EECS 70 Discrete Mathematics and Probability Theory
Spring 2014 Anant Sahai

Note 13

Introduction

At this point, we have seen enough examples that it is worth just taking stock of our model of probability and many of the key definitions. We are going to formalize some tools to deal with combinations of events.

Probability Recap

The most basic thing is the sample space Ω representing all the distinct possibilites of what the random experiment could yield. Doing the random experiment results in exactly one outcome $\omega \in \Omega$ being selected by nature. (And nature here is non-adversarial to us — it is not out to cause us trouble on purpose.) Ω itself might have some internal structure to it. The most common case is that Ω consists of tuples — lists — and Ω itself can be viewed as a Cartesian product $\Omega_1 \times \Omega_2 \times \cdots \times \Omega_n$, where sometimes the individual Ω_i can be thought of as sub-experiments that are all done simultaneously as a part of the larger experiment.

However, we are interested not just in individual outcomes, but in sets of possible outcomes. Sets of outcomes are called events, and formally, these must be subsets of Ω. The null subset is an allowable event, as is the entire set Ω itself. Probability is a function on events. It obeys certain natural properties, which are sometimes called the axioms of probability.

- If A is an event (subset of Ω), $\Pr[A] \geq 0$. This property is called non-negativity.
- If A is an event (subset of Ω), $\Pr[A] \leq 1$ with $\Pr[\Omega] = 1$. This property is called normalization.
- If A and B are events, and the events are disjoint (i.e. $A \cap B = \emptyset$), then $\Pr[A \cup B] = \Pr[A] + \Pr[B]$. This property is called additivity. By induction, it can easily be extended to what is called finite additivity. If A_i for $i = 1, 2, \ldots, n$ are events that are all disjoint (i.e. for all $i \neq j$, $A_i \cap A_j = \emptyset$), then

$$\Pr[\bigcup_{i=1}^{n} A_i] = \sum_{i=1}^{n} \Pr[A_i].$$

  This is complemented, for technical reasons, by a second additivity axiom that deals with countably infinite collections of events. If A_i for $i = 1, 2, \ldots$ are events that are all disjoint (i.e. for all $i \neq j$, $A_i \cap A_j = \emptyset$), then

$$\Pr[\bigcup_{i=1}^{\infty} A_i] = \sum_{i=1}^{\infty} \Pr[A_i].$$

  For the purposes of EECS70, these two forms of additivity can just be viewed together. The only important thing is that additivity requires you to be able to list all the events in question in order, one at a time. This will only become an important restriction later when we consider continuous probability.

These properties alone give rise to various useful properties that are largely inherited from set theory, and we will talk about them in the next section.

Nontrivial combinations of events

In most applications of probability in EECS, we are interested in things like $\Pr[\bigcup_{i=1}^{n} A_i]$ and $\Pr[\bigcap_{i=1}^{n} A_i]$, where the A_i are simple events (i.e., we know, or can easily compute, the $\Pr[A_i]$). The intersection $\bigcap_i A_i$ corresponds to the logical AND of the events A_i, while the union $\bigcup_i A_i$ corresponds to their logical OR. As an example, if A_i denotes the event that a failure of type i happens in a certain system, then $\bigcup_i A_i$ is the event that the system fails.

In general, computing the probabilities of such combinations can be very difficult. In this section, we discuss some situations where it can be done. Let's start with independent events, for which intersections are quite simple to compute.

Independent Events

Definition 13.1 (independence): Two events A, B in the same probability space are independent if $\Pr[A \cap B] = \Pr[A] \times \Pr[B]$.

One intuition behind this definition is the following. Suppose that $\Pr[B] > 0$. Then we have

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

Thus independence has the natural meaning that "the probability of A is not affected by whether or not B occurs." (By a symmetrical argument, we also have $\Pr[B|A] = \Pr[B]$ provided $\Pr[A] > 0$.) For events A, B such that $\Pr[B] > 0$, the condition $\Pr[A|B] = \Pr[A]$ is actually *equivalent* to the definition of independence.

A deeper intuition is that independence is the way to capture the essence (as far as inference goes) of the property that two completely unrelated subexperiments have to each other. Knowing something about one tells you nothing about the other. In fact, several of our previously mentioned random experiments consist of independent events. For example, if we flip a coin twice, the event of obtaining heads in the first trial is independent to the event of obtaining heads in the second trial. The same applies for two rolls of a die; the outcomes of each trial are independent.

The above definition generalizes to any finite set of events:

Definition 13.2 (mutual independence): Events $A_1, \ldots, A_n$ are mutually independent if for every subset $I \subseteq \{1, \ldots, n\}$,

$$\Pr[\bigcap_{i \in I} A_i] = \prod_{i \in I} \Pr[A_i].$$

Note that we need this property to hold for *every* subset I.

