Topics of Mathematics in Cryptology

Probability Theory

Dominique Unruh

February 16, 2017

1 Events and probabilities

A *probability* is an indicator how likely something is to happen. For example, if we roll a die, we may ask what the probability is that we roll a six. In other words, there is a certain *event* $A :=$ "we roll a six", and we ask for the probability $\Pr[A]$ that this event occurs. In our example, for a fair die, $\Pr[A] = \frac{1}{6}$ because there are six equally likely numbers we can roll.

probability

event

So, an event is a description of something that may or may not happen, and the probability of that event tells us how often it happens. (E.g., if $\Pr[A] = 1/n$, then A happens one out of n times.)

From this intuitive description, we can easily see the following basic laws of probability:

Lemma 1 (Elementary properties of probabilities)

(a) $0 \leq \Pr[A] \leq 1$.

(No event happens less often than never, no event happens more often than always.)

(b) If A is always true, then $\Pr[A] = 1$.

(E.g., $\Pr[$*1 is an integer*$] = 1$ *or* $\Pr[$*we roll a number greater than* 0 *(with a six-sided die)*$] = 1$*.)*

(c) If A is never true, then $\Pr[A] = 0$.

(E.g., $\Pr[\frac{3}{2}$ *is an integer*$] = 0$ *and* $\Pr[$*we roll a seven*$] = 0$*.)*

Some examples of events:

- $A :=$ "We roll an even number". ($\Pr[A] = \frac{1}{2}$)
- $A :=$ "The number of people born on December 31, 2025 is greater than 20 millions". (Here, and in the next two, we do not know $\Pr[A]$.)
- $A :=$ "Tomorrow it rains". (Assuming a precise definition of raining.)
- $A :=$ "You will pass this course". (Hopefully, $\Pr[A]$ is close to 1.)

- Assume some particular cryptographic system and some attacker that attacks the protocol. $A :=$ "The attacker logs into the system without having been authorized". (Given a precise description of the system and the attacker, $\Pr[A]$ could, at least in principle, be computed.)

1.1 Mutually exclusive events

mutually exclusive

An important special case is that of *mutually exclusive* events. We call two events A, B mutually exclusive if $A \wedge B$ is impossible. For example, if $A :=$ "we roll a five" and $A :=$ "we roll a six", then A and B are mutually exclusive because it is impossible to roll five and six simultaneously.

Other examples of mutually exclusive events:

- $A :=$ "tomorrow it rains", $B :=$ "tomorrow it does not rain".
- Assume a cryptographic system where one can only login once. $A :=$ "the adversary logs in as Alice", $B :=$ "the adversary logs in as Bob".
- $A :=$ "$x > 10$", $B :=$ "$x < 5$".

The importance of the concept of mutually exclusive events stems from the following fact:

Lemma 2 (Mutually exclusive events) *If A and B are mutually exclusive, then* $\Pr[A \vee B] = \Pr[A] + \Pr[B]$.

In many cases, this lemma allows us to compute $\Pr[A \vee B]$ more easily. (Or sometimes, we know $\Pr[A \vee B]$ and $\Pr[A]$ and can then compute $\Pr[B]$.)

We illustrate the usefulness of this lemma by a few examples:

Example: Obviously, for any event A, the events A and $\neg A$ are mutually exclusive. Hence $1 \stackrel{(*)}{=} \Pr[A \vee \neg A] \stackrel{(**)}{=} \Pr[A] + \Pr[\neg A]$ where $(*)$ uses Lemma 1 (b) and $(**)$ uses Lemma 2. From this we immediately derive:

Lemma 3 *For any event A, we have* $\Pr[A] = 1 - \Pr[\neg A]$.

Example: Assume the following game of chance: Two random numbers x, y are picked from $1, \ldots, 100$. We win if $x \neq y$ and $x, y > 90$. What is the probability of winning?

It is easy to see that $\Pr[x > 90, y > 90] = 0.1 \cdot 0.1 = 0.01$. (Formally, this can be seen using the tools from the next subsection.) Similarly, $\Pr[x = y, x, y > 90] = 0.1/100 = 0.001$ (the probability that $x > 90$ is 0.1, and with probability $1/100$, y has the same value). Notice that $(x \neq y \wedge x, y > 90)$ and $(x = y \wedge x, y > 90)$ are mutually exclusive events. Thus

$$\begin{aligned} 0.01 = \Pr[x, y > 90] &= \Pr[x \neq y \wedge x, y > 90] + \Pr[x = y \wedge x, y > 90] \\ &= \Pr[x \neq y \wedge x, y > 90] + 0.001 \end{aligned}$$

and hence $\Pr[win] = \Pr[x \neq y \wedge x, y > 90] = 0.01 - 0.001 = 0.009$.

