Razão de distorção

- Consideremos uma fonte sem memória e alfabeto $X = \{x_i : i = 1, \ldots, q\}$, com respetivas probabilidades $\{P(x_i) : i = 1, \ldots, q\}$.
- Na sua transmissão por um canal, os símbolos x_i fonte serão em geral transformados em novos símbolos y_i de um alfabeto $Y = \{y_j : j = 1, \ldots, r\}$.

- Em geral a recuperação da mensagem inicial a partir do resultado da transmissão da sua versão codificada não será total por
  - na própria codificação das mensagens, a representação dos símbolos originais poderá não ser exacta, pelo que a recuperação da mensagem original pode ser sujeita a erros;
  - a redundância introduzida na codificação para a transmissão da mensagem pelo canal poderá ser insuficiente de forma que a informação transmitida exceda a capacidade do canal.
- A teoria da razão de distorção, introduzida por C. E. Shannon em 1959, visa a determinação da informação mútua mínima que o canal deverá possuir, para uma distribuição de probabilidades dada para a fonte à entrada, de forma que a distorção média não exceda uma dada tolerância D.

- Suponhamos que cada uma de M possíveis mensagens fonte é codificada por uma palavra de comprimento n. Seja $H(M)$ a entropia da mensagem fonte. A *razão de codificação* R é definida por

$$R = \frac{H(M)}{n}.$$

- Note-se que, no caso de todas as mensagens terem a mesma probabilidade, $H(M) = \log M$.
- No caso geral, $H(M)$ é o número médio de bits transmitido para cada mensagem fonte.

Medidas de distorção

- A *medida de distorção de uma letra* x_i, $d(x_i, y_j)$, é a medida do custo da representação do símbolo fonte x_i pelo símbolo y_j.
- A *distorção média*, $\overline{d}$ é definida como sendo a média pesada dos $d(x_i, y_j)$:

$$\overline{d} = \sum_{i=1}^{q} \sum_{j=1}^{r} P(x_i)P(y_j|x_i)d(x_i, y_j)$$

  onde os $P(y_j|x_i)$ são as probabilidades de transição do canal.
- Fixada a distribuição de probabilidades para a entrada, dizemos que as probabilidades de transição do canal são *D-admissíveis* se a correspondente distorção média satisfizer a desigualdade $\overline{d} \leqslant D$.

- O conjunto das probabilidades de transição D-admissíveis representa-se por $\mathcal{P}_D$.
- A cada elemento $\mathbf{P}$ de $\mathcal{P}_D$ corresponde a informação mútua do canal, dada por

$$I(X;Y) = \sum_{i=1}^{q} \sum_{j=1}^{r} P(x_i) P(y_j|x_i) \log \frac{P(y_j|x_i)}{P(y_j)},$$

onde $P(y_j) = \sum_{i=1}^{q} P(y_j|x_i) P(x_i)$, como habitualmente.

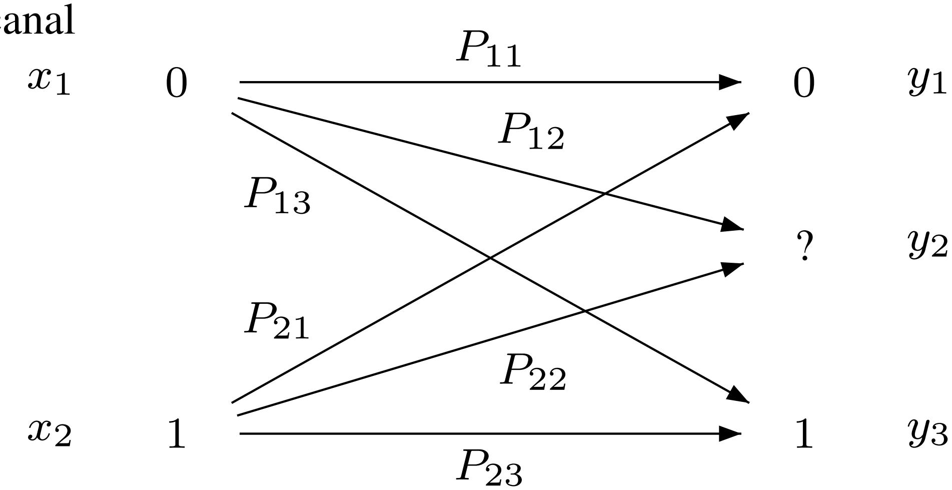
- Dada uma distribuição de probabilidades $\{P(x_i) : i = 1, \ldots, q\}$ para os símbolos da fonte, a *função razão de distorção* associa a cada D o seguinte valor:

