MTAT.07.003 Cryptology II
Spring 2010 / Exercise Session III

1. Let $\mathcal{X}_0$ be a uniform distribution over $\mathbb{Z}_{16}$ and let $\mathcal{X}_1$ be a uniform distribution over $\{0, 2, 4, 6, 8, 10, 12, 14\}$.
   (a) What is the statistical difference between $\mathcal{X}_0$ and $\mathcal{X}_1$?
   (b) Find an distinguishing strategy $\mathcal{A}$ that minimises the ratio of false positives $\beta(\mathcal{A})$ and achieves false negative rate $\alpha(\mathcal{A}) = 0\%$.
   (c) Find an distinguishing strategy $\mathcal{A}$ that minimises the ratio of false positives $\beta(\mathcal{A})$ and achieves false negative rate $\alpha(\mathcal{A}) \le 50\%$.
   (d) Generalise the distinguishing strategy and find minimal ratio of false positives $\beta(\mathcal{A})$ for all bounds $\alpha(\mathcal{A}) \le \alpha_0$.

2. Normally, it is impossible to compute computational distance between two distributions directly since the number of potential distinguishing algorithms is humongous. However, for really small time-bounds it can be done. Here, we assume that all distinguishers $\mathcal{A} : \mathbb{Z}_{16} \to \{0, 1\}$ are implemented as Boolean circuits consisting of NOT, AND and OR gates and the corresponding time-complexity is just the number of logic gates. For example, $\mathcal{A}(x_3x_2x_1x_0) = x_1$ has time-complexity 0 and $\mathcal{A}(x_3x_2x_1x_0) = x_1 \vee \neg x_3 \wedge x_2$ has time-complexity 3.
   (a) Let $\mathcal{X}_0$ be a uniform distribution over $\mathbb{Z}_{16}$ and let $\mathcal{X}_1$ be a uniform distribution over $\{0, 2, 4, 6, 8, 10, 12, 14\}$. What is $\mathsf{cd}_x^1(\mathcal{X}_0, \mathcal{X}_1)$?
   (b) Find a uniform distribution $\mathcal{X}_2$ over some 8 element set such that $\mathsf{cd}_x^1(\mathcal{X}_0, \mathcal{X}_2)$ is minimal. Compute $\mathsf{cd}_x^2(\mathcal{X}_0, \mathcal{X}_2)$ and $\mathsf{cd}_x^3(\mathcal{X}_0, \mathcal{X}_2)$.
   (c) Find a uniform distribution $\mathcal{X}_3$ over some 8 element set such that $\mathsf{cd}_x^1(\mathcal{X}_0, \mathcal{X}_3) + \mathsf{cd}_x^1(\mathcal{X}_0, \mathcal{X}_3)$ is minimal.
   (d) Estimate for which value of t the distances $\mathsf{cd}_x^t(\mathcal{X}_0, \mathcal{X}_1)$ and $\mathsf{sd}_x(\mathcal{X}_0, \mathcal{X}_1)$ coincide for all distributions over $\mathbb{Z}_{16}$.

3. Let $\mathcal{A}$ be a t-time distinguisher and let $\alpha(\mathcal{A}) = \Pr[\mathcal{A} = 1|\mathcal{H}_0]$ and $\beta(\mathcal{A}) = \Pr[\mathcal{A} = 0|\mathcal{H}_1]$ be the ratios of false negatives and false positives. Show that for any c there exists a $t + \mathrm{O}(1)$-time adversary $\mathcal{B}$ such that
$$\alpha(\mathcal{B}) = (1 - c) \cdot \alpha(\mathcal{A}) \qquad \text{and} \qquad \beta(\mathcal{B}) = c + (1 - c) \cdot \beta(\mathcal{A}) \ .$$
   Are there any practical settings where such trade-offs are economically justified? Give some real world examples.

   Hint: What happens if you first throw a fair coin and run $\mathcal{A}$ only if you get tail and otherwise output 0?

