CS 880: Pseudorandomness and Derandomization 5/6/2013

Lecture 28: Pseudorandomness for Half-spaces

Instructors: Holger Dell and Dieter van Melkebeek Scribe: Nick Pappas

DRAFT

1 Preliminaries

Here are the notions that we need:

Definition 1 (Half-Space). *A half-space is defined as a function $H_{w,\theta} : \{-1,1\}^n \to \{-1,1\}$ for some vector $w \in \mathbb{R}^n$ and some constant threshold value $\theta \in \mathbb{R}$. For any vector $x \in \mathbb{R}^n$, we define $H_{w,\theta}(x) = sign(\langle w, x\rangle - \theta)$.*

I.e, we report 1 if the projection of x onto v is greater than θ, and we report -1 if the projection is less than θ.We assume that the length of w will be 1 ($\|w\|_2 = 1$).

Definition 2 (Regular Half-Space). *(We remind the definition of the supreme norm denoted as $\|w\|_\infty$):*
$\|w\|_\infty \doteq \max |w_i|$

A half-space is regular if $\|w\|_\infty \leq \epsilon$

Definition 3 (Supreme distance). *We define the supreme distance denoted as $d_\infty(A, B)$: $d_\infty(A, B) \doteq \sup_{t\in R} |\Pr(A < t) - \Pr(B < t)|$*

Theorem 1. *Barry-Esseen Theorem*
Let $Y_1, \ldots, Y_n$ be independent random variables with the following properties:

1. $\mathrm{E}[Y_i] = 0$
2. $\sum_i E[Y_i^2] = 1$
3. $\sum_i E[Y_i^4] \leq \varepsilon^2$

Let $S_n = Y_1 + \ldots + Y_n$ and let $N(01)$ denote the normal distribution with mean 0 and variance 1. Then the supreme distance $d_\infty(S_n, N(01)) \leq \varepsilon$

Corollary 2. *Corollary of the Barry-Esseen Theorem*
$d_\infty(< w, U_n >, N(01)) \leq \varepsilon$

2 PRG construction attempt

Let $Y_i = wX_i$ where X_i is uniform in -1,1. We pick X_is from a 4-wise uniform distribution and we end up with :
$X_1, X_2, ..., X_n$
However picking X_is like that gives us a problem; the Y_is will fail the third condition of the Barry -Esseen Theorem. Therefore we split them in t groups:

$$\underbrace{\left(X_1, ..., X_{\frac{n}{t}}\right)}_{D^1} \underbrace{\left(X_{\frac{n}{t}+1}, ...\right)}_{D^2} ...... \underbrace{\left(.., X_{\frac{n}{t}}\right)}_{D^t}$$

The seed length is $\approx t \log \frac{n}{t}$

Theorem 3. $d_\infty(S_n, N(01)) \leq \varepsilon$

Proof. We use the B-E theorem:

$$w = \underbrace{\left(w_1.....w_{\frac{n}{t}}\right)}_{w^1} \underbrace{\left(w_{\frac{n}{t}}.....\right)}_{w^2} ...... \underbrace{(.....w_n)}_{w^t}$$

Now let :
$Y_i \doteq < D^i, w^i >$ then $S_t =< w, D >$ What we have now is the following:

1. $E[Y_i] = \underbrace{\sum_j {w^i}_j E[{D^i}_j]}_{0} = 0$

2. $\sum_i E[{Y_i}^2] = \sum_i \sum_{j_1+j_2} \underbrace{E[{w^i}_{j_1} {D^i}_{j_1}] E[{w^i}_{j_2} {D^i}_{j_2}]}_{j_1 \neq j_2 \rightarrow 0} + \sum_i \sum_j \left({w^i}_{j_1}\right)^2 \underbrace{E[\left({D^i}_j\right)^2]}_{1} = \sum_k {w_k}^2 = ||w|| = 1$

3. $\sum_i E[{Y_i}^4] = \sum_{j_1+j_2+j_3+j_4} {w_{j_1}}^i ... {w_{j_4}}^i E[Dj_1...Dj_4] \leq 3 \sum_{j,k} \left(w^i_j\right)^2 \left(w^i_k\right)^2 = 3\left(\sum_j \left(w^i_j\right)^2\right)^2 = 3||w^i||_2{}^4$

For the last part we require that $\sum_t ||w^i||_2{}^4 \leq \varepsilon^2$ for which it is sufficient that $||w^i||_2{}^4 \leq \frac{\varepsilon^2}{t}$

For example: $w^1 = (\varepsilon.....\varepsilon)$ will give us :

$$||w^i||_2{}^4 = \left(\frac{n}{t}\right)^2 \varepsilon^4 \leq \frac{\varepsilon^2}{t} \Leftrightarrow t \geq n^2 \varepsilon^2$$

∎

Idea: We pick the t partitions of the w^is at random using a hash function $h[n] \rightarrow [t]$. Now the last part becomes :

$$....... \leq 3 \sum_{j,k} \underset{h}{E} \left(w^i_j\right)^2 \left(w^i_k\right)^2 = 3\left(\sum_j \underset{h}{E} \left(w^i_j\right)^2\right)^2 = 3 \underset{h}{E} ||w^i||_2{}^4$$

Theorem 4. $\underset{h}{E}\left[||w^i||_2{}^4\right] \leq \Theta\left(\frac{\varepsilon^2}{t}\right)$

Proof. Let $H_j = \begin{cases} 1 & h(j) = i \\ 0 & otherwise \end{cases}$
Then

$$\underset{h}{E}\left[||w^i||_2{}^4\right] = \sum_j^n H_j{}^4 w_j{}^4 + \sum \underbrace{E\left[H_j H_k\right] w_j{}^2 w_k{}^2}_{\frac{1}{t^2}} \leq \frac{\varepsilon^2}{t} + \frac{1}{t^2}$$

and so finally:

$$||w^i||_2{}^4 \leq \frac{\varepsilon^2}{t} + \frac{1}{t^2} \underbrace{=}_{t=\frac{1}{\varepsilon^2}} \Theta\left(\frac{\varepsilon^2}{t}\right)$$

□

Now PRG $D' = (D, h)$ satisfies

$$d_\infty\left(< w, D' >, N(01)\right) \leq \varepsilon \Rightarrow d\left(H_{w,\Theta}(U), H_{w,\Theta}(D')\right) \leq O(\varepsilon)$$

and the seed length of D' is

$$t \log \frac{n}{t} + \log(nt) \sim \varepsilon^2 \log \frac{1}{\varepsilon} \log n$$