Mathematics in the Real World
Math 107

Lectures 14 & 15: Independent events. Conditional probability.

Laws of probability. Disjoint events.

Example 1. Let us roll two dice. What are the chances that the sum of the outcomes is at most 3?
Solution: Note that the outcomes both have to be 1 or more, so the sum of the outcomes cannot be 1. It must be 2 or 3. How can it be 2? This happens via only one possible outcome: (1,1). How can it be 3? The dice rolled yield (1,2) or (2,1). Thus, our event is now $\{(1,1),(1,2),(2,1)\}$, and the total number of outcomes is $6 \times 6 = 36$ (since each die has 6 outcomes, two dice rolled independently yield 36 outcomes). Thus, the probability of getting 2 or 3 is P(2 or 3) = 3/36 = 1/12.

Here is another way of viewing the last example: as the probability of a union of two disjoint events. Two events are *disjoint* if there are no outcomes common to both events. In our example, the event of getting at most 3 is the *disjoint union* of two events: getting *exactly 2* and *exactly 3* as the sum of the two dice. In case of disjoint events, the probability of their union is the sum or their probabilities. The outcomes that yield 2 are simply (1,1), so the probability of getting 2 is 1/36. The outcomes that yield 3 are (1,2) and (2,1), so $P(3) = 2/36 = 1/18$. Adding them up,

$$P(2 \text{ or } 3) = P(2) + P(3) = \frac{1}{36} + \frac{2}{36} = \frac{1}{12}.$$

$\square$

Let us formalize this observation. Recall that we denoted by Ω the set of all outcomes of an experiment. An event is a subset of all outcomes: $E \subset \Omega$. Then the probability of an event E is the proportion of the outcomes in E to all outcomes:

$$P(E) = \frac{\#E}{\#\Omega}.$$

Below we formulate the basic *laws of probability:*

- $0 \leq P(E) \leq 1$ for all events $E \subset \Omega$. Moreover, $P(E) = 0$ if and only if E is the *empty set*, meaning that the event can never happen (such as getting heads *and* tails in a single coin toss). And $P(E) = 1$ if and only if $E = \Omega$.
- If $E \subset E'$ (e.g., E is getting a spade when you draw a card, and E' is getting a spade or a club), then $P(E) \leq P(E')$.

- If E and E' are disjoint, then their union $E \cup E'$ (meaning that an outcome in E *or* in E' is favorable) has the probability: $P(E \cup E') = P(E) + P(E')$.
- If E^c is the *complementary event* to E (i.e., the set of all outcomes that are not in E), then $P(E^c) = 1 - P(E)$.

Here are a couple of examples of how to apply the laws to compute probabilities.

Example 2. Suppose we roll two dice. What are the chances that the sum of the outcomes is at least 4?
Solution: We know from above that the chances of getting the sum to be at most 3, is 1/12. The complementary event to getting at most 3 (i.e., 2 or 3) is to get *more than 3* - i.e., at least 4. Thus,
P(at least 4) = 1 - P(at most 3) = 1-(1/12) = 11/12.

□

Example 3. What is the probability of drawing a 7 or a red card from a deck?
Solution: The probability of drawing a 7 from a deck of 52 is $P_1 = \frac{4}{52} = \frac{1}{13}$. The probability of drawing a red card, *but not* 7, is $P_2 = \frac{26-2}{52} = \frac{24}{52} = \frac{6}{13}$. The events are disjoint, therefore the probability of their union is $P = P_1 + P_2 = \frac{1}{13} + \frac{6}{13} = \frac{7}{13}$. Try adding the probability of drawing a red card and the probability of drawing a 7, but not red - you will get the same answer!

□

Independent events. Two events are *independent* if the occurrence of one has no influence on the probability of the other. For example, tosses of a fair coin are independent: independently of the result of the previous tosses, in each toss you get heads and tails with probability $\frac{1}{2}$.

Two events A and B are independent if and only if the probability of both of them happening (probability of their intersection) equals to the product of their probabilities:

$$P(A \cap B) = P(A) \cdot P(B).$$

Thus, the probability of 3 heads in 3 coin tosses is $\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$.

