

Teoria da Informação

Licenciatura em Engenharia Informática

Miguel Barão

Escola de Ciência e Tecnologia
Universidade de Évora

Resumo

Nestes slides estudam-se canais discretos sem memória. Estes canais permitem o envio de símbolos de um alfabeto discreto \mathcal{X} sendo recebidos símbolos de um alfabeto \mathcal{Y} . A transmissão está sujeita a erros aleatórios. É definida a *capacidade do canal* segundo Shannon e são apresentados vários casos particulares para os quais a capacidade pode ser determinada explicitamente.

Finalmente é apresentado o teorema da codificação de canal de Shannon, um resultado central em teoria da informação.

Canais de comunicação

- Transmissão de informação por um canal ruidoso

- Definição de canal discreto sem memória

Capacidade do canal

- Definição

- Propriedades

- Casos particulares com solução explícita

- Algoritmos iterativos para calcular a capacidade

- Concatenação de canais

Teorema da codificação de canal

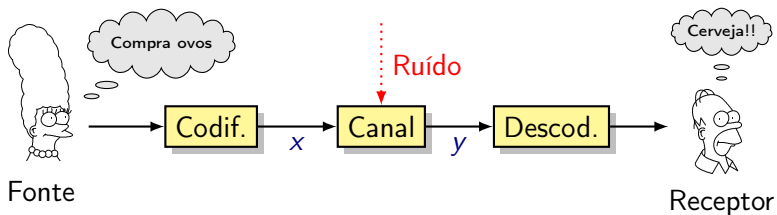
- Funções de codificação e decodificação

- Probabilidades de erro

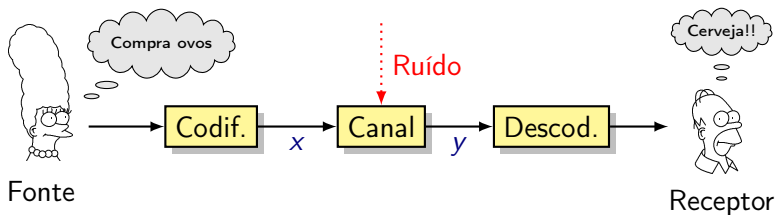
- Ritmo de transmissão

- Teorema da codificação de canal

Transmissão de informação por um canal ruidoso

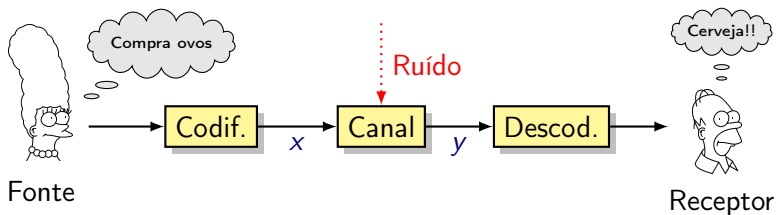


Transmissão de informação por um canal ruidoso



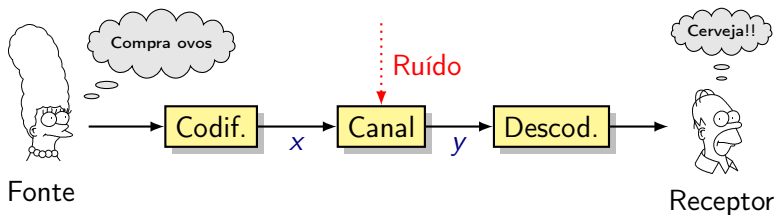
- 1 A fonte gera uma mensagem.

Transmissão de informação por um canal ruidoso



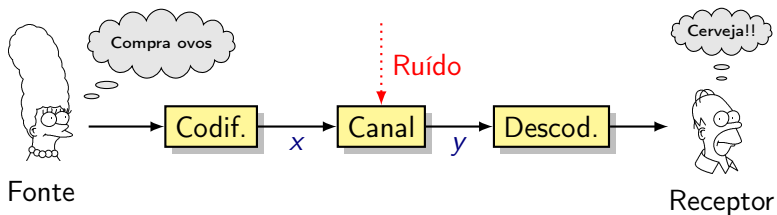
- 1 A fonte gera uma mensagem.
- 2 O codificador transforma a mensagem para a adaptar ao canal usado.

