Lecture Notes
Basic concepts and statistical distance

S. Wehner

October 22, 2012

1 Basic probability and notation

Before embarking on our study of cryptography, let us establish some notation concerning probability distributions. I'm assuming that you are all very familiar with probabilities - if not, then please review this section carefully. Throughout, we will use capital letters X to denote random variables. The range, i.e. the set that the distribution is taken over will be denoted by cursive $\mathcal{X}$[1]. As usual, $|\mathcal{X}|$ denotes the number of elements in the set $\mathcal{X}$. We will also write $P_X(x)$ for the probability that outcome $x \in \mathcal{X}$ occurs, sometimes abreviated $p_x := P_X(x)$. When considering the probability of multiple outcomes from a set $\mathcal{S} \subseteq \mathcal{X}$ we also use the shorthand $P_X(\mathcal{S}) = \sum_{x \in \mathcal{S}} P_X(x)$. The probability distribution over $\mathcal{X} = \{1, \ldots, N\}$ will be written as a list, or vector,

$$\vec{\rho}_X := (p_1, p_2, \ldots, p_N) \ . \tag{1}$$

As an example, consider a fair die. It has possible outcomes $\mathcal{X} = \{1, 2, 3, 4, 5, 6\}$ each of which occurs with equal probability. That is $p_x = 1/6$ for all $x \in \mathcal{X}$, or in vector form

$$\vec{\rho}_X = (1/6, 1/6, 1/6, 1/6, 1/6, 1/6) \ . \tag{2}$$

A distribution in which all elements are equally likely is also known as the *uniform* distribution. If $\mathcal{S} = \{1, 2\}$, then $P_X(\mathcal{S}) = 2/6$.

1.1 Joint probabilities

Frequently, we will also consider *joint* probability distributions. That is, distributions over two sets $\mathcal{X} \times \mathcal{E}$. For convenience sake we will often abbreviate the joint distribution as $p_{xe} := P_{XE}(x, e)$ and write $\vec{\rho}_{XE}$ for the corresponding probability vector. Recall that a joint distribution is called a *product distribution* if P_{XE} is of the form $P_{XE}(x, e) = P_X(x) \cdot P_E(e)$ for all x and e. In terms of probability vectors this is expressed as $\vec{\rho}_{XE} = \vec{\rho}_X \otimes \vec{\rho}_E$. From the viewpoint of classical cryptography, this is slightly non standard notation, but it will make our later transition to quantum cryptography all the more easy.

As an example, consider our fair die from above. In addition, consider a distribution over colors red 'r' and blue 'b' - $\mathcal{E} = \{r, b\}$ Suppose that the distribution over colors is such that red occurs if

[1] All sets in this class are finite.

and only if (iff) the die took on values $1, 2$ or 3, and b, and blue occurs iff values $4, 5$ or 6 occured. The joint distribution then looks like

$$P_{XE}(x, r) = \begin{cases} 1/6 & \text{if } x \leq 3 \ , \\ 0 & otherwise \ . \end{cases} \tag{3}$$

$$P_{XE}(x, b) = \begin{cases} 1/6 & \text{if } x \geq 4 \ , \\ 0 & otherwise \ . \end{cases} \tag{4}$$

1.2 Conditional probabilities

Finally, we will refer to conditional probability distributions. That is, distributions over some set $\mathcal{X}$ *conditioned* on the fact that some other event occured. Given the joint distribution P_{XE} over $\mathcal{X} \times \mathcal{E}$ the distribution $P_{X|E}(x|e)$ is the distribution over $\mathcal{X}$ alone given that $E = e$. Note that by Bayes' rule we can write this distribution as

$$P_{X|E}(x|e) = \frac{P_{XE}(x, e)}{P_E(e)} \ . \tag{5}$$

An easy way to remember this is to consider how you might express P_{XE} itself if you only know P_E and $P_{X|E}$. Note that we have $P_{XE}(x, e) = P_{X|E}(x|e)P_E(e)$, which gives Bayes' rule by rearranging terms. An elementary but useful trick to remember is that also $P_{XE}(x, e) = P_{E|X}(y|x)P_X(x)$, allowing you to rexpress conditioning on E by conditioning on X. If you're feeling somewhat unsure about all of this please

