INFO THEORY

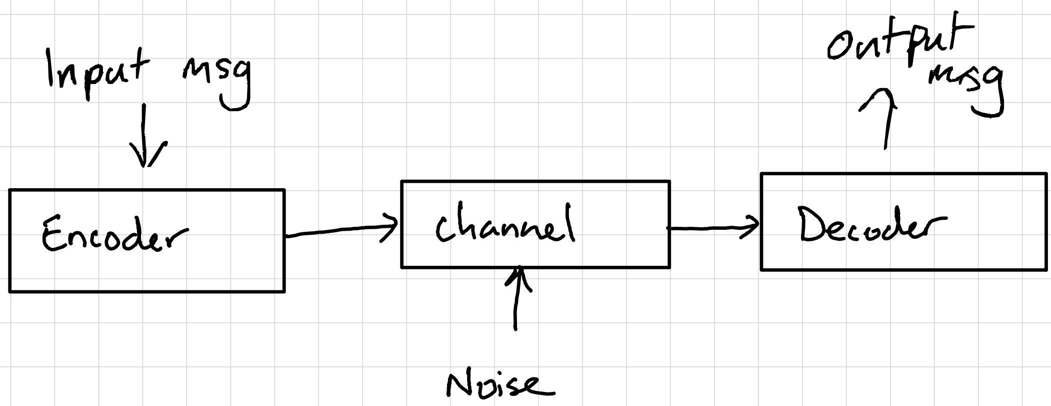
Lectures

There are two in-depth proofs to study

- These will be pre-recorded. to allow you to really focus on them offline.

- L3 Virtual November 12
- L9 Virtual November 26

Information Theory



Mostly our goal is output = input

Info Theory concerns itself with:

- Efficiency ← Noiseless channels (Compression) — Strip redundancy
- Reliability/Quality ← Noisy channels — Put redundancy back in

"Channels"

Info Theory talks of channels over which information flows

Flowing over space

- → WiFi
- → Audio
- → Post

Flowing over time

- → APD
- → Letter

What is information?

1 2 3 4 5 6 7 8

Q1	≤ 4	✓	Q1′	$=4$	✓
Q2	≤ 2	✗			
Q3	$=4$	✓			

Q1′ = Q3 Same literal question, same answer... but different information ??!

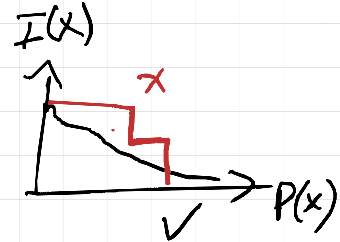
Core Intuition : $\boxed{\text{Information links to SURPRISE}}$

↳ It is a probabilistic entity

↳ $I(X=x) = f(P(X=x))$

Unknown function

Shannon's Desiderata



i) Continuity — As $P(X=x)$ varies smoothly so must $I(X=x)$

ii) Symmetry — Information from two events, X and Y $I(X=x; Y=y)$ must not depend on the order they occur.

iii) Additive — $I(Q1') = I(Q_1) + I(Q_2) + I(Q_3)$

What form?

$$I(X=x, Y=y) = f(P(X=x, Y=y))$$

$$= f(P(X=x)P(Y=y)) \quad \text{when } X, Y \text{ are independent.}$$

$$= I(X=x) + I(Y=y) \quad \text{since additive.}$$

$$= f(P(X=x)) + f(P(Y=y))$$

i.e. $f(P(X=x)P(Y=y)) = f(P(X=x)) + f(P(Y=y))$

This is a property of Logarithms

$$\Rightarrow \quad I(X=x) = k \log(g(P(X=x)))$$

Shannon Information

$$\boxed{h(x) = -\log_2 P(x) = \log_2 \frac{1}{P(x)}}$$

Game of 8

Q1 ≤ 4 $\qquad I_1 = -\log\left(\frac{1}{2}\right) = 1$ bit

Q2 ≤ 2 $\qquad I_2 = -\log\left(\frac{1}{2}\right) = 1$ bit

Q3 $= 4$ $\qquad I_3 = -\log\left(\frac{1}{2}\right) = 1$ bit

$\underline{\overline{3 \text{ bits}}}$

Q1' $= 4$ $\qquad I_{1'} = -\log \frac{1}{2^3} = \underline{3 \text{ bits}}$

When bits aren't bits

Q_1: $P(4) = \frac{1}{8}$ $I(4) = 3$ bits

Q_1: $P(4) = \frac{1}{2}$ $I(4) = 1$ bit

Q_1: $P(4) = \frac{1}{32}$ $I(4) = 5$ bits

Binary Digit (bit) is not the same as info content.

Shannon bits.

Towards Entropy

Information of events ⟵ Past tense

$$h(x) = -\log P(x)$$

Possible info we could get ⟵ Future tense

↳ Expected value

$$\boxed{H(X) = -\sum_{x \in A_X} P(x) \log_2 P(x)}$$

"Entropy"

Alphabet A_X

R.V. X

Event x

Example Game of 8, equiprob. $\Rightarrow H(X) = 8 \times \sum \frac{1}{8} \log 8 = \underline{\underline{3 \text{ bits}}}$

What is entropy?

More strictly $\boxed{\text{Discrete Entropy}}$

$$\boxed{\text{Entropy quantifies the average level of uncertainty in a system}}$$

Measure of disorder.