1.2 Independent events

independent events

Another very important concept is that of *independent events* (a.k.a. *stochastically independent* events). Intuitively, two events A and B are independent if knowing whether A occurs does not tell us anything about whether B occurs. This happens, e.g., when A and B are the result of separate random processes. For example, when Alice and Bob each roll a fair die, then $A :=$ "Alice rolls a one" and $B :=$ "Bob rolls a six" are independent.

Our intuition tells us that if A happens in one out of n cases, and B happens in one out of m cases, and they are independent, then $A \wedge B$ happens in one out of $n \cdot m$ cases. This leads to the following mathematical definition of independence:

Definition 1 *Two events A and B are* independent *iff* $\Pr[A \wedge B] = \Pr[A] \cdot \Pr[B]$.

In the example where Alice and Bob roll a fair die each, we thus have

$$\Pr[\text{Alice rolls 1, Bob rolls 6}] = \Pr[A \wedge B] = \Pr[A] \cdot \Pr[B] = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}.$$

Some examples for independent events are:

- When Alice picks a number x in some way. (We do not care how.) Bob then picks and announces a random number y. $A :=$ "$x < 5$" and $B :=$ "$y < 5$".
- Let x, y be random bits, each chosen using a fair coin flip. $A :=$ "$x = 0$" and $B :=$ "x=y". A and B are independent because $\Pr[A \wedge B] = \Pr[x = 0 \wedge x = y] = \Pr[x = 0 \wedge y = 0] = \frac{1}{4} = \Pr[A] \cdot \Pr[B]$.

Some examples of events that are not independent:

- Bob picks and announces a random number y. Then Alice picks a number x in some way. (We do not care how.) $A :=$ "x<5" and $B :=$ "y<5". Here A, B are not independent (at least for some strategy of Alice) because Alice may choose x depending on y. (E.g., $x := y$.)
- Let x be the outcome of a die roll. $A :=$ "$x < 3$", $B :=$ "$x < 5$".

The following elementary laws hold for independent events:

Lemma 4 (Independent events)

- *If A and B are independent, then also A and $\neg B$ are independent.*
- *If A always holds (or never holds), then A and B are independent.*

1.3 Exercises

Exercise 1 *Formulate three events involving a lottery.*

Exercise 2 *Assign probabilities to the following events:*
(A) Alice's die roll is odd.
(B) The sum of Alice's two die rolls is at least 11*.*
(C) $x = y$, where x and y are die rolls.

Exercise 3 *Which of the following events are mutually exclusive, which are not, and for which is it unknown? (Here x, y are the results of two fair die rolls.)*
A. Tomorrow's temperature is below zero.
B. $x > 4$.
C. $y > 6$.
D. $x < 4$.
E. Tomorrow, the sun is shining.
F. Tomorrow's temperature is above the melting point of water.

Exercise 4 *Assume that you have the following information: With probability* 0.1*, my new friend is female and at most* 29 *years old. With probability* 0.3*, my new friend is female and at least* 30 *years old.*

What $\Pr[$*my new friend is female*$]$*? (And which lemma do we use?)*

Why don't we know $\Pr[$*my new friend is female*$]$ *in the following case? With probability* 0.1*, my new friend is female and at most* 50 *years old. With probability* 0.3*, my new friend is female and at least* 30 *years old.*

Exercise 5 *Which of the events in Exercise 3 are independent? Why? Why not?*

Exercise 6 *In a (6 out of 49) lottery, the probability of winning is* $1/13\,983\,816$*. The probability that* 49 *is one of the numbers if* $\frac{6}{49}$*.*

You play in the lottery and at the same time roll a die. What is the probability that you win the lottery and *roll a six?*

Why is it not possible to use analogous reasoning to answer the following question? What is the probability that you win the lottery and *one of the lottery-numbers if* 49*?*

Exercise 7 *Let A and B be independent events. What is* $\Pr[A \vee B]$*? (Hint: Notice that $A \vee B$ is the same as $(A \wedge \neg B) \vee (A \wedge B) \vee (\neg A \wedge B)$. Use Definition 1, Lemma 4, Lemma 2, and Lemma 3.)*

2 Random variables and distributions

So far, we have seen how to talk about events and their probabilities. What we have not discussed is how events are described (we have just used informal textual discussions) and how probabilities can be computed in the first place.