Transmissão de informação por um canal ruidoso



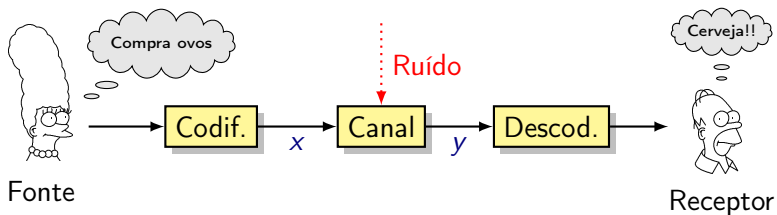
- 1 A fonte gera uma mensagem.
- 2 O codificador transforma a mensagem para a adaptar ao canal usado.
- 3 A mensagem codificada é enviada pelo canal. O canal não é fiável e pode corromper a mensagem.

Transmissão de informação por um canal ruidoso

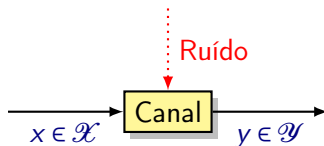


- 1 A fonte gera uma mensagem.
- 2 O codificador transforma a mensagem para a adaptar ao canal usado.
- 3 A mensagem codificada é enviada pelo canal. O canal não é fiável e pode corromper a mensagem.
- 4 O decodificador recebe a mensagem do canal e tenta
 - 4.1 detectar erros de transmissão;
 - 4.2 corrigir os erros.

Transmissão de informação por um canal ruidoso



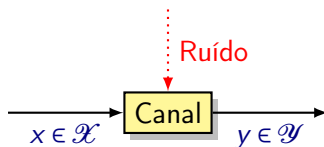
- 1 A fonte gera uma mensagem.
- 2 O codificador transforma a mensagem para a adaptar ao canal usado.
- 3 A mensagem codificada é enviada pelo canal. O canal não é fiável e pode corromper a mensagem.
- 4 O decodificador recebe a mensagem do canal e tenta
 - 4.1 detectar erros de transmissão;
 - 4.2 corrigir os erros.
- 5 A mensagem decodificada é entregue ao receptor.



Definição (Canal discreto sem memória)

Consiste em:

- alfabeto de entrada \mathcal{X}
- alfabeto de saída \mathcal{Y}
- probabilidades de transição $p(y|x)$



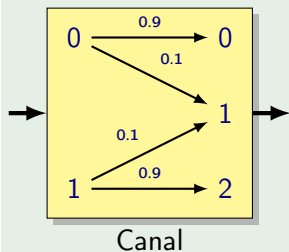
Definição (Canal discreto sem memória)

Consiste em:

- alfabeto de entrada \mathcal{X}
- alfabeto de saída \mathcal{Y}
- probabilidades de transição $p(y|x)$

O canal é **sem memória** se a distribuição de probabilidade da saída depende da entrada e é condicionalmente independente das entradas e saídas passadas: $p(y_t|x_t, x_{t-1}, \dots, y_{t-1}, \dots) = p(y_t|x_t)$.

Exemplo

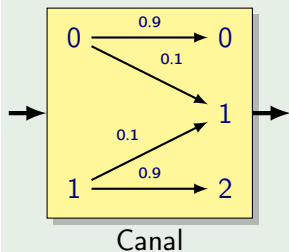


Este canal é caracterizado por:

- alfabeto de entrada $\mathcal{X} = \{0, 1\}$
- alfabeto de saída $\mathcal{Y} = \{0, 1, 2\}$
- probabilidades de transição:

$p(y x)$	$x = 0$	$x = 1$
$y = 0$	0.9	0.0
$y = 1$	0.1	0.1
$y = 2$	0.0	0.9

Exemplo



Este canal é caracterizado por:

- alfabeto de entrada $\mathcal{X} = \{0, 1\}$
- alfabeto de saída $\mathcal{Y} = \{0, 1, 2\}$
- probabilidades de transição:

$p(y x)$	$x = 0$	$x = 1$
$y = 0$	0.9	0.0
$y = 1$	0.1	0.1
$y = 2$	0.0	0.9

Note que os alfabetos de entrada e saída podem ser diferentes!

