CS276 Scribe Notes

Adam Chlipala <adamc@cs.berkeley.edu>

January 28, 2004 (part 1)

1 Augmenting a PRG's output with random bits

Let $G : X \to Y$ be a (t, ϵ)-PRG that runs in time t'.

Define a new PRG H that is represented by the following diagram:

$$\begin{array}{ccccc} u_X & \longrightarrow & \boxed{G} & \longrightarrow & G(u_X) \\ u_Y & & \longrightarrow & & u_Y \end{array}$$

Formally, $H : X \times Y \to Y \times Y$ is defined by:

$$H(x, y) = (G(x), y)$$

Claim. H is a $(t - O(1), \epsilon)$-PRG.

1.1 First proof

Proof. Suppose not. Assume we can break H and show that this means we can break G.

Suppose B $(t - O(1), \epsilon)$-breaks H. Formally, there exists algorithm $B : Y \times Y \to \{0, 1\}$ such that:

$$\text{Adv}B = |\Pr_{x \leftarrow X, y \leftarrow Y}[B(G(x), y) = 1] - \Pr_{y, y' \leftarrow Y}[B(y, y') = 1]| > \epsilon$$

We would like to construct algorithm A that (t, ϵ)-breaks G. We can define such an $A : Y \to \{0, 1\}$ like this:

$A(y)$:

1. Pick $y' \leftarrow u_Y$.
2. Output $B(y, y')$.

We haven't given the details of our computational model, but we'll assume that choosing an element uniformly from Y can be accomplished in constant time. Thus, since the first step of A runs in $O(1)$ time and the second runs in $t - O(1)$ time, we have that A's worst case running time is t, where we consider the $O(1)$ factors to all be the fixed time to choose a value uniformly.

We can complete the proof by proving that A's advantage is greater than ϵ:

$$\begin{aligned}
\mathrm{Adv}A &= |\Pr_{x \to X}[A(G(x)) = 1] - \Pr_{y \to Y}[A(y) = 1]| \\
&= |\Pr_{x \to X, y \to Y}[B(G(x), y) = 1] - \Pr_{y, y' \to Y}[B(y, y') = 1]| \\
&= \mathrm{Adv}B \\
&> \epsilon
\end{aligned}$$

Thus, A (t, ϵ)-breaks G. This contradicts the initial assumption, so H is a $(t - O(1), \epsilon)$-PRG. □

1.2 A shorter proof

Fact 1. Let $\sim$ be an informal notion of similarity between distributions according to efficient statistical tests. If $D \sim D'$ and f is an efficiently computable (possibly randomized) function, then $f(D) \sim f(D')$. A more rigorous version of this fact is given in Homework 1 in terms of a distance measure on distributions: Where L is any (possibly randomized) algorithm running with resources R', $d_R(L(D), L(D')) \le d_{R+R'}(D, D')$.

Proof. Exercise 1(a) on Homework 1. □

Now we can construct a much shorter proof that H is a $(t - O(1), \epsilon)$-PRG.

Alternate (informal) proof. We are given $G(u_X) \sim u_Y$.

Define $f : Y \to Y \times Y$ with $f(y) = (y, y')$ where $y' \leftarrow Y$. So:

$$\begin{aligned}
G(u_X) &\sim u_Y \\
f(G(u_X)) &\sim f(u_Y) \text{ (by Fact 1)} \\
(G(u_X), u'_Y) &\sim (u_Y, u'_Y) \\
H(u_X, u'_Y) &\sim (u_Y, u'_Y)
\end{aligned}$$

So H is a PRG. □

Conclusion a. $(G(u_X), u'_Y) \sim (u_Y, u'_Y)$

Alternate (formal) proof. By definition, since G is a (t, ϵ)-PRG, $d_t(G(u_X), u_Y) \le \epsilon$. Using the definition of f from above along with the formal version of Fact 1, we have:

$$\begin{aligned}
d_{t-O(1)}(H(u_X, u'_Y), (u_Y, u'_Y)) &= d_{t-O(1)}((G(u_X), u'_Y), (u_Y, u'_Y)) \\
&= d_{t-O(1)}(f(G(u_X)), f(u_Y)) \\
&\le d_t(G(u_X), u_Y) \\
&\le \epsilon
\end{aligned}$$

