University of Maryland
CMSC 456 — Introduction to Cryptography
Professor Jonathan Katz

Notes on Discrete Probability[1]

1 Notation

The notation introduced here is standard, and will be used throughout the course. Most of it should be familiar from Discrete Math.

For a set S, we let $|S|$ represent the size of S; i.e., the number of elements contained in S. The notation $s \in S$ denotes that element s is a member of set S. When S has no members, we say that S is the *empty set*, and denote it by $\emptyset$. For sets S, T, the notation $S \subset T$ means that every member of S is also a member of T; formally, for all $a \in S$ it is the case that $a \in T$. For any set T, it is the case that $\emptyset \subset T$. $S \cup T$ (the *union* of S and T) denotes the set of elements which are in *either* S or T. $S \cap T$ (the *intersection* of S and T) denotes the set of elements which are in *both* S and T.

We define $S \times T$ as the *cross product* of two sets S and T, which consists of ordered pairs of elements in which the first element is from S and the second element is from T. Formally, $S \times T = \{(a, b) : a \in S, b \in T\}$ (where the colon ":" should be read as "such that"). This can be extended to any (finite) number of sets $S_1 \times \cdots \times S_n$. We abbreviate $S \times S$ by S^2, and so on for any finite number of copies of S. S^0 is defined as the empty set. As an example, consider the set $\{0, 1\}$. Then $\{0, 1\}^n$ is shorthand for n copies of $\{0, 1\}$, and thus $\{0, 1\}^n$ can be viewed as the set of all bit strings of length exactly n.

The notation $\binom{a}{b}$ denotes the number of ways of choosing b elements from a set of s elements, where order is unimportant. An elementary result shows that $\binom{a}{b} = \frac{a!}{b!(a-b)!}$.

A *bit* b is a value in $\{0, 1\}$. The XOR operation $\oplus$ is defined by: $0 \oplus 0 = 0; 0 \oplus 1 = 1 \oplus 0 = 1$; and $1 \oplus 1 = 0$. This naturally extends to the XOR of any two bit strings of equal length.

The phrase iff means "if and only if".

2 Definitions

In cryptography, we typically want to prove that an adversary that tries to break some protocol has only very small probability of success. In order to prove such results, we need to be able to compute the probability that certain events happen. In order to model a probabilistic system, we define a *sample space* and a *probability distribution*. The sample space is the set of all possible *elementary events*; i.e., things that can occur. A probability distribution is a function which assigns a real number in $[0, 1]$ to each elementary event, this number being the probability that the event occurs. The probabilities of all events must sum to 1. Formally:

[1] Adapted in part from Luca Trevisan's "Notes on Discrete Probability".

Definition 1 For a finite sample space Ω and a function $\mathbf{Pr} : \Omega \to [0, 1]$, we say that $\mathbf{Pr}$ is a probability distribution if: $\sum_{a\in\Omega} \mathbf{Pr}(a) = 1$. ■

A *uniform distribution* is one in which $\Pr(a) = \Pr(b)$ for all $a, b \in \Omega$. In other words, each elementary event is equally likely to occur. It is easy to show that this implies $\Pr(a) = 1/|\Omega|$, for all $a \in \Omega$.

We note that the definitions are more complicated in the case of continuous probability, where Ω may be infinite. However, since cryptography (and, in general, computer science) deals with "real world" events (in which the set of possibilities may be large, but is always finite), we only need to be concerned with discrete probability distributions.

As our running example, we will consider the case of a sequence of three random bits (which may be viewed as tosses of a coin, with $1 =$ *heads* and $0 =$ *tails*). Thus, our sample space is $\{0, 1\}^3$, and the probability distribution assigns equal probability 1/8 to each elementary event of this space. This is an example of a uniform distribution.