To understand how to describe events, we first look at how we formulated events. Examples are: "Alice rolls a six" or "tomorrow it rains". If we want to write these events in a more mathematical way, we can introduce variables for the things we do not know about and then express the events in terms of these. For example: Let a denote the outcome of Alice's die roll. Let w denote tomorrows weather (i.e., $w \in \{rain, sun, clouds\}$.). Then the above events can be written as $a = 6$ and $w = rain$. (In probability theory, capital letters are typically used for random variables. In crypto, however, also small letters are used.)

We see that we can split the description of events into two parts. First, we define some variables that refer to some random values (such as the outcome a of a die roll or the weather w), and then we can express an event as a precise mathematical statement. The variables representing random values are called *random variables*.

random variables

Then, formally, an event is just any mathematical statement (i.e., Boolean formula) involving random variables. But that still does not allow us to compute the probability of a given event. To compute, e.g., $\Pr[x > 10]$, we need to know which value x takes with which probability. To know this, we need to specify the *distribution* of x. A distribution is a mathematical object that, for some given random variables $x_1, \dots, x_n$, tells us, for any possible values $v_1, \dots, v_n$ that $x_1, \dots, x_n$ can have, what the probability of those values are. That is, a distribution $\mathcal{D}$ is a function that maps $v_1, \dots, v_n$ to $\mathcal{D}(v_1, \dots, v_n) = \Pr[x_1 = v_1, \dots, x_n = v_n]$.

distribution

Some examples:

- If x, y are random variables describing Alice's and Bob's die rolls, then the corresponding distribution is $\mathcal{D}(v_x, v_y) := \frac{1}{36}$ for $v_x, v_y \in \{1, \dots, 6\}$ (and $\mathcal{D}(v_x, v_y) := 0$ if v_x or v_y is not $1, \dots, 6$).
- If the weather forecast says that there is a 10 % chance of rain, 20 % of clouds, and 70 % of sun, then they mean the distribution of the random variable describing the weather is $\mathcal{D}(\mathit{rain}) = 0.1$, $\mathcal{D}(\mathit{clouds}) = 0.2$, $\mathcal{D}(\mathit{sun}) = 0.7$. (We ignore here that several weathers can occur together.)
- If Bob rolls a fair die (random variable x) and Alice has a magic die that always rolls a number y at least as high as Bob's die roll (and all these numbers are equally likely), then this situation is described by a distribution $\mathcal{D}(v_x, v_y) := \frac{1}{6} \cdot \frac{1}{7 - v_x}$ for $v_x, v_y \in \{1, \dots, 6\}$, $v_y \geq v_x$. (Here $\frac{1}{6}$ is the probability to get $x = v_x$, and $\frac{1}{7-x}$ is the probability to get a $y = v_y$ in that case.)
- One particular important example of a distribution (on a single random variable x) is the *uniform distribution*. The uniform distribution $\mathcal{U}$ on a finite set M assigns to all elements of M the same probability. That is, $\mathcal{D}(v) = \frac{1}{|M|}$ for all $v \in M$ and $\mathcal{D}(v) = 0$ for all $v \notin M$. (Here $|M|$ denotes the size of M.)

uniform distribution

Once we have a distribution $\mathcal{D}$ that describes the random variables $x_1, \dots, x_n$, we can compute the probability of any event E that contains only these random variables. Namely, we need to compute the probability that the random variables take values such that E becomes true. Formally:

$$\Pr[E] = \sum_{\substack{v_1,\dots,v_n \\ E(v_1,\dots,v_n) \text{ is true}}} \mathcal{D}(v_1, \dots, v_n) \tag{1}$$

Here $E(v_1, \dots, v_n)$ denotes the result of replacing, in the Boolean formula E, each x_i by the value v_i.

Example: If x, y are fair die rolls (i.e., $\mathcal{D}(v_x, v_y) = \frac{1}{36}$), then

$$\Pr[x > y] = \sum_{\substack{v_x, v_y \\ v_x > v_y}} \mathcal{D}(v_x, v_y) = \sum_{\substack{v_x, v_y \\ v_x > v_y}} \frac{1}{36} \overset{(*)}{=} 15 \cdot \frac{1}{36} = \frac{5}{12}$$

(Here $(*)$ uses that there are 15 values of $v_x, v_y \in \{1, \dots, 6\}$ that satisfy $v_x > v_y$.)

