W01 ~ Information theory

+ "A mathematical theory of communication"

Claude Shannon, 1948

(1912-2001)

Mt. Auburn Cementery

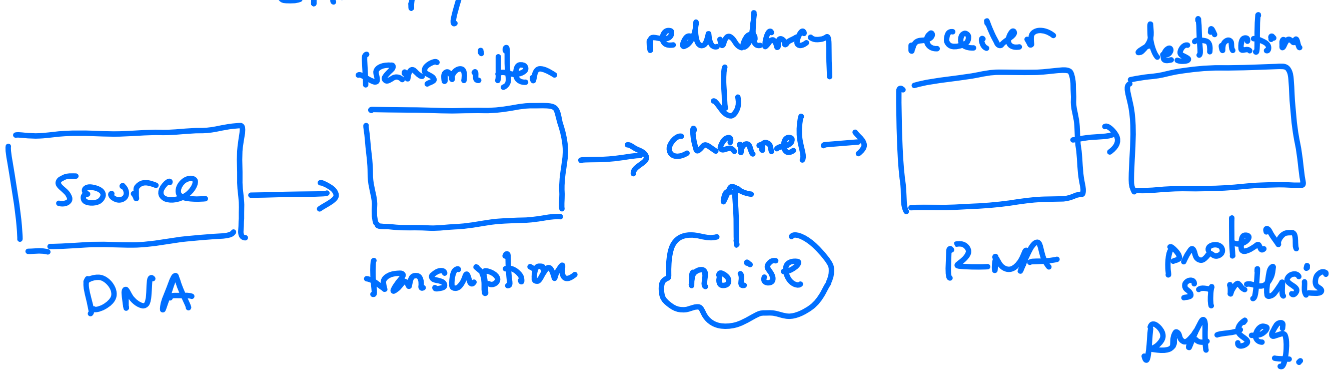
Betty Shannon

p1,2

section 6

+ MacKay chapter 2, lectures 1 an 2

"Entropy or Shannon entropy"



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

Need to quantify "information"

1. Today I got dressed
2. Today is not my birthday
3. Today it's my birthday

4. Today I rescued a lost turtle

To quantify information:

"the significant aspect is that the actual message is one selected from a set of possible messages"

Source $\begin{cases} o_1 \\ o_2 \\ \vdots \\ o_n \end{cases}$ → information by selecting o_4.

.

+ What is your major?

  biology

  CS

  math

  pysics

  history **

  no major * * *

+ How many languajes you speak?

  1

  2

  3

  ⋮

  n * *

+

Outcomes/events have probabilities

$$\left\{ \begin{matrix} O_1 & \dots & O_n \\ P_1 & \dots & P_n \end{matrix} \right\} \quad \sum_{i=1}^{n} P_i = 1$$

Information of observing O_i:

i) $I(o) \propto \frac{1}{p}$ $\quad p \uparrow \quad I \downarrow$

* the rarest the event, the lowest it's probability, and the more information you obtain by "seeing" it
* the more possible outcomes (assume all eq. likely) the more ignorant about outcome, and the more information by observing on $I \propto n$ (# of outcomes)

ii) Information should be additive

if o_1 and o_2 are independent,

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

Shannon proposed Information content

$$I(o) = \log \frac{1}{p_o} = -\log p_o$$

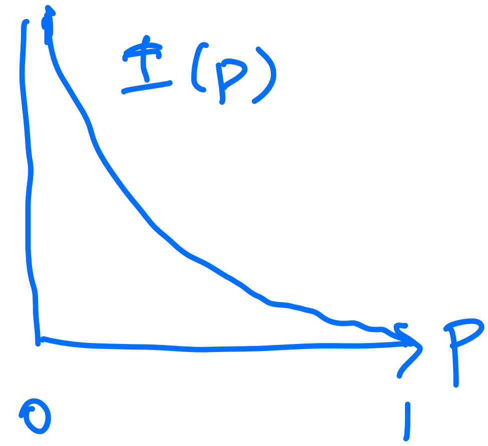
In shannon words:

'useful'
'intuitive'
'suitable'

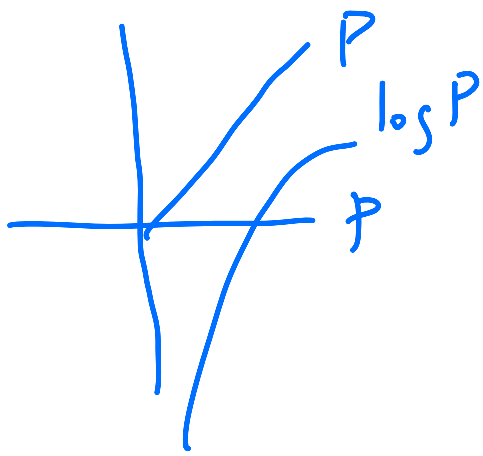
} subjective!

