ETH Zürich, D-INFK
Spring 2018

Prof. Ueli Maurer
Christian Badertscher
Fabio Banfi

Cryptography Foundations
Solution Exercise 8

8.1 Changing the Distribution of Bit-Guessing Problems

a) Recall the solution to Exercise 1.3 a): We defined the set $\mathcal{X}^* := \{x \in \mathcal{X} \mid \mathsf{P}_X(x) \geq \mathsf{P}_Y(x)\}$ and proved that $\delta(X, X') = \Pr[X \in \mathcal{X}^*] - \Pr[X' \in \mathcal{X}^*]$. Stated differently, $\mathcal{X}$ is the set of elementary events and $\mathcal{X}^* \subseteq \mathcal{X}$ is a particular event and we write $\delta(X, X') = \Pr^X[\mathcal{X}^*] - \Pr^{X'}[\mathcal{X}^*]$. As explained in that solution, the maximum likelihood method allowed us to derive this event $\mathcal{X}^*$ which maximizes the term $\Pr^X[\mathcal{X}^*] - \Pr^{X'}[\mathcal{X}^*]$. Briefly, adding an additional element $x \in \mathcal{X} \setminus \mathcal{X}^*$ to the set $\mathcal{X}^*$ would only decrease the term (because $\mathsf{P}_X(x) - \mathsf{P}_{X'}(x) < 0$) and also removing an element from the set $\mathcal{X}^*$ would only decrease the term (because $\mathsf{P}_X(x) - \mathsf{P}_{X'}(x) \geq 0$).

More generally, for two probability spaces $(\Omega, \mathcal{F}, \mathsf{P}_X)$ and $(\Omega, \mathcal{F}, \mathsf{P}_{X'})$ we have that

$$\delta(X, X') = \sup_{\mathcal{B} \in \mathcal{F}} \left| \Pr^X[\mathcal{B}] - \Pr^{X'}[\mathcal{B}] \right|,$$

which is another common formulation of the statistical distance. (The text above actually is the proof for the special case in which we have the (finite) sample space $\Omega = \mathcal{X}$ and event set $\mathcal{F} = 2^\Omega$, i.e., full information. This is the typical case in this lecture.)

It is not hard to see that for any event $\mathcal{A} \in \mathcal{F}$ we have

$$\Pr^X[\mathcal{A}] - \Pr^{X'}[\mathcal{A}] \leq \sup_{\mathcal{B} \in \mathcal{F}} \left(\Pr^X[\mathcal{B}] - \Pr^{X'}[\mathcal{B}] \right) \leq \sup_{\mathcal{B} \in \mathcal{F}} \left| \Pr^X[\mathcal{B}] - \Pr^{X'}[\mathcal{B}] \right| = \delta(X, X').$$

b) Exercise 4.4 in the lecture notes asks to show that for a bit-guessing problem (S, B) and a distinguisher D for it, if one changes the instance distribution of (S, B) by at most d in terms of statistical distance, then the performance of D changes by at most $2d$. The performance of D is measured in terms of its advantage $\Lambda^D\big((S, B)\big)$. Changing the instance distribution of (S, B) as described above means considering a new bit-guessing problem (S', B') such that $d = \delta((S, B), (S', B'))$. We assume without loss of generality that the output bit B of S is a deterministic function of S and thus the statistical distance of $\delta((S, f(S)), (S', f(S'))$ is no greater than $\delta(S, S')$ as we know from a previous exercise.

