]evensidemargin=0.15in

6.842 Randomness and Computation May 5, 2008

Lecture 23

Lecturer: Ronitt Rubinfeld *Scribe: Alex Cornejo*

Recall

Definition 1 *Function* $\varepsilon(n)$ *is* negligible *if* $\varepsilon(n) < \frac{1}{n^c}\ \ \forall c$. *Let* $f : \{\pm 1\} \to \mathbb{R}$ *then* $L_1(f) = \sum_S \left|\hat{f}(S)\right|$

Definition 2 $L \in BPP$ *if there exists a probabilistic polynomial time algorithm* $\mathcal{A}$ *such that.*

- $x \in L \Rightarrow \Pr[\mathcal{A}(x) \text{ accepts}] \geq \frac{2}{3}$
- $x \notin L \Rightarrow \Pr[\mathcal{A}(x) \text{ accepts}] \leq \frac{1}{3}$

Definition 3 Statistical distance

$$\Delta(X, Y) = \max_{T \subseteq S} |\Pr[x \in T] - \Pr[x \in S]|$$

Plan for Today

- Computational Indistinguishability
- Pseudorandom Generators (and derandomizing BPP)
- Unpredictability

  n random bits. → PRG → $m >> n$ random bits.

  How should we measure the amount of randomness? L_1 distance?, Kolmogorov Complexity?, we will focus on computational indistinguishability.

Computational Indistinguishability

Definition 4 (Computational Indistinguishability (C.I.)) *Let* X_n *and* Y_n *be sequences of random variables on* $\{0,1\}^n$. *We say the collections* $\{X_n\}, \{Y_n\}$ *are "*$\varepsilon(n)$*-indistinguishable for time* $t(n)$*" if* $\forall$ *probabilistic* $t(n)$*-time algorithm* T, $|\underbrace{\Pr[T(X_n) = 1] - \Pr[T(Y_n) = 1]}_{\text{advantage of } T}| \leq \varepsilon(n)$, $\forall n > n_0$ *for some* n_0.

- *If* $\varepsilon(n)$ *not specified then* $\varepsilon(n) = \frac{1}{t(n)}$
- $X_n \stackrel{c}{\equiv} Y_n$ *used for C.I.*
- *It is stronger to say that* T *is nonuniform, i.e.* $t(n)$ *size circuits.*
- *N.C.I. used for non-uniform C.I., which means that it also holds when given* $\leq t(n)$ *bits of advice.*

Definition 5 (Pseudorandom (P.R.)) X_n *is pseudo-random if* $X_n \stackrel{c}{\equiv} U_n$.

Some nice theorems:

Theorem 6 *If* X_n, Y_n *are N.C.I., then* $\forall k = poly(n)$ $\underbrace{X_n^k, Y_n^k}_{k \text{ independent copies}}$ *are N.C.I.*

Theorem 7 *If* X_n, Y_n *are C.I., and* X_n, Y_n *are polytime sampleable then* $X_n^k \stackrel{c}{\equiv} Y_n^k$.

Definition 8 (PRG) *[Blum-Micali-Yao]* $G : \{0,1\}^{\ell(n)} \to \{0,1\}^n$ *is a pseudo-random generator if* $\ell(n) < n$ *and* $G(U_{\ell(n)}) \stackrel{c}{\equiv} U_n$. G *is "efficient" if computable in time poly*(n).

Theorem 9 *If there is an efficient PRG against time* n *with seed length* $\ell(n)$ *then* $BPP \subseteq \bigcup_c DTIME(2^{\ell(n^c)} n^c)$.

In particular, using this theorem we get several interesting results by assuming different values of $\ell(n)$, for example:

Theorem 10 *There exists a PRG against nonuniform time $t(n)$ with seed length $O(\log t(n))$.*

However, notice that the theorem does not say if it is efficiently computable, and therefore it does not imply that $BPP = P$.

Theorem 11 *If there exists an efficient PRG then $P \neq NP$.*

Proof We prove the contrapositive of the statement, that is if $P = NP$ then no efficient PRG exists.

Fix G and define $T(x)$ as:

$$T(x) = \begin{cases} 1 & \text{if } \exists y \text{ such that } G(y) = x \\ 0 & \text{otherwise} \end{cases}$$

The test $T(x)$ is such that $Pr_{x \in G(U_{\ell(n)})}[T(x) = 1] = 1$ and $Pr_{x \in U_n}[T(x) = 1] \leq \frac{2^{\ell(n)}}{2^n} \leq \frac{1}{2}$. Therefore T distinguishes distributions $G(U_{\ell(n)})$ and U_n with advantage $\geq \frac{1}{2}$.

If we assume that G is efficiently computable, notice that $T \in NP$ since we can guess y and verify $G(y) = x$ in polynomial time since G. Therefore if $P = NP$ then T is an efficiently computable test that distinguishes G from the uniform distribution, which means that G is not efficiently computable – a contradiction. ■

Next-bit Unpredicatble

Definition 12 *Next-bit unpredictable Let $\mathbb{X} = (X_1, \ldots, X_n)$ be a distribution on $\{0,1\}^n$. We say $\mathbb{X}$ is next bit unpredictable if for every probabilistic polynomial time algorithm A there is a negligible function $\varepsilon(n)$ such that.*

$$\Pr_{x,i,coins\ of\ P}[P(X_1, \ldots, X_n) = X_i] \leq \frac{1}{2} + \varepsilon(n)$$

Notice that if $\mathbb{X}$ where the uniform distribution then $\varepsilon(n) = 0$.

Theorem 13 *$\mathbb{X}$ is pseudo-random if $\mathbb{X}$ is next-bit unpredictable.*

Proof

- If $\mathbb{X}$ is next-bit unpredictable $\Rightarrow$ $\mathbb{X}$ is not pseudo-random.

  Assume

  $$\Pr_{x,i,\text{coins of P}}[P(X_1, \ldots, X_n) = X_i] \geq \frac{1}{2} + \frac{1}{n^k}$$

  In particular this means that $\exists i$ such that

  $$\Pr_{x,\text{coins of P}}[P(X_1, \ldots, X_n) = X_i] \geq \frac{1}{2} + \frac{1}{n^k}$$

  We know define the statistical test $T(y_1, \ldots, y_n)$ as

  $$T(y_1, \ldots, y_n) = \begin{cases} 0 & \text{if } P(y_1, \ldots, y_n) \neq y_i \\ 1 & \text{if } P(y_1, \ldots, y_n) = y_i \end{cases}$$

  So the probability that T passes is $\Pr_{y \in \mathbb{X}}[T \text{ passes}] \geq \frac{1}{2} + \frac{1}{n^k}$, and $\Pr_{y \in U_n}[T \text{ passes}] = \frac{1}{2}$.

  Therefore T distinguishes $\mathbb{X}$ and U_n with advantage $\geq \frac{1}{n^c}$, which means that $\mathbb{X}$ is not pseudo-random.

- If $\mathbb{X}$ is not pseudo-random $\Rightarrow$ exists next-bit test.

  Not enough time to prove in this lecture, but here is the outline:

  - Use **hybrid** argument to construct next-bit predictor P
  -
  $$\begin{array}{rrlll} U = & D_0 = & U_1, & \ldots, & U_n \\ & D_1 = & X_1, U_2, & \ldots, & U_n \\ & \vdots & & & \\ X = & D_n = & X_1, & \ldots, & X_n \end{array}$$
  - If distance between U and X is ε then there exists D_i, D_j with distance $\geq \varepsilon/n$.

■