Probability II

Margaret M. Fleck

9 November 2009

This lecture continues the discussion of probability, finishing section 6.1 of Rosen and covering parts of 6.2 and 6.4.

1 Announcements

We're hoping to have the exams done by Wednesday, but I'm not sure if we'll make that or not.

2 Basic terminology and examples

Recall from last class:

- Sample space: set of possible outcomes of experiment.
- Event: subset of sample space, containing the outcomes we are interested in.

For all the examples used in this class, the sample space will be finite. When you take a probability and statistics course, or when you see formulas for analyzing some types of scientific data, you'll see infinite sample spaces. The ideas stay basically the same in the infinite case, but the finite sums are replaced with integrals.

If our sample space is S and our event is E, and we assume that all outcomes are equally likely, then the probability of E is $Pr(E) = \frac{|E|}{|S|}$.

You'll see this probability written as either $Pr(E)$ or $P(E)$, depending on the author and their mood.

For example, suppose we want to know the probability of getting the sum 8 if we roll two dice. Then our sample space S contains all ordered pairs of numbers between 1 and 6. There are 36 such pairs. There are five pairs whose sum is 8: $(2, 6)$, $(6, 2)$, $(3, 5)$, $(5, 3)$, $(4, 4)$. Our event E is the set containing these five pairs. So $Pr(E) = \frac{|E|}{|S|} = \frac{5}{36}$.

Suppose that we roll 3 dice. Again, what's the probability of getting the sum 8? This time, our sample space contains all triples of numbers between 1 and 6, of which there are $6^3 = 216$. To figure out how many of these triples have sum 8, we adapt the trick we used last lecture to analyze combinations with repetition.

Let's represent our sum 8 as a string of 8 stars:

$* * * * * * * *$

To divide 8 up among three dice, we insert two separators, e.g. the outcome $(2, 3, 3)$ would look like:

$* * \# * * * \# * * *$

Since our dice don't have the value 0, we need to ensure that each group of stars contains at least one star. So we need to put each separator between two stars and we can't put two separators next to one another. So there are 7 places we can put a separator, from which we need to choose 2. So we have $\binom{7}{2} = 21$ ways to divide up the total 8, i.e. 21 outcomes in our event.

So the probability that our three dice sum to 8 is $\frac{21}{216}$.

3 Using familiar counting tricks

Suppose we generate random bit-strings with 10 bits. What's the chance of getting a string that contains at least one zero? Our sample space contains

2^{10} outcomes. But it's less obvious how to count the members of our event, i.e. the 10-bit strings containing at least one zero.

A useful trick in this case is to count the outcomes that are **not** in our event. In general, $|\overline{E}| = |S| - |E|$, so $P(\overline{E}) = 1 - P(E)$. In this example, $\overline{E}$ contains only one string: the string containing ten 1's. So $|E| = 2^{10} - 1$. And therefore $P(E) = \frac{2^{10}-1}{2^{10}}$.

Now, suppose that we randomly choose a positive integer no larger than 100. What's the probability that it is divisible by either 2 or 5?

In this case, $|S|$ is 100. To compute $|E|$, we count the number of integers between 1 and 100 that are divisible by 2, which is 50. Then add the number divisible by 5, which is 20. But we've now double-counted the ones that are divisible by both 2 and 5, so we need to subtract the number of those, which is 10. (Remember the "Inclusion-exclusion principle"?) So $|E| = 50+20-10 = 60$. And therefore $P(E) = \frac{60}{100} = \frac{3}{5}$.

4 Non-uniform probability

So far, we've assume that all outcomes in our sample space were equally likely. If that's not true, we need to assign a probability $p(s)$ to each outcome s in the sample space. $p(s)$ measures the inherent probability of the outcome s. We then adjust our formulas to use sums that are weighted by p.

Formally, a *probability distribution* p is a function from the sample space S to the real numbers that satisfies two conditiosn:

- $0 \leq p(s) \leq 1$ for every $s \in S$.
- $\sum_{s \in S} p(s) = 1$

Notice the new use of the summation notation. Our set of outcomes don't come with handy integer indices. Rather than forcing an indexing system on this set, we simply say that we'll sum over all elements of the set S. If you need to apply operations such as sum and union to an infinite set, this notation is sometimes essential because there's no good way to create an indexing system.

To calculate the probability of an event when we have a probability distribution for the outcomes, we sum the probabilities of all outcomes in the event:

$$P(E) = \sum_{s \in E} p(s)$$

For example, suppose that our sample space contains six movies: “The Good, the Bad, and the Ugly,” “My neighbor Totoro,” “Mulan,” “Alien 3,” “Spirited Away,” and “Ponyo.” The probabilities of watching each movie might be

The Good, the Bad, and the Ugly: $\frac{1}{10}$

My neighbor Totoro: $\frac{3}{10}$

Mulan: $\frac{1}{10}$

Alien 3: $\frac{1}{10}$

Spirited Away: $\frac{2}{10}$

Ponyo: $\frac{2}{10}$

Notice that these probabilities add up to exactly 1.

