Last Time: Info Theory

Source Coding Theorem: Let $(X_k)_{k \geq 1} \overset{iid}{\sim} P_X$. For any $\varepsilon > 0$, the vector $X_1, \ldots, X_n$ can be encoded using $\leq H(X) + \varepsilon$ bits/symbol on avg for all n sufficiently large.

Conversely, any encoding using $< H(X)$ bits/symbol will incur loss. (can't reproduce the source)

$H(X)$: entropy; tells us ultimate limit of compression.

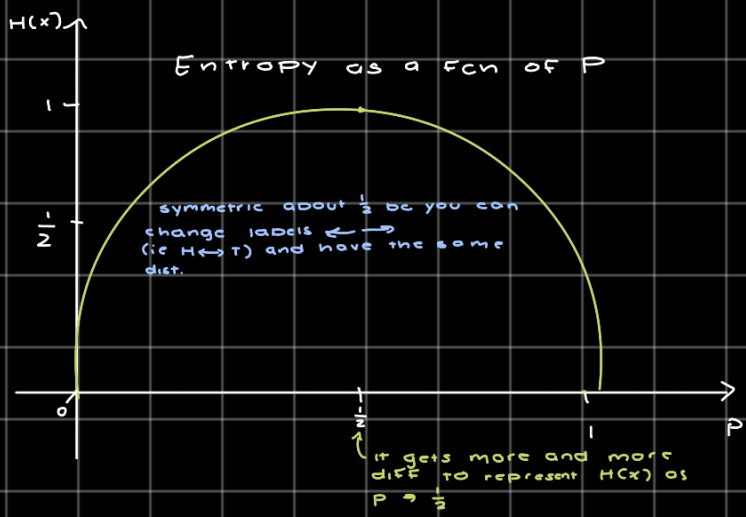
↳ tells us how many bits we need to represent an encoding losslessly

Result has 2 parts:

1. "Achievability": Descriptions using $\leq H(X) + \varepsilon$ bits/symbol on avg exist. (not always)
2. "Impossibility": Lossless descriptions using $< H(X)$ bits/symbol on avg don't exist.

- **Example of Near-Optimal Scheme**: If I know P_X (prob: in reality don't usually know P_X), I can design a "Huffman code" requiring $\leq H(X) + \frac{1}{n}$ bits on avg to compress sequences of length n.

For $X \sim \text{Bern}(p)$



Q/ How are we going to show we can compress down to the entropy?

A/ Probability?! specifically concentration. (From randomness comes structure)

For a sequence $(x_1, x_2, \ldots x_n) \in \mathcal{X}^n$, let prob of observing it be:

$$P(x_1 \cdots x_n) = \prod_{i=1}^{n} P_X(x_i)$$

i.e., $X_1 \cdots X_n \overset{iid}{\sim} P_X$

Asymptotic Equipartition Property (AEP) Thm:

If $(X_n)_{n\geq 1} \overset{iid}{\sim} P_X$, then

$$-\frac{1}{n}\log P(X_1,\dots,X_n) \to H(X) \text{ in probability}$$

Intuitively: With overwhelming probability,

$$P(X_1,\dots,X_n) \approx 2^{-nH(X)}$$

Pf:

$$-\frac{1}{n}\log P(X_1,\dots,X_n) = \frac{1}{n}\sum_{i=1}^{n}\log\frac{1}{P_X(X_i)} \xrightarrow{\text{WLLN}} \mathbb{E}\left[\log\frac{1}{P_X(X)}\right] = H(X)$$

(taking empirical avg)

Now, want to prove "achievability" of SCT using AEP. But first, need to introduce new concept (typical set):

Typical set: For $\varepsilon > 0$ For each $n \geq 1$ define the "typical set":

↳ sequences in the set that you'd typically see

$$A_\varepsilon^{(n)} := \left\{(x_1,\dots x_n) : p(x_1,\dots x_n) \geq 2^{-n(H(X)+\varepsilon)}\right\} \subset \mathcal{X}$$

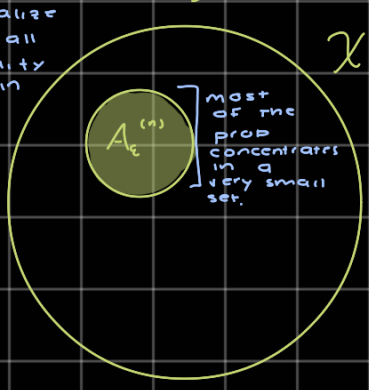
Properties: Let $(X_n)_{n\geq 1} \overset{iid}{\sim} P_X$

