W01 ~ Information Theory

Claude Shannon (1912-2001)

Mt. Auburn

Betty Shannon.

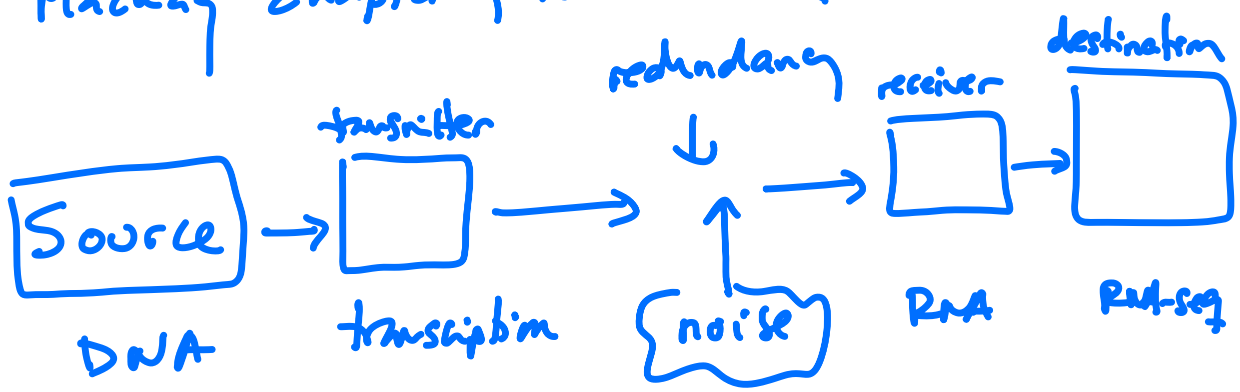
1948

"A mathematical theory of communication"

p1,2

section 6

MacKay chapter 2, lectures 1,2



How much information you need to send so that the message is reproduced faithfully?

Need to quantify Information

1 – I had breakfast this morning

2 – Today is NOT my birthday

3 – Today IS my birthday

4 – I ran the 1st Cambridge ½ (2016)

5 – I " 1st + 2nd " " (2017)

6 – I " 1+2+3 " " (2019)

(4th) 17 NOV 2021

7 – I've been to Antartica.

Source: O_1, O_2, $\vdots$, O_n $\rightarrow$ Information by love selecting O_4

Outcome / Event has prob p

Information of observing O?

(1) $I(O) \propto \frac{1}{p}$,, $p \uparrow$ $I \downarrow$

* The rarer the event, the lowest its probability, and the more info you would obtain by sampling it

* The more possibilities (assuming all equally likely) the more ignorant about outcome, the more info by observing one

$I \sim n$ (# of outcomes)

(2) Information should be additive

if O_1, O_2 are independents,

$$I(O_1, O_2) = I(O_1) + I(O_2)$$

→ Shannon proposed

$$\boxed{I(O) = \log \frac{1}{p} = -\log p}$$

In Shannon's words:

'useful',
'intuitive',
'suitable" } subjective!

