09/27/2016
Tuesday

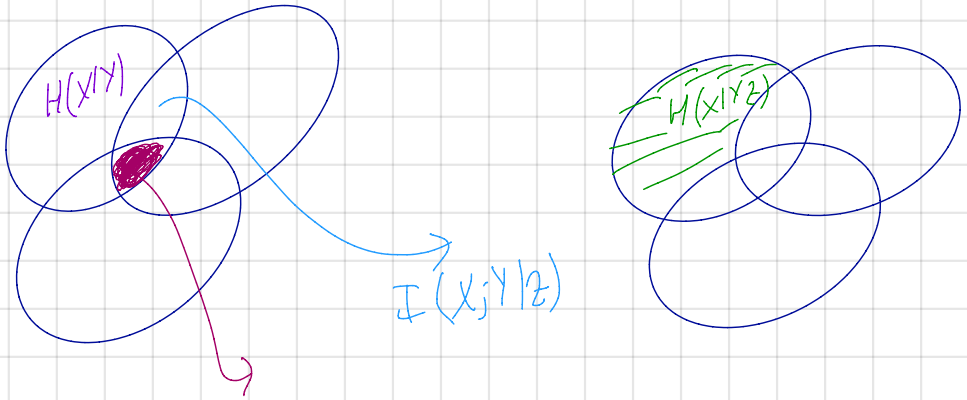
Last time:

$I(X;Y) \lessgtr I(X,Y|Z)$

If $X \perp\!\!\!\perp Z$ $\leq$

If $X - Y - Z$ $\geq$

How to remember:



$$H(X,Y,Z) - H(X,Y) - H(X,Z) - H(Y,Z) + H(X) + H(Y) + H(Z)$$

$= I(X;Y) - I(X,Y|Z)$

$= I(Y;Z) - I(Y;Z|X)$

$= I(X;Z) - I(X;Z|Y)$

$I(X;Y)$ is concave in P_X → How to show this using $I(X;Y) \geq I(X,Y|Z)$ when $X-Y-Z$?

$I(X;Y)$ " convex in $P_{Y|X}$

$W \sim Ber(p)$

$P_{X|W=1} = P_1$

$P_{X|W=0} = P_1$

$W - X - Y$

fix $P_{Y|X}$

$I(X;Y) \geq I(X;Y|W)$

↓ mixture P_X (input)

↳ convex combination of result.

Fano's Inequality

Suppose you can estimate X from Y with high probability ($H(X|Y)$ is small.)

Suppose $X - Y - \hat{X}$ and $\mathbb{P}[X \neq \hat{X}] = P_e$

$H(X|Y) \leq H(P_e) + P_e \log |\mathcal{X}|$

Proof

Define $E = 1_{X \neq \hat{X}} = \begin{cases} 1 & \text{when } X \neq \hat{X} \\ 0 & \text{when } X = X \end{cases}$

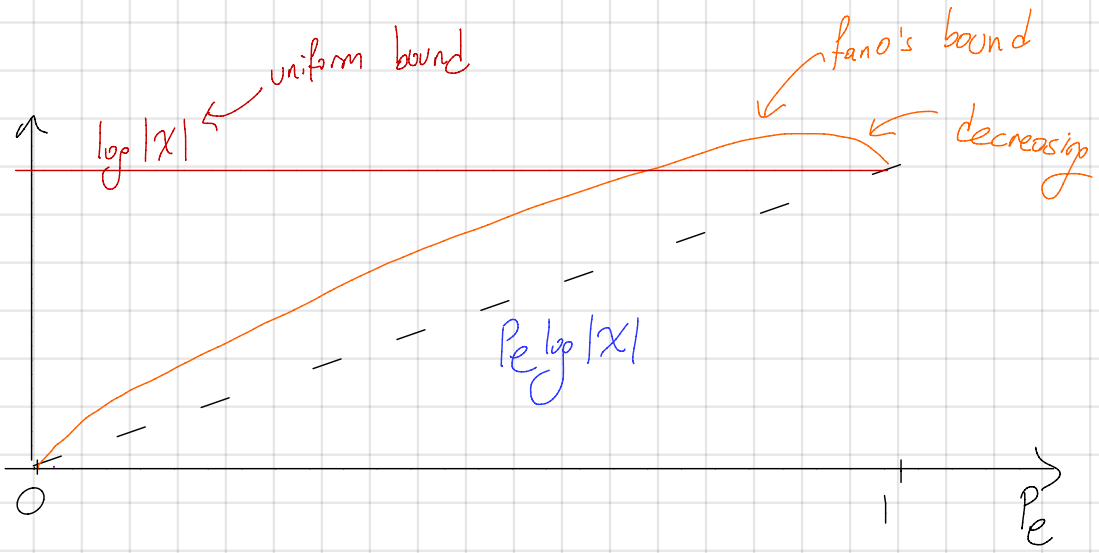
$$H(X,E|\hat{X}) = H(E|\hat{X}) + H(X|E,\hat{X})$$
$$= H(X|\hat{X}) + H(E|X,\hat{X})$$

