Information theory 1,

Information Theory

INFORMATION & UNCERTAINTY

Definition

$$I(A) \triangleq \mathrm{ld}\,\frac{1}{P(A)} = -\,\mathrm{ld}(P(A)) \quad [\text{bits}]$$

where :

- A – an event
- $P(A)$ – Probability of A occuring
- $\mathrm{ld}(x)$ – Logarithm base 2
- $I(A)$ – The information in getting to know A occured
- **bit** – unit of measure , older units : nat, decit

The binary logarithm

$$\mathrm{ld}(x) = \log_2(x) = \mathrm{lb}(x) = \frac{\log x}{\log 2} = \frac{\ln x}{\ln 2}$$

$$= 1{,}443 \cdot \ln x$$

EX.1

$\mathrm{ld}\,1 = 0$ $\quad \mathrm{ld}\,\frac{1}{2} = -1$

$\mathrm{ld}\,2 = 1$ $\quad \mathrm{ld}\,\frac{1}{4} = -2$

$\mathrm{ld}\,4 = 2$

$\mathrm{ld}\,1024 = 10$

Simplified notation

$\mathrm{ld}\,0 = -\infty$

$0 \cdot \mathrm{ld}\,0 = 0$

For independent events:

$$P(A \cap B) = P(A)P(B) \quad [\cdot]$$

$$I(A \cap B) = I(A) + I(B) \quad [\text{bits}]$$

EX. 2

Heads and tails:

$$I(\text{tail}) = -\operatorname{ld}\frac{1}{2} = \operatorname{ld} 2 = 1 \text{ bit}$$

Casino wheel:

$$I('0') = -\operatorname{ld}\frac{1}{37} = \operatorname{ld} 37 = 5,21 \text{ bit}$$

Sources of Information

Two kinds of sources:

1. Independent sources (memory less)
2. Sources with memory (Markov sources)

Sthocastic processes

Model for information sources.

$$\underset{\text{sthocastic process}}{X} \longrightarrow x_k \qquad \underbrace{\begin{Bmatrix} x_1 \\ \vdots \\ x_n \end{Bmatrix}}_{\text{Alphabet } \mathcal{A}}$$

$N = \# \mathcal{A}$

Alphabet size

Cardinality of $\mathcal{A}$

x_k are symbols (generic name)

- Letters
- characters
- signs
- outcomes
- events
- tosses

examples

Simplified notation

$P(x_k)$ means

$P(X = x_k)$

Entropy

For an independent source an entropy is defined.

$$H(X) = E[-\mathrm{ld}\, P(X_k)] =$$
$$= E[I(X_k)]$$
$$= \sum p(X_k) \cdot \mathrm{ld}\, P(X_k) \quad [\text{bits/symbol}]$$

The average information contents per symbol.

EX.3

Heads and tails

$$H(X) = -\sum_{k=1}^{2} P \cdot \mathrm{ld}\, P = -\sum_{k=1}^{2} \frac{1}{2} \cdot \mathrm{ld}\, \frac{1}{2} = 1 \text{ bit/toss}$$

The binary entropy function

A process with $N = 2$

$\Psi(P)$

The maximum value of $H(X)$

It can be shown:

$$H(X) = H_{MAX} = \mathrm{ld}\, N \quad \text{for} \quad P(X_k) = \frac{1}{N}$$

Hence we have:

$$H(X) \leq \mathrm{ld}\, N$$

EX 4

DK alphabet : $N = 30$ (29 letters + space)

$$H_{.} = \mathrm{ld}\, 30 = 4.91 \text{ bit/letter}$$

Conditional probability

A variety of Bayes rule says:

$$P(X,Y) = P(X|Y)\,P(Y) = P(Y|X)\,P(X)$$

Notation:

$$P(X_k) = P(X = x_k)$$
$$P(X_k, Y_n) = P(X = x_k \wedge Y = y_n)$$

$P(X)$ The distribution of X

$P(X|Y)$ The conditional distribution

$P(X,Y)$ The joint distribution

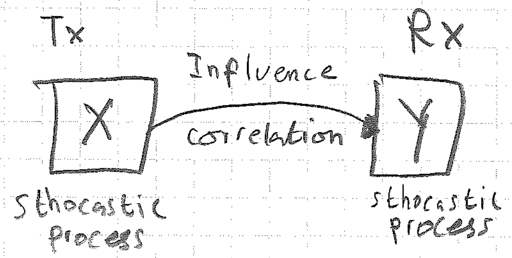
For independent processes:

