Lecture 2: Some simple probability calculations with cards and beans; independent events

January 23, 2019

1 Cards; sampling without and with replacement

Select two cards from a shuffled deck. What is the probability of the event 'both cards are aces'? Of the event 'both cards are the same rank?'

What is the sample space? Model the deck of cards by the set $\{1, 2, \ldots, 52\}$. The aces are 1,2,3,4; the twos are 5,6,7,8,*etc.* Model the outcome of the experiment by the sample space:

$$S = \{(i, j) : i, j \in \{1, 2, , \ldots, 52\}; i \neq j\}.$$

In effect, we think of it as drawing the first card, then drawing the second card. There is no reason that any sequence of two cards should be more likely than any other, so we use a uniform distribution.

$$|S| = 52 \times 51 = 2652.$$

(Think of tree diagram—52 possibilities for first pick, then for each of these possibilities, 51 possibilities for second pick.)

$$E_1 = \text{ 'both cards are aces' } = \{(i, j) : i, j \in \{1, 2, 3, 4\}; i \neq j\}$$

By above argument, $|E_1| = 4 \times 3 = 12$. So

$$P(E_1) = \frac{12}{2652} = \frac{1}{13 \times 17} \approx 0.0045.$$

In this account there are 12 different outcomes constituting the event. But shouldn't the answer be 6? We are treating '$A\heartsuit A\clubsuit$' as a different outcome from '$A\clubsuit A\heartsuit$'. Is that OK?

An alternative way to solve the problem: Treat each pair of outcomes in the model above as a single outcome. In other words, each outcome is a *set* of two cards. Thus each outcome in the new model corresponds to exactly two outcomes in the original one. Both the numerator $|E_1|$ and the denominator $|S|$ are divided by two, so we get the same probability.

Let $E_2 =$ 'both cards have the same rank'. E_2 is the union of the events 'both cards are Aces', 'both cards are 2's', etc. Each of these 12 events has the same probability as E_1, and they are pairwise mutually exclusive, so we can add their probabilities to get $P(E_2)$. Thus

$$P(E_2) = 13 \cdot P(E_1) = \frac{1}{17} \approx 0.059.$$

What if we replace the first card before drawing the second card? We can model the outcomes as the set of all ordered pairs $\{1, \ldots, 52\} \times \{1, \ldots 52\}$, so $|S| = 52^2$. We then get $|E_1| = 4^2$. Under the assumption of uniform distribution we get

$$P(E_1) = \frac{4^2}{52^2} = \frac{1}{13^2},$$

and thus $(E_2) = \frac{1}{13} \approx 0.077$.

Alternatively we could again try to model an outcome by just specifying the set of cards that was drawn, instead of the ordered pair. So again we are treating '$A\heartsuit A\clubsuit$' as the same hand as '$A\clubsuit A\heartsuit$'. This works, in a sense, but is not terribly useful because the distribution is no longer uniform, and computing the probability as $\frac{|E_1|}{|S|}$ leads to an incorrect result. It is exactly the same problem as with the dice: A '2 and a 3' is twice as likely as 'two 3's'.

2 Independent events.

Events E_1 and E_2 in a sample space S are *independent* if

$$P(E_1 \cap E_2) = P(E_1) \cdot P(E_2).$$

One way to think of this is that E_2 is no more or less likely to occur when E_1 occurs than when E_1 does not occur.

Example 1 Let $S = \{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\}$, with uniform distribution, representing as usual the roll of two dice. Let E_1 be the event 'the first die is odd', and let E_2 be the event 'the second die is even'. Then

$$E_1 = \{1,3,5\} \times \{1,2,3,4,5,6\}.$$

$$E_2 = \{1,2,3,4,5,6\} \times \{2,4,6\}.$$

$$E_1 \cap E_2 = \{1,2,3\} \times \{4,5,6\}.$$

So $|E_1| = |E_2| = 18$, $|E_1 \cap E_2| = 9$, and thus

$$P(E_1 \cap E_2) = 9/36 = 1/4 = 1/2 \cdot 1/2 = P(E_1) \cdot P(E_2).$$

So the events are independent.

Example 2 Same S and P as above, E_1 is 'first die is odd' again, E_2 is 'sum of the two dice is 5'. You can check that

$$P(E_1 \cap E_2) = 2/36 = 1/18,$$

$$P(E_1) = 1/2, P(E_2) = 1/9,$$

so the two events are independent (maybe unexpectedly). But let E_3 be the event 'the sum of the two dice is 6'. You can check now that

$$P(E_1 \cap E_2) = 1/12 \neq 5/72 = P(E_1) \times O(E_2),$$

so the events are not independent.

3 Beans

This is a variant of the cards problem.

A jar contains 100 navy beans, 100 pinto beans, and 100 black beans. You reach in and pull out three beans. What is the probability that the three beans all have different colors?

This is really not different from the cards problems above. We can number the beans in the jar 1,..,300. First let's consider a simpler version of the problem, where we use sampling with replacement—that is, we throw the bean back in the jar after each draw. The sample space is

$$S = \{1, \ldots, 300\} \times \{1, \ldots, 300\} \times \{1, \ldots, 300\}$$

which has cardinality

$$300^3 = 2.7 \times 10^7.$$

The event E consists of all the triples (i, j, k) where i, j, k denote beans of three different colors. We can have any of the 300 possible values for i, but once that is chosen we only have 200 valid possibilities for j, and once *that* is chosen, only 100 possibilities for k, so

$$|E| = 300 \times 200 \times 100 = 6 \times 10^6,$$

so

$$P(E) = \frac{6}{27} = 0.2222.$$

Observe that having 100 beans of each kind in the jar is irrelevant—as long as the number of each kind of bean is the same, then we will get the same result. (It is as though you rolled a 3-sided die three times and asked the probability of getting a different outcome on each roll.)

Now consider the problem in the form in which it was originally intended, using sampling without replacement. Here the sample space consists of all triples (i, j, k) of *distinct* integers from $\{1, \ldots, 300\}$, By the same reasoning we've been using up until now, this set has cardinality $300 \times 299 \times 298$, but the event E is exactly the same as what it was in the sampling-with-replacement version. So now

$$P(E) = 6 \times 10^6/(300 \times 299 \times 298) = 0.2245.$$

There's not much difference from the result we got for sampling with replacement, and if we had even more beans of each type, say ten thousand of each kind of bean, then the difference would be even smaller, agreeing to four decimal places. This is what you would expect: when there are so many beans and we are pulling out so few, it makes very little difference if we throw a bean back in the jar after we sample it.

4 NumPy `choice` function

The `choice` function in NumPy lets you produce a random sample from a list or an array of elements. There are options for both sampling with replacement and sampling without replacement The posted code illustrates its use by simulating the experiments described above.