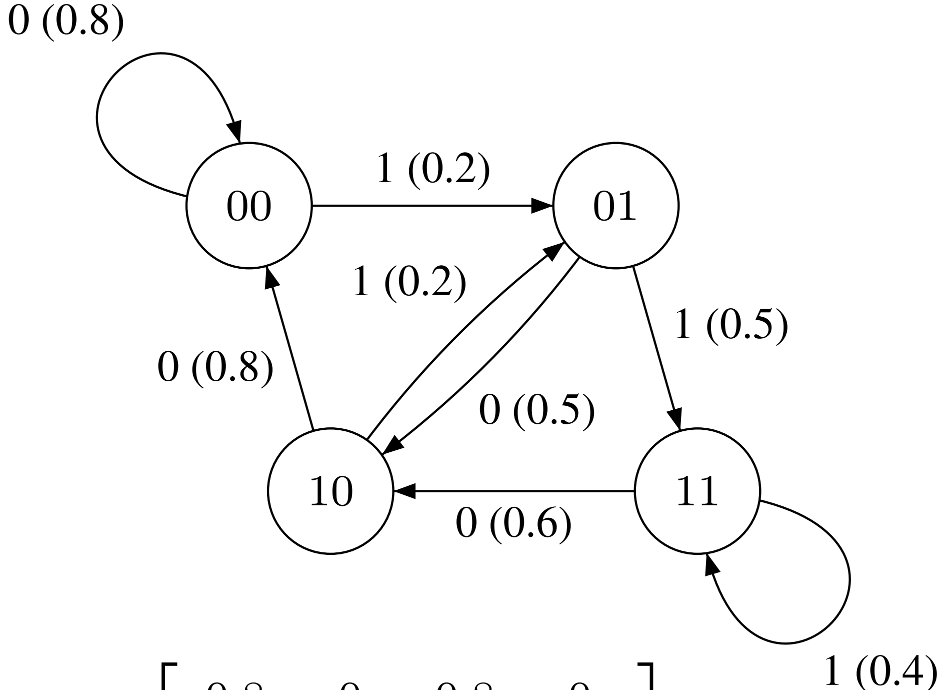
Exemplo

Consideremos a fonte Markov de ordem 2 dada por



cuja matrix de transição é $\Pi = \begin{bmatrix} 0.8 & 0 & 0.8 & 0 \\ 0.2 & 0 & 0.2 & 0 \\ 0 & 0.5 & 0 & 0.6 \\ 0 & 0.5 & 0 & 0.4 \end{bmatrix}$.

Exemplo

Para obter uma fonte estacionária, resolvemos o sistema

$$\begin{aligned}
0.8w_1 + 0w_2 \quad + 0.8w_3 + 0w_4 \quad &= w_1 \\
0.2w_1 + 0w_2 \quad + 0.2w_3 + 0w_4 \quad &= w_2 \\
0w_1 + 0.5w_2 \quad + 0w_3 + 0.6w_4 &= w_3 \\
0w_1 + 0.5w_2 \quad + 0w_3 + 0.4w_4 &= w_4
\end{aligned}$$

sujeito às restrições $w_i \geqslant 0$ $(i = 1, 2, 3, 4)$ e $w_1 + w_2 + w_3 + w_4 = 1$.
Obtemos a distribuição estacionária

$$P(00) = \frac{28}{46},\ P(01) = \frac{6}{46},\ P(10) = \frac{7}{46},\ P(11) = \frac{5}{46}.$$

Assim, as probabilidades da fonte adjunta são

$$P(0) = P(0|00)P(00) + P(0|01)P(01) + P(0|10)P(10) + P(0|11)P(11) = \frac{17}{23}$$
$$P(1) = P(1|00)P(00) + P(1|01)P(01) + P(1|10)P(10) + P(1|11)P(11) = \frac{6}{23}$$

Note-se que, apesar das probabilidades de emissão de um dado símbolo serem as mesmas na fonte de Markov e na fonte adjunta, a probabilidade de emissão de palavras de comprimento maior do que 1 já não é a mesma.
Por exemplo, na fonte original, $P(000) = P(0|00) = 0.8$
enquanto na fonte adjunta, $P(000) = P(0)^3 \simeq 0.4038$, uma vez que na fonte adjunta se supõe a independência da emissão de símbolos.

