Reading — Chapter 2 of Cover & Thomas
Opt — Shannon's original Papers

Entropy

$$S(A) = -\sum_x P_A(x) \log P_A(x)$$

Clausius (1855) $dS = \frac{\delta Q}{T}$

Boltzmann (1875) $S = \ln W$

Plank (1900) $S = k_B \ln W$

Gibbs (1902) $S = -\sum P_A(x) \ln P(x)$

Shannon (1948) $H = -\sum p(x) \log p(x)$

E.T. Jaynes (1956)

Landuer (1961)

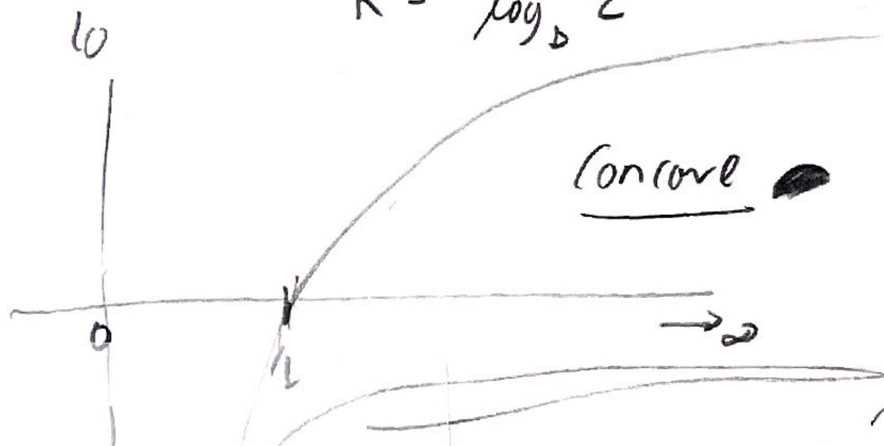
$\log_b x = y \qquad b^y = x$

$\log_b (xy) = \log_b x + \log_b y$

$\log_b x^n = n \log_b x$

$\log_b x = K_b \log_c x$

$K = \log_b c$



$0 \ln 0 = 0 \quad 0^0 = 1$

Knuth 1992

(2)

$$S(A) = -\sum_x P_A(x) \log_b P_A(x)$$

$$0 \leq S(A) \leq \log W$$

Units b

Base 2	bits		lg
Base e	nats		ln
Base 10	bans	decibans $= \frac{1}{10}$ ban	log

Conversion

$$2^{10} \approx e^7 \approx 10^3$$

$$1024 \approx 1096.6\ldots \approx 1000$$

10 bits $\approx$ 7 nats $\approx$ 3 bans

1 nat = 0.7 bits

1 deciban = $\frac{1}{3}$ bit

Bit – fundamental indivisible unit of information

(examples – any physical system with 2 stable states)

example 10 sided Dice

Roll 6 times 2 7 1 7 3 6

10 Digits 10 bans

10^6 possibility

$$S = \log_2 10^N \approx \log_2 2^{20}$$

$N = 6$

≈ 20 bits

6 digit number

$$P(\text{Prime}) \approx \frac{1}{\ln 10^6} \approx \frac{1}{\ln e^{14}} \approx \frac{1}{14}$$

1 byte = 8 bits

Kilo byte = 10^3 bytes $\approx 2^{10}$ bytes Kibi 2^{10}

~~Mega~~ 10^6 $\approx 2^{20}$ bytes mebi 2^{20}

giga 10^9 $\approx 2^{30}$ bytes gibi 2^{30}

Tera byte 10^{12} $\approx 2^{40}$ bytes Tebi 2^{40}

10^{15} 10^{18} 10^{21}

[story about hard drive nomenclature pebi

| ChatGPT 45 Tera $\times 10^{12}$ bytes $\approx 4 \times 10^{14}$ bits
| $45 \times 2^{40} \times 8$ bits $= 2^{48}$ bits