Now, suppose that our event is watching a movie by Studio Ghibli. This event contains three of our movies: “My neighbor Totoro,” “Spirited Away,” and “Ponyo.” So the probability of this event is the sum of their probabilties, i.e. $\frac{3}{10} + \frac{2}{10} + \frac{2}{10} = \frac{7}{10} = 0.7$.

5 Bernoulli Trials

Suppose that you run an experiment that has two possible outcomes: it succeeds with probability p and it fails with probability $1 - p$. This is called a Bernoulli trial. Many practical problems can be modelled as a sequence of Bernoulli trials. The key question for these problems is: what is the probability of (exactly) k successes in a sequence of n Bernoulli trials?

For example, suppose that when I play magic against my 10-year-old, my chance of winning an individual game is 0.7. What is my chance of winning exactly 5 games if I play him 10 times?

Consider a particular sequence of 5 wins and 5 losses, e.g. $WLLWWLLLWW$. The chance of getting this sequence is

$$(0.7)(0.3)(0.3)(0.7)(0.7)(0.3)(0.3)(0.3)(0.7)(0.7) = (0.7)^5(0.3)^5$$

There are $\binom{10}{5}$ ways to construct a sequence containing exactly 5 wins and 5 losses. Each of these sequences has probability $(0.7)^5(0.3)^5$. My chance of winning 5 games is the sum of these probabilities, i.e. $\binom{10}{5}(0.7)^5(0.3)^5$

In general, the probability of winning k times in n Bernoulli trials with probability of winning p is

$$\binom{n}{k}p^k(1-p)^{n-k}$$

This formula is **much** easier to remember if you understand how it was derived, because it's not hard to reconstruct. But you do need to remember what the term "Bernoulli trial" means!

To figure out the probability of winning at least k times, you would need to add up several copies of this formula.

If we consider this formula as a function of k and fix the rest of the parameters, we get a function called the binomial distribution:

$F(k) = \binom{n}{k}p^k(1-p)^{n-k}$

See Wikipedia for plots of this distribution. Notice that it's weighted towards the center like a standard bell curve, but the shape isn't quite the same. And the shape varies with the choice of n and p.

6 Random variables

A *random variable* is a function from the sample space to the real numbers.

Up, this definition looks a lot like that of a probability distribution. They are both functions from the sample space to the reals. They differ mostly in that they play different roles in a probability analysis.

Also notice that the name "random variable" is a really bad piece of terminology. First, it's a function not a variable. Second, it's deterministic; the randomness lies elsewhere in the model.

For example, suppose that we flip a coin three times. So an element of our sample space might look like HHT. Let's set up a random variable X such that $X(s)$ is the number of heads in s. E.g. $X(HHT) = 2$. We can make a table of the values for X for all outcomes in our sample space:

$$X(HHH) = 3$$

$$X(HHT) = 2$$

$$X(HTH) = 2$$

$$X(HTT) = 1$$

$$X(THH) = 2$$

$$X(THT) = 1$$

$$X(TTH) = 1$$

$$X(TTT) = 0$$

Now, suppose we want to calculate the average value of X as we do a sequence of random coin flips. This is called the *expected value* of X. It is calculated by taking the sum of the X values for all outcomes, weighted by the probability of the outcomes. That is:

$$E(X) = \sum_{s \in S} X(s)p(s)$$

In the above example, p is the same for all outcomes: $\frac{1}{8}$. So we compute $\frac{3}{8} + \frac{2}{8} + \frac{2}{8} + \frac{1}{8} + \frac{2}{8} + \frac{1}{8} + \frac{1}{8} = \frac{12}{8} = 1.5$.

Alternatively, we can coelesce some of our outcomes, grouping together the ones with the same value for X. This gives us four groups: $X = 3$, $X = 2$, $X = 1$, $X = 0$. Their probabilities are:

$P(X = 3) = \frac{1}{8}$

$P(X = 2) = \frac{3}{8}$

$P(X = 1) = \frac{3}{8}$

$P(X = 0) = \frac{1}{8}$

With this approach, our expected value is

$$3\frac{1}{8} + 2\frac{3}{8} + 1\frac{3}{8} + 0\frac{1}{8} = \frac{12}{8} = 1.5$$

Suppose that we roll a single die, what's its expected face value? Again, we add up the possible numbers, weighted by likely each is to occur. Since all six numbers are equally likely, we get

$$1\frac{1}{6} + 2\frac{1}{6} + 3\frac{1}{6} + 4\frac{1}{6} + 5\frac{1}{6} + 6\frac{1}{6} = \frac{21}{6} = 3.5$$