Definição (Capacidade de um canal discreto sem memória)

$$C = \max_{p(x)} I(X; Y)$$

Observações:

- Um canal em que a saída Y é independente de X não permite transmitir informação.
- Se é possível tornar X e Y mais dependentes, então a capacidade do canal aumenta.
- A capacidade é calculada como um problema de otimização: encontrar $p(x)$ que maximiza a informação mútua $I(X; Y)$.

- $C \geq 0$ uma vez que $I(X; Y) \geq 0$.

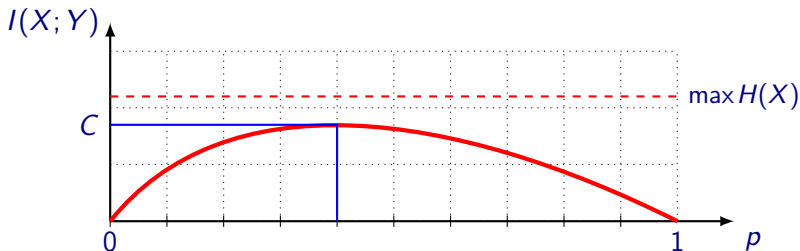
- $C \geq 0$ uma vez que $I(X; Y) \geq 0$.
- $C \leq \log|\mathcal{X}|$ uma vez que
 $C = \max I(X; Y) \leq \max H(X) \leq \log|\mathcal{X}|$.

- $C \geq 0$ uma vez que $I(X; Y) \geq 0$.
- $C \leq \log|\mathcal{X}|$ uma vez que
$$C = \max I(X; Y) \leq \max H(X) \leq \log|\mathcal{X}|.$$
- $C \leq \log|\mathcal{Y}|$ pelo mesmo motivo.

- $C \geq 0$ uma vez que $I(X; Y) \geq 0$.
- $C \leq \log|\mathcal{X}|$ uma vez que
 $C = \max I(X; Y) \leq \max H(X) \leq \log|\mathcal{X}|$.
- $C \leq \log|\mathcal{Y}|$ pelo mesmo motivo.
- $I(X; Y)$ é uma função contínua de $p(x)$.

- $C \geq 0$ uma vez que $I(X; Y) \geq 0$.
- $C \leq \log|\mathcal{X}|$ uma vez que
 $C = \max I(X; Y) \leq \max H(X) \leq \log|\mathcal{X}|$.
- $C \leq \log|\mathcal{Y}|$ pelo mesmo motivo.
- $I(X; Y)$ é uma função contínua de $p(x)$.
- $I(X; Y)$ é uma função concava de $p(x)$.

- $C \geq 0$ uma vez que $I(X; Y) \geq 0$.
- $C \leq \log |\mathcal{X}|$ uma vez que
 $C = \max I(X; Y) \leq \max H(X) \leq \log |\mathcal{X}|$.
- $C \leq \log |\mathcal{Y}|$ pelo mesmo motivo.
- $I(X; Y)$ é uma função contínua de $p(x)$.
- $I(X; Y)$ é uma função concava de $p(x)$.



Canal binário sem erros

A capacidade do canal é:

0 \longrightarrow 0

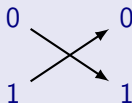
1 \longrightarrow 1

$$\begin{aligned} C &= \max_{p(x)} I(X; Y) \\ &= \max_{p(x)} \left(H(Y) - \underbrace{H(Y|X)}_{=0} \right) \\ &= 1 \text{ bit.} \end{aligned}$$

O máximo é atingido com $p(x) = \frac{1}{2}$.

Canal binário inversor

A capacidade do canal é:



$$\begin{aligned}
 C &= \max_{p(x)} I(X; Y) \\
 &= \max_{p(x)} \left(H(Y) - \underbrace{H(Y|X)}_{=0} \right) \\
 &= 1 \text{ bit.}
 \end{aligned}$$

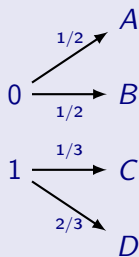
O máximo é atingido com $p(x) = \frac{1}{2}$.