Example 4.
(1) What is the probability of getting (HHT) (in this order) in 3 tosses?
Solution: It is also $\frac{1}{8}$, since the probability of getting any specified outcome in each toss is $\frac{1}{2}$.
(2) What is the probability of getting exactly two heads in three tosses?
Solution: The desired outcomes are $(HHT), (HTH), (THH)$. There are $2^3 = 8$ possible outcomes. Therefore, the probability is $\frac{3}{8}$.
There is another way to compute this probability: the number of positive

outcomes equals to the number of choices of 2 out of 3 items - in this case, 2 out of 3 "places" where H will occur. This number is ${}_3C_2 = \frac{3!}{2!1!} = 3$. Dividing by the number of all possible outcomes, we get $\frac{3}{8}$.

(3) The use of combinations allows to compute a probability of a particular number of heads in a large number of tosses. For example, what is the probability of getting 20 heads in 40 tosses? Attention: it's not $\frac{1}{2}$! To find the number of favorable outcomes, compute

$$ {}_{40}C_{20} = \frac{40!}{20!20!} = 137846528820. $$

The total number of outcomes is 2^{40}. Therefore, the probability of getting exactly 20 heads in 40 coin tosses is

$$ P = \frac{{}_{40}C_{20}}{2^{40}} \simeq 0.125. $$

The reason this probability is so small is that we required a very precise result: the number of heads is equal to the number of tails. But it is just about as likely that they differ by one: 19 heads and 21 tails, or the opposite. You can check that a probability of getting anywhere between 18 and 22 heads in 40 tosses is greater than $\frac{1}{2}$. The formula you need is

$$ \frac{2 \cdot {}_{40}C_{18} + 2 \cdot {}_{40}C_{19} + {}_{40}C_{20}}{2^{40}}, $$

where we took into account that ${}_{40}C_{22} = {}_{40}C_{18}$ and ${}_{40}C_{21} = {}_{40}C_{19}$.

□

To check if two events A and B are independent, we have to compute the probabilities $P(A)$, $P(B)$ and $P(A \cap B)$ and check if the formula

$$ P(A \cap B) = P(A) \cdot P(B) $$

holds. Many random events are not independent.

Example 5. For example, suppose you draw two cards randomly from a deck. Let A denote the event that the first card is a spade, and B, that the second card is a spade. Are they independent?

Clearly, $P(A) = \frac{13}{52} = \frac{1}{4}$. What about $P(B)$, the probability of drawing a spade from the deck that is already missing a card? If the first card drawn was not a spade (probability $\frac{3}{4}$), then we have 13 spades and 51 total cards left; if the first card was a spade (probability $\frac{1}{4}$), then we have only 12 spades and still 51 cards left. These situations are disjoint (the first card is either a spade or not a spade). Therefore we have

$$ P(B) = \frac{3}{4} \cdot \frac{13}{51} + \frac{1}{4} \cdot \frac{12}{51} = \frac{39+12}{4 \cdot 51} = \frac{51}{4 \cdot 51} = \frac{1}{4}. $$

Now we compute $P(A \cap B)$. This is the event that both cards are spades, and its probability is:

$$P(A \cap B) = \frac{{}_{13}P_2}{{}_{52}P_2} = \frac{13 \times 12}{52 \times 51} = \frac{12}{4 \times 51} = \frac{1}{17}.$$

On the other hand,

$$P(A) \cdot P(B) = (1/4) \cdot (1/4) = 1/16.$$

Since this is not equal to $P(A \cap B)$, the events are not independent.

Example 6. Draw two cards randomly from a deck (without replacement). A is the event that the first card is a diamond ($\diamond$); B, that the second card is a queen (Q).
Solution: Compute the probabilities:

$$P(\diamond) = \frac{13}{52} = \frac{1}{4}. \quad P(Q) = \frac{4}{52} = \frac{1}{13}.$$

Note that just as in the example with the spades above, if we don't know what the first card was, the probability of the second being a queen is the same as the probability of drawing a queen from a full deck.
To count the number of outcomes with the first card being a diamond and the second a queen we have to consider two cases: (1) the first card is diamond and not a queen: 12 possibilities, and (2) the first card is the queen of diamonds: 1 possibility. In the first case, the second card can be any of the 4 queens; in the second, any of the 3 remaining queens. Totally, we have $12 \cdot 4 + 1 \cdot 3 = 51$ desired outcomes. There are ${}_{52}P_2 = 52 \cdot 51$ possible outcomes of drawing two cards from the deck without replacement. The probability of $A \cap B$ is

$$P(A \cap B) = \frac{51}{52 \cdot 51} = \frac{1}{52} = \frac{1}{4} \cdot \frac{1}{13} = P(A) \cdot P(B).$$

Therefore, the events A and B are independent.

