TD1 - Définir la sécurité

(Indistinguishability) **Exercise 1.** *Two definitions of indistinguishability*

On considère deux distributions D_0 et D_1 sur $\{0,1\}^n$.

1. Rappelez la définition donnée en cours pour la notion de *distingueur* et *indistinguabilité* de D_0 et D_1.

On considère maintenant l'expérience suivante.

$\mathcal{C}$	$\mathcal{A}$
tire $b \hookleftarrow U(0,1)$	
tire $x \hookleftarrow D_b$	
envoie x à $\mathcal{A}$	
	calcule un bit b'
	envoie b' à $\mathcal{C}$
Si $b = b'$, retourner "Win", sinon, retourner "Lose".	

On dit qu'un algorithme PPT $\mathcal{A}$ est un *distingueur* s'il existe un ε non négligeable tel que, dans cette expérience, $\mathbb{P}\mathrm{r}[\texttt{Win}] \geqslant 1/2 + \varepsilon$. Les distributions D_0 et D_1 sont dites *indistinguables* s'il n'existe aucun distingueur.

2. Montrez que cette definition d'indistinguabilité est équivalente à celle rappelée à la question précédente.

3. Un étudiant rebelle decide de définir un distingueur comme un algorithm PPT $\mathcal{A}$ avec $\mathbb{P}\mathrm{r}[\texttt{Win}] \leq 1/2 - \varepsilon$ dans l'expérience précédente (plutôt que $\geq 1/2 + \varepsilon$). Est-ce une idée révolutionnaire ?

4.

(SD) **Exercise 2.** *Statistical Distance*

Definition 1 (Statistical distance)**.** *Soient X et Y deux variables aléatoires discrètes sur un ensemble dénombrable A. La distance statistique entre X et Y est définie comme:*

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

La distance statistique vérifie les propriétés usuelles d'une distance, à savoir : c'est une fonction symétrique définie positive qui satisfait l'inégalité triangulaire :

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

1. Soient X et Y deux variables aléatoires discrètes sur un ensemble dénombrable A, et soit Z une troisième variable aléatoire sur un ensemble B. Montrez que si Z est statistiquement indépendant de X et Y alors
$$\Delta((X,Z),(Y,Z)) = \Delta(X,Y).$$

2. Soient $X_1, \ldots, X_k$ et $Y_1, \ldots, Y_k$ deux listes de variables totalement indépendantes. Montrez que

$$\Delta((X_1, \ldots, X_k), (Y_1, \ldots, Y_k)) \leqslant \sum_{i=1}^{k} \Delta(X_i, Y_i).$$

3. Soient X et Y deux viriables aléatoires sur un ensemble commun A. Montrez que pour toute (possiblement random) fonction f de domaine A,

$$\Delta(f(X), f(Y)) \leqslant \Delta(X, Y).$$

Donnez une condition sur f pour que l'égalité soit vérifiée.

Indice: Considérez d'abord le cas des fonctions déterministes.

4. Montrez que pour tout adversaire $\mathcal{A}$, $\mathrm{Adv}_{\mathcal{A}}(X, Y) \leqslant \Delta(X, Y)$.

(SD2) Exercise 3. *Statistical Distance (2)*

Definition 2 (Statistical distance)**.** *Soit X et Y deux variables aléatoires discrètes sur un ensemble dénombrable S. La distance statistique entre X et Y est la quantité*

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

La distance statistique vérifie les propriétés usuelles d'une distance, à savoir, c'est une fonction symétrique définie positive qui satisfait l'inégalité triangulaire :

- $\Delta(X, Y) \geqslant 0$, avec égalité si et seulement si X et Y sont identiquement distribuées,
- $\Delta(X, Y) = \Delta(Y, X)$,
- $\Delta(X, Z) \leqslant \Delta(X, Y) + \Delta(Y, Z)$.

1. Montrez que si $\Delta(X, Y) = 0$, alors pour tout adversaire $\mathcal{A}$ on a $\mathrm{Adv}_{\mathcal{A}}(X, Y) = 0$.

On rappel la propriété suivante : si X et Y sont deux variable aléatoire sur un ensemble commun A, alors pour toute function (possiblement aléatoire) f de domaine S on a

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

de plus, si f est injective alors on a égalité.

2. Montrez que pour tout adversaire $\mathcal{A}$, $\mathrm{Adv}_{\mathcal{A}}(X, Y) \leqslant \Delta(X, Y)$.

(Yaostep) Exercise 4. *Back to Yao*

Soit $G : \{0,1\}^s \to \{0,1\}^n$ un générateur pseudo-aléatoire. Soit $\mathcal{A}$ un adversaire PPT tel que

$$\mathrm{Adv}_{\mathcal{A}}\left(G\left(U(\{0,1\}^s)\right), U\left(\{0,1\}^n\right)\right) \geqslant \varepsilon.$$

Pour $1 \leq j \leq n$, on definie la distribution D_j sur $\{0,1\}^n$ comme suit: les j premiers bits sont les mêmes que ceux de la sortie de $G(k)$, et les $n - j$ derniers bits sont tirés uniformément.

1. En utilisant un argument hybrid, montrez qu'il existe i tel que $\mathrm{Adv}_{\mathcal{A}}\left(D_i, D_{i+1}\right) \geqslant \varepsilon/n$.

2. Rappelez l'expression de l'avantage de l'adversaire à distinguer D_i de D_{i+1} $\mathrm{Adv}_{\mathcal{A}}(D_i, D_{i+1})$

On souhaite construire un algorithme PPT qui trouve i tel que $\mathrm{Adv}_{\mathcal{A}}(D_i, D_{i+1}) \geqslant \frac{\varepsilon}{2n}$ avec probabilité $\geqslant 1 - 2^{-n}$. On note $p_j = \mathbb{P}\mathbf{r}[\mathcal{A}(D_j) = 1]$ pour tout j.

3. En utilisant une méthode pour approcher la valeur de p_j par $\hat{p}_j$, montrez l'égalité suivante

$$\mathbb{P}\mathbf{r}[|\hat{p}_j - p_j \geqslant v] \leqslant 2e^{-2v^2t} \quad \forall v > 0$$

Hint: Use the Hoeffding bound.

4. Montrez qu'avec probabilité $\geqslant 1 - 4e^{-2v^2t}$,

$$||\hat{p_{j+1}} - \hat{p}_j| - |p_{j+1} - p_j|| \leqslant 2v$$

5. En choisissant $v = \frac{\varepsilon}{8n}$, concluez.