1. $P\{(X_1,\dots X_n) \in A_\varepsilon^{(n)}\} \to 1$ as $n \to \infty$
2. $|A_\varepsilon^{(n)}| \leq 2^{n(H(X)+\varepsilon)}$ ← if unfair coin flip, this is <1

In context of coin flips: entropy usually <1 (unless fair coin)

increasing n decreases size of set, but realize that almost all the probability concentrates in this set

by AEP



Pf of (1)

$$P\{(x_1, \dots, x_n) \in A_\varepsilon^{(n)}\} = P\{p(x_1, \dots, x_n) \geq 2^{-nH(x)+\varepsilon}\}$$

$$= P\{-\tfrac{1}{n} \log p(x_1 \dots x_n) \leq H(x) + \varepsilon\}$$

$$\longrightarrow 1 \quad \text{(by AEP)}$$

Pf of (2)

$$1 \geq \sum_{(x_1,\dots,x_n) \in A_\varepsilon^{(n)}} p(x_1, \dots, x_n) \geq \sum_{(x_1,\dots,x_n) \in A_\varepsilon^{(n)}} 2^{-n(H(x)+\varepsilon)} = |A_\varepsilon^{(n)}| \, 2^{-n(H(x)+\varepsilon)}$$

Q/ Suppose I have N objects. What's the max # of bits needed to represent each object?

A/ $\lceil \log_2(N) \rceil$

Back to Source Coding!

Protocol for Source Coding:

- If I observe $(x_1, \dots, x_n) \in A_{\varepsilon/2}^{(n)}$, I describe it via bitstring

  (1, x x x ... x) — the 1 is a flag, i.e. "I observed a # in the typical set"; the x's take $\lceil \log_2 |A_\varepsilon^{(n)}| \rceil$ bits

  $\leq 2 + \log |A_{\varepsilon/2}^{(n)}|$ bits

- If I observe $(x_1, \dots, x_n) \notin A_{\varepsilon/2}$, I describe it via bitstring:

  (0, x x x ... x) — the 0 is a flag; the x's take $\lceil \log |\mathcal{X}|^n \rceil$ bits → bc all bits in set $\mathcal{X}^n$

Q/ What is performance? (i.e. avg # bits needed per symbol observed)

$$\frac{1}{n} \mathbb{E}[\text{\# of bits in representation}]$$

Use Law of Total Prob

↓

$$\leq \frac{1}{n}(2 + n(H(x)+\varepsilon))P(\{x_1,\dots,x_n\} \in A_{\varepsilon/2}^{(n)}\}$$

$$+ \frac{1}{n}(2 + n\log|\mathcal{X}|)\, P\{(x_1,\dots x_n) \notin A_\varepsilon^{(n)}\}$$

↳ upper bound by 1

$$\leq H(x) + \frac{\varepsilon}{2} + \frac{4}{n} + \log|\mathcal{X}|\, P\{(x_1,\dots,x_n) \notin A_\varepsilon^{(n)}\}$$

$< \frac{\varepsilon}{2}$ $\forall n$ sufficiently large

$$\leq H(x) + \varepsilon \qquad \forall n \text{ large enough}$$

Key idea: leveraged probability & concentrations to compress sequences to almost entropy. This showed achievability of SCT.

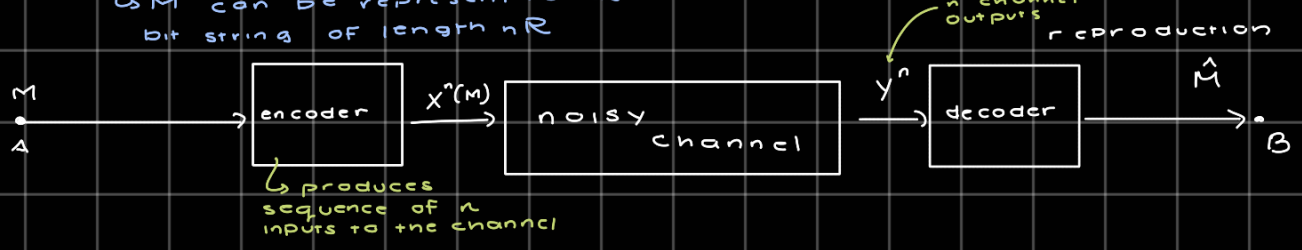
Channel Coding (Information Transmission)

↳ Step 1: Create a model of this problem

Model:

message $M \sim \text{Unif}(\{1,\dots 2^{nR}\})$ ← rate

↳ M can be represented as a bit string of length nR



$X^{(n)}(M) = (X_1(M) \cdots X_n(M))$ ↳ n channel inputs