1 Indistinguishability

We'll now start a major unit on *indistinguishability* and *pseudorandomness*. These concepts are a cornerstone of modern cryptography, underlying several foundational applications such as pseudorandom generators, secure encryption, "commitment" schemes, and much more.

For example, our most immediate application of indistinguishability will be to construct cryptographically strong *pseudorandom bit generators*. These are algorithms that produce many "random-looking" bits, while using very little "true" randomness. (In particular, the bits they output will necessarily be "very far" from truly random, in a statistical sense.) One easy-to-imagine application would be to use the pseudorandom bit stream as a one-time encryption pad, which would allow the shared secret key to be much smaller than the message. But what does it *mean* for a string of bits to be "random-looking"? And how can we be confident that using such bits does not introduce any unforeseen weaknesses in our system?

More generally, the motivating question for our study is:

When can two (possibly different) objects be considered *effectively the same*?

The answer:

When they can't be told apart!

Though seemingly glib, this answer encapsulates a very powerful mindset that will serve us well as we go forward.

1.1 Statistical Indistinguishability

We use probability theory to model (in)distinguishability. If two distributions are identical, then they certainly should be considered indistinguishable. We relax this condition to define *statistical* indistinguishability, for when the *statistical distance* between the two distributions is negligible. The statistical distance between two distributions X and Y over a domain Ω is defined as[1]

$$\Delta(X,Y) := \sup_{A \subseteq \Omega} |X(A) - Y(A)|,$$

where $X(A) = \sum_{w \in A} \Pr[X = w]$ is the probability that a draw from X lands in A, and likewise for $Y(A)$. We can view $A \subseteq \Omega$ as a statistical "test" that has some probability of "passing" when given an element drawn from X, or from Y; the statistical distance is essentially the maximum difference between these two probabilities, taken over all tests. Note that A and $\bar{A}$ are effectively the same test, since

$$|X(\bar{A}) - Y(\bar{A})| = |1 - X(A) - (1 - Y(A))| = |Y(A) - X(A)| = |X(A) - Y(A)|.$$

Lemma 1.1. *For distributions X, Y over a finite domain Ω,*

$$\Delta(X,Y) = \frac{1}{2} \sum_{w \in \Omega} |X(w) - Y(w)|.$$

[1] For a set S of real numbers, its *supremum* $\sup(S)$ is the least upper bound of the elements of S. The suprenum can be thought of as a generalization of the maximum element to (infinite) sets where no such element may exist. For example, there is no maximum element of the set $[0,1) \subseteq \mathbb{R}$, but the suprenum is 1 since it is the smallest real number that is larger than every element of $[0,1)$. When S is finite, the supremum is simply the maximum.

Proof. Let the test $A = \{w \in \Omega \ : \ X(w) > Y(w)\}$. This makes $X(A) - Y(A)$ as large as possible, so $\Delta(X,Y) = X(A) - Y(A) = \sum_{w \in A} |X(w) - Y(w)|$. As noted above, we also have $\Delta(X,Y) = Y(\bar{A}) - X(\bar{A}) = \sum_{w \in \bar{A}} |X(w) - Y(w)|$. Summing the two equations, we have $2\Delta(X,Y) = \sum_{w \in \Omega} |X(w) - Y(w)|$, as desired. □

Statistical distance is very robust, which enhances its usefulness. Using Lemma 1.1, the following facts are straightforward to prove.

Lemma 1.2. *Let f be a function (or more generally, randomized procedure) on the domain of X, Y. Then $\Delta(f(X), f(Y)) \leq \Delta(X,Y)$.*

In other words, statistical distance cannot be increased by the application of any (randomized) procedure.

Lemma 1.3. *Statistical distance is a metric; in particular, it satisfies the following three properties:*

- *Identity of indiscernibles:* $\Delta(X,Y) = 0 \iff X = Y$
- *Symmetry:* $\Delta(X,Y) = \Delta(Y,X)$
- *Subadditivity (triangle inequality):* $\Delta(X,Z) \leq \Delta(X,Y) + \Delta(Y,Z)$

Question 1. Justify why it is the case that $\Delta(X,Y) = 0 \iff X = Y$.

Statistical distance lets us say when two (sequences of) distributions are "essentially the same," in an asymptotic sense.