Summarizing: Once the distribution $\mathcal{D}$ of the random variables $x_1, \dots, x_n$ is fixed, we can at least in principle compute the probability $\Pr[E]$ of any event involving only these random variables by using (1).

Derived random variables. In the preceding section, we have introduced the concept of a random variable. The random variables we have seen so far were random variables described by a variable name x (and an associated distribution tells us how probable particular values of x are). In the following, we call such random variables elementary random variables.[1]

However, there is also a second kind of random variables, namely any expression depending on elementary random variables. We call these derived random variables.[2] For example, when we have two random variables x and y that describe the outcome of rolling die 1 and 2, then $x + y$ is the random variable that describes the sum of the two dice.

Examples for derived random variables:

- $x + y$, the sum of two die rolls.
- If t is tomorrow's weather temperature in degrees Celsius, then $t \cdot \frac{9}{5} + 32$ is the random variable describing tomorrows weather in degrees Fahrenheit.
- Consider an execution of a given cryptographic protocol. Let r be the random variable that describes all the random choices the protocol parties and the adversary make. Let f be the Boolean function that computes whether given random choices r, the adversary successfully breaks the protocol (such a function is well-defined but typically quite complex). Then $f(r)$ is a random variable describing whether the adversary breaks the protocol.
- Given random variables x, y, the pair (x, y) is a random variable. Such joint random variables are important to express things like z does not depend on x or y. This would be expressed as z and (x, y) are independent (see below).

Independence of random variables. Above, we have introduce the concept of independent events. An event A is independent of an event B if we don't learn anything about whether A holds from whether B holds. Similarly, we can define *independent random variables*. Intuitively, x and y are independent if x gives no information about y and vice versa.

independent random variables

Definition 2 *Two random variables x and y (with respect to some distribution $\mathcal{D}$ that assigns probabilities to both) are independent if for all v_x, v_y we have* $\Pr[x = v_x \wedge y = v_y] = \Pr[x = v_x] \Pr[y = v_y]$.

Examples:

- If x and y describe the outcome of fair dice rolls, then x and y are independent.
- If Alice chooses some value x (in whatever way she wishes), and then Bob afterwards chooses x uniformly from some set M, then x and y are independent.
- If Alice picks a secret key k, then k is independent from any value x that the adversary might produce before Alice uses k for the first time.

[1] This is not an established name.
[2] This is not an established name.

- If x and y are independent random bits (i.e., $\Pr[x = a, y = b] = \frac{1}{4}$ for all $a, b \in \{0, 1\}$), then x and $x \oplus y$ are independent. (You can check that $\Pr[x = a, x \oplus y = b] = \frac{1}{4} = \Pr[x = a] \Pr[x \oplus y = b]$ for all $a, b \in \{0, 1\}$.)

Independent random variables lead to independent events:

Lemma 5 *If the event A depends only on $x_1, \ldots, x_n$ (i.e., A is a Boolean formula containing only those variables), and B depends only on $y_1, \ldots, y_m$, and $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$ are independent, then A and B are independent.*

Example: Let x and y be independent dice rolls of Alice and Bob. Then x and $x+y \bmod 6$ are independent (similar to the example with the independent random bits above). Hence $A :=$ "$x < 5$" and $B :=$ "$x + y \bmod 6 < 3$" are independent by Lemma 5.

2.1 Exercises

Exercise 8 *For each event described in Section 1, restate the event in the following form: List the random variables, describe the distribution (if sufficient information is available to do so), define the event (as a formula involving only the random variables).*

Exercise 9 *Let x, y be random variables with value $1, \ldots, 6$.*

(i) Find a distribution $\mathcal{D}_1$ such that $\Pr[x = i] = \Pr[y = i] = \frac{1}{6}$ for all i, and such that $\Pr[x = y] = 1$.

(ii) Find a distribution $\mathcal{D}_2$ such that $\Pr[x = i] = \Pr[y = i] = \frac{1}{6}$ for all i, and such that $\Pr[x = y] = \frac{1}{6}$.

