Lecture 20:

Pseudorandomness

Pseudorandom Generators

Given n random bits, generate $m >> n$ bits that "look random"

n random bits $\longrightarrow$ $\boxed{\text{PRG}}$ $\xrightarrow{m >> n \text{ "random looking" bits}}$

What is "random looking"?

- L_1 distance to uniform is small $\leftarrow$ impossible
- Useful for randomized algorithms
  - e.g. · pairwise / Kwise independent $\leftarrow$ doesn't suffice for all algorithms
  - · "Computational Indistinguishability" by ptime algorithms $\leftarrow$ good for all ptime algs but based on complexity assumption
- Kolmogorov Complexity $\approx$ incompressible $\leftarrow$ Success is undecidable

Computational Indistinguishability

sequence of bits → X "looks random" if no <u>efficient</u> (ptime, small space, low depth ckt...) algorithm can distinguish X from uniform Y, ie.

statistical test

$$\Delta(X,Y) = \max_{\substack{T \text{ that} \\ \text{is "efficiently computable"}}} \left| \Pr[X \in T] - \Pr[Y \in T] \right| \text{ is } \underbrace{\text{small}}_{\text{need to define this}}$$

<u>def.</u> Computational Indistinguishability (C.I.)

Let X_n, Y_n be sequences of r.v.'s on $\{0,1\}^n$ ← or poly(n)

We say $\{X_n\}$ $\{Y_n\}$ are $\varepsilon(n)$- indistinguishable

for time $t(n)$

if $\forall$ probabilistic poly time algorithm T (test) running in time $t(n)$

Turing machine (uniform) or circuits? (nonuniform)

$$\underbrace{\left| \Pr[T(X_n)=1] - \Pr[T(Y_n)=1] \right|}_{\text{advantage of } T} \le \varepsilon(n)$$

$\forall n$ large enough.

(Probabilities are over X_n, Y_n, coin tosses of T)

Comments

- if $\varepsilon(n)$ not specified then $\varepsilon(n) = \frac{1}{t(n)}$

- $X_n \overset{c}{\equiv} Y_n$ is notation for C.I. (w/o ε)

  if $\frac{1}{n^c}$ - indistin for time n^c $\forall c$

  equivalently:

  $\forall$ ppt T, $\exists\ \varepsilon(n) = n^{-\omega(1)}$ s.t.

  $$\left| \Pr[T(X_n)=1] - \Pr[T(Y_n)=1] \right| \leq \varepsilon(n)$$

- X_n, Y_n C.I. in nonuniform model time $t(n)$ (NCI)

  if also holds when given $\leq t(n)$ $\underbrace{\text{advice bits}}_{\text{i.e. encode in circuit}}$

Def. $\varepsilon(n)$ "negligible" if $\varepsilon(n) < \frac{1}{n^c}$ $\forall c$

Def. "Pseudorandom" X_n is p.r. if $X_n \overset{c}{\equiv} U_n$

A nice theorem: CI in nonuniform model $\Rightarrow$ K reps CI in nonuniform model

<u>Thm</u> X_n, Y_n NCI
then $\forall K = poly(n)$ X_n^K, Y_n^K are NCI
K independent copies

<u>Pf</u>.
By induction on K

$H_i = X_n^{k-i} Y_n^i$ so $H_0 = X_n^K$, $H_K = Y_n^K$ "hybrid distributions"

Consider $H_0 \dots H_n$:

$H_0 = X_n X_n \cdots X_n X_n$
$H_1 = X_n X_n \cdots X_n Y_n$
$X_n X_n \cdots X_n Y_n Y_n$
$\vdots$

$X_n Y_n \cdots Y_n$
$H_K = Y_n \cdots\cdots Y_n$

Assume for contradiction that $\exists$ ppt T s.t.

$$|\Pr[T(X_n^k)=1] - \Pr[T(Y_n^k)=1]| > \varepsilon$$

Then we'll construct ppt T' distinguishing X_n & Y_n
for infinitely many n's
(if finite, there is largest n_0
for which doesn't work so no contradiction)

How?