$H(E|X,\hat{X}) = 0$

$$H(X|\hat{X}) = H(E|\hat{X}) + H(X|E,\hat{X})$$
$$\leq \underbrace{H(E)}_{H(P_e)} + \mathbb{P}[E=0] \cdot 0 + \underbrace{\mathbb{P}[E=1]}_{P_e} \log |\mathcal{X}|$$

$\mathbb{P}[E=0] \cdot 0 = 0$

~ we are being slightly loose could write $\log(|\mathcal{X}|-1)$



Asymptotic Equipartition Property (A.E.P)

L.L.N. If $X_1,\dots,X_n$ i.i.d $\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mathbb{E}[X]$

A.E.P $\frac{1}{n}\log\frac{1}{P_{X_1\cdots X_n}(X_1,\dots,X_n)} \xrightarrow{P} H(X)$

in other words: $P_{X_1\cdots X_n}(X_1,\dots,X_n) = 2^{-nH(X)}$

"almost all events are almost equally surprising."

Types of Convergences:

Convergence in probability: $Z_n \xrightarrow{P} Z$: $\forall \epsilon>0 \quad \mathbb{P}[|Z_n - Z|>\epsilon] \to 0$ as $n\to\infty$

Almost sure convergence: $Z_n \xrightarrow{a.s.} Z$: $\mathbb{P}[\lim_{n\to\infty} Z_n \neq Z] = 0$

L_2 convergence: $Z_n \xrightarrow{L_2} Z$: $\mathbb{E}[(Z_n - Z)^2] \to 0$

proof of AEP:

$$\frac{1}{n}\log\frac{1}{P_{X_1\cdots X_n}(X_1\cdots X_n)} = \frac{1}{n}\log\frac{1}{\prod_{i=1}^{n} P_X(X_i)}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\log\frac{1}{P_X(X_i)}$$

$$\xrightarrow{P} \mathbb{E}\log\frac{1}{P_X(X)} = H(X)$$

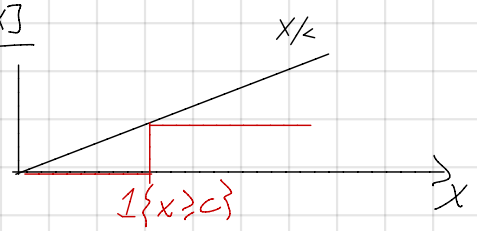
(by LLN)

Markov Inequality: If $X\geq 0$ w.p. 1

$$\mathbb{P}[X\geq c] \leq \frac{\mathbb{E}[X]}{c}$$

proof:

$$\mathbb{E}[1_{\{X\geq c\}}] \leq \mathbb{E}\left[\frac{X}{c}\right]$$



Chebychev's Inequality: $\mathbb{P}[|X-\mathbb{E}[X]|\geq c] \leq \frac{\sigma^2}{c^2}$

proof: Let $Y = (X-\mathbb{E}[X])^2$ $Y\geq 0$

$$\mathbb{P}[Y\geq d] \leq \frac{\mathbb{E}[Y]}{d} = \frac{\sigma_X^2}{d}$$

$$= \mathbb{P}[|X-\mathbb{E}[X]|^2 \geq d]$$

$$= \mathbb{P}[|X-\mathbb{E}[X]| \geq \sqrt{d}]$$

proof of LLN: Assume: X has finite variance. Use Chebychev's inequality.

$$\mathbb{P}\left[\left|\frac{1}{n}\sum X_i - \mathbb{E}[X]\right| > \varepsilon\right] \leq \frac{\frac{1}{n^2}\cdot n\cdot \mathrm{Var}(X)}{\varepsilon^2} \longrightarrow 0$$

Typical Sets: Let $X^n = (X_1, \ldots, X_n)$ (write $X_i^j = (X_i, \ldots, X_j)$)

$$A_\varepsilon^{(n)} = \left\{x^n : \left|\frac{1}{n}\log\frac{1}{P_{X^n}(x^n)} - H(X)\right| \leq \varepsilon\right\} \quad \text{(implicitly a function of } P_X \text{ and } P_{X^n} = \Pi P_X\text{)}$$

Thm: 1) $\forall x^n \in A_\varepsilon^{(n)}$: $P_{X^n}(x^n) \in \left[2^{-n(H(X)+\varepsilon)}, 2^{-n(H(X)-\varepsilon)}\right]$

