Probability Lecture #1 - Written Notes

Most simple view of a probability question is that we have some options that are equally likely (sample space), and some of those options are successes, and then probability of success = (# of successful options)/(total # of options)

Roll a standard 6 sided die…what's the probability you will get a 6? $\frac{1}{6}$.

What's the probability of rolling a prime number? $\frac{3}{6} = \frac{1}{2}$, since 2,3, 5 are prime numbers.

Class has 10 boys, 12 girls. A committee of 3 students is chosen? What is the probability that the committee is all girls?

Sample Space = $\binom{22}{3}$, # of possible committees
Successes = $\binom{12}{3}$

Given probability is $\frac{\binom{12}{3}}{\binom{22}{3}} = \frac{12\times11\times10}{22\times21\times20} = \frac{1}{7}$.

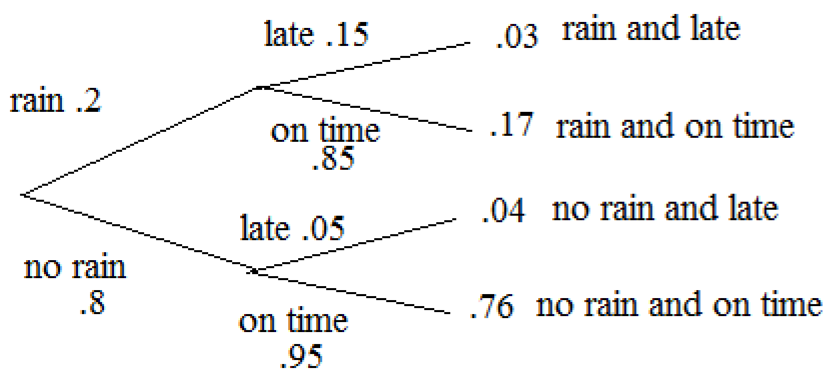
A lot probability boils down to two counting questions - 1 for the sample space and one for the # of success.

Probability Trees

We can think about events occurring in sequence, and if we know something has happened, then we adjust the probability (formally this is known conditional probabilities).

Rains 20% of the time.
If it rains, Maya is late to school 15% of the time.
If it does not rain, Maya is late to school 5% of the time.

Make tree probability diagram that encapsulates this situation:



What is the probability on a randomly chosen day that Maya is late?

P(late) = p(rain and late) + p(no rain and late) = .03 + .04 = 7%

General rule:
If p(A and B) = 0, then p(A or B) = p(A) + p(B).

In ALL situations the Inclusion-Exclusion Principle Applies:

P(A or B) = p(A) + p(B) - p(A and B)

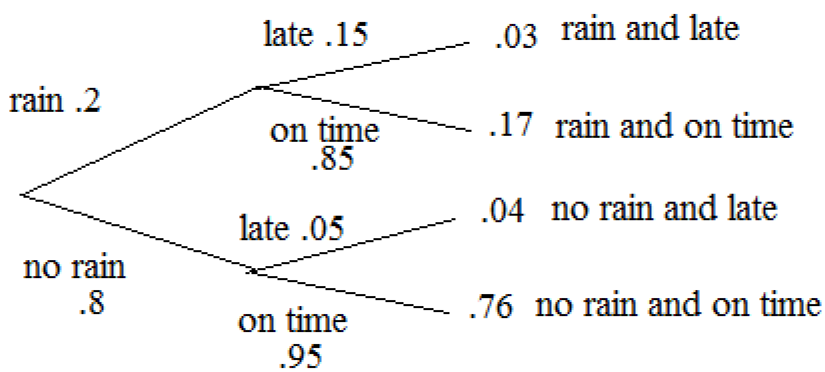
A conditional probability is one that is based on something already happening. We denote it as follows:

P(A | B) is read as "probability of event A occurring given that event B has occurred."

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

Given that Maya is late on a randomly chosen day, what is the probability that it was raining?

$$p(raining|late) = \frac{p(rain \cap late)}{p(late)} = \frac{.03}{.07} = \frac{3}{7}$$



An intuitive way of thinking about this chart is:

Out of 100 days, it rained on 20 and didn't rain on 80. In those 20 rainy days, Maya was late on 3 of them (15%) and on time on 17 of them (85%). On those 80 non-rainy days, Maya was late for 4 of them and on time for 76 of them. It's clear she was late 7 days, and of those 7 days it was raining 3 of them, hence the answer 3/7.

Monty Hall

Game show…3 doors…car behind one door, goats behind the other two.

Initially the contestant picks a door.
Host will reveal that behind a different door there is a goat.

