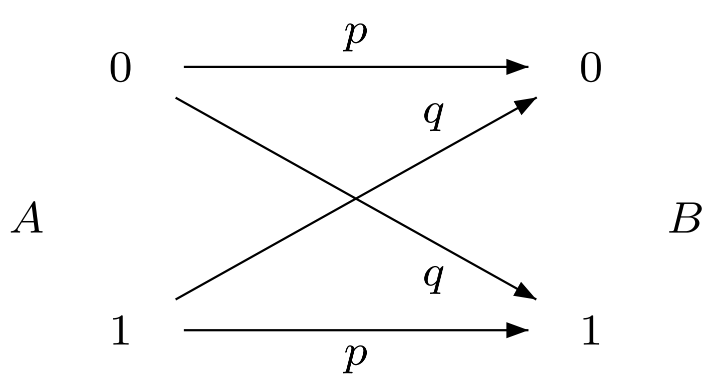
BSC

- Nos canais de transmissão de informação habituais, a informação é transmitida na forma de sequências de dígitos binários, isto é sequências de 0's e 1's, ou de *bits*.
- Num *canal simétrico binário (BSC)* a probabilidade não nula do bit recebido ser invertido é independente do bit, ou seja

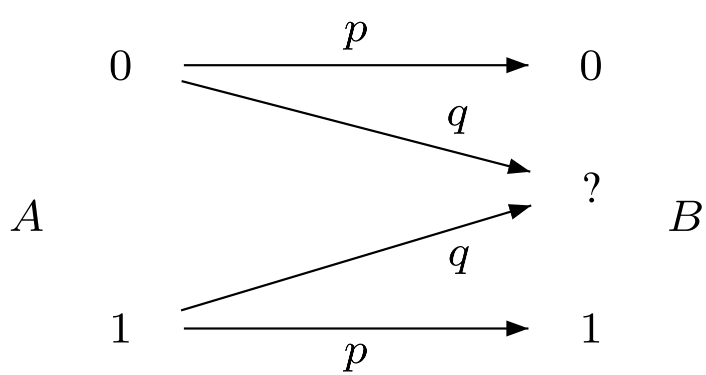
$$P(0|1) = P(1|0) = q,\ P(0|0) = P(1|1) = p = 1 - q.$$



BEC

- Num *canal binário com apagamento (BEC)* o ruído impede a emissão do bit recebido, emitindo a indicação de incapacidade de o identificar através de um novo símbolo ?, sendo

$$P(?|0) = P(?|1) = q,\ P(0|0) = P(1|1) = p = 1 - q.$$



Informação mútua

- O comportamento de um canal depende não só da matriz de probabilidades de transmissão como da informação que lhe é fornecida para transmitir.
- Devemos assim considerar uma distribuição de probabilidades $P(A) = \{P(a_1), \ldots, P(a_r)\}$ para as entradas.
- A resultante distribuição de probabilidades $P(B) = \{P(b_1), \ldots, P(b_s)\}$ para as saídas é dada por

$$P(b_j) = \sum_{i=1}^{r} P(b_j|a_i)P(a_i).$$

- A probabilidade $P(b_j|a_i)$ chama-se por vezes de *probabilidade para a frente*, sendo aqui o *sentido* aquele em que se concebe à partida o funcionamento do canal de informação.
- Podemos também definir a *probabilidade para trás* $P(a_i|b_j)$, que indica a probabilidade de ter sido emitido o símbolo a_i sabendo que foi recebido o símbolo b_j.
- Usando o Teorema de Bayes, obtemos

$$P(a_i|b_j) = \frac{P(a_i, b_j)}{P(b_j)} = \frac{P(b_j|a_i)P(a_i)}{P(b_j)} = \frac{P(b_j|a_i)P(a_i)}{\sum_{k=1}^{r} P(b_j|a_k)P(a_k)}$$

sendo $P(a_i, b_j)$ a probabilidade conjunta de ser emitido o símbolo a_i e recebido o símbolo b_j.

