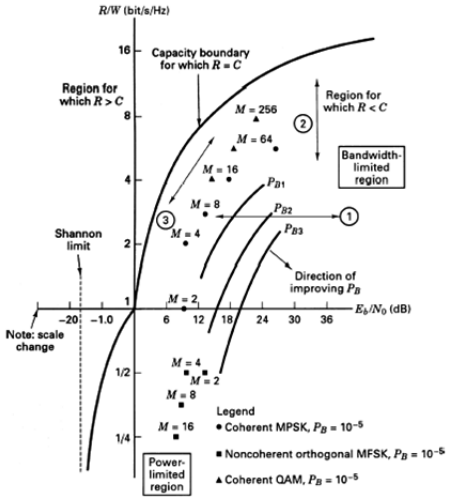
Notes

Limite de Shannon



Notes

Limite de Shannon



Notes

Notes