- Este exemplo é análogo ao canal binário sem erros.
- O símbolo transmitido pode ser obtido invertendo o símbolo recebido.

Canal com saídas não sobrepostas

A capacidade do canal é:

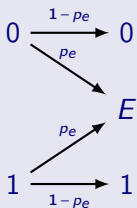
$$C = \max_{p(x)} I(X; Y) = 1 \text{ bit.}$$



O máximo é atingido com distribuição uniforme $p(x) = \frac{1}{2}$.

Na realidade este problema é semelhante ao canal binário sem erros, uma vez que observando um dos símbolos $\{A, B, C, D\}$ é possível saber inequivocamente qual o símbolo emitido.

Canal binário com perdas

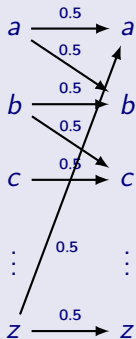


A capacidade do canal é:

$$C = \max_{p(x)} I(X; Y) = 1 - p_e$$

O máximo é atingido com $p(x) = \frac{1}{2}$.

Máquina de escrever ruidosa



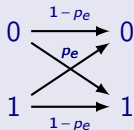
A capacidade do canal é:

$$C = \max_{p(x)} I(X; Y) = \log 13$$

O máximo é atingido com distribuição uniforme
 $p(x) = \frac{1}{26}$.

Canal binário simétrico

A capacidade do canal é:



$$C = \max_{p(x)} I(X; Y) = 1 - H(p_e)$$

onde $H(p_e) \triangleq -p_e \log p_e - (1 - p_e) \log(1 - p_e)$.
O máximo é atingido com $p(x) = \frac{1}{2}$.

Canal simétrico

É um canal em que todas as linhas e colunas são permutações umas das outras, por exemplo

$$\mathbf{P} = \begin{bmatrix} p_1 & p_2 & p_3 \\ p_2 & p_3 & p_1 \\ p_3 & p_1 & p_2 \end{bmatrix}.$$

A capacidade neste caso é

$$C = \log |\mathcal{Y}| - H(\text{"uma linha da matriz"}).$$

O máximo é atingido com distribuição uniforme.

Canal simétrico

É um canal em que todas as linhas e colunas são permutações umas das outras, por exemplo

$$\mathbf{P} = \begin{bmatrix} p_1 & p_2 & p_3 \\ p_2 & p_3 & p_1 \\ p_3 & p_1 & p_2 \end{bmatrix}.$$

A capacidade neste caso é

$$C = \log |\mathcal{Y}| - H(\text{"uma linha da matriz"}).$$

O máximo é atingido com distribuição uniforme.



O canal binário simétrico é um caso particular deste.

Canal fracamente simétrico

É um canal em que todas as linhas são permutações umas das outras e as colunas têm a mesma soma

$$\sum_x p(y|x).$$

A capacidade é também

$$C = \log |\mathcal{Y}| - H(\text{“uma linha da matriz”}).$$

O máximo é atingido com distribuição uniforme.



Atenção:

Não existe, em geral, uma fórmula explícita para a determinação da capacidade C , sendo necessário usar um método iterativo para estimar a capacidade.

Atenção:

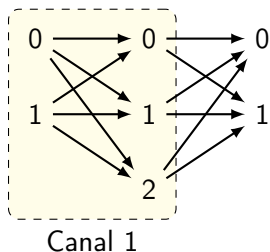
Não existe, em geral, uma fórmula explícita para a determinação da capacidade C , sendo necessário usar um método iterativo para estimar a capacidade.

Algoritmos iterativos “clássicos” para determinar a capacidade:

-  S. Arimoto, “*An algorithm for computing the capacity of arbitrary discrete memoryless channels*”, IEEE Transactions on Information Theory, vol. 18 (1) pp. 14–20, 1972.
-  R. Blahut, “*Computation of channel capacity and rate-distortion functions*”, IEEE Transactions on Information Theory, vol. 18 (4) pp. 460–473, 1972.

