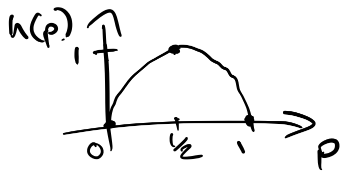
Revision – Chapter 1

Bernoulli rv $X = \begin{cases} 0 & \text{with prop. } p \\ 1 & \text{" } 1-p \end{cases}$ $p \in [0,1]$

$$H(X) = \boxed{h(p) := p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p}}$$

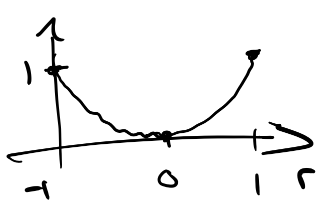


Mutual information, two binary rv X and Y with joint pmf

X\Y	0	1
0	$\frac{1}{4}(1+r)$	$\frac{1}{4}(1-r)$
1	$\frac{1}{4}(1-r)$	$\frac{1}{4}(1+r)$

$r \in [-1, 1]$

$$I(X:Y) = H(X) - H(X|Y)$$
$$= 1 - \sum_y P_Y(y) H(X|Y=y)$$
$$= 1 - h\left(\frac{1+r}{2}\right)$$



Cond. mutual information

Chinese whispers

$$
\begin{tikzcd}
 & X \arrow[dl] \arrow[dr] & \\
Y_1 \arrow[d] & & Z_1 \arrow[d] \\
\vdots \arrow[d] & & \vdots \arrow[d] \\
Y_n & & Z_n
\end{tikzcd}
$$

$I(Y_n : Z_n)$?

$I(Y_n : Z_n | X) = 0$!

forms a Markov chain

Chapter 2 : Source coding

2.1 Problem setup and definitions

2.11 Data Source : infinite sequence of rvs

$$\mathbb{X} = X_1, X_2, X_3, \dots, X_k, \dots$$

Properties : X_i is discrete, takes values in $\mathcal{X}$, $|\mathcal{X}| < \infty$

X_i are independent and identically distributed (i.i.d.)

$$P_{X_1 X_2 \dots X_k \dots}(x_1, x_2, \dots, x_k, \dots) = P_X(x_1) P_X(x_2) \dots P_X(x_k) \dots$$

$$x_1, x_2, \dots, x_k \in \mathcal{X}$$

Such a source is called a discrete memoryless source (DMS).

2.1.2 Source code (or code)

Def. A code is a map C from $\mathcal{X}$ to $\{0,1\}^*$
$\Rightarrow C(x) \in \{0,1\}^*$
We denote by $\ell(x)$ the length of $C(x)$.

<u>Properties</u>:

- A code is called fixed-length if $\ell(x) = \ell$ is constant, otherwise it is called variable length.

[NS] A code is called non-singular if C is injective, i.e. $C(x) \neq C(x')$ for $x \neq x'$

[P] A code is called a prefix code if for any $x \neq x'$: $C(x)$ is not a prefix of $C(x')$

(01 is a prefix of 01011,
01 and 101 are not prefixes of each other)

[U] A code is uniquely decodable if there exists a decoder that for any $n \in \mathbb{N}$ and any sequence $x^n \in \mathcal{X}^n$ can recover x^n from $C(x_1)C(x_2)\ldots C(x_n)$

[1] A code is called instantaneous if it is UD and if the decoder can deduce x_k as soon as it has seen the bitstring

$$C(x_1)C(x_2)\ldots C(x_k)C(x_{k+1})$$

up until (and including $C(x_k)$).

$$\begin{tikzcd}
P \arrow[r, Leftrightarrow] \arrow[d, Rightarrow] & I \arrow[d, Rightarrow] \\
NS \arrow[r, Rightarrow, "/" description] & U \arrow[l, Rightarrow] \arrow[u, Rightarrow, bend right, "/" description]
\end{tikzcd}$$

Examples:

- Let $\mathcal{X} = \{0, 1, 2, 3\}$

$C(0) = 0$, $C(1) = 1$ $C(2) = 10$, $C(3) = 11$

This a NS code, but

$$110 \begin{cases} = C(1)C(2) \\ = C(1)C(1)C(0) \\ = C(3)C(0) \end{cases}$$

$\Rightarrow$ not uniquely decodable!

