DATE 07 · 08 · 2017

Information Theory (P Vijay Kumar)

Lecture 1.

Text: Cover & Thomas : 2-5 chapters
Elements of Information theory 7-9 "

Ref: Basic concepts of Information theory & Coding by Golomb Peile Scholtz.

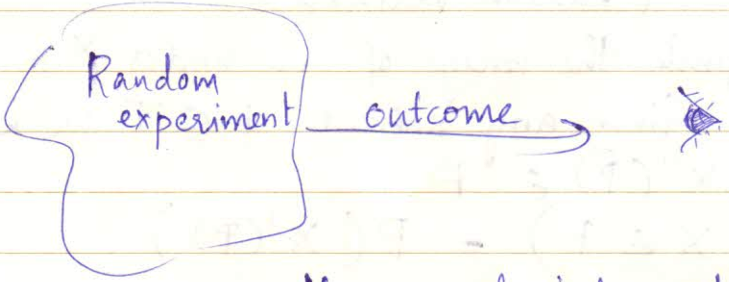
Out of Campus

MW 9-10.30 am
F 8.30-10am

{ Aug 29 - Sep 9
Sep 25 - Oct 2

Applications
- data compression
- channel capacity
- kolomogorov complexity
- Mathematics (Statistics)

Measure of information



· How much information did I get?

Probability Space

(Ω, F, P)

Ω → Sample space (set of all possible outcomes of experiment)

F → events of interest (σ-algebra)

P → Probability measure.

Natural

Will assume for now that Ω is finite (Eg: $\Omega = \{1,2,3,4,5,6\}$ throws of a die)

F = collection of subsets of Ω satisfying

1) $\Omega \in F$
2) if $A \in F$ then $A^c \in F$
3) $\{E_i\} \in F \Rightarrow \bigcup_{i=1}^{n} E_i \in F$ (can be countable union as well)

Eg: F = all subsets of Ω.

Eg: $F = \{ \phi, \Omega, \{1,3,5\}, \{2,4,6\} \}$

P = Probability measure on F

$P : F \longrightarrow [0,1]$

a) $P(\Omega) = 1$ $\qquad E_i \in F$

b) If $\{E_i\}$ are disjoint events, then

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i)$$

Eg: $P(\{i\}) = \frac{1}{6} \quad \forall\, 1 \le i \le 6$

$X : \Omega \longrightarrow R^1$ (Random Variable)

Since Ω is finite the image of Ω under X is finite, say T. Given any subset $P \subseteq T$ we will require that $X^{-1}(P) \in F$

then $P(X \in P) = P(X^{-1}(P))$

let $|\Omega| = n$ let $X(\Omega) = \{x_1\, x_2 \ldots x_n\}$ say distinct

let $X^{-1}(x_i) = A_i$

Then $A_i \in F$, $\bigcup_{i=1}^{n} A_i = \Omega$ and A_i are disjoint.

Let $P_r(X = x_1) = P_X(x_1) = p_1$ etc

$\therefore p_i = P(A_i)$

$H(X)$ entropy of the random variable X $\triangleq$ $\left\{ \begin{array}{l} \text{the average uncertainity} \\ \text{dispelled when } X \text{ is} \\ \text{observed.} \end{array} \right.$

Start with some simple cases:

We will by some abuse of notation write $H(p_1, p_2, \dots p_n)$ in place of $H(X)$

Axiom 1: If all events are equally likely, the uncertainity function increases as the number of events increases.

Meaning: Suppose $P_X(x_i) = \frac{1}{n}$ $\forall i, 1 \le i \le n$

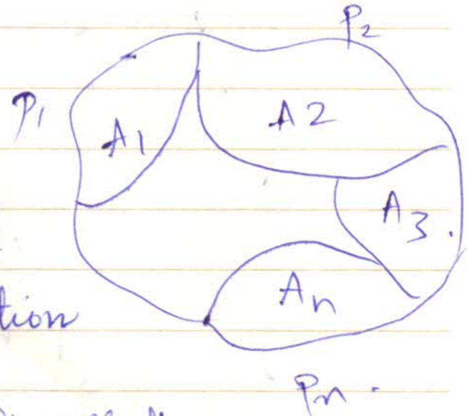
then $H(X) \Uparrow$ as $n \Uparrow$.

$h(n) = H(\underbrace{\tfrac{1}{n}, \dots \tfrac{1}{n}}_{n \text{ terms}})$.

Then $h(\cdot)$ increases monotonically with n.