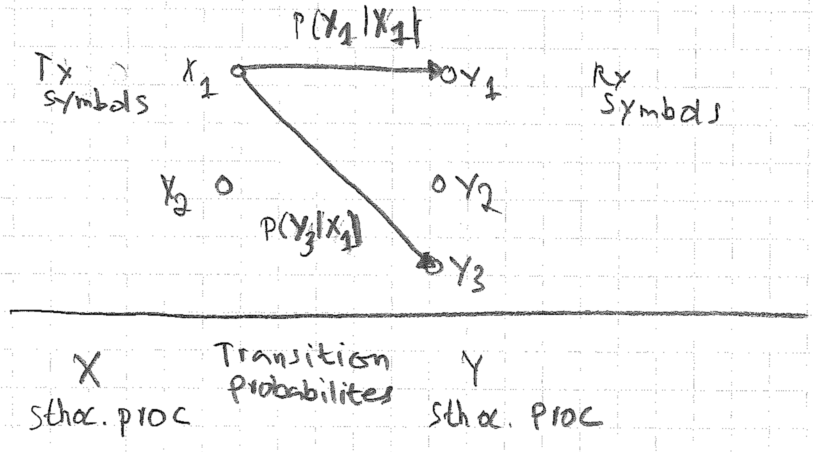
$$P(X,Y) = P(X)\,P(Y)$$
$$P(X|Y) = P(X)$$

The marginal distribution

$$P(X) = \sum_n P(X_k|Y_n)\,P(Y_n)$$
$$= \sum_n P(X_k, Y_n)$$

Channel model





Properties:

1. $\sum_k P(X_k) = 1$
2. $\sum_n P(Y_n) = 1$
3. $\sum_n P(Y_n | X_k) = 1$

Conditional entropy

The information in » X_K given Y_n «

$$I(X_k | Y_n) = -\mathrm{ld}\, P(X_k | Y_n) \quad [\text{bit}]$$

The entropy of X given Y_n:

$$H(X | Y_n) = -\sum_k P(X_k | Y_n) \cdot \mathrm{ld}\, P(X_k | Y_n) \quad [\text{bit/symb.}]$$

Summing over all Y_n's

$$H(X|Y) = \sum P(Y_n)\, P(X_k | Y_n) \cdot \mathrm{ld}\, P(X_k | Y_n) \quad [\text{bit/symbol}]$$

missing information in X

The » equivocation of X wrt. Y «

$$H(X|Y) = -\sum_n \sum_k \underbrace{P(x_k, y_n)}_{\substack{P(x_k|y_n)\,P(y_n) \\ P(y_n|x_k)\,P(x_k)}} \cdot \operatorname{ld} P(x_k|y_n) \quad [\text{bit/symbol}]$$

Normally we know: $P(X)$ and $P(Y|X)$

$$P(Y) = \sum_k P(Y|X)\,P(X)$$

$$P(X|Y) = \frac{P(X,Y)}{P(Y)} = \frac{P(Y|X)\,P(X)}{P(Y)}$$

$$= P(Y|X) \cdot \frac{P(X)}{P(Y)}$$

Mutual Information

Mutual entropy, transinformation

$$I(X;Y) \triangleq H(X) - H(X|Y) \quad [\text{bits/symbol}]$$

other notations: $I(X \rightarrow Y)$ or $I(X \leftrightarrow Y)$

It can be shown:

$$I(X;Y) = I(Y;X) = H(X) - H(X|Y)$$

$$= H(Y) - H(Y|X) \quad [\text{bit/symb}]$$

$\boxed{Y|X \quad \text{X given Y}}$

For the system as a whole:

Joint entropy

$$H(X,Y) = -\sum_n \sum_k p(X_k, Y_n) \cdot \mathrm{ld}\, p(X_k, Y_n) \quad [\text{bit/symb}]$$

Theorem

$$H(X,Y) = H(X) + H(Y|X)$$
$$= H(Y) + H(X|Y)$$
$$= H(X|Y) + H(Y|X) + \underbrace{I(X;Y)}_{\text{amount of symbols transferred}} \quad [\text{bit/symb}]$$

Entropy rate

$$R_{XY} = I(X;Y) \cdot r_s \quad [\text{bit/s}]$$

where:

$I(X;Y)$ is the mutual entropy [bit/symb]

r_s is the symbol rate (average) [symbols/s]