TD 1: Play with definitions

Notation. For $n > 0$, we write $\mathbb{Z}_n$ the ring $\mathbb{Z}/n\mathbb{Z}$ of integers modulo n.

Exercise 1. *Distributions and (in)dinstinguishability*

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* and *indistinguishability* of D_0 and D_1.

Now, we consider the following experiment.

$\mathcal{C}$	$\mathcal{A}$
sample $b \hookleftarrow U(0,1)$	
sample $x \hookleftarrow D_b$	
send x to $\mathcal{A}$	
	compute a bit b'
	send b' to $\mathcal{C}$
If $b = b'$, say "Win", else say "Lose".	

We say that a PPT algorithm $\mathcal{A}$ is a *distinguisher* if there exists a non-negligible ε such that, in this experiment, $\Pr[\texttt{Win}] \geq 1/2 + \varepsilon$. The distributions D_0 and D_1 are said to be *indistinguishable* if there is no such distinguisher.

2. Show that this definition of indistinguishability is equivalent to the one recalled in the previous question.

3. A rebellious student decides to define a distinguisher as a PPT algorithm $\mathcal{A}$ with $\Pr[\texttt{Win}] \leq 1/2 - \varepsilon$ in the above experiment (rather than $\geq 1/2 + \varepsilon$). Is this a revolutionary idea?

Exercise 2. *Statistical distance*

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

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

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 adversary $\mathcal{A}$ we have $\mathrm{Adv}_{\mathcal{A}}(X,Y) = 0$.

We also recall the following property: if X and Y are two random variables over a common set A, then for any (possibly randomized) function f with domain S we have

$$\Delta(f(X), f(Y)) \leq \Delta(X,Y);$$

besides, if f is injective then the equality holds.

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

3. 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 $\max_{\mathcal{A}\text{ PPT}} \mathrm{Adv}_{\mathcal{A}}(G(U(\{0,1\}^s)), U(\{0,1\}^n))$.

Exercise 3. *Introduction to Computational Hardness Assumptions*

Definition 2 (Decisional Diffie-Hellman distribution)**.** *Let $\mathbb{G}$ be a cyclic group of prime order q, and let g be a publicly known generator of $\mathbb{G}$. The* decisional Diffie-Hellman *distribution (DDH) is,* $D_{\mathrm{DDH}} = (g^a, g^b, g^{ab}) \in \mathbb{G}^3$ *with a, b sampled independently and uniformly at random in $\mathbb{Z}_q$.*

Definition 3 (Decisional Diffie-Hellman assumption)**.** *The decisional Diffie-Hellman assumption states that there exists no probabilistic polynomial-time distinguisher between D_{DDH} and (g^a, g^b, g^c) with a, b, c sampled independently and uniformly at random in $\mathbb{Z}_q$.*

1. Does the DDH assumption hold in $\mathbb{G} = (\mathbb{Z}_p, +)$ for $p = \mathcal{O}(2^\lambda)$ prime?
2. Same question for $\mathbb{G} = (\mathbb{Z}_p^\star, \times)$ of order $p - 1$.
3. Now we take $\mathbb{Z}_p$ such that $p = 2q + 1$ with q prime (also called a *safe-prime*). Let us work in a subgroup $\mathbb{G}$ of order q in $(\mathbb{Z}_p^\star, \times)$.
   - **(a)** Given a generator g of $\mathbb{G}$, propose a construction for a function $\hat{G} : \mathbb{Z}_q \to \mathbb{G} \times \mathbb{G}$ (which may depend on public parameters) such that $\hat{G}(U(\mathbb{Z}_q))$ is computationally indistinguishable from $U(\mathbb{G} \times \mathbb{G})$ based on the DDH assumption on $\mathbb{G}$ (where, in $\hat{G}(U(\mathbb{Z}_q))$, the probability is also taken over the public parameters of $\hat{G}$).
   - **(b)** What is the size of the output of $\hat{G}$ given the size of its input?
   - **(c)** Why is it not a pseudo-random generator from $\{0,1\}^\ell$ to $\{0,1\}^{2\ell}$ for $\ell = \lceil \lg q \rceil$?

Exercise 4. *Let us go post-quantum!*

Definition 4 (Learning with Errors)**.** *Let $\ell < k \in \mathbb{N}$, $n < m \in \mathbb{N}$, $q = 2^k$, $B = 2^\ell$, $\mathbf{A} \hookleftarrow U(\mathbb{Z}_q^{m \times n})$. The Learning with Errors (*LWE*) distribution is defined as follows: $D_{\mathrm{LWE},\mathbf{A}} = (\mathbf{A}, \mathbf{A} \cdot \mathbf{s} + \mathbf{e} \bmod q)$ for $\mathbf{s} \hookleftarrow U(\mathbb{Z}_q^n)$ and $\mathbf{e} \hookleftarrow U\left(\left[-\frac{B}{2}, \frac{B}{2} - 1\right]^m \cap \mathbb{Z}^m\right)$.*

Note. In this setting, the vector **s** is called the secret, and **e** the noise.
The *LWE assumption* states that, given suitable parameters k, ℓ, m, n, it is computationally hard to distinguish $D_{\mathrm{LWE},\mathbf{A}}$ from the distribution $(\mathbf{A}, U(\mathbb{Z}_q^m))$.
Let us propose the following generator: $G_{\mathbf{A}}(\mathbf{s}, \mathbf{e}) = \mathbf{A} \cdot \mathbf{s} + \mathbf{e} \bmod q$.

1. Given the binary representation of **s**, **e**, compute the bitsize of the input and the output of the function G with respect to k, ℓ, m, n.
2. Evaluate the cost of a bruteforce attack to retrieve the input **s**, **e** in terms of arithmetic operations in $\mathbb{Z}_q$.
3. What happens if $B = 0$? ☞ *This bound can prove useful:* $\prod_{i=1}^{n}(1 - 2^{-i}) > 0.288$.
4. Given the previous question, refine the bruteforce attack of question 2. What does it mean for the security of the generator G?
5. What happens if $\ell = k$?
6. Given suitable ℓ, k, n, m such that the LWE problem holds in this setting, show that $G_{\mathbf{A}}$ is a pseudo-random generator.