This shows that knowing the distribution of the individual random variables (the so-called marginal distribution*, i.e., $\Pr[x = i]$ and $\Pr[y = i]$) is not enough, one always needs to know the joint distribution of all random variables together.*

marginal distribution

Exercise 10 *Let x be a (uniformly) random die roll. Let y be a random bit (independent of x). What is the distribution of x, y?*

Exercise 11 *Let x be a (uniformly) random die roll. Let y be 1 with probability $1/x$, and 0 otherwise. What is the distribution of x, y?*

Exercise 12 *Let $\mathcal{D}(v, w) = \frac{1}{36}$ for all $v, w \in \{1, \ldots, 6\}$ and and $\mathcal{D}(v, w) := 0$ otherwise. Assume x, y are distributed according to $\mathcal{D}$. What is $\Pr[x = y]$?*

Exercise 13 *Let $\mathcal{D}(v, w) := \frac{1}{10v}$ for all $v, w \in \{1, \ldots, 10\}$ with $w \leq v$ and $\mathcal{D}(v, w) := 0$ otherwise. Assume x, y are distributed according to $\mathcal{D}$. What is $\Pr[x = y]$?*

Exercise 14 *Let x be uniformly distributed on $\{1, \ldots, 6\}$. What is the distribution of x^2? What is the distribution of $(x - 3)^2$? (Hint: To get the distribution of a derived random variable d, just compute $\Pr[d = i]$ for all i.)*

Exercise 15 *In which of the following cases are the two random variables independent? Why?*

- *x, y are two fair die rolls.*
- *x is a fair die roll, y is uniformly chosen from $1, \ldots, n$.*
- *a, b is are (independent) random bits. $x := a \oplus b$ and $y := a$.*
- *a, b is are (independent) random bits. $x := a \oplus b$ and $y := a \cdot b$.*

3 Conditional probabilities

So far, we have seen how probabilities of particular are modeled if these probabilities hold unconditionally. However, often one has to ask "what if"-questions. E.g., assuming the sun shines tomorrow, what is the probability that the temperature is at least 10 degrees Celsius. (At the time of this writing, the answer is "pretty low".) To answer such questions, we use *conditional probabilities*. A conditional probability $\Pr[A|B]$ tells us what the probability of the event A is under the condition that the event B is known to happen.

conditional probability

Examples:

- $\Pr[x + y = 12|x = 6] = \frac{1}{6}$ where x, y are independent fair die rolls. The reason is: If you already know that $x = 6$, then you just need to know what the probability of $y = 6$ is because $x = y = 6$ is the only case that leads to $x + y = 12$. In contrast, if we do not know what x is, we have $\Pr[x + y = 12] = \frac{1}{36}$.
- Assume a cancer patient has 81 % chance of surviving 5 years ($\Pr[\text{lives 5 years}] = 0.81$), and a 77 % chance of surviving 10 years ($\Pr[\text{lives 10 years}] = 0.77$). Then we may ask what the chance is to live to the 10th year assuming one has already survived 5 years. (I.e., after 5 years, if the patient is still alive, what is the probability of surviving another 5.) Using Definition 3 below, we can compute $\Pr[\text{lives 10 years}|\text{lives 5 years}] \approx 0.95$.
- The probability that we have more than 10 degrees Celcius ($t > 10$) tomorrow under the condition that the sun shines ($w = \textit{sun}$) is written $\Pr[t > 10|w = \textit{sun}]$. I cannot tell what this probability is, but it is likely that $\Pr[t > 10|w = \textit{sun}] > \Pr[t > 10]$. (Meaning that if the sun shines, the probability of warm weather is higher than on average.)

But mathematically, how do we compute conditional probabilities? It turns out that the formal definition directly gives us a means for computing conditional probabilities:

Definition 3 (Conditional probability) *Let A and B be events (with respect to some distribution $\mathcal{D}$). Assume that $\Pr[B] > 0$. Then $\Pr[A|B] := \Pr[A \wedge B]/\Pr[B]$.*

If one thinks about it, this definition is quite natural. If B happens in 1 out of n cases ($\Pr[B] = 1/n$), and A happens in 1 out of m of the cases in which B happens ($\Pr[A|B] = 1/m$), then $A \wedge B$ happens in 1 out of nm cases. Thus $\Pr[A \wedge B] = \Pr[A|B] \cdot \Pr[B]$, and from this Definition 3 follows.

Notice that we require that $\Pr[B] > 0$. (Otherwise $\Pr[A \wedge B]/\Pr[B]$ would not be defined.) This means that we cannot ask "what if"-questions that talk about impossible situations. (E.g., "could I fly if I had wings?": $\Pr[\text{I can fly}|\text{I have wings}]$ is undefined because $\Pr[\text{I have wings}] = 0$, I just don't have wings, no probability involved.)