- $P(b_j)$ é a probabilidade *a priori* de ser emitido à saída o símbolo b_j, sem conhecer o símbolo que foi emitido à entrada;
- $P(b_j|a_i)$ é a probabilidade *a posteriori* de ser emitido à saída o símbolo b_j, sabendo que foi emitido à entrada o símbolo a_i;
- $P(a_i)$ é a probabilidade *a priori* de ser emitido à entrada o símbolo a_i, sem conhecer o símbolo que é emitido à saída;
- $P(a_i|b_j)$ é a probabilidade *a posteriori* de ser emitido à entrada o símbolo a_i, sabendo que é emitido à saída o símbolo b_j.

- A *entropia a priori de A* é

$$H(A) = -\sum_{a \in A} P(a) \log(P(a)),$$

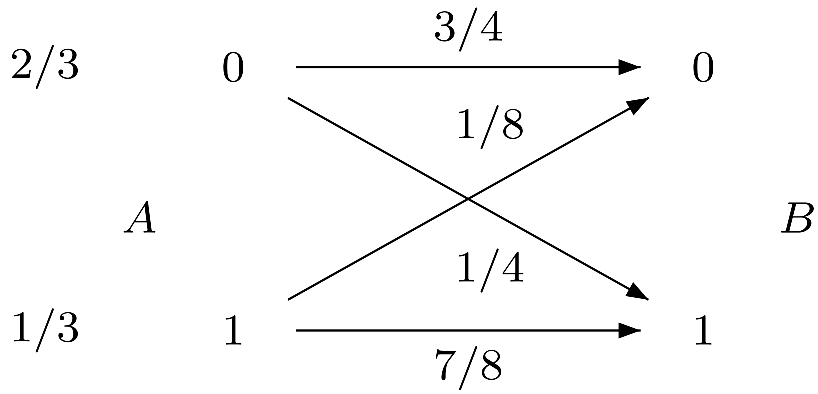
a média pesada da incerteza da entrada, sem conhecer a saída.

- A *entropia a posteriori de A, dado b_j* é

$$H(A|b_j) = -\sum_{a \in A} P(a|b_j) \log(P(a|b_j)),$$

a média pesada da incerteza da entrada, sabendo que à saída é emitido b_j.

Exemplo. Considere-se o canal binário descrito pelo diagrama



Fazendo as contas, obtemos

$$P(b=0) = \frac{13}{24},\ P(b=1) = \frac{11}{24}$$
$$P(a=0|b=0) = \frac{12}{13},\ P(a=1|b=0) = \frac{1}{13},\ P(a=1|b=1) = \frac{7}{11},\ P(a=0|b=1) = \frac{4}{11}$$
$$H(A) \simeq 0.918,\ H(A|0) \simeq 0.391,\ H(A|1) \simeq 0.946,$$
$$H(B) \simeq 0.995,\ H(B|0) \simeq 0.811,\ H(B|1) \simeq 0.544.$$

Note-se que a incerteza aumenta quando observamos a saída 1: apesar da probabilidade da entrada, sem conhecer a saída, aponte para a expetativa de obter 0, essa expetativa diminui graças às probabilidades de erros de transmisão.

Em contrapartida, a média pesada das entropias à entrada a posteriori é

$$\sum_{j=1,2} P(b_j) H(A|b_j) \simeq 0.645 < H(A).$$

Informação mútua

- O *equívoco* de A em relação a B é

$$H(A|B) = \sum_{j=1}^{s} P(b_j) H(A|b_j),$$

  e mede a incerteza acerca da entrada conhecida a saída.

- Assim, $I(A;B) = H(A) - H(A|B)$ mede a diminuição da incerteza de A resultante da utilização do canal e chama-se a *informação mútua*.

Alguns tipos de canais:

- *sem ruído*: $H(A|B) = 0$, pelo que $I(A;B) = H(A)$, ou seja a incerteza acerca da entrada não depende da utilização do canal;
- *com ruído*: $H(A|B) > 0$ mas $H(A|B) < H(A)$, caso em que a quantidade de informação, ou seja a incerteza quanto à entrada, diminui com a utilização do canal;
- *ambíguo*: $H(A|B) = H(A)$, ou seja, o ruído é tanto que a incerteza desaparece por completo, o resultado à saída deixando de depender do que foi emitido à entrada.