$(-p^2)$ other proposed options

i) $I = -\log P$



$P\uparrow \quad \log P\uparrow \quad I\downarrow$



ii) $O_1 \perp O_2$

$$P(O_1, O_2) = P(O_1)\cdot P(O_2)$$

$$I(O_1, O_2) = -\log P(O_1, O_2)$$
$$= -\log\left[P(O_1)\cdot P(O_2)\right]$$
$$= -\log P(O_1) - \log P(O_2)$$
$$= I(O_1) + I(O_2)$$

little math refresher

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

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

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

the base is arbitrary

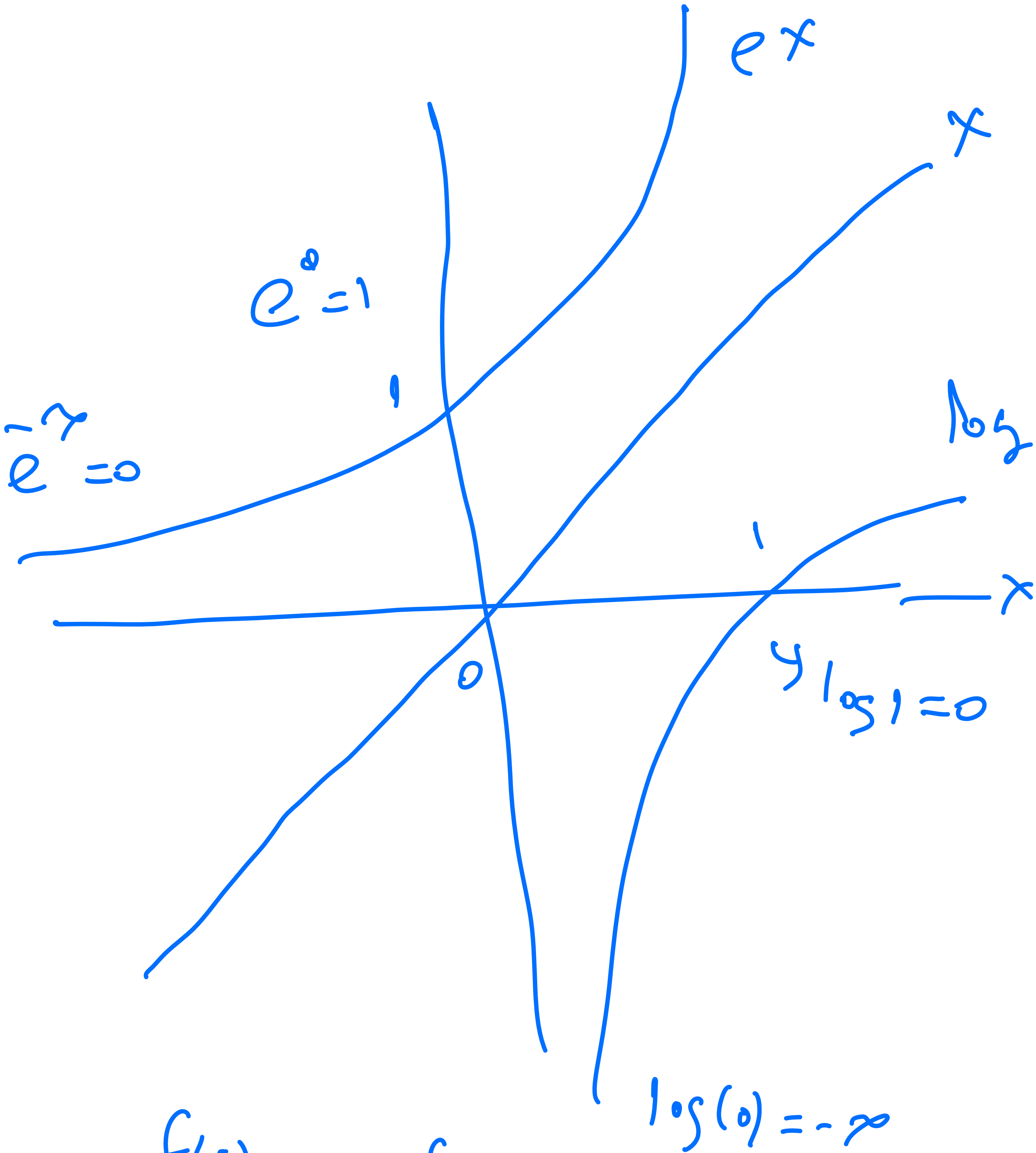
(natural logs) $\log a = b \qquad a = e^{b}$