An example for a computation of a conditional probability: Let x and y be fair dice rolls. We may ask what the probability is that $x = 6$ assuming $x + y > 6$. We count that there are 21 possibilities for (x, y) that lead to $x + y > 6$. Thus $\Pr[x + y > 6] = \frac{21}{36}$ since each combination has the same probability $\frac{1}{36}$. There are 6 possibilities such that $x + y > 6$ and $x = 6$. Hence $\Pr[x = 6 \wedge x + y > 6] = \frac{1}{6}$. Thus

$$\Pr[x = 6|x + y > 6] = \frac{\Pr[x = 6 \wedge x + y > 6]}{\Pr[x + y > 6]} = \frac{1}{6} \cdot \frac{36}{21} = \frac{2}{7} \approx 0.29.$$

Notice that $\Pr[A|B]$ and $\Pr[B|A]$ are quite different things. For example, the probability that a given person will possess more than 100000 Euro (event A) under the condition that he wins the jackpot in the lottery (event B) is quite high ($\Pr[A|B] \approx 1$). But the probability that someone wins the lottery jackpot under the condition that he will possess more than 100000 Euro is not very high (because there are many more likely causes for getting 100000 Euro than winning in the lottery). Hence $\Pr[B|A] \approx 0$. One can, however, in certain cases compute $\Pr[A|B]$ from $\Pr[B|A]$ (and vice versa) using the following theorem (which follows directly by Definition 3):

Theorem 1 (Bayes' law) *Let A and B be events with $\Pr[A], \Pr[B] > 0$. Then*

$$\Pr[A|B] = \frac{\Pr[B|A]\Pr[A]}{\Pr[B]}.$$

Another interesting fact about conditional probabilities is that it gives a new view on independent events. We said that intuitively, events A and B are independent, if knowing whether A occurs does not tell us whether B occurs. That is, the probability that B occurs should be the same as the probability that B occurs under the condition that A occurs. Mathematically, $\Pr[B] = \Pr[B|A]$. This, indeed, is an alternative but equivalent definition of independence as the following lemma shows (which again follows very easily from Definition 3 and Definition 1):

Lemma 6 *Let A and B be events. Assume that $\Pr[A] > 0$. Then A and B are independent if and only if $\Pr[B] = \Pr[B|A]$.*

3.1 Exercises

Exercise 16 *Write the following questions using conditional probabilities and answer them.*

(a) When the sun shines, the temperature is above 10 degrees with 90 % chance. And tomorrow the sun shines with 10 % probability. How likely is it that the sun shines and *the temperature is above 10 degrees?*

(b) I have a pair of weird dice. Tests show: Each of them individually rolls a six with probability $\frac{1}{6}$. But the probability that both *roll a six is $\frac{1}{10}$ (and not $\frac{1}{36}$ as would be expected from fair dice). What is the probability that the second die rolls a six when the first did.*

Exercise 17 *A random person has probability $\frac{1}{1000}$ of being in Hospital. A random person has probability $\frac{1}{50}$ of being ill. A random person in Hospital has probability 9/10 of being ill. What is the probability that a random ill person is in Hospital?*

Exercise 18 *Identify the two implicit statements about conditional probabilities in Figure 1 and write them down as conditional probabilities.*