Peta 10^{15} pebi 2^{60}

~~Exa~~ 10^{18} exbi 2^{70}

Zetta 10^{21} zebi 2^{80}

STORAGE

~100 Zetta bytes

$\sim 100 \times 10^{21} \times 8$ bits

8×10^{22} bits

Doubling every ~4 years

A mol of bits within 10 years

UNITS

$$S = - \cancel{K_B} \sum_x P(x) \ln P(x)$$

natural log

Boltzmann's constant

Entropy of water

$= 16.8$ cal K^{-1} mol^{-1}

4.2 J per cal

$= 70.6$ J K^{-1} mol^{-1}

Plank 1900

$K_B = 1.38 \times 10^{-23}$ J K^{-1} [per nat]

(ROMAN)

nats

$$\frac{\partial^n S}{\partial E} = \frac{1}{K_B T}$$ ← Kelvin

Joules

At Thermodynamic equilibrium

$$\beta = \frac{1}{K_B T} \text{ nats/Joule}$$

$S_{H_2O} = 70.6$ J K^{-1} mol^{-1}

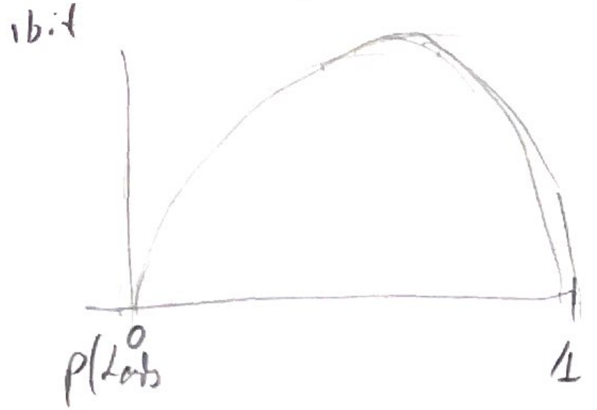
~~70.6 × 6 × 10^23~~

$$\frac{70.6 \;\cancel{\times 1.38 \times 10^{-23}}}{(6 \times 10^{23})(1.38 \times 10^{-23})} = 8.5 \text{ nats/molecule}$$

≈ 6 bits/molecule

[Only I am pendantic about This.]

$S = -\sum P(x) \log P(x)$



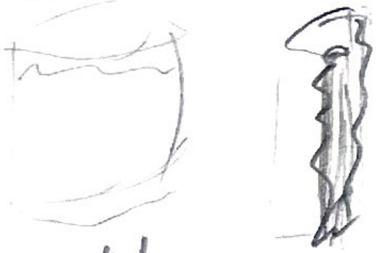
What is Entropy?

Dispersion, Information? Disorder?

→ Entropy is the solution to many problems.
Same thing. But probabilities change, names change

Bit is overloaded.

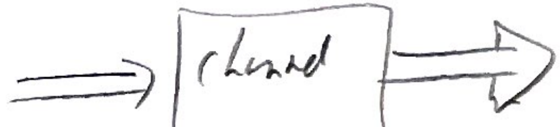
Both unit of entropy
And physical thing
(Bucket or water)



Smaller bucket

Can't have half a bucket. Can have half a bucket of water.

Shannon

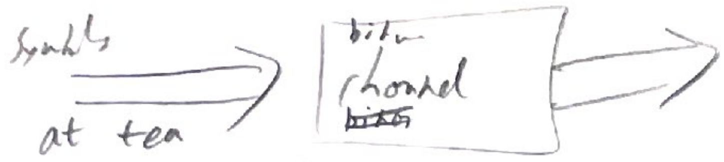


	P
_	1/2
e	1/4
t	1/8
a	1/8

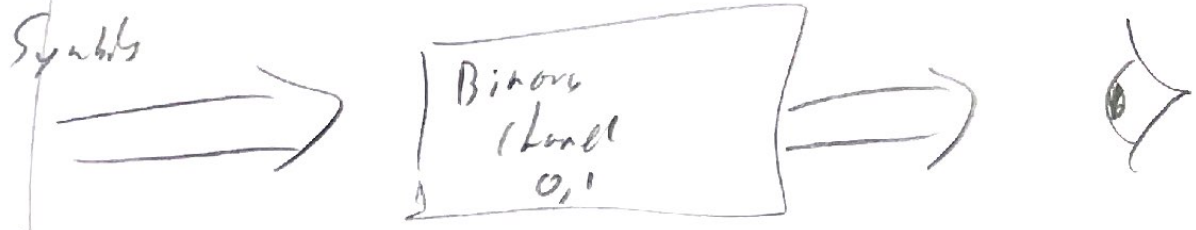
$S = \sum_n P \log P =$

5b)