Note $\left|\Pr[T(X_n^k)=1) - \Pr(T(Y_n^k)=1)\right|$

$$= \left| \sum_{i=1}^{k} \left(\Pr(T(H_{i-1})=1) - \Pr(T(H_i)=1) \right) \right|$$

telescoping sum

$> \varepsilon$

$$\Rightarrow \exists\, i \text{ s.t. } \left| \Pr[T(H_{i-1})=1] - \Pr[T(H_i)=1] \right| > \varepsilon/k$$

$$T(H_{i-1}) \equiv T(X_n^{k-i} X_n Y_n^{i-1}) \qquad T(H_i) \equiv T(X_n^{k-i} Y_n Y_n^{i-1})$$

wlog assume absolute value positive

$\Rightarrow$ via averaging,

$\exists\ x_1 \cdots x_{k-i}, y_{k-i+2} \cdots y_k$ s.t.

$$\Pr[T(x_1 \cdots x_{k-i} X_n y_{k-i+2} \cdots y_k) = 1]$$
$$- \Pr[T(x_1 \cdots x_{k-i} Y_n y_{k-i+2} \cdots y_k) = 1] > \varepsilon/k$$

So define $T'(z) = T(x_1 \cdots x_{k-i}\, z\, y_{k-i+2} \cdots y_k)$

- nonuniform - put $i, x_1 \cdots x_{k-i}, y_{k-i+2} \cdots y_k$ into circuit
  i.e. many such ckts, one for each possible advice, + we are claiming at least one works so $\exists$ poly size ckt which distinguishes
- time (T') = time $(T) + O(nk)$ $\leftarrow$ extra time for advice
- $\varepsilon' = \text{adv}(T') > \varepsilon/k$

So T' distinguishes X_n + Y_n with nonnegligible probability $\rightarrow\leftarrow$

An almost nice theorem: CI in uniform model + sampleable $\Rightarrow$ k reps CI in uniform model

Thm.

if $X_n \overset{c}{\equiv} Y_n$ UCI (CI in uniform model)

+ X_n, Y_n ptime sampleable ($\exists$ ppt M s.t. $M(1^n) = X_n$)

then $X_n^k \overset{c}{\equiv} Y_n^k$

Pf. Assume $\exists$ ptime T s.t. $\Pr[T(X_n)=1] - \Pr[T(Y_n)=1] > \varepsilon$

$T'(z)$: choose $i \in_R \{1..k\}$

output $T(\overbrace{X_n^{k-i}\, z\, Y_n^{i-1}}^{H_i})$

generate via M

if $X_n^k \not\overset{c}{\equiv} Y_n^k$ then

$$\Pr[T'(X_n)=1] - \Pr[T'(Y_n)=1]$$

$$= \left(\frac{1}{k} \sum_{i=1}^{k} \Pr[T(H_{i-1})=1]\right) - \left(\frac{1}{k} \sum_{i=1}^{k} \Pr[T(H_i)=1]\right)$$

$$= \frac{1}{k} \left[\Pr[T(H_0)=1] - \Pr[T(H_k)=1]\right]$$

$$> \frac{\varepsilon}{k}$$

+ $\text{time}(T') = \text{time}(T) + k(n)\, q(n)$

time to sample X_n, Y_n

$\rightarrow\!\leftarrow$

■

Def. [Blum- Micali - Yao]

$G: \{0,1\}^{\ell(n)} \to \{0,1\}^n$ is a pseudorandom generator (PRG)

if (1) $\ell(n) < n$
(2) $G(U_{\ell(n)}) \stackrel{c}{\equiv} U_n$

G is "efficient" if poly $\underbrace{\text{time}}_{\text{in } n,}$ computable since $\ell(n)$ could be as small as 1

Comments :

- must fool all ptime statistical tests T, even those with runtime >> G's } useful for crypto
- can generalize to ε-PRG against nonuniform machines running in time $t(n)$
- create 1-time-pads
- recall earlier lecture :

method of enumeration (try all random seeds + take majority vote)

$$\Rightarrow \text{BPP} \subseteq \bigcup_c \text{DTIME}\left(2^{\ell(n)^c} n^c\right)$$

i.e. if $\ell(n) = O(\log n)$ BPP = P

Can we prove existence of PRG's?
well, if we could, then we would also be able to prove $P \neq NP$!

Thm efficient PRGs exist $\Rightarrow P \neq NP$

Pf. (show $P = NP \Rightarrow$ no PRG's)
Assume $G: \{0,1\}^{\ell(n)} \to \{0,1\}^n$ + can be computed in ptime + $\ell(n) < n$
(G is proposed efficient prg)
Given G let

$$T(x) = \begin{cases} 1 & \text{if } \exists y \text{ st. } G(y) = x \\ 0 & \text{o.w.} \end{cases}$$

note, $T \in NP$ since can guess y + verify via G since G is ptime

$$\Pr_{x \in U_{\ell(n)}}[T(G(x)) = 1] = 1$$

$$\Pr_{y \in U_n}[T(y) = 1] \leq \frac{2^{\ell(n)}}{2^n} \leq \frac{1}{2} \qquad \text{since } \ell(n) < n$$

if $P = NP$ then $T \in P$ + can distinguish the two distributions
but then G is not PRG $\rightarrow\leftarrow$

■

so need complexity theoretic assumptions to construct PRG's.