Today’s Topic: Pseudo-Random Distributions.

Throughout, $X$ will be a probability distribution over $\Sigma^n$ and $U$ will be the uniform distribution over $\Sigma^n$.

Intuitively, how do we check that $X = U$? We do a statistical test on elements of $\Sigma^n$ produced by $X$, and see whether the behavior is uniform. For example, is the parity zero 50% of the time, and one the other 50% of the time? $X$ “passes” a statistical test if the outcome of the test is “close” to what the outcome would be if the test were performed on $U$.

Formally, a test will be a function $\text{Test} : \Sigma^n \rightarrow \{0, 1\}$. We then look at the size of the difference

$$\left| \Pr_{x\in U} [\text{Test}(x) = 1] - \Pr_{x\in X} [\text{Test}(x) = 1] \right|.$$

If that difference is small, we say $X$ is “random,” i.e., close to the uniform distribution. Again intuitively, if $X$ passes “all possible tests” it is close to random. But what does “all possible” mean?

**Definition 1.** A distribution is $(s, \epsilon)$-pseudorandom if for all circuits $C$ of size $\leq s$

$$\left| \Pr_{y\in X} [C(y) = 1] - \Pr_{y\in U} [C(y) = 1] \right| \leq \epsilon.$$

**Example:** The output distribution of pairwise independent generator we have considered in several lectures does beat certain tests. For example, it beats the test that looks at the first two bits $x_0x_1$ of a string, and outputs 1 if $x_0x_1 = 00$ and outputs 0 otherwise. If the test looks at three bits, though, the pairwise independent generator will fail.

A circuit can perform more complicated operations than the pairwise independent generator. Think of $s$ as “small.” Clearly, by definition, uniform distribution is $(s, \epsilon)$ pseudo-random. The interesting question is whether there exist distributions that are “far-away” from being uniform, and yet are $(s, \epsilon)$ pseudo-random. The answer is “Yes”. we can build such distributions by diagonalizing against circuits of size $s$. Pseudo-random distributions do exist, but the question now is: can they be effectively generated?
Definition 2. A family of distributions \( \{X_n\}_{n \in \mathbb{N}} \) is \((s, \epsilon)\)-pseudorandom if for every \( n \) and for all circuits \( C \) if size \( \leq s(n) \) it is the case that

\[
\left| \Pr_{y \in X_n} \left[ C(y) = 1 \right] - \Pr_{y \in U_n} \left[ C(y) = 1 \right] \right| \leq \epsilon.
\]

In addition to running tests on a distribution, we can assess the distribution’s randomness by trying to predict the bits it generates. This motivates the following definition.

Definition 3. A distribution \( X \) is \((s, \epsilon)\)-unpredictable if for all circuits \( C \) of size \( \leq s \) and for all \( i < n \) it is the case that

\[
\Pr_{y_1 \ldots y_i \in X} [c(y_1 \ldots y_k) = y_{i+1}] \leq \frac{1}{2} + \epsilon.
\]

It turns out that the definition of pseudorandom distribution and unpredictable distribution are equivalent.

Theorem 1. If a distribution \( X \) is \((s, \epsilon)\)-pseudorandom, then it is \((s', \epsilon)\)-unpredictable, where \( s' = s - \mathcal{O}(\log n) \).

Proof. We prove the contrapositive. Suppose \( X \) is \((s', \epsilon)\)-predictable. There exists an \( i \) and a \( C \) of size \( \leq s' \) such that when we randomly pick a string from \( X \)

\[
\Pr_{y_1 \ldots y_i \in X} [c(y_1 \ldots y_i) = y_{i+1}] \geq \frac{1}{2} + \epsilon.
\]

We will build a new circuit \( D \) to run a test on \( X \), as follows:

1. Input \( y_1 \ldots y_n \)
2. Run \( C \)
3. If \( C(y_1 \ldots y_i) = y_{i+1} \), output 1.
4. Else output 0.

If we start with \( U \), \( D \) will output 1 exactly half the time. Therefore

\[
\Pr_{y_1 \ldots y_n \in U_n} [D(y_1 \ldots y_n) = 1] = \frac{1}{2}.
\]

However, because of the predictability of \( X \)

\[
\Pr_{y_1 \ldots y_n \in X_n} [D(y_1 \ldots y_n) = 1] \geq \frac{1}{2} + \epsilon.
\]

Therefore \( X \) is not \((s, \epsilon)\)-pseudorandom. \( \square \)
We can also go the other way.