Thus, $d_{t-O(1)}((G(u_X), u'_Y), (u_Y, u'_Y)) \le \epsilon$, so H is a $(t - O(1), \epsilon)$-PRG. For convenience in a later example, we also note that since $t - t' \le t - O(1)$, $d_{t-t'}((G(u_X), u'_Y), (u_Y, u'_Y)) \le d_{t-O(1)}((G(u_X), u'_Y), (u_Y, u'_Y)) \le \epsilon$. □

2 Example

Let G be the same PRG from the last section.

Claim. $(G(u_X), G(u'_X)) \sim (G(u_X), u_Y)$.

Informal proof. $G(u'_X) \sim u_Y$ follows from the fact that G is a PRG.

Define $f : Y \to Y \times Y$ by $f(y) = (G(x), y)$ where $x \leftarrow X$.

$$\begin{array}{rcl} G(u'_X) & \sim & u_Y \\ f(G(u'_X)) & \sim & f(u_Y) \text{ (by Fact 1)} \\ (G(u_X), G(u'_X)) & \sim & (G(u_X), u_Y) \end{array}$$

□

Conclusion b. $(G(u_X), G(u'_X)) \sim (G(u_X), u_Y)$

Formal proof. This formal proof is very similar to the last one. Let H refer to the function defined by $H(y, y') = (G(x), y')$ where $x \leftarrow X$. The running time of f is $O(1)$ to choose a random x plus t' to run G, which we will consider asymptotically so that the total time is t'.

$$\begin{array}{rcl} d_{t-t'}(H(u_Y, u'_Y), (u_Y, u'_Y)) & = & d_{t-t'}((G(u_X), u'_Y), (u_Y, u'_Y)) \\ & = & d_{t-t'}(f(G(u'_X)), f(u_Y)) \\ & \leq & d_t(G(u_X), u_Y) \\ & \leq & \epsilon \end{array}$$

So $d_{t-t'}((G(u_X), u'_Y), (u_Y, u'_Y)) \leq \epsilon$. □

3 Example

Fact 2. If $D \sim D'$ and $D' \sim D''$, then $D \sim D''$. Formally, $d_R(D, D'') \leq d_R(D, D') + d_R(D', D'')$.

Proof. Exercise 1(b) on Homework 1. □

Consider the same G from the last two examples. Define H in terms of G as indicated by this diagram:

$$\begin{array}{l} u_X \longrightarrow \boxed{G} \longrightarrow G(u_X) \\ u'_X \longrightarrow \boxed{G} \longrightarrow G(u'_X) \end{array}$$

Formally, $H : X \times X \to Y \times Y$ is defined by:

$$H(x, x') = (G(x), G(x'))$$

Claim. H is a *PRG*, i.e., $(G(u_X), G(u'_X)) \sim (u_Y, u'_Y)$.

Informal proof. This follows directly by Fact 2 from Conclusions a and b. In "diagram form":

$$(G(u_X), G(u'_X)) \sim^b (G(u_X), u'_Y) \sim^a (u_Y, u'_Y)$$

□

Formal proof. Using the formal versions of Fact 2 and Conclusions a and b, we have:

$$\begin{aligned} d_{t-t'}((G(u_X), G(u'_X)), (u_Y, u'_Y)) &\le d_{t-t'}((G(u_X), G(u'_X)), (G(u_X), u'_Y)) + d_{t-t'}((G(u_X), u'_Y), (u_Y, u'_Y)) \\ &\le \epsilon + \epsilon \\ &= 2\epsilon \end{aligned}$$

Thus, $d_{t-t'}((G(u_X), G(u'_X)), (u_Y, u'_Y)) \le 2\epsilon$, so H is a $(t-t', 2\epsilon)$-PRG. $\square$