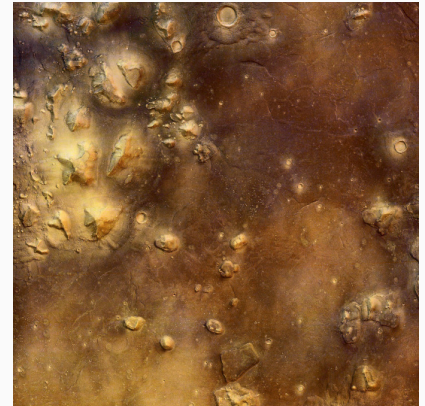


Teoria da Informação e Codificação

Teoria da informação: Codificação de canal



Prof. Roberto Wanderley da Nóbrega

Instituto Federal de Santa Catarina

Introdução

Information theorist's coat of arms

Codificação de fonte

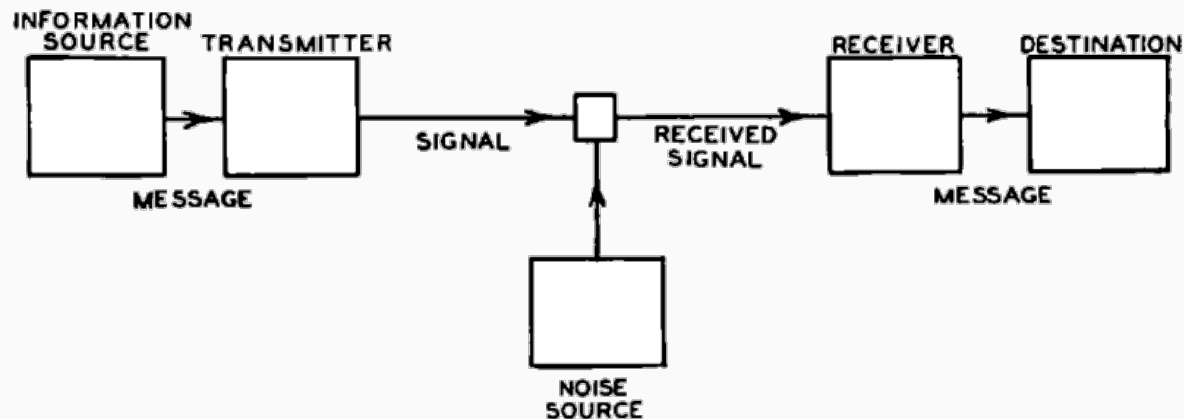
Representar com **fidelidade** uma *fonte de informação* através de uma *sequência de bits* com a **menor taxa possível**.

Conceito chave: **Entropia da fonte**.

Codificação de canal

Transmitir com **confiabilidade** uma *sequência de bits* por um *canal de comunicação* com a **maior taxa possível**.

Conceito chave: **Capacidade do canal**.



Fonte: Shannon–Weaver [1].

Entropia condicional e informação mútua

Entropia conjunta

Sejam X e Y duas v.a.s discretas, com alfabetos $\mathcal{X}$ e $\mathcal{Y}$, respectivamente.

A **entropia conjunta de X e Y** é definida por

$$H(X, Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log \frac{1}{p(x, y)},$$

onde $p(x, y)$ é a pmf conjunta de X e Y .

Observação: Note o abuso de notação: escreve-se $p(x, y)$, $p(x)$ e $p(x|y)$ em vez de $p_{X,Y}(x, y)$, $p_X(x)$ e $p_{X|Y=y}(x)$, sendo o argumento que identifica a distribuição.

Entropia condicional

Sejam X e Y duas v.a.s discretas, com alfabetos $\mathcal{X}$ e $\mathcal{Y}$, respectivamente.

A **entropia condicional de X dado $Y = y$** é definida pela entropia da distribuição condicional associada. Ou seja,

$$H(X \mid Y = y) = \sum_{x \in \mathcal{X}} p(x|y) \log \frac{1}{p(x|y)},$$

onde $p(x|y)$ é a pmf condicional de X dado $Y = y$.

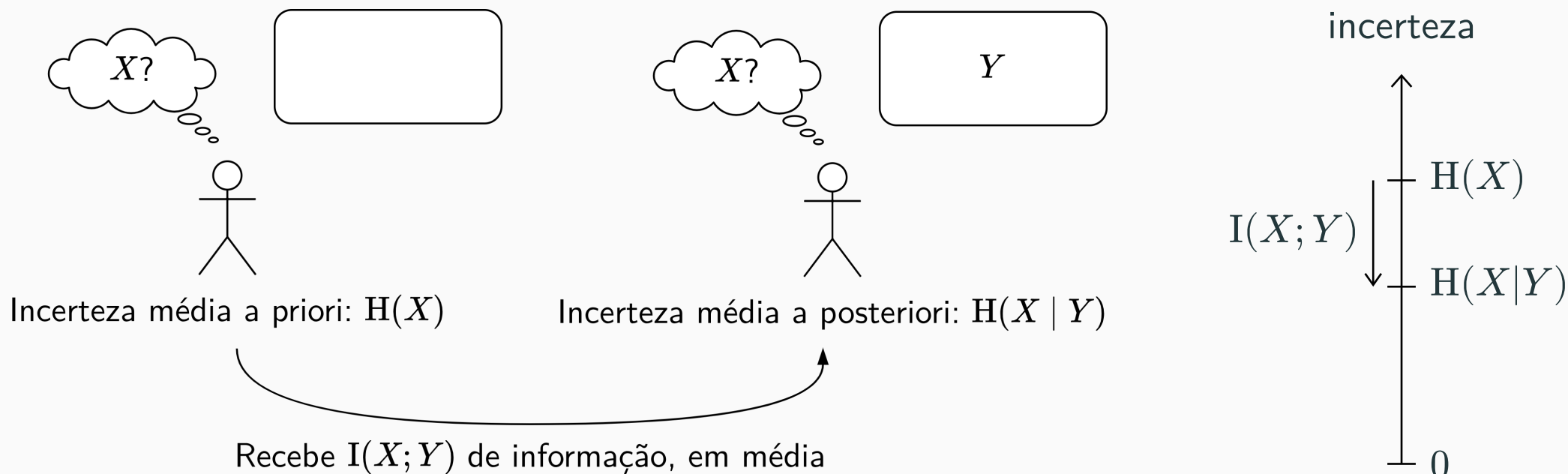
A **entropia condicional de X dado Y** é o valor médio da quantidade acima, com a média sendo calculada sobre Y :

$$H(X \mid Y) = \sum_{y \in \mathcal{Y}} p(y) H(X \mid Y = y).$$

Informação mútua

A **informação mútua** entre X e Y é definida por

$$I(X; Y) = H(X) - H(X | Y).$$



Exercício

Sejam X e Y conjuntamente distribuídas conforme abaixo.

	$x = 0$	$x = 1$
$y = 0$	0	$1/8$
$y = 1$	$3/4$	$1/8$

Determine:

- (a) As entropias de X e de Y .
- (b) A entropia conjunta de X e Y .
- (c) A entropia condicional de X dado Y , e a de Y dado X .
- (d) A informação mútua entre X e Y , e a entre Y e X .

Exercício

Sejam X e Y conjuntamente distribuídas conforme abaixo.

	$x = 0$	$x = 1$
$y = 0$	0	$1/8$
$y = 1$	$3/4$	$1/8$

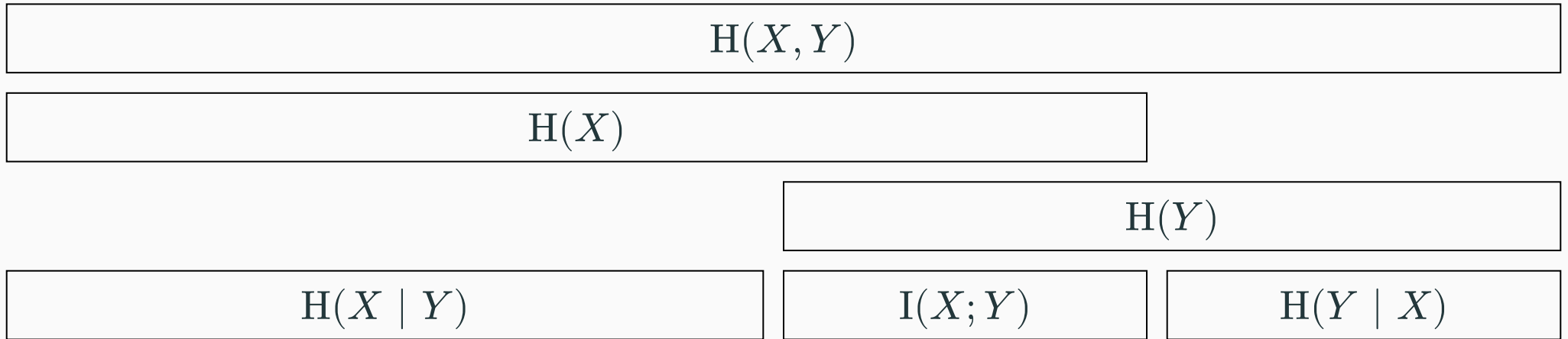
Respostas:

- (a) $H(X) = 0.811$ bits e $H(Y) = 0.544$ bits.
- (b) $H(X, Y) = 1.061$ bits.
- (c) $H(X | Y) = 0.518$ bits e $H(Y | X) = 0.25$ bits.
- (d) $I(X; Y) = I(Y; X) = 0.294$ bits.