Note-se que

$$
\begin{aligned}
I(A;B) &= H(A) - H(A|B) \\
&= -\sum_{a\in A} P(a)\log(P(a)) + \sum_{a\in A}\sum_{b\in B} P(a|b)P(b)\log(P(a|b)) \\
&= -\sum_{a\in A}\sum_{b\in B} P(a,b)\log(P(a)) + \sum_{a\in A}\sum_{b\in B} P(a,b)\log(P(a|b)) \\
&= \sum_{a\in A}\sum_{b\in B} P(a,b)\log\frac{P(a|b)}{P(a)}
\end{aligned}
$$

Resultado 2.1 *Tem-se*

$$
\begin{aligned}
I(A;B) &= \sum_{a\in A}\sum_{b\in B} P(a,b)\log\frac{P(a,b)}{P(a)P(b)} \\
&= \sum_{a\in A}\sum_{b\in B} P(a)P(b|a)\log\frac{P(b|a)}{P(b)}. \;\square
\end{aligned}
$$

Propriedades

Recorde-se que, para duas distribuições de probabilidade sobre o mesmo conjunto, p_i e q_i, tem-se

$$\sum_{i=1}^{N} p_i \log \frac{q_i}{p_i} \leqslant 0,$$

com igualdade se e só se $p_i = q_i$ $(i = 1, \ldots, N)$.
Fazendo $p_i = P(a,b)$, $q_i = P(a)P(b)$, e $N = |A \times B|$, obtemos

$$-I(A;B) = \sum_{a \in A} \sum_{b \in B} P(a,b) \log \frac{P(a)P(b)}{P(a,b)} \leqslant 0 \tag{4}$$

Resultado 2.2 *Tem-se $I(A;B) \geqslant 0$, verificando-se a igualdade se e só se $P(a,b) = P(a)P(b)$ para quaisquer $a \in A$ e $b \in B$, i.e., a informação à entrada e a informação à saída são estatisticamente independentes.*$\Box$

- Por simetria da expressão (4), na entrada e na saída, temos $I(A;B) = I(B;A)$, ou seja

$$H(A) - H(A|B) = H(B) - H(B|A). \tag{5}$$

- $H(B|A)$ designa-se por *equívoco de B em relação a A* e mede a incerteza da saída dada a entrada.

  Resultado 2.3 *Tem-se $H(A|B) \leqslant H(A)$ e $H(B|A) \leqslant H(B)$, ou seja, informalmente, a incerteza diminui, em média, quando acrescentamos informação (no caso presente, sobre informação que potencialmente influencia aquela que estamos interessados em observar).* $\square$