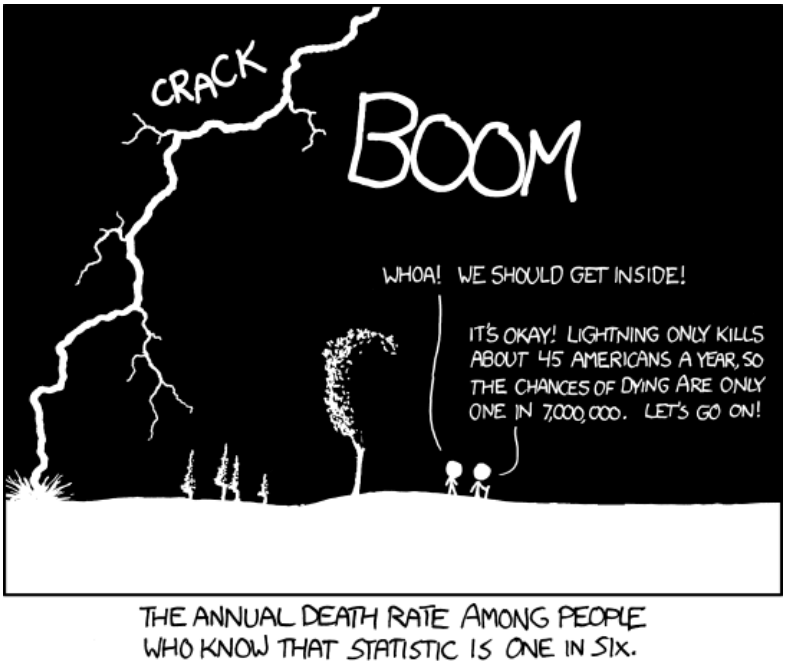

Figure 1: xkcd comic #795 by Randall Munroe, Creative Commons by-nc 2.5.

4 Expected value

If we know the distribution of some random variable x, we can ask the question: what is the average value of that random variable? (Of course, this makes only sense if the random variable is a number, not, e.g., for the weather-example with $w \in \{rain, sun, clouds\}$.) We call this average value the *expected value* $\mathrm{E}[x]$ of x, and we can compute it according to the following definition:

expected value

Definition 4 (Expected value) *Let x be a random variable (whose possible values are real numbers). Then the expected value $\mathrm{E}[x]$ of x is defined as*

$$\mathrm{E}[x] = \sum_v v \cdot \Pr[x = v]$$

where v ranges over all possible values of x.

For example, if we wish to know the average value of a die roll x, we compute

$$\mathrm{E}[x] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{7}{2}.$$

The expected value can also be computed from derived random variables. For example, for dice rolls x, y, we can ask what their expected sum $\mathrm{E}[x + y]$ is. We could compute this by explicitly evaluating $\mathrm{E}[x + y] = \sum_{v_x=1}^{6} \sum_{v_y=1}^{6} (x + y) \cdot \frac{1}{36}$, but Lemma 7 below gives a better method.

Another example: An adversary wishes to find out a secret key $k \in \{1, \ldots, 2^n\}$. All he can do is to try out keys and see whether he got the right one or not. k has been chosen uniformly. The adversary tries all keys systematically (i.e., $k = 1$, then $k = 2$, ...).

What is the average number t of tries he needs? We have $t = k$ because he needs k tries to reach the key k. Thus

$$\mathrm{E}[t] = \mathrm{E}[k] = \sum_{v=1}^{2^n} k \cdot \Pr[k = v] = \sum_{v=1}^{2^n} k \cdot \frac{1}{2^n} = \frac{2^n(2^n+1)}{2}\frac{1}{2^n} = 2^{n-1} + \frac{1}{2}.$$

So the adversary needs approximately 2^{n-1} tries on average.

Lemma 7 (Basic properties of the expected value) *For any random variables x and y, and any real-number c, we have:*

(a) $\mathrm{E}[c \cdot x] = c \cdot \mathrm{E}[x]$.

(b) $\mathrm{E}[x + y] = \mathrm{E}[x] + \mathrm{E}[y]$.

(c) If $\Pr[x \geq y] = 1$, *then* $\mathrm{E}[x] \geq \mathrm{E}[y]$.

4.1 Exercises

Exercise 19 *We roll a fair die x. What is* $\mathrm{E}[x^2]$*?*

Exercise 20 *You roll two fair die x, y. You get $p := x^2 + y$ points. What is* $\mathrm{E}[p]$*? (Hint: Use that you already know* $\mathrm{E}[x^2]$ *and* $\mathrm{E}[y]$*.)*

Exercise 21 *You roll one fair die x. You get $p := x^2 + x$ points. What is* $\mathrm{E}[p]$*?*

Exercise 22 *You throw a die until you roll one. What is the expected number of tries? (Hint: First compute* $\Pr[$*the number of tries is i*]$ *for all i.)*

5 Game notation

So far, we have always assumed that the distribution of the random variables is explicitly given, namely by assigning a probability for any possible combination of values of the random variables. Except for the simplest situations, however, this is inconvenient. Computing all these probabilities explicitly leads to complicated and uninsightful formulas.

Instead, at least in cryptography, the distribution of random variables is usually given by describing a kind of little "program" (often called a *game*) that describes how the random variables' values are chosen. This implicitly specifies their distribution. **game**

For example, we can describe a distribution for the random variables x, y, z by saying: First, x is chosen uniformly from $1, \ldots, 10$. Then y is chosen uniformly from $1, \ldots, x$. Then z is computed by running the algorithm A on input $x \cdot y$.