$$o_1 \perp o_2$$

$$P(o_1, o_2) = P(o_1) \cdot P(o_2)$$

$$I(o_1, o_2) = -\log P(o_1, o_2) =$$

$$= -\log P(o_1) \cdot P(o_2)$$

$$= -\log P(o_1) - \log P(o_2)$$

$$= I(o_1) + I(o_2)$$

$$-p_i^2$$

$$\log(a \cdot b) = \log a + \log b$$

$$\log(a/b) = \log a - \log b$$

$$\log(a^n) = n \log a$$

The log base

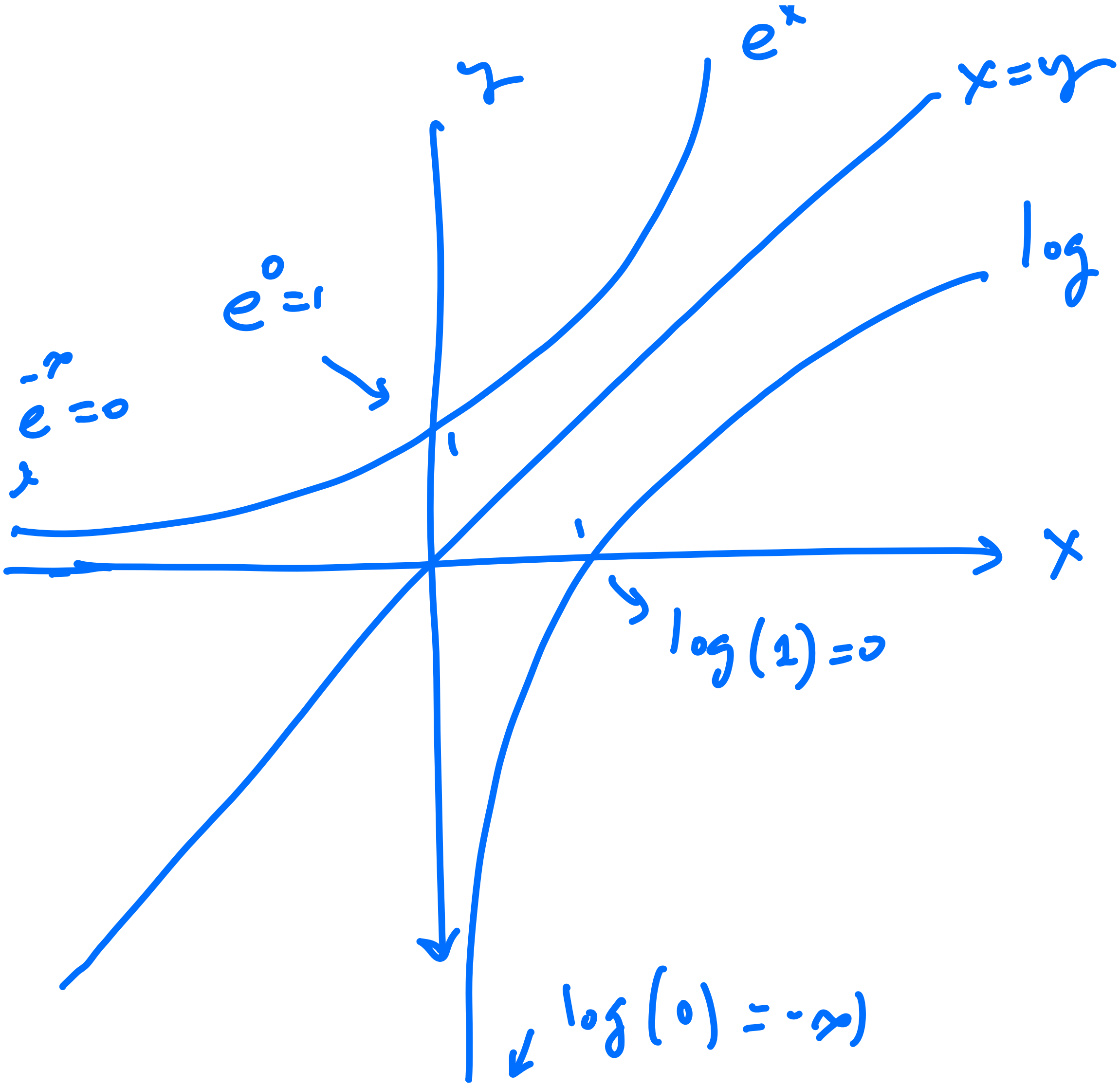
(natural logs) $\log a = b \Rightarrow a = e^{b}$

(nats) $\log_{10} a = b_{10} \Rightarrow a = 10^{b_{10}}$

(bits) $\log_2 a = b_2 \Rightarrow a = 2^{b_2}$

$$e^{b} = 2^{b_2} \Rightarrow b = \log(2^{b_2})$$
$$= b_2 \log 2$$

$$b = b_2 \log 2$$



$$\frac{d}{dx} e^{\alpha x} = \alpha e^{\alpha x}$$

$$\frac{d}{dx} \log(ax) = \frac{1}{ax} \times a = \frac{1}{x}$$

$$I(o, p) = -\log p$$

	p	I	bits
I had breakfast	1	$\log(1)$	0
My birthday	$\frac{364}{365}$	$\log\frac{364}{365}$	2.6
Not my birthday	$\frac{1}{365}$	$-\log 365$	8.5
1st Cambridge 1/2	$\frac{4500}{4.7\cdot 10^3}$		10
1+2 " "	$\frac{4500}{4.7\cdot 10^3}\ \frac{6500}{4.7\cdot 10^3}$		13.5

Entropy

Average Information of a prob distribution.

