LECTURE 16

TODAY

POLAR CODES

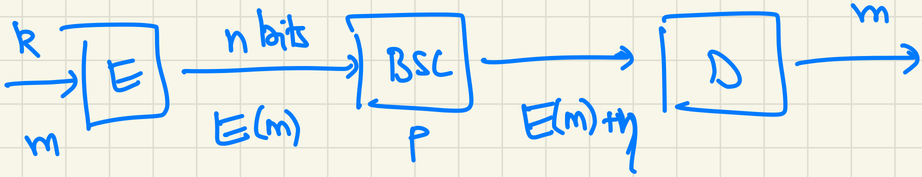
- Motivation: Shannon & Gap to Capacity
- Construction:
  0. Reduction to Linear Compression
  1. Polarizing Tranformation + Inf. Th. basics.
  2. Polarization + Theorem
  3. Encoding + Decoding
  4. Proof of Polarization Theorem. (some skipped)

ADMIN

- PS4 Due this week
- Weekly reports every week
- PS5 out tomorrow
- "Practice PS6" might be released later
- 4 best PS out of 6.
- Ask if you have questions.

Use Chat during lectures!

Motivation



$$\Pr[\hat{m} = m] \geq 1 - o(1)$$

- Can do this with Rate $(k/n) \to 1 - H(p)$

  $p \triangleq$ Bit flip Prob.

- if we want to be at rate $1 - H(p) - \epsilon$ then can achieve error prob $\exp(-\epsilon' \cdot n)$

- Shanon - Non constructive

- PS 3 ? - Can make this poly time .....

  Concatenation "disappeared" the problem ?

- Running are now poly(n)

  But there is a constant in front which depends on ϵ.

- e.g. $2^{1/\epsilon^2} \cdot n^2$ $\Leftarrow$ needed because blocks are of size $1/\epsilon^2$ ... exponential in block.

Problem Formulation: Given ϵ, determine k, n

s.t. $\frac{k}{n} \geq 1 - H(p) - \epsilon$

& $\exists$ $E_k: \{0,1\}^k \to \{0,1\}^n$

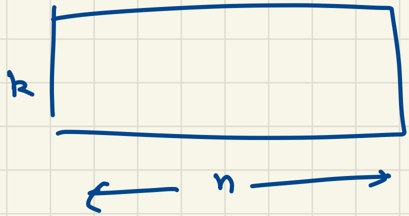
$D_k: \{0,1\}^n \to \{0,1\}^k$

$$\boxed{\text{s.t. } \Pr_{m, BSC}\left[D(BSC(E(m))) \neq m \right] \leq 1 - o(1)}$$

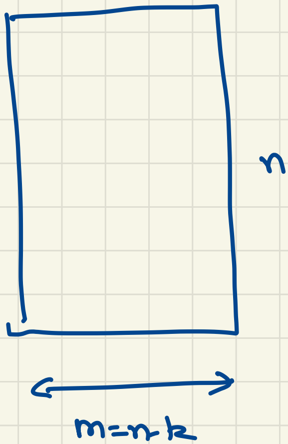
& running time of E_k, D_k are poly$(1/\epsilon)$

History:
- Q. raised by Luby, Mitzenmacher, Shokrollahi, Spielman '95
- 2008: Arikan - proposed POLAR CODES
- 2013: - Guruswami, Xia
  - Hassani, Alishahi, Urbanke

$\Big\}$ resolved using POLAR codes.

Codes speified by $G =$ 

or by

$H =$ 

G is generator of a ... good code (corrects errors)

$\iff$ H is a <u>good</u> linear compressor of $\text{Bern}(p)^n$

$$\text{Bern}(p) = Z = \begin{cases} 0 & \text{w.p. } 1-p \\ 1 & \text{w.p. } p \end{cases}$$

$\text{Bern}(p)^n$ = n independent copies $Z_1 \dots Z_n$

$Z_i \sim \text{Bern}(p)$.

~~When~~ is H a <u>good</u> compressor?

H good compresses $\mathrm{Bern}(p)^n$ if

① $H = n \times m$ with $m \le (H(p)+\epsilon)\cdot n$

② $\exists$ Decoding alg. D

$$\Pr_Z\Big[D(\underbrace{ZH}_{\text{compression } Z}) \ne Z\Big] = o(1)$$

③ D should run in time $\mathrm{poly}(1/\epsilon)$.

$$\underset{\text{parity}}{H} \equiv \underset{\text{genetor}}{G}$$

$$X \xrightarrow{\text{Enc}} XG \xrightarrow{\text{Channel}} \underbrace{XG + Z}_{Y} \longrightarrow$$ Can get Z & hence X whp.