1. $I(X; Y) = I(Y; X)$.
2. $I(X; Y) \geq 0$ ou, equivalentemente, $H(X | Y) \leq H(X)$,
com igualdade se e somente se X e Y são independentes.
Desta propriedade vem “*information can't hurt*” (em média).
3. $H(X, Y) = H(X) + H(Y) - I(X; Y)$.

Diagrama

Relações entre as entropias e a informação mútua de X e Y .



Fonte: MacKay [3].

[3, Exercício 8.10, p. 142] Considere três cartas: uma com ambas as faces brancas; uma com ambas as faces pretas; e uma com uma face branca e a outra preta. As três cartas são embaralhadas dentro de uma urna, da qual uma delas é sorteada e colocada sobre a mesa com orientação aleatória, de modo que apenas uma das faces fica visível. Quanta informação a cor da face visível fornece sobre a cor da face oculta?

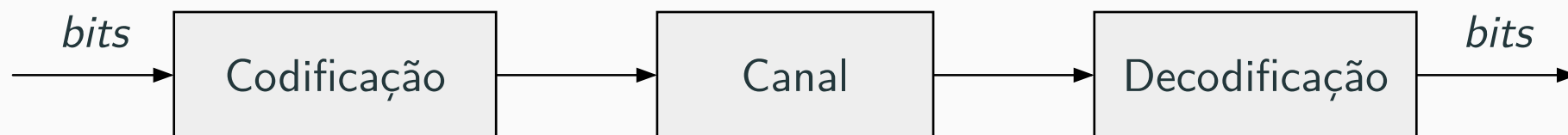
[3, Exercício 8.10, p. 142] Considere três cartas: uma com ambas as faces brancas; uma com ambas as faces pretas; e uma com uma face branca e a outra preta. As três cartas são embaralhadas dentro de uma urna, da qual uma delas é sorteada e colocada sobre a mesa com orientação aleatória, de modo que apenas uma das faces fica visível. Quanta informação a cor da face visível fornece sobre a cor da face oculta?

Resposta: $1 - \mathcal{H}(1/3) = 0.0817$ bits.

Canais discretos sem memória

Codificação de canal

Objetivo: Transmitir com *confiabilidade* uma sequência de bits por um canal de comunicação com a *maior taxa* possível.



Confiabilidade = ??? Por enquanto, pense em “probabilidade de erro baixa”.

Começaremos por um modelo simples, em que o canal é *discreto* e *sem memória*.

O canal mais importante em telecomunicações, o **canal AWGN**, ficará para depois.

Canal discreto sem memória

Um **canal discreto sem memória** (DMC) é caracterizado por:

- Um *alfabeto de entrada* finito, denotado por $\mathcal{X}$;
- Um *alfabeto de saída* finito, denotado por $\mathcal{Y}$;
- Uma *matriz de transição*, denotada por $p_{Y|X}$.

A cada *uso do canal* a saída $Y \in \mathcal{Y}$ é determinada (aleatoriamente) a partir da entrada $X \in \mathcal{X}$ de acordo com

$$\Pr[Y = y \mid X = x] = p_{Y|X}(y \mid x).$$

O canal é *sem memória*: a entrada atual é (estatisticamente) suficiente para a saída atual.

Convenção: Na matriz de transição $p_{Y|X}$, as linhas são referentes à entrada x e as colunas são referentes à saída y .

Exercício

Considere o DMC com alfabeto de entrada $\mathcal{X} = \{0, 1, 2\}$, alfabeto de saída $\mathcal{Y} = \{0, 1, 2, 3\}$ e matriz de transição

$$p_{Y|X} = \begin{bmatrix} 0.7 & 0.1 & 0.1 & 0.1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0.4 & 0.6 \end{bmatrix}.$$

Determine a pmf da saída supondo que a pmf da entrada seja $p_X = [0.4, 0.2, 0.4]$.

Exercício

Considere o DMC com alfabeto de entrada $\mathcal{X} = \{0, 1, 2\}$, alfabeto de saída $\mathcal{Y} = \{0, 1, 2, 3\}$ e matriz de transição

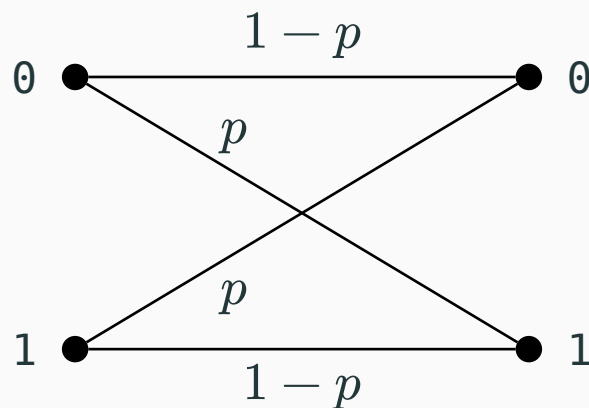
$$p_{Y|X} = \begin{bmatrix} 0.7 & 0.1 & 0.1 & 0.1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0.4 & 0.6 \end{bmatrix}.$$

Determine a pmf da saída supondo que a pmf da entrada seja $p_X = [0.4, 0.2, 0.4]$.

Resposta:

$$p_Y = [0.28, 0.04, 0.20, 0.48].$$

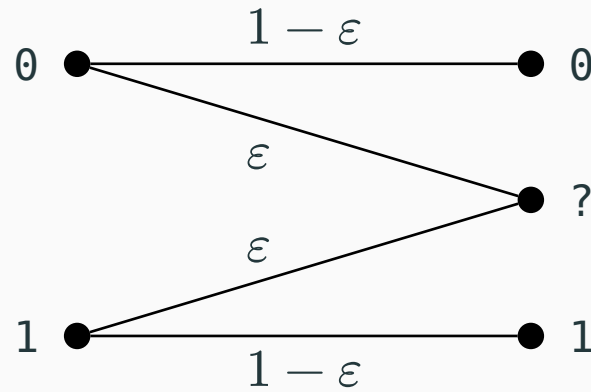
Canal binário simétrico, denotado por $\text{BSC}(p)$.



Modelo algébrico: $Y = X + Z$, com $Z \sim \text{Bernoulli}(p)$ e a soma é módulo-2 (XOR).

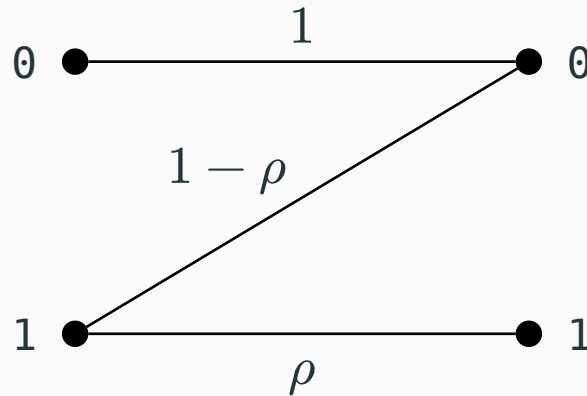
Observação: É suficiente considerar $0 \leq p \leq 1/2$ (por quê?).

Canal binário com apagamento, denotado por $\text{BEC}(\varepsilon)$.



Observação: Nesse caso, $\text{BEC}(0.2)$ é equivalente a $\text{BEC}(0.8)$?

Canal Z.



Modelo algébrico: $Y = AX$, com $A \sim \text{Bernoulli}(\rho)$.

Exercício

Considere o canal Z com $\rho = 0.8$. Supondo que a pmf a priori de X seja

$$p(x) = \{0 : 0.9, 1 : 0.1\},$$

determine as pmfs a posteriori de X dado $Y = 0$ e dado $Y = 1$.

Considere o canal Z com $\rho = 0.8$. Supondo que a pmf a priori de X seja

$$p(x) = \{0 : 0.9, 1 : 0.1\},$$

determine as pmfs a posteriori de X dado $Y = 0$ e dado $Y = 1$.

Resposta:

- $p(x \mid Y = 0) \approx \{0 : 0.978, 1 : 0.022\}.$
- $p(x \mid Y = 1) = \{0 : 0, 1 : 1\}.$

Exercício

Considere o canal composto por uma cascata de dois $\text{BSC}(p)$, com $0 \leq p \leq 1/2$.

- (a) Mostre que o canal resultante é um $\text{BSC}(q)$ e expresse q em função de p .
- (b) Mostre que $q \geq p$. Ou seja, a cascata é um BSC “piorado”.

Exercício

Considere o canal composto por uma cascata de dois $\text{BSC}(p)$, com $0 \leq p \leq 1/2$.

- (a) Mostre que o canal resultante é um $\text{BSC}(q)$ e expresse q em função de p .
- (b) Mostre que $q \geq p$. Ou seja, a cascata é um BSC “piorado”.