($\star$) Let the time-complexity of distinguishing algorithms be defined as in Exercise 2. Find disjoint distributions $\mathcal{X}_0$ and $\mathcal{X}_1$ over $\mathbb{Z}_{256}$ such that their computational distance is minimal. Tabulate the results for time-bounds $0, 1, \dots, 16$. More precisely, find the optimal distribution pair for each time-bound and their computational distance for all time-bounds.

4. Consider the following game, where an adversary $\mathcal{A}$ gets three values x_1, x_2 and x_3. Two of them are sampled from the efficiently samplable distribution $\mathcal{X}_0$ and one of them is sampled from the efficiently samplable distribution $\mathcal{X}_1$. The adversary wins the game if it correctly determines which sample is taken from $\mathcal{X}_1$.

   (a) Find an upper bound to the success probability if distributions $\mathcal{X}_0$ and $\mathcal{X}_1$ are (t, ε)-indistinguishable.

   (b) How does the bound on the success change if we modify the game in the following manner. First, the adversary can first make its initial guess i_0. Then the challenger reveals $j \neq i_0$ such that x_j was sampled from $\mathcal{X}_0$ and then the adversary can output its final guess i_1.

   Hint: How well the adversary can perform if the challenger gives no samples to the adversary? How can you still simulate the game to the adversary who expects these samples?

5. Recall that a game is a two-party protocol between the challenger $\mathcal{G}$ and an adversary $\mathcal{A}$ and that the output of the game $\mathcal{G}^{\mathcal{A}}$ is always determined by the challenger. Prove the following claims:

   (a) Any hypothesis testing scenario $\mathcal{H}$ can be formalised as a game $\mathcal{G}$ such that $\Pr[\mathcal{A} = b|\mathcal{H}] = \Pr[\mathcal{G}^{\mathcal{A}} = b]$ for all adversaries $\mathcal{A}$.

   (b) For two simple hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$, there is a game $\mathcal{G}$ such that

$$\mathsf{cd}^t_\star(\mathcal{H}_0, \mathcal{H}_1) = 2 \cdot \max_{\mathcal{A} \text{ is } t\text{-time}} \left|\Pr[\mathcal{G}^{\mathcal{A}} = 1] - \tfrac{1}{2}\right| \enspace .$$

   (c) The computational distance between games defined as follows

$$\mathsf{cd}_\star(\mathcal{G}_0, \mathcal{G}_1) = \max_{\mathcal{A} \text{ is } t\text{-time}} \left|\Pr[\mathcal{G}_0^{\mathcal{A}} = 1] - \Pr[\mathcal{G}_1^{\mathcal{A}} = 1]\right| \enspace .$$

   Show that this quantity is indeed a pseudo-metric:

$$\begin{aligned} \mathsf{cd}^t_\star(\mathcal{G}_0, \mathcal{G}_1) &= \mathsf{cd}^t_\star(\mathcal{G}_1, \mathcal{G}_0) \enspace , \\ \mathsf{cd}^t_\star(\mathcal{G}_0, \mathcal{G}_2) &\leq \mathsf{cd}^t_\star(\mathcal{G}_0, \mathcal{G}_1) + \mathsf{cd}^t_\star(\mathcal{G}_1, \mathcal{G}_2) \enspace . \end{aligned}$$

   When is the computational distance a proper metric, i.e.,

$$\mathsf{cd}^t_\star(\mathcal{G}_0, \mathcal{G}_1) \neq 0 \qquad \Leftrightarrow \qquad \mathsf{sd}_\star(\mathcal{G}_0, \mathcal{G}_1) \neq 0 \ ?$$

6. Let $\mathcal{X}_0$ and $\mathcal{X}_1$ efficiently samplable distributions that are (t, ε)-indistinguishable. Show that distributions $\mathcal{X}_0$ and $\mathcal{X}_1$ remain computationally indistinguishable even if the adversary can get n samples.