In summary: what we want to prove is in this case

$$\Lambda^D\big((S, B)\big) = \Lambda^D\big((S', B')\big) + 2 \cdot \delta(S, S').$$

Consider the random experiment $D(S, B)$, i.e., a distinguisher D interacting with system S (which outputs bit B) and outputs a guess Z, as a probability space where the elementary events correspond to sampling D and sampling S. All properties, including the event $\mathcal{A} \coloneqq Z = B$ are deterministic functions when given these (sampled) problem instance and distinguisher. From subtask **a)**, we conclude that

$$\begin{aligned}
\Lambda^D\big((S, B)\big) - \Lambda^D\big((S', B')\big) &= 2 \cdot \Pr^{D(S,B)}[Z = B] - 1 - (2 \cdot \Pr^{D(S',B')}[Z' = B'] - 1) \\
&= 2 \cdot (\Pr^{D(S,B)}[\mathcal{A}] - \Pr^{D(S',B')}[\mathcal{A}]) \\
&\leq 2 \cdot \delta((D, S), (D, S')) \leq 2 \cdot \delta(S, S').
\end{aligned}$$

Note that $Z = B$ and $Z' = B'$ denote the same event in the two experiments (expressed as a function of D and S)[1]. The final step that $\delta((D,S),(D,S')) \leq \delta(S,S')$ follows from a simple property of the statistical distance (analog to one of the properties proven on the previous exercise sheet) since by definition of the random experiment, D and S (resp. S') are sampled independently.

8.2 Amplifying the Performance of a Worst-Case Solver

Let X_i for $i \in \{1,\dots,q\}$ be the binary random variable that is 1 if the ith invocation of S returns the correct bit. Since S has performance ϵ, we have $p := \Pr[X_i = 1] = \frac{\epsilon}{2} + \frac{1}{2}$. Note that all X_i are independent and that the solver T outputs the wrong bit if and only if S outputs more wrong than correct bits. That is, the probability that T outputs the wrong bit is $\Pr\big[\sum_{i=1}^{q} X_i < \frac{q}{2}\big]$. Let $\alpha := \frac{\epsilon}{2} = p - \frac{1}{2}$. We then obtain for the probability that T outputs the wrong bit using Hoeffding's inequality

$$\Pr\left[\sum_{i=1}^{q} X_i < \frac{q}{2}\right] = \Pr\left[\sum_{i=1}^{q} X_i \leq (p-\alpha)q\right] \leq e^{-2\alpha^2 q} = e^{-q\epsilon^2/2}.$$

For $q \geq \frac{2}{\epsilon^2} \cdot \log\frac{2}{\delta}$, we have

$$e^{-q\epsilon^2/2} \leq e^{-\log(2/\delta)} = e^{\log(\delta/2)} = \frac{\delta}{2}.$$

Hence, the success probability of T for such q is at least $1 - \frac{\delta}{2}$, and the performance of T is at least $1 - \delta$.

8.3 The Next Bit Test

Recall that for an integer i the notation a^i denotes the sequence $a_1,\dots,a_i$, and that we denote its concatenation with another sequence b^j (namely, the sequence $a_1,\dots,a_i,b_1,\dots,b_j$) as $a^i\, b^j$. For this task we further introduce the following notation: for integers $i \leq j$, we write $a^{i\,:\,j}$ to denote the sequence $a_i, a_{i+1},\dots,a_j$ (note that $a^{j\,:\,i}$ would correspond to the empty sequence). We now describe how to construct a predictor P_i, with $i \in \{1,\dots,\ell\}$, for the i-th bit of an arbitrarily distributed bit-string X^ℓ. First, P_i receives the (partial) bit-string X^{i-1}. Then it samples the bit-string $U^{i\,:\,\ell}$ uniformly at random (i.e., each bit $U_i,\dots,U_\ell$ is distributed independently and uniformly at random). P_i then proceeds by invoking D on input the bit-string $X^{i-1}\, U^{i\,:\,\ell}$. Upon D outputting a guess bit Z, P_i outputs as its guess for X_i the bit $Z \oplus U_i$.

Before analyzing the advantage of the predictor P_i, let introduce the following hybrid sequences:

$$\mathbf{H}_k := X^k\, U^{k+1\,:\,\ell} \tag{1}$$

Note that for the extreme cases we have

$$\mathbf{H}_0 = U^\ell \qquad \text{and} \qquad \mathbf{H}_\ell = X^\ell. \tag{2}$$

[1]This means that we can identify the subset of pairs of deterministic systems from the product space $\mathcal{D} \times \mathcal{S}$ for which the output bit of the distinguisher equals the bit of the bit-guessing problem.

