Lecture 8: Derandomization

(Mon, 2/24/2025 - Lecture 1)

Ex. Let $\{X_i\}_{i=1}^n$ be iid random variables with $\mathbb{E}[X_i] = \mu$ and $\sigma^2 = \mathbb{V}[X_i]$

Then approximate $\frac{1}{n}\sum_{i=1}^n X_i$ by μ with error $\sigma/\sqrt{n}$ (w.p. $\frac{3}{4}$)

Derandomization: Can we do this without randomness?

Yes, we can! Let $\{X_i(x)\}_{i=1}^n$ be a family of pairwise independent random variables, where $x \in \{0,1\}^n$ is the seed.

Then $\mathbb{E}[X_i(x)] = \mu$ and $\mathbb{V}[X_i(x)] = \sigma^2$

Let $\mathbb{E}[X_i(x)X_j(x)] = \mu^2 + \sigma^2 \cdot \mathbb{1}[i=j]$

Then $\mathbb{E}_x\left[\left(\frac{1}{n}\sum_{i=1}^n X_i(x)\right)^2\right] = \mu^2 + \frac{\sigma^2}{n}$

So $\mathbb{V}_x\left[\frac{1}{n}\sum_{i=1}^n X_i(x)\right] = \frac{\sigma^2}{n}$

So we can still do error $\sigma/\sqrt{n}$ (w.p. $\frac{3}{4}$) with 2^n seeds.

Idea: Use $\{X_i(x)\}_{i=1}^n$ to derive $\{Y_i(x)\}_{i=1}^n$ that are iid.

Then $\mathbb{E}[Y_i(x)] = \mu$ and $\mathbb{V}[Y_i(x)] = \sigma^2$

Then we can still do error $\sigma/\sqrt{n}$ (w.p. $\frac{3}{4}$) with $2^{\log n}$ seeds.

Idea: $\{Y_i(x)\}_{i=1}^n = \{X_i(x)\}_{i=1}^n$

Then $\mathbb{E}[Y_i(x)] = \mu$ and $\mathbb{V}[Y_i(x)] = \sigma^2$

But $\mathbb{E}[Y_i(x)Y_j(x)] = \mu^2 + \sigma^2 \cdot \mathbb{1}[i=j]$

So $\mathbb{V}_x\left[\frac{1}{n}\sum_{i=1}^n Y_i(x)\right] = \frac{\sigma^2}{n}$

Pairwise independence

Let $x \in \{0,1\}^n$ be a seed.

Let $\{X_i(x)\}_{i=1}^n$ be pairwise independent random variables.

Define $Y_i(x) = \sum_{j \in S_i} X_j(x)$ where $S_i \subseteq [n]$ and $|S_i| = \log n$

Then $\mathbb{E}[Y_i(x)] = \sum_{j \in S_i} \mathbb{E}[X_j(x)] = \mu \log n$

Also $\mathbb{V}[Y_i(x)] = \sum_{j \in S_i} \mathbb{V}[X_j(x)] = \sigma^2 \log n$

Also $\mathbb{E}[Y_i(x)Y_j(x)] = \mu^2 \log^2 n + \sigma^2 \log n \cdot \mathbb{1}[i=j]$

So $\mathbb{V}_x\left[\frac{1}{n}\sum_{i=1}^n Y_i(x)\right] = \frac{\sigma^2 \log n}{n}$

So we can use $\log n$ seeds.

Idea: Use $\{X_i(x)\}_{i=1}^n$ to derive $\{Y_i(x)\}_{i=1}^n$ that are iid.

Then $\mathbb{E}[Y_i(x)] = \mu \log n$ and $\mathbb{V}[Y_i(x)] = \sigma^2 \log n$

But $\mathbb{E}[Y_i(x)Y_j(x)] = \mu^2 \log^2 n + \sigma^2 \log n \cdot \mathbb{1}[i=j]$

So $\mathbb{V}_x\left[\frac{1}{n}\sum_{i=1}^n Y_i(x)\right] = \frac{\sigma^2 \log n}{n}$