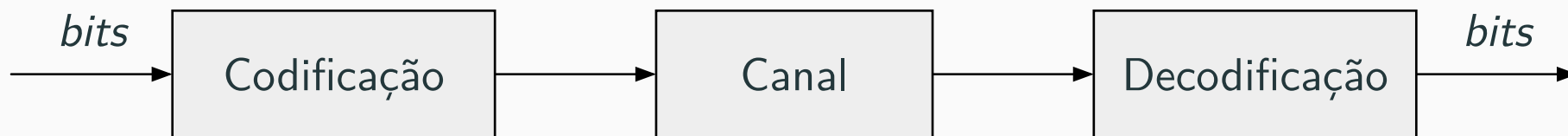
Respostas:

- (a) $q = 2p\bar{p}$, onde $\bar{p} = 1 - p$.
- (b) Dica: Compare $q - p$ com 0.

Codificação para um DMC

Codificação de canal

Lembrando o modelo de codificação de canal.



Ideias:

- Utilizar o canal múltiplas vezes.
- Adicionar redundância à informação $\iff$ restringir as sequências transmitidas.

Esquemas de codificação

- Um **esquema de codificação (de bloco)** para um DMC com alfabetos $\mathcal{X}$ e $\mathcal{Y}$ é dado por
- um **codificador**, que associa uma *mensagem* $m \in [0 : M)$ a uma *palavra-código* $x \in \mathcal{X}^n$,
 - um **decodificador**, que associa cada possível *palavra recebida* $y \in \mathcal{Y}^n$ a uma *estimativa da mensagem* $\hat{m} \in [0 : M)$.

Observação: Aqui, $\log_2 M$ é o número de bits a serem comunicados e n é número de usos do canal.

A **taxa** do esquema é definida por

$$R = (\log_2 M)/n \quad [\text{bits/uso do canal}]$$

e a **probabilidade de erro de bloco** do esquema é definida por

$$P_B = \Pr[\hat{m} \neq m],$$

isto é, a probabilidade de errar a mensagem m , suposta uniforme sobre $[0 : M)$.

Exemplo: Esquemas de codificação simples para o BEC — Sem codificação

Sem codificação.

Notação: $\bar{\varepsilon} = 1 - \varepsilon$.

Codificador			Decodificador		
m	x	$p(x)$	y	$\hat{m}$	$p(y)$
0	0	$1/2$	0	0	$\bar{\varepsilon}/2$
1	1	$1/2$	1	1	$\bar{\varepsilon}/2$
			?	0	ε

- Taxa: $R = 1$ bit/uso do canal.
- Probabilidade de erro: $P_B = \varepsilon \cdot (1/2) = \varepsilon/2$.

Observação: Para $y = ?$ o decodificador chuta $\hat{m} = 0$. Declarar falha ($\hat{m} = \perp$) contaria como erro; o chute acerta com probabilidade $1/2$.

Exemplo: Esquemas de codificação simples para o BEC — Repetição

No **código de repetição de comprimento n** , tem-se sempre $M = 2$, de modo que o alfabeto da mensagem é $[0 : M) = \{0, 1\}$. A codificação da mensagem $m \in \{0, 1\}$ consiste em converter m para bit e repetir esse bit um total de n vezes.

Exemplo: Esquemas de codificação simples para o BEC — Repetição

No **código de repetição de comprimento n** , tem-se sempre $M = 2$, de modo que o alfabeto da mensagem é $[0 : M) = \{0, 1\}$. A codificação da mensagem $m \in \{0, 1\}$ consiste em converter m para bit e repetir esse bit um total de n vezes.

Exemplo: Considere o código de repetição com $n = 4$ sobre o BEC.

- (a) Qual a sequência de palavras-código referente à sequência de mensagens 0, 0, 1, 0?
- (b) Suponha agora que, em outra transmissão, a sequência observada na saída do canal seja 111? 0000 ???? 1?1? ???1. Como você decodificaria esta sequência?

Exemplo: Esquemas de codificação simples para o BEC — Repetição

No **código de repetição de comprimento n** , tem-se sempre $M = 2$, de modo que o alfabeto da mensagem é $[0 : M) = \{0, 1\}$. A codificação da mensagem $m \in \{0, 1\}$ consiste em converter m para bit e repetir esse bit um total de n vezes.

Exemplo: Considere o código de repetição com $n = 4$ sobre o BEC.

- (a) Qual a sequência de palavras-código referente à sequência de mensagens 0, 0, 1, 0?
- (b) Suponha agora que, em outra transmissão, a sequência observada na saída do canal seja 111? 0000 ???? 1?1? ???1. Como você decodificaria esta sequência?

Respostas:

- (a) 0000 0000 1111 0000.
- (b) 1, 0, 0 (ou 1), 1, 1.

Exemplo: Esquemas de codificação simples para o BEC — Repetição

Repetição com $n = 2$.

Codificador			Decodificador					
m	x	$p(x)$	y	$\hat{m}$	$p(y)$	y	$\hat{m}$	$p(y)$
0	00	1/2	00	0	$\bar{\epsilon}^2/2$	10	—	0
1	11	1/2	01	—	0	11	1	$\bar{\epsilon}^2/2$
			0?	0	$\bar{\epsilon}\epsilon/2$	1?	1	$\bar{\epsilon}\epsilon/2$
						??	0	ϵ^2

$\log_2 M = 1$ bit
 $n = 2$ usos do canal

- Taxa: $R = 0.5$ bits/uso do canal.
- Probabilidade de erro: $P_B = \epsilon^2 \cdot (1/2) = \epsilon^2/2$.

Exemplo: Esquemas de codificação simples para o BEC — Repetição

Repetição com $n = 3$.

m	x	$p(x)$	y	$\hat{m}$	$p(y)$	y	$\hat{m}$	$p(y)$	y	$\hat{m}$	$p(y)$
0	000	1/2	000	0	$\bar{\varepsilon}^3/2$	100	—	0	?00	0	$\bar{\varepsilon}^2\varepsilon/2$
1	111	1/2	001	—	0	101	—	0	?01	—	0
			00?	0	$\bar{\varepsilon}^2\varepsilon/2$	10?	—	0	?0?	0	$\bar{\varepsilon}\varepsilon^2/2$
			010	—	0	110	—	0	?10	—	0
			011	—	0	111	1	$\bar{\varepsilon}^3/2$	?11	1	$\bar{\varepsilon}^2\varepsilon/2$
			01?	—	0	11?	1	$\bar{\varepsilon}^2\varepsilon/2$	?1?	1	$\bar{\varepsilon}\varepsilon^2/2$
			0?0	0	$\bar{\varepsilon}^2\varepsilon/2$	1?0	—	0	??0	0	$\bar{\varepsilon}\varepsilon^2/2$
			0?1	—	0	1?1	1	$\bar{\varepsilon}^2\varepsilon/2$	??1	1	$\bar{\varepsilon}\varepsilon^2/2$
			0??	0	$\bar{\varepsilon}\varepsilon^2/2$	1??	1	$\bar{\varepsilon}\varepsilon^2/2$	???	0	ε^3

- Taxa: $R = 0.33\bar{3}$ bits/uso do canal.
- Probabilidade de erro: $P_B = \varepsilon^3 \cdot (1/2) = \varepsilon^3/2$.

Exemplo: Esquemas de codificação simples para o BEC — Paridade simples

No **código de paridade simples de comprimento n** , tem-se $M = 2^{n-1}$, de modo que o alfabeto da mensagem é $[0 : M) = \{0, 1, \dots, 2^{n-1} - 1\}$. A codificação da mensagem $m \in [0 : M)$ consiste em converter m para binário ($n - 1$ bits) e adicionar um último bit de modo que o número total de bits 1 seja par.

Exemplo: Esquemas de codificação simples para o BEC — Paridade simples

No **código de paridade simples de comprimento n** , tem-se $M = 2^{n-1}$, de modo que o alfabeto da mensagem é $[0 : M) = \{0, 1, \dots, 2^{n-1} - 1\}$. A codificação da mensagem $m \in [0 : M)$ consiste em converter m para binário ($n - 1$ bits) e adicionar um último bit de modo que o número total de bits 1 seja par.

Exemplo: Considere o código de paridade com $n = 4$ sobre o BEC.

- (a) Qual a sequência de palavras-código referente à sequência de mensagens 5, 7, 6, 0?
- (b) Suponha agora que, em outra transmissão, a sequência observada na saída do canal seja 111? 0000 1??? 0??? 1????. Como você decodificaria esta sequência?

Exemplo: Esquemas de codificação simples para o BEC — Paridade simples

No **código de paridade simples de comprimento n** , tem-se $M = 2^{n-1}$, de modo que o alfabeto da mensagem é $[0 : M) = \{0, 1, \dots, 2^{n-1} - 1\}$. A codificação da mensagem $m \in [0 : M)$ consiste em converter m para binário ($n - 1$ bits) e adicionar um último bit de modo que o número total de bits 1 seja par.

Exemplo: Considere o código de paridade com $n = 4$ sobre o BEC.

- (a) Qual a sequência de palavras-código referente à sequência de mensagens 5, 7, 6, 0?
- (b) Suponha agora que, em outra transmissão, a sequência observada na saída do canal seja 111? 0000 1??? 0??? 1????. Como você decodificaria esta sequência?

Respostas:

- (a) 1010 1111 1100 0000.
- (b) 7, 0, 5 (ou 6), 1 (ou 2 ou 4 ou 7), 0 (ou qualquer outro).