   (a) First estimate computational distances between following games

$$\begin{array}{lll} \mathcal{G}_{00}^{\mathcal{A}} & \mathcal{G}_{01}^{\mathcal{A}} & \mathcal{G}_{11}^{\mathcal{A}} \\ \left[\begin{array}{l} x_0 \leftarrow \mathcal{X}_0 \\ x_1 \leftarrow \mathcal{X}_0 \\ \textbf{return } \mathcal{A}(x_0, x_1) \end{array}\right. & \left[\begin{array}{l} x_0 \leftarrow \mathcal{X}_0 \\ x_1 \leftarrow \mathcal{X}_1 \\ \textbf{return } \mathcal{A}(x_0, x_1) \end{array}\right. & \left[\begin{array}{l} x_0 \leftarrow \mathcal{X}_1 \\ x_1 \leftarrow \mathcal{X}_1 \\ \textbf{return } \mathcal{A}(x_0, x_1) \end{array}\right. \end{array}$$

Hint: What happens if you feed a sample $x_0 \leftarrow \mathcal{X}_0$ together an unknown sample $x_1 \leftarrow \mathcal{X}_i$ to $\mathcal{A}$ and use the reply to guess i.

(b) Generalise the argumentation to the case, where the adversary $\mathcal{A}$ gets n samples from a distribution $\mathcal{X}_i$. That is, define the corresponding sequence of games $\mathcal{G}_{00\ldots0}, \ldots, \mathcal{G}_{11\ldots1}$.

(c) Why do we need to assume that distributions $\mathcal{X}_0$ and $\mathcal{X}_1$ are efficiently samplable?

($\star$) Usually, the statistical distance $\mathsf{sd}_\star(\mathcal{G}_0, \mathcal{G}_1)$ is defined as a limiting value $\mathsf{sd}_\star(\mathcal{G}_0, \mathcal{G}_1) = \lim_{t\to\infty} \mathsf{cd}^t_\star(\mathcal{G}_0, \mathcal{G}_1)$. Express the statistical distance in terms of the distributions of challenger replies

$$p_i(y_i|x_1, y_1, \ldots, x_i) = \Pr\left[\begin{array}{l}\mathcal{G}_i \text{ sends } y \text{ as the } i\text{th message to } \mathcal{A} \text{ given}\\ \text{that preceding messages were } x_1, y_1, \ldots, x_i\end{array}\right]$$

where x_1 be the first message sent by the challenger $\mathcal{G}_i$, y_1 the corresponding reply from the adversary $\mathcal{A}$ and the last message y_n corresponds to the output of the game. Note that there are essentially two types of games. In the interactive hypothesis testing games, the output of $\mathcal{G}_i$ is determined by the last reply x_n of $\mathcal{A}$, i.e., $y_n = x_n$. In other more general types of games, y_n can arbitrarily depend on the previous messages $x_1, \ldots, x_n$ received by the challenger $\mathcal{G}_i$.

($\star$) Prove that (t, ε)-pseudorandom generators $f : \{0,1\}^n \to \{0,1\}^m$ exist for sufficiently big values of m and n, if we do not limit the computational complexity of the function f. Give an interpretation to this result.

Hint: First prove that there are only finite number of t-time adversaries and that these adversaries can perfectly distinguish only a fixed number functions $f : \{0,1\}^n \to \{0,1\}^m$ for any number of m, n.

($\star$) Let $f : \mathcal{S} \to \{0,1\}^*$ be an efficiently predictable from $f(s)$. That is, there exists a t-time algorithm that achieves

$$\mathsf{Adv}^{\mathsf{sem}}_{f,f}(\mathcal{A}) = \Pr\left[s \leftarrow \mathcal{S} : \mathcal{A}(f(s)) = f(s)\right] - \Pr\left[s \leftarrow \mathcal{S} : f(s) = f(s)\right] \geq \varepsilon$$

for some probability distribution over $\mathcal{S}$. Prove that there exist a $2t$ algorithm $\mathcal{B}$ and two states $s_0, s_1 \in \mathcal{S}$ such that $\mathsf{Adv}^{\mathsf{ind}}_{f(s_0),f(s_1)}(\mathcal{B}) \geq \varepsilon$. Conclude that f cannot be deterministic and $\Pr\left[f(s) = y\right] \leq \varepsilon$ for an invertible random function f. State the last result in terms of min-entropy.