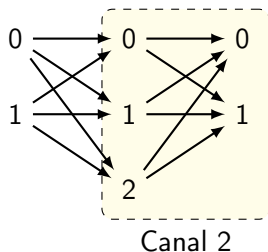
Concatenação de vários canais

O que acontece quando se concatenam vários canais?
Por exemplo:



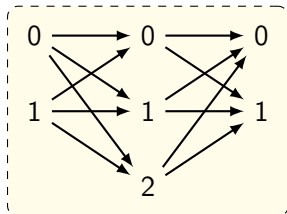
Concatenação de vários canais

O que acontece quando se concatenam vários canais?
Por exemplo:



Concatenação de vários canais

O que acontece quando se concatenam vários canais?
Por exemplo:

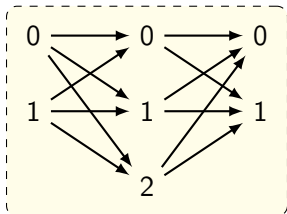


Concatenação dos dois canais

Concatenação de vários canais

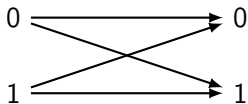
O que acontece quando se concatenam vários canais?

Por exemplo:



Concatenação dos dois canais

A concatenação destes dois canais é equivalente a um canal binário



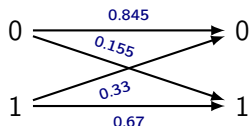
É necessário calcular as probabilidades de transição do canal equivalente.

Se as matrizes de transição são

$$\mathbf{P}_1 = \begin{bmatrix} 0.9 & 0.1 \\ 0.05 & 0.7 \\ 0.05 & 0.2 \end{bmatrix} \quad \text{e} \quad \mathbf{P}_2 = \begin{bmatrix} 0.9 & 0.2 & 0.5 \\ 0.1 & 0.8 & 0.5 \end{bmatrix}$$

então as probabilidades de transição do canal equivalente são

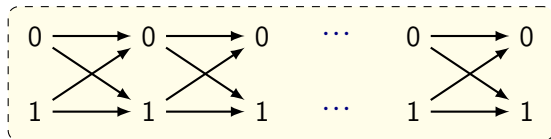
$$\mathbf{P}_{\text{eq}} = \mathbf{P}_2 \mathbf{P}_1 = \begin{bmatrix} 0.9 & 0.2 & 0.5 \\ 0.1 & 0.8 & 0.5 \end{bmatrix} \begin{bmatrix} 0.9 & 0.1 \\ 0.05 & 0.7 \\ 0.05 & 0.2 \end{bmatrix} = \begin{bmatrix} 0.845 & 0.33 \\ 0.155 & 0.67 \end{bmatrix}$$



Neste exemplo, o canal equivalente é binário, mas não é simétrico. A capacidade tem de ser calculada iterativamente.

Concatenação de canais idênticos

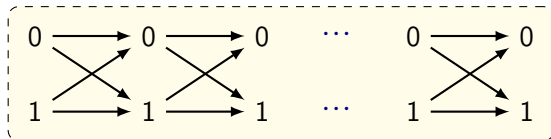
O que acontece quando se concatenam vários canais idênticos, cada um com matriz de transição P ?



Concatenação de n canais idênticos

Concatenação de canais idênticos

O que acontece quando se concatenam vários canais idênticos, cada um com matriz de transição \mathbf{P} ?

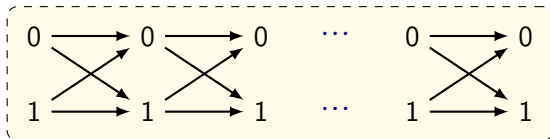


Concatenação de n canais idênticos

O canal equivalente é um canal binário com matriz de transição

$$\mathbf{P}_{\text{eq}} = \mathbf{P}^n$$

O que acontece quando se concatenam vários canais idênticos, cada um com matriz de transição \mathbf{P} ?



Concatenação de n canais idênticos

O canal equivalente é um canal binário com matriz de transição

$$\mathbf{P}_{\text{eq}} = \mathbf{P}^n$$



A concatenação de canais forma uma cadeia de Markov!

- A transmissão de informação por um canal ruidoso está geralmente sujeita a erros.