Exemplo: Esquemas de codificação simples para o BEC — Paridade simples

Paridade simples com $n = 2$.

Codificador

m	x	$p(x)$
0	00	$1/2$
1	11	$1/2$

- Coincide com o código de repetição com $n = 2$!

Exemplo: Esquemas de codificação simples para o BEC — Paridade simples

Paridade simples com $n = 3$.

m	x	$p(x)$	y	$\hat{m}$	$p(y)$	y	$\hat{m}$	$p(y)$	y	$\hat{m}$	$p(y)$
0	000	1/4	000	0	$\bar{\epsilon}^3/4$	100	—	0	?00	0	$\bar{\epsilon}^2\epsilon/4$
1	011	1/4	001	—	0	101	2	$\bar{\epsilon}^3/4$	?01	2	$\bar{\epsilon}^2\epsilon/4$
2	101	1/4	00?	0	$\bar{\epsilon}^2\epsilon/4$	10?	2	$\bar{\epsilon}^2\epsilon/4$	?0?	0	$\bar{\epsilon}\epsilon^2/2$
3	110	1/4	010	—	0	110	3	$\bar{\epsilon}^3/4$	?10	3	$\bar{\epsilon}^2\epsilon/4$
$\log_2 M = 2$ bits $n = 3$ usos do canal			011	1	$\bar{\epsilon}^3/4$	111	—	0	?11	1	$\bar{\epsilon}^2\epsilon/4$
			01?	1	$\bar{\epsilon}^2\epsilon/4$	11?	3	$\bar{\epsilon}^2\epsilon/4$	?1?	1	$\bar{\epsilon}\epsilon^2/2$
			0?0	0	$\bar{\epsilon}^2\epsilon/4$	1?0	3	$\bar{\epsilon}^2\epsilon/4$	??0	0	$\bar{\epsilon}\epsilon^2/2$
			0?1	1	$\bar{\epsilon}^2\epsilon/4$	1?1	2	$\bar{\epsilon}^2\epsilon/4$	??1	1	$\bar{\epsilon}\epsilon^2/2$
			0??	0	$\bar{\epsilon}\epsilon^2/2$	1??	2	$\bar{\epsilon}\epsilon^2/2$	???	0	ϵ^3

- Taxa: $R = 0.66\bar{6}$ bits/uso do canal.
- Probabilidade de erro: $P_B = 6 \cdot \bar{\epsilon}\epsilon^2/2 \cdot (1/2) + 1 \cdot \epsilon^3 \cdot (3/4) = 3\epsilon^2/2 - 3\epsilon^3/4$.

Exemplo: Esquemas de codificação simples para o BEC — Comparação

Exemplo para $\varepsilon = 0.2$.

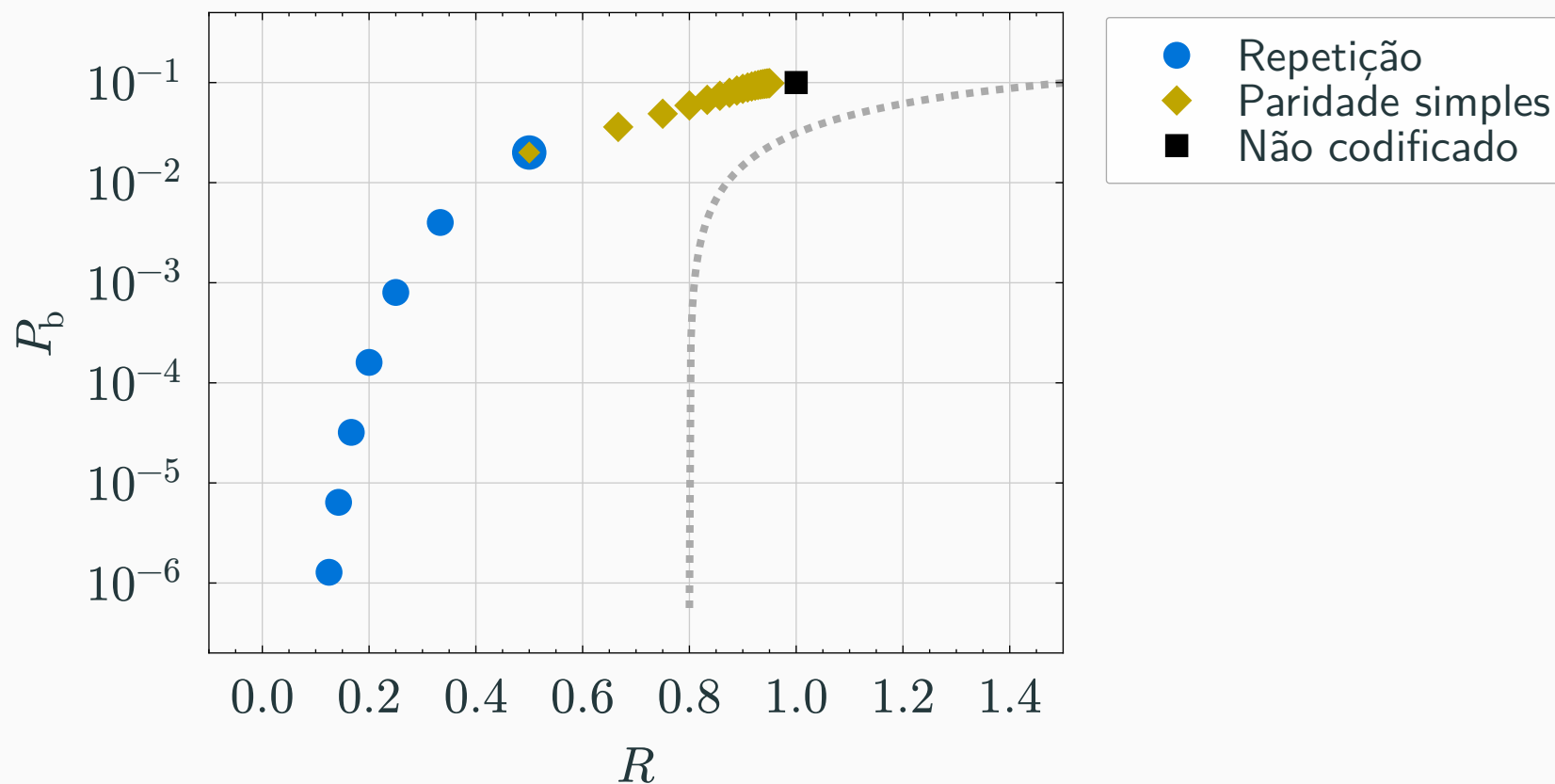
Código	R	P_B	P_b
Não codificado	1	10%	10%
Paridade simples $n = 3$	$0.66\bar{6}$	5.4%	3.6%
Repetição $n = 2$	0.5	2%	2%
Repetição $n = 3$	$0.33\bar{3}$	0.4%	0.4%

Observação: O cálculo da probabilidade de erro de bit P_b não será abordado aqui. No entanto, é possível mostrar que

$$\frac{P_B}{\log_2 M} \leq P_b \leq P_B.$$

Exemplo: Esquemas de codificação simples para o BEC — Comparação

Exemplo para $\varepsilon = 0.2$.



Capacidade de um DMC

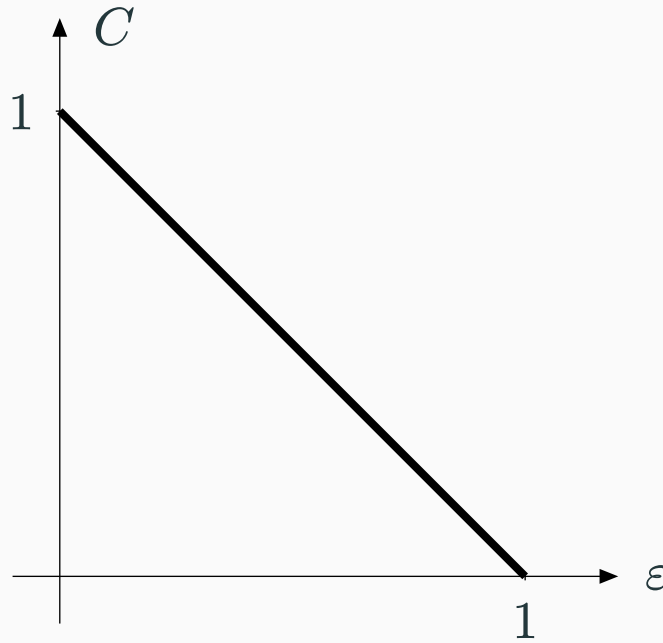
Teorema. (*C. E. Shannon, 1948*) Considere um DMC. Defina a **capacidade** do canal como

$$C = \max_{p_X} I(X; Y).$$

1. Se $R < C$, então para todo $\delta > 0$ (por menor que seja) existe um esquema de codificação (com n suficientemente grande) de taxa R ou maior e $P_B < \delta$.
2. Se $R > C$, não existe tal esquema. Mais precisamente, para todo $R > C$ existe $\delta > 0$ tal que todo esquema de taxa R ou maior tem $P_B > \delta$.