- $\mathcal{X} = \{\text{'a'}, \text{'b'}, \text{'c'}\}$,

$C(\text{'a'}) = 1$, $C(\text{'b'}) = 10$, $C(\text{'c'}) = 00$

This code is not (1) because the sequence 100 can be $C('a')C('c')$ or part of $C('b')C('c')$

However, if we know the sequence is complete we can decode uniquely (by looking at parity of # of 0's between two ones).

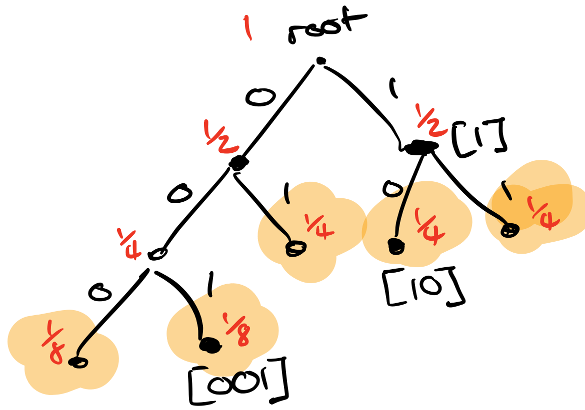
Proposition: A code is instantaneous iff it is a prefix code.

Proof: Prefix => Instantaneous:

If you see $C(x_1)C(x_2)\dots$, once $C(x_1)$ is fully read you can decode x_1 because $C(x_1)$ is not a prefix of any other code-word. Moreover, we could not have decoded anything else because $C(x_1)$ does not contain any other codewords as prefixes.

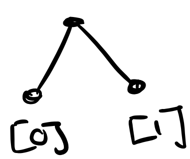
Instantaneous => Prefix $\square$

Any (NS) code can be represented as a binary tree where codewords are given by the path from root to the node.



A prefix code has all its codeword on leaves* (*assuming it has minimal depth).

Simplest example!



2.2. Variable-length codes

2.2.1. Optimal codeword lengths

<u>Prop.</u> McMillan-Kraft inequality:

$$\sum_{x \in \mathcal{X}} 2^{-\ell(x)} \le 1 \qquad (1)$$

For any (u) code. In turn, if $\ell(x)$ satisfy Eq. (1), then there exists a prefix code with these codeword lengths.

<u>Proof:</u> For prefix codes $\longrightarrow$ look at the tree!

2.2.2. Optimal expected codeword length

Def.: Expected codeword length

$$\bar{\ell}(C) := \mathbb{E}[\ell(X)] = \sum_{x\in\mathcal{X}} P_X(x)\,\ell(x)$$

<u>Prop</u>: For any $\mathcal{U}$ code for a DMS $X \leftarrow P_X$, we have

$$\bar{\ell}(C) \geq H(X).$$

Moreover, this is saturated iff $\ell(x)$ saturate Eq. (1) and $P_X(x) = 2^{-\ell(x)}$.