For mutually independent events $A_1, \ldots, A_n$, it is not hard to check from the definition of conditional probability that, for any $1 \le i \le n$ and any subset $I \subseteq \{1, \ldots, n\} \setminus \{i\}$, we have

$$\Pr[A_i | \bigcap_{j \in I} A_j] = \Pr[A_i].$$

Note that the independence of every pair of events (so-called *pairwise independence*) does *not* necessarily imply mutual independence. For example, it is possible to construct three events A, B, C such that each *pair* is independent but the triple A, B, C is *not* mutually independent.

Pairwise Independence Example

Suppose you toss a fair coin twice and let A be the event that the first flip is H's and B be the event that the second flip is H's. Now let C be the event that both flips are the same (i.e. both H's or both T's). Of course A and B are independent. What is more interesting is that so are A and C: given that the first toss came up H's, there is still an even chance that the second flip is the same as the first. Another way of saying this is that $P[A \cap C] = P[A]P[C] = 1/4$ since $A \cap C$ is the event that the first flip is H's and the second is also H's. By the same reasoning B and C are also independent.

The fact that A should be independent of C is not intuitively obvious at first glance. This is the power of the definition of independence. It tells us something nonobvious.

On the other hand, A, B and C are not mutually independent. For example if we are given that A and B occurred then the probability that C occurs is 1. So even though A, B and C are not mutually independent, every pair of them are independent. In other words, A, B and C are pairwise independent but not mutually independent.

Intersections of events

Computing intersections of independent events is easy; it follows from the definition. We simply multiply the probabilities of each event. How do we compute intersections for events which may not be independent? From the definition of conditional probability, we immediately have the following product rule (sometimes also called the chain rule) for computing the probability of an intersection of events.

Theorem 13.1: **[Product Rule]** For any events A, B, we have

$$\Pr[A \cap B] = \Pr[A]\Pr[B|A].$$

More generally, for any events $A_1, \dots, A_n$,

$$\Pr[\textstyle\bigcap_{i=1}^n A_i] = \Pr[A_1] \times \Pr[A_2|A_1] \times \Pr[A_3|A_1 \cap A_2] \times \cdots \times \Pr[A_n|\textstyle\bigcap_{i=1}^{n-1} A_i].$$

Proof: The first assertion follows directly from the definition of $\Pr[B|A]$ (and is in fact a special case of the second assertion with $n = 2$).

To prove the second assertion, we will use induction on n (the number of events). The base case is $n = 1$, and corresponds to the statement that $\Pr[A] = \Pr[A]$, which is trivially true. For the inductive step, let $n > 1$ and assume (the inductive hypothesis) that

$$\Pr[\textstyle\bigcap_{i=1}^{n-1} A_i] = \Pr[A_1] \times \Pr[A_2|A_1] \times \cdots \times \Pr[A_{n-1}|\textstyle\bigcap_{i=1}^{n-2} A_i].$$

Now we can apply the definition of conditional probability to the two events A_n and $\bigcap_{i=1}^{n-1} A_i$ to deduce that

$$\begin{aligned}\Pr[\textstyle\bigcap_{i=1}^n A_i] = \Pr[A_n \cap (\textstyle\bigcap_{i=1}^{n-1} A_i)] &= \Pr[A_n|\textstyle\bigcap_{i=1}^{n-1} A_i] \times \Pr[\textstyle\bigcap_{i=1}^{n-1} A_i] \\ &= \Pr[A_n|\textstyle\bigcap_{i=1}^{n-1} A_i] \times \Pr[A_1] \times \Pr[A_2|A_1] \times \cdots \times \Pr[A_{n-1}|\textstyle\bigcap_{i=1}^{n-2} A_i],\end{aligned}$$

where in the last line we have used the inductive hypothesis. This completes the proof by induction. □

The product rule is particularly useful when we can view our sample space as a sequence of choices. The next few examples illustrate this point.

Examples

Coin tosses.

Toss a fair coin three times. Let A be the event that all three tosses are heads. Then $A = A_1 \cap A_2 \cap A_3$, where A_i is the event that the ith toss comes up heads. We have

$$\begin{aligned}\Pr[A] &= \Pr[A_1] \times \Pr[A_2|A_1] \times \Pr[A_3|A_1 \cap A_2]\\ &= \Pr[A_1] \times \Pr[A_2] \times \Pr[A_3]\\ &= \tfrac{1}{2} \times \tfrac{1}{2} \times \tfrac{1}{2} = \tfrac{1}{8}.\end{aligned}$$

The second line here follows from the fact that the tosses are mutually independent. Of course, we already know that $\Pr[A] = \frac{1}{8}$ from our definition of the probability space in an earlier lecture note. Another way of looking at this calculation is that it justifies our definition of the probability space, and shows that it was consistent with assuming that the coin flips are mutually independent.