Exercício: Capacidade do BEC

Mostre que a capacidade do $\text{BEC}(\varepsilon)$ é dada por $C = 1 - \varepsilon$, em bits por uso do canal, alcançada com entrada uniforme.



Solução: Capacidade do BEC

Vamos pelo caminho

$$I(X; Y) = H(X) - H(X | Y).$$

Assuma uma distribuição de entrada genérica: $p_X = \{0 : \bar{\alpha}, 1 : \alpha\}$, com $\alpha + \bar{\alpha} = 1$.

Então, $H(X) = \mathcal{H}(\alpha)$. Quanto à $H(X | Y)$, temos:

- $H(X | Y = 0) = H(X | Y = 1) = 0$. (Por quê?)
- $H(X | Y = ?) = H(X) = \mathcal{H}(\alpha)$. (Por quê?)

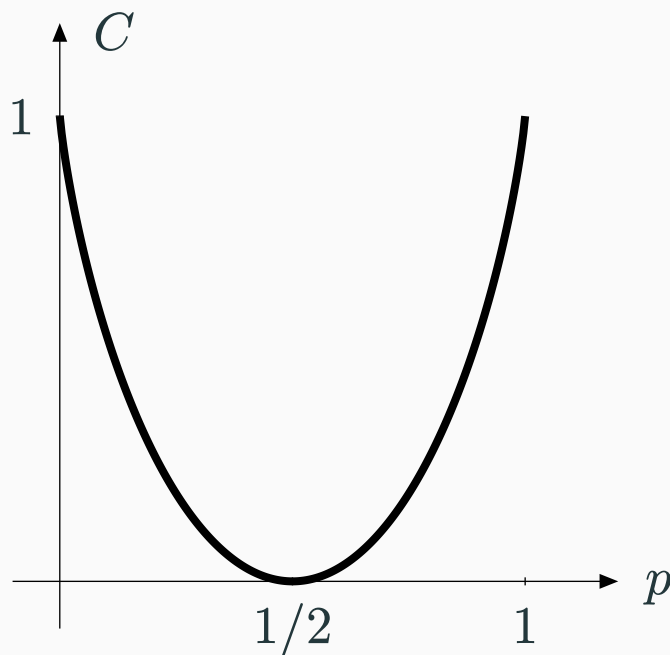
Assim, $H(X | Y) = \sum_y p(y) H(X | Y = y) = \varepsilon \mathcal{H}(\alpha)$. Portanto,

$$I(X; Y) = \mathcal{H}(\alpha) - \varepsilon \mathcal{H}(\alpha) = (1 - \varepsilon) \mathcal{H}(\alpha).$$

Esse valor é maximizado quando $\alpha = 1/2$, com valor máximo $C = 1 - \varepsilon$.

Exercício: Capacidade do BSC

Mostre que a capacidade do $\text{BSC}(p)$ é dada por $C = 1 - \mathcal{H}(p)$, em bits por uso do canal, alcançada com entrada uniforme.



Solução: Capacidade do BSC

Desta vez, vamos pelo caminho

$$I(X; Y) = H(Y) - H(Y | X).$$

Novamente, assuma distribuição de entrada $p_X = \{0 : \bar{\alpha}, 1 : \alpha\}$, com $\alpha + \bar{\alpha} = 1$.

Temos $H(Y | X = 0) = H(Y | X = 1) = \mathcal{H}(p)$, de modo que $H(Y | X) = \mathcal{H}(p)$.

Temos também $p_Y = \{0 : \bar{\alpha}\bar{p} + \alpha p, 1 : \dots\}$, de modo que $H(Y) = \mathcal{H}(\bar{\alpha}\bar{p} + \alpha p)$. Portanto,

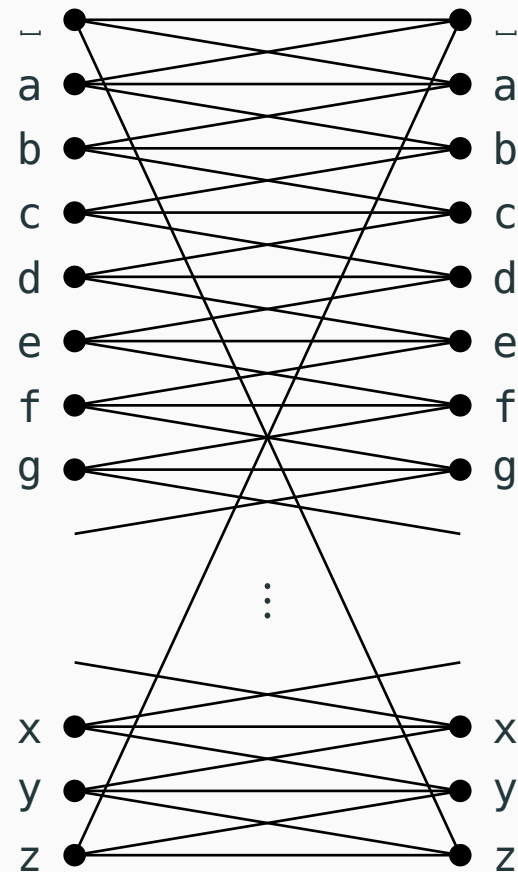
$$I(X; Y) = \mathcal{H}(\bar{\alpha}\bar{p} + \alpha p) - \mathcal{H}(p).$$

Esse valor é maximizado quando $\alpha = 1/2$, com valor máximo $C = 1 - \mathcal{H}(p)$.

Exercício: Máquina de escrever ruidosa

Determine a capacidade do DMC ao lado, cujos alfabetos de entrada e de saída são $\mathcal{X} = \mathcal{Y} = \{_, a, b, \dots, z\}$, ou seja, as 26 letras mais o espaço.

As letras estão dispostas em círculo: ao se tentar digitar a, o que sai é $_$, a ou b, com probabilidade $1/3$ cada; ao se digitar b, o que sai é a, b ou c e assim por diante, sendo a última letra, z, adjacente à primeira, $_$.



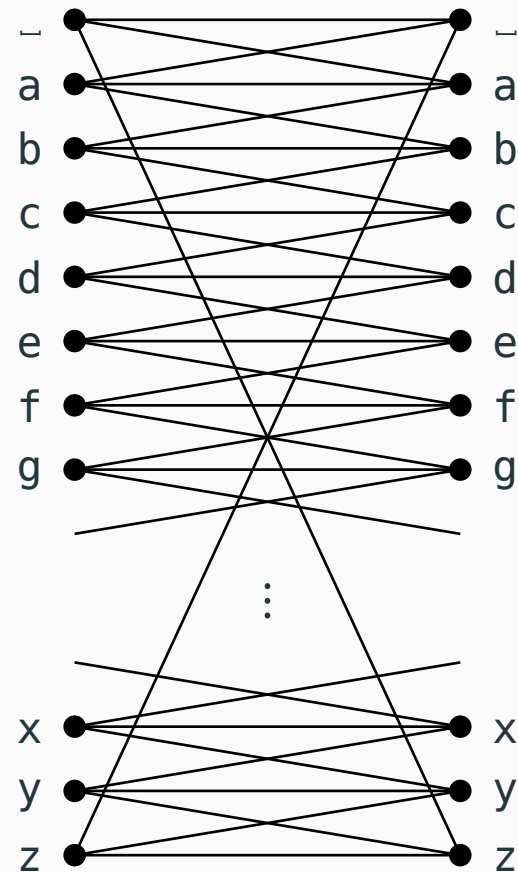
Exercício: Máquina de escrever ruidosa

Determine a capacidade do DMC ao lado, cujos alfabetos de entrada e de saída são $\mathcal{X} = \mathcal{Y} = \{_, a, b, \dots, z\}$, ou seja, as 26 letras mais o espaço.

As letras estão dispostas em círculo: ao se tentar digitar a, o que sai é $_$, a ou b, com probabilidade $1/3$ cada; ao se digitar b, o que sai é a, b ou c e assim por diante, sendo a última letra, z, adjacente à primeira, $_$.