Ex.

$\boxed{\text{non linear compressor easy}}$

① $Y \cdot H = XGH + ZH = ZH$

② $D(ZH) = \hat{Z} = Z$ whp

③ $Y - \hat{Z} =$ whp our codeword.

Aside: Linear compression equivalent to our problem if we want linear code.

Rest of Polar Coding: Linear Compression.

Arikan's

- Two bits & compress them:

$$(U, V) \longrightarrow (U+V, V)$$

- Move entropy around

$U+V$ more "uncertain" than U or V

$H(U+V) > H(U)$

- V no more/less entropic V?

"Conditional Entropy" $H(V \mid U+V) < H(V \mid U) = H(V)$

> **Exercise**
>
> Determine Dist. of $U+V$, given $U \sim Bern(p_1)$, $V \sim Bern(p_2)$

Entropy: of random variable X dist. on $[M]$

~~has~~ with $\Pr[X=i] = p_i$

$$H(X) \triangleq \sum_{i=1}^{M} p_i \log_2 \frac{1}{p_i}$$

Entropy says how effectively we can compress n ind. copies of X; amortized

[Describing $X_1 \ldots X_n$ $\quad X_j \sim X$ i.i.d. takes roughly $H(X) \cdot n$ bits]

- Suppose $U, V \sim \text{Bern}(.01)$

- $U+V \sim \text{Bern}(.01999)$

- $U+V = 0 \Rightarrow \Pr[U=1 \mid U+V=0] = .000101\ldots$

$\uparrow \Pr[\cdot] \approx .98$

- $U+V=1 \Rightarrow \Pr[U=1 \mid U+V=1] = \frac{1}{2}$

$\Pr[\ ] \approx .02$

But in expectation U will be "less random" given $U+V$.

Conditional Entropy

- (X, Y) jointly dist. over $[M] \times [N]$

$\Pr[X=i \wedge Y=j] = P_{ij}$; P_X marginal on X
P_Y " on Y

- $$H(Y|X) = \mathop{\mathbb{E}}_{x \sim P_X}\left[H(Y|X=x)\right]$$

---

① **Chain Rule**

$$H(X,Y) = H(X) + H(Y|X)$$

bits needed to describe Y if you know X.

② "Conditioning "reduces" entropy: $H(Y|X) \leq H(Y)$.

- f: is a one-to-one function
  $H(X) = H(f(X))$

- $H(U,V) = H(U+V, V)$ — ①

- (By calc / Exercise): $H(U+V) > H(U)$ — ②

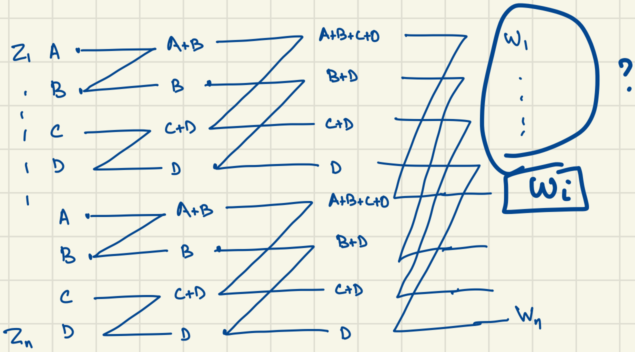
- $H(V|U+V) = H(U+V, V) - H(U+V)$ Chain Rule
  $= H(U,V) - H(U+V)$ by ①
  $< H(U,V) - H(U)$ by ②
  $= H(V|U)$ Chain
  $= H(V)$ [since V & U indenpent].

Polar coding idea

- lets iterate this process many times, moving conditional entropies around.

- "Polarization": At end every bit will conditional Entropy close to 0, or 1.

- Compression: throw away all bits with 0 entropy.



$$H(W_i \mid W_1 \ldots W_{i-1}) = \text{very close to } 0 .$$

or very close to 1

$$S \triangleq \{ i \mid H(W_i \mid W_{<i}) \to 1 \}$$

Compression of $Z = W|_S$

① This is linear ✓

② $|S| \approx H(p) \cdot n$ → follows from all entropies are 0 or 1 ?

③ Given $W|_S$ can compute W and then Z efficiently.

All together ⇒ Polarization process good compressor

⇒ Gives our theorem.