$$R(D) = \min_{[P(y_j|x_i)]_{i,j} \in \mathcal{P}_D} I(X;Y),$$

  ou seja, a informação mútua mínima que o canal poderá proporcionar não excedendo a tolerância D para a distorção média.
- Trata-se de um problema de otimização de uma função contínua não linear definida no compacto $\mathcal{P}_D$. O mínimo existe, mas é difícil de calcular.

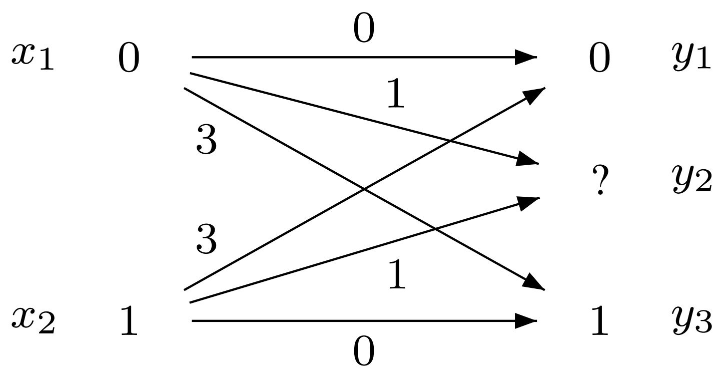
Exemplos

- Consideremos o canal



Note-se que se trata de uma generalização dos BSC e dos BEC.

- Tomemos a medida de distorção para o canal dada por:
  - $d(x_1, y_1) = d(x_2, y_3) = 0$, indicando que nesses casos a transmissão é fiel e não há erro na representação dos símbolos à partida como resultado da utilização do canal;
  - $d(x_1, y_2) = d(x_2, y_2) = 1$, ou seja atribuímos custo 1 à perda do símbolo transmitido;
  - $d(x_1, y_3) = d(x_2, y_1) = 3$, o que significa que atribuímos custo 3 à inversão de bit.



Supondo que $P(x_1) = P(x_2) = 0.5$ e as probabilidades $P(y_j|x_i)$ do canal são dadas pela seguinte matriz

$$\mathbf{P} = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.1 & 0.2 & 0.7 \end{bmatrix},$$

podemos calcular $P(y_1) = 0.4$, $P(y_2) = 0.2$, $P(y_3) = 0.4$,

$$\begin{aligned} I(X;Y) &= \frac{1}{2}(0.7\log(0.7/0.4) + 0.1\log(0.1/0.4) + 0.2\log(0.2/0.2) \\ &\quad + 0.2\log(0.2/0.2) + 0.1\log(0.1/0.4) + 0.7\log(0.7/0.4)) \\ &\simeq 0.365, \end{aligned}$$

e a distorção média

$$\begin{aligned}\overline{d} &= \sum_{i=1}^{2}\sum_{j=1}^{3} P(x_i)P(y_j|x_i)\, d(x_i, y_j) \\ &= 2 \times \frac{1}{2}[(0.7)(0) + (0.2)(1) + (0.1)(3)] \\ &= 0.5\end{aligned}$$

- Assim, para $D = 0.5$, a matriz do canal dada por $\mathbf{P}$ é D-admissível e tem informação mútua $I(X;Y) \simeq 0.365$.
- Haverá alguma matriz do canal com $\overline{d} \leqslant 0.5$ e $I(X;Y) < 0.365$?
- Intuitivamente, é de esperar que o mínimo da informação mútua seja atingido quando se tomar a distorção máxima admissível $\overline{d} = 0.5$.

