# [ Information Theory ]

Transmission of Information – Ralph Heartely 1924

$H = n \log S$ (Heartley)

$S$ ← # of possible symbols

$n$ ← # of symbols in a transmission

Transmission of Information[1]

By R. V. L. HARTLEY

SYNOPSIS: A quantitative measure of "information" is developed which is based on physical as contrasted with psychological considerations. How the rate of transmission of this information over a system is limited by the

$$H = n \log s \tag{9}$$

$$= \log s^n. \tag{10}$$

What we have done then is to take as our practical measure of information the logarithm of the number of possible symbol sequences.

⇐ measure of average uncertainty = entropy = H = bit = rareness

= how many questions = Sending Info = surprisal

= randomness ∵= message space ∵= $n$ ∵= questions

If same probability

1 | 2 | 3 | 4 | 5 6 7 8

③ ② ①

$2^3 = 8$

$3 = \log_2 8$

$2^n = W$

$n = \log_2 W$

(unit) entropy = $\log_2$ message space

If different probability

1 2 3 4 5 6 7 8

$$H(P) = P_1 \times \log \frac{1}{P_1} + P_2 \times \log \frac{1}{P_2} + \cdots\cdots$$

$$= \sum_{i=1}^{n} P_i \times \log\left(\frac{1}{P_i}\right)$$

= $\Sigma$ unit probability × unit entropy

= $\propto \frac{1}{\text{probability}}$ $\propto$ rareness ∵ { rare Info = much Info gain

$$= -\sum p \times \log(P)$$

# [ redundancy ]

redundancy = 1 − relative entropy $\left( = \dfrac{\text{Real } H}{\text{Max } H} \right)$

ex) alphabet 27

$H_{alphabet} = \log_2 27 = 4.784$ (where P. per cha is same)

$= 4.08$ (where P. per cha is diff)

$\therefore$ redundancy $= 1 - \dfrac{4.08}{4.784}$

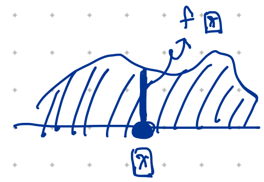
$= 1 - 0.848$

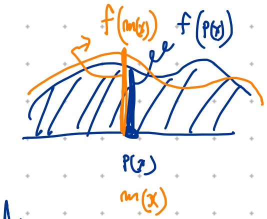
$= 0.152$

# [ Cross entropy ]

$$H(p) = \sum p \times \log\left(\frac{1}{p}\right) = -\sum p \log p$$

$$H(p,m) = \sum p \times \log\left(\frac{1}{m}\right) = -\sum p \log m$$




$$\mathbb{E} = \int x \times f(x)\, dx$$

$$H(X) = \mathbb{E}_x[I(x)] = -\sum_{x \in X} p(x) \cdot \log p(x)$$

$$H_b(p) = H(p) + H(1-p)$$
$$= p \times \log \frac{1}{p} + (1-p) \log\left(\frac{1}{1-p}\right)$$
$$= -p \times \log(p) - (1-p)\log(1-p)$$

$$H_b(p,m) = H(p,m) + H(1-p,m)$$
$$= p \times \log \frac{1}{m} + (1-p) \times \log \frac{1}{1-m}$$
$$= -p \log m - (1-p)\log(1-m)$$

$$H(p,m) = \underline{p}\ \underline{\log\left(\frac{1}{m}\right)}$$

$$\log \frac{1}{m} > \log \frac{1}{p}$$

arbitrary probability distribution — true probability distribution

$$D_{KL}(p(x) \,\|\, q(x)) = H(p,q) - H(p)$$

(= information gain
= information divergence
= relative entropy)

$$= -\sum p \log q - \left(-\sum p \log p\right)$$
$$= \sum p \log \frac{p}{q}$$