# CS 744: Pseudorandomness Generators

Lecture 9: Basic Derandomization Techniques - IV

January 25, 2015

## 1 Pairwise Independent Hash Functions

**Theorem 1.** *Let $\mathbb{F}$ be a finite field. Then the family of functions*

$$\mathcal{H} = \{h_{a,b} : \mathbb{F} \to \mathbb{F} \mid a, b \in F\},$$

*where $h_{a,b} = ax + b$, is an explicit family of pairwise independent hash functions.*

*Proof.* Given $(x_1, y_1)$ and $(x_2, y_2)$ in $\mathbb{F}^2$, there is exactly one line passing through them - the one with slope $a = (y_2 - y_1)/(x_2 - x_1)$ and $y$-intercept $b = y_1 - ax_1$. There are $|\mathbb{F}|^2$ distinct lines, and the probability of picking this line $ax + b$ as the hash function is precisely $\frac{1}{|\mathbb{F}|^2}$. Thus $\mathcal{H}$ is pairwise independent.

Homework: Show that $\mathcal{H}$ is explicit. □

## 2 Randomness-efficient error reduction and sampling

For a BPP algorithm, we can reduce the error probability to $2^{-k}$ by using $k$ repetitions, as we can prove using the Chernoff bound. Suppose we have an algorithm that works with pairwise independent random sources. Can we reduce its error by repeating it independently many times? The following tail inequality answers this question.

**Lemma 2.** *Pairwise Independence Tail Inequality Let $X_1, \ldots, X_n$ be a sequence of pairwise independent random variables with values in $[0, 1]$. If we denote by $S$, their average, $\sum X_i/n$, then*

$$Pr[|X - EX| \geq \epsilon] \leq \frac{1}{n\epsilon^2}.$$

*Proof.* The variance of $\sum_i X_i$ is the sum of the variances of the individual $X_i$s, by pairwise independence. Hence $Var(S)$ is equal to $\frac{1}{n^2}\sum_i Var(X_i)$, which is at most $1/n$ since $X_i$s are bounded above in value by 1. The inequality follows by the Chebyshev inequality. □

**Discussion**

## 2.1 Sampling

We consider the following problem of *sampling a function*: Given a black-box access to a function $f : \{0,1\}^m \to [0,1]$, is it possible to efficiently approximate $Ef$ to within an additive error of $\epsilon > 0$? One general strategy is to sample $f$ at a few inputs, and output the sample average as an approximation to $Ef$.

It is possible to sample a set of size $O(\log(1/\delta)/\epsilon^2)$ to solve the problem with probability greater than $1-\delta$. We can prove using Chernoff bounds, that this strategy works. This sampling involves truly random bits.

We now try to replace the truly random bits by pairwise independent bits. We will need a sample of $t = 1/(\epsilon^2\delta)$ pairwise independent samples from $\{0,1\}^m$. To generate $t$ pairwise independent samples of $m$ bits each, we require $O(m + \log t) = O(m + \log(1/\epsilon) + log(1/\delta))$ pure random bits.

The following sampling technique will be nonadaptive — that is, the sequence of questions asked depend only on the random coin tosses, and in particular, questions asked will not depend on the answers to previous questions. Moreover, the answer will simply be the average of the outputs. No further postprocessing is done.

**Definition 3.** *A* sampler *is a function $\sigma : [N] \to [M]^t$ for a set $[M]$ maps a sequence of coin-tosses $x$ drawn independently at random from $[N]$ to a sequence of samples $z_1, \ldots, z_t \in [M]$.*

*We say that a sampler $\sigma : [N] \to [M]^t$ is called an* averaging sampler *if for every function $f : [M] \to [0,1]$, the following holds.*

$$Prob_{z_1,\ldots,z_t \tilde{\sigma}(U_{[N]})} \left[ \frac{\sum_{i=1}^t f(z_i)}{t} \; > \; Ef + \epsilon \right] \; \leq \; \delta.$$

Note that we require only a one-sided error guarantee - the sample average may be lesser than $Ef - \epsilon$ with a significant probability.