Uma vez que a passagem à fonte adjunta reduz as restrições na emissão de símbolos, aumenta a incerteza sobre as sequências geradas, pelo que a entropia aumenta:

Teorema 1.13 *Se $\bar{M}$ é a adjunta da fonte de Markov M, então as suas entropias verificam a desigualdade $H(M) \leqslant H(\bar{M})$.*

Prova. [Para fontes de Markov de ordem m.] Seja M uma fonte de Markov de ordem m com alfabeto $\{a_1, \ldots, a_q\}$.

Representamos os estados, formados pelas palavras de comprimento m, por α_I, com $1 \leqslant I \leqslant q^m$.

Assumimos que M possui uma distribuição estacionária $(w_I)_{1 \leqslant I \leqslant q^m}$.

As probabilidades de emissão dos símbolos são dadas por

$$P(a_i) = \sum_I w_I p(a_i|\alpha_I).$$

A entropia da fonte adjunta é dada por

$$\begin{aligned} H(\bar{M}) &= -\sum_{i=1}^{q} P(a_i)\log(P(a_i)) \\ &= -\sum_{i=1}^{q}\sum_{I} w_I P(a_i|\alpha_I)\log(P(a_i)) \\ &= -\sum_{I} w_I \sum_{i=1}^{q} P(a_i|\alpha_I)\log(P(a_i)). \end{aligned}$$

A entropia do I-ésimo estado é dada por $H(P_I) = -\sum_{i=1}^{q} P(a_i|\alpha_I)\log(P(a_i|\alpha_I))$ pelo que a entropia de M é

$$\begin{aligned} H(M) &= -\sum_{I}\sum_{i=1}^{q} w_I P(a_i|\alpha_I)\log(P(a_i|\alpha_I)) \\ &= -\sum_{I} w_I \sum_{i=1}^{q} P(a_i|\alpha_I)\log(P(a_i|\alpha_I)) \end{aligned}$$

Finalmente, basta aplicar o Lema 1.1 às somas sobre i. Note-se que se tem a igualdade se e só se $P(a_i) = P(a_i|\alpha_I)$, isto é a própria fonte de Markov dada não tem memória. $\square$

Citação

When Shannon had invented his quantity and consulted von Neumann on what to call it, von Neumann replied: 'Call it entropy. It is already in use under that name and besides, it will give you a great edge in debates because nobody knows what entropy is anyway'.

(Denbigh 1990)

Extensões de fontes

- A *n-ésima extensão* de uma fonte estacionária sem memória S sobre um alfabeto A é a fonte estacionária sem memória S^n tal que
  - o alfabeto é A^n;
  - a probabilidade de emissão de $w \in A^n$ por S^n é a probabilidade de emissão da sequência w pela fonte S.

Teorema 1.14 *Sendo S^n a n-ésima extensão da fonte estacionária sem memória S, tem-se a seguinte igualdade:*

$$H(S^n) = nH(S).$$

Prova. Seja $A = \{a_1, \ldots, a_q\}$ o alfabeto de S e seja $P(a_i)$ a probabilidade de emissão do símbolo a_i. Então a entropia de S é dada por

$$H(S) = -\sum_{i=1}^{q} P(a_i) \log(P(a_i)).$$

Como a probabilidade de emissão por S^n do símbolo $a_{i_1} \cdots a_{i_n}$ é

$$P(a_{i_1} \cdots a_{i_n}) = P(a_{i_1}) \cdots P(a_{i_n}),$$

temos

$$
\begin{aligned}
H(S^n) &= -\sum_{i_1=1}^{q} \cdots \sum_{i_n=1}^{q} P(a_{i_1} \cdots a_{i_n}) \log(P(a_{i_1} \cdots a_{i_n})) \\
&= -\sum_{i_1=1}^{q} \cdots \sum_{i_n=1}^{q} P(a_{i_1} \cdots a_{i_n}) \sum_{j=1}^{n} \log(P(a_{i_j})) \\
&= -\sum_{j=1}^{n} \sum_{i_1=1}^{q} \cdots \sum_{i_n=1}^{q} P(a_{i_1} \cdots a_{i_n}) \log(P(a_{i_j})) \\
&= -\sum_{j=1}^{n} \sum_{i_1=1}^{q} P(a_{i_1}) \cdots \sum_{i_j=1}^{q} P(a_{i_j}) \log(P(a_{i_j})) \cdots \sum_{i_n=1}^{q} P(a_{i_n}) \\
&= -\sum_{j=1}^{n} \sum_{i_j=1}^{q} P(a_{i_j}) \log(P(a_{i_j})) \\
&= \sum_{j=1}^{n} H(S) \\
&= nH(S).
\end{aligned}
$$