- De (5) segue que $H(A) + H(B|A) = H(B) + H(A|B)$. O valor comum é

$$H(A,B) = -\sum_{a \in A} \sum_{b \in B} P(a,b) \log(P(a,b)).$$

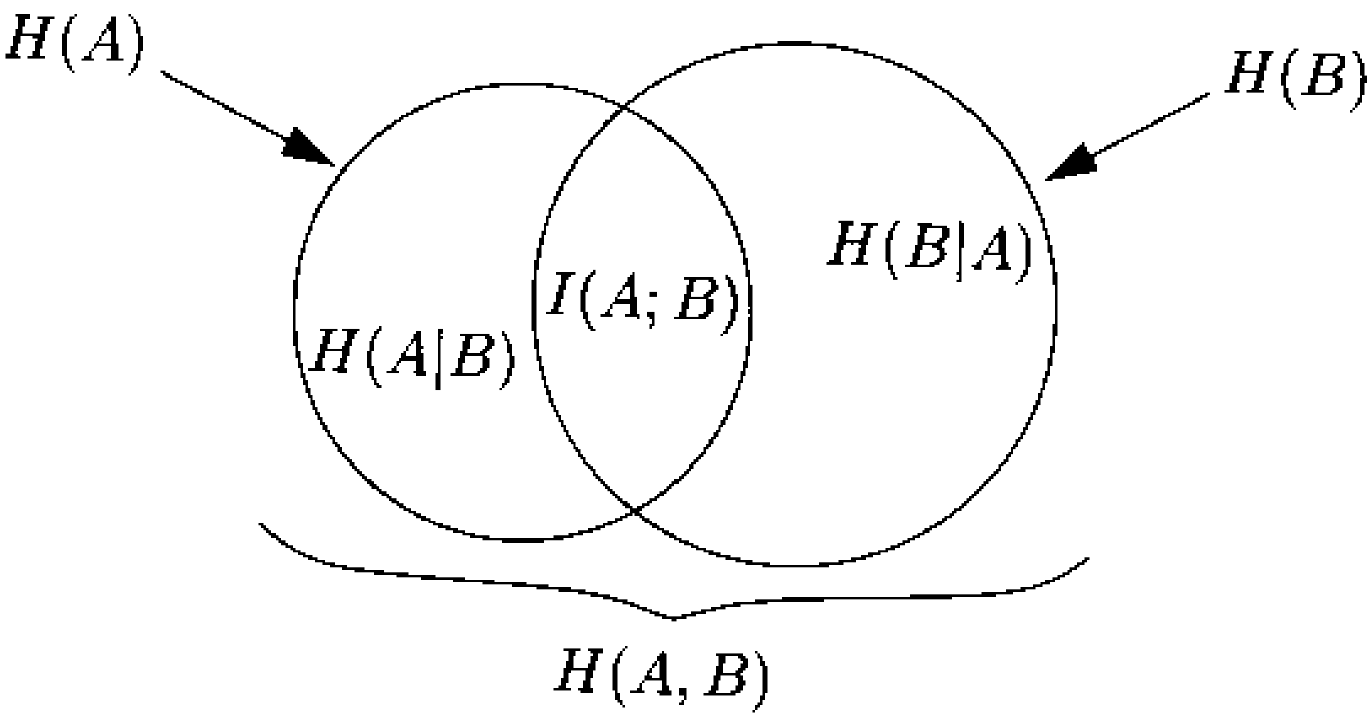
- Temos portanto

$$H(A,B) = H(A) + H(B|A) = H(B) + H(A|B)$$
$$= H(A) + H(B) - I(A;B).$$

- Em termos informais, $H(A,B) = H(A) + H(B) - I(A;B)$ significa que a incerteza conjunta de A e B é a soma das incertezas de A e B menos a informação fornecida pelo canal.
- Por sua vez, $H(A,B) = H(A) + H(B|A)$ significa que a incerteza conjunta de A e B é a incerteza de A acrescida da incerteza de B depois de conhecido A.

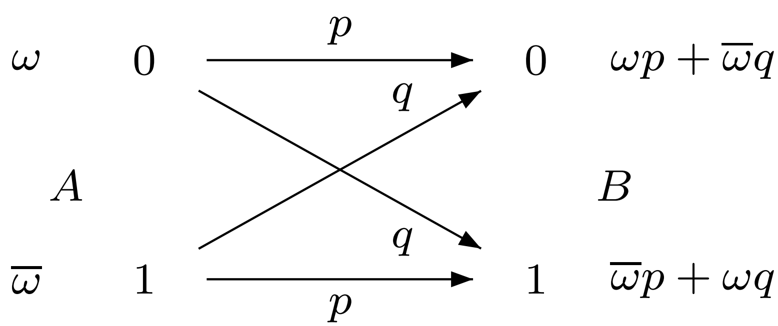
Relações entre a entropia e a informação mútua

$$\begin{aligned} H(A,B) &= H(A) + H(B|A) = H(B) + H(A|B) \\ &= H(A) + H(B) - I(A;B). \end{aligned}$$



Informação mútua de um BSC

Sejam $\omega + \overline{\omega} = p + q = 1$ e considere-se o BSC dado por



$$H(B) = (\omega p + \overline{\omega} q) \log \frac{1}{\omega p + \overline{\omega} q} + (\overline{\omega} p + \omega q) \log \frac{1}{\overline{\omega} p + \omega q},$$

$$\begin{aligned} H(B|A) &= \sum_{a \in A} P(a) \sum_{b \in B} P(b|a) \log \frac{1}{P(b|a)} \\ &= \omega(q \log \frac{1}{q} + p \log \frac{1}{p}) + \overline{\omega}(q \log \frac{1}{q} + p \log \frac{1}{p}) \\ &= q \log \frac{1}{q} + p \log \frac{1}{p}, \end{aligned}$$

sendo $I(A;B) = H(B) - H(B|A)$.