(nats

bits $\qquad \log_{10} a = b_{10} \qquad a = 10^{b_{10}}$

$\log_2 a = b_2 \qquad a = 2^{b_2}$

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



$$\frac{d}{dx} e^{f(x)} = \frac{df}{dx} e^{f(x)}$$

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

$$\frac{d}{dx} \log[f(x)] = \frac{1}{f(x)} \frac{df}{dx}$$

Quantify Information $\doteq$ Shannon Information content

	P	I	bits
I got dressed today	1	$\log_2(1)$	0
Not my birthday	$\frac{364}{365}$	$\log_2(\ )$	0.04
if it's my birthday	$\frac{1}{365}$ 0.0027	$-\log_2(365)$	8.5
Speaking more than 4 languages (3%)	$p=0.03$	$\log_2 \frac{1}{0.03}$	5.6
being a nobel laureate ~1000	$p = \frac{1000}{8\cdot 10^9} = 1.25\cdot 10^{-7}$		22.9

Entropy average shannon inf content of a prob. distribution

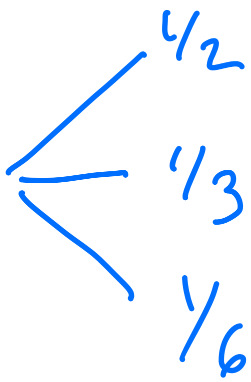
$a \in X \to I(a) = \log \frac{1}{P_X(a)}$

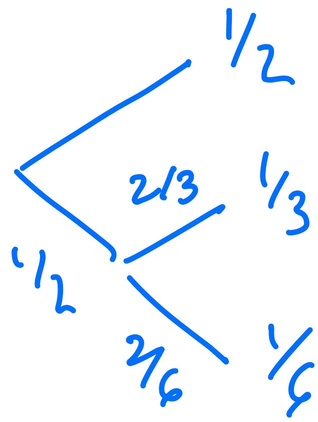
$$H(X) = \sum_i P_i \log \frac{1}{P_i} = -\sum_i P_i \log P_i$$
discrete

$$H(X) = \int_X P_X(x) \log \frac{1}{P_X(x)} dx$$
cont

$H(X) \geq 0$, $H(X) = 0 \iff P_i = 0$ except for 1 value

Important property





$$H(1/2\ 1/3\ 1/6) = H(1/2, 1/2) + \\ + \frac{1}{2} H(2/3, 1/3)$$

The composition law

Same information no matter how the choices are broken down.

$$H(1/2\ 1/3\ 1/6) = \frac{1}{2}\log 2 + \frac{1}{3}\log 3 + \frac{1}{6}\log 6.$$

$$\frac{1}{2}\log 2 + \frac{1}{2}\log 2 + \frac{1}{2}\times\frac{2}{3}\log\frac{3}{2} + \frac{1}{2}\times\frac{1}{3}\log 3$$

$$= \frac{1}{2}\log 2 + \frac{1}{2}\log 2 + \frac{1}{3}\log\frac{3}{2} + \frac{1}{6}\log 3$$

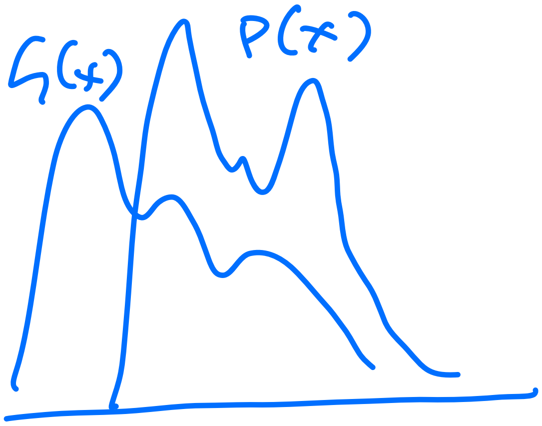
$$= \frac{1}{2}\log 2 + \frac{1}{3}\log 3 + \frac{1}{2}\log 2 - \frac{1}{3}\log 2 + \frac{1}{6}\log 3$$

$$= \frac{1}{2}\log 2 + \frac{1}{3}\log 3 + \underbrace{\frac{1}{6}\log 2 + \frac{1}{6}\log 3}_{\frac{1}{6}\log(2\cdot 3)}$$