**Theorem 2.** If distribution $X$ is $(s, \epsilon)$-unpredictable, then $X$ is $(s', n\epsilon)$-pseudorandom, where $s' = s - \mathcal{O}(n)$.

**Proof.** Again, we prove the contrapositive. Suppose $X$ is not $(s', \epsilon)$-pseudorandom. Then there exists a circuit $C$ of size $\leq s'$ such that

$$\left| \Pr_{y \in X} [C(y) = 1] - \Pr_{y \in U} [C(y) = 1] \right| \geq \epsilon n.$$

We proceed using a method called the *Hybrid Technique* or *Hybrid Argument*.

Define a distribution $H_i$ as follows:

- Randomly pick $y_1 \ldots y_n \in X$.
- Uniformly at random pick an $(n - i)$-bit string $r$ from $\Sigma^{n-i}$.
- Output $y_1 \ldots y_i r$.

Note that $H_0 = U$, $H_n = X$ and $H_1$ is one bit from $X$ followed by $n-1$ bits from $U$. We can make the following analysis.

$$\Pr[C(H_n) = 1] - \Pr[C(H_0) = 1] \geq \epsilon n$$

$$= \Pr[C(H_n) = 1] - \Pr[C(H_{n-1}) = 1] + \Pr[C(H_{n-1}) = 1] - \Pr[C(H_{n-2}) = 1] + \Pr[C(H_{n-2}) = 1] - \cdots$$

So there is some $i$ such that $\Pr[C(H_{i+1}) = 1] - \Pr[C(H_i) = 1] \geq \epsilon$. Therefore, we can build a probabilistic circuit $D$ as follows:

- **Input** $y_1 \ldots y_k$.
- Randomly pick $b \in \{0, 1\}$.
- Uniformly at random pick an $n - (i + 1)$-bit string $r$.
- If $C(y_1 \ldots y_i b r) = 1$ output $b$ Else output $\overline{b}$ (i.e., $1 - b$).

Intuitive observation: Randomness does not really help when we come to circuits. We can convert $D$ to a deterministic circuit. We name probabilities $P_i$ by

$$\Pr[C(H_i) = 1] = P_i$$

$$\Pr[C(H_{i+1}) = 1] = P_{i+1}$$

and define $\overline{P}_{i+1}$ by picking a string according to $H_{i+1}$, flipping the $(i + 1)^{\text{st}}$ bit and outputting the resulting string.

$H_i$ can be generated as follows.
Toss a coin.
If Heads, act according to $H_{i+1}$.
If Tails, act according to $H_{i+1}$.
We define probabilities $P'_i$ by $\Pr[(\overline{H}_{i+1}) = 1] = P'_i$. Note then that
$$P_i = \frac{P_{i+1} + P'_i}{2}.$$  
We will now engage in a technical analysis of these probabilities in order to show that $D$ is a witness that $X$ is not unpredictable. Note first that we can decompose the probability that $D$ will output a particular answer by
$$\Pr[D(y_1 \ldots y_k) = y_{k+1}] = \Pr[C(y_1 \ldots y_k|b) = y_{k+1}]$$
$$= \Pr[C(y_1 \ldots y_k|b) = 1 \mid b = y_{k+1}] \cdot \Pr[b = y_{k+1}]$$
$$+ \Pr[C(y_1 \ldots y_k|b) = 0 \mid b \neq y_{k+1}]$$
$$\cdot \Pr[b \neq y_{k+1}].$$
Second, we observe that
$$\Pr[C(y_1 \ldots y_k|b) = 1 \mid b = y_{k+1}] = \Pr[C(y_1 \ldots y_{k+1}|b) = 1]$$
$$= \Pr[C(H_i) = 1]$$
$$= P_i$$
and further that
$$\Pr[C(y_1 \ldots y_k|b) = 0 \mid b \neq y_{k+1}] = \Pr[C(\overline{H}_{i+1}) = 0]$$
$$= 1 - \Pr[C(\overline{H}_{i+1}) = 1].$$
Combining this all together, we get
$$\Pr[D(y_1 \ldots y_k) = y_{k+1}] = \frac{P_i + 1 - P'_i}{2}$$
$$P'_i = 2P_i - P_{i+1}$$
$$= \frac{1 + P_{i+1} - P_i}{2}$$
$$\geq \frac{1}{2} + \frac{\epsilon}{2}.$$ 
This shows that the distribution $X$ is not $(s, \epsilon)$-unpredictable, thus proving the contrapositive of the theorem statement. We are done.\qed