- A transmissão de informação por um canal ruidoso está geralmente sujeita a erros.
- É possível construir códigos que permitam a detecção e eventual correcção de erros. Esses códigos requerem a transmissão de informação adicional.

- A transmissão de informação por um canal ruidoso está geralmente sujeita a erros.
- É possível construir códigos que permitam a detecção e eventual correcção de erros. Esses códigos requerem a transmissão de informação adicional.
- A redundância adicionada no código permite baixar a probabilidade de erro, mas faz baixar o ritmo de transmissão pois é necessário transmitir mais símbolos pelo canal por cada símbolo da fonte.

- A transmissão de informação por um canal ruidoso está geralmente sujeita a erros.
- É possível construir códigos que permitam a detecção e eventual correcção de erros. Esses códigos requerem a transmissão de informação adicional.
- A redundância adicionada no código permite baixar a probabilidade de erro, mas faz baixar o ritmo de transmissão pois é necessário transmitir mais símbolos pelo canal por cada símbolo da fonte.
- Shannon mostrou em 1948 que é possível construir códigos com probabilidade de erro arbitrariamente baixa apenas se o ritmo de transmissão estiver abaixo da capacidade do canal.

- A transmissão de informação por um canal ruidoso está geralmente sujeita a erros.
- É possível construir códigos que permitam a detecção e eventual correcção de erros. Esses códigos requerem a transmissão de informação adicional.
- A redundância adicionada no código permite baixar a probabilidade de erro, mas faz baixar o ritmo de transmissão pois é necessário transmitir mais símbolos pelo canal por cada símbolo da fonte.
- Shannon mostrou em 1948 que é possível construir códigos com probabilidade de erro arbitrariamente baixa apenas se o ritmo de transmissão estiver abaixo da capacidade do canal.
- Este resultado é conhecido como *Teorema da codificação de canal*. Shannon mostrou que era possível a construção destes códigos mas não mostrou como se poderiam construir.

Definição (Código para o canal)

Um código para o canal $(\mathcal{X}, p(y|x), \mathcal{Y})$ consiste no seguinte:

Definição (Código para o canal)

Um código para o canal $(\mathcal{X}, p(y|x), \mathcal{Y})$ consiste no seguinte:

- 1 Um conjunto de índices $\{1, 2, \dots, M\}$. Os índices correspondem às M possíveis mensagens que se podem transmitir.

Definição (Código para o canal)

Um código para o canal $(\mathcal{X}, p(y|x), \mathcal{Y})$ consiste no seguinte:

- 1 Um conjunto de índices $\{1, 2, \dots, M\}$. Os índices correspondem às M possíveis mensagens que se podem transmitir.
- 2 Uma função de codificação $X^n: \{1, 2, \dots, M\} \rightarrow \mathcal{X}^n$ que fornece as palavras de código $X^n(1), X^n(2), \dots, X^n(M)$.
O índice (mensagem) a transmitir é codificado como uma sequência de n símbolos do alfabeto \mathcal{X} .

Definição (Código para o canal)

Um código para o canal $(\mathcal{X}, p(y|x), \mathcal{Y})$ consiste no seguinte:

- 1 Um conjunto de índices $\{1, 2, \dots, M\}$. Os índices correspondem às M possíveis mensagens que se podem transmitir.
- 2 Uma função de codificação $X^n: \{1, 2, \dots, M\} \rightarrow \mathcal{X}^n$ que fornece as palavras de código $X^n(1), X^n(2), \dots, X^n(M)$.
O índice (mensagem) a transmitir é codificado como uma sequência de n símbolos do alfabeto \mathcal{X} .
- 3 Uma função de decodificação $g: \mathcal{Y}^n \rightarrow \{1, 2, \dots, M\}$, que é uma regra determinística que “adivinha” o índice a partir da sequência de símbolos recebida \mathcal{Y}^n .
A sequência \mathcal{Y}^n formada por n símbolos do alfabeto \mathcal{Y} é decodificada pela função g que devolve um índice de $\{1, \dots, M\}$ correspondente à mensagem que se julga ter sido enviada.

Exemplo (Código para o canal)



Codif.

Canal

Descod.