Relative Entropy

The Kullback–Liebler divergence

to compare 2 probability distributions



$$D_{KL}(P\|q) = \int_x P(x) \log \frac{P(x)}{q(x)}$$

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

ii) $D_{KL} \geq 0$ $\quad D_{KL} = 0 \iff q = P$

Mutual Information

two random variables X, Y

P_{XY}

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

$$P_X(a) = \int_y P(a, y)\, dy$$

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

$$MI(X, Y) = D_{KL}(P_{XY} \| P_X P_Y)$$

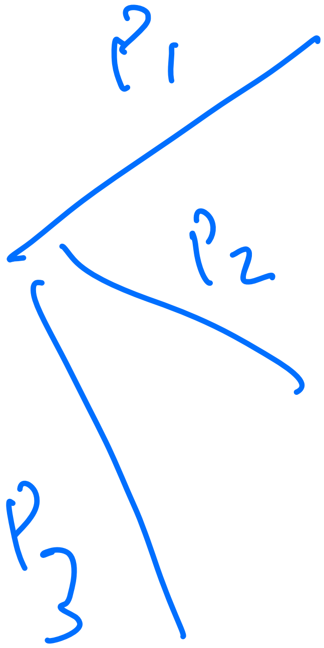
$$= \int_{xy} P(x, y) \log \frac{P(x, y)}{P(x) P(y)}\, dy$$

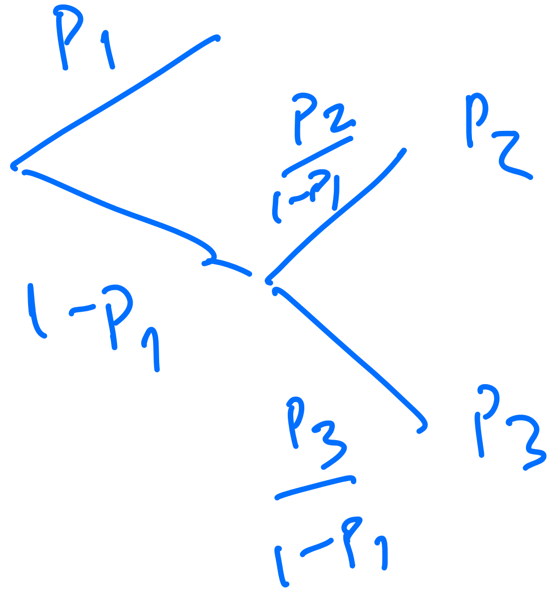
$MI \geq 0$

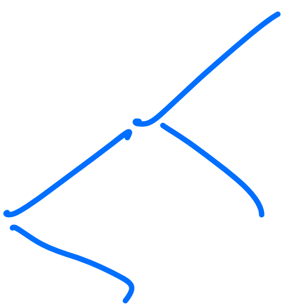
$X \perp Y \quad P(X, Y) = P(X)P(Y)$

$MI = 0$

$$H(p_1, p_2, p_3) = H(p_1, 1-p_1) + (1-p_1)\,H\left(\frac{p_2}{1-p_1}, \frac{p_3}{1-p_1}\right)$$





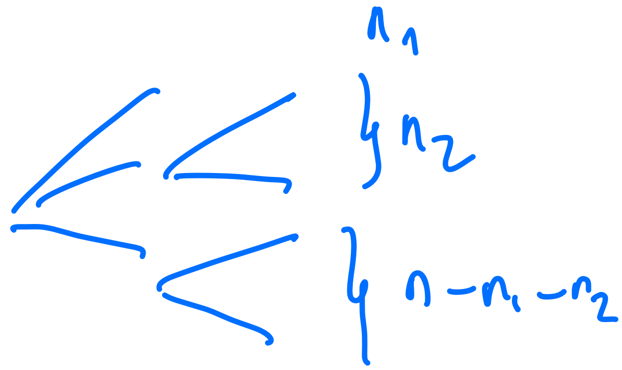


p_i

$(p_1, 1-p)$

$(p_2$


p_1
p_n


n_1
n_2
$n-n_1-n_2$