Math-UA.233: Theory of Probability
Lecture 5

Tim Austin

tim@cims.nyu.edu
cims.nyu.edu/∼tim

Conditional probability (Ross Ch 3)

Sometimes we obtain partial information about an experiment, and want to compute probabilities in the light of that information.

Example

Toss 2 fair dice. If the first die shows 3, what is the probability that the sum of the 2 values equals 8?

Initially, the sample space is $S = \{(i, j) : 1 \leq i, j \leq 6\}$. But we are told that the first die shows 3, so we know that most of the outcomes in S are irrelevant. The six remaining possibilities are

$$(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6),$$

and exactly one of these — $(3, 5)$ — gives a sum of 8.

These six remaining possibilities were initially equally likely, so a sensible answer is $1/6$.

The previous example isn't purely a math problem. It again involves a *modeling choice*. If all outcomes in S were initially equally likely, and we are given the information that a certain event F happens (in the above, $F = \{$first die shows 3$\}$), then we *assume* that the individual outcomes within F are still equally likely.

This idea has an important generalization.

Consider any experiment with sample space S and probability distribution P. Let $F \subset S$ be an event, and imagine that we are told that 'F happens', but are not given any other information about the outcome.

In the light of this information, we should update our choice of the probability values: P won't be a sensible choice any more. For instance, if $E \subset S$ is mutually exclusive with F, then the updated probability of E should be zero, even if we initially had $P(E) \neq 0$.

The natural choice is to replace them with the ratios of probability value computed *within the event F*:

Definition

*If $F \subset S$ and $P(F) > 0$, and also $E \subset S$, then the **conditional probability of** E **given** F **is***

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

Here, F is the event that we condition *on*, and E is the other event whose conditional probability we are computing.

Observations:

- $P(F \mid F) = P(F)/P(F) = 1$
  ("conditioned on the event that F happens, it is certain that F happens").

- if $E \cap F = \emptyset$, then $P(E \mid F) = P(\emptyset)/P(F) = 0$
  ("if E and F are mutually exclusive, then conditioned on the event that F happens, E has zero probability of happening).

So far, this makes the value $P(E \mid F)$ look like a sensible choice for the probability of E knowing that F has happened.

Even more is true. If we want to update our model of the experiment using the knowledge that F happens, then we can regard this as a whole *new* experiment. We keep the original sample space S, but now the probability values for $E \subset S$ are the numbers $P(E \mid F)$. This works because:

Theorem (Ross Prop 3.5.1)

Fix $F \subset S$ with $P(F) > 0$. Then the values $P(E \mid F)$ for all possible events $E \subset S$ still satisfy the three axioms of probability.

Remark

Since $P(F \mid F) = 1$, the new conditioned probabilities put all the 'weight' of probability on outcomes inside F: any events which lie outside F (i.e., are mutually exclusive with F) get probability zero.

So, ALTERNATIVE: replace S with the smaller sample space F, and just consider the probability values $P(E \mid F)$ for $E \subset F$. Either way is correct, but one may be more convenient than the other in a given situation.

Let's check that these conditional probabilities do indeed generalize the 'dice' example that we started with.

Proposition

Suppose that S is finite, that $F \subset S$ is not empty, and that P is the uniform distribution on S. Then the conditional probabilities $P(E \mid F)$ agree with the uniform distribution on F.

Example (Ross E.g. 3.2a)

Joe is 80 percent certain that his missing key is in one of the two pockets of his hanging jacket, being 40 percent certain it is in the left-hand pocket and 40 percent certain it is in the right-hand pocket. If a search of the left-hand pocket does not find the key, what is the conditional probability that it is in the other pocket?

NOTEWORTHY FEATURE: The rules of conditional probability make it clear how to solve this problem using the data we are given. But you actually don't need to specify the individual outcomes in the sample space!

Example (Ross E.g. 3.2b)

A fair coin is flipped twice. What is the conditional probability that both flips give heads, given that

(a) *the first flip gives heads?*
(b) *at least one flip gives heads?*

Conditional probability is defined in terms of unconditional probability.

But for many experiments, it is more natural to *first* choose how to model some conditional probabilities, and *then* derive unconditioned probabilities *from them*.

Mathematically, this change in viewpoint means we simply re-arrange the definition of conditional probability:

$$P(E \cap F) = P(F)P(E \mid F).$$

Of course, we can easily write down versions of the multiplication rule involving E^c in place of E or F^c in place of F. A nice way to visualize all these together is a **tree diagram**.
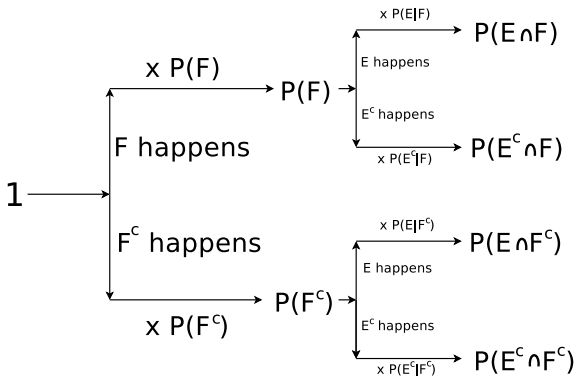
Example (Ross E.g. 3.2e)

An urn contains 3 red and 4 white balls. We draw 2 at random without replacement.