Definition 1.4. Let $\mathcal{X} = \{X_n\}_{n \in \mathbb{N}}$ and $\mathcal{Y} = \{Y_n\}_{n \in \mathbb{N}}$ be sequences of probability distributions, called *ensembles*. We say that $\mathcal{X}$ and $\mathcal{Y}$ are *statistically indistinguishable*, written $\mathcal{X} \overset{s}{\approx} \mathcal{Y}$, if

$$\Delta(X_n, Y_n) = \mathrm{negl}(n).$$

Example 1.5. Let X_n be the uniform distribution over $\{0,1\}^n$, and let Y_n be the uniform distribution over the nonzero strings $\{0,1\}^n \backslash \{0^n\}$. An optimal test A is the singleton set $A = \{0^n\}$, yielding $\Delta(X_n, Y_n) = 2^{-n} = \mathrm{negl}(n)$, so $\mathcal{X} \overset{s}{\approx} \mathcal{Y}$. (This can also be seen by calculating the summation in Lemma 1.1.) The analysis extends similarly to any Y_n that leaves out a $\mathrm{negl}(n)$ fraction of $\{0,1\}^n$. From this we can say that such ensembles $\mathcal{Y}$ are "essentially uniform," or *statistically pseudorandom*.

Can a statistically peudorandom generator exist? This depends on the definition of "generator" (which we give below), but for any meaningful definition of the term, it isn't possible! This can be shown by explicitly demonstrating an appropriate subset (or test) that distinguishes strings output by the generator from uniformly random ones; see Question 3 below.

1.2 Computational Indistinguishability

We can define a natural analogue of statistical distance in the computational setting, where the "test" is implemented by an *efficient algorithm*. Namely, for distributions X and Y and an algorithm $\mathcal{A}$ (possibly randomized), define $\mathcal{A}$'s *distinguishing advantage* between X and Y as

$$\mathbf{Adv}_{X,Y}(\mathcal{A}) = |\Pr[\mathcal{A}(X) = 1] - \Pr[\mathcal{A}(Y) = 1]|.$$

(The output of $\mathcal{A}$ can be arbitrary, but we interpret 1 as a special output indicating that the test implemented by $\mathcal{A}$ is "satisfied," and any other output as "not satisfied.") We extend this to ensembles $\mathcal{X}$ and $\mathcal{Y}$, making $\mathbf{Adv}_{\mathcal{X},\mathcal{Y}}(\mathcal{A})$ a function of $n \in \mathbb{N}$.

Definition 1.6. Let $\mathcal{X} = \{X_n\}$ and $\mathcal{Y} = \{Y_n\}$ be ensembles, where X_n and Y_n are distributions over $\{0,1\}^{l(n)}$ for $l(n) = \mathrm{poly}(n)$. We say that $\mathcal{X}$ and $\mathcal{Y}$ are *computationally indistinguishable*, written $\mathcal{X} \overset{c}{\approx} \mathcal{Y}$, if $\mathbf{Adv}_{\mathcal{X},\mathcal{Y}}(\mathcal{A}) = \mathrm{negl}(n)$ for all non-uniform PPT algorithms $\mathcal{A}$. We say that $\mathcal{X}$ is (computationally) *pseudorandom* if $\mathcal{X} \overset{c}{\approx} \{U_{l(n)}\}$, the ensemble of uniform distributions over $\{0,1\}^{l(n)}$.

The basic facts about statistical distance also carry over to computational indistinguishability, where we restrict all functions/tests to be *efficient*. We call the analogue of Lemma 1.2 the *composition lemma*, because it involves the composition of an applied procedure and a distinguisher. We call the analogue of the triangle inequality the *hybrid lemma*, because (as we will see soon) we usually invoke it on a sequence of "hybrid" distributions that interpolate step-by-step between two distributions we ultimately care about.

Lemma 1.7 (Composition lemma). *Let $\mathcal{X} = \{X_n\}$ and $\mathcal{Y} = \{Y_n\}$ be ensembles, $\mathcal{S}$ be any non-uniform PPT algorithm, and $\mathcal{X}' = \{X'_n = \mathcal{S}(X_n)\}$ and $\mathcal{Y}' = \{Y'_n = \mathcal{S}(Y_n)\}$. Then for any non-uniform PPT algorithm $\mathcal{D}'$, we have that $\mathbf{Adv}_{\mathcal{X}',\mathcal{Y}'}(\mathcal{D}') = \mathbf{Adv}_{\mathcal{X},\mathcal{Y}}(\mathcal{D}' \circ \mathcal{S})$. In particular, if $\mathcal{X} \overset{c}{\approx} \mathcal{Y}$, then $\mathcal{X}' \overset{c}{\approx} \mathcal{Y}'$.*