$\square$

Conditional probability. What if the events A and B are not independent? Can we have a formula that relates the probabilities of A, B and $A \cap B$? Consider the following example. Suppose we draw a card, and we *already know* that this card is red. What are the odds that the card is a diamond?

When you compute probabilities, you are to divide the number of desired outcomes (in the event) by the number of *possible* outcomes. But now, given that the card drawn is red, the number of possible outcomes is reduced to only 26 of the 52 cards. And among them, the number of diamonds is 13. Thus, we compute:

P(diamond, *given* that the card drawn is red) $= 13/26 = 1/2$.

More generally, the *conditional probability* of A given B is the probability of an event A occurring, *given* that an event B has occurred. It is denoted by $P(A|B)$ and equals $P(A \cap B)/P(B)$. Here, $A \cap B$ is the *intersection* of A and B, the outcomes common to both A and B. In the above example, say A is getting a diamond and B is getting a red card. Then $P(A \cap B)$ is the probability that a card is red and a diamond, which means that it is a diamond, $P(A \cap B) = 1/4$. The probability of a card being red is $P(B) = 1/2$. Therefore, the probability of a card being a diamond, given that it is red, is:

$$P(A|B) = P(A \cap B)/P(B) = \frac{1/4}{1/2} = \frac{1}{2},$$

same answer as we obtained before.

In general, for any two events we have:

$$P(A \cap B) = P(B) \cdot P(A|B), \quad \text{and}$$
$$P(A \cap B) = P(A) \cdot P(B|A),$$

since the probability of intersection (A *and* B happening) is symmetric with respect to A and B, so if the first formula above is true, then the second should be also true.

Let us reconsider the example from above: A is drawing a spade; B is drawing another spade (without replacement). Then $(A \cap B)$ is drawing two spades in a row and we computed its probability to be equal to $\frac{1}{17}$. Now we can calculate this number using conditional probabilities. Clearly $P(A) = \frac{1}{4}$. The conditional probability $P(B|A)$ is easy to compute: if we know that the first card was a spade, then there are only 12 choices left for the second card to be a spade, out of 51 cards total:

$$P(B|A) = \frac{12}{51} = \frac{4}{17}, \qquad \text{and } \; P(A \cap B) = P(A) \cdot P(B|A) = \frac{1}{4} \cdot \frac{4}{17} = \frac{1}{17}.$$

Bayes' law of conditional probabilities. Comparing the two formulas for the joint probability of two events,

$$P(A \cap B) = P(B) \cdot P(A|B), \quad \text{and}$$
$$P(A \cap B) = P(A) \cdot P(B|A),$$

we can conclude:

$$P(A) \cdot P(B|A) = P(B) \cdot P(A|B).$$

This can be rewritten as

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)}.$$

This is the *Bayes' law* of conditional probabilities that holds for any two events A and B with positive probabilities.

Note: If A and B are independent, then $P(B|A) = P(B)$ and $P(A|B) = P(A)$ and the Bayes' law is trivially true.