Exemplos

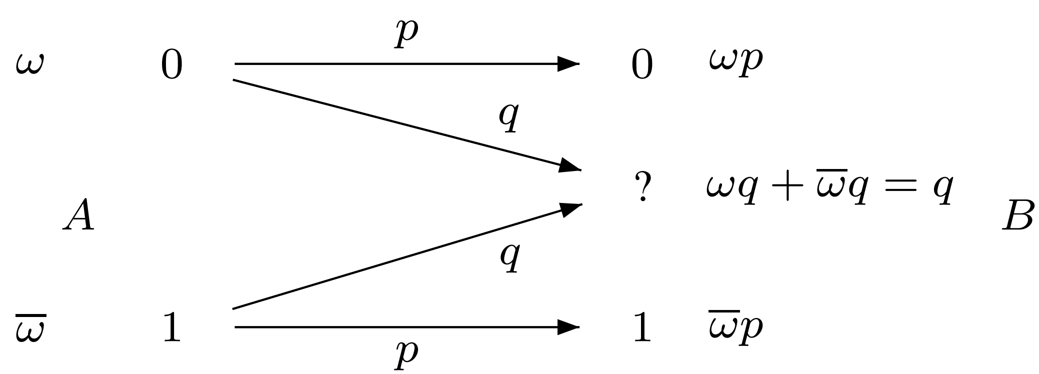
$$H(B) = (\omega p + \overline{\omega} q) \log \frac{1}{\omega p + \overline{\omega} q} + (\overline{\omega} p + \omega q) \log \frac{1}{\overline{\omega} p + \omega q}$$

$$H(B|A) = q \log \frac{1}{q} + p \log \frac{1}{p}$$

- $p = q = 1/2$: $H(B) = 1$, ou seja incerteza total na saída, e o equívoco é também $H(B|A) = 1$; logo a informação mútua é $I(A;B) = 0$.
- $p = 1$: o canal não tem ruído, $H(B) = H(A) = \omega \log \frac{1}{\omega} + \overline{\omega} \log \frac{1}{\overline{\omega}}$, $H(B|A) = H(A|B) = 0$ e a informação mútua é $I(A;B) = H(A) = H(B)$.
- $\omega = \overline{\omega} = 1/2$: a fonte à entrada, e também à saída, tem entropia máxima $H(A) = H(B) = 1$ e a informação mútua é $I(A;B) = 1 - (q \log \frac{1}{q} + p \log \frac{1}{p})$.
- $\omega = 1$: a fonte à entrada não contém qualquer informação, $H(A) = 0$, e $H(B) = H(B|A) = q \log \frac{1}{q} + p \log \frac{1}{p}$, pelo que a informação mútua é $I(A;B) = 0$, não havendo qualquer informação a ser fornecida ao canal.

Informação mútua de um BEC

Considere-se o BEC dado por



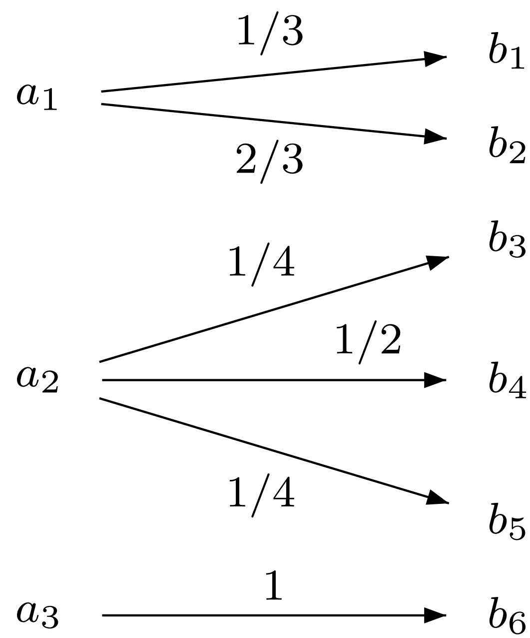
Tem-se então

$$
\begin{aligned}
H(B) &= \omega p \log \frac{1}{\omega p} + \overline{\omega} p \log \frac{1}{\overline{\omega} p} + q \log \frac{1}{q} \\
H(B|A) &= q \log \frac{1}{q} + p \log \frac{1}{p} \\
I(A;B) &= H(B) - H(B|A) = \omega p \log \frac{1}{\omega p} + \overline{\omega} p \log \frac{1}{\overline{\omega} p} - p \log \frac{1}{p}
\end{aligned}
$$

