

Lecture 3. Some Probability.

Recall that a die (“die” is the singular of “dice”) is cube with dots on each of its six faces with from 1 to 6 dots on each face. (Check Figure 1 on page 192 if you are unsure of what a die is.) Suppose that we toss such a die over and over many times recording how many dots show after each toss. So we’ll wind up with a scrambled list of the numbers 1, 2, 3, 4, 5, 6. What do we “expect” to find with this list? If our list includes the results of many tosses, say 1000, then we probably would expect to see that each of the six numbers shows up essentially the same number of times as any other number. (Of course, this requires that our die be a “fair” one, that is, one that is not loaded to favor one or two faces.) Assuming that our expectation is met, then we’d likely say something like, “The probability of any one number showing is $1/6$.” Here we want to try to make this notion of probability more precise in some particularly simple settings. You will undertake a more extensive study of probability next term when you get into Chapter 9.

The typical setup that we shall deal with is one where we perform some experiment, such as tossing a die, and recording or noting the outcome. The totality of outcomes is called the **sample space** and is usually denoted by the capital Greek letter Ω . In this introduction we will assume that the sample space Ω is finite. Then each subset E of Ω is called an **event**.

For example, suppose our experiment is tossing a single die. Then our sample space is $\Omega = \{1, 2, 3, 4, 5, 6\}$. Of the 2^6 events (the subsets of Ω) a couple of them is the simple event $E = \{\text{the outcome is a 4}\} = \{4\}$, and the event $E = \{\omega \in \Omega \mid \omega \text{ is even}\} = \{2, 4, 6\}$.

Given such an experiment and sample space Ω , we would like to have some way of measuring the “likelihood” that for some event E an experiment will have an outcome in E . Probability is a technique to provide such a measure. So a probability function P for such an experiment is a function that assigns to each event E a real number $P(E)$ on the interval $[0, 1]$. In particular, we require that $P(E) = 0$ iff the event E is impossible, and we require that $P(E) = 1$ iff the event is certain. For example, if for some outcome a probability function P assigns $P(E) = 1/2$, then this is stating that the probability (for that particular probability function) an experiment will have an outcome ω that falls in E is probably $1/2$. More later.

So exactly what does determine what functions are probability functions. The requirements are simple. Let Ω be a finite set, the sample space. Then a function P defined on the power set $\mathcal{P}(\Omega)$ to the real interval $[0, 1]$ is a **probability function** on Ω if it satisfies the two axioms:

$$\mathbf{P}_1 \quad P(\Omega) = 1;$$

$$\mathbf{P}_2 \quad \text{If } E \text{ and } F \text{ are disjoint events, then } P(E \cup F) = P(E) + P(F).$$

Before we see some more specific examples of sample spaces and probability functions, let’s list four key properties of probability functions that we will use extensively. First, though, if Ω is a sample space and E is an event in this space, then the complement E^c of E is the set

$$E^c = \Omega \setminus E, \quad \text{so } E \text{ and } E^c \text{ are disjoint and } \Omega = E \cup E^c.$$

We are now going to restate the Theorem given on page 191 of the text, but we will omit the proof. However, we expect you to expend a little effort to follow and understand its proof — it is not only fun in its own right, but it illustrates some very important properties of sets and probability functions.

Theorem 3.1. *Let P be a probability function on a finite sample space Ω . Then*

- (a) $P(\emptyset) = 0$;
- (b) $P(E^c) = 1 - P(E)$ for every event E in Ω ;
- (c) $P(E \cup F) = P(E) + P(F) - P(E \cap F)$ for every pair of events E and F ;
- (d) For every pairwise disjoint set of events E_1, E_2, \dots, E_n

$$P(E_1 \cup E_2 \cup \dots \cup E_n) = P(E_1) + P(E_2) + \dots + P(E_n).$$

Here is one of the great advantages of having only a finite sample space Ω . For then each single outcome $\omega \in \Omega$ is an event and each event $E \subseteq \Omega$ is the disjoint union of its elements. Hence, by (d) of the Theorem, the probability of an event E is simply the sum of the probability of its elements. (For infinite sample spaces the Theorem is still correct, and all events will be the disjoint unions of their elements, but for such sample spaces those unions are not likely to be finite. In fact they can be quite ugly, so that in these cases we need a more delicate notion of probability.)

Example 3.1. Suppose that our “experiment” is to choose two cards (without repetition) from a standard 52 card bridge deck. Assuming that our deck is “fair” with no marked or otherwise funny cards, then we usually choose the probability function that says that the probability of drawing any two cards is the same as drawing any other two, in which case we say that any two pairs are **equally likely**. Now the total number of possible pairs of two cards is simply the number of two card sets in the deck, namely the number

$$\binom{52}{2} = \frac{52!}{2!50!} = 1326.$$

Since we assume that drawing any pair is equally likely, by Axiom P_1 the sum of these 1326 equal probabilities must be 1. That is, if ω_0 is the pair consisting of, say, the ace of hearts and the ace of spades, then by the equally likely assumption

$$\sum_{\omega \in \Omega} P(\omega) = (1326)P(\omega_0) = 1.$$

So for any pair ω of cards we have

$$P(\omega) = P(\omega_0) = \frac{1}{1326}.$$

Therefore, in this setting we know the probability of drawing any one pair, and so the task of finding the probability $P(E)$ of any event boils down to counting the number of elements in E . ■

This important fact deserves a formal statement, so here it is:

Equally Likely Probabilities. *Let Ω be a finite sample space and let P be an equally likely probability function on Ω . Then for each event $E \subseteq \Omega$*

$$P(E) = \frac{|E|}{|\Omega|}.$$

Example 3.2. Let's return to the equally likely setup we had in Example 3.1. Here, the sample space, of all possible two card draws, has 1326 elements. So, first, say we find the probability that if we draw two cards, we'll get a pair of aces. That is, we want $P(E)$ where E is the event of "two aces". Well, to count E is easy. There are 4 aces and we want two of them, so there are exactly

$$\binom{4}{2} = \frac{4!}{2!2!} = 6,$$

such pairs. Thus, the probability of drawing two aces is

$$P(E) = \frac{|E|}{|\Omega|} = \frac{6}{1326} = 0.00452.$$

So being so lucky would be very remote.