- Consideremos as duas matrizes

$$P_{BSC} = \begin{bmatrix} \frac{5}{6} & 0 & \frac{1}{6} \\ \frac{1}{6} & 0 & \frac{5}{6} \end{bmatrix} \qquad P_{BEC} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix},$$

 correspondendo respetivamente a um BSC e a um BEC.

- Fazendo as contas, obtém-se em ambos os casos distorção média $\overline{d} = 0.5$ enquanto $I_{BSC}(X;Y) \simeq 0.350$ e $I_{BEC}(X;Y) \simeq 0.5$.

Propriedades de $R(D)$

- Passamos a estudar o comportamento da função $D \mapsto R(D)$.

  Resultado 2.9 *A função* $R(D)$ *é uma função decrescente de* D *com valores no intervalo* $[0, H(X)]$. *O máximo é atingido para o valor mínimo possível da distorção,* $D_{\min}$*, e existe um valor mínimo* $D_{\max}$ *tal que* $R(D) = 0$ *para* $D \geqslant D_{\max}$.

- Note-se que se $D_{\min}$ é o mínimo para a distorção média $\overline{d}$, então $\mathcal{P}_D = \emptyset$ para $D < D_{\min}$.
- Note-se também que a função $R(D)$ é obviamente decrescente pois o conjunto de matrizes sobre o qual se toma o mínimo de $I(X;Y)$ aumenta com D.

- Vejamos então qual é o valor $D_{\min}$ para

$$\overline{d} = \sum_{i=1}^{q} \sum_{j=1}^{r} P(x_i) P(y_j|x_i) d(x_i, y_j).$$

  Recorde-se que os $d(x_i, y_j)$ são medidas de distorção fixadas e as probabilidades $P(x_i)$ são dadas. A variável em relação à qual se pretende minimizar a distorção média $\overline{d}$ é a matriz $\mathbf{P} = [P(y_j|x_i)]_{i,j}$ do canal.

- Uma vez que todos os coeficientes são não negativos, para um dado x_i a melhor estratégia para minimizar $\overline{d}$ consiste em concentrar a probabilidade $P(y_j|x_i) = 1$ num y_j para o qual $d(x_i, y_j)$ for mínimo.

- Fazendo $d(x_i) = \min_j d(x_i, y_j) = d(x_i, y_{j(x_i)})$, temos então

$$D_{\min} = \sum_{i=1}^{q} P(x_i) d(x_i).$$

- Calculando $I(X;Y)$, obtemos

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &= H(Y) - \sum_i \sum_j P(x_i)P(y_j|x_i)\log P(y_j|x_i) \\ &= H(Y) \end{aligned}$$

o que mostra que $R(D_{\min}) = H(Y) = I(X;Y) \leqslant H(X)$.

- Para mostrar que $R(D_{\max}) = 0$, começamos por observar que, sendo $R(D)$ decrescente, se $R(D) = 0$ for atingido para um dado D, também será $R(D') = 0$ para $D' \geqslant D$.
- Ora mostrámos que $I(X;Y) = 0$ implica que a entrada e a saída são independentes, i.e., $P(y_j) = P(y_j|x_i)$ para todos os i e j, ou ainda cada coluna da matriz $\mathbf{P}$ tem as entradas todas iguais.
- Assim, mediante a condição $I(X;Y) = 0$, obtemos

