CS 276 – Cryptography Sept 17, 2014

Lecture 4: Chebyshev's Inequality and Pseudorandom Generators

Instructor: Sanjam Garg *Scribe: Aaron Schild*

1 Pairwise independence and Chebyshev's Inequality

Last class, we needed to generate a large number of pairwise independent random samples from a small (logarithmic) number of truly independent random bits.

Definition 1 *A collection of random variables* $\{X_1, \ldots, X_n\}$ *is said to be* pairwise independent *if for all tuples of values* $(v_1, \ldots, v_n)$*,*

$$Pr[X_i = v_i, X_j = v_j] = P[X_i = v_i]P[X_j = v_j]$$

for all $i \neq j$*.*

Now we are ready for Chebyshev's Inequality:

Lemma 1 *Let* $X_1, \ldots, X_m$ *be pairwise independent and identically distributed binary random variables. In particular, suppose that* $Pr[X_i = 1] = p$ *for some* $p \in [0, 1]$ *and all* $i \in [m] = \{1, 2, \ldots, m\}$*. Then*

$$Pr[|\sum_i X_i - pm| > \delta m] < \frac{1}{4\delta^2 m}$$

Proof. Let $Y = \sum_i X_i$. Then

$$\begin{aligned} Pr[|\sum_i X_i - pm| > \delta m] &= Pr[(\sum_i X_i - pm)^2 > \delta^2 m^2] \\ &< \frac{E[|Y - pm|^2]}{\delta^2 m^2} \\ &= \frac{E[Y^2] - p^2 m^2}{\delta^2 m^2} \\ &= \frac{Var(Y)}{\delta^2 m^2} \end{aligned}$$

Note that

$$\begin{aligned} Var(Y) &= E[Y^2] - (E[Y])^2 \\ &= \sum_{i,j} (E[X_i Y_i] - E[X_i]E[Y_i]) \\ &= \sum_i E[X_i^2] - (E[X_i])^2 \\ &= mp(1-p) \end{aligned}$$

where the transition from the second to the third line holds by pairwise independence. Therefore,

$$Pr[|\sum_i X_i - pm| > \delta m] \leq \frac{p(1-p)}{\delta^2 m} \leq \frac{1}{4\delta^2 m}$$

as desired.

2 Computational Indistinguishability

What should it mean for a computationally bounded adversary to be able to distinguish two distributions from each other? It is tricky to define for a single pair of distributions because the length of the output of a random variable is a constant. Therefore, in order for "computationally bounded" adversaries to make sense, we have to work with infinite families of probability distributions.

Definition 2 *An* ensemble *of probability distributions is a sequence of random variables* $\{X_n\}_{n\in\mathbb{N}}$. *Two ensembles of probability distributions* $\{X_n\}_n$ *and* $\{Y_n\}_n$ *are said to be* computationally indistinguishable *if for all PPT machines A (that are PPT in* $\ell(n)$*, the length of any value of the variable* X_n*), the quantities*

$$p(n) := Pr[A(X_n) = 1] = \sum_x Pr[X_n = x]Pr[A(x) = 1]$$

and

$$q(n) := Pr[A(Y_n) = 1] = \sum_y Pr[Y_n = y]Pr[A(y) = 1]$$

differ by a negligible amount; i.e. $|p(n) - q(n)|$ *is negligible in* n.
This equivalence is denoted by

$$\{X_n\}_n \approx \{Y_n\}_n$$

Now, we can define pseudorandom generators, which intuitively generator a polynomial number of bits that are indistinguishable from being uniformly random:

Definition 3 *A function* $G : \{0,1\}^n \to \{0,1\}^{n+m}$ *with* $m = poly(n)$ *is called a* pseudorandom generator *if*

- *for all $x \in \{0,1\}^n$, $G(x)$ is PPT-computable.*
- *$U_{n+m} \approx G(U_n)$, where U_k denotes the uniform distribution on $\{0,1\}^k$.*

We now prove some properties of computationally indistinguishable ensembles that will be useful later on.

Lemma 2 (Sunglass Lemma) *If $\{X_n\}_n \approx \{Y_n\}_n$ and P is a PPT-machine, then*

$$\{P(X_n)\}_n \approx \{P(Y_n)\}_n$$

Proof. Consider an adversary A that can distinguish $\{P(X_n)\}_n$ from $\{P(Y_n)\}_n$ with nonnegligible probability. Then the adversary $A \circ P$ can distinguish $\{X_n\}_n$ from $\{Y_n\}_n$ with the same nonnegligible probability. Since P and A are both PPT-machines, the composition is also a PPT-machine. This proves the contrapositive of the lemma.