Then for any $i \in \{1, \ldots, \ell\}$ we have:

$$\begin{aligned}
\Lambda^{P_i}\big((X^{i-1}, X_i)\big) &= 2 \cdot \Pr^{P_i\,(X^{i-1},X_i)}[Z' = X_i] - 1 \\
&= 2 \cdot \Pr^{P_i\,(X^{i-1},X_i)}[Z \oplus U_i = X_i] - 1 \\
&= 2 \cdot \Big(\Pr^{P_i\,(X^{i-1},X_i)}[Z = X_i \oplus U_i \,|\, U_i = X_i] \cdot \frac{1}{2} \\
&\qquad + \Pr^{P_i\,(X^{i-1},X_i)}[Z = X_i \oplus U_i \,|\, U_i \neq X_i] \cdot \frac{1}{2}\Big) - 1 \\
&= \Pr^{D\,(X^{i-1}\,X_i\,U^{i+1\,:\,\ell})}[Z = 0] + \Pr^{D\,(X^{i-1}\,\overline{X_i}\,U^{i+1\,:\,\ell})}[Z = 1] - 1 \\
&= \Pr^{D\,(X^{i-1}\,\overline{X_i}\,U^{i+1\,:\,\ell})}[Z = 1] - \Pr^{D\,(X^{i-1}\,X_i\,U^{i+1\,:\,\ell})}[Z = 1] \\
&= \Delta^D(X^{i-1}\,X_i\,U^{i+1\,:\,\ell}, X^{i-1}\,\overline{X_i}\,U^{i+1\,:\,\ell}).
\end{aligned}$$

Now consider a (probabilistic) system S which outputs the sequence $X^{i-1}\,U^{i+1\,:\,\ell}$. Recall from Exercise 1.3 b) that for a bit B correlated with S and an independent and uniformly distributed bit U, we have

$$\Delta^D\big((S, B), (S, U)\big) = \frac{1}{2} \cdot \Delta\big((S, B), (S, \overline{B})\big). \tag{3}$$

Therefore, since X_i is indeed correlated with S, whereas U_i is independent and uniformly distributed, from (3) we get

$$\Delta^D(X^{i-1}\,X_i\,U^{i+1\,:\,\ell}, X^{i-1}\,\overline{X_i}\,U^{i+1\,:\,\ell}) = 2 \cdot \Delta^D(X^i\,U^{i+1\,:\,\ell}, X^{i-1}\,U^{i\,:\,\ell}).$$

Putting things together, using (1) we have

$$\Lambda^{P_i}\big((X^{i-1}, X_i)\big) = 2 \cdot \Delta^D(\mathbf{H}_i, \mathbf{H}_{i-1}).$$

Finally, using (a slight variation of) Lemma 2.2 and (2), we have

$$\sum_{i=1}^{\ell} \Lambda^{P_i}\big((X^{i-1}, X_i)\big) = 2 \cdot \sum_{i=1}^{\ell} \Delta^D(\mathbf{H}_i, \mathbf{H}_{i-1}) = 2 \cdot \Delta^D(\mathbf{H}_\ell, \mathbf{H}_0) = 2 \cdot \Delta^D(X^\ell, U^\ell),$$

and thus it follows that not all predictors P_i can have advantage less than $\frac{2}{\ell} \cdot \Delta^D(X^\ell, U^\ell)$. Turned around, this means that there exists an $i \in \{1, \ldots, \ell\}$ and a predictor P_i for X^ℓ such that

$$\Lambda^{P_i}\big((X^{i-1}, X_i)\big) \geq \frac{2}{\ell} \cdot \Delta^D(X^\ell, U^\ell).$$