Proof. Consider the composed algorithm $\mathcal{D} = \mathcal{D}' \circ \mathcal{S}$, i.e., on input x, it runs $\mathcal{D}(\mathcal{S}(x))$ and outputs the same answer. By construction, we have that

$$\begin{aligned}\mathbf{Adv}_{\mathcal{X}',\mathcal{Y}'}(\mathcal{D}') &= |\Pr[\mathcal{D}'(X'_n) = 1] - \Pr[\mathcal{D}'(Y'_n) = 1]| \\ &= |\Pr[\mathcal{D}'(\mathcal{S}(X_n)) = 1] - \Pr[\mathcal{D}'(\mathcal{S}(Y_n)) = 1]| \\ &= \mathbf{Adv}_{\mathcal{X},\mathcal{Y}}(\mathcal{D}),\end{aligned}$$

as claimed. In particular, if $\mathcal{X} \overset{c}{\approx} \mathcal{Y}$, then because $\mathcal{D}$ is non-uniform PPT by construction, the right-hand side is $\mathrm{negl}(n)$, hence so is the left-hand side. □

Lemma 1.8 (Hybrid lemma). *Let X^i for $i \in \{0, 1, \ldots, m\}$ be distributions for some positive integer m. Then for* any *algorithm $\mathcal{D}$ (regardless of efficiency),*

$$\mathbf{Adv}_{X^0,X^m}(\mathcal{D}) \leq \sum_{i \in [m]} \mathbf{Adv}_{X^{i-1},X^i}(\mathcal{D}).$$

In particular, let $\mathcal{X}^i = \{X^i_n\}$ be m ensembles, where m is a constant independent of n. If $\mathcal{X}^{i-1} \overset{c}{\approx} \mathcal{X}^i$ for every $i \in [m]$, then $\mathcal{X}^0 \overset{c}{\approx} \mathcal{X}^m$.

Proof. For each i, let $p_i = \Pr[\mathcal{D}(X^i) = 1]$. By the triangle inequality, we can write $\mathcal{D}$'s advantage as

$$\mathbf{Adv}_{X^0,X^m}(\mathcal{D}) = |p_0 - p_m| \leq \sum_{i \in [m]} |p_{i-1} - p_i| = \sum_{i \in [m]} \mathbf{Adv}_{X^{i-1},X^i}(\mathcal{D}).$$

Now consider the ensembles $\mathcal{X}^i$, and restrict to non-uniform PPT $\mathcal{D}$. By assumption, each $\mathbf{Adv}_{\mathcal{X}^{i-1},\mathcal{X}^i}(\mathcal{D}) = \nu_i(n)$, where $\nu_i(n)$ is a negligible function that may be *different for each* $i \in [m]$. Fortunately, the sum of any *constant* number of negligible functions is indeed negligible: letting $\nu(n) = \sum_i \nu_i(n)$, we need to show that for all $c > 0$, there exists some constant n_0 such that $\nu(n) \leq n^{-c}$ for all $n \geq n_0$. We know that for each $i \in [m]$, there is a constant n_i such that $\nu_i(n) \leq n^{-(c+1)} \leq n^{-c}/m$ for all $n \geq n_i$. Letting $n_0 = \max_{i \in [m]} n_i$ be the largest of these—which is a constant (independent of n) because m is constant—it follows that $\nu(n) \leq n^{-c}$ for all $n \geq n_0$, as desired. □

Remark 1.9. In the proof of the lemma, we used the triangle inequality on the quantities $\mathbf{Adv}_{X^{i-1},X^i}(\mathcal{D})$ to conclude something about $\mathbf{Adv}_{X^0,X^m}(\mathcal{D})$. Syntactically this is unremarkable, but observe closely what we have done: even though $\mathcal{D}$'s "goal in life" is to distinguish between X^0 and X^m, by referring to the quantities $\mathbf{Adv}_{X^{i-1},X^i}(\mathcal{D})$, we are implicitly considering how $\mathcal{D}$ behaves on *all* the hybrid distributions X^i—even ones on which $\mathcal{D}$ was never "designed" to run! Yet because $\mathcal{D}$ is "just an algorithm," we can run it and use it for whatever purposes we like. The hybrid lemma says that in order for $\mathcal{D}$ to distinguish (with non-negligible advantage) between the ensembles $\mathcal{X}^0$ and $\mathcal{X}^m$, it must also distinguish between $\mathcal{X}^i$ and $\mathcal{X}^{i+1}$ for some i, which is impossible by hypothesis.

