Math-UA.233: Theory of Probability
Lecture 6

Tim Austin

tim@cims.nyu.edu
cims.nyu.edu/~tim

From last time... 1

If $F \subset S$ is an event with $P(F) > 0$, and $E \subset S$ is another event, then

$$\boxed{\underbrace{P(E \mid F) = \frac{P(E \cap F)}{P(F)}}_{\text{definition}}.}$$

STORY: As the experiment is performed, we find out that the outcome lies in F, but no other information about it. For any other event E, $P(E \mid F)$ is the new probability that E occurs, now that we know F has occurred.

From last time... 2

To reconstruct unconditioned probabilities from conditional ones, can use the **multiplication rule**

$$\boxed{P(E \cap F) = P(F)P(E \mid F)}$$

(or its more general version for a sequence of n events).

STORY: Now we have an experiment where we know how to model the conditional probabilities given F, and we can use these to figure out the unconditional probabilities of other events intersected with F.

From last time... 3

If we know the conditional probabilities of E occurring given *either F or F^c*, then we don't need to know whether or not F occurs to compute $P(E)$:

$$\boxed{P(E) = P(F)P(E \mid F) + P(F^c)P(E \mid F^c).}$$

NEXT TOPIC: an important generalization of this formula.

Definition

If S is the sample space, then a ***partition*** *of S is a (finite or infinite) sequence of events $F_1, F_2, \ldots$ such that*

- *they are mutually exclusive, and*
- $\bigcup_i F_i = S$.

Another way of saying these two conditions: every outcome lies in exactly one of the events F_i.

(Ross uses this notion but doesn't give it a name: see p68. More in Pishro-Nik: Subsections 1.2.2 and 1.4.2)

Example

A partition into two events must consist of some event F_1 *and then* $F_2 = F_1^{\mathrm{c}}$*.*

Partitions into three or more events can be more complicated.

Theorem (Law of total probability, ‘LOTP’; Ross Eq. (3.4))

If $F_1, F_2, \ldots$ are a partition of S, and $E \subset S$ is any other event, then

$$\boxed{P(E) = \sum_i P(F_i)P(E \mid F_i),}$$

where this is either a finite sum or a convergent series.

IDEA: Write E as the union of the mutually exclusive events $E \cap F_i$, and use axiom 3.

CAVEAT: If $P(F_i) = 0$ for some i, then $P(E \mid F_i)$ is technically not defined. In that case, interpret “$P(E \mid F_i)P(F_i)$” as 0.

Example (Another 'waiting for something to happen' example)

Alice and Bob play a game with a fair die. They roll it until either an even number appears, in which case Alice wins, or the number 3 appears, in which case Bob wins. What is the probability of Alice winning?

IDEA (A BIT TRICKY): Let A be the event that Alice wins, and use LOTP by partitioning according to the *value of the first roll* (F_i, $i = 1, 2, \ldots, 6$).

Bayes' formula (Ross section 3.3)

So far we have used conditional probability in two kinds of situation:

- When we know the unconditioned probability values, but get new information about a certain event occurring — then use the definition.
- When we know conditional probabilities on a certain event (and maybe its complement), and want to recover some unconditioned probabilities — then use the multiplication rule or LOTP.

In statistical applications, we often meet a third, more complicated situation.

Suppose we have an experiment, so we have S and P, and also:

- An event F that we really *care* about, and for which we know $P(F)$, but which we can't *observe* directly, and
- An event E that isn't important by itself, but which is influenced by F in a simple way so that we know $P(E \mid F)$ and $P(E \mid F^c)$, and which we can observe directly.

If we observe that E occurs, then we want to compute the probability that F occurred in light of that information, i.e. to compute $P(F \mid E)$. We can work this out from the other information that we have.

Theorem (Bayes' formula)

$$P(F \mid E) = \frac{P(E \mid F)P(F)}{P(E)} = \frac{P(F)P(E \mid F)}{P(F)P(E \mid F) + P(F^c)P(E \mid F^c)}.$$

IDEA: The numerator comes from the multiplication rule, and the denominator comes from the LOTP.

WARNING: On the left we condition on E, but on the right the conditioning is on F or F^c.

Bayes' formula has a natural interpretation in the 'subjective' ('degree of belief') view of probability.

$$\boxed{P(F \mid E) = P(F)\frac{P(E \mid F)}{P(E)}}$$

- F is a *proposed event* — we can't determine directly whether it happened or not.
- E is *evidence*, which we can determine directly.
- $P(F)$ is the *prior* ("before") probability — our initial degree of belief in F.
- $P(F \mid E)$ is the *posterior* ("after") probability — our modified degree of belief in F once we know that E happens.
- the quotient $\frac{P(E \mid F)}{P(E)}$ represents the *support* that E provides for believing in F.