$\square$

Extensões de fontes de Markov de ordem m

- A *n-ésima extensão* de uma fonte de Markov M de ordem m sobre o alfabeto A é a fonte de Markov M^n de ordem $p = \lceil m/n \rceil$ tal que
  - o alfabeto é A^n,
  - a probabilidade da transição

$$(v_1, \ldots, v_p) \xrightarrow{v_{p+1}} (v_2, \ldots, v_{p+1})$$

é definida pela probabilidade condicionada do correspondente trajeto de M, onde $r = pn - m$ e $v_i = s_{i,1} \cdots s_{i,n}$:

$s_{1,1} \cdots s_{1,n} s_{2,1} \cdots s_{2,n} \cdots \cdots s_{p,1} \cdots s_{p,n-r}$,

$s_{1,2} \cdots s_{1,n} s_{2,1} \cdots s_{2,n} \cdots \cdots s_{p,1} \cdots s_{p,n-r+1}$,

$\ldots$,

$s_{2,r+1} \cdots s_{2,n} s_{3,1} \cdots s_{3,n} \cdots \cdots s_{p,1} \cdots s_{p,n} s_{p+1,1} \cdots s_{p+1,n}$

Teorema 1.15 *Sendo M^n a n-ésima extensão da fonte de Markov M, as suas entropias verificam a igualdade*

$$H(M^n) = nH(M).$$

Prova. Exercício. $\Box$

- Note-se que, para $m < n$ e $H(M) > 0$, tem-se

$$H(M^m) = mH(M) < nH(M) = H(M^n).$$

- Se M é de ordem m, e $\overline{M^n}$ é a adjunta da n-ésima extensão M^n, então

$$H(\overline{M^n}) \geqslant H(M^n) = nH(M).$$

Espaços de amostragem infinitos

- Consideremos o espaço de amostragem $\mathbb{N} = \{0, 1, 2, \ldots\}$, dos números naturais e a distribuição de probabilidade dada por $P(n) = 2^{-n-1}$. A entropia é ainda definida como sendo o valor esperado da surpresa, dada pela função $-\log P(n)$:

$$H(P) = -\sum_{n=0}^{\infty} 2^{-n-1}(-n-1) = \sum_{n=1}^{\infty} \frac{n}{2^n} = 2.$$

Seja $f(x) = \frac{1}{2}\sum_{n=1}^{\infty} x^n = \frac{x}{2(1-x)}$ para $|x| < 1$. Então, derivando termo a termo a série de potências, obtemos $f'(x) = \frac{1}{2(1-x)^2} = \frac{1}{2}\sum_{n=1}^{\infty} n x^{n-1}$ para $|x| < 1$. Em particular, tomando $x = 1/2$, concluímos que $\sum_{n=1}^{\infty} \frac{n}{2^n} = f'(1/2) = 2$.

- No caso dum espaço de amostragem contínuo, em vez de somas estão envolvidos integrais e, em vez duma distribuição de probabilidade consideramos uma função de densidade de probabilidade f, digamos definida em $\mathbb{R}^n$. A probabilidade dum subconjunto S de $\mathbb{R}^n$ é então $P(S) = \int_S f(x)\,dx$ caso o integral esteja definido. Por exemplo, no caso de $n = 1$, $P([a,b]) = \int_a^b f(x)\,dx$.
- Para uma outra função $g : \mathbb{R}^n \to \mathbb{R}^m$, o *valor esperado* de g é $\int_{\mathbb{R}^n} g(x) f(x)\,dx$. Note-se que quando este integral estiver definido, o seu valor é um elemento de $\mathbb{R}^m$.