Canais isentos de ruído

- Um *canal isento de ruído* é um canal em que cada símbolo à saída só pode ser produzido por um símbolo específico à entrada.
- Por outras palavras, cada coluna $[P(b_j|a_i)]_i$ da matriz do canal tem exactamente uma entrada não nula, correspondendo à entrada a_i que produziu a saída b_j.

- No entanto, várias saídas podem resultar da mesma entrada:



com matriz

$$P = \begin{bmatrix} 1/3 & 2/3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/4 & 1/2 & 1/4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Informação mútua de um canal isento de ruído

Para um canal isento de ruído, a cada b_j corresponde um único $a_{\varphi(j)}$ que lhe dá origem, sendo $P(a_{\varphi(j)}|b_j) = 1$ e $P(a_i|b_j) = 0$ para $i \neq \varphi(j)$. Assim,

$$\begin{aligned} H(A|B) &= \sum_{a \in A} \sum_{b \in B} P(a,b) \log \frac{1}{P(a|b)} \\ &= \sum_{b_j \in B} P(b_j) \log \frac{1}{P(a_{\varphi(j)}|b_j)} \\ &= 0 \end{aligned}$$

Resultado 2.4 *A informação mútua de um canal isento de ruído é dada por*

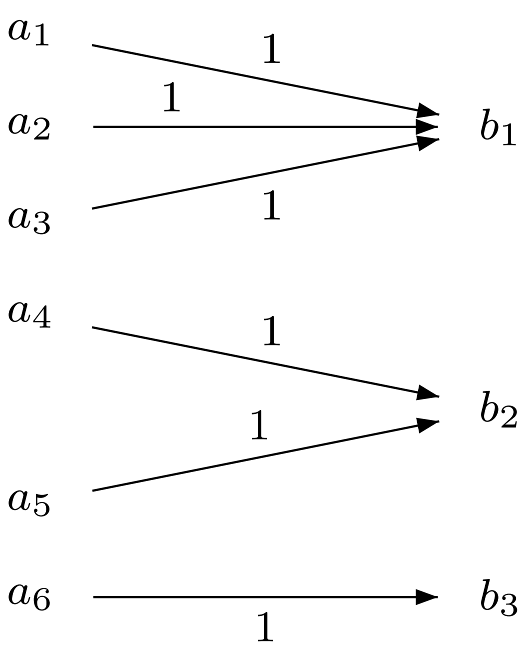
$$I(A;B) = H(A)$$

ou seja, a quantidade de informação dada pelo canal é a mesma que lhe é fornecida à entrada.

Canais determinísticos

- Um canal em que a cada símbolo à entrada corrresponde um único símbolo à saída diz-se um *canal determinístico*.
- A propriedade que caracteriza a matriz de um tal canal é que cada linha tem exactamente uma entrada não nula, a qual tem necessariamente de ser 1.

Exemplo



com matriz $P = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

é um canal determinístico.

Informação mútua de um canal determinístico

Para um canal determinístico, sendo transmitido o símbolo a_i, é recebido um único possível símbolo $b_{\psi(i)}$, pelo que $P(b_{\psi(i)}|a_i) = 1$ sendo os restantes $P(b_j|a_i) = 0$ para $j \neq \psi(i)$. Assim,

$$\begin{aligned} H(B|A) &= \sum_{b\in B}\sum_{a\in A} P(a,b)\log\frac{1}{P(b|a)} \\ &= \sum_{a_i\in A} P(a_i)\log\frac{1}{P(b_{\psi(i)}|a_i)} \\ &= 0 \end{aligned}$$

Resultado 2.5 *A informação mútua de um canal determinístico é dada por*

$$I(A;B) = H(B)$$

ou seja, a quantidade de informação dada pelo canal é a mesma que ele produz à saída.