Example 7. Rolling three dice, what is the probability of getting a 3 at the first roll, if the sum of the three outcomes is 6?
Solution: Let A be getting the sum of three outcomes equal to 6, and B be getting a 3 on the first roll. We need to find $P(B|A)$. By definition,

$$P(B|A) = \frac{P(A \cap B)}{P(A)}.$$

The outcomes of A are $(2, 2, 2), (1, 1, 4)(4, 1, 1), (1, 4, 1)$, and the six permutations of $(1, 2, 3)$, totally 10 outcomes. (Why six? The number of permutations of 3 elements is ${}_3P_3 = 3! = 6$.) Then we have $P(A) = 10/6^3 = 5/108$. The outcomes of $(A \cap B)$ are $(3, 2, 1)$ and $(3, 1, 2)$, so $P(A \cap B) = 2/6^3 = 1/108$. Then

$$P(B|A) = \frac{1/108}{5/108} = \frac{1}{5}.$$

We could have used the Bayes' law to compute the same probability:

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)}.$$

Then we need to know $P(B)$ and $P(A|B)$. The first is easy: the probability of getting a 3 in one roll of a die is $P(B) = 1/6$. The second it the probability of getting the sum or three outcomes equal to 6 given that the first outcome is 3. This is the same as the probability of getting the sum of two outcomes in two rolls equal to 3. There are two possibilities: $(1, 2)$ and $(2, 1)$. Therefore, $P(A|B) = 2/36 = 1/18$. The Bayes' law gives

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)} = \frac{\frac{1}{18} \cdot \frac{1}{6}}{\frac{5}{108}} = \frac{1}{5}.$$

□

The Bayes' Law comes useful when you want to find a conditional probability $P(B|A)$ and the reverse conditional probability $P(A|B)$ is given, or easy to find. Here is an example.

Example 8. Suppose you have box of chocolates with three types of fillings: cherry, almond and coffee. You know that 1/4 of all chocolates have cherry filling, 1/4 coffee and 1/2 almond. The chocolates are wrapped in colorful foil according to the following rule: (1) all cherry-filled chocolates are wrapped in red foil; (2) of coffee-filled chocolates, 1/2 are wrapped in yellow and 1/2 in brown; and (3) of almond-filled chocolates, 1/3 are wrapped in red, 1/3 in yellow and 1/3 in blue. Given that a chocolate wrapped in red, what are the chances that it has cherry filling?

Solution: This is a classic example where the Bayes' Law works best. Denote by A the probability that a chocolate is wrapped in red; and by B the probability that it has cherry filling. We need to find $P(B|A)$. The reverse probability $P(A|B)$ is given: if a chocolate holds cherry filling, it is always red, so $P(A|B) = 1$. The probability

$P(B)$ is also given: 1/4 of the chocolates have cherry filling, therefore $P(B) = 1/4$. Let us compute $P(A) = P(\text{wrapped in red})$:

$$P(A) = \frac{1}{4} \cdot 1 + \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{4} + \frac{1}{6} = \frac{5}{12}.$$

Here the first summand accounts for the cherry-filled chocolates wrapped in red and the second summand accounts for the almond-filled chocolates wrapped in red. Finally, the Bayes' law gives:

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)} = \frac{1 \cdot \frac{1}{4}}{\frac{5}{12}} = \frac{3}{5}.$$

$\square$

Disease detection and drug testing. An important application of Bayes' law is in assessing the results of various medical tests. In the next example, we will use the Bayes' Law to solve problems involving *false positives* in disease detection or drug testing.

Here is an example: suppose a test to detect a rare disease claims to detect people with the disease nine times out of ten, and for people without the disease, it correctly says they do not have it, 95% of the time. Suppose these figures were obtained by having tested a population of 1000 subjects, of whom only 100 have the disease (this was verified independently of the test). This is the information we have on the reliability of the test. Let us see how it translates into practice.

Suppose a subject undergoes the test, and the result is positive. What are the chances that the subject does have the disease? It is tempting to assume that the chances are 90%, because the test said it was 90% reliable on sick patients.

In fact, the probability of the subject having the disease is much lower. To see this, let us name specific events: let $D, N, +, -$ denote the events that the subject has the disease, does not have the disease, is diagnosed to have the disease (by the test), and is diagnosed to not have the disease, respectively. (Note that the diagnosis being positive is one thing, and suffering from the disease is a completely different event.)

Now what can we say from the above data? It allows us to form certain conclusions. For example, given that a patient has the disease, (s)he is diagnosed to have it nine times out of ten. Thus,

$$P(+|D) = 0.9.$$

Similarly, $P(-|D) = 1 - 0.9 = 0.1$, and $P(D) = 100/1000 = 0.1$. How about if the patient does not have the disease? Then we have: $P(N) = 0.9$ and $P(-|N) = 0.95$, which means that $P(+|N) = 0.05$.