(a) *If we assume that at each draw, each ball in the urn is equally likely to be chosen, what is the probability that both balls drawn are red?*

KEY POINT: We assume that all pairs of distinct balls are equally likely to be chosen. This is the same as assuming that, for the second draw, all *remaining* outcomes are equally likely *conditioned on the colour of the first ball*. So this can be viewed as an assumption about conditional probabilities.

An urn contains 3 red and 4 white balls. We draw 2 at random without replacement.

(b) *Now suppose the red balls have weight r and the white balls have weight w, and that the probability that a given ball in the urn is the next one chosen is proportional to its weight. Now what is the probability that both balls drawn are red?*

This time, we are *not* assuming equally likely outcomes in any way. Now we are forced to answer the question in terms of conditional probabilities.

The previous product formula has an important generalization.

Proposition (The multiplication rule; Ross p59)

$$P(E_1 \cap E_2 \cap \cdots \cap E_n)$$
$$= P(E_1) \times P(E_2 \mid E_1) \times P(E_3 \mid E_1 \cap E_2) \times \cdots$$
$$\cdots \times P(E_n \mid E_1 \cap E_2 \cap \cdots \cap E_{n-1}).$$

We can visualize this using a tree with n 'levels' of branching.

CAVEAT: This only makes sense provided all the probabilities

$$P(E_1), \ P(E_1 \cap E_2), \ \ldots, \ P(E_1 \cap E_2 \cap \cdots \cap E_{n-1})$$

are not zero.

The multiplication rule clarifies our previous analysis of experiments in which we are waiting for something to happen.

Example (Repeated from lecture 6)

An urn contains 2 indistinguishable red and 2 indistinguishable blue balls. They are withdrawn one-by-one at random (and not replaced) until a red ball is obtained.

We saw previously that the probabilities of requiring 1, 2 or 3 steps until a red ball is obtained are

$$P(\{1\}) = \frac{2}{4} = \frac{1}{2}, \quad P(\{2\}) = \frac{2 \cdot 2}{4 \cdot 3} = \frac{1}{3}, \quad P(\{3\}) = \frac{2 \cdot 1 \cdot 2}{4 \cdot 3 \cdot 2} = \frac{1}{6}.$$

NEW JUSTIFICATION FOR THIS MODEL: We compute using the multiplication rule and 'equally likely conditional outcomes' at each step.

Consider again the n people randomly exchanging their hats. We showed before that the probability of no-one getting their own hat is

$$P_n = \sum_{i=0}^{n} (-1)^i / i! \quad (\approx 1/e \text{ if } n \text{ is large}).$$

What is the probability that exactly k of the n people get their own hat?

(See also E.g. 3.5d)

An ordinary deck of 52 playing cards is randomly divided into 4 piles of 13 cards each. Compute the probability that each pile has exactly 1 ace.

This problem can also be solved with or without conditional probability. It doesn't matter which you choose, but comparing the two methods sheds more light on the idea of conditional probability.

So far we have used the multiplication rule to obtain $P(E \cap F)$ from $P(F)$ and $P(E \mid F)$. But often we want to obtain $P(E)$ when $E \not\subset F$.

To do this, it can help to decompose E into two sub-events, according to whether F occurs or not:

$$E = (E \cap F) \cup (E \cap F^c).$$

This leads to

$$P(E) = P(E \cap F) + P(E \cap F^c) = P(E \mid F)P(F) + P(E \mid F^c)P(F^c).$$

We can now find $P(E)$ if we know $P(F)$ and the conditional probabilities of E given whether F or F^c occurs.

Tree diagrams give a nice way to visualize this formula: we add up over all the branches for which the event E occurs.

Example (Ross E.g. 3.3a)

An insurance company believes that people can be classified into "clumsy" and "dexterous". According to the company's statistics, within a fixed 1-year period, a clumsy person will have an accident with probability 40%, and a dexterous person will have one with probability 20%. We assume that 30% of the population is clumsy. What is the probability that a new policyholder will have an accident within the first year of owning a policy?

NOTEWORTHY FEATURE: Again, here it's tricky, and unnecessary, to actually specify the individual outcomes in the sample space!

Example (Ross E.g. 3.3h)

Twins can be either identical or fraternal. Identical twins always have the same sex, whereas the sexes of fraternal twins have roughly a 50% chance of being the same. Hospital data from Los Angeles County shows that roughly 64% of pairs of newborn twins have the same sex. Use this to estimate the percentage of pairs of newborn twins that are identical.

(The answer — 28% — comes out of Ross. I don't know his source, but it matches other values I've seen for the US. This statistic does apparently vary widely between countries. In China and Japan, twins are much rarer overall, and identical twins are substantially *more* common than fraternal twins. See, for example, http://www.twinstwice.com/twins.html)