- Convince yourself that the *marginal* distribution over $\mathcal{X}$ alone is given by

$$P_X(x) = \sum_{y \in \mathcal{E}} P_E(e) P_{X|E}(x|e) \ . \tag{6}$$

As an example of a conditional probability distribution, consider again our die and colors from above. For the joint distribution P_{XE} above, how can we write down the conditional probability distribution $P_{X|E}(x|e)$? Intuitively, we might guess that $P_{X|E}(x|r) = 1/3$ if $x \leq 3$, and $P_{X|E}(x|r) = 0$ if $x \geq 4$ - after all, the color red is only possible if $x \leq 3$ hence ruling out larger values of x. Second, our die was fair, so all values should be equally likely. Let's use the rules from above to convince ourselves that our guess is correct. Note that since we're rolling a fair die $P_X(x) = 1/6$. We can hence compute using Bayes' rule

$$P_{E|X}(r|x) = \frac{P_{XE}(x, r)}{P_X(x)} = \begin{cases} 1 & \text{if } x \leq 3 \ , \\ 0 & otherwise \ . \end{cases} \tag{7}$$

Thus by (6), $P_E(r) = \sum_x P_X(x) P_{E|X}(r|x) = 1/2$. Again applying Bayes' rule

$$P_{X|E}(x|r) = \frac{P_{XE}(x, r)}{P_E(r)} = \begin{cases} 1/3 & \text{if } x \leq 3 \ , \\ 0 & otherwise \ . \end{cases} \tag{8}$$

1.3 Expectation values

Another concept you should be familiar with is the *expectation value* $\mathbb{E}(X) = \sum_{x \in \mathcal{X}} p_x x$. As an example, consider again the fair die. It's easy to see that $\mathbb{E}(X) = 7/2$. By the law of large numbers the expected value is equal to the sample mean - i.e. the average outcome of the die roll - if we were to roll the die infinitely often. We will return to this fact later on.

1.4 Functions of random variables

Occasionally, we will not be content merely to make statements about some X, but rather we will apply a function $f : \mathcal{X} \to \mathcal{R}$ to X and then only consider the outcome $f(X)$. Note that this induces a new probability distribution - this time over $\mathcal{R}$. We will use $P_{f(X)}(f(x))$ and $\vec{\rho}_{f(X)}$ to remind ourselves how this new distribution arose.

As an example, consider again our far die. Imagine we roll the die and then apply the function

$$f(x) = \begin{cases} 0 & x \leq 4 \ , \\ 1 & x \geq 5 \ . \end{cases} \tag{9}$$

This yields a distribution $P_{f(X)}(0) = 4/6 = 2/3$ and $P_{f(X)}(1) = 1/3$. Similarly, we will denote the expectation value as $\mathbb{E}(f(X))$.

To make matters worse - we will sometimes even apply functions that are themselves chosen at random! I.e. the function F is itself a random variable. We also call this a randomized function. Imagine a set of two possible functions $\{f_1, f_2\}$ where $f_1 = f$ from above, but f_2 is defined as f but with 0 and 1 outcomes inverted. Imagine that $P_F(f_1) = P_F(f_2) = 1/2$. Can you write down $P_{F(X)}(1)$?

2 What does it mean to be ignorant?

We are now ready to turn to cryptography itself! Let us start by trying to formalize some of our basic intuitions. Quite often we say that the adversary (attacker) "does not know" or "does not learn" something during the course of the protocol. But what does this really mean?

2.1 Complete ignorance

To see what we might need to formalize this notion, let's return to our example of the die and the colors above. Imagine that I roll the die secretly behind my hand without letting you see anything about the outcome X. I claim that you "don't know anything" about X. Do you agree? How about if I tell you the corresponding color E?