Importantly, the asymptotic part of Lemma 1.8 considers only a *constant* number of hybrid ensembles, because it critically relies on the fact that the sum of a constant number of negligible functions is negligible. However, in many applications we will need to consider a number of hybrid distributions that *grows with the security parameter* n. This is considerably more subtle: first, since the distributions in an ensemble are themselves indexed by n, it is not so clear how to meaningfully define a number of ensembles that varies with n. Second, it turns out that the sum of $\mathrm{poly}(n)$-many *different* negligible functions may be non-negligible![2]

The following corollary side-steps the first issue by considering ensembles for only the "endpoint" distributions, not the intermediate hybrids, and deals with the second issue by requiring that each advantage term be bounded by the *same* negligible function. This will indeed be the case in all the applications we consider, because all such terms will arise from applying the composition lemma (Lemma 1.7) for the *same pair* of indistinguishable "source" ensembles $\mathcal{X}, \mathcal{Y}$, with the *same (composed) distinguisher* $\mathcal{D} = \mathcal{D}' \circ \mathcal{S}$. (Or, slightly more generally, a constant number of indistinguishable pairs, with one composed distinguisher for each.)

Corollary 1.10. *Let $\mathcal{X} = \{X_n\}$ and $\mathcal{Y} = \{Y_n\}$ be ensembles, and H_n^i be "hybrid" distributions for $i \in \{0, \ldots, m\}$ where $m = \mathrm{poly}(n)$ may vary with n, and where $H_n^0 = X_n$ and $H_n^m = Y_n$. If $\mathbf{Adv}_{H_n^{i-1},H_n^i}(\mathcal{D}) \leq \nu(n)$ for all $i \in [m]$ and a* single *negligible function $\nu(n) = \mathrm{negl}(n)$, then $\mathbf{Adv}_{\mathcal{X},\mathcal{Y}}(\mathcal{D}) = \mathrm{negl}(n)$ as well.*

Proof. This follows immediately from (the first part of) Lemma 1.8 and the fact that $m \cdot \nu(n) = \mathrm{negl}(n)$, because $m = \mathrm{poly}(n)$. □

2 Pseudorandom Generators

Definition 2.1. A *deterministic* function $G : \{0,1\}^* \to \{0,1\}^*$ is a *pseudorandom generator* (PRG) with output length $\ell(n) > n$ if:

- G can be computed by a polynomial-time algorithm,
- $|G(x)| = \ell(|x|) > |x|$ for all $x \in \{0,1\}^*$, and
- the ensemble $\{G(U_n)\}$ is (computationally) pseudorandom.

This last property essentially says that $\{G(U_n)\}$ and $\{U_{\ell(n)}\}$ are computationally indistinguishable. By the composition lemma, it follows that $G(U_n)$ can be used in place of $U_{\ell(n)}$ in *any* (efficient) application!

[2] Consider the sequence of functions ν_i for $i \in \mathbb{N}$, where $\nu_i(n) = 1$ if $n = i$, and $\nu_i(n) = 0$ otherwise. Then each individual ν_i is negligible, because $\nu_i(n) = 0$ for all $n > n_0 := i$. However, defining $\nu(n) = \sum_{i=1}^{n} \nu_i(n)$ as the sum of the first n functions in the sequence, we have that $\nu(n) = 1$ for all n, which is clearly not negligible. The proof of Lemma 1.8 fails here because the n_0 defined therein is not a constant, but instead depends on n.

Question 2. Let G be a PRG. Is $H(x) := \overline{G(x)}$ *necessarily* a PRG as well? *Hint:* think about the composition lemma.

Question 3. Prove (or at least sketch a proof) that a *statistically* pseudorandom generator cannot exist. *Hint:* think about the sizes of $G(\{0,1\}^n)$ and $\{0,1\}^{\ell(n)}$, and define an appropriate statistical test (i.e., subset of $\{0,1\}^{\ell(n)}$).