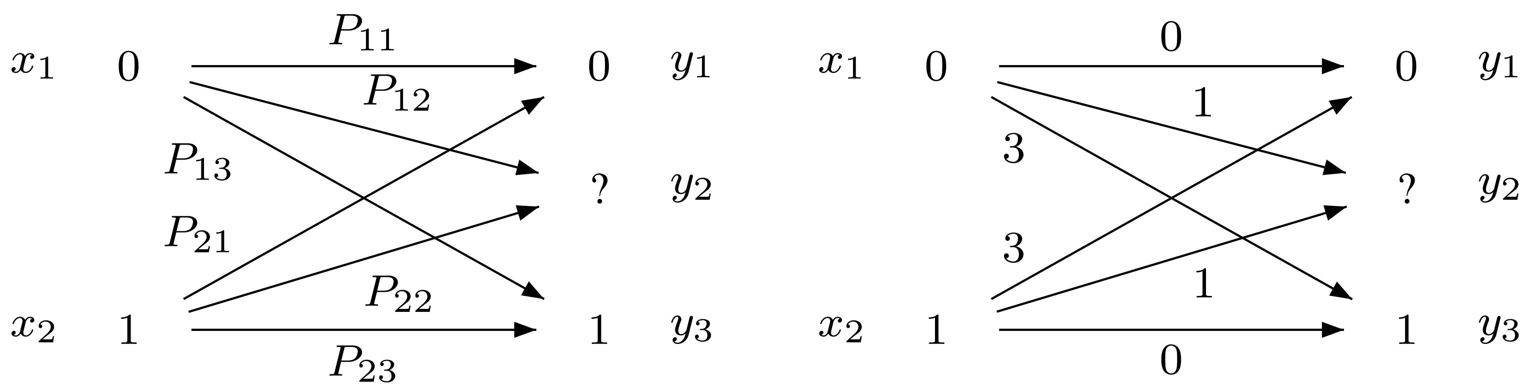
$$\overline{d} = \sum_j P(y_j) \sum_i P(x_i)\, d(x_i, y_j). \tag{6}$$

- Procuramos portanto o menor $\overline{d}$ desta forma, ou seja pretendemos minimizar (6), sendo aqui a variável a distribuição $P(y_j)$ (a qual, mediante as restrições acima, determina a matriz do canal).
- Seja j^* o índice para o qual $\sum_i P(x_i)\, d(x_i, y_j)$ é mínimo. Concentrando em y_{j^*} a probabilidade, i.e., tomando $P(y_{j^*}) = 1$, obtemos o valor mínimo para $\overline{d}$. Concluímos que $D_{\max}$ é dado pela fórmula:

$$D_{\max} = \min_{j=1,\ldots,r} \sum_{i=1}^{q} P(x_i)\, d(x_i, y_j).$$

Exemplo

- Recorde-se o exemplo anterior:



- Consideremos o canal deste tipo com matriz de probabilidades

$$\mathbf{P} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}.$$

- Como $P(y_j|x_i) = P(y_j)$, $I(X;Y) = 0$.
- A distorção correspondente é

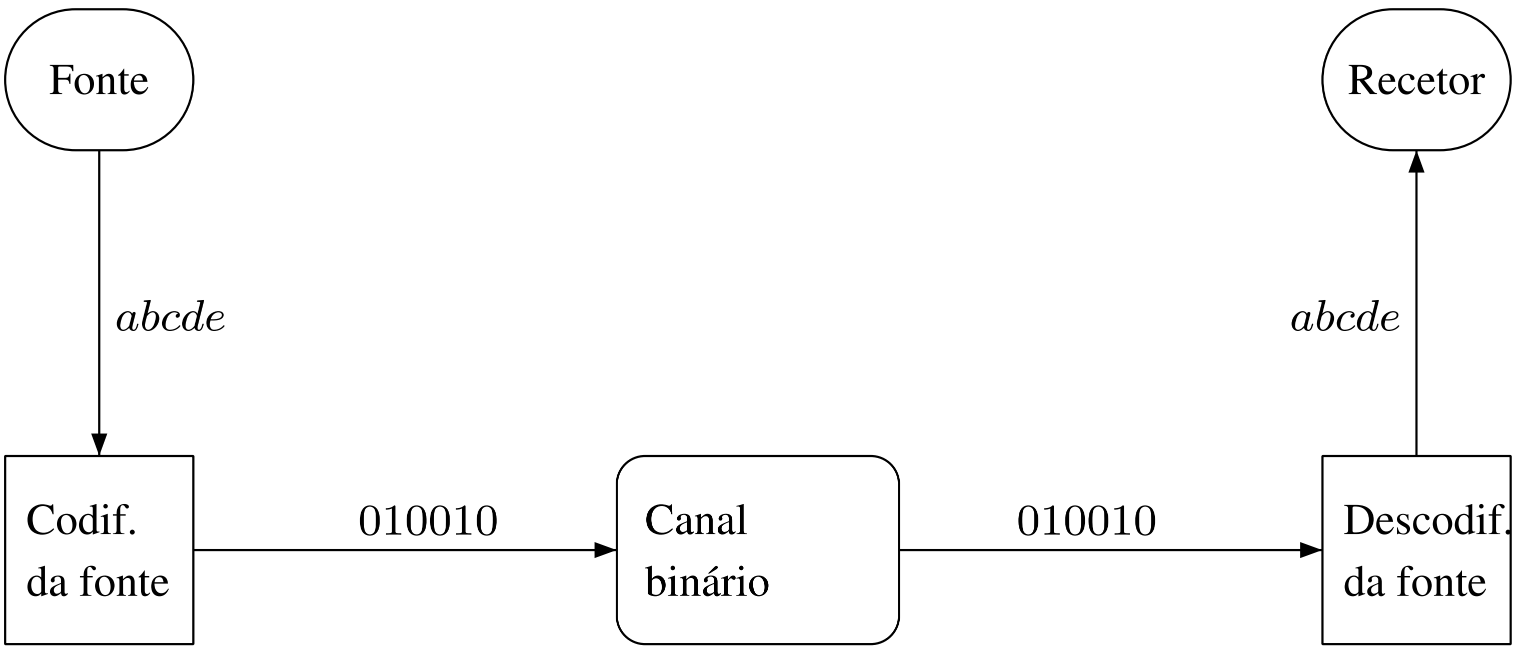
$$\overline{d} = \frac{1}{2}\,\frac{1}{3}\,(0 + 1 + 3 + 3 + 1 + 0) = \frac{4}{3}.$$

