Today
- Error Redn (rnd vs pairwise)
- Sampling
- Fooling linear tests.

CS3.413.1
Pseudorandomness
Lecture 05 (2021-9-7)
Instructor: Prahladh Harsha.

Recap from last time

(1) Const of pw hash families.

Thm. $\forall m,n$, there exists a pw rnd family of hash functions $\mathcal{H}_{m,n}$ that requires at most $2\max\{m,n\}$ bits to specify any fn $h \in \mathcal{H}_{m,n}$

$$\mathcal{H}_{m,n} \subseteq \{h : \{0,1\}^n \to \{0,1\}^m\}.$$

(2) Tail Bounds:

$X_1, \ldots X_t$ - $[0,1]$-valued r.v

$$\bar{X} = \sum X_i / t; \quad \mu = \mathbb{E}[\bar{X}]$$

Chernoff: X_i's independent

$$\Pr[|\bar{X} - \mu| > \varepsilon] \le 2e^{-t\varepsilon^2/4}$$

Chebyshev: X_i's are pairwise independent

$$\Pr[|\bar{X}-\mu| > \varepsilon] \le \frac{1}{t\varepsilon^2}.$$

Error Reduction of BPP Algorithms

BPP Alg that uses m-random bits
error $\le 1/3$

$\downarrow$

Reduce error $1/3$ to $1/2^k$

$(1/3 \to 2^{-k})$

	#repetitions	#random bits
Independent	$O(k)$	$O(km)$
Pairwise Independent	$O(2^k)$	$\underbrace{O(k+m)}_{(*)}$.

(*) t-pairwise independent samples

$H_{m,n}$ $t = 2^n$

$h: \{0,1\}^n \to \{0,1\}^m$

$t = O(2^k)$;

random bits

$= O(m+n)$

$= O(m + \log t)$

$= O(m + O(k))$

$X_1 \dots X_t \in \{0,1\}^m$ - pw ind

$H \leftarrow \mathcal{H}_{m,n}$

$$\underbrace{H(0^n),\ H(0 0 \dots 1),\ \dots\ H(1 \dots 1)}$$

t - pairwise independent samples in $\{0,1\}^m$

use $\mathcal{H}_{m, \log t}$.

Sampling:

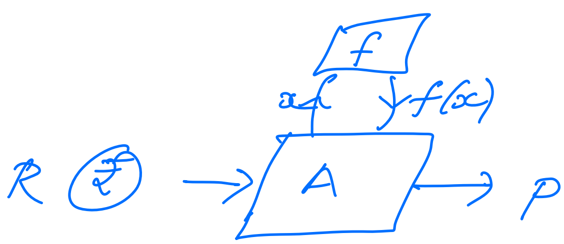
Problem: Given an oracle

$$f : \{0,1\}^m \to [0,1]$$

Compute (estimate)

$$\mu = \mathbb{E}[f(U_m)] = \mathbb{E}_{x \leftarrow_u \{0,1\}^m}[f(x)]$$

to within an additive approximation of ε.



Guarantee:

$p \in (\mu - \varepsilon, \mu + \varepsilon)$.

(with high probability)

A is ε-additive approximate sampler.

Ind. Sampler:

1. Choose $x_1 \dots x_t \leftarrow_U \{0,1\}^m$
2. Query Oracle f at $x_1 \dots x_t$
3. Output $\sum_{i=1}^{t} f(x_i) / t$.

Chernoff Bound.

$$\Pr_{x_1 \dots x_t}\left[\left|\frac{\sum f(x_i)}{t} - \mu\right| > \varepsilon\right] \le 2e^{-t\varepsilon^2/4}$$

$$= \delta$$

(set $t = O(\frac{1}{\varepsilon^2} \log \frac{1}{\delta})$).

Thm: Ind-sampler. has the following property.

- $\Pr_{R = x_1 \dots x_t}\left[\left|\text{Ind-Sampler}^f(R) - \mu\right| > \varepsilon\right] \le \delta.$

using $t = O(\frac{1}{\varepsilon^2} \log \frac{1}{\delta})$ - samples

- $O(\frac{1}{\varepsilon^2} \log(\frac{1}{\delta}) \cdot m)$ - random bits.

Pairwise Independent Sampler.

1. Set n s.t. $t = 2^n$
2. Pick $H \leftarrow \mathcal{H}_{m,n}$.
3. Set $x_1 \ldots x_t \leftarrow H(0^n), \ldots H(1^n)$
4. Query f at $x_1 \ldots x_t$.
5. Output $\sum f(x_i)/t$.

Thm: Pairwise Ind sampler.

$$\Pr_{(x_1 \ldots x_t) \; R} \left[\left| \text{Pairwise-Sampler}^f(R) - \mu \right| > \varepsilon \right] \leq \delta$$

by $t = \frac{1}{\varepsilon^2 \delta}$ - samples.

$\# \; O\left(m + \log\left(\frac{1}{\varepsilon}\right) + \log\left(\frac{1}{\delta}\right)\right)$ - random bits.

Epsilon-biased Distributions.

Recall: MAXCUT examples

Analysis: pairwise independence of underlying r.v.s.

Obs: Suff to sample random coins from a pw ind dist rather than independent.

