Lecture 5: Sept 6, 2021

Lecturer: Eshan Chattopadhyay *Scribe: Andrew Chen*

1 Lower bound on randomness for k-wise independence

Let D be any k-wise independent distribution on $\{0,1\}^n$. Define $D(x) = \Pr[D = x]$ and

$$\sup(D) = \{x \in \{0,1\}^n : D(x) > 0\}$$

We claim that $|\sup(D)| \geq n^{k/2}$. In particular, this means that we need $\geq \frac{1}{2}k \log n$ random bits to generate D. Compare this to our construction!

Proof. For brevity, let $S = \sup(D)$. View the distribution as a real valued function $D : S \to \mathbb{R}^+$. Let

$$V = \{f : S \to \mathbb{R}\}$$

be the vector space of functions from S to $\mathbb{R}$. Clearly $\dim(V) = |S|$, since the collection of indicator functions $\{e_y\}_{y \in S}$ given by

$$e_y(x) := \begin{cases} 1 & \text{if } x = y \\ 0 & \text{else} \end{cases}$$

is a basis for V.
Define an inner product $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ by

$$\langle f, g \rangle := \mathbb{E}_{x \sim D}[f(x) \cdot g(x)] = \sum_{x \in S} D(x) f(x) g(x).$$

We can easily verify that it is an inner product; for any $\alpha, \beta \in \mathbb{R}$ and $f_1, f_2, g \in V$, we have

$$\langle \alpha f_1 + \beta f_2, g \rangle = \sum_{x \in S} D(x) \left(\alpha f_1 + \beta f_2\right) g(x)$$
$$= \alpha \sum_{x \in S} D(x) f_1(x) g(x) + \beta \sum_{x \in S} D(x) f_2(x) g(x) = \alpha \langle f_1, g \rangle + \beta \langle f_2, g \rangle,$$

and similarly for the second coordinate. It is also clear that

$$\langle f, f \rangle = \sum_{x \in S} D(x) f(x)^2 \geq 0$$

since $D(x) \geq 0$, with equality if and only if $f(x) = 0$ for all $x \in \sup(D)$ (i.e. if $f = 0$).
Next we define a collection of $|S|$ orthogonal functions in V with respect to the inner product. For all subsets $T \subseteq [n]$ with $|T| \leq k/2$, define $\chi_T : S \to \mathbb{R}$ by

$$\chi_T(x) := \prod_{i \in T} (-1)^{x_i}$$

recalling that $x = (x_1, x_2, \cdots, x_n) \in \{0,1\}^n$ is an n-tuple. We claim that the collection $\{\chi_T : T \subseteq [n], |T| \leq k/2\}$ is orthogonal. To see this, select distinct $T_1, T_2 \subseteq [n]$ with $|T_1|, |T_2| \leq k/2$. Expanding the definition of the inner product gives

$$\langle \chi_{T_1}, \chi_{T_2} \rangle = \mathbb{E}_{x\sim D}[\chi_{T_1}(x)\chi_{T_2}(x)] = \mathbb{E}_{x\sim D}\left[\prod_{i\in T_1}(-1)^{x_i}\prod_{j\in T_2}(-1)^{x_j}\right].$$

Observe that the terms with $i \in T_1 \cap T_2$ will cancel each other out; in particular, we have

$$\mathbb{E}_{x\sim D}\left[\prod_{i\in T_1}(-1)^{x_i}\prod_{j\in T_2}(-1)^{x_j}\right] = \mathbb{E}_{x\sim D}\left[\prod_{i\in T_1\Delta T_2}(-1)^{x_i}\right].$$

Now note that $|T_1 \Delta T_2| \leq |T_1| + |T_2| \leq \frac{k}{2} + \frac{k}{2} = k$. Since D is k-wise independent, we have

$$\mathbb{E}_{x\sim D}\left[\prod_{i\in T_1\Delta T_2}(-1)^{x_i}\right] = \left[\prod_{i\in T_1\Delta T_2}\mathbb{E}_{x_i\sim D_i}(-1)^{x_i}\right] = 0$$

as D_i, the marginal distribution of D on the i-th component, is uniform on $\{0,1\}$.
Since the collection $\{\chi_T\}$ is orthogonal, its cardinality provides a lower bound on the dimension of V. In particular,

$$|S| = \dim(V) \geq \binom{n}{k/2} \geq n^{k/2}.$$

□

Remark 1.1. *In fact, there are explicit constructions which show that this bound is tight.*

2 Pseudorandom Generators

The motivating idea behind pseudorandom generators is to provide a derandomization black-box for algorithm design.

Definition 2.1. *A family of pseudorandom generators (PRG) is given by the collection*

$$\{G_n : \{0,1\}^{s(n)} \to \{0,1\}^n\}_{n\in\mathbb{N}}.$$

Definition 2.2. *A family (class) of Boolean functions is*

$$\mathcal{F} = \bigcup_{n\geq 0} F_n, \qquad F_n \subseteq \{f : \{0,1\}^n \to \{0,1\}\}.$$

Definition 2.3. *For $\varepsilon = \varepsilon(n)$, we say that $\{G_n\}$ is an ε-PRG for the Boolean family $\{F_n\}$ with seed length $s(n)$ if for all $n \geq 0$ and $f \in F_n$,*

$$\left|\mathbb{E}[f(U_n)] - \mathbb{E}\left[f(G_n(U_{s(n)}))\right]\right| < \varepsilon(n).$$

Computing PRGs

We say that A is mildly explicit if A runs in $\text{poly}(n, 2^{s(n)})$. (This is sufficient for our purposes, since our derandomization technique is to run G on all $2^{s(n)}$ inputs.)
We say that A is strongly (or fully) explicit if A runs in $\text{poly}(n, S(n))$. (This one is necessary for cryptographic purposes.)

Example 2.4. *(k-Juntos) For fixed $n \geq 1$, and consider the collection of functions $F_n \subseteq \{f : \{0,1\}^n \to \{0,1\}\}$ which depends on at most k input bits. Denote this family by $F^{k\text{-Junta}}$. We claim that k-wise independence fools $F^{k\text{-Junta}}$.*
Indeed, if $f \in F_n$ for some $n \geq 1$,we have

$$\mathbb{E}[f(U_n)] = \mathbb{E}[f(G_n(U_{s(n)}))]$$

since f ignores all but k bits (by its membership in F_n).