Start @ Winkler

Lec 3. [Reference NC . Sec 12-2]

[Capital: rv
Small: outcomes
values
realization.]

X: random variable

Ω: sample space , $|\Omega| = m$ (m outcomes)

p : prob distⁿ of X.

$$p : \Omega \to [0, 1]$$

$$x \longmapsto p(x) , \quad \sum_{x \in \Omega} p(x) = 1.$$

← X takes value in Ω.

eg. a loaded dice

$\Omega = \{1, 2, \ldots, 6\} = [6]$

$p(1) = \frac{1}{2}, \quad p(2) = \cdot \quad = p(6) = \frac{1}{10}.$

- A (discrete) <u>source</u> is a sequence of r.v.'s $X_1, X_2, \ldots,$ with a common sample space Ω (source alphabet). eg. n draws have m^n outcomes.

eg. can roll dice countably many times

eg. $\Omega = \{$ sun, cloud, rain, snow $\}$.

Can record the weather of every day → gives a source.

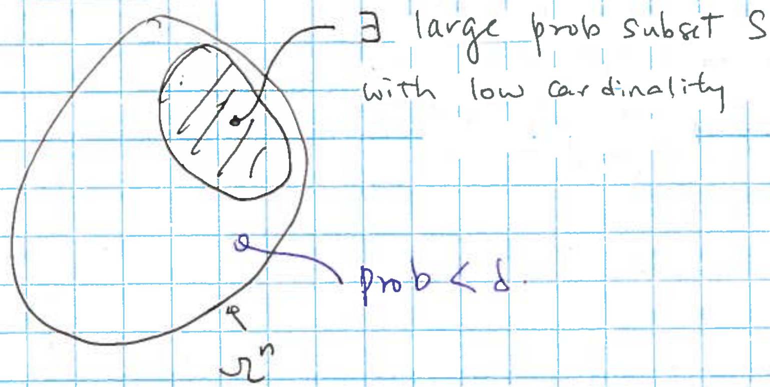
- In general, X_i's need not be indep or identical.
- If X_i's iid (independent & identically distributed) we have an "iid source" - what we focus on mostly.

For iid source:

Want: - quantify info carried by a source

- min physical requirement needed to store an info source.

Idea:



Keep

How does S look like?

Consider arbitrary $x^n = x_1 x_2 \cdots x_n$ ← $X^n = X_1 X_2 \cdots X_n$.

$p(x^n) = p(x_1) \cdot p(x_2) \cdots p(x_n)$ by independence.

$= 2^{\log p(x_1)} \, 2^{\log p(x_2)} \cdots 2^{\log p(x_n)}$ ← log based 2.

$$= 2^{-n\underbrace{\left[\frac{1}{n}\sum_i (-)\log p(x_i)\right]}_{\text{Empirical ave of } -\log p(x)}} \xrightarrow{\text{LLN}} 2^{-n\underbrace{\left[-\sum_{x\in\Omega} p(x)\log p(x)\right]}_{-\mathbb{E}_p \log p(x) = H(X)}}$$

S contains x^n with $p(x^n) \approx 2^{-nH(X)}$.

Bonus: elts of S ≈ equiprobable

$x^n \in T_{n,\varepsilon} \iff p(x^n) = 2^{-n\left[\frac{1}{n}\sum_i (-)\log p(x_i)\right]} \in \left[2^{-n(H(X)+\varepsilon)}, 2^{-n(H(X)-\varepsilon)}\right]$

$\iff \frac{1}{n}\sum_i \log p(x_i) \in \left[-H(X)-\varepsilon, -H(X)+\varepsilon\right]$

Def [Shannon Entropy]:

$$H(X) \text{ or } H(p) := -\sum_{x \in \Omega} p(x) \log p(x)$$

eg. for the loaded dice:

$$H(X) = -\left[\tfrac{1}{2}\log\tfrac{1}{2} + 5 \times \tfrac{1}{10}\log\tfrac{1}{10}\right]$$

$$= \tfrac{1}{2} + \tfrac{1}{2}\log 10 \doteq 2.1610$$

for a fair dice:

$$H(X) = -\left[6 \times \tfrac{1}{6}\log\tfrac{1}{6}\right] = \log 6 \doteq 2.5850.$$

Def [typical sequence]. ["atypical" if not]

- x^n is ε-typical

if $\left| -\frac{1}{n}\log p(x^n) - H(X) \right| \leq \varepsilon.$

For loaded dice:

Say $\varepsilon = 0.2$, $n = 20$

(1) $x^n = 11111111123344455666$

$$-\tfrac{1}{20}\left[9 \times \log 0.5 + 11 \times \log 0.1\right]$$

$= 2.2771$ $\therefore$ ε-typical.