- Em particular, a *média* e a *variância* da função de densidade de probabilidade f em $\mathbb{R}^n$ são dadas respetivamente por

$$\mu = \int_{\mathbb{R}^n} x f(x)\, dx \quad \text{e} \quad \sigma^2 = \int_{\mathbb{R}^n} \|x - \mu\|^2 f(x)\, dx,$$

  onde se considera a norma Euclidiana.

- Por sua vez, a *entropia (diferencial)* de f é o valor esperado da função *surpresa* $-\log f(x)$, ou seja

$$H(f) = -\int_{\mathbb{R}^n} f(x) \log f(x)\, dx.$$

Teorema 1.16 *A entropia da função de densidade Gaussiana com média μ e variância σ^2 é $\ln(\sigma\sqrt{2\pi e})$ nits.*

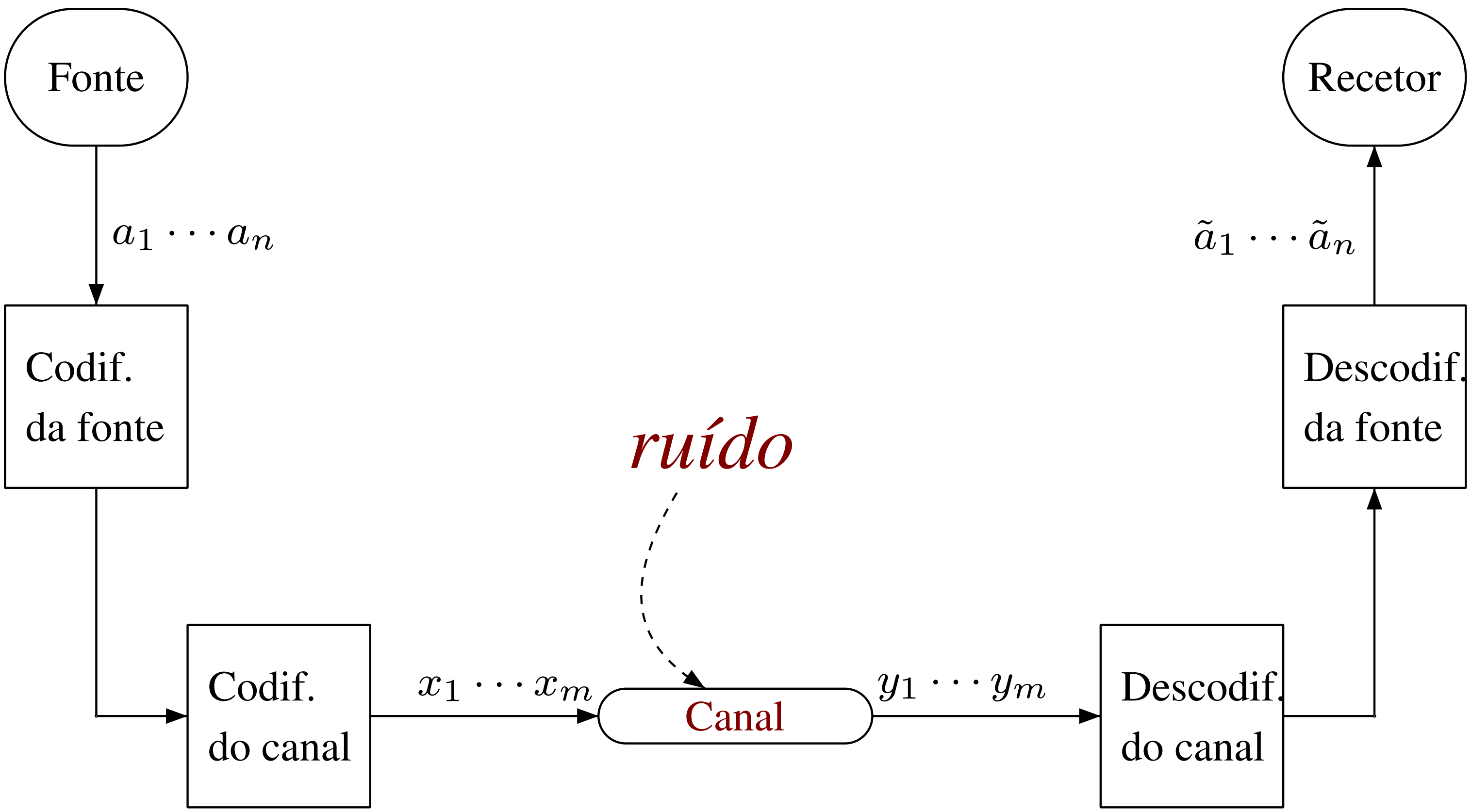
Prova. Por definição, a função de densidade Gaussiana em causa é dada por

$$g(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{\|x-\mu\|^2}{2\sigma^2}}.$$