Compra	Índice	$X^n(\cdot)$
pão	1	000
ovos	2	010
carne	3	101
cerveja	4	111

\mathcal{Y}^n	$g(\mathcal{Y}^n)$	Compra
000	1	pão
001	1	pão
010	2	ovos
011	2	ovos
100	3	carne
101	3	carne
110	4	cerveja
111	4	cerveja

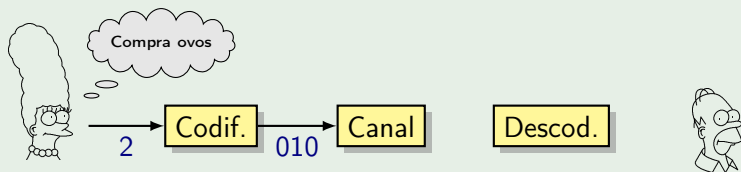
Exemplo (Código para o canal)



Compra	Índice	$X^n(\cdot)$
pão	1	000
ovos	2	010
carne	3	101
cerveja	4	111

\mathcal{Y}^n	$g(\mathcal{Y}^n)$	Compra
000	1	pão
001	1	pão
010	2	ovos
011	2	ovos
100	3	carne
101	3	carne
110	4	cerveja
111	4	cerveja

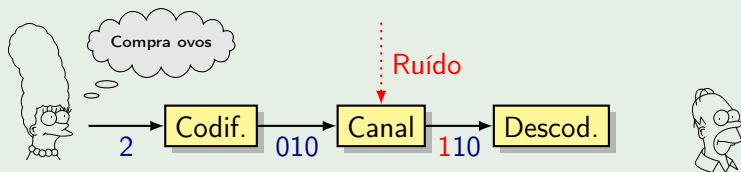
Exemplo (Código para o canal)



Compra	Índice	$X^n(\cdot)$
pão	1	000
ovos	2	010
carne	3	101
cerveja	4	111

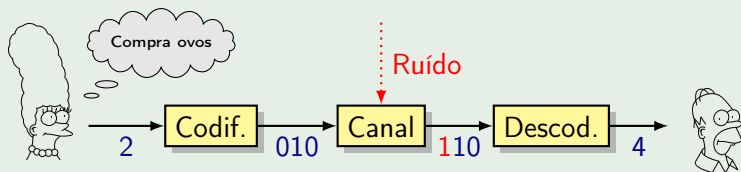
\mathcal{Y}^n	$g(\mathcal{Y}^n)$	Compra
000	1	pão
001	1	pão
010	2	ovos
011	2	ovos
100	3	carne
101	3	carne
110	4	cerveja
111	4	cerveja

Exemplo (Código para o canal)



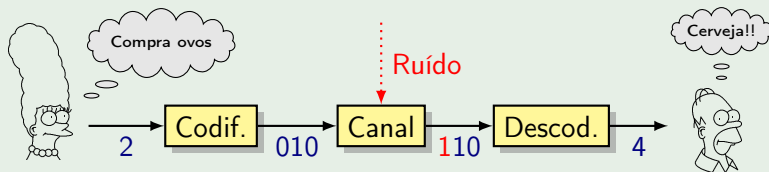
Compra	Índice	$X^n(\cdot)$	\mathcal{Y}^n	$g(\mathcal{Y}^n)$	Compra
pão	1	000	000	1	pão
ovos	2	010	001	1	pão
carne	3	101	010	2	ovos
cerveja	4	111	011	2	ovos
			100	3	carne
			101	3	carne
			110	4	cerveja
			111	4	cerveja

Exemplo (Código para o canal)



Compra	Índice	$X^n(\cdot)$	\mathcal{Y}^n	$g(\mathcal{Y}^n)$	Compra
pão	1	000	000	1	pão
ovos	2	010	001	1	pão
carne	3	101	010	2	ovos
cerveja	4	111	011	2	ovos
			100	3	carne
			101	3	carne
			110	4	cerveja
			111	4	cerveja

Exemplo (Código para o canal)



Compra	Índice	$X^n(\cdot)$	\mathcal{Y}^n	$g(\mathcal{Y}^n)$	Compra
pão	1	000	000	1	pão
ovos	2	010	001	1	pão
carne	3	101	010	2	ovos
cerveja	4	111	011	2	ovos
			100	3	carne
			101	3	carne
			110	4	cerveja
			111	4	cerveja

