BAYES
THEOREM
EXAMPLES

A Visual
Guide
For Beginners

SCOTT HARTSHORN

Bayes Theorem Examples

A Visual Guide For Beginners

By Scott Hartshorn

Thank You!

Thank you for getting this book! This book contains examples of different probability problems worked using Bayes Theorem. It is intended to be direct and to give easy to follow example problems that you can duplicate, without getting bogged down in a lot of theory or specific probability functions.

Most of the problems have been solved using Excel, which is a useful tool for these types of probability problems.

If you want to help us produce more material like this, **then please leave a positive review for this book on Amazon**.   It really does make a difference!

Your Free Gift

As a way of saying thank you for your purchase, I'm offering this free Bayes Theorem cheat sheet that's exclusive to my readers.

This cheat sheet contains information about the Bayes Theorem and key terminology, 6 easy steps to solve a Bayes Theorem Problem, and an example to follow.   This is a PDF document that I encourage you to print, save, and share. You can download it by going here

http://www.fairlynerdy.com/bayes-theorem-cheat-sheets/

Bayes Theorem Overview

Bayes theorem describes the probability of an event based on other information that might be relevant. Essentially, you are estimating a probability, but then updating that estimate based on other things that you know.

This is something that you already do every day in real life. For instance, if your friend is supposed to pick you up to go out to dinner, you might have a mental estimate of if she will be on time, be 15 minutes late, or be a half hour late. That would be your starting probability. If you then look outside and see that there are 8 inches of new snow on the ground, you would update your probabilities to account for the new data.

Bayes theorem is a formal way of doing that.

This book is designed to give you an intuitive understanding of how to use Bayes Theorem. It starts with the definition of what Bayes Theorem is, but the focus of the book is on providing examples that you can follow and duplicate. Most of the examples are calculated in Excel, which is useful for updating probability if you have dozens or hundreds of data points to roll in. You can download all the Excel examples, for Free, here
http://www.fairlynerdy.com/bayes-theorem-examples/

Table of Contents

Bayes Theorem Definition & Intuitive Explanation
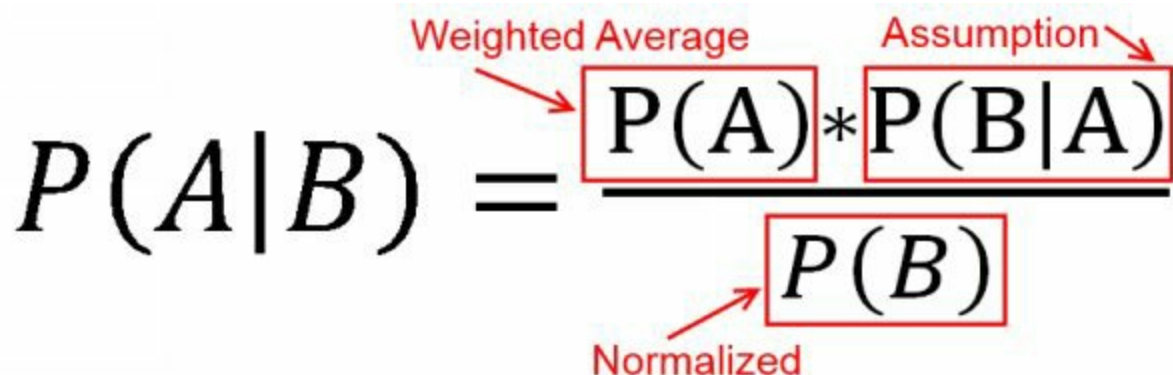
The equation for Bayes Theorem is

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

Where
- A & B are events
- P(A) and P(B) are the probabilities of A and B without regard for each other
- P(A|B) is the conditional probability, the probability of A given that B is true
- P(B|A) is the probability of B given that A is true

The equation is somewhat complicated, but using the equation really isn't. What it ends up being is just a normalized weighted average, given an assumption.   One of the goals of this book is to develop your intuition for these problems, so we will be discussing many of the problems from the normalized weight average perspective.

To solve all of these problems we will follow these steps

1. Determine what we want the probability of, and what we are observing
2. Estimate initial probabilities for all of the possible answers
3. For each of the initial possible answers, assume it is true and calculate the probability of getting our observation with that possibility being true
4. Multiply the initial probabilities (Step 2) by the probabilities based on our observation (Step 3) for each of the initial possible answers
5. Normalize the results (divide the each probability by the sum of the total probabilities so that the new total probability is 1)
6. Repeat Steps 2-5 over and over for each new observation

Example 1 - Simple Example With Dice

Suppose that your friend has 3 dice. One has 4 sides, one has 6 sides, and one has 8 sides. He draws one die at random, rolls it one time without showing you, and reports the result as having rolled a 2. How would you calculate the probability that the die was the 4 sided die, the probability for the 6 sided die, and the probability for the 8 sided die

Step 1 – Determine desired probability, and what is being observed

What we want is the probability of three different possibilities, 4 sided, 6 sided, or 8 sided. What we observe is the number that is rolled.

Step 2 – Estimate initial probabilities

Since we started with 3 dice, and we think the die that was rolled was randomly chosen, we will assume that the initial probability that the die was 4 sided is 1/3, the initial probability that the die was 6 sided is 1/3, and the initial probability that the die was 8 sided is 1/3. Essentially, this step is the same as if our friend had pulled the die out of the bag, held it in a closed hand without rolling it, and we had to bet on which die it was with no other information.

Step 3 – For each initial possibility, calculate the chance of getting our observation

Our observation was that a 2 was rolled. If we assume that our friend had the 4 sided die, the odds that die would roll a 2 is 1/4. If we assume that our friend had the 6 sided die, the odds that die would roll a 2 is 1/6. If we assume that our friend had the 8 sided die, the odds that die would roll a 2 is 1/8

Step 4 – Get the weighted average, multiply initial probabilities by each

probability based on the assumption.

- For the 4 sided die, initially it had a 1/3 chance and there was a 1/4 chance it would roll a 2, so the total probability our friend would draw the 4 sided die, and roll a 2 is 1 / 12
- For the 6 sided die, initially it had a 1/3 chance and there was a 1/6 chance it would roll a 2, so the total probability our friend would draw the 6 sided die, and roll a 2 is 1 / 18
- For the 8 sided die, initially it had a 1/3 chance and there was a 1/8 change it would roll a 2, so the total probability that our friend would draw the 8 sided die, and roll a 2 is 1 / 24

Step 5 – Normalize the Results

The total probability of the picking one of the 3 dice, and rolling a 2 is 1/12 + 1/18 + 1/24 = 13 / 72.   So we know that the odds that we selected any of the dice and rolled a 2 are 13 / 72. That number is less than 1 because the odds of rolling a 2 are less than 1.   However we know that we have already rolled the 2. So we need to divide the odds of each die by 13 / 72 so that sum of all the odds for all the die are 1. That process is known as Normalizing.

Normalizing each of the dice we find the probability that that die is the one we selected

- 4 sided die = (1/12) / (13/72) = (1 * 72) / (12 * 13) =   6/13
- 6 sided die = (1/18) / (13/72) = (1 * 72) / (18 * 13) =   4/13
- 8 sided die = (1/24) / (13/72) = (1 * 72) / (24 * 13) =   3/13

And that's the answer.

When we started the problem, we assumed that each die had a 33.3% chance that it was the one selected from the bag. After rolling 1 number, and rolling a 2, we determined that the odds the 4 die was the one selected was 46.1%, the odds the 6 die was the one selected was 30.8%, and the odds the 8 die was the one selected was 23.1%.

If we had more die rolls we could incorporate those, using the new percentages as our starting percentages and really refine the numbers

If you have a single data point of information to update your initial assumptions, then putting everything in one horizontal table is a concise format that works really well

Step 1	Step 2	Step 3	Step 4	Step 5
Die	Initial Probabilities	Chance of Roll	New Probability	Normalized Final Result
4	1/3	1/4	1/12	6/13
6	1/3	1/6	1/18	4/13
8	1/3	1/8	1/24	3/13
		Total Probability	13/72	

Some things to take note of on this example

There are a few points that we didn't hit in this example, but are important to know

- When we calculated the chance of getting a roll for each die, the odds were 1 divided by the number on the die. But that is true only because we rolled a 2, which is a number that could be rolled by any of the dice. The true odds for each roll are 1 divided by the number of the die **IF** the number of the roll is less than or equal to the number on the die. Otherwise the odds for that die are zero. For instance, if we had rolled a 7, then step 3 would show the odds of getting that result with the 4 sided die was zero, the odds with the 6 sided die was zero, and the odds with the 8 sided die was 1 / 8. When we normalized that result in Step 5, then obviously there would be a 100% chance that the die that rolled the 7 was the 8 sided die
- For this problem we worked with fractions, because they are fairly clean for only 3 possible dice, and only 1 roll. For most problems with more possibilities and more data, fractions become a pain and it is easier to work in decimals.
- It is usually just as easy to find the probability for all possibilities as for a single possibility. i.e. If you want to find the probability of the 4 sided die for this problem, you might as well solve for the 6 and 8 sided dice at the same time, since you have to do that anyway to get the total probability

What is in the rest of the book

At this point you may feel you have a good grasp of Bayes' theorem, and wonder if it is worth reading the rest of the book. And it's a good question because overall Bayes' theorem is very simple, so if you want to put the book down and kick back with some coffee I can't blame you. For the most part, everything that is more complicated than what we just covered falls into only a few different buckets, either

- Making the initial probabilities more complicated to cover cases where you have detailed initial information or
- Making the conditional probabilities more complicated to cover real life scenarios that are more complicated than rolling a dice or flipping a coin. (Like say, estimating the number of goals scored in a soccer game)

For the most part this book is going to avoid doing either of those and focus on more easy to understand, intuitive examples. None-the-less there are some interesting intricacies that we'll cover that may not be obvious from the first example such as

- How to include more than one piece of data
- How to handle possible errors in the data
- Non-intuitive final probabilities in the real life example of drug testing

Example 2 – More Dice, More Rolls

For this problem, we will stick with trying to predict the probabilities of dice drawn at random. This time we are going have 6 possible dice, one with 4 sides, one with 6 sides, one with 8 sides, one with 10 sides, one with 12 sides and one with 20 sides. We are going to roll the die 15 times and calculate what the probability for each die that it was the one that was drawn.

For this problem, since there is a lot more data, we will set up the problem in Excel to generate the probability tables. The tables that we will generate are like the table at the end of example 1, except transposed so that the possible dice run across the top row, and each new die roll is a new row. This makes for a nice format in excel for multiple rolls, and makes it easy to just set up the equations in one row and drag them down for the rest of the rolls.

Generating The Random Roll

For problem 1, I chose that the die rolled a 2 as an illustrative example. However people are, in general, bad at picking truly random numbers, so instead of picking 15 "Random" die rolls I have input the number of sides on the die "randomly" chosen from the bag, and let excel choose random numbers to simulate the roll

	A	B	C
8		Input: Actual Dice Sides	
9		8	
10			
11			
12	Roll #		
13	0	Random Number Updated Every Time	Random Number Copied From Column B So it doesn't update every time
14	1	4	6
15	2	3	3
16	3	8	2
17	4	6	6
18	5	2	8
19	6	7	5
20	7	1	6
21	8	5	7
22	9	3	2
23	10	2	7
24	11	4	7
25	12	4	1
26	13	2	5
27	14	7	7
28	15	8	6

The random number is generated in column B using the function

= RANDBETWEEN(1,B9)

To generate a random number between 1 and 8. (The cell B9 references the 8 sided die chosen). Since Excel has an annoying habit of updating random numbers in its cells every time you press enter, I've copied the values as a paste special into column C so that those values wouldn't change. The values in column C are what will be used for the rolls for this problem. (The copy also updated the values in column B, which is why they don't match column C)

Initial Probabilities

The next step is to decide the initial probabilities.   Since there are six dice, and we are assuming they are randomly drawn from a bag,   I set the initial probability of drawing each die to be 1/6, which is .1667

Dice Possibilities	4	6	8	10	12	20
Initial Probability	0.16667	0.16667	0.16667	0.16667	0.16667	0.16667

Probability Of The Roll For Each Die

The next step is to set up the equation for the dice probability, given a roll. Like we saw at the end of example 1

- If the die roll is greater than whichever die we are looking at, the probability that die would roll that number is 0.   (i.e. if we rolled an 9, the probability that a 4, 6, or 8 sided die would roll that number is zero)
- If the die roll is less than whichever die we are looking at, the probability that die would roll that number is 1 divided by the number of sides on the die

We are going to combine steps 3 & 4 that we had in the previous problem, so we are going to get our probability for a roll, and multiply by the previous probability in a single cell.

We could manually put in the equation for each roll, but it is easier to use an IF statement in excel. The equation that we want to use is

= IF (equation to evaluate, value if true, value if false)

So we want

= IF (roll > die # ,   0,   1 / die number * previous normalized probability)

This says, if the roll is greater than the die number, set the probability equal to zero, otherwise set the probability equal to 1 divided by the die number, multiplied by the previous normalized probability.

=IF($C14>E$12, 0, 1/E$12 * M13)
    roll   die #   die #   normalized probability from previous roll

	C	D	E	F	G	H	I	J
12		Dice Possibilities	4	6	8	10	12	20
13	Random Number Copied From Column B So it doesn't update every time	Initial Probability	0.166667	0.166667	0.166667	0.166667	0.166667	0.166667
14	6	Roll 1	* M13)	0.028	0.021	0.017	0.014	0.008
15	3	Roll 2	0.000	0.053	0.030	0.019	0.013	0.005
16	2	Roll 3	0.000	0.074	0.031	0.016	0.009	0.002
17	6	Roll 4	0.000	0.093	0.029	0.012	0.006	0.001
18	8	Roll 5	0.000	0.000	0.026	0.009	0.003	0.000
19	5	Roll 6	0.000	0.000	0.085	0.022	0.007	0.000
20	6	Roll 7	0.000	0.000	0.092	0.019	0.005	0.000
21	7	Roll 8	0.000	0.000	0.098	0.017	0.004	0.000
22	2	Roll 9	0.000	0.000	0.104	0.014	0.003	0.000
23	7	Roll 10	0.000	0.000	0.108	0.012	0.002	0.000
24	7	Roll 11	0.000	0.000	0.111	0.010	0.001	0.000
25	1	Roll 12	0.000	0.000	0.114	0.008	0.001	0.000
26	5	Roll 13	0.000	0.000	0.116	0.006	0.001	0.000
27	7	Roll 14	0.000	0.000	0.118	0.005	0.000	0.000
28	6	Roll 15	0.000	0.000	0.119	0.004	0.000	0.000

If you are careful to put dollar signs ($) in front of the appropriate roll/column references you can drag this equation to work for all the dice, and all of the rolls

Normalize The Results

The final step is to normalize the results after each roll, which will give the total probability up until that roll