A common notation for such a game is: $x \xleftarrow{\$} \{1, \ldots, 10\}$, $y \xleftarrow{\$} \{1, \ldots, x\}$, $z \leftarrow A(x \cdot y)$.

That is, we use the following symbols: $x \xleftarrow{\$} M$ picks x uniformly at random from the set M. $x \leftarrow A(e_1, \ldots, e_n)$ evaluates the expressions $e_1, \ldots, e_n$ (which may contain random variables already assigned), gives the result to the algorithm A, and assigns the output of A to x. And $x \leftarrow e$ just evaluates the expression e and assigns the result to x.

Examples:

- $x \xleftarrow{\$} \{1, \ldots, 6\}$, $y \xleftarrow{\$} \{1, \ldots, 6\}$: Two fair die rolls.

- $k \xleftarrow{\$} K$, $m \xleftarrow{\$} M$, $c \leftarrow E(k,m)$, $m' \leftarrow A(c)$: A key k is chosen at random from a key space K, a message m at random from M, then the ciphertext c is computed by encrypting m with k, and finally the adversary A is given c and outputs m'. (We may then ask questions such as how probable it is that the adversary guesses m, i.e., what is $\Pr[m = m']$?)

This new notation allows us to very compactly write probabilities. Namely, we when writing $\Pr[E : G]$ we ask for the probability $\Pr[E]$ of the event E when the distribution of the random variables is given by the game G. For example:

- $\Pr[x = y : x \xleftarrow{\$} \{1,\ldots,6\}, y \xleftarrow{\$} \{1,\ldots,6\}] = \frac{1}{6}$. (The probability that two independent die rolls are equal.)
- $\Pr[m = m' : k \xleftarrow{\$} K, m \xleftarrow{\$} M, c \leftarrow E(k,m), m' \leftarrow A(c)]$. (The probability that the adversary guesses the plaintext m. Of course, this probability depends on the encryption function E and on the adversary A.)
- $\Pr[x = y : x \xleftarrow{\$} \{1,\ldots,2^n\}, y \xleftarrow{\$} A()] \leq 2^{-n}$. (The probability that the adversary guesses a totally unknown value x.)

In some cases, it may be necessary to explicitly compute the distribution $\mathcal{D}$ on $x_1,\ldots,x_n$ defined by a game G. (Most of the time, however, one fares best with the intuitive description above.) This can be done recursively by using the following formulas: Let G' be a game defining $x_1,\ldots,x_{n-1}$.

$$\Pr[x_1 = v_1,\ldots,x_n = v_n : G', x_n \xleftarrow{\$} M]$$
$$= \Pr[x_1 = v_1,\ldots,x_{n-1} = v_{n-1} : G'] \cdot \begin{cases} \frac{1}{M^*} & M^* \\ 0 & v_n \notin M^* \end{cases} \tag{2}$$
$$\Pr[x_1 = v_1,\ldots,x_n = v_n : G', x_n \leftarrow e]$$
$$= \Pr[x_1 = v_1,\ldots,x_{n-1} = v_{n-1} : G'] \cdot \begin{cases} 1 & v_n = e^* \\ 0 & v_n \neq e^* \end{cases} \tag{3}$$
$$\Pr[x_1 = v_1,\ldots,x_n = v_n : G', x_n \leftarrow A(e_1,\ldots,e_m)]$$
$$= \Pr[x_1 = v_1,\ldots,x_{n-1} = v_{n-1} : G'] \cdot \Pr[A(e_1^*,\ldots,e_m^*) = v_n] \tag{4}$$

Here $M^*, e^*, e_1^*, \ldots, e_m^*$ stand for $M, e, e_1, \ldots, e_m$ with all x_i replaced by v_i. And $\Pr[A(e_1^*,\ldots,e_m^*) = v_n]$ is the probability that the algorithm A outputs v_n on inputs $e_1^*,\ldots,e_m^*$.

5.1 Exercises

Exercise 23 *Find five probabilities that in this document that are not written in game notation and rewrite them in game notation.*

Exercise 24 *Write the following in game notation: "x and y are fair die rolls. z is chosen uniformly from $\{1,\ldots,xy\}$. What is the probability that $z = xy$?"*

Exercise 25 *Explain why the equations* (2)–(4) *hold.*

Exercise 26 *What is* $\Pr[x = z : x \xleftarrow{\$} \{1,\ldots,6\}, y \xleftarrow{\$} \{1,\ldots,3\}, z \leftarrow xy]$*?*