(2) So is $1^9 2^{11}$

(3) Not $1^8 2^3 3^2 4^3 5^2 6^2$

Def [typical set]

- $T_{n,\varepsilon} = \{x^n : x^n \ \varepsilon\text{-typical}\}$

Note perm sym.

$$2.3932 = -\tfrac{1}{20}\left[8\log 0.5 + 12\log 0.1\right]$$

Asymptotic Equipartition thm (AEP):

$\forall \varepsilon > 0 \quad \forall \delta > 0 \quad \exists n_0$ s.t. $\forall n \geq n_0$

① $p(T_{n,\varepsilon}) \geq 1-\delta$

② $(1-\delta)\, 2^{n(H(x)-\varepsilon)} \leq |T_{n,\varepsilon}| \leq 2^{n(H(x)+\varepsilon)}$

③ $\forall A \subseteq \Omega^n, \quad p(A) \geq 1-\delta \Rightarrow |A| \geq (1-2\delta)\, 2^{nH(|X|-\varepsilon)}$

ie. $\forall \delta > 0$, large prob means $> 1-\delta$

① The typical set is a large prob set

② Typical set are $\approx$ equiprobable

③ Every large prob set cannot be much smaller.

Pf ① Essentially from LLN on $Y = \log p(X)$.

Induced r.v. $\forall x \in \Omega$

$Y = \log p(x)$. wp $p(x)$, $\mathbb{E}Y = \sum_x p(x)\log p(x) = -H(X)$.

The iid source $X_1 X_2 \dots X_n$

induces iid - - $Y_1 Y_2 \dots Y_n$.

Let $x^n = x_1 x_2 \dots x_n \in \Omega^n$, $y_i = \log p(x_i)$.

Then $x^n \notin T_{n,\varepsilon} \iff \left| \frac{1}{n}\sum_{i=1}^{n} y_i - \mathbb{E}Y \right| > \varepsilon$. (*)

Chebyshev's ineq for any rv Z :

$$\Pr\{ |Z - \mathbb{E}Z| \geq k\sqrt{\operatorname{Var} Z} \} \leq \frac{1}{k^2}$$

Choose $Z = \frac{1}{n}\sum_{i=1}^{n} Y_i$

$$n \geq n_0 = \frac{\operatorname{Var} Y}{\varepsilon^2 \delta}$$

$$\operatorname{Var} Z = \frac{1}{n} \underbrace{\operatorname{Var} Y}_{\text{fixed}} \leq \frac{1}{n_0}\operatorname{Var} Y = \varepsilon^2 \delta$$

we can choose

$$k = \frac{\varepsilon}{\sqrt{\operatorname{Var} Z}}$$

$$\Pr\left\{ \left| \frac{1}{n}\sum_{i=1}^{n} y_i - \mathbb{E}Y \right| \geq \varepsilon \right\} \leq \frac{\operatorname{Var} Z}{\varepsilon^2} \leq \delta \quad (**)$$

$\therefore \Pr(x \notin T_{n,\varepsilon}) < \delta$. by (*) & (**)

[NB $T_{n,\varepsilon}$ "large prob" i.e. $\geq 1-\delta$]

Pf ② $\quad 1-\delta \overset{①}{\boxed{\leq}} p(T_{n,\varepsilon}) \leq 1$

$$|T_{n,\varepsilon}|\, 2^{-n(H(X)+\varepsilon)} \leq \sum_{x^n \in T_{n,\varepsilon}} p(x^n) \boxed{\leq} |T_{n,\varepsilon}|\, 2^{-n(H(X)-\varepsilon)}$$

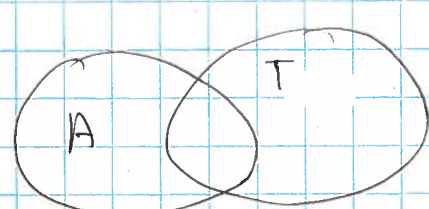
(by Def of typicality).

$\therefore |T_{n,\varepsilon}| \leq 2^{n(H(X)+\varepsilon)}$

$|T_{n,\varepsilon}| \geq (1-\delta)\, 2^{n(H(X)-\varepsilon)}$

NB $\quad \dfrac{|T_{n,\varepsilon}|}{|\Omega|} \leq 2^{-n[\underbrace{\log m - H(X) - \varepsilon}_{\text{+ve for most } X}]}$. $\quad$ [$\therefore$ $T_{n,\varepsilon}$ has "small cardinality"].