Definição (Probabilidade de erro)

A probabilidade condicional de erro dado que foi transmitido o símbolo i é

$$\lambda_i \triangleq \Pr \{g(Y^n) \neq i \mid X^n = X^n(i)\}$$

Definição (Probabilidade de erro)

A probabilidade condicional de erro dado que foi transmitido o símbolo i é

$$\lambda_i \triangleq \Pr \{g(Y^n) \neq i \mid X^n = X^n(i)\}$$

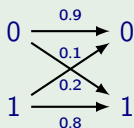
Definição (Probabilidade de erro máxima)

A probabilidade de erro máxima para um código de comprimento n é dada por

$$\lambda^{(n)} \triangleq \max_{i \in \{1, \dots, M\}} \lambda_i.$$

Exemplo

Considere o código e o canal binário (não simétrico) seguinte:

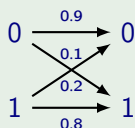


Msg.	$X^n(\cdot)$	\mathcal{Y}^n	$g(\mathcal{Y}^n)$
1	000	2 ou mais zeros	1
2	111	2 ou mais uns	2

Quais as probabilidades de erro quando são transmitidas as mensagens 1 e 2?

Exemplo

Considere o código e o canal binário (não simétrico) seguinte:



Msg.	$X^n(\cdot)$	\mathcal{Y}^n	$g(\mathcal{Y}^n)$
1	000	2 ou mais zeros	1
2	111	2 ou mais uns	2

Quais as probabilidades de erro quando são transmitidas as mensagens 1 e 2?

$$\lambda_1 = \Pr\{g(Y^n) \neq 1 \mid X^n = X^n(1) = 000\} = 0.1^3 + 3 \times 0.1^2 \times 0.9 = 0.028$$

$$\lambda_2 = \Pr\{g(Y^n) \neq 2 \mid X^n = X^n(1) = 111\} = 0.2^3 + 3 \times 0.2^2 \times 0.8 = 0.104$$

A probabilidade de erro máxima é $\max_i \lambda_i = 0.104$.

Definição (Ritmo de um código)

O ritmo de um código de comprimento n para M índices é definido por

$$R = \frac{\log M}{n}$$

Definição (Ritmo de um código)

O ritmo de um código de comprimento n para M índices é definido por

$$R = \frac{\log M}{n}$$

Exemplo

Compra	Índice	$X^n(\cdot)$
pão	1	000
ovos	2	010
carne	3	101
cerveja	4	111

Este código tem ritmo

$$R = \frac{\log 4}{3} = \frac{2}{3} \approx 0.667.$$

Definição (Ritmo atingível)

Um ritmo R diz-se **atingível** se for possível construir uma sequência de códigos progressivamente maiores tal que a probabilidade de erro máxima $\lambda^{(n)}$ tende para 0 quando $n \rightarrow \infty$.

Definição (Ritmo atingível)

Um ritmo R diz-se **atingível** se for possível construir uma sequência de códigos progressivamente maiores tal que a probabilidade de erro máxima $\lambda^{(n)}$ tende para 0 quando $n \rightarrow \infty$.

Teorema (Codificação de canal, Shannon 1948)

Todos os ritmos abaixo da capacidade do canal são atingíveis. Isto é, para todos os ritmos $R < C$, é possível construir uma sequência de códigos progressivamente maiores tais que a probabilidade de erro máxima $\lambda^{(n)} \rightarrow 0$.

Exemplo

Suponhamos que um canal binário simétrico tem capacidade $C = 0.5$ bits. Pretende-se construir um código que permita obter uma probabilidade de erro arbitrariamente baixa. Qual o tamanho mínimo que as palavras de código têm de ter, em média?

Probabilidade de erro arbitrariamente baixa implica $R < C$.

Portanto

$$R = \frac{\log M}{n} < 0.5$$

ou seja,

$$n > \frac{\log M}{0.5} = 2 \log M$$

Considerando $M = 2$, conclui-se que são necessários mais de 2 bits por cada bit a transmitir.

(Note que este resultado não nos diz como construir o código...)