Define an *event* as an arbitrary subset $A \subset \Omega$. Note the distinction between *elementary events* and *events*. The former are things which actually occur, while the latter are sets of things which occur (which may all have some desired feature). The probability of an event is defined as one might expect:

$$\Pr[A] = \sum_{a\in A} \Pr(a).$$

We use square brackets to remind us that we are now considering a different function: $\mathbf{Pr}(\cdot)$ is a function whose inputs are *elements* of the sample space, while $\mathbf{Pr}[\cdot]$ is a function whose inputs are *subsets* of the sample space (of course, a one-element subset may be viewed as an element, but formally these are different objects). Note that $\Pr[\emptyset] = 0$ and $\Pr[\Omega] = 1$.

For any event E (in an implicit sample space Ω), we define $\overline{E}$ as the set $\Omega - E$; that is, all elements not in E.

As an example, consider the event $E =$ "there are exactly two 1's". Clearly, we have: $E = \{110, 101, 011\}$. Thus, the probability of event E is 3/8. Very often, as in this case, computing the probability of an event reduces to counting the size of a set.

Given two samples spaces Ω_1 and Ω_2, with probability distributions $\Pr_1$ and $\Pr_2$, we may define the sample space $\Omega = \Omega_1 \times \Omega_2$ with probability distribution $\Pr$ defined by:

$$\Pr((a, b)) \stackrel{\text{def}}{=} \Pr_1(a) \cdot \Pr_2(b).$$

Note that Ω may be viewed as an event from Ω_1 followed by an event from Ω_2. Then the above simply states that the probability that *both* a and b occur is given by the product of their individual probabilities. The above definition extends to events as well, so that $\Pr[(E_1, E_2)] = \Pr_1[E_1] \cdot \Pr_2[E_2]$.

As an example, note that we may view $\{0, 1\}^3$ as $\{0, 1\} \times \{0, 1\}^2$. Thus, $\Pr(001) = \Pr_1(0) \cdot \Pr_2(01) = 1/2 \cdot 1/4 = 1/8$, as expected.

2.1 Conditioning and Mutual Independence

There is another way to view the previous example. Consider sample space Ω with probability distribution $\Pr$. E_1, E_2 be events in Ω. We define the *conditional probability* of E_2

given E_1, denoted $\Pr[E_2|E_1]$, as follows:

$$\Pr[E_2|E_1] \stackrel{\text{def}}{=} \frac{\Pr[E_2 \bigcap E_1]}{\Pr[E_1]}.$$

(This assumes $\Pr[E_1] \neq 0$.) Note that if $E_2 \cap E_1 = \emptyset$, then $\Pr[E_2|E_1] = 0$. In other words, if there is no way for events E_1 and E_2 to occur simultaneously, then the probability of E_2, given the fact that E_1 has already occurred, is 0.

We say that events E_1 and E_2 are *independent* if $\Pr[E_1|E_2] = \Pr[E_1]$. Note that when two events are independent then $\Pr[E_1 \cap E_2] = \Pr[E_1] \cdot \Pr[E_2]$. When we have several events, we can define a generalized notion of independence:

Definition 2 Let $E_1, \ldots, E_n$ be events. We say that they are mutually independent if, for every subset of indices $I \subseteq \{1, \ldots, n\}, I \neq \emptyset$, we have:

$$\Pr[\bigcap_{i \in I} E_i] = \Pi_{i \in I} \Pr[E_i].$$

■

As an example, let $E_2 = \{001\}$ and let $E_1 =$ "the first bit of the sequence is 0". Note that $E_2 \bigcap E_1 = E_2$. Thus:

$$\Pr[E_2|E_1] = \frac{\Pr[E_2 \bigcap E_1]}{\Pr[E_1]} = \frac{1/8}{1/2} = 1/4.$$

Another way of looking at the above result is to ask: given that the first bit of the sequence is 0, what is the probability that the entire sequence is 001? But this is equivalent to asking: given that the first bit of the sequence is 0, what is the probability that the next two bits are 01? And since we assume that each bit is independent of all the others, this probability is simply 1/4. This agrees with the above calculation.