~~Coder~~ Communication



	P	code	ave. bits
␣	1/2	0	1/2
e	1/4	10	1/2
t	1/8	110	3/8
a	1/8	111	3/8
			1.75 bits



$\underbrace{111}_{a}\ \underbrace{110}_{t}\ \underbrace{0}_{␣}\ \underbrace{110}_{t}\ \underbrace{10}_{e}\ \underbrace{111}_{a}$

at tea !

15 bits ÷ 6 characters

$\frac{15}{6} = 2\frac{1}{2}$ bits/character

$S = \frac{1}{2} + \frac{1}{2} + \frac{6}{8} = 1.75$ bits

$S(A) = -\sum P_A(x) \log P_A(x) \leftarrow$

$s_A(x) = -\log P_A(x)$ ← point entropy

(still function of ensemble, not system)

((Can talk about the entropy of a single realization!)

Break

Lecture 2b

Entropies

Joint $P_{AB}(a,b)$ $\quad S(A,B) = -\sum_{a,b} P(a,b) \log P(a,b)$

$S(A) = -\sum_{a} P(a) \log P(a) \quad P(a) = \sum_{b} P(a,b)$

!

Conditional $\quad S(A|B) = -\sum_{a,b} P(a,b) \log P(a|b)$

$= -\sum P(a,b) \log \frac{P(a,b)}{P(b)}$

$= S(A,B) - S(B)$

$$\boxed{S(A) \geq S(A|B)}$$

Chain Rule

$\cancel{\geq 0}$ (but not quantum)

~~Conditioning Reduces Entropy~~ $\cancel{S(A|B) \leq S(A)}$

$\cancel{S(A,B) \leq S(B) - S(A) \leq 0}$ ← later)

Mutual Information

[Note Sign]

$$I(A:B) = \sum P(a,b) \log \frac{P(a,b)}{P(a)\,P(b)} \geq 0$$

(note colon rather than semicolon)

$$I(A:B) = S(A) + S(B) - S(A,B)$$

(Implies $S(A,B) \leq S(A) + S(B)$

equality for independence)

⟹ Coin example where is the information.

	Coin	Head
1/2	tails	tail
1/2	heads	heads

$S(C) = 1 \quad S(H) = 1$

$S(C,H) = 1$

$I(C,H) = 1 + 1 - 1 = 1$ bit

$$I(A:B) = S(A) - S(A|B) \geq 0$$

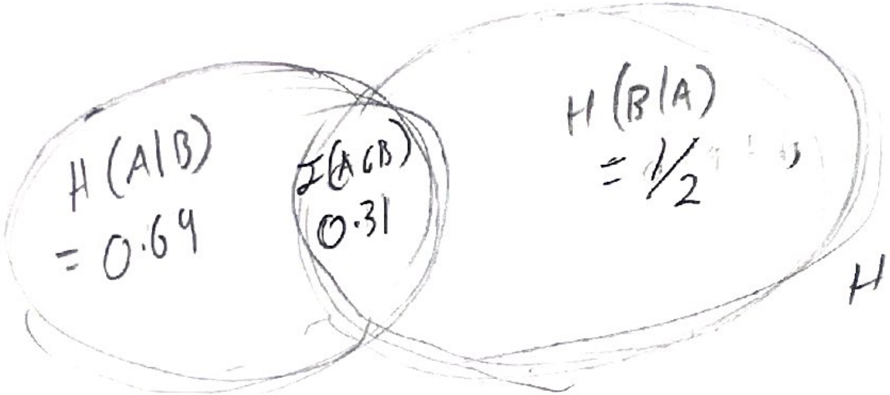
$$S(A) \geq S(A|B)$$

Conditioning reduces entropy

[But not quantum!]