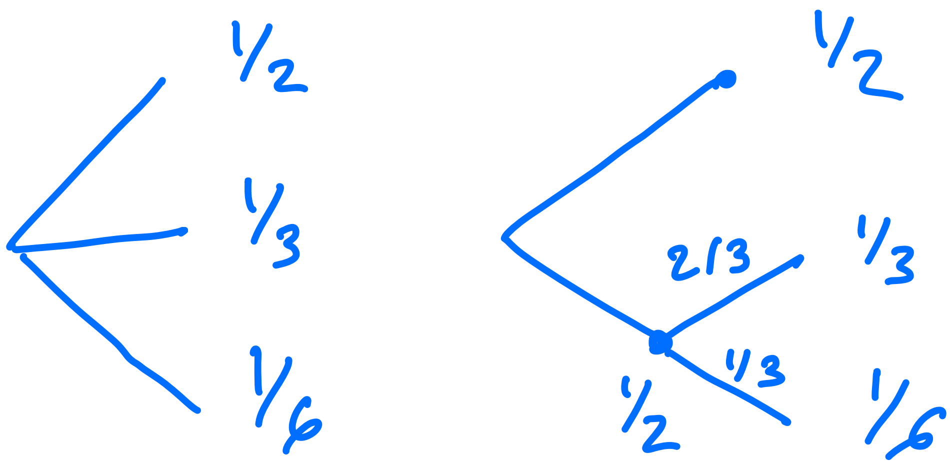
$a \in X \qquad I(a) = -\log p_X(a)$

$$H(X) = \int_a p_X(a) \cdot \log \frac{1}{p_X(a)} \, da$$

$$= \sum_i p_i \log \frac{1}{p_i} = -\sum_i p_i \log p_i$$

$H(x) \geq 0 \quad " \quad 0 \Longleftrightarrow p_i = 0$ except for 1 value

Important property



$$H\left(\tfrac{1}{2}, \tfrac{1}{3}, \tfrac{1}{6}\right) = H\left(\tfrac{1}{2}, \tfrac{1}{2}\right) + \frac{1}{2} H\left(\tfrac{2}{3}, \tfrac{1}{3}\right)$$

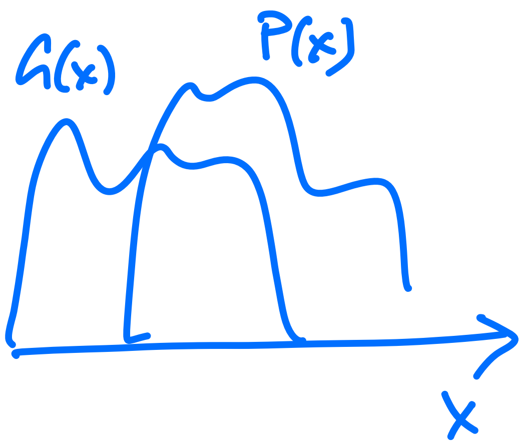
The composition Law

Same Information no matter how the choices are broken

Relative entropy

the Kullback-Liebler divergence

To compare 2 probability distributions on X



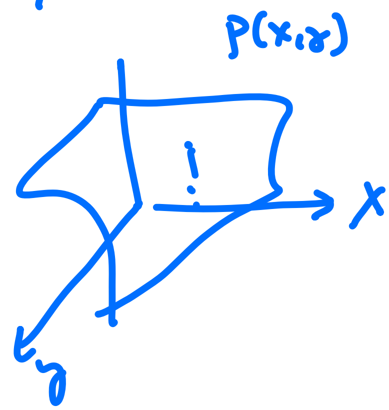
$$D_{KL}(P \| Q) = \int_X P(x) \log \frac{P(x)}{Q(x)}$$

i) $D_{KL}(P \| Q) \neq D_{KL}(Q \| P)$

ii) $D_{KL} \geq 0 \qquad = 0 \iff Q = P$

Mutual Information

two random variables X, Y



P_{XY}

$\left.\begin{array}{l} P_X \\ P_Y \end{array}\right\}$ marginals

$$P_X(x) = \int_y p(x,y)\,dy$$

$$P_Y(y) = \int_x p(x,y)\,dx$$

$$MI(X,Y) = D_{KL}(P_{XY} \| P_X P_Y) = \int_{xy} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}\,dx\,dy$$

$MI \geq 0$

if $X \perp Y$ $\quad p(x,y) = P_X(x) P_Y(y)$

$MI = 0$