

LECTURE 22TODAY: "ENTROPY IN COMPUTATIONAL COMPLEXITY"

→ Statistical Difference Problem (SD)

-  $SD \equiv \overline{SD}$



Probability Distributions in (T)CS vs. IT:

- TCS focus: One sample from distribution on large support
- IT focus: Many samples from distribution on small support

Resulting Concepts:

- IT:  $H(X)$ , KL Divergence  $D(P||Q)$ - CS:  $H_\infty(X)$ ; Statistical distance  $\mathcal{E}(P, Q)$ .

Definitions:

$$\bullet H_\infty(X) = \min_{w \in \Sigma} \log_2 \frac{1}{\Pr[X=w]} \quad \text{vs. } H(X) = \mathbb{E}_{w \in \Sigma} \left[ \log_2 \frac{1}{\Pr[X=w]} \right]$$

• Very popular in "randomness extraction" ...

(Not today's focus).

-  $S(P, Q) = \frac{1}{2} \sum_{w \in \Omega} |P(w) - Q(w)| = \max_{T: \Omega \rightarrow \{0,1\}} \left\{ \begin{array}{l} \Pr_{X \sim P} [T(X)=1] \\ - \Pr_{X \sim Q} [T(X)=1] \end{array} \right\}$

max. distinguishing prob.

- Strong in CS: Distributions that are not very distinguishable  
 $S(P, Q) \rightarrow 0$  ( $\frac{1}{n^{100}}, \frac{1}{2\sqrt{n}}, \dots$ )

- Useful features: -  $\Delta$ -inequality  $\Delta(P, Q) + \Delta(Q, R) \geq \Delta(P, R)$   
 "indistinguishability" is "transitive".

- Computational Indist.

$S^c(P, Q) = \max_{T: \Omega \rightarrow \{0,1\}} \left\{ \begin{array}{l} E[T(X)] - E[T(Y)] \\ \text{Efficiency} \\ \text{Computable} \end{array} \right\}$

$S^c(P, Q) \leq S(P, Q)$

$S^c(P, Q) \leq S^c(P, R) + S(R, Q)$  (as an example).

Comparisons: ① Pinsker:  $S(\cdot)$  vs.  $D(\cdot)$

②  $H_2(X^n) \approx n H_2(X)$  for  $X^n$ ,  $\exists \bar{X} \approx_{d(\cdot)} X^n$  with

$H_\infty(\bar{X}) = n \cdot H(X)$ .



Historically: ① Prob. Encryption [Goldwasser-Micali] "Comp. Indst" ③  
↓ (in a few steps)

② [GMW]: Graph-Non-Isomorphism ∈ "(Statistical) Zero Knowledge"

↓ (many years later)

③ Statistical Difference is SZK-complete

④  $SD \approx \overline{SD}$  ← This is what we'll define & prove.

- Given some idea of steps above: (including idea of definitions).

- [GMW] GNI ∈ SZK (TLA-galore).

Question: How can I prove to you that two  $n$ -vertex graphs are not isomorphic

Defn:  $G = (V, E)$   $H = (W, F)$  are isomorphic

if  $\exists \pi: V \rightarrow W$  1-1

st.  $(u, v) \in E \iff (\pi(u), \pi(v)) \in F$

- Proving isomorphism easy: I can send  $\pi$  to you.

- Non-isomorphism? Can send all  $n!$   $\pi$ 's to you?

No good!



# What else can be Proof (Interactive, Zero Knowledge):

Verifiers (You)

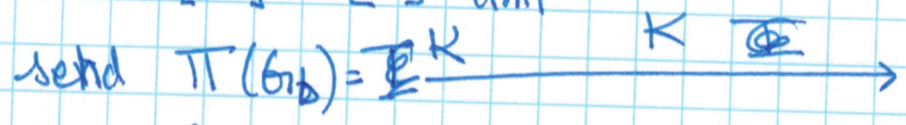
Prover (me)

$$G_0 \neq H \quad G = G_0 = ([n], E)$$

$$H = G_1 = ([n], F)$$

↓  
- Pick  $b \in_{\mathcal{R}} \{0, 1\}$

-  $\Pi: [n] \rightarrow [n]$  unit.



(Challenge: Guess  $\delta$ )

Accept if  $b = \hat{\delta}$ .

Assertions: (1) if  $G \neq H$  I can (with  $\exp(n)$  time) guess  $b$  from  $\mathbb{F}^K$  ( $\mathbb{F}^K$  isomorphic to  $G_b$  only)

$$\Rightarrow \Pr[\text{accept}] = 1$$

(2) if  $G \neq H$  I can't do better than random guessing. formally

$$\left\{ \mathbb{F}^K \mid b=0, G, H \right\} \cong \left\{ \mathbb{F}^K \mid b=1, G, H \right\}$$

or equiv  $I(K; b \mid G, H) = 0$ .

(3) Verifier learns nothing except  $G \neq H$  (if so.).

"Zero Knowledge" - No formalism here!



Question: What other statements can be proved like this.

Theorem [SV]: SD is SZK-complete.