Information Diagrams

	B	$3/4$	$1/4$
A	$\frac{1}{2}$	$\frac{1}{4}$	$1/4$
	$\frac{1}{2}$	$1/2$	0



$H(A) = 1$

$H(A,B) = 1\frac{1}{2}$

$H(B) =$ [illegible] ?

$$= \underbrace{-\tfrac{1}{4}\log\tfrac{1}{4}}_{1/2} \underbrace{- \tfrac{3}{4}\log\tfrac{3}{4}}_{0.31}$$

$= 0.81$ bits

3 Variables

a	b	c	
0	0	0	1/4
0	1	1	1/4
1	0	1	1/4
1	1	0	1/4

$H(A,B,C) = 2$

$H(A) = 1$

$H(B) = 1$ $\quad H(A,B) = 2$

$H(C) = 1$

$I(A:B) = 0$

$I(A:C) = 0$

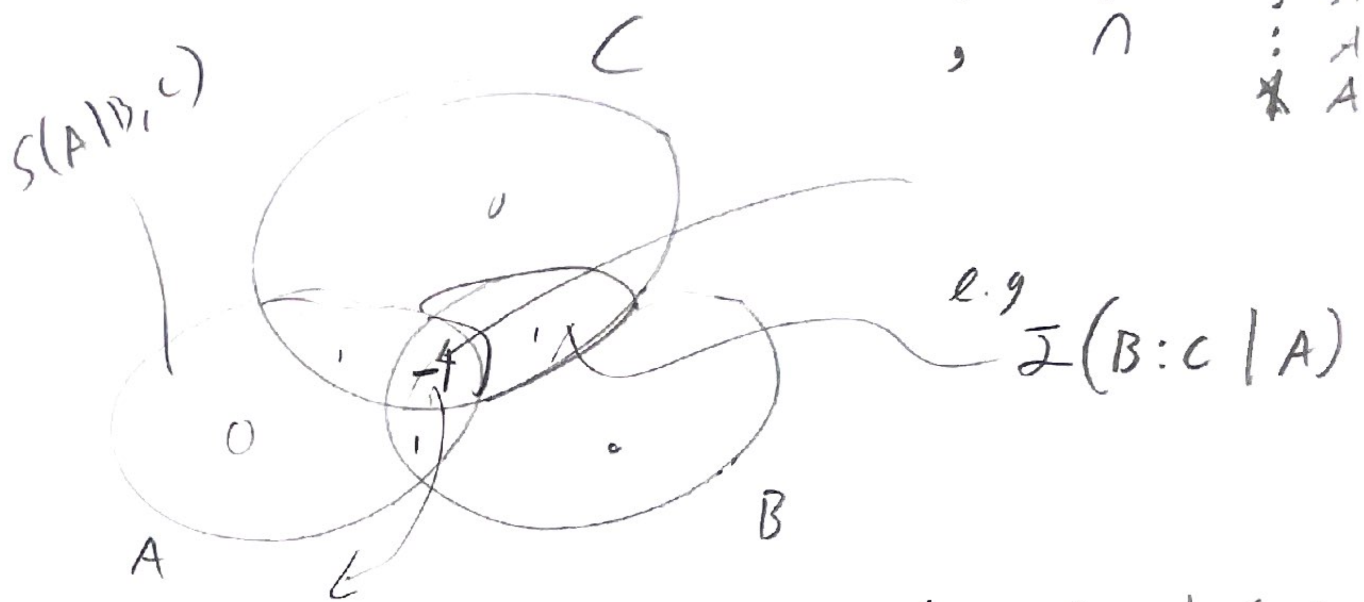
$I(B:C) = 0$

$I(B:C|A) = 1$

$, \quad \cup \quad , \quad A \cup B$

$; \quad \cap \quad : \quad A \cap B$

$\setminus \quad A \setminus B$



$$I(A:B:C) = \sum_{abc} P(a,b,c) \ln \frac{P(A,B)P(B,C)P(A,C)}{P(A,B,C)P(A)P(B)P(C)}$$

$$= +S(A,B,C) - S(A,B) - S(B,C) - S(A,C) + S(A) + S(B) + S(C)$$

$$= -2 + 2 - 2 - 2 + 1 + 1 + 1$$

$$= -1 \text{ bit} \quad !$$