Lemma 3 (Hybrid Argument) *For a polynomial $t : \mathbb{Z}^+ \to \mathbb{Z}^+$ let the t-product of $\{Z_n\}_n$ be*

$$\{Z_n^{(1)}, Z_n^{(2)}, \ldots, Z_n^{(t(n))}\}_n$$

where the $Z_n^{(i)}$s are independent copies of Z_n. If

$$\{X_n\}_n \approx \{Y_n\}_n$$

then

$$\{X_n^{(1)}, \ldots, X_n^{(t)}\}_n \approx \{Y_n^{(1)}, \ldots, Y_n^{(t)}\}_n$$

as well.

Proof. Consider the set of tuple random variables

$$H_n^{(i,t)} = (Y_n^{(1)}, \ldots, Y_n^{(i)}, X_n^{(i+1)}, X_n^{(i+2)}, \ldots, X_n^{(t)}$$

for integers $0 \leq i \leq t$. Assume, for the sake of contradiction, that there is a PPT adversary A that can distinguish between $\{H_n^{(0,t)}\}_n$ and $\{H_n^{(t,t)}\}_n$ with nonnegligible probability difference $r(n)$. Suppose that A returns 1 with probability ϵ_i when it runs on samples from $H_n^{(i,t)}$. By definition, $|\epsilon_t - \epsilon_0| \geq r(n)$. By the Triangle Inequality and the Pidgeonhole Principle, there is some index k for which
$|\epsilon_{k+1} - \epsilon_k| \geq r(n)/t$.
This is equivalent to trying to distinguish the ensembles $\{(X_n, T_n)\}_n$ from $\{(Y_n, T_n)\}_n$, where T_n is independent of X_n and Y_n (T_n is the random variable representing all coordinates but the k-th coordinate). Note that

$$
\begin{aligned}
r(n)/t &\leq |Pr[A(Y_n, T_n) = 1] - Pr[A(X_n, T_n) = 1]| \\
&= |\sum_{x,t} (Pr[Y_n = x, T_n = t] - Pr[X_n = x, T_n = t]) Pr[A(x,t) = 1]| \\
&= |\sum_{t} Pr[T_n = t] \sum_{x} (Pr[Y_n = x] - Pr[X_n = x]) Pr[A(x,t) = 1]| \\
&\leq \sum_{t} Pr[T_n = t] \sum_{x} |Pr[Y_n = x] - Pr[X_n = x]| Pr[A(x,t) = 1]
\end{aligned}
$$

so by the probabilistic method there is a t_0 for which $r(n)/t \leq \sum_x |Pr[Y_n = x] - Pr[X_n = x]| Pr[A(x, t_0) = 1]$. This means that X_n can be distinguished from Y_n with probability difference $r(n)/t$, which is nonnegligible (a contradiction).
Finally, we show that any pseudorandom generator that produces one bit of randomness can be extended to create a polynomial number of bits of randomness.

Theorem 4 *Consider $F : \{0,1\}^n \to \{0,1\}^{n+1}$. Construct a function $G : \{0,1\}^n \to \{0,1\}^{n+m}$ as follows. Let $T : \{0,1\}^{n+1} \to \{0,1\}^n$ be the function that truncates a string of length $n+1$ down to its first n digits. Let $S : \{0,1\}^{n+1} \to \{0,1\}$ be the function that outputs the last bit in a string of length $n+1$.*
Now, let G be the function for which $G(x)_i = (S \circ (F \circ T)^{i-1} \circ F)(x)$ for $i \in m$ and $G(x)_{m+j} = (T \circ (F \circ T)^{m-1} \circ F)(x)_j$ for all $j \in n$.
G is a pseuodorandom generator if F is also a pseudorandom generator.

Proof. Let $H_i : \{0,1\}^n \to \{0,1\}^{n+m}$, $0 \leq i \leq m$ be the function that replaces the input to the $i+1$th function call with U_n and populates the first i bits of the output with U_i. Note that $H_0(U_n) = G(U_n)$ and that $H_m(U_n) = U_n$.
Suppose, for the sake of contradiction, that there is a PPT machine A that distinguishes $H_0(U_n)$ from $H_m(U_n)$ with nonnegligible probability difference $r(n)$. Then, there is some index k for which $|Pr[A(H_k(U_n)) = 1] - Pr[A(H_{k-1}(U_n)) = 1]| \geq r(n)/m$ by the Pidgeonhole Principle and the Triangle Inequality. These functions can only differ on one bit, though. In particular, the must differ in distribution only on the last bit of the output of the kth call to F. In particular, A must distinguish $F(U_n)$ from U_n with probability difference at least $r(n)/m$ (since these are the two different inputs to the $k+1$th call to f. $r(n)/m$ is nonnegligible, which is a contradiction.

Figure 1: The construction of G. The output of G is the $m+n$ bit vector depicted as a rectangular column with a red border.