This leads to a very useful technique for computing probabilities. Let $E = E_1 \cup \cdots \cup E_n$, where the E_i are pairwise disjoint (i.e., for all i, j, $E_i \cap E_j = \emptyset$). Note that, by simply considering the individual elements of E, we have:

$$\Pr[E] = \sum_i \Pr[E_i]. \tag{1}$$

Now, let $B_1, \ldots, B_n$ be events which are pairwise disjoint, and such that $\Omega = B_1 \cup \cdots \cup B_n$. It is then the case that the events $E \cap B_1, \ldots, E \cap B_n$ are pairwise disjoint and $E = (E \cap B_1) \cup \cdots \cup (E \cap B_n)$. Therefore:

$$\begin{aligned} \Pr[E] &= \sum_i \Pr[E \cap B_i] \\ &= \sum_i \Pr[E|B_i] \Pr[B_i]. \end{aligned}$$

This can be very useful in practice. As an "easy" example, Let $E =$ "the sequence is 001" and $B =$ "the first bit of the sequence is 0". Note that the pair $B, \overline{B}$ satisfy the above criteria. Thus:

$$\Pr[E] = \Pr[E|B]\Pr[B] + \Pr[E|\overline{B}]\Pr[\overline{B}] = 1/4 \cdot 1/2 + 0 \cdot 1/2 = 1/8.$$

Finally, we give a generalization of equation (1). For two *arbitrary* sets E_1, E_2 we have:

$$\Pr[E_1 \cup E_2] = \Pr[E_1] + \Pr[E_2] - \Pr[E_1 \cap E_2].$$

(Note that in equation (1) we have $E_1 \cap E_2 = \emptyset$ and therefore the final term is 0.) The proof, as before, follows by simply considering the individual elements of E_1 and E_2. Can you find the generalization to n arbitrary sets?

We note that weaker definitions of independence are also possible. For example, one may define *pairwise independence* as follows:

Definition 3 Let $E_1, \ldots, E_n$ be events. We say they are pairwise independent it, for all $i, j \in \{1, \ldots, n\}, i \neq j$, we have that E_i and E_j are independent. ■

It is important to note that pairwise independence does not imply mutual independence (although the converse is obviously true). For example, flip two bits and let E_1 be the event that the first bit was a 1, E_2 be the event that the second bit was a 1, and E_3 be the event that the XOR of the bits is 1. Clearly, these events are not mutually independent, since knowledge of any two of these events automatically implies knowledge about the remaining one. However, one can check that each of these events are pairwise independent (can you prove it?).

3 Random Variables and Expectation

Very often, we are interested in studying some value that depends on the elementary events that take place. For example, when we play dice, the elementary event "roll a 6" corresponds to the value 6, and we may be interested in the average value we get from this game. In a cryptographic setting, we often study a randomized algorithm which terminates after an arbitrary number of steps; we want to find out, on average, how many steps it takes. The notion of a random variable gives us a tool to formalize questions of this nature.

A *random variable* X is (formally) a function $X : \Omega \to R$, where R is the *range* of X, and is typically the real numbers or the integers. One should think of a random variable as a function that on input an elementary event returns as output some value associated with that event. Let Ω be a sample space, **Pr** a probability distribution, and X a random variable. The expression $X = v$ then denotes an event, namely the event $\{a : X(a) = v\}$; therefore, the expression $\Pr[X = v]$ is well-defined. Of course, more complicated examples are also possible.

We can define the *expectation* of a random variable X, denoted $E[X]$, as a "weighted average" of the values the function takes. Formally:

$$E[X] = \sum_{a \in \Omega} \Pr(a) \cdot X(a).$$

When the range R is finite (for example, a subset of the integers), we may also write:

$$E[X] = \sum_{r \in R} r \cdot \Pr[X = r].$$

This is the form most frequently used.

As an example, consider a single throw of a fair die. Then $\Omega = \{1, 2, 3, 4, 5, 6\}$ and **Pr** is the uniform distribution. Let X be the value of the number on the die. Then:

$$E[X] = \sum_{1 \le i \le 6} i \cdot \Pr[X = i] = \sum_{1 \le i \le 6} i/6 = 3.5.$$

One very useful property of expectation is that it is *linear*:

Theorem 1 *Let X be a random variable and c a real number. Then $E[cX] = c \cdot E[X]$. Furthermore, let $X_1, \ldots, X_n$ be random variables over the same sample space. Then $E[X_1 + \cdots + X_n] = E[X_1] + \cdots + E[X_n]$.*

Note that no assumption of independence is needed.

