Homework 1

This problem set is due in class on **Tuesday, January 31.**

It contains one "long answer" question, where you are requested to provide full proofs, and two "short answer" questions, where you are expected to check on your own that you could do a careful proof if you had to, but where you only need to write down a one-word or one-sentence answer.

1 Long Answer: Computational indistinguishability

We define a distance measure on probability distributions as follows:

$$d_R(D, D') = \max_A |\Pr_{x \leftarrow D}[A(x) = 1] - \Pr_{x' \leftarrow D'}[A(x') = 1]|$$

where the maximum is taken over all algorithms running with resources R. Here R denotes a set of resources, which is used to restrict the adversary. For instance, R might list a bound on number of steps of computation, on queries to the oracle, on the amount of memory used, or something else. If algorithm A runs with resources R and A' with R', then let $R + R'$ denote the resources used by running first A then A'. If you like, you can assume that $R + R' = R' + R$.

Note that D, D' are (t, e)-indistinguishable if and only if $d_t(D, D') \leq e$, so we can see that this distance measure is of fundamental interest in cryptography. This problem will ask you to study basic properties of this distance measure.

Provide careful formal proofs of each of the following.

(a) [*Data processing inequality*] Let L be any (possibly randomized) algorithm running with resources R', and let $L(D)$ denote the distribution obtained by picking x from D and then outputting $L(x)$. Prove: $d_R(L(D), L(D')) \leq d_{R'+R}(D, D')$.

(b) [*Triangle inequality*] Let D' be any distribution whatsoever. Prove: $d_R(D, D'') \leq d_R(D, D') + d_R(D', D'')$.

(c) [*Multiple samples*] If D_1, D_2 are distributions, let $D_1 \times D_2$ denote the distribution on pairs whose first component is chosen according to D_1 and second component according to D_2 (independently), i.e., $\Pr_{D_1 \times D_2}[(x_1, x_2)] = \Pr_{D_1}[x_1] \times \Pr_{D_2}[x_2]$.

Prove: $d_R(D \times D, D' \times D') \leq 2 d_R(D, D')$.

(d) [*Information-theoretic security and the variation distance*] The variation distance is defined by $V(D_1, D_2) = \frac{1}{2} \sum_x |\Pr_{D_1}[x] - \Pr_{D_2}[x]|$. When used as a resource bound, let ∞ denote that there are no restrictions on the adversary (e.g., the algorithm can have unbounded running time).

Prove: $d_\infty(D, D') = V(D, D')$.

(e) [*Information-theoretic security*] Prove: $d_R(D, D') \leq d_\infty(D, D')$.

(f) [*Conditioning*] If E is an event, let $D|E$ represent the distribution of the output of D conditioned on E occurring, i.e., $\Pr_{D|E}[x] = \Pr_D[x|E] = \Pr_D[x \text{ and } E] / \Pr[E]$. Let $\neg E$ denote the complementary event to E, i.e., that E does not occur.

Prove: $d_R(D, D') \leq d_R(D|E, D'|E) + \Pr[\neg E]$.

(g) [*A XOR lemma*] If D_1, D_2 are distributions on k-bit strings, let $D_1 \oplus D_2$ denote the distribution on $x_1 \oplus x_2$ when x_1 is chosen from D_1 and x_2 is chosen from D_2 (independently). Here $\oplus$ denotes the xor operator. In other words, $\Pr_{D_1 \times D_2}[y] = \sum \Pr_{D_1}[x_1] \times \Pr_{D_2}[x_2]$, where the sum is taken over all pairs (x_1, x_2) such that $x_1 \oplus x_2 = y$. Let U denote the uniform distribution on k-bit strings.

Prove: $d_R(D \oplus D', U) \le 2 d_\infty(D, U) d_\infty(D', U)$.

2 Short Answer: Indistinguishability and guessing

Recall that the distributions D_0, D_1 are said to be (t, ϵ)-indistinguishable (for $0 \le \epsilon \le 1$) if

$$|\Pr_{x_0}[x_0 \stackrel{\$}{\leftarrow} D_0; A(x_0) = 1] - \Pr_{x_1}[x_1 \stackrel{\$}{\leftarrow} D_1; A(x_1) = 1]| \le \epsilon$$

holds for all adversaries A running in time at most t, where the random variable x_0 is distributed according to D_0 and x_1 is distributed like D_1.

Now, let's call the distributions D_0, D_1 (t, ϵ)-*inseparable* just if

$$\frac{1}{2} - \frac{\epsilon}{2} \le \Pr_{b,x}[b \stackrel{\$}{\leftarrow} \{0,1\}; x \stackrel{\$}{\leftarrow} D_b; A(x) = b] \le \frac{1}{2} + \frac{\epsilon}{2}$$

holds for all adversaries A running in time at most t, where the random variable b is a uniformly random bit and where the random variable x is distributed according to D_b. This is a very natural notion, because it talks about our chances of guessing correctly which distribution x came from, and whether we can do much better than simply flipping a coin.

It turns out that there is a universal constant $c > 0$ so that D_0, D_1 are (t, ϵ)-indistinguishable if and only if they are $(t, c\epsilon)$-inseparable. (Hence the notion of inseparability is redundant.)

What is the value of c?

(You should check on your own that you are able to prove the "if and only if" statement, but there is no need to turn in your proof.)

3 Short Answer: Correlation

Let X, Y, Z be three random variables. Suppose that $X \sim Y$ (i.e., X, Y are (t, ϵ)-indistinguishable for some t and some ϵ that is not too large).

True or false: It necessarily follows that $(X, Z) \sim (Y, Z)$ (i.e., (X, Z) and (Y, Z) are (t', ϵ')-indistinguishable for some t' and some ϵ' that is not too large).

Give a concise one-sentence explanation why or why not. (You should check on your own that you are able to prove your answer, but there is no need to turn in your proof.)