Resposta:

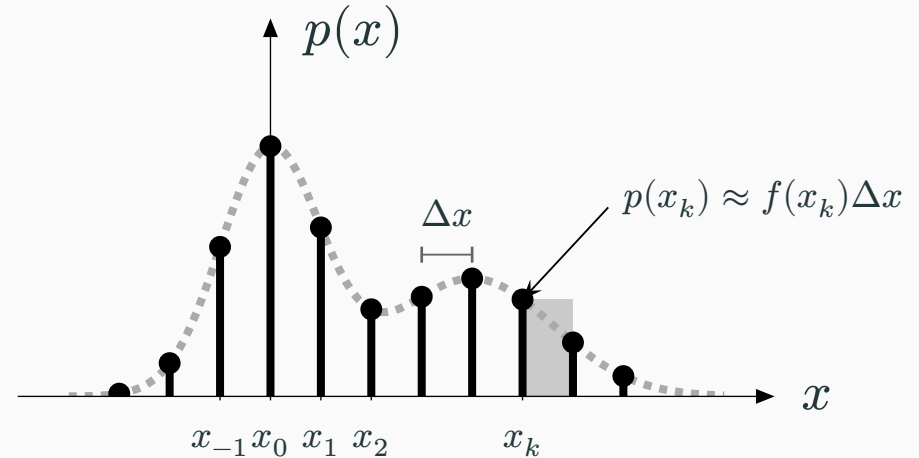
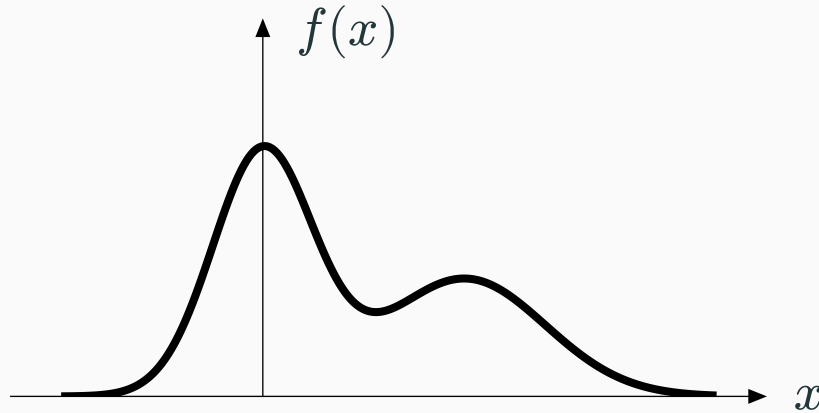
$C = \log 9 = 2 \text{ trits/uso do canal} \approx 3.17 \text{ bits/uso do canal}.$



Entropia diferencial

Entropia de uma variável aleatória contínua?

Seja X uma variável aleatória contínua com pdf $f(x)$.



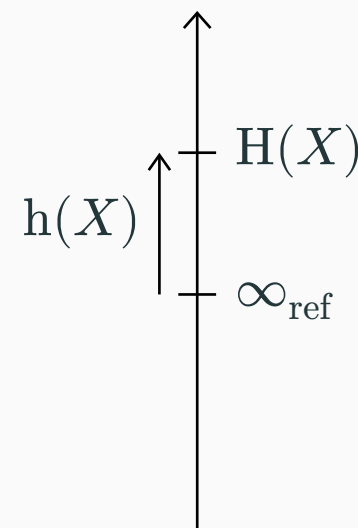
A entropia de X é dada por

$$H(X) = \lim_{\Delta x \rightarrow 0} \sum_{k=-\infty}^{\infty} p(x_k) \log \frac{1}{p(x_k)}.$$

Entropia de uma variável aleatória contínua?

Continuando:

$$\begin{aligned} H(X) &= \lim_{\Delta x \rightarrow 0} \sum_{k=-\infty}^{\infty} p(x_k) \log \frac{1}{p(x_k)} \\ &= \lim_{\Delta x \rightarrow 0} \sum_{k=-\infty}^{\infty} f(x_k) \Delta x \log \frac{1}{f(x_k) \Delta x} \\ &= \lim_{\Delta x \rightarrow 0} \left(\sum_{k=-\infty}^{\infty} f(x_k) \log \frac{1}{f(x_k)} \Delta x + \sum_{k=-\infty}^{\infty} f(x_k) \log \frac{1}{\Delta x} \Delta x \right) \\ &= \underbrace{\int_{-\infty}^{\infty} f(x) \log \left(\frac{1}{f(x)} \right) dx}_{\stackrel{\text{def}}{=} h(X)} + \underbrace{\left(\lim_{\Delta x \rightarrow 0} \log \frac{1}{\Delta x} \right)}_{\text{referência comum } (\infty_{\text{ref}})} \underbrace{\int_{-\infty}^{\infty} f(x) dx}_{=1} \\ &= \infty_{\text{ref}} + h(X). \end{aligned}$$

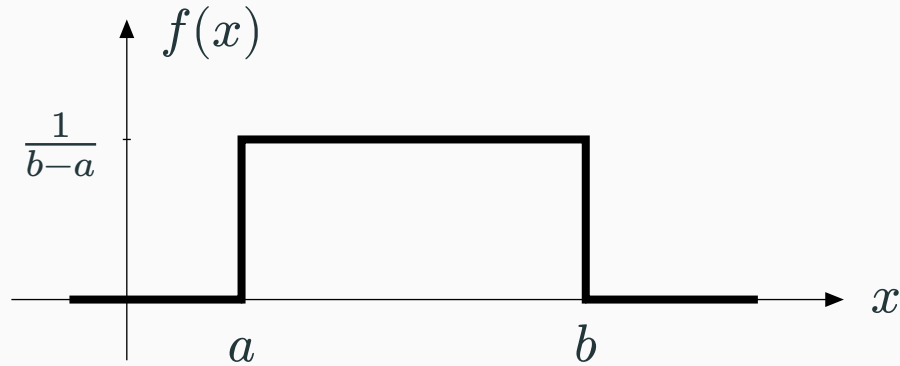


Define-se a **entropia diferencial de X** como

$$h(X) = \int_{-\infty}^{\infty} f(x) \log \frac{1}{f(x)} dx.$$

Exemplo: Distribuição uniforme contínua

Seja $X \sim \text{Uniform}(a, b)$.

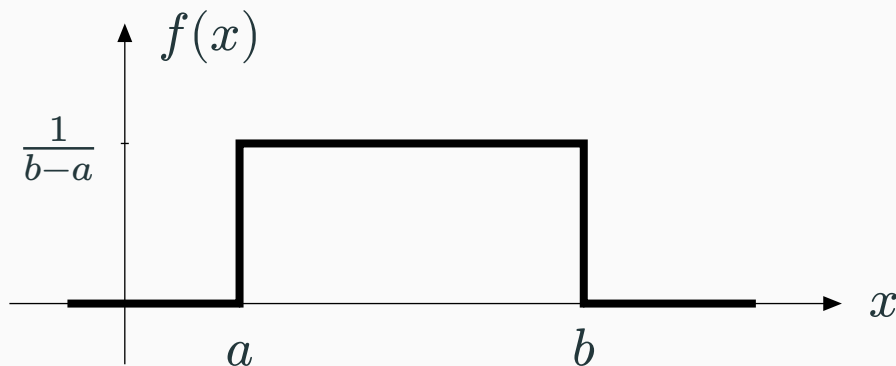


Mostre que

$$h(X) = \log(b - a).$$

Exemplo: Distribuição uniforme contínua

Seja $X \sim \text{Uniform}(a, b)$.



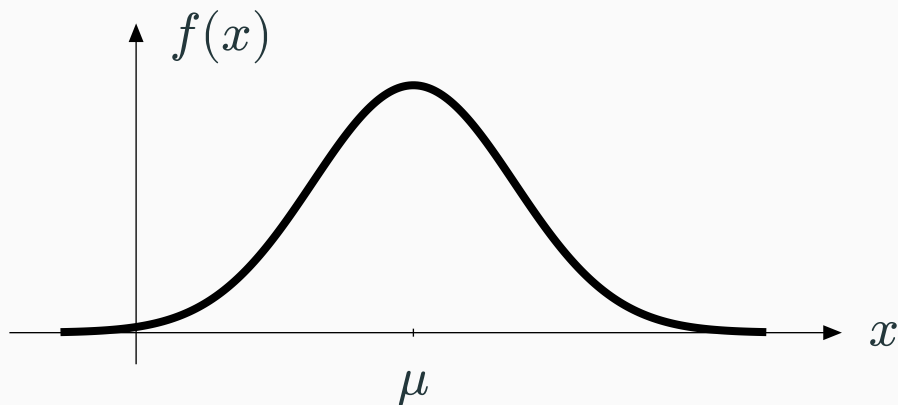
Mostre que

$$h(X) = \log(b - a).$$

Pergunta: Como ficam os casos $X \sim \text{Uniform}(0, 2)$, $X \sim \text{Uniform}(0, 1)$ e $X \sim \text{Uniform}(0, 0.5)$, em bits? Conclua.

Exemplo: Distribuição normal

Seja $X \sim \text{Normal}(\mu, \sigma^2)$.



Mostre que

$$h(X) = \frac{1}{2} \log(2\pi e\sigma^2).$$

Algumas distribuições de entropia máxima

1. Dentre todas as distribuições restritas ao suporte $[a, b]$, a uniforme é aquela de maior entropia diferencial.

Demonstração: Ver [2, Exercise 9.26, p. 623].

2. Dentre todas as distribuições com a mesma variância, a gaussiana é aquela de maior entropia diferencial.

Demonstração: Ver [2, Example 9.8, p. 595].

Observação: O resultado 2 é o mais importante. Será usado no cálculo da capacidade do canal AWGN.

Informação mútua entre duas variáveis aleatórias contínuas

Temos:

$$\begin{aligned} I(X; Y) &= H(X) - H(X | Y) \\ &= \infty_{\text{ref}} + h(X) - \infty_{\text{ref}} - h(X | Y) \\ &= h(X) - h(X | Y), \end{aligned}$$

onde $h(X | Y)$ é a **entropia diferencial condicional de X dado Y** , definida de maneira análoga ao caso discreto.

Observação: As propriedades da informação mútua se mantêm! Por exemplo, $I(X; Y) = I(Y; X)$ e $I(X; Y) \geq 0$.