Clearly, before I told you the color your best guess for my outcome was pretty hard - all values are equally likely, given by $P_X(x) = 1/6$. You know nothing. Yet, if I told you the color was red, you know that $x \leq 3$, given by $P_{X|E}(x|e)$ above. In other words, E does tell you something about X, i.e., you do know something about X. The first aspect in capturing the notion of ignorance formally is to demand that the two situations are the same. That is, $P_{X|E}(x|e) = P_X(x)$ for all x and y.

Is this already enough? Let's consider a different die in which outcome '6' occurs with probability $p_6 = 1$. Again, imagine that I secretly roll the die to obtain X. Would you say that you know nothing about X? Generally, when we say an attacker "does not know" we also mean that the distribution is uniform. Summarizing, we can thus define what it means to be ignorant.

Definition 2.1. *An adversary holding information E* does not know *X if and only if*

1. $P_X(x) = 1/|\mathcal{X}|$ *for all* $x \in \mathcal{X}$.
2. $P_{X|E}(x|e) = P_X(x)$ *for all* $x \in \mathcal{X}$ *and* $y \in \mathcal{E}$.

To fully understand this definition, convince yourself that the distribution P_{XE} given in the example above satisfies (1) but not (2). Instead of saying that the adversary "does not know", we will also use informal terms like "is completely ignorant" about X, or simple "is ignorant".

Since we will refer to this notion rather frequently, let us introduce a convenient shorthand. We will write $\mathrm{unif}(\mathcal{X})$ to denote the uniform distribution over $\mathcal{X}$ in vector notation. We can thus state (1) and (2) in the definition above more succintly as

$$\vec{\rho}_{XE} = \mathrm{unif}(\mathcal{X}) \otimes \vec{\rho}_E \ . \tag{10}$$

It should be noted that there are protocols in which we demand only (2), but desire a different distribution in (1). In this case, however, we will not talk about complete ignorance.

2.2 A notion of security

Having defined a meaningful notion of what it means to be ignorant, we can now define security of protocols more formally. To make a very simple, yet illustrative example, let's suppose that we want to define the security of a "Die tossing protocol". It takes nothing as input, and outputs a value $x \in \mathcal{X}$ chosen uniformly at random to the honest party. It's only "security requirement" is that any outside adversary cannot learn the outcome X.

Definition 2.2. *A Die Tossing scheme is a one-party protocol for Alice such that Alice obtains $X = x \in \mathcal{X}$ and for any adversary holding E at the end of the protocol*

$$\vec{\rho}_{XE} = \mathrm{unif}(\mathcal{X}) \otimes \vec{\rho}_E \ . \tag{11}$$

We then call any protocol that satisfies (11) a *secure* Die Tossing protocol. Why am I inventing such a complicated definition for a simple task? Note that we have effectively stated what the protocol should achieve by specifying properties of the output distributions of X and E. Almost all cryptographic tasks can be defined by what we want such output distributions to look like and we will return to this idea later on.

3 Statistical distance

In any real protocol, ensuring that the adversary remains completely ignorant about our secrets can almost never be achieved. Instead, we typically have to accept a small probability ε that our protocol fails in some disastrous manner and leaks all information to the adversary. Yet, if this probability is sufficiently small we may still want to call the protocol secure. Note, however, that since the adversary learns all information with probability ε, our definition of complete ignorance no longer applies. Even a die tossing protocol would not be considered secure according to our stringent definition. To deal with such practical issues, we would thus like to extend our notions of ignorance and security to the approximate case. I.e., we would like to have a notion of "almost completely ignorant".