Mathematically the defn is ave. info. but Entropy is not itself ave information.

Information resolves entropy.

Principle of max. Entropy (PoME)

$$\boxed{\text{The prob. dist. that best represents the current state of a system is the one that maximises the entropy}}$$

"Maximum ignorance".

This is useful when problems are underconstrained...

But first, some Maths

· We need to maximise entropy under constraints, which you may not have seen before!

E.g. 1 Maximise $f(x,y) = x + y^2$ s.t. $x + y = 5$

$$= x + (5-x)^2 = x^2 - 9x + 25$$

$$\frac{df}{dx} = 2x - 9 = 0 \Rightarrow \boxed{\begin{aligned} x &= 4.5 \\ y &= 0.5 \end{aligned}}$$

E.g. 2 Max. $f(x,y,z) = x + y^2 - z^3$ s.t. $x+y-z = 3$

Errr....

Lagrangian Multiplier

When you want to maximise subject to constraints, encode the constraint into a modified equation (the Lagrangian) s.t. it forms a penalty if it is not satisfied

$$\boxed{L(x, y, \dots) = f(x, y, \dots) - \sum \lambda_i (\text{constraint}_i)}$$

E.g. $f(x, y, z) = x + y^2 - z^3$ s.t. $x + y - z = 3$

$$L(x, y, z) = f(x, y, z) - \lambda \underbrace{(x + y - z - 3)}$$

Lagrange multiplier (a constant we don't know)

Constraint encoded so it has no effect when satisfied

Now proceed as normal on $L(x,y,z)$ rather then $f(x,y,z)$

$$L(x,y,z) = x + y^2 - z^3 - \lambda(x + y - z - 3)$$

$$\left.\begin{aligned} \frac{\partial L}{\partial x} &= 1 - \lambda = 0 \\ \frac{\partial L}{\partial y} &= 2y - \lambda = 0 \\ \frac{\partial L}{\partial z} &= -3z^2 + \lambda = 0 \end{aligned}\right\}$$

4 equations, 4 unknowns

$\lambda = 1$

$y = 0.5$

$z = \frac{1}{\sqrt{3}}$

$x = 2.5 + \frac{1}{\sqrt{3}}$

$$\frac{\partial L}{\partial \lambda} = x + y - z = 3$$

x

Example

You sell pies and charge differently for different groups

Children – £1
Students – £2
Faculty – £3

You take £$\frac{17}{7}$ on average but want to know the prob's of each group: P_c, P_s, P_F

We know

$$\left.\begin{aligned} P_c + P_s + P_F &= 1 \\ P_c + 2P_s + 3P_F &= \frac{17}{7} \end{aligned}\right\} \begin{array}{l} 3 \text{ unknowns} \\ 2 \text{ constraints} \\ \quad\Downarrow \\ \text{underconstrained} \end{array}$$

PoME $\Rightarrow$ $H(x) = -P_c \log P_c - P_s \log P_s - P_F \log P_F$

Max. this, subject to constraints above

So:

$$L(P_i) = \underbrace{-P_1 \log P_1 - P_2 \log P_2 + P_3 \log P_3}_{\text{Entropy we want to maximise}} - \underbrace{\lambda_1\left(\sum P_i - 1\right)}_{\text{constraint 1}} - \underbrace{\lambda_2\left(\sum x_i P_i - \frac{17}{7}\right)}_{\text{constraint 2}}$$

$$\boxed{\begin{array}{c}\text{When maximising entropy under a constraint of average value} \\ P_i \sim \dfrac{e^{x_i \lambda}}{\sum e^{x_i \lambda}} = \dfrac{e^{x_i \lambda}}{Z}\end{array}}$$

You prove this in the examples sheet.

$$\Rightarrow \quad P_1 = \frac{e^{\lambda}}{Z} = \frac{\alpha}{Z} \qquad \text{where } \alpha = e^{\lambda}$$

$$Z = e^{\lambda} + e^{2\lambda} + e^{3\lambda} = \alpha + \alpha^2 + \alpha^3$$

$$P_2 = \frac{\alpha^2}{7}$$

$$P_3 = \frac{\alpha^3}{Z}$$

$$\therefore \quad P_1 + 2P_2 + 3P_3 = \frac{\alpha + 2\alpha^2 + 3\alpha^3}{Z} = \frac{17}{7}$$

$$\Rightarrow \quad 7\alpha + 14\alpha^2 + 3\alpha^3 = 17\alpha + 17\alpha^2 + 17\alpha^3$$

Collect terms $\Rightarrow$ $\quad x(4\alpha^2 - 3\alpha - 10) = 0$

Solve $\Rightarrow$ $\quad \alpha = 0$ trivial, meaningless

$$\alpha = \frac{3 \pm \sqrt{9 + 160}}{8} = \frac{3 \pm 13}{8} = 2$$

($\alpha = e^{\lambda}$ so must be > 0)

So:

$$Z = 2 + 2^2 + 2^3 = 14.$$

$$P_1 = \frac{\alpha}{Z} = \frac{2}{14} = 1/7$$

$$P_2 = \frac{\alpha^2}{Z} = \frac{4}{14} = \frac{2}{7}$$

$$P_3 = \frac{\alpha^3}{Z} = \frac{8}{14} = \frac{4}{7}$$

Best $\{P_i\}$ given the info we have. It retains the most uncertainty so means we haven't made assumptions