The question that we would like to solve is: what is $P(D|+)$? This is different from the 90% claimed by the test, which was $P(+|D)$.

To compute our desired probability, we now use Bayes's Rule, and the above information:

$$P(D|+) = \frac{P(+|D) \cdot P(D)}{P(+)} = \frac{0.9 \times 0.1}{P(+)} = \frac{0.09}{P(+)}.$$

What is $P(+)$? The number of users who are *diagnosed* to have the disease is: 90% of the 100 users who actually have the disease, and 5% of the 900 people who do not have the disease. Computing the proportion out of a thousand people,

$$P(+) = P(+|D) \cdot P(D) + P(+|N) \cdot P(N) = 0.9 \cdot \frac{100}{1000} + 0.05 \cdot \frac{900}{1000} = 0.09 + 0.045 = 0.135.$$

Hence $P(D|+) = 0.09/0.135 = 2/3$! What this means is, the chances are not as high as 9 in 10 that the subject has the disease; it's closer to 67%, or 2 in 3.

Similarly, one can have examples where even fewer subjects in the "tested group" had the disease to start with - and in that case, it may be that $P(D|+) < 0.5$ - which would mean that a "+" means that it is *more* likely that the patient does *not* have the disease.

The moral of the story: do not confuse one conditional probability for the reverse conditional probability; rather, compute one from the other using Bayes' Law.

Example 9. Suppose you have two tests for disease D: Test A claims to diagnose correctly both diseased and healthy people in 99% of cases and was carried out in a sample of subjects where only 1% had the disease. Test B claims to diagnose correctly both diseased and healthy people in 95% of cases and was carried out in a sample of subjects where only 5% had the disease. Which test is more reliable?

Solution: Let us denote as above having the disease by D and not having it by N; testing positive by $+$ and testing negative by $-$. We will use indices A and B for probabilities obtained for test A and B, respectively. To compare reliability, we need to compare two conditional probabilities: $P_A(D|+)$ and $P_B(D|+)$.

We have $P_A(+|D) = 99/100, P_A(+|N) = 1/100$ and $P_A(D) = 1/100$. To compute $P_A(D|+)$ by the Bayes' law, we need $P_A(+)$. Proceeding as before, we have

$$P_A(+) = P_A(+|D) \cdot P_A(D) + P_A(+|N) \cdot P_A(N) = \frac{99}{100} \cdot \frac{1}{100} + \frac{1}{100} \cdot \frac{99}{100}.$$

Then by the Bayes' law we compute the probability that a person has the disease if they test positively using test A:

$$P_A(D|+) = \frac{P_A(+|D) \cdot P_A(D)}{P_A(+)} = \frac{\frac{99}{100} \cdot \frac{1}{100}}{\frac{99}{100} \cdot \frac{1}{100} + \frac{1}{100} \cdot \frac{99}{100}} = \frac{1}{2}.$$

Does it get any better with test B? We have $P_B(+|D) = 95/100, P_B(+|N) = 5/100$ and $P_B(D) = 5/100$. To compute $P_B(D|+)$ by the Bayes' law, we need $P_B(+)$. We have

$$P_B(+) = P_B(+|D) \cdot P_B(D) + P_B(+|N) \cdot P_B(N) = \frac{95}{100} \cdot \frac{5}{100} + \frac{5}{100} \cdot \frac{95}{100}.$$

Then by the Bayes' law we compute the probability that a person has the disease if they test positively using test B:

$$P_B(D|+) = \frac{P_B(+|D) \cdot P_B(D)}{P_B(+)} = \frac{\frac{95}{100} \cdot \frac{5}{100}}{\frac{95}{100} \cdot \frac{5}{100} + \frac{5}{100} \cdot \frac{95}{100}} = \frac{1}{2}.$$

The tests are equally (un)reliable.

In general, to have a reliable test for a disease or a drug, you need to get high $P(+|D)$ probability on a *large* sample of diseased subjects (and of course a small number of mistakes $P(+|N)$ on healthy people), which would ensure that

$$P(+|D) \cdot P(D) > P(+|N) \cdot P(N).$$