If the coin is biased with heads probability p, we get, again using independence,

$$\Pr[A] = \Pr[A_1] \times \Pr[A_2] \times \Pr[A_3] = p^3.$$

And more generally, the probability of any sequence of n tosses containing r heads and $n-r$ tails is $p^r(1-p)^{n-r}$. This is in fact the reason we defined the probability space this way in the previous lecture note: we defined the sample point probabilities so that the coin tosses would behave independently.

Monty Hall

Recall the Monty Hall problem from an earlier lecture: there are three doors and the probability that the prize is behind any given door is $\frac{1}{3}$. There are goats behind the other two doors. The contestant picks a door randomly, and the host opens one of the other two doors, revealing a goat. How do we calculate intersections in this setting? For example, what is the probability that the contestant chooses door 1, the prize is behind door 2, and the host chooses door 3?

Let A_1 be the event that the contestant chooses door 1, let A_2 be the event that the prize is behind door 2, and let A_3 be the event that the host chooses door 3. We would like to compute $\Pr[A_1 \cap A_2 \cap A_3]$. By the product rule:

$$\Pr[A_1 \cap A_2 \cap A_3] = \Pr[A_1] \times \Pr[A_2|A_1] \times \Pr[A_3|A_1 \cap A_2]$$

The probability of A_1 is $\frac{1}{3}$, since the contestant is choosing the door at random. The probability A_2 given A_1 is still $\frac{1}{3}$ since they are independent. The probability of the host choosing door 3 given events A_1 and A_2 is 1; the host cannot choose door 1, since the contestant has already opened it, and the host cannot choose door 2, since the host must reveal a goat (and not the prize). Therefore,

$$\Pr[A_1 \cap A_2 \cap A_3] = \frac{1}{3} \times \frac{1}{3} \times 1 = \frac{1}{9}.$$

Observe that we did need conditional probability in this setting; had we simply multiplied the probabilities of each event, we would have obtained $\frac{1}{27}$ since the probability of A_3 is also $\frac{1}{3}$ (can you figure out why?). What if we changed the situation, and instead asked for the probability that the contestant chooses door 1, the prize is behind door 1, and the host chooses door 2? We can use the same technique as above, but our final answer will be different. This is left as an exercise.

Another useful exercise is to use conditional probability to analyze the case when the two goats have different genders and Monty is committed to always reveal the location of the female goat. (This could be the door that you yourself had chosen). In this case, what happens when Monty reveals the female goat is behind one of the other two doors? What is the conditional probability of winning for switching vs not switching?

Poker Hands

Let's use the product rule to compute the probability of a flush in a different way. This is equal to $4 \times \Pr[A]$, where A is the probability of a Hearts flush. Intuitively, this should be clear since there are 4 suits; we'll see why this is formally true in the next section. We can write $A = \bigcap_{i=1}^{5} A_i$, where A_i is the event that the ith card we pick is a Heart. So we have

$$\Pr[A] = \Pr[A_1] \times \Pr[A_2|A_1] \times \cdots \times \Pr[A_5|\bigcap_{i=1}^{4} A_i].$$

Clearly $\Pr[A_1] = \frac{13}{52} = \frac{1}{4}$. What about $\Pr[A_2|A_1]$? Well, since we are conditioning on A_1 (the first card is a Heart), there are only 51 remaining possibilities for the second card, 12 of which are Hearts. So $\Pr[A_2|A_1] = \frac{12}{51}$. Similarly, $\Pr[A_3|A_1 \cap A_2] = \frac{11}{50}$, and so on. So we get

$$4 \times \Pr[A] = 4 \times \frac{13}{52} \times \frac{12}{51} \times \frac{11}{50} \times \frac{10}{49} \times \frac{9}{48},$$

which is exactly the same fraction we computed in the previous lecture note.

So now we have two methods of computing probabilities in many of our sample spaces. It is useful to keep these different methods around, both as a check on your answers and because in some cases one of the methods is easier to use than the other.

Unions of events

You are in Las Vegas, and you spy a new game with the following rules. You pick a number between 1 and 6. Then three dice are thrown. You win if and only if your number comes up on at least one of the dice.