To accomplish this, we need to introduce a measure of similarity betwen two probability distributions. Many such measures exist, but the one that is most relevant in cryptography is given by the so-called *statistical distance*. Formally, it is defined as

Definition 3.1. *Let $\vec{\rho}_X$ and $\vec{\rho}_Y$ denote probability distributions over a finite set $\mathcal{R}$ respectively. The* statistical distance *(also known as the* L1-distance*) between the two distributions is defined as*

$$\Delta(\vec{\rho}_X, \vec{\rho}_Y) := \frac{1}{2} \sum_{r \in \mathcal{R}} |P_X(r) - P_Y(r)| \ . \tag{12}$$

We will also say that two distributions $\vec{\rho}_X$ and $\vec{\rho}_Y$ are ε-close iff

$$\Delta(\vec{\rho}_Z, \vec{\rho}_Y) \leq \varepsilon. \tag{13}$$

Equivalently, we also write $\vec{\rho}_Z \approx_\varepsilon \vec{\rho}_Y$.

At first glance it may seem a restriction to demand that both X and Y are distributed over the same set $\mathcal{R}$. Note, however, that we may always take $\mathcal{R} = \mathcal{X} \cup \mathcal{Y}$ and assign '0' probability on values outside the original set.

The statistical distance satisfies many appealing properties which justify its use in cryptography. We will prove some of them here - the rest is left as a homework exercise for you!

Theorem 3.2. *The statistical distance satisfies the following properties for all $\vec{\rho}_X$ and $\vec{\rho}_Y$*

1. *$\Delta(\cdot, \cdot)$ is a metric.*
2. *$\Delta(\cdot, \cdot) \leq 1$.*
3. *$\Delta(\vec{\rho}_X, \vec{\rho}_Y) = \max_{\mathcal{S} \subseteq \mathcal{R}} P_X(\mathcal{S}) - P_Y(\mathcal{S})$ where the maximization is taken over all subsets $\mathcal{S} \subseteq \mathcal{R}$.*
4. *$\Delta(\vec{\rho}_X, \vec{\rho}_Y) = \max_f |\mathbb{E}(f(X)) - \mathbb{E}(f(Y))|$, where the maximization is taken over all functions $f : \mathcal{R} \to [0, 1]$.*
5. *For any subset $\mathcal{S} \subseteq \mathcal{R}$, $P_X(\mathcal{S}) \leq P_Y(\mathcal{S}) + \Delta(\vec{\rho}_X, \vec{\rho}_Y)$.*
6. *For any randomized function F, $\Delta(\vec{\rho}_{F(X)}, \vec{\rho}_{F(Y)}) \leq \Delta(\vec{\rho}_X, \vec{\rho}_Y)$.*
7. *For any product distributions $\vec{\rho}_{X_1 X_2} = \vec{\rho}_{X_1} \otimes \vec{\rho}_{X_2}$ and $\vec{\rho}_{Y_1 Y_2} = \vec{\rho}_{Y_1} \otimes \vec{\rho}_{Y_2}$,*

$$\Delta(\vec{\rho}_{X_1 X_2}, \vec{\rho}_{Y_1 Y_2}) \leq \Delta(\vec{\rho}_{X_1}, \vec{\rho}_{X_2}) + \Delta(\vec{\rho}_{Y_1}, \vec{\rho}_{Y_2}) \ . \tag{14}$$

Looking at the first property, you may be wondering - what is a metric? In mathematics, the term metric means that Δ does indeed correspond to our intuitive notion of distance. This means that it satisfies the following conditions for all $\vec{\rho}_X$ and $\vec{\rho}_Y$

- $\Delta(\vec{\rho}_X, \vec{\rho}_Y) \geq 0$. Intuitvely, this means that distances can only be positive. It's easy to see that this is indeed satisfied since all the terms appearing in the sum are positive.
- $\Delta(\vec{\rho}_X, \vec{\rho}_Y) = 0$ iff $\vec{\rho}_X = \vec{\rho}_Y$. Intuitvely, this means that the distance can only be 0, if we are indeed at the same spot. Again, it's easy to convince yourself that this is satisfied since all terms appearing in the sum are positive, and each one individually can only vanish (be 0) iff $P_X(r) = P_Y(r)$.
- $\Delta(\vec{\rho}_X, \vec{\rho}_Y) = \Delta(\vec{\rho}_Y, \vec{\rho}_X)$. Intuively, this means that distances are the same both ways. Again, it's easy to see that this is the case due to the absolute value in each term.