Example (Ross E.g. 3.3b)

Recall the insurance company that classifies the population into "clumsy" and "dexterous".

$$P(\textit{accident within one year} \mid \textit{clumsy}) = 40\%$$

$$P(\textit{accident within one year} \mid \textit{dexterous}) = 20\%$$

$$P(\textit{a random person is clumsy}) = 30\%.$$

Suppose that a new policyholder <u>does</u> have an accident in their first year. What is the probability that they are a clumsy person?

In this example, the company *really* wants to know whether the person is clumsy or not (in order to estimate their chance of future accidents, perhaps). But the only information available to them is that an accident happened this year.

Example (Ross E.g. 3.3i)

An urn contains two type-A coins and one type-B coin. A type-A coin is (1/4)-biased towards heads, and a type-B coin is (3/4)-biased towards heads. One of the coins is randomly chosen from the urn and flipped, and it gives heads. What is the probability that it was a type-A coin?

Observe the format of the example:

- There's an 'internal' feature (the coin being type-A or type-B) which we want to guess at, but which we can't observe directly.
- There's an 'observable' feature (whether the flip gives heads or tails) whose probability is influenced by the 'internal' feature in a way that we understand.

(Ross uses this example to illustrate the alternative 'odds' notation, but I'll leave that aside.)

Bayes' formula is an extremely powerful tool. But it must also be used very carefully, because the results of applying it (so-called 'Bayesian calculations') are often not intuitive. *There is no substitute for carefully using the formula*.

Some of the most common, important, and misunderstood applications of Bayes' formula resemble the following.

Example (Ross E.g. 3.3d)

A blood test is available for cooties. If a person has cooties, the test is positive with probability 95%. But if they don't have cooties, the test may still give a "false positive" with probability 1%. We know that 0.5% of the population carries cooties. A person comes to a clinic, takes the blood test, and gets a positive result. What is the probability that they actually have cooties?

ANS = $95/294 \approx 0.323$

Many people find the previous example unintuitive: the test seems very reliable, so surely if it is positive then the probability of having cooties is high.

This intuition is wrong, because the result is skewed by the fact that *so few people have cooties to begin with*. In this example, it's actually much *more* likely that this person is one of the unfortunate false positives, than that they're one of the *very* unfortunate people who actually have cooties.

(See Pishro-Nik E.g. 1.26 for a similar example illustrated using a tree diagram.)

The prior probability (the fraction 0.5% in the previous example) is also called the *base rate*, and neglecting its effect is called the *base rate fallacy*.

This issue is very important in the real world, because the tests themselves may be expensive or dangerous, and a false positive can lead to patient distress or even more expensive and dangerous testing.

Another common source of examples is criminal cases.

Example (Ross E.g. 3.3f)

At some point in her investigation, Inspector Z is 60% convinced that Joe Bloggs is guilty. But then she finds out that the criminal was left-handed. Joe Bloggs is left-handed, but so are 20% of the population as a whole. What is Z's updated estimate of the probability that Bloggs is the criminal?

IMPORTANT REMINDER: We can solve this problem *only* if we already have an initial, unconditioned probability that Bloggs is guilty: that is, the prior $P(F)$. *You cannot apply Bayes' formula without it*.

We don't know how Inspector Z decided on her prior of 60%, but at some point it must have come from outside pure probability theory.

POSSIBLE SOURCE OF CONFUSION: In this example, '20%' is *not* the base rate (i.e., the prior). It's the entry $P(E \mid F^c)$ in Bayes' formula. This time, the statistic about the general population is not playing the same role as in the 'cooties' example.

Sometimes, instead of a single event F that we care about, the sample space is divided up into several possibilities, and we care about which one of them occurs. That is, we have a partition. Then there is a more general version:

Theorem (General Bayes' formula; Ross Prop 3.1)

If $F_1, F_2, \ldots$ are a partition of S, and $E \subset S$ is another event, then for each j we have

$$P(F_j \mid E) = \frac{P(E \cap F_j)}{P(E)} = \frac{P(F_j)P(E \mid F_j)}{\sum_i P(F_i)P(E \mid F_i)}.$$

It is essential that $F_1, F_2, \ldots$ are a partition; this doesn't work for arbitrary sequences of events.

Here is another example where the answer is not obvious, and one needs to be careful about using the formula.

Example (Ross E.g. 3l)

I have three cards, identical except that the first has an X on both sides, the second has an X on one side and an O on the other, and the last has an O on both sides. I mix up the cards, in order and orientation, then select one at random and show you one face of it. Suppose it shows an X (respectively, an O). What is the probability that the other face of that card also has an X (respectively, an O)?

Example (Ross E.g. 3n)

A bin contains 3 types of disposable flashlights: 20% of them are type-1, 30% are type-2, and 50% are type-3. The probability that a type-1 flashlight will last more than 100 hours is 70%, and the corresponding probabilities for types 2 and 3 are 40% and 30%, respectively.

(a) *What is the probability that a randomly-chosen flashlight will last more than 100 hours?*

(b) *Given that the flashlight does last more than 100 hours, what is the conditional probability that it was of type-1?*

ANS: (a) = 0.41; (b) = (0.7)(0.2)/(0.41) = 14/41