(Grouping Axiom)

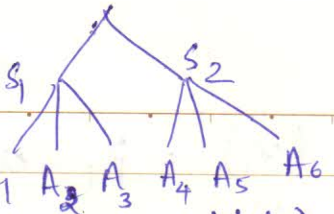
Axiom 2: The total uncertainity eliminated by indicating which event actually occurred does not depend on the method of indication



Events: $A_1\ A_2\ A_3\ A_4\ A_5\ A_6$

$S_1 = \{A_1, A_2, A_3\}$ $\left\{ \begin{array}{l} \text{assume} \\ \text{all equally} \\ \text{likely} \end{array} \right)$

$S_2 = \{A_4, A_5, A_6\}$



$$h(6) = h(2) + \frac{1}{2}h(3) + \frac{1}{2}h(3)$$
$$= h(2) + h(3)$$

$h(2*3) = h(2) + h(3)$ $\longrightarrow$ looks like logarithmic function

<u>Thm</u>: It follows from the grouping Axiom that
$h(mn) = h(m) + h(n)$, m, n integers

<u>Claim</u>: $h(m) = \lambda \log m$ (where λ is some constant)

<u>Pf</u>: Let a, b, c be integers > 1
$\exists$ d an integer s.t

$$c^d \leq a^b < c^{d+1}$$

$$\therefore \quad d \log c \leq b \log a < (d+1) \log c$$

$$\boxed{\frac{d}{b} \leq \frac{\log a}{\log c} < \frac{d+1}{b}} \longrightarrow ①$$

<u>From Axiom 1</u> (monotonicity)

$$h(c^d) \leq h(a^b) < h(c^{d+1})$$

$$\therefore \quad d\,h(c) \leq b\,h(a) < (d+1)\,h(c).$$

$$\boxed{\therefore \quad \frac{d}{b} \leq \frac{h(a)}{h(c)} < \frac{d+1}{b}} \longrightarrow ②.$$

From ① & ② it follows that

$$\left| \frac{h(a)}{h(c)} - \frac{\log a}{\log c} \right| < \frac{1}{b}$$

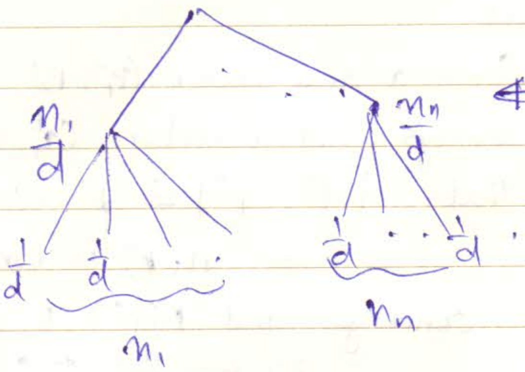
Since b is an arbitrary integer, it must be true that $\dfrac{h(a)}{h(c)} = \dfrac{\log a}{\log c}$

$$\Rightarrow \frac{h(a)}{\log a} = \lambda = \frac{h(c)}{\log c}$$

Next suppose $p_i = \frac{n_i}{d}$ (all prob are rational numbers & d is the common denominator)

$$p_1 \quad p_2, \quad \dots \quad p_n$$
$$\frac{n_1}{d} \quad \frac{n_2}{d} \quad \dots \quad \frac{n_n}{d}$$

$$\sum_{i=1}^{n} \frac{n_i}{d} = 1$$
$$\Rightarrow \sum n_i = d$$



interested in ENTROPY here

Suppose consider a source which has uniform probability across $\sum n_i = d$ events.

$$h(d) = \lambda \log d.$$
$$h(d) = h(p_1, \dots p_n) + \sum_{i=1} \frac{n_i}{d} h(n_i).$$
$$h(d) = h(p_1, \dots p_n) + \frac{\lambda}{d} \sum_i n_i \log(n_i).$$
$$\Rightarrow h(p_1, \dots p_n) = \lambda \sum_i \frac{n_i}{d} \log\left(\frac{d}{n_i}\right)$$
$$= \lambda \sum_i p_i \log\left(\frac{1}{p_i}\right).$$
$$= -\lambda \sum_i p_i \log p_i$$

Can absorb the constant λ into the base of logarithm & forget about it.

DATE

<u>Conclusion</u>: Can WLOG assume that for a source with rational probabilities, $\{\frac{n_i}{d}\}$

$$H\left(\left\{\frac{n_i}{d}\right\}_{i=1}^{n}\right) = -\sum \left(\frac{n_i}{d}\right) \log\left(\frac{n_i}{d}\right)$$

<u>Axiom 3</u>: $H(p_1, p_2, \dots p_n)$ is a continuos function in its arguments.

Given a real vector $(p_1^*, p_2^* \dots, p_n^*) = p^*$
can always find a vector $(q_1^i, q_2^i, \dots q_n^i) = \underline{q}^i$ where $q_j^i \in \mathbb{Q}$
suct that $|\underline{q}^i - p^*| \leq \frac{1}{i}$. $\Rightarrow q^i \longrightarrow p^*$
as $H(p)$ is continuos $H(q^i)$ ~~conv~~
converges and $H(p^*)$ is the limit of the sequenc

$p_i \leftarrow \frac{a_{ik}}{b_{ik}}$

$$\therefore H(p_1, \dots p_n) = H\left(\left\{\lim_{k\to\infty} \frac{a_{ik}}{b_{ik}}\right\}_{i=1}^{n}\right)$$

$$= \lim_{k\to\infty} H\left(\left\{\frac{a_{ik}}{b_{ik}}\right\}_{i=1}^{n}\right)$$

$$= \lim_{k\to\infty} -\sum_{i=1}^{n}\left(\frac{a_{ik}}{b_{ik}}\right)\log\left(\frac{a_{ik}}{b_{ik}}\right)$$

$$= -\sum_{i=1}^{n} p_i \log(p_i) \quad ///$$

Natura