In general, "property" of random coin used by alg.

Qn: Is there a smaller space of random coins that has this property?

Property: Linear tests / Linear functions

Linear function:

$\ell : \{0,1\}^n \to \{0,1\}$ is a linear function if there exists a set $S \subseteq [n]$ st

$$\ell(x_1, \ldots x_n) = \bigoplus_{i \in S} x_i$$

$$= \sum_{i \in S} x_i \pmod 2$$

(In this case, we will denote ℓ by ℓ_S)

$\ell_\emptyset \equiv 0$.

$$\Pr_{x_1 \ldots x_n \leftarrow U_n}\left[\ell_S(x_1 \ldots x_n) = 0\right] = \begin{cases} 1 & \text{if } S = \phi \\ 1/2 & \text{if } S \neq \phi \end{cases}$$

Distribution D on $\{0,1\}^n$ is ε-biased if for all linear functions ℓ_S.

$$\left| \Pr_{x_1 \ldots x_n \leftarrow U_n}\left[\ell_S(x_1 \ldots x_n) = 0\right] - \Pr_{x_1 \ldots x_n \leftarrow D}\left[\ell_S(x_1 \ldots x_n) = 0\right] \right| \leq \varepsilon.$$

ie,
$$\left| \Pr_{x_1 \ldots x_n \leftarrow D}\left[\ell_S(x_1 \ldots x_n) = 0\right] - \frac{1}{2} \right| \leq \varepsilon \quad \forall S \neq \phi$$

U_n - 0-biased.
but ~~it~~ needs n-random bits to generate.

—

Construction of ε-biased distributions [Alon, Goldreich, Håstad, Peralta]

Uses finite fields $GF(2^k)$

① **Degree Mantra:**

$p(x) \in \mathbb{F}_{\le d}[X]$ - degree d univariate non-zero poly

$$\Pr_{x \leftarrow \mathbb{F}}[P(x)=0] \le \frac{d}{|\mathbb{F}|}$$

② **Non-trivial linear functions are unbiased.**

$S \neq \phi$ $$\Pr_{\bar{x} \leftarrow U_n}[\ell_S(x_1 \ldots x_n)=0] = \frac{1}{2}$$

AGHP Construction of ε-biased $D \sim \{0,1\}^n$

Input: n, ε

0. Choose a field $\mathbb{F} = GF(2^k)$ where $2^k \ge \frac{n}{\varepsilon}$

\# rand bits $= 2k = O(\log n + \log \frac{1}{\varepsilon})$

1. Pick $Y, Z \leftarrow_U \mathbb{F}$ ($2k$ bits of randomness)

2. Set $x_0, \ldots x_{n-1}$ as follows

$$x_i = \langle Y^i, Z \rangle$$

$$\ell_S(x) = \langle \sum_{i \in S} Y^i, x \rangle$$

$$Y^i = \underbrace{Y \cdot Y \cdots Y}_{i\text{-times multiplication}}$$

$$Y^i \in GF(2^k) \cong \{0,1\}^k \qquad \langle Y, Z \rangle$$

$$Z \in GF(2^k) \cong \{0,1\}^k \qquad = \sum Y_i \cdot Z_i \pmod 2$$

$$GF(2^k) = \mathbb{F}_2[X] / \langle f(X) \rangle \quad \text{where}$$

$f \in \mathbb{F}_2[X]$ is a deg k irreducible.

$$\mathbb{F}_2^k$$

$$\emptyset \neq S.$$

$$\ell_S(X) = \sum_{i \in S} X_i = \sum_{i \in S} \langle Y^i, Z \rangle$$

$$= \left\langle \sum_{i \in S} Y^i, Z \right\rangle$$

$$\Pr_{X \leftarrow D} [\ell_S(X) = 0] = \Pr_{Y,Z} \left[\left\langle \sum_{i \in S} Y^i, Z \right\rangle = 0 \right]$$

$$p \triangleq \Pr_{Y} \left[\sum_{i \in S} Y^i = 0 \right] \leq \frac{n}{|\mathbb{F}|}$$

(degree mantra since $S \neq \emptyset$)

$$\Pr_{Y,Z}\left[\left\langle \sum_{i\in S} Y^i, Z\right\rangle = 0\right]$$

$$= \Pr_{Y}\left[\sum_{i\in S} Y^i = 0\right]\cdot \Pr_{Z}\left[\left\langle \sum_{i\in S} Y^i, Z\right\rangle = 0 \,\middle/\, \sum_{i\in S} Y^i = 0\right]$$

$$+ \Pr_{Y}\left[\sum_{i\in S} Y^i \neq 0\right]\cdot \Pr_{Z}\left[\left\langle \sum_{i\in S} Y^i, Z\right\rangle = 0 \,\middle/\, \sum_{i\in S} Y^i \neq 0\right]$$

$$= p\cdot 1 + (1-p)\cdot \frac{1}{2}$$

$$= \frac{1}{2} + \frac{p}{2}.$$

Hence

$$\left|\Pr\left[g(X) = 0\right] - \frac{1}{2}\right| = \frac{p}{2} \leq \frac{\varepsilon}{2}.$$

Thus X is ε-biased $\boxtimes$