Pf ③. Let A has $p(A) > 1-\delta$, $T = T_{n,\varepsilon}$



$$|A| \geq |A \cap T|$$

$$\geq \frac{\text{Prob}(A \cap T)}{\max_{a \in A\cap T} \{p(a)\}}.$$

$$\geq \frac{\text{Prob}(A \cap T)}{2^{-n[H(X)-\varepsilon]}} \qquad (\because a \in T)$$

$$\text{Prob}(A \cap T) = 1 - \text{Prob}(\neg A \cup \neg T)$$

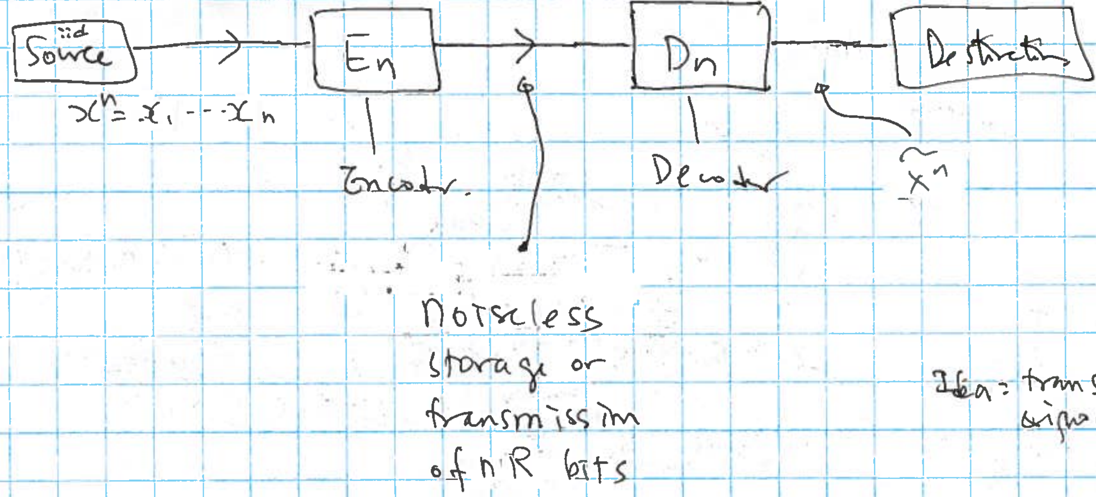
$$\geq 1 - [\underbrace{\text{Prob}(\neg A)}_{\leq \delta} + \underbrace{(\text{Prob}\, \neg T)}_{\leq \delta}] \qquad \text{Union bdd.}$$

$$\geq 1 - 2\delta$$

$\therefore |A| \geq (1-2\delta)\, 2^{n[H(X)-\varepsilon]}$.

App: Data compression

AKA: Shannon's noiseless coding thm.



Idea: transmit typical set & ignore the rest.

Thm: ① $\forall \delta > 0$, $\forall R > H(x)$

$\exists n_0$ s.t. $\forall n \geq n_0$

(depends on δ) $\exists$ En, Dn

s.t. $\Pr(\hat{x}^n \neq x^n) < \delta$ above.

② $\forall R < H(x)$

$\exists n_0$ s.t. $\forall n \geq n_0$

$\forall$ En, Dn,

$\Pr(\hat{x}^n = x^n) \leq (1-2\delta)\, 2^{n\left[\frac{H(x)-R}{2}\right]}$

• Proof reading

~~Re to this~~

Pf (1) Let $\varepsilon = R - H(X)$, $Y = \log p(X)$

$\forall n > n_0 = \frac{\operatorname{Var} Y}{\varepsilon^2 \delta}$, so AEP holds.

Assign a unique positive integer in binary rep $f(x^n)$ to every $x^n \in T_{n,\varepsilon}$

E_n: check if input $x^n \in T_{n,\varepsilon}$.

If yes, output $f(x^n) = b$

if no, output error symbol "E"

D_n: check if input = "E".

If so output "E"

Else, apply f^{-1}.

$\Pr(\hat{X}_n \neq X_n) = \Pr(X_n \notin T_{n,\varepsilon}) \leq \delta$.

Pf (2). WLOG E_n deterministic

[a random E_n that works

works for a specific value of random coins].

So only 2^{nR} symbols present.

Let $\varepsilon = \frac{H(X)-R}{2}$.