Canais em cascata

- Vejamos como, encadeando vários canais de informação se comporta a informação mútua.

$$\xrightarrow{\quad A \quad} \boxed{\text{Canal } AB} \xrightarrow{\quad B \quad} \boxed{\text{Canal } BC} \xrightarrow{\quad C \quad}$$

  digamos sendo $|A| = r$, $|B| = s$, $|C| = t$.

- Se ao canal AB é fornecido o símbolo a_i à entrada, e ele produz b_j à saída, então o símbolo b_j é fornecido à entrada do canal BC, que por sua vez produzirá um símbolo c_k à saída.
- Supomos os canais independentes, pelo que c_k depende somente de b_j e não de a_i, ou seja

$$P(c_k|a_i, b_j) = P(c_k|b_j) \quad (\forall i, j, k).$$

- Analogamente, o símbolo a_i fornecido à entrada do canal AB depende somente do b_j produzido à sua saída e não do c_k que venha a resultar da sua transmissão pelo canal BC:

$$P(a_i|b_j, c_k) = P(a_i|b_j) \quad (\forall i, j, k).$$

- O nosso próximo objetivo é comparar $I(A; B)$ com $I(A; C)$, i.e., em que medida a introdução de um canal de informação adicional afeta a informação mútua entre transmissor e recetor.
- Começamos por mostrar que $H(A|C) - H(A|B) \geqslant 0$.

Note-se que $P(a|b) = P(a|b,c)$, $P(a,b,c) = P(b,c)P(a|b,c)$, e $\ln \frac{1}{x} \geqslant 1 - x$, com igualdade se e só se $x = 1$.

$$
\begin{aligned}
H(A|C) - H(A|B) &= \sum_{a \in A} \sum_{c \in C} P(a,c) \log \frac{1}{P(a|c)} - \sum_{a \in A} \sum_{b \in B} P(a,b) \log \frac{1}{P(a|b)} \\
&= \sum_{a \in A} \sum_{b \in B} \sum_{c \in C} P(a,b,c) \log \frac{P(a|b)}{P(a|c)} \\
&\geqslant \frac{1}{\ln 2} \sum_{a \in A} \sum_{b \in B} \sum_{c \in C} P(b,c)P(a|b,c) \left(1 - \frac{P(a|c)}{P(a|b,c)}\right) \\
&\geqslant \frac{1}{\ln 2} \sum_{b \in B} \sum_{c \in C} P(b,c) \left(\sum_{a \in A} P(a|b,c) - \sum_{a \in A} P(a|c)\right) \\
&\geqslant \frac{1}{\ln 2} \sum_{b \in B} \sum_{c \in C} P(b,c)(1-1) = 0
\end{aligned}
$$

com igualdade se e só se $P(a|b) = P(a|c)$ sempre que $P(b,c) \neq 0$.
Notando que $I(A;B) = H(A) - H(A|B)$ e $I(A;C) = H(A) - H(A|C)$, obtemos:

Informação mútua para canais em cascata

Resultado 2.6 *Para a cascata dos canais AB e BC, tem-se $I(A;B) \geqslant I(A;C)$, verificando-se a igualdade se e só se $P(a|c) = P(a|b)$ sempre que $P(b,c) \neq 0$.*
Por outras palavras, a informação resultante da utilização de uma cascata de canais não pode exceder a informação resultante da utilização do primeiro canal.

Vejamos em seguida que, se o canal BC for isento de ruído, então $I(A;B) = I(A;C)$.
Na hipótese sobre BC, se $P(b,c) \neq 0$ então $P(b|c) = 1$.
Logo

$$P(a|c) = \sum_{b \in B} P(a|b,c)P(b|c) = \sum_{b \in B} P(a|b)P(b|c) = P(a|b)$$

pois, sendo o canal BC isento de ruído, $P(b,c) \neq 0 \Rightarrow P(b|c) = 1$ e $P(b,c) = 0 \Rightarrow P(b|c) = 0$.
Logo $I(A;B) = I(A;C)$. $\square$