- $\Delta(\vec{\rho}_X, \vec{\rho}_Y) \leq \Delta(\vec{\rho}_X, \vec{\rho}_Z) + \Delta(\vec{\rho}_Z, \vec{\rho}_Y)$. Intuitvely, this means that the distance from $\vec{\rho}_X$ to $\vec{\rho}_Y$ is never greater than the distance you would have to walk from $\vec{\rho}_X$ to $\vec{\rho}_Z$ and then onwards from $\vec{\rho}_Z$ to $\vec{\rho}_Y$ (after all - that's a way to get to $\vec{\rho}_Y$!) This inequality is also known as the *triangle inequality* and is extremely useful. It's only slighly more difficult to prove:

$$\begin{aligned}
\Delta(\vec{\rho}_X, \vec{\rho}_Y) &= \frac{1}{2} \sum_{r \in \mathcal{R}} |P_X(r) - P_Y(r)| && (15)\\
&= \frac{1}{2} \sum_{r \in \mathcal{R}} |P_X(r) - P_Z(r) + P_Z(r) - P_Y(r)| && (16)\\
&\leq \frac{1}{2} \sum_{r \in \mathcal{R}} |P_X(r) - P_Z(r)| + \frac{1}{2} \sum_{r \in \mathcal{R}} |P_Z(r) - P_Y(r)| && (17)\\
&= \Delta(\vec{\rho}_X, \vec{\rho}_Z) + \Delta(\vec{\rho}_Z, \vec{\rho}_Y) && (18)
\end{aligned}$$

I will leave it as an optional exercise for you to prove 2. Properties 3, 4 and 6 are extremely useful - you will prove them in your homework and explore their cryptographic significance. To give you a little flavor of proofs, let us convince ourselves that 5 holds. Note that to prove 5, we are allowed to make use of the earlier properties - just as you may do in your homework. Looking at the list above, 3 seems particularly useful to us in order to prove 5. Indeed, we have for any set $\mathcal{S}$

$$P_X(\mathcal{S}) - P_Y(\mathcal{S}) \leq \Delta(\vec{\rho}_Z, \vec{\rho}_Y) \ , \tag{19}$$

by property 3, and 5 is a simple rewriting of this equation.

Even though 5 looks rather trivial, it has one very important consequence. Let's suppose that we were to run cryptographic protocol or an algorithm which takes as inputs elements Y drawn according to a distribution P_Y. Furthermore, suppose for simplicity that there is some set $\mathcal{S}$ of inputs for which our protocol/algorithm is sure to fail - otherwise it always succeeds. What is the probability that protocols fails? Clearly, it is given by $P_Y(\mathcal{S})$. Imagine now that already P_Y is flawed in the sense that we are actually selecting elements according to the distribution P_X with

$$\vec{\rho}_X \approx_\varepsilon \vec{\rho}_Y \ . \tag{20}$$

What is the probability that we will fail this time? Note that property 5 tells us that the probability of failure obeys $P_X(\mathcal{S}) \leq P_Y(\mathcal{S}) + \varepsilon$. That is, for small ε the difference is indeed very small! The statistical distance is thus a very nice measure in this case, since it does indeed have an operational meaning: if the two distributions are ε-close it matters very little which one we use.

In the homework, you will explore this idea in more detail and investigate why the statistical distance is so useful in cryptography. Yet, it should already be clear how we're going to use it: Instead of demanding complete ignorance as above, we are now able to capture "almost complete ignorance" by demanding that

$$\vec{\rho}_{XE} \approx_\varepsilon \mathrm{unif}(\mathcal{X}) \otimes \vec{\rho}_E \ . \tag{21}$$

Similarly, we say that a protocol is ε-secure iff the output distribution it generates is ε-close to the ideal one. For example, a protocol is an ε-secure die tossing protocol iff $\vec{\rho}_{XE} \approx_\varepsilon \mathrm{unif}(\mathcal{X}) \otimes \vec{\rho}_E$.