- No entanto o valor de $D_{\max}$ poderá ser inferior, pois não começámos por minimizar $\overline{d}$ sujeito à condição $I(X;Y)=0$.
- De facto,

$$D_{\max} = \frac{1}{2}\min_j(d(x_1,y_j)+d(x_2,y_j)) = \frac{1}{2}\min\{0+3, 1+1, 3+0\} = 1.$$

  Logo $D_{\max}=1$, sendo $I(X;Y)=0$ com $\overline{d}=1$ obtido com $\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}$.
- Concluímos que, se admitirmos uma distorção de 1, então é possível obter informação mútua nula, a qual é atingida tomando o canal que produz à saída sempre o símbolo $y_2 = ?$.

3 Codificação da fonte



Sistemas de comunicação digital

- Em sistemas de comunicação digital, um problema fundamental consiste no armazenamento e transmissão eficientes da informação.
- Para começar, precisamos de representar a informação na forma adequada ao seu tratamento digital, ou seja em código binário.
- Assim, pretende-se obter sistemas de codificação que satisfaçam as seguintes condições:
  - a partir da mensagem codificada, deve ser possível recuperar sem ambiguidade a mensagem original;
  - uma mensagem dada deve poder ser transformada eficientemente numa sua representação binária o mais curta possível e dela ser recuperada também eficientemente.

- Para a transmissão por um canal com ruído, devemos tomar cuidados adicionais, aumentando a redundância da versão codificada transmitida de forma a recuperar da perturbação da mensagem introduzida pelo canal.
- De momento vamos concentrarmo-nos na codificação da fonte, supondo que a mensagem codificada recebida pelo recetor reproduz fielmente a mensagem codificada emitida pela fonte.
- Considere-se um alfabeto $S = \{s_i : i = 1, \ldots, q\}$ da fonte e um alfabeto $X = \{x_j : j = 1, \ldots, r\}$ dito o *alfabeto do código*.
- Uma *palavra-código* é uma palavra do alfabeto do código.