The casino claims that your odds of winning are 50%, using the following argument. Let A be the event that you win. We can write $A = A_1 \cup A_2 \cup A_3$, where A_i is the event that your number comes up on die i. Clearly $\Pr[A_i] = \frac{1}{6}$ for each i. Therefore,

$$\Pr[A] = \Pr[A_1 \cup A_2 \cup A_3] = \Pr[A_1] + \Pr[A_2] + \Pr[A_3] = 3 \times \frac{1}{6} = \frac{1}{2}.$$

Is this calculation correct? Well, suppose instead that the casino rolled six dice, and again you win iff your number comes up at least once. Then the analogous calculation would say that you win with probability $6 \times \frac{1}{6} = 1$, i.e., certainly! The situation becomes even more ridiculous when the number of dice gets bigger than 6.

The problem is that the events A_i are *not disjoint*: i.e., there are some sample points that lie in more than one of the A_i. (We could get really lucky and our number could come up on two of the dice, or all three.) So if we add up the $\Pr[A_i]$ we are counting some sample points more than once.

Fortunately, there is a formula for this, known as the Principle of Inclusion/Exclusion:

Theorem 13.2: **[Inclusion/Exclusion]** For events $A_1, \ldots, A_n$ in some probability space, we have

$$\Pr[\bigcup_{i=1}^{n} A_i] = \sum_{i=1}^{n} \Pr[A_i] - \sum_{\{i,j\}} \Pr[A_i \cap A_j] + \sum_{\{i,j,k\}} \Pr[A_i \cap A_j \cap A_k] - \cdots \pm \Pr[\bigcap_{i=1}^{n} A_i].$$

[In the above summations, $\{i,j\}$ denotes all unordered pairs with $i \neq j$, $\{i,j,k\}$ denotes all unordered triples of distinct elements, and so on.]

I.e., to compute $\Pr[\bigcup_i A_i]$, we start by summing the event probabilities $\Pr[A_i]$, then we *subtract* the probabilities of all pairwise intersections, then we *add* back in the probabilities of all three-way intersections, and so on.

We won't prove this formula here; but you might like to verify it for the special case $n = 3$ by drawing a Venn diagram and checking that every sample point in $A_1 \cup A_2 \cup A_3$ is counted exactly once by the formula. You might also like to (hint: do this) prove the formula for general n by induction (in similar fashion to the proof of the Product Rule above).

Taking the formula on faith, what is the probability we get lucky in the new game in Vegas?

$$\Pr[A_1 \cup A_2 \cup A_3] = \Pr[A_1] + \Pr[A_2] + \Pr[A_3] - \Pr[A_1 \cap A_2] - \Pr[A_1 \cap A_3] - \Pr[A_2 \cap A_3] + \Pr[A_1 \cap A_2 \cap A_3].$$

Now the nice thing here is that the events A_i are mutually independent (the outcome of any die does not depend on that of the others), so $\Pr[A_i \cap A_j] = \Pr[A_i]\Pr[A_j] = (\frac{1}{6})^2 = \frac{1}{36}$, and similarly $\Pr[A_1 \cap A_2 \cap A_3] = (\frac{1}{6})^3 = \frac{1}{216}$. So we get

$$\Pr[A_1 \cup A_2 \cup A_3] = \left(3 \times \tfrac{1}{6}\right) - \left(3 \times \tfrac{1}{36}\right) + \tfrac{1}{216} = \tfrac{91}{216} \approx 0.42.$$

So your odds are quite a bit worse than the casino is claiming!

When n is large (i.e., we are interested in the union of many events), the Inclusion/Exclusion formula is essentially useless because it involves computing the probability of the intersection of every non-empty subset of the events: and there are $2^n - 1$ of these! Sometimes we can just look at the first few terms of it and forget the rest: note that successive terms actually give us an overestimate and then an underestimate of the answer, and these estimates both get better as we go along.

However, in many situations we can get a long way by just looking at the <u>first</u> term:

1. **Disjoint events.** If the events A_i are all *disjoint* (i.e., no pair of them contain a common sample point — such events are also called *mutually exclusive*), then

   $$\Pr[\textstyle\bigcup_{i=1}^n A_i] = \sum_{i=1}^n \Pr[A_i].$$

   [Note that we have already used this fact several times in our examples, e.g., in claiming that the probability of a flush is four times the probability of a Hearts flush — clearly flushes in different suits are disjoint events.]

2. **Union bound.** Always, it is the case that

   $$\Pr[\textstyle\bigcup_{i=1}^n A_i] \leq \sum_{i=1}^n \Pr[A_i].$$

   This merely says that adding up the $\Pr[A_i]$ can only *over*estimate the probability of the union. Crude as it may seem, in the next lecture note we'll see how to use the union bound effectively in a core EECS example.