2) $P_{X^n}(A_\varepsilon^{(n)}) = \mathbb{P}\left[X^n \in A_\varepsilon^{(n)}\right] \longrightarrow 1$ as $n\to\infty$ $\forall \varepsilon > 0$

3) $|A_\varepsilon^{(n)}| \leq 2^{n(H(X)+\varepsilon)}$

4) $|A_\varepsilon^{(n)}| \geq (1-\varepsilon)\, 2^{n(H(X)-\varepsilon)}$ for n large enough.

proof:

1.) By defn of $A_\varepsilon^{(n)}$

2) By AEP theorem

3) $$1 = \sum_{x^n \in \mathcal{X}^n} P_{X^n}(x^n)$$
$$\geq \sum_{x^n \in A_\varepsilon^{(n)}} P_{X^n}(x^n)$$
$$\geq \sum_{x^n \in A_\varepsilon^{(n)}} 2^{-n(H(X)+\varepsilon)} = |A_\varepsilon^{(n)}|\, 2^{-n(H(X)+\varepsilon)}$$

4) For n large enough: $\sum_{x^n \in A_\varepsilon^{(n)}} P_{X^n}(x^n) = P_{X^n}(A_\varepsilon^{(n)}) \geq 1-\varepsilon$

$$\sum_{x^n \in A_\varepsilon^{(n)}} 2^{-n(H(X)-\varepsilon)} = |A_\varepsilon^{(n)}|\, 2^{-n(H(X)-\varepsilon)}$$

Lossless Data Compression:

$X_1, \ldots, X_n \sim$ i.i.d. P_X

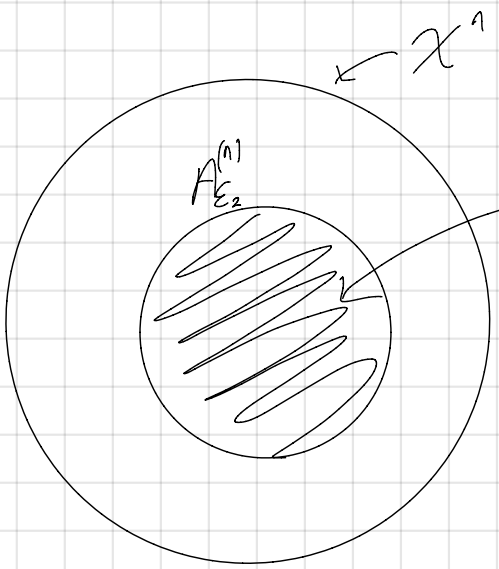
Block encoding: $X^n \to$ [Enc] $\to M \in \{1, \ldots, 2^{nR}\} = [2^{nR}]$ $\to$ [Dec] $\to \hat{X}^n$

(R bits per symbol)

Goal: $\mathbb{P}[X \neq \hat{X}] < \varepsilon$

Achievability: R is achievable if $\forall \varepsilon > 0$, exists n, encoder, decoder such that $\mathbb{P}[X^n \neq \hat{X}^n] < \varepsilon$

Theorem: Minimum achievable rate for lossless compression is $H(X)$



$$|\mathcal{X}^n| = |\mathcal{X}|^n = 2^{n \log |\mathcal{X}|}$$

Typical set

$$|A_{\varepsilon_2}^{(n)}| \leq 2^{n(H(X)+\varepsilon)}$$

As long as $R > H(X) + \varepsilon_2$ is achievable

→ You can turn this into 0-error scheme by putting an indicator bit: If typical indicator bit is 1 and we do AEP encoding. If indicator is 0 then we encode the whole sequence x^n with $n \log |\mathcal{X}|$ and Average length: $\mathbb{P}[A_\varepsilon^{(n)}](n(H(X)+\varepsilon)) + (1 - P(A_\varepsilon^{(n)}))(n \log |\mathcal{X}|)$

What about a converse result?

Given n, enc., dec. s.t. $\mathbb{P}[X^n \neq \hat{X}^n] < \varepsilon$ for ε arbitrarily small.

$$X^n - M - \hat{X}^n$$

$$nR \geq H(M) \geq I(X^n; M) \overset{\text{Data Proc. ineq.}}{\geq} I(X^n; \hat{X}^n) = H(X^n) - H(X^n|\hat{X}^n) = nH(X) - H(X^n|\hat{X}^n) \geq nH(X) - H(\varepsilon) - \varepsilon n \log |\mathcal{X}|$$

$$\Rightarrow R \geq H(X) - \varepsilon \log |\mathcal{X}| - \frac{H(\varepsilon)}{n} \quad \text{as } \varepsilon \to 0 \quad R \geq H(X)$$