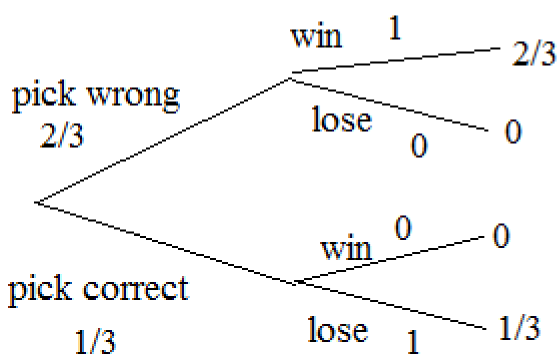
Then the host gives the contestant the choice if she wants to keep her door, or switch to the third door which wasn't revealed.

Question: what should our contestant do? Should she switch?

Given that she switches, what is her probability of winning?
Given that she stays, what is her probability of winning?

We go in assuming that we'll switch…
Our chance of picking the wrong door is 2/3.
Then, the host, as long as I picked a wrong door, is **<u>FORCED</u>** to reveal the other door. So, that means the third remaining door is guaranteed to have the car…so as long as I switch, I win.

If you stay, you only had a 1/3 chacne to pick the right door when you began. If you did pick the right door, the host can reveal either of the two remaining doors.



Switching probability for Monty Hall Game

Recitation Question

Instead of 3 doors, you have n doors.
Instead of revealing one goat, the host reveals k goats.
Then you want to switch, what is your probability of winning.

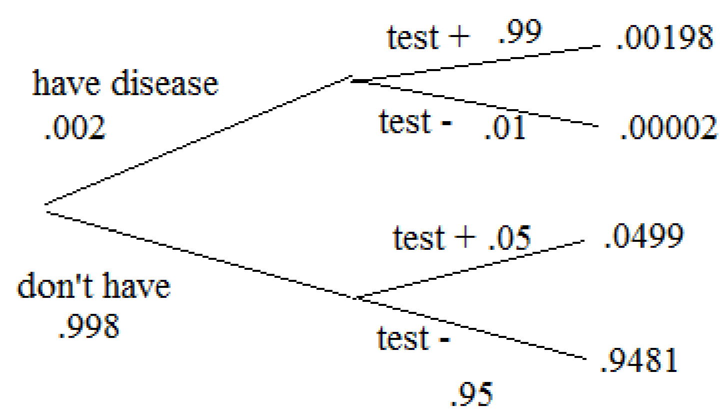
One More Conditional Probability Question

Disease .2% of the population has it.

There's a test for the disease.
Given that a person has the disease, the test correctly identifies this 99% of the time.

Given that a person does NOT have the disease, the test correctly identifies this 95% of the time.

You have just tested positive for the disease, what is the probability that you actually have it?



P(test+) = .00198 + .0499 = .05188
P(disease|test+) = p(disease and test+)/p(test+) = .00198/.05188 = 198/5188 ~ .03816

The answer to the question is less than 4%.

Key Lesson: If something is rare, then it's easy for the false positives to dilute the signal. So, if something is rare, you really need a test that reduces the chance of false positives!!!

If your test doesn't have false positives, then you won't unnecessarily scare or put someone through unnecessary treatment.

If your test has lots of false negatives, then you'll forget to treat people who have the disease!!!

Two more things I want to get to:

1. Rolling Multiple Dice
2. Lottery Example

<u>Pair of Dice</u>
Sum is in between 2 and 12, but each of these sums isn't equally likely.

Try it for yourself…roll a pair of dice a bunch and see if you get 7 more often or 2 as the sum.

(x, y)

Sample space is all ordered pairs of the form (x, y), where 1 <= x,y <= 6

There are 36 surch ordered pairs.

R1/r2	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

P(2) = 1/36, P(7) = 6/36 = 1/6

A simple intuitive way to see this is as follows:

If I want to get a 7, no matter what I roll on my first die, I always have a single target value I want to hit…and I have a 1/6 chacne of getting it.

1→6
2→5
3→4
4→3
5→2
6→1

This is true for no other #. (Example, if I am going for a 6 and I roll a 6 on my first die, I can't get a sum of 6…If I aim for 10 and roll a 1 on the first die, I can't get 10.)

<u>Lottery Problem</u>
N = 52 numbers you can choose from
You choose k = 6.

Prize matching 3, 4 5 or 6…

Sample Space = $\binom{52}{6}$.

# of ways of matching exactly 3 numbers = $\binom{6}{3}\binom{46}{3}$
P(matching 3) = $\binom{6}{3}\binom{46}{3} / \binom{52}{6}$

In general, in a lottery of n numbers, where you must select k of them, the probability of matching exactly m numbers is:

$$\binom{k}{m}\binom{n-k}{k-m} / \binom{n}{k}$$

Let's say the wining combo is 3, 12, 29, 31, 44, 51
And you picked let's say, 2, 8, 29, 33, 44 and 51, you matched 3 numbers…

One quick note: this is identical to our committee problem that I went over previously about committees with some number of men and women…

Next time:

1. Craps
2. Discrete Random Variables
3. Start doing some problems!