Lecture 07: More on Probability and Hybrid Arguments

More Examples

Claim

Let $\alpha \in [0,1]$ *and* $\overline{\alpha} = 1 - \alpha$. *For distributions* A, B *and* C *over the sample space* Ω, *the following holds:*

$$\mathrm{SD}\left(\alpha A + \overline{\alpha} B, \alpha C + \overline{\alpha} B\right) = \alpha \mathrm{SD}\left(A, C\right)$$

$$\begin{aligned}
\mathrm{SD}\left(\alpha A + \overline{\alpha} B, \alpha C + \overline{\alpha} B\right) &= \frac{1}{2} \sum_{x \in \Omega} \left|(\alpha A + \overline{\alpha} B)(x) - (\alpha C + \overline{\alpha} B)(x)\right| \\
&= \frac{1}{2} \sum_{x \in \Omega} \left|(\alpha A(x) + \overline{\alpha} B(x)) - (\alpha C(x) + \overline{\alpha} B(x))\right| \\
&= \alpha \cdot \frac{1}{2} \sum_{x \in \Omega} \left|A(x) - C(x)\right| \\
&= \alpha \cdot \mathrm{SD}\left(A, C\right)
\end{aligned}$$

More Examples

Claim

Let A and B be distribution over Ω and C over Ω' be independent distributions. Then the following holds:

$$\mathrm{SD}\left((A, C), (B, C)\right) = \mathrm{SD}\left(A, B\right)$$

$$\begin{aligned}
\mathrm{SD}\left((A, C), (B, C)\right) &= \frac{1}{2} \sum_{\substack{x \in \Omega \\ y \in \Omega'}} \left|(A, C)(x, y) - (B, C)(x, y)\right| \\
&= \frac{1}{2} \sum_{x \in \Omega} \sum_{y \in \Omega'} \left|A(x)C(y) - B(x)C(y)\right| \\
&= \frac{1}{2} \sum_{x \in \Omega} \left|A(x) - B(x)\right| \sum_{y \in \Omega'} C(y) \\
&= \frac{1}{2} \sum_{x \in \Omega} \left|A(x) - B(x)\right| = \mathrm{SD}\left(A, B\right)
\end{aligned}$$

Pseudorandom Generators

Definition (Pseudorandom Generators (First Attempt))

A pseudorandom generator is a function $G\colon \{0,1\}^n \rightarrow \{0,1\}^{n+\ell}$, for $\ell \geqslant 1$, such that:

$$\mathrm{SD}\left(G(U_{\{0,1\}^n}), U_{\{0,1\}^{n+\ell}}\right) \leqslant \text{"small"}$$

Its input is called seed, and ℓ is called the stretch of the PRG.

Intuition: Given a small-length uniformly random seed, the PRG extends it to a longer "random-looking" string.

Impossibility against Unbounded Adversaries

Lemma

$$\mathrm{SD}\left(G(U_{\{0,1\}^n}), U_{\{0,1\}^{n+\ell}}\right) \geqslant 1 - \frac{1}{2^\ell}$$

- Let $Z = \{y \colon y \in \{0,1\}^{n+\ell}, \exists x \in \{0,1\}^n \text{ s.t. } G(x) = y\}$. Note that $|Z| \leqslant 2^n$.
- Then consider the following manipulation:

$$\begin{aligned}\mathrm{SD}\left(G(U_{\{0,1\}^n}), U_{\{0,1\}^{n+\ell}}\right) &= \sum_{y \in Z} \frac{\left|f^{-1}(y)\right|}{2^n} - \frac{1}{2^{n+\ell}} \\ &= \frac{\sum_{y\in Z}\left|f^{-1}(y)\right|}{2^n} - \frac{|Z|}{2^{n+\ell}} \\ &\geqslant 1 - \frac{1}{2^\ell}\end{aligned}$$

Change in Definition

Instead of any adversary (which includes adversaries with unbounded computational power) we restrict to adversaries that have bounded computational power. Then PRGs are believed to exist.

Example of Hybrid Argument

- Consider the experiment where and adversary $\mathcal{A}$ has to predict whether the sample was generated using the distribution $A^{(0)}$ or $A^{(1)}$.
- Note that we are interested in finding the distribution:

  $$\widetilde{B} = \mathcal{A}(\frac{1}{2} \cdot A^{(0)} + \frac{1}{2} \cdot A^{(1)})$$

  We do not understand this behavior.
- But consider a related distribution:

  $$\widetilde{B'} = \mathcal{A}(\frac{1}{2} \cdot A^{(0)} + \frac{1}{2} \cdot A^{(0)})$$

  That is, independent of the random bit b, we sample according to the distribution $A^{(0)}$.
- Suppose $\mathrm{SD}\left(A^{(0)}, A^{(1)}\right) = \varepsilon$, then $\mathrm{SD}\left(\widetilde{B}, \widetilde{B'}\right) \leqslant \varepsilon/2$ (using the examples we proved today and data-processing inequality)

Example Continued

- Consider the function $f(x) = (b == x)$, i.e. the function that tests the equality of x and the secret bit b chosen by the honest challenger
- We know that $\mathrm{SD}\left(f(\widetilde{B}), f(\widetilde{B'})\right) \leqslant \mathrm{SD}\left(\widetilde{B}, \widetilde{B'}\right) \leqslant \varepsilon/2$ (by data-processing inequality)
- Note that $f(\widetilde{B}) = U_{\{0,1\}}$, i.e., the uniform distribution over one bit
- So, $f(\widetilde{B'})$ is at most $\varepsilon/2$ close to the uniform distribution over one-bit. Thus, the advantage of the adversary is at most $\varepsilon/2$.

Another Example

- Suppose there exists two messages $m^{(0)}$ and $m^{(1)}$ such that the distribution of their respective ciphertexts $C^{(0)}$ and $C^{(1)}$ have statistical distance ε
- Prove using the above strategy that the advantage of an adversary to correctly predict the bit b in the security game is at at most $\varepsilon/2$