Discrete probability

Most of the serious cryptographic protocols we consider will involve some degree of randomisation, particularly in the choice of keys. So it is necessary to have a basic knowledge of discrete probability in order to understand their analysis. Fortunately, the knowledge required is pretty basic indeed, and these notes (together with the video references and links their) should suffice.

Probability distributions and events

A *probability distribution* on a finite set U is a function $p : U \to \mathbb{R}$ with the following properties:

$$p(u) \geq 0 \text{ for all } u \in U,$$
$$\sum_{u \in U} p(u) = 1.$$

In the second condition, Σ stands for *sum* and the subscript denotes the set of things we are taking the sum over – that is we are taking the sum of all the values $p(u)$ as u runs over U (think of a for loop or an iterator!) Note also that the conditions imply that $0 \leq p(u) \leq 1$ for all $u \in U$.

A particularly important probability distribution is the *uniform* distribution where $p(u)$ is the same for all $u \in U$. In that case $p(u) = 1/|U|$, where $|U|$ represents the *cardinality* or number of elements in U (generically, absolute value bars around something denote its size – which is supposed to be understood or have a unique sensible interpretation).

Given a probability distribution on U, an *event* is just a subset $A \subseteq U$ and we extend the function p from individual elements to events by setting $p(A) = \sum_{a \in A} p(a)$. Then $0 \leq p(A) \leq 1$ for any event A. This terminology comes from thinking of U as "things that might happen". In that context we are frequently interested in avoiding something bad happening, so there might be one of more bad events. Ideally we'd like to be able to show that the bad events have low probability and to this end the *union bound* can be useful:

$$p(A_1 \cup A_2) \leq p(A_1) + p(A_2).$$

The event $A_1 \cup A_2$ is just A_1 or A_2 (or both). The union bound follows by noting that the right hand side certainly contains all the terms that are required for the sum represented by the left hand side, and moreover contains those corresponding to elements of $A_1 \cap A_2$ (the intersection of A_1 with A_2) twice. In fact:

$$p(A_1 \cup A_2) = p(A_1) + p(A_2) - p(A_1 \cap A_2).$$

However, the moral meaning of the union bound is that if there are a limited number of bad things that might happen, and each of them individually is of low probability then they are collectively of low probability (though not quite as low as the individual ones).

Random variables

In everyday discussions about probability we don't really tend to think about distributions but rather about events whose outcome is uncertain. So when we say "The chance of rolling a 6 on a fair die is $1/6$" we are thinking about a *random variable*, R (the result of rolling a fair die), and the probability that it takes a particular value. Similarly if we say "Choose a byte uniformly at random" we are again thinking about an underlying probability distribution on bytes (the uniform one where each has probability $1/256$) and then an experiment where we generate a byte from this distribution.

We will use capital letters to denote random variables and corresponding lower case letters (generally) to denote their observed (or potentially observed) values. We can also describe events in terms of random variables. For instance if B is the "choose a byte uniformly at random" random variable then we can define the event "the low order bit of B is equal to 1". Call this event L. What is the probability of L? It is the sum over all bytes whose low order bit is equal to 1 of the probability of that byte being chosen. Symbolically

$$p(L) = \sum_{b=c1} p(B = b) = \sum_{b=c1} 1/256 = 128/256 = 1/2.$$

Here c stands for the first seven bits of b. Since c is arbitrary, there are 128 such c and that gives the (obvious in advance) answer. In the uniform case, working out probabilities is always just a matter of counting how many elements make up the event in question.

Independence

Consider two random variables: T the result of a fair coin toss, and R the result of rolling a fair die. If we consider the two results together we can ask questions like "what is the probability that the result of the coin toss is heads, and the result of the die roll is a 3?" The two random variables are *independent* if this probability is what we naively expect it to be - namely the product of the individual probabilities. Formally, two random variables X and Y are independent if for all x and y

$$p(X = x \text{ and } Y = y) = p(X = x) \times p(Y = y).$$

Mostly it's obvious when things are independent or not (by construction).

A special context

Almost all of the arguments we will be needing to use about probability deal with probability on the universe of "strings of n bits". We will denote this space by $\mathbf{2}^n$. Furthermore, we will be using an operation on this space called XOR (for exclusive or) and denoted $\oplus$. On single bits:

$$0 \oplus 0 = 1 \oplus 1 = 0 \quad 0 \oplus 1 = 1 \oplus 0 = 1.$$

On strings of bits we just form the XOR by taking the XOR of the corresponding individual bits. XOR has some important properties that are easily verified:

$$x \oplus x = 00 \cdots 0$$
$$x \oplus y = y \oplus x$$
$$x \oplus (y \oplus z) = (x \oplus y) \oplus z.$$

In turn these imply that all three of the following are equivalent:

$$a \oplus b = c \quad b \oplus c = a \quad c \oplus a = b.$$

The key construction that we will use again and again is the following. Let Z be any random variable on $\mathbf{2}^n$ (its distribution is arbitrary and we may not even know it). Let X be a uniform random variable on $\mathbf{2}^n$ which is independent of Z (we generate bit strings uniformly at random from a source under our control and independent of Z). Then the distribution of $X \oplus Z$ is uniform.

In other words by taking the XOR of an arbitrarily distributed bit string with one which is independent and uniformly distributed we hide all the information that might have been present in the original string (since every possible string is now equally likely). That's worth a proof.

Let w be any element of $\mathbf{2}^n$. Then, using independence:

$$p(X \oplus Z = w) = \sum_{z \in \mathbf{2}^n} p(Z = z \text{ and } w = X \oplus z) = \sum_{z \in \mathbf{2}^n} p(Z = z) p(X = w \oplus z).$$

But $p(X = w \oplus z) = 1/2^n$ since X is uniform. So

$$p(X \oplus Z = w) = (1/2^n) \sum_{z \in \mathbf{2}^n} p(Z = z) = 1/2^n$$

since the sum is now 1 (Z has to take on some value!)