<u>Proof</u>: We write

$$\begin{aligned}\bar{\ell}(C) - H(X) &= \mathbb{E}\left[\ell(X) - \log\frac{1}{P_X(X)}\right]\\ &= \mathbb{E}\left[\log\frac{P_X(X)}{2^{-\ell(X)}}\right]\\ &\geq \mathbb{E}\left[\log\frac{t\,P_X(x)}{2^{-\ell(x)}}\right]\\ &= D(P_X \,\|\, Q_X) \geq 0\end{aligned}$$

$t = \sum_x 2^{-\ell(x)} \leq 1$ by Eq. (1)

positivity of rel. entropy

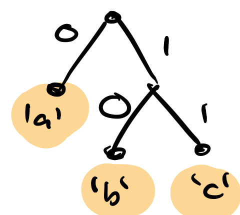
$$Q_X(x) = \frac{2^{-\ell(x)}}{t}$$

2.2.3 Shannon code

Example: $\mathcal{X} = \{\text{'a'}, \text{'b'}, \text{'c'}\}$

$P_X(\text{'a'}) = \frac{1}{2}$, $P_X(\text{'b'}) = P_X(\text{'c'}) = \frac{1}{4}$

Code C



$$\bar{\ell}(C) = \frac{1}{2}\cdot 1 + \frac{1}{2}\cdot 2 = \frac{3}{2}$$

$$H(X) = \frac{1}{2}\log 2 + \frac{1}{4}\log 4 + \frac{1}{4}\log 4 = \frac{3}{2}$$

Prop.: For any DMS $X \leftarrow P_X$ there exists a code C with $\bar{\ell}(C) \leq H(X) + 1$.
(This is the Shannon code.)

Proof: By construction:

Choose $\ell(x) = \left\lceil \log \frac{1}{P_X(x)} \right\rceil$.

Then:
$$\sum_x 2^{-\ell(x)} = \sum_x 2^{-\left\lceil \log \frac{1}{P_X(x)} \right\rceil}$$
$$\leq \sum_x 2^{-\log \frac{1}{P_X(x)}}$$
$$= \sum_x P_X(x) = 1 \checkmark$$

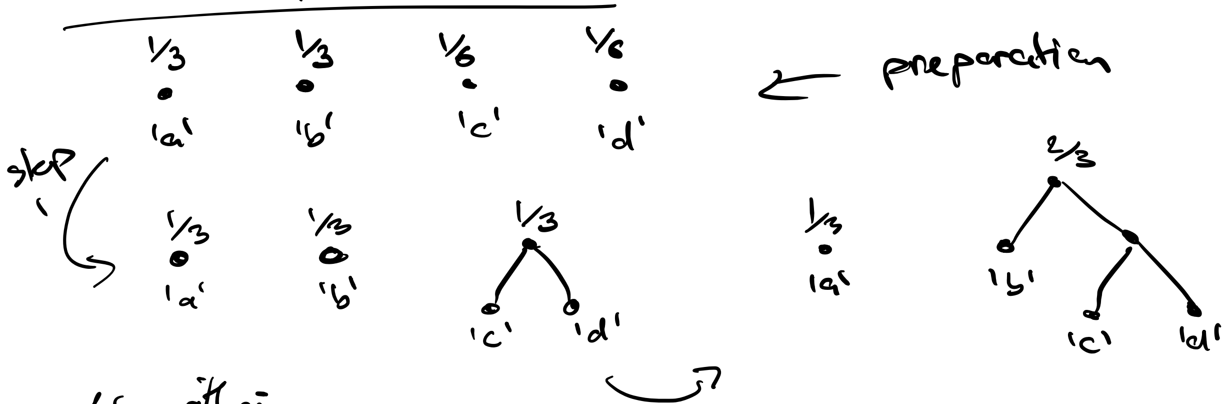
$\Rightarrow$ there exists a prefix code with these codeword lengths

Moreover $\bar{\ell}(C) = \mathbb{E}\left[\left\lceil \log \frac{1}{P_X(x)} \right\rceil\right]$

$$\leq \mathbb{E}\left[\log \frac{1}{P_X(x)} + 1\right]$$

$$= H(X) + 1$$

2.2.4 Huffman codes



Algorithm

while {there is more than 1 tree in the forest}

- pick 2 trees with smallest probabilities
- join them on a new root, that tree now has the sum of the probabilities

end