Mostra-se que se trata duma função de densidade, pelo que se tem $\int_{\mathbb{R}^n} g(x)\,dx = 1$ e a sua média e variância são efetivamente μ e σ^2, respetivamente. Podemos então calcular a entropia como segue:

$$\begin{aligned}
H(g) &= -\int_{\mathbb{R}^n} g(x)\ln g(x)\,dx \\
&= -\int_{\mathbb{R}^n} g(x)\left(-\ln(\sigma\sqrt{2\pi}) - \frac{\|x-\mu\|^2}{2\sigma^2}\right) dx \\
&= \int_{\mathbb{R}^n} g(x)\ln(\sigma\sqrt{2\pi})\,dx + \int_{\mathbb{R}^n} g(x)\frac{\|x-\mu\|^2}{2\sigma^2}\,dx \\
&= \ln(\sigma\sqrt{2\pi})\int_{\mathbb{R}^n} g(x)\,dx + \frac{1}{2\sigma^2}\int_{\mathbb{R}^n} g(x)\|x-\mu\|^2\,dx \\
&= \ln(\sigma\sqrt{2\pi}) + \frac{\sigma^2}{2\sigma^2} = \ln(\sigma\sqrt{2\pi}) + \ln\sqrt{e} = \ln(\sigma\sqrt{2\pi e}).\square
\end{aligned}$$

2 Canais de informação



- No capítulo anterior, estudámos alguns modelos matemáticos de fontes de informação e como medir a quantidade de informação.
- Mas, a informação que fica na fonte é essencialmente inútil, havendo que a comunicar através dos meios apropriados.
- No modelo que vamos considerar dos canais de informação, vamos supô-los
    - *estacionários*: não há variação do comportamento estatístico do canal ao longo do tempo
    - *sem memória*: o comportamento do canal num dado instante não depende do seu comportamento anterior.

Canais de informação

- Um *canal de informação* é um terno (A, B, P) onde
  - $A = \{a_i : i = 1, \ldots, r\}$ é o *alfabeto de entrada*
  - $B = \{b_j : j = 1, \ldots, s\}$ é o *alfabeto de saída*
  - P é a *probabilidade do canal* que mede a probabilidade $P(b_j|a_i)$ de ser emitido à saída o símbolo b_j quando à entrada recebe o símbolo a_i.

Notas

- A razão de usar diferentes alfabetos à entrada e à saída é para modelar o fenómeno do ruído que altera a informação, eventualmente conduzindo a símbolos que nunca seriam emitidos.
- Do ponto de vista do modelo estatístico, o que está em causa é a relação entre duas variáveis aleatórias (símbolo à entrada, símbolo à saída).

Matriz do canal

- A *matriz* **P** *do canal* é a matriz

$$\mathbf{P} = \begin{bmatrix} P(b_1|a_1) & P(b_2|a_1) & \cdots & P(b_s|a_1) \\ P(b_1|a_2) & P(b_2|a_2) & \cdots & P(b_s|a_2) \\ \vdots & \vdots & \ddots & \vdots \\ P(b_1|a_r) & P(b_2|a_r) & \cdots & P(b_s|a_r) \end{bmatrix}$$

cuja entrada $P(b_j|a_i)$ também poderemos representar por P_{ij}.

- Assumimos que, para $i = 1, 2, \ldots, r$,

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

ou seja, se o canal recebe algum símbolo, então, com probabilidade 1 vai emitir algum símbolo.

- No caso do canal não ter ruído, o símbolo emitido coincide com o símbolo recebido.
- Para um canal com ruído, a probabilidade do símbolo recebido ser alterado é positiva.