- Uma *codificação* é uma correspondência $s_i \mapsto C(s_i)$ associando a cada letra de S uma palavra-código, i.e., uma sua representação como uma palavra sobre o alfabeto do código.
- Uma *tabela de codificação* ou simplesmente uma *codificação da fonte* é uma lista $\{s \mapsto C(s) : s \in S\}$ representando a codificação de cada símbolo da fonte.
- Dado um alfabeto A, representemos por A^* o conjunto de todas as palavras (finitas) nas letras de A.
- A *concatenação* define uma operação sobre A^*:

$$(a_1 a_2 \ldots a_m) \cdot (b_1 b_2 \ldots b_n) = a_1 a_2 \ldots a_m b_1 b_2 \ldots b_n.$$

  Note-se que esta operação é associativa e que A^* é um monoide, sendo a palavra vazia ε o elemento neutro.

- Naturalmente, as letras são vistas como palavras com um só símbolo.

Proposição 3.1 *O monoide A^* tem a seguinte* propriedade universal*: dado um monoide M e uma função $\varphi : A \to M$, existe uma única extensão de φ a um homomorfismo $\hat{\varphi} : A^* \to M$, i.e., tal que o diagrama seguinte comuta:*

$$\begin{tikzcd} A \arrow[r, hook] \arrow[dr, "\varphi"'] & A^* \arrow[d, "\hat{\varphi}"] \\ & M \end{tikzcd}$$

Prova. Basta tomar $\hat{\varphi}(a_1 a_2 \ldots a_m) = \varphi(a_1)\varphi(a_2)\ldots\varphi(a_m)$. $\square$

- Uma codificação não é mais do que uma função $C : S \to X^*$.

- Dizemos que a codificação C é *não-singular* se C for injetiva.
- Dizemos que a codificação C é *descodificável* (ou um *código*, segundo alguns autores) se a extensão $\hat{C}$ a S^* for injetiva.
- Uma codificação diz-se *binária, ternária, r-ária* se o alfabeto da codificação for respetivamente um alfabeto com 2, 3, r símbolos. Nos dois primeiros casos, é frequente tomar $X = \{0, 1\}$ e $X = \{0, 1, 2\}$, respetivamente.

Exemplo

Sejam $S = \{s_1, s_2, s_3, s_4\}$ e $X = \{0, 1\}$. Consideremos os seguintes esquemas de codificação:

Fonte	**A**	**B**	**C**
s_1	0	0	00
s_2	11	11	01
s_3	00	00	10
s_4	11	010	11

- **A** não é não-singular;
- **B** é não-singular mas não é descodificável: $\mathbf{B}(s_3) = 00 = \hat{\mathbf{B}}(s_1 s_1)$;
- **C** é descodificável.

Códigos instantâneos

- Um esquema de codificação C diz-se *instantâneo* se a descodificação de uma mensagem parcial pode ser feita logo que seja reconhecida uma palavra do código, independentemente do que se lhe siga.
- De forma equivalente, se $C(s_i)$ é *prefixo* (i.e., factor à esquerda no monoide X^*) de $C(s_j)$ então $s_i = s_j$.
  Uma codificação com esta propriedade diz-se também um *código prefixo*.
- Em particular, note-se que todo o código prefixo é descodificável.
- Dualmente, se C é tal que, sempre que $C(s_i)$ é *sufixo* (i.e., factor à direita no monoide X^*) de $C(s_j)$, tem-se $s_i = s_j$, então C diz-se um *código sufixo*. [Exercício: mostre que os “códigos sufixos” são códigos.]

Exemplo

- Considere-se os seguintes esquemas de codificação:

Fonte	**A**	**B**	**C**
s_1	0	0	0
s_2	10	01	01
s_3	110	011	011
s_4	1110	0111	111

- Os três esquemas são descodificáveis: sendo o primeiro um código prefixo, o segundo e o terceiro são códigos sufixos.
- Para provar que **C** é descodificável, basta notar que se tivermos uma mensagem em código $x_{i_1} x_{i_2} \ldots x_{i_m} \in \hat{\mathbf{C}}(S^*)$, as posições com 0 determinam a factorização em palavras-código: conforme a classe de congruência módulo 3 do número de 1's entre 0's consecutivos, teremos s_1, s_2 ou s_3 seguido de uma potência de s_4.
- Note-se também que, dado um prefixo $0111111\ldots$ de uma mensagem codificada, por muito longo que seja, não sabemos qual é a primeira letra da mensagem original até que apareça o próximo 0.