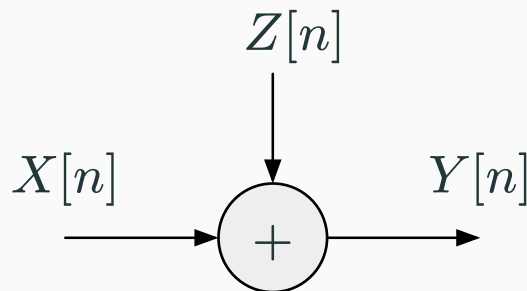
Capacidade do canal AWGN

Canal em tempo discreto

Relembrando o **canal AWGN em tempo discreto**:

$$Y[n] = X[n] + Z[n],$$

onde $X[n]$ é a *entrada*, $Y[n]$ é a *saída* e $Z[n]$ é o *ruído*, com $X[n]$ e $Z[n]$ independentes.



$$Z[n] \stackrel{\text{iid}}{\sim} \text{Normal}(0, \sigma_Z^2).$$

Além disso, impõe-se a *restrição de potência*

$$\mathbb{E}[X^2[n]] = \sigma_X^2,$$

sem a qual o problema fica trivial (por quê?).

Vamos pelo caminho $I(X; Y) = h(Y) - h(Y | X)$. Temos:

$$\begin{aligned} I(X; Y) &= h(Y) - h(Y | X) \\ &= h(Y) - h(X + Z | X) \\ &= h(Y) - h(Z | X) \\ &= h(Y) - h(Z). \end{aligned}$$

Como $Z[n] \stackrel{\text{iid}}{\sim} \text{Normal}(0, \sigma_Z^2)$, obtemos

$$I(X; Y) = h(Y) - \frac{1}{2} \log_2(2\pi e \sigma_Z^2).$$

Capacidade do canal em tempo discreto

Maximizando sobre todas as possíveis distribuições de entrada:

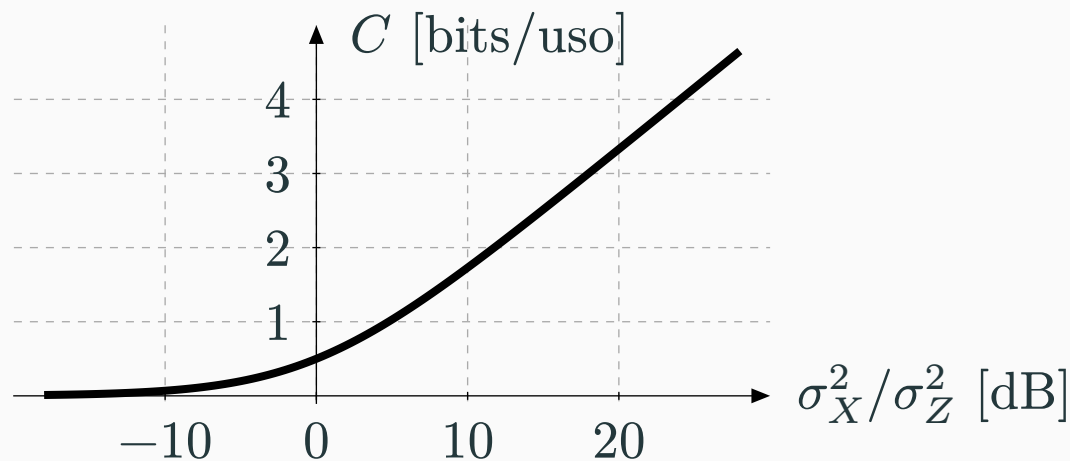
$$\begin{aligned} C &= \max_{f_X: \mathbb{E}[X^2] = \sigma_X^2} I(X; Y) \\ &= \max_{f_X: \mathbb{E}[X^2] = \sigma_X^2} \left[h(Y) - \frac{1}{2} \log_2(2\pi e \sigma_Z^2) \right] \\ &= \left[\max_{f_X: \mathbb{E}[X^2] = \sigma_X^2} h(Y) \right] - \frac{1}{2} \log_2(2\pi e \sigma_Z^2) \\ &= \frac{1}{2} \log_2(2\pi e (\sigma_X^2 + \sigma_Z^2)) - \frac{1}{2} \log_2(2\pi e \sigma_Z^2) \\ &= \frac{1}{2} \log_2 \left(\frac{\sigma_X^2 + \sigma_Z^2}{\sigma_Z^2} \right). \end{aligned}$$

Capacidade do canal em tempo discreto

Finalmente:

$$C = \frac{1}{2} \log_2 \left(1 + \frac{\sigma_X^2}{\sigma_Z^2} \right), \quad [\text{bits/uso}]$$

alcançada quando $X \sim \text{Normal}(0, \sigma_X^2)$.

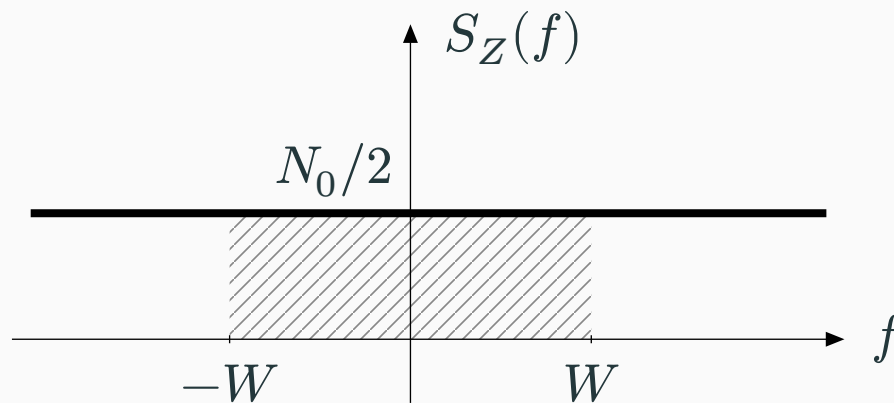
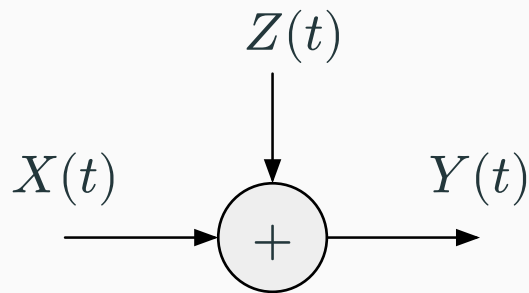


Canal em tempo contínuo e limitado em banda

Relembrando o **canal AWGN em tempo contínuo** e limitado em banda:

$$Y(t) = X(t) + Z(t),$$

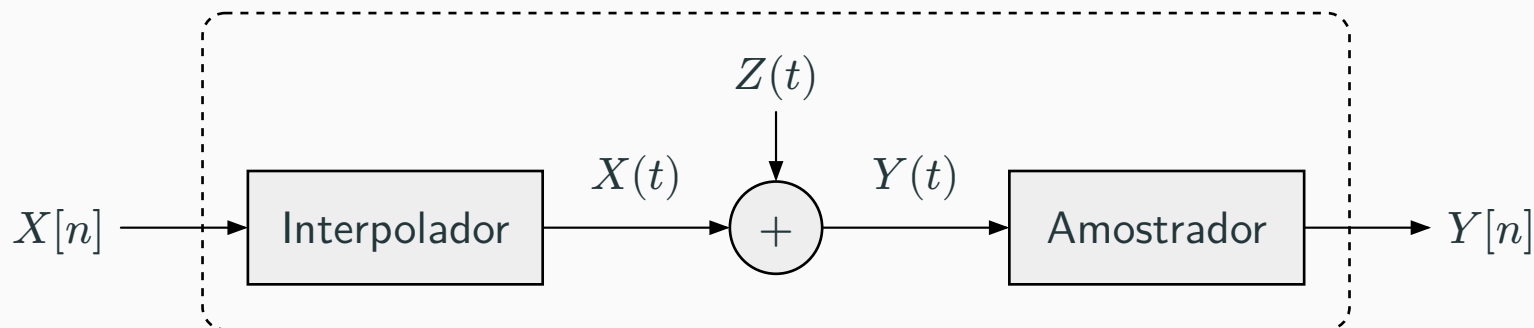
onde a entrada $X(t)$ é limitada à faixa $[-W, W]$ e tem potência P_X , e o ruído $Z(t)$ é branco e gaussiano, com densidade espectral de potência $S_Z(f) = N_0/2$.



A potência do ruído na faixa de interesse é $P_Z = N_0 W$.

Intuição

Pelo teorema da amostragem, amostras à taxa $2W$ determinam um sinal limitado a $[-W, W]$. Assim, o canal contínuo equivale a um canal discreto com $2W$ usos por segundo.



- Amostras do sinal: Potência $\sigma_X^2 = P_X$.
- Amostras do ruído: Potência $\sigma_Z^2 = P_Z = N_0 W$.
- Razão sinal-ruído: $\text{SNR} = P_X / P_Z$.

Cada uso do canal fornece $\frac{1}{2} \log_2(1 + \text{SNR})$ bits e há $2W$ usos por segundo.

Teorema. (*C. E. Shannon, 1948*) A capacidade do canal AWGN, em bits/s, é dada por

$$C = W \log_2(1 + \text{SNR}),$$

onde $\text{SNR} = P_X / (N_0 W)$.