Next let's ask for the probability of drawing a pair of clubs. This time let F_1 be the event consisting of all pairs of clubs. We are looking for

$$P(F_1) = \frac{|F_1|}{|\Omega|}.$$

Well, the number of clubs is 13, so the number of pairs of clubs, the size of F_1 , is

$$|F_1| = \binom{13}{2} = 78,$$

so the desired probability is

$$P(F_1) = \frac{|F_1|}{|\Omega|} = \frac{78}{1326} = 0.0588.$$

That is, assuming a fair deck, we would expect to draw a pair of clubs about 6% of the time.

Next let's ask for drawing two cards of the same suit. Since each suit contains exactly 13 cards, the number of pairs of any one suit is the same as the number of pairs of clubs, that is, 78. If we let F_1, F_2, F_3, F_4 be the four events of two clubs, diamonds, hearts, and spades, then these four events are pairwise disjoint and each has the same probability, $P(F_1) = 0.0588$. Now we are looking for the probability of the disjoint union F of these four events, and by part (d) of the Theorem, we have

$$\begin{aligned} P(F) &= P(F_1 \cup F_2 \cup F_3 \cup F_4) = P(F_1) + P(F_2) + P(F_3) + P(F_4) \\ &= 4P(F_1) = 4 \times (0.0588) = 0.235. \end{aligned}$$

In other words, you'd expect to draw two cards of the same suit about 23% of the time.

Finally, let's turn that around and find the probability that if we draw two cards, we get cards of different suits. Of course, now that's pretty easy. Our event this time is simply the set $F^c = \Omega \setminus F$. So by part (b) of the Theorem, we have

$$P(F^c) = 1 - P(F) = 1 - 0.235 = 0.765.$$

Thus, we'd expect to draw two cards of different suits slightly more than 75% of the time. ■

Achtung! Next, we'd particularly like to call your attention to Example 5 of Section 5.2, on pages 192,3. Here the event is to roll a pair of dice, in the example, one black and one red. If the goal is to record the value of the red and the value of black die on each roll as an ordered pair, then the appropriate sample space has the 36 elements listed in display (a) at the top of page 193. But often, if not usually, the focus is on the **sum** of the two dice. Since the possible sums are the numbers 2, 3, 4, . . . , 11, 12, these 11 numbers make an appropriate sample space. But here is the rub. If we work with the 36 element sample space of display (a), then the probability of any one of these 36 outcomes equals that of any other, that is, this probability is equally likely. On the other hand, however, if our interest is in the sums, then using the 11 element sample space makes sense. But here the events are not equally likely! The probability of rolling a 2 is just 0.028 while the probability of rolling a 7 is 0.167. In either case, we still have that the probability of any event is the sum of the probabilities of its individual outcomes, but the probabilities of these outcomes need not be equal for different outcomes. This is no big deal; use whatever sample space is most appropriate, But, please, be warned that by changing sample spaces you may very well lose (or gain!) equal likelihood.

Let's finish by checking out some more challenges that you should see if you can answer. Again, if necessary, we can discuss them in class.

Example 3.3. Suppose that we roll a single fair die over and over. Then the probability of rolling any value is $1/6$. Say, for instance, we toss the die five times. Then an appropriate sample space would be the set $S = D^5$ of all 5-sequences $(d_1, d_2, d_3, d_4, d_5)$ with each d in the set $D = \{1, 2, 3, 4, 5, 6\}$ of outcomes for a single toss. Then the sample space satisfies

$$|S| = 7776.$$

Okay, let's see what we can do with this.

- (a) What's the probability of rolling a 1 on the first two tosses? [Hint: This corresponds to the event $E = \{(1, \star, \star, \star, \star) : \star \in D\}$. So $|E| = 6^4$, and $P(E) = ?$.]
- (b) What is the probability of rolling a 1 on the second toss, but not on the first? [Hint: Here the event would be $F = \{(d_1, 1, \star, \star, \star) \in S : 2 \leq d_1 \leq 6 \text{ and } \star \in D\}$. The answer is that $P(F) = 5/36 = \frac{5}{6} \cdot \frac{1}{6}$. Why?]
- (c) What is the probability of rolling the first 1 on the second toss and the first 2 on the fourth toss?
- (d) What is the probability of rolling at least one 1 in the 5 tosses? ■

Remark. Those of you who know a little probability will wonder why we attacked the last problem the way we did. Recognizing that rolling a die and then rolling it again are two "*independent*" experiments render the questions of this last Example quite easy. Unfortunately, this course is organized so that this one little excursion into the subject here in Section 5.2 is not really a course in probability so much as an important example of the value of knowing how to count things. If you'd like to explore this further, check out Section 9.1 of the text.