Interpretation of entropy: Shannon source coding theorem

surprisal: $-k_B \ln p_n(t)$: associated w/ state n, but depends on whole ensemble

entropy = avg. surprisal $S(t) = -k_B \sum_n p_n(t) \ln p_n(t)$

Relate to Shannon's source coding theorem [1948]:

imagine scenario: Alice (A) is a prof.
Bob (B) is a grad. student

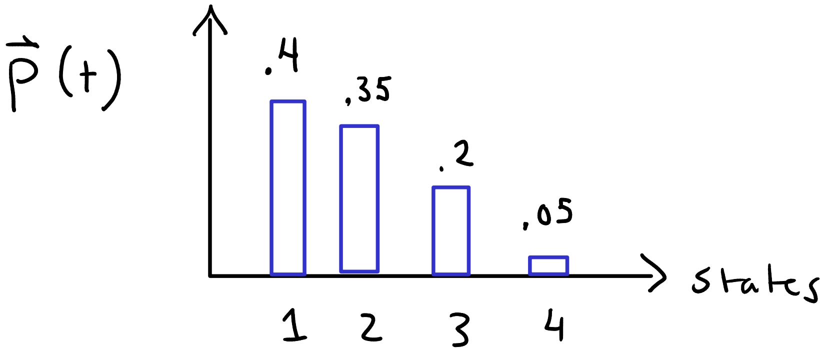
experiment:
- a system ($N=4$ states)

repeat it many times
- B prepares it in initial distrib. $\vec{p}(0)$
- B lets it evolve for a time t
- B takes measurement, records state

output: series of measurements
i.e. 2, 4, 1, 3, 2, ...

A knows: system trans. matrix W
prob $\vec{p}(0)$
hence $\vec{p}(t) = W^t \vec{p}(0)$
+ she tells B what $\vec{p}(t)$

A doesn't know B's measurement (specific sequence) + she asks for it to be sent



case I: Bob is lazy (or dumb?)

state	code
1	00
2	01
3	10
4	11

easy to read

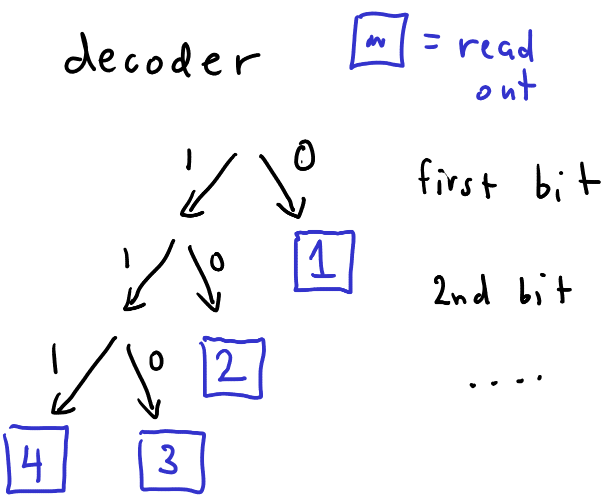
mean message length $= 2$ bits per state ($N=4$)

$= \log_2 N$ in general

case II: Bob is clever: a variable length code based on a binary decision tree ("prefix-free": no stop markers b/t states)

tree: decoder

one example (not unique)



state	code	# bits length ℓ_n
1	0	1
2	10	2
3	110	3
4	111	3

100111110 ..... $\Rightarrow$ 2 1 4 3 .....

avg. # of bits per state (cost of sending code)

$$B = \sum_n p_n(t)\, \ell_n = 0.4\times 1 + 0.35\times 2 + 0.2\times 3 + 0.05\times 3$$

$$= 1.85 \text{ bits / state}$$

Shannon proved: among such codes (all such binary decision trees) there is a lower bound (optimum)

$$B \geq B_{min} = -\sum_n p_n(t)\log_2 p_n(t)$$

units: "bits"

"information entropy" ~ "information"

compare: Gibb's entropy formula $-k_B \sum_n p_n(t) \ln p_n(t) = S(t)$

$$B_{min} = \frac{S(t)}{k_B \ln 2}$$

convert from $\frac{J}{K}$ to bit

examples: $p_n(t) = \delta_{n,1}$ (message: 1 1 1 1 1 1 ...)

$B_{min} = 0$ (no info needs to be sent)

$p_n(t) = \frac{1}{4}$

$B_{min} = 2$ bits (need 2 bits/state & can't do any better)

in general: $0 \leq B_{min} \leq \log_2 N$