Definition: Given boolean circuit  $C: \{0,1\}^m \rightarrow \{0,1\}^n$  bits

with  $m, |C| = \text{poly}(n)$ ,

we say  $C$  "represents" a sampleable distribution

(also called)  $C$  supported on  $\{0,1\}^n$  given by  $\{C(y) \mid y \text{ uniform on } \{0,1\}^m\}$

("circuits" are "sampleable distributions")

$SD^{c,f}$ :

$$CLOSE^c = \{ (C_1, C_2) \mid \delta(C_1, C_2) \leq c \}$$

$$FAR^f = \{ (C_1, C_2) \mid \delta(C_1, C_2) > s \}$$

Problem: Given  $(C_1, C_2) \in CLOSE^c \cup FAR^f$  decide if  $(C_1, C_2) \in CLOSE^c$   
or  $(C_1, C_2) \in FAR^f$

Claims (1)  $SD^{1/3, 2/3}$  is SZK-complete [won't define or prove]

$$(2) \quad SD^{1/3, 2/3} \equiv SD^{2^{-n}, 1-2^{-n}}$$



Hint of SD  $2^{-n\epsilon}, 1-2^{-n\epsilon} \in \text{SZK}$

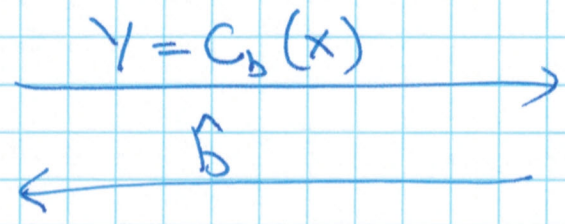
Verifier

Prover

$C_0, C_1$

Pick  $b \in \{0,1\}$  at random

$Y \in \{0,1\}^m$  at random



Completeness:  $\delta(C_0, C_1) \geq 1-2^{-n\epsilon} \Rightarrow$  Prover accepted w.p.  $1-2^{-n\epsilon}$

Soundness:  $\delta(C_0, C_1) \leq 2^{-n\epsilon} \Rightarrow$  Prover accepted w.p.  $\leq \frac{1}{2} + 2^{-n\epsilon}$

"ZK":  ~~$\Pr[b = \hat{b}] \geq 1-2^{-n\epsilon} = 1$~~

~~Case~~  
FAR

So prover ~~is~~ <sup>is</sup> at gets <sup>the expected</sup> an answer;  
No knowledge gained.

[ Note: if FAR = FAR<sup>2/3</sup> then  $\Pr[b = \hat{b} | \text{FAR}] \geq \frac{2}{3}$

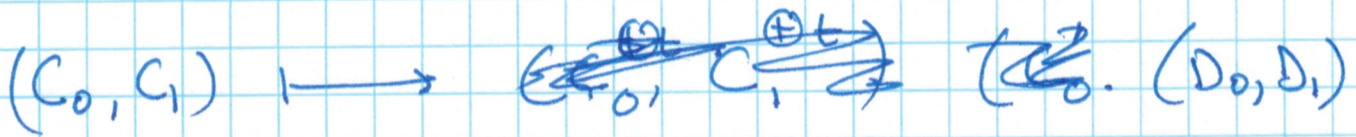
↑  
Some knowledge could be gained!

so amplification essential!

Amplifying SD:



①  $SD^{c,f} \leq SD^{c^t, f^t}$  "XOR REDUCTION"



$$D_i(b_1 \dots b_{t-1}, X_1 \dots X_t) = (C_{b_1}(X_1), \dots, C_{b_{t-1}}(X_{t-1}), C_b(X_t))$$

where  $b_t = t \oplus \left( \bigoplus_{j=1}^{t-1} b_j \right)$

[so mixes odd/even #  $C_0$ 's with rest  $C_1$ 's]

- makes it very hard to distinguish

Exercise :  $\delta(D_0, D_1) = \delta(C_0, C_1)^t$

②  $SD^{c,f} \leq SD^{t:c, \underbrace{1-2e^{-t^2/2}}_{\substack{\downarrow \\ \text{"BPP amplifier"} \\ \Delta\text{-inequality}}}}$

Putting ~~①~~ ~~②~~ ~~①~~ ① + ② + ① together

$\Rightarrow SD^{1/3, 2/3} \leq SD^{2^{-n^6}, 1-2^{-n^6}}$



Main Result for next lecture:

$SD \leq \overline{SD}$  : [Notice surprising switch]

Main tool:

•  $ED^{\Delta}$ :  ~~$(C_0, C_1)$~~  (LOW, HIGH)

HIGH =  $(C_0, C_1)$  s.t.  $H(C_0) \leq H(C_1) - \Delta$

LOW =  $(C_0, C_1)$  s.t.  $H(C_0) \geq H(C_1)$

Task: Given  $(C_0, C_1) \in$  HIGH  $\cup$  LOW  
decide which.

• Main Results

①  $SD \leq ED \rightarrow$  [mostly simple]

②  $ED \leq \overline{ED} \rightarrow$  [trivial]

③  $ED \leq SD \rightarrow$  [next time]



$$\underline{SD \leq ED}$$

$$(C_0, C_1) \longrightarrow (P, Q)$$

$$P \equiv \left( \begin{array}{l} P(x, b, s) = (C_s(x), s) \\ Q(x, b, s) = (C_s(x), b) \end{array} \right)$$

if  $(C_0, C_1)$  - far then  $C_s(x)$  reveals  $S$

$$\text{so } H(P) = V + o(1)$$

$$V \triangleq \frac{H(C_0) + H(C_1)}{2}$$

$$H(Q) = V + 1$$

if  $(C_0, C_1)$  - close then  $C_s(x)$  does not reveal  $S$

$$\Rightarrow H(P) \approx H(Q) \approx V + 1.$$