Para uma discussão mais cuidadosa, ver [4, Chapter 9].

Exercício: Dois exemplos

Considere dois canais AWGN, ambos com densidade espectral de potência do ruído $N_0/2$ fixa:

- **Canal telefônico:** $W = 3.3$ kHz, $\text{SNR} = 33$ dB.
- **Enlace de espaço profundo:** $W = 1$ MHz, $\text{SNR} = -29$ dB.

- (a) Determine a capacidade de cada um.
- (b) Havendo recursos para dobrar *ou* a potência transmitida *ou* a largura de banda, mas não as duas, qual escolher em cada caso?

Exercício: Dois exemplos

Considere dois canais AWGN, ambos com densidade espectral de potência do ruído $N_0/2$ fixa:

- **Canal telefônico:** $W = 3.3$ kHz, $\text{SNR} = 33$ dB.
- **Enlace de espaço profundo:** $W = 1$ MHz, $\text{SNR} = -29$ dB.

- (a) Determine a capacidade de cada um.
- (b) Havendo recursos para dobrar *ou* a potência transmitida *ou* a largura de banda, mas não as duas, qual escolher em cada caso?

Respostas:

	Original	Potência dobrada	Banda dobrada
Telefônico	36.2 kbit/s	39.5 kbit/s	65.8 kbit/s
Espaço profundo	1.815 kbit/s	3.628 kbit/s	1.816 kbit/s

Regime de alta SNR

Quando $\text{SNR} \gg 1$, tem-se $\log_2(1 + \text{SNR}) \approx \log_2 \text{SNR}$, e portanto

$$C \approx W \log_2 \text{SNR} = W \log_2 \frac{P_X}{N_0 W}.$$

- Dobrar a potência (+3 dB) aumenta a capacidade em W bits/s. (Ganho logarítmico.)
- Dobrar a banda, mantendo a potência, quase dobra a capacidade. (Ganho quase linear.)

Este caso também é chamado de **regime limitado em banda**: o recurso escasso é a banda. Usam-se constelações de alta ordem (muitos bits por símbolo) e códigos de taxa alta.

É o caso do canal telefônico.

Regime de baixa SNR

Quando $\text{SNR} \ll 1$, tem-se $\log_2(1 + \text{SNR}) \approx \text{SNR} \log_2 e$, e portanto

$$C \approx W \text{ SNR} \log_2 e = \frac{P_X}{N_0} \log_2 e.$$

- Dobrar a potência (+3 dB) dobra a capacidade. (Ganho linear.)
- Aumentar a banda não ajuda.

Este caso também é chamado de **regime limitado em potência**: o recurso escasso é a potência. Usam-se constelações de baixa ordem (poucos bits por símbolo) e códigos de taxa baixa.

É o caso do enlace de espaço profundo.

Limite de Shannon

Escreva $P_X = E_b R_b$, onde E_b é a energia por bit e R_b é a taxa de bits, e sejam

- $\rho = R_b/W$ a **eficiência espectral** [bits/s/Hz];
- $\gamma = E_b/N_0$ a **razão sinal–ruído de bit**.

Note que o produto das duas é a razão sinal–ruído do canal:

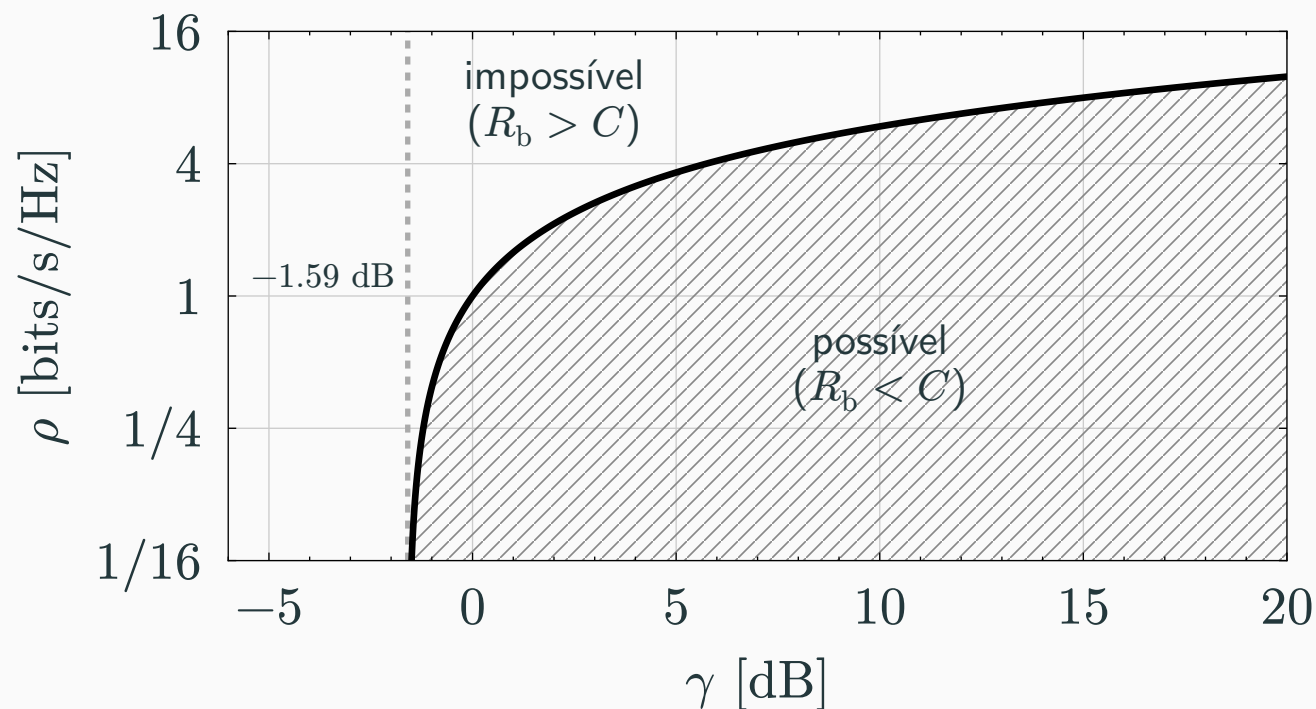
$$\gamma\rho = \frac{E_b}{N_0} \cdot \frac{R_b}{W} = \frac{P_X}{P_Z} = \text{SNR}.$$

A condição “ $R < C$ ” para comunicação confiável fica

$$R_b < W \log_2(1 + \text{SNR}) \iff \rho < \log_2(1 + \gamma\rho) \iff \gamma > \frac{2^\rho - 1}{\rho}.$$

Limite de Shannon

A curva $\gamma = (2^\rho - 1)/\rho$ é conhecida como **limite de Shannon**.



Para $\rho \rightarrow 0$, tem-se $\gamma \rightarrow \ln 2 \approx -1.59$ dB: valor mínimo para comunicação confiável, mesmo com banda infinita.

Curiosidade: Invertendo o limite de Shannon

Considere o limite de Shannon

$$\gamma = \frac{2^\rho - 1}{\rho}.$$

Como isolar ρ ? Isso pode ser feito através da função W de Lambert, definida pelos ramos da relação inversa de $w \mapsto we^w$. Convém usar sua versão na base b , dada por

$$W_k^{(b)}(x) = \frac{W_k(x \ln b)}{\ln b},$$

que inverte $w \mapsto wb^w$. Tomando $b = 1/2$ obtém-se a inversa de $w \mapsto w2^{-w}$, e para $\gamma > \ln 2$ a solução é

$$\rho = W_{-1}^{(1/2)}\left(\frac{2^{-1/\gamma}}{\gamma}\right) - \frac{1}{\gamma}.$$

Exercício: Dois exemplos, novamente

Retomando os dois canais AWGN, agora com sistemas reais operando neles:

- **Telefônico:** $W = 3.3$ kHz e $\text{SNR} = 33$ dB; modem V.34 com $R_b = 33.6$ kbit/s.
- **Espaço profundo:** $W = 1$ MHz e $\text{SNR} = -29$ dB; sonda com $R_b = 1.0$ kbit/s.

(a) Determine ρ e γ de cada sistema.

(b) Determine a distância de cada um até o limite de Shannon, em dB.

Exercício: Dois exemplos, novamente

Retomando os dois canais AWGN, agora com sistemas reais operando neles:

- **Telefônico:** $W = 3.3$ kHz e $\text{SNR} = 33$ dB; modem V.34 com $R_b = 33.6$ kbit/s.
- **Espaço profundo:** $W = 1$ MHz e $\text{SNR} = -29$ dB; sonda com $R_b = 1.0$ kbit/s.

(a) Determine ρ e γ de cada sistema.

(b) Determine a distância de cada um até o limite de Shannon, em dB.

Respostas:

	ρ	γ	$(2^\rho - 1)/\rho$	Distância
Telefônico	10.2 bits/s/Hz	22.9 dB	20.6 dB	2.4 dB
Espaço profundo	0.001 bits/s/Hz	1.0 dB	-1.6 dB	2.6 dB

Referências

- [1] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*. University of Illinois Press, 1949.
- [2] S. Haykin, *Communication Systems*, 4th ed. John Wiley & Sons, 2001.
- [3] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.
- [4] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley-Interscience, 2006.