As an example of how useful this can be, consider the problem of flipping n fair coins, and let X be the random variable counting the number of heads. We want to find the expected value of X. First, let's work with the definition of expectation. Clearly, the number of heads is between 0 and n, inclusive. We can compute $\Pr[X = i]$ as follows: the probability of any given elementary event is simply 2^{-n} (since we assume a uniform distribution; i.e., a fair coin). The number of ways of getting exactly i heads in n tosses is $\binom{n}{i}$. Therefore, $\Pr[X = i] = \binom{n}{i}2^{-n}$. Thus, we can write:

$$E[X] = \sum_{0 \le i \le n} i \cdot \Pr[X = i] = \sum_{0 \le i \le n} i \cdot \binom{n}{i} 2^{-n}.$$

Now what? Note that getting an explicit solution for the value of the sum is difficult.

Instead, define random variable X_i for $1 \le i \le n$ as 1 iff the ith coin came up heads. Notice that $X = X_1 + \cdots + X_n$. Note also that $E[X_i] = 1/2$. Thus, $E[X] = E[X_1] + \cdots + E[X_n] = n/2$. We just saved ourselves a lot of trouble!

Finally, we can also define a notion of independence for random variables.

Definition 4 We say that random variables X and Y (over the same sample space) are independent if, for any two values x and y, the events $X = x$ and $Y = y$ are independent. ■

4 The Birthday Problem

This section[2] is motivated by the following, well-known party[3] question: given q people chosen at random, what is the probability that, for some pair of people, their birthdays are equal? Put a different way: how large should N be so that there is a 50% chance that two people will have identical birthdays?

To study this question, we consider an abstract game in which we have q balls and N boxes, where $N \ge q$. We throw the balls, one by one, at random into the boxes. In other words, for each ball, the probability that it lands in any particular box is $1/N$, and this probability is independent of the outcome for any other balls. We say event Collision occurs

[2]Adapted in part from the Appendix to the CSE 107 lecture notes by Mihir Bellare.

[3]Of course, this depends on the kind of parties you go to.

if there exists some box which ends up containing more than one ball. Note that this is a restatement of the "birthday" game above, in which $N = 365$ (the number of possible birthdays).

We want to compute $\Pr_{q,N}[\mathsf{Collision}]$. The following theorem bounds this probability:

Theorem 2 *Let $q \leq N$ and $q \geq 1$. Then:*

$$\Pr_{q,N}[\mathsf{Collision}] \leq \frac{q(q-1)}{2N},$$

and furthermore:

$$\Pr_{q,N}[\mathsf{Collision}] \geq 1 - e^{q(q-1)/2N},$$

and for $1 \leq q \leq \sqrt{2N}$:

$$\Pr_{q,N}[\mathsf{Collision}] \geq \frac{0.3q(q-1)}{N}.$$

In answer to our original question (the "birthday" game), note that the probability of a collision is about 50% when $q = O(\sqrt{N}) = 24$. This is lower than most people expect (you can win a lot of bets this way...)!

5 Useful Facts

One fact which comes up time and again is the following, which holds for all real x:

$$1 + x \leq e^x. \tag{2}$$

As an example of its usefulness, consider a randomized algorithm which has probability ϵ of success. How many times do we have to run the algorithm (using new random coins each time) before we have a probability of succeeding of at least 3/4? The probability that the algorithm never succeeds in k different (independent) runs is $(1-\epsilon)^k$ (why?). But this is less than $e^{-\epsilon k}$, by (2). Choosing $k = 2/\epsilon$, the probability of k consecutive failures is less than $e^{-2} < 1/4$, and therefore the probability of succeeding (at least once) is at least 3/4.