2.1 Expansion of a PRG

From the definition, it is easy to see that the "weakest" PRG we could ask for would be one that stretches its input by just 1 bit, i.e., $\ell(n) = n + 1$. Is there an upper limit on how much a PRG can stretch? The following theorem says that there is (effectively) *no limit*: if you can stretch by even just 1 bit, then you can stretch by essentially any (polynomial) amount!

Theorem 2.2. *Suppose there exists a PRG G with expansion $\ell(n) = n + 1$. Then for any polynomial $t(\cdot) = \mathrm{poly}(n)$, there exists a PRG $G_t : \{0,1\}^n \to \{0,1\}^{t(n)}$.*

We will prove this theorem in the next lecture.

Remark 2.3. This theorem says something extremely strong. Observe that the image $\{G_t(s) : s \in \{0,1\}^n\}$ of G_t is an extremely small fraction $2^{n-t(n)}$ of its range set $\{0,1\}^{t(n)}$. Yet no computationally bounded algorithm can distinguish a random element from this small subset, from a truly random one over the whole space!

Answers

Question 1. Justify why it is the case that $\Delta(X, Y) = 0 \iff X = Y$.

Answer. In the forward direction, because $\Delta(X, Y) = 0$ and all the terms in the sum from Lemma 1.1 are non-negative, we have that $|X(w) - Y(w)| = 0$ for all $w \in \Omega$, i.e., $X(w) = Y(w)$. Hence, $X = Y$. In the other direction, if $X = Y$, then by definition $X(A) - Y(A) = 0$ for all A, so $\Delta(X, Y) = \sup(\{0\}) = 0$.

Question 2. Let G be a PRG. Is $H(x) := \overline{G(x)}$ *necessarily* a PRG as well? *Hint:* think about the composition lemma.

Answer. Yes. The intuition here is that if $G(x)$ looks random, then its complement $\overline{G(x)}$ should look like the complement of a uniformly random string, which itself is uniformly random.

The formal proof is a straightforward application of the composition lemma. Since G is a PRG, we have $G(U_n) \overset{c}{\approx} U_{\ell(n)}$, where $\ell(n)$ is G's expansion. Consider the efficient algorithm $\mathcal{S}(x) := \overline{x}$. Observe that $H(x) = \overline{G(x)} = \mathcal{S}(G(x))$ for any x. By the composition lemma, we have

$$H(U_n) = \mathcal{S}(G(U_n)) \overset{c}{\approx} \mathcal{S}(U_{\ell(n)}) \equiv U_{\ell(n)},$$

where the last equivalence holds because any bijective function applied to the uniform distribution yields the uniform distribution. Thus, $H(U_n) \overset{c}{\approx} U_{\ell(n)}$, as needed.

Question 3. Prove (or at least sketch a proof) that a *statistically* pseudorandom generator cannot exist. *Hint:* think about the sizes of $G(\{0,1\}^n)$ and $\{0,1\}^{\ell(n)}$, and define an appropriate statistical test (i.e., subset of $\{0,1\}^{\ell(n)}$).

Answer. Since G has n-bit seeds (of which there are 2^n), there are at most 2^n possible outputs of G (which reside in $\{0,1\}^{\ell(n)}$). However, $\{0,1\}^{\ell(n)}$ is significantly larger than this; it has $2^{\ell(n)} \geq 2^{n+1}$ elements, making it at least twice as big as the image of G. Informally, we can think of our statistical test as checking whether or not its given string is in the image of G. When the string is an output of G, the check always succeeds; when the string is uniformly random, there is a significant (at least $1/2$) probability that the check fails.

More formally and in the language of statistical indistinguishably: let $A = G(\{0,1\}^n) \subseteq \{0,1\}^{\ell(n)}$, i.e., the image of G, as suggested above. Because G is deterministic, we have $|A| \leq 2^n$. Then

$$\begin{aligned} |G(U_n)(A) - U_{\ell(n)}(A)| &= |1 - U_{\ell(n)}(A)| \\ &= 1 - |A| \cdot 2^{-\ell(n)} \\ &\geq 1 - 2^{n-\ell(n)} \\ &\geq 1 - 2^{-1} = 1/2, \end{aligned}$$

which is (very much) non-negligible. Since we have demonstrated a particular A for which $|G(U_n)(A) - U_{\ell(n)}(A)| \geq 1/2$, the suprenum over all possible A cannot be smaller, so $\Delta(G(U_n), U_{\ell(n)}) \geq 1/2$.