Any set A with $p(A) \geq 1-\delta$ overlap with $T_{\varepsilon,n}$

on at least $(1-2\delta)\, 2^{n(H(X)-\varepsilon)} = (1-2\delta)\, 2^{n(R+\varepsilon)}$

elements

Can transmit them w/ 2^{nR} symbols only. (C2).

Comments:

- Allowing an arbitrary small error (δ)
  makes a big diff in required resources.

  (e.g. $n \log m \to n H(X)$ bits).

- Data compression gives the Shannon entropy $H(X)$ an operational meaning
  - how much is needed to rep each symbol asymptotically, or how uncertain we are without knowing a symbol.

- NB: data compression gives a rate per symbol which is indep of the block length, though the code is a block code.

Cover & Thomas.

- We've consider "block codes" with fixed input & output size for E_n, and an inefficient E_n.
  requires entire x^n to output $f(x^n)$

  There are schemes (e.g. Huffman code) with no error, and variable length outputs, and no need for brute force table lookup for $f(x^n)$. [Note E_n is hard to implement, though all we need the here is an info th result - so OK]

- $\exists$ Universal compression alg requiring only an upper bdd on $H(X)$ without knowing much about X itself.

Quantum source, entropy and data compression:

- Imagine a device, such that every time you press a button, it emits a quantum system in state ρ_x wp $p(x)$
  d-dim

[the description for $p(x)$, ρ_x known.]

- This is "one draw" from the ensemble $\mathcal{E} = \{ p(x), \rho_x \}$. ← set of pairs.

- Can be modeled as, ① a r.v. X drawn, $\sim p(x)$ ② conditional on x, create ρ_x.

- The state after one draw is:

$$\Lambda = \sum_x p(x)\, |x\rangle\langle x|_D \otimes \rho_{x\,S}$$

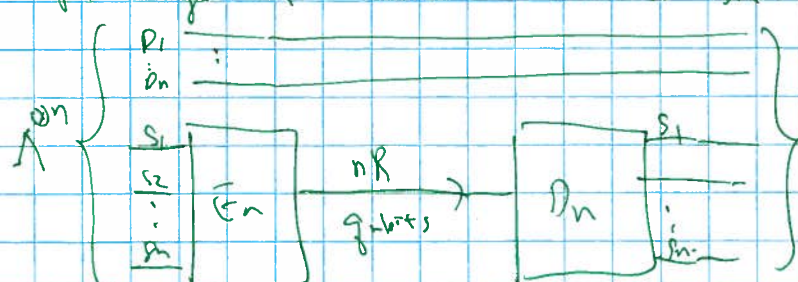
$\mathrm{tr}_S \Lambda = \sum_x p(x)\, |x\rangle\langle x|$ is one draw of the classical r.v. X w/ dist^n $p(x)$.

$\rho := \mathrm{tr}_D \Lambda = \sum_x p(x)\, \rho_{x\,S}$ is called the "average state" of the ensemble $\mathcal{E}$

- The state after n draws: keep on board

$$\Lambda^{\otimes n} = \sum_{x_1 x_2 \cdots x_n} p(x_1)\, p(x_2) \cdots p(x_n)\, \underbrace{|x_1 x_2 \cdots x_n\rangle\langle x_1 \cdots x_n|}_{D_1 D_2 \cdots D_n} \otimes \underbrace{\rho_{x_1} \otimes \rho_{x_2} \otimes \cdots \otimes \rho_{x_n}}_{S_1 S_2 \cdots S_n}$$

- Qn: for large n, what's the min R s.t. $\exists\, \mathcal{E}_n, \mathcal{D}_n$. skip?? keep.



$\approx \Lambda^{\otimes n}$?

eg. BB 84:

$\Omega = \{0, 1, 2, 3\}$, $f(x) = \frac{1}{4}$ $\forall x \in \Omega$.

$\rho_0 = |0\rangle\langle 0|$, $\rho_1 = |1\rangle\langle 1|$, $\rho_2 = |+\rangle\langle +|$, $\rho_3 = |-\rangle\langle -|$

$|+\rangle = \frac{|0\rangle + |1\rangle}{\sqrt{2}}$, $|-\rangle = \frac{|0\rangle - |1\rangle}{\sqrt{2}}$.

In this case, $D_1 \ldots D_n$ stays with Alice. (purification w/ $f_{rv}(c)$)

$S_1 \ldots S_n$ is transmitted to Eve then to Bob.

eg. B92

$\Omega = \{0, 1\}$, $f(x) = \frac{1}{2}$ $\forall x \in \Omega$.

$\rho_0 = |0\rangle\langle 0|$, $\rho_1 = |+\rangle\langle +|$.

Similar to above.