Probability

I. Why do we need to look at this?

- probability is fundamental to statistics.

- briefly, we use statistics to figure out how “probable” something is. In other words, if we observe something (an event of some kind), we can use these methods to figure out what the probability of that event is. Then, if that probability is very low, we tend to think something is going on.

- for example:

  - flip a coin 10 times:

    - 5 heads, 5 tails
    - 4 heads, 6 tails
    - 3 heads, 7 tails
    - 2 heads, 8 tails
    - 1 head, 9 tails
    - 0 heads, 10 tails

- you’d probably think the coin was unfair if you got one of the results near the bottom of the list. In other words, the probability of getting 0 heads and 10 tails is very low, and you get suspicious.

  (not convinced? suppose I give you a dollar for every head, and you give me a dollar for every tail, and we flip the coin 10 times (my coin!) and get the result at the bottom. You think I’m being honest??)

- Another example:

  - we try out a new medicine for blood pressure. In 85% of the people, we see a reduction in blood pressure. Is this a random event, or did the medicine do something? (Better: what is the probability the medicine did something?)

- We’ll learn how to evaluate something like this soon, but for now we need to look a little bit at probability so that we can understand “why” statistical estimates and tests work.

II. Some very basic probability.

- We won’t learn a lot of probability here. Just enough so that you understand a couple of basics.
- Probability: a number that describes the chance of an event happening. We note probability by:

\[ \Pr\{E\} \] (sometimes also just \( P\{E\} \))

- this means “the probability of the event ‘\( E \)’“

- so what is \( E \)?

- some examples:

- \( E_1 \): Heads, when tossing a coin
- \( E_2 \): the # 6, when rolling a die
- \( E_3 \): the # 7, when rolling two dice
- \( E_4 \): the Orioles (or Nationals) winning the World Series

so we have:

- \( \Pr\{E_1\} = 1/2 \)
- \( \Pr\{E_2\} = 1/6 \)
- \( \Pr\{E_3\} = 1/6 \) (more on this soon)
- \( \Pr\{E_4\} = ??? \) probably very close to zero.

- see also section 3.3 in your text.

- \( \Pr\{E\} \) is always between 0 and 1.

- Your text is a little strange on the top of p. 90 \{80\} \{86\} \{85\}. Let’s think about things this way instead:

\[
\Pr\{E\} = \frac{\text{the number of ways the event can occur}}{\text{the number of possible outcomes}}
\]

so for \( E_1 \) above,

\( \Pr\{E_1\} = 1/2 \) because there is only 1 way of getting a head, but two possible outcomes (heads or tails).

\( \Pr\{E_2\} = 1/6 \) again, because there is only 1 way of getting a 6, but there are six possible outcomes (the #’s 1-6).

\( \Pr\{E_3\} = 1/6 \) this one is trickier. First, figure out how many different ways you can get 7:
So there are six different ways you can get a 7. How many possible outcomes are there? If you make up a list like above, you’ll get 36 different ways of adding up the numbers (try it). So we have 6/36, which simplifies to 1/6.

\[ \Pr\{E_4\} = 0? \]

I won’t even try to illustrate this (I can’t), but this one would be difficult to calculate. In any case, it seems pretty obvious that neither the Orioles or the Nationals will win the world series (think bout it this way - the numerator would be 0).

Some more examples:

For a single die, what is the probability of getting a 2, 3, or 4?

- there are three possible ways to get what we want (2, 3, or 4), so our numerator is 3.

- there are six possible outcomes (the numbers 1-6).

so we have \( \Pr\{2 \leq Y \leq 4\} = 3/6 = 1/2 \). See also your text on page 92 - 93 [96 - 97] for this and other examples. Note the different way of writing this event; \( E = 2 \leq Y \leq 4 \). Remember that \( Y \) signifies our random variable.

III. Combining probabilities:

- again, we’re just scratching the surface here.

- Suppose you wanted to figure out the probability of getting a 6 and heads (you’re rolling one die and tossing one coin).

- Well, you can count up all the possible outcomes, and do it that way, or you can multiply some probabilities. For instance:
Pr\{Y_1 = 6\} = 1/6 \text{ (see above)}
Pr\{Y_2 = \text{heads}\} = 1/2 \text{ (see above)}

so \Pr\{Y_1 = 6 \text{ and } Y_2 = \text{heads}\} = 1/6 \times 1/2 = 1/12.

- Your book uses “probability trees”, which are neat but can get out of hand real fast (try doing three dice as in the example above).

[illustrate above example using probability tree]

- one important point - the two (or three or however many) events can not influence each other, or this method doesn’t work. Or is statistical language, they must be INDEPENDENT.

(for example, if you have “somehow” connected your coin and die so that whenever you roll a 6 you ALWAYS get heads, then the above doesn’t work)

- Example 3.19 [3.16] in text:

- Newborns sometimes suffer from a serious respiratory ailment known as “hypoxic respiratory failure. A possible alternative treatment to surgery is Nitric Oxide.

- Does Nitric Oxide work?

- newborns with the disease were assigned into two equal sized groups. One group was given nitric oxide, the other was not.

- in the group given nitric oxide, 54.4% did not need surgery. In the control group, only 36.4% did not need surgery.

- so what is the overall probability of needing surgery?

- in the control group, 1-.364 = .636 or 63.6% needed surgery, in the treatment group 1-.544 = .456 or 45.6% needed surgery. So the overall probability of needing surgery becomes:

\[ \Pr\{\text{surgery}\} = .5 \times .456 + .5 \times .636 = .228 + .318 = .546, \text{ or } 54.6\%

(Caution: if you have the 2nd edition, there's a slight math mistake here)

- note that we multiply the different probabilities (i.e. the probability of being in one or the other group by the probability of needing surgery) and
then add up the outcomes we’re interested in.

- (illustrate with probability tree)

Conditional probability:

When we talk about conditional probability, what we’re interested in is “what is the probability that Event A happens GIVEN that Event B has happened.

For example:

\[ \Pr\{\text{rolling an 8 with 2 dice GIVEN that one of the two die shows a 3}\} \]

We usually write this like this:

\[ \Pr\{\text{rolling an 8 with 2 dice | that one of the two die shows a 3}\} \]

where the vertical bar ("|") means “given”.

So lets solve the above problem:

- one die shows a 3
- this implies the other die MUST come up 5.
- so what is the \( \Pr\{\text{rolling a 5}\} \)? Easy, just 1/6.

But notice the following (one way of looking at independence):

\[ \Pr\{\text{rolling an 8 with 2 dice }\} = \frac{5}{36} \]

(you can do the math yourself, just add up all the possible ways of getting an 8 with 2 dice and divide by 36)

This is NOT the same answer we got above. This, incidentally, tells us that the events “rolling an 8 with 2 dice” and “one of the two die shows a 3” are not independent.

A good way of thinking about conditional probability is to "redefine" your universe:

For example, in the above dice rolling experiment, we are no longer interested in all possible outcomes with two dice, JUST THOSE in which the first die you've rolled shows a 3. So you only look at those results, and ignore all others (you're not interested in ANYTHING if the first die is a 2).
Now, finally, let's do the examples in the book on p. 98 and 99 [86 & 87] [92 & 93] [91 & 92].

First, let's list some probabilities (see Fig. 3.6 [3.2.5] [3.2.5]):

\[
\begin{align*}
\Pr\{\text{having disease and (+) test} \} &= .076 \\
\Pr\{\text{having disease and (-) test} \} &= .004 \\
\Pr\{\text{not having disease and (+) test} \} &= .092 \\
\Pr\{\text{not having disease and (-) test} \} &= .828
\end{align*}
\]

For example 3.21 [3.18] [3.2.11] [3.2.11] what we really want is:

\[
\Pr\{\text{having disease | (+) test}\}
\]

So we get \( \Pr\{ (+) \text{ test} \} = .076 + .092 = .168 \)

(just adding up everything above with a (+) test. This is our new universe - we’re no longer interested in anything with a (-) test)

Then we get \( \Pr\{\text{having disease and (+) test} \} = .076 \)

(this is straight from above)

Finally, we divide .076/.168 = .452

So we can say that

\( \Pr\{\text{having disease | (+) test}\} = .452 \)

Which is pretty awful. It means that even if you have a positive test, chances are you DON’T have the disease.

Let’s take the example one step further and figure out:

\[
\begin{align*}
\Pr\{\text{no disease | (-) test} \}
\end{align*}
\]

very quickly now:

\[
.004 + .828 = .832
\]

so \( \Pr\{\text{no disease | (-) test} \} = .828/.832 = .995 \)

This is much better. If you have a negative test, you almost
certainly don’t have the disease.

(Note that your text doesn't use conditional probability for this problem, which makes it a bit confusing following it)

IV. Getting back to coins:

- Let’s return to our problem from right at the beginning, flipping a coin 10 times.

- So what IS the probability of getting 8 tails if the coin is really fair?

- let’s figure out the number of possible outcomes:

- all possible ways of getting 0 tails: HHHHHHHHHH,

- all possible ways of getting 1 tail: THHHHHHHHH,
  HTHHHHHHHH,
  HHTHHHHHHH,
  etc.

- then you need to list all possible ways of getting 2 tails, then 3 tails, and so forth. You could do it this way, but you’d be at it for a very long time.

- Let’s consider this:

In a situation like this where we have a bunch of trials (tosses), we can figure out how many different possible ways we can get what we want by using something called the binomial coefficient:

\[
\binom{n}{j} = \binom{n}{j} = \frac{n!}{j!(n-j)!}
\]

the symbol ! means “factorial”. This is defined as follows (for any positive integer):

\[x! = x(x-1)(x-2)...(2)(1)\]

so, for example, \(3! = 3 \times 2 \times 1 = 6\), and \(5! = 5 \times 4 \times 3 \times 2 \times 1 = 120\)

- note that this gets big FAST (try 10! or 20!)

- also, \(0! = 1\) (you can actually show this, even though it doesn’t seem reasonable, but we won’t do it in here).

- so what does all the other stuff mean?
\[ n = \text{the number of trials} \]
\[ k = \text{the number of “successes”} \]

- so we have:

\[
\binom{10}{8} = \frac{10!}{8!(10-8)!}
\]

(demonstrate how to simplify this.)

- So now we know that there are 45 different ways of getting 8 tails if we toss a coin 10 times.

- we could now go through and calculate the number of different ways of getting 0 tails, 1 tail, 2 tails, etc. and then divide 45 by the total number of possible outcomes. But this is still tedious (it DOES work, though, if you want to try it).

- So let’s multiply some probabilities instead:

The probability of getting a tail = .5, so if we want 8 tails, we can multiply:

\[ .5 \times .5 \times .5 \times .5 \times .5 \times .5 = .5^8 = 0.00390625 \]

But if we have 10 trials, that means we have 2 heads, so we multiply that in as well:

\[ .5 \times .5 = .5^2 = .25 \]

Now if we multiply everything together, we get:

\[ 0.00390625 \times 0.25 = 0.0009765625 \]

(Remember, what we have now is the Probability of one way of getting 8 heads and 2 tails)

BUT there are 45 ways of getting this, so we have:

\[ 45 \times 0.00097625625 = 0.0439 \]

So, finally, the probability of getting 8 heads in 10 tosses is 0.0439

(Comment: we’ll talk about significant figures soon, for now just round stuff off to something that seems reasonable, but don’t do it until you’re all done with whatever you’re trying to calculate)
This is an application of the binomial distribution formula, which is given as follows:

\[
\binom{n}{j} p^j q^{n-j}
\]

where \( p \) = probability of success
\( q = 1 - p \) = probability of failure
\( n \) = the number of trials
\( j \) = the number of successes

in the simple example above,

\( p = \) probability of tail = .5  
\( q = 1 - p = \) probability of head = .5  
\( n = \) the number of tosses = 10  
\( j = \) the number of successes = 8

so we have, (much easier now):

\[
P(8 \text{ tails in 10 tosses}) = \binom{10}{8} .5^8 .5^2 = 0.0439
\]

Let’s finish this by going through example 3.29 in your text:

- we are sampling 5 individuals from a large population (basically, sampling with replacement; we’ll talk more about this later)

- 39% of the individuals in the population are mutants.

- so the probability of getting three mutants is:

\[
P(3 \text{ mutants in a sample of 5}) = \binom{5}{3} .39^3 .61^2 = 10 - 0.022 = .22
\]

- your book lists all possible combinations for a sample of 5 in table 3.5. You should look at this and make sure you know how they got the results.

- your book also lists the binomial coefficient in table 2 of the appendix. To use it, figure out \( n \) and \( j \), then look down the row for \( n \), and the column for \( j \) (if \( n = 5 \) and \( j = 3 \), the table gives 10, which is what was used above.

Note that your text presents the binomial distribution in a slightly different way than I do (there’s nothing wrong with it, it’s just different).