TD 1: Play with definitions

Exercise 1. *Statistical distance*

Definition 1 (Statistical distance). *Let X and Y be two discrete random variables over a countable set A. The statistical distance between X and Y is the quantity*

$$\Delta(X, Y) = \frac{1}{2} \sum_{a \in A} |\Pr[X = a] - \Pr[Y = a]|.$$

The statistical distance verifies usual properties of distance function, i.e., it is a positive definite binary symmetric function that satisfies the triangle inequality:

- $\Delta(X, Y) \geq 0$, with equality if and only if X and Y are identically distributed,
- $\Delta(X, Y) = \Delta(Y, X)$,
- $\Delta(X, Z) \leq \Delta(X, Y) + \Delta(Y, Z)$.

1. Show that if $\Delta(X, Y) = 0$, then for any deterministic adversary $\mathcal{A}$, we have $\mathrm{Adv}_{\mathcal{A}}(X, Y) = 0$.

In the next question, we will prove the *data processing inequality* for the statistical distance.

2. Let X, Y be two random variables over a common set A.

   (a) Let $f : A \to S$ be a deterministic function with domain S. Show that
   $$\Delta(f(X), f(Y)) \leq \Delta(X, Y).$$

   (b) Let Z be another random variable with domain $\mathcal{Z}$, statistically independent from X and Y. Show that
   $$\Delta((X, Z), (Y, Z)) = \Delta(X, Y).$$

   (c) Let f be a (possibly probabilistic) function with domain S. Define f' a deterministic function and R a random variable independent from X and Y such that for any input x, we have $f'(x, R) = f(x)$. The random variable R is the internal randomness of f. Using f' and R, show that $\Delta(f(X), f(Y)) = \Delta(f'(X, R), f'(Y, R)) \leq \Delta(X, Y)$.

3. Show that for any (possibly probabilistic) adversary $\mathcal{A}$, we have $\mathrm{Adv}_{\mathcal{A}}(X, Y) \leq \Delta(X, Y)$.

4. Assuming the existence of a secure PRG $G : \{0,1\}^s \to \{0,1\}^n$, show that $\Delta(G(U(\{0,1\}^s)), U(\{0,1\}^n))$ can be much larger than $\max_{\mathcal{A}} \mathrm{Adv}_{\mathcal{A}}(G(U(\{0,1\}^s)), U(\{0,1\}^n))$.

Exercise 2. *A weird distinguisher...*

We consider two distributions D_0 and D_1 over $\{0,1\}^n$.

1. Recall the definitions that were given in class for the notions of *distinguisher*, *advantage* and *indistinguishability* of D_0 and D_1.

You found a distinguisher $\mathcal{A}$ on internet. However, you cannot find anywhere in the documentation its performances!

2. Assuming that you have access to as many samples as you like from D_0 and D_1 (you can for instance assume that you can sample yourself from these distributions), how would you estimate the advantage of $\mathcal{A}$? *Hint: use the Chernoff Bound:* $\Pr(|X - np| \geq nt) \leq 2\exp(-2nt^2)$, *where* X *follows a binomial distribution with parameters* (n, p).

By convention, you want to design a distinguisher such that it outputs 1 when it thinks the sample comes from D_1 and 0 otherwise. However, because of the definition of advantage, it is also possible to design distinguishers that do the reverse, and still have the same advantage. For instance, the above distinguisher $\mathcal{A}$ may often be "wrong". This could be troublesome if your aim is to use its output to do further computations. Luckily, there exists a way to transform $\mathcal{A}$ into a distinguisher that is more often right than wrong, whatever it previously did.

3. The definition of advantage from question 1 may be called Absolute Advantage, for the purpose of this exercise. In this question, we define the Positive Advantage of $\mathcal{A}$ as

$$\mathsf{Adv}_P(\mathcal{A}) := \Pr(\mathcal{A} \xrightarrow{Exp_1} 1) - \Pr(\mathcal{A} \xrightarrow{Exp_0} 1).$$

Given a distinguisher $\mathcal{A}$ with Absolute Advantage ε, we build a distinguisher $\mathcal{A}'$ that does the following:

1. Upon receiving a sample $y \hookleftarrow D_b$, it runs $b' \leftarrow \mathcal{A}(y)$.
2. It samples $x_0 \hookleftarrow D_0$ and $x_1 \hookleftarrow D_1$ and runs $b_0 \leftarrow \mathcal{A}(x_0)$ and $b_1 \leftarrow \mathcal{A}(x_1)$.
3. It returns b' if $b_0 = 0$ and $b_1 = 1$. It returns $1 - b'$ if $b_0 = 1$ and $b_1 = 0$.
4. In any other cases, it returns a uniform bit.

Prove that the Positive Advantage of $\mathcal{A}'$ is ε^2.

Exercise 3. *Bit-flip of a PRG*

Let G a pseudo-random generator (PRG) of input range $\{0,1\}^s$ and output range $\{0,1\}^n$. We define $\bar{G}$ as follows:

$$\forall x \in \{0,1\}^s, \bar{G}(x) := 1^n \oplus G(x),$$

where $\oplus$ denotes the XOR operation. This corresponds to flipping every bit of the output of G.

1. Prove that $\bar{G}$ is secure if and only if G is secure.

Exercise 4. *(Bonus) Variable-length OTP is not secure*

A *variable length one-time pad* is a cipher (E, D), where the keys are bit strings of some fixed length L, while messages and ciphertexts are variable length bit strings, of length at most L. Thus, the cipher (E, D) is defined over $(\mathcal{K}, \mathcal{M}, \mathcal{C})$, where

$$\mathcal{K} := \{0,1\}^L \text{ and } \mathcal{M} := \mathcal{C} = \{0,1\}^{\leq L}$$

for some parameter L. Here, $\{0,1\}^{\leq L}$ denotes the set of all bit strings of length at most L (including the empty string). For a key $k \in \{0,1\}^L$ and a message $m \in \{0,1\}^{\leq L}$ of length ℓ, the encryption function is defined as follows:

$$E(k, m) := k[0 \mathinner{.\,.} \ell - 1] \oplus m$$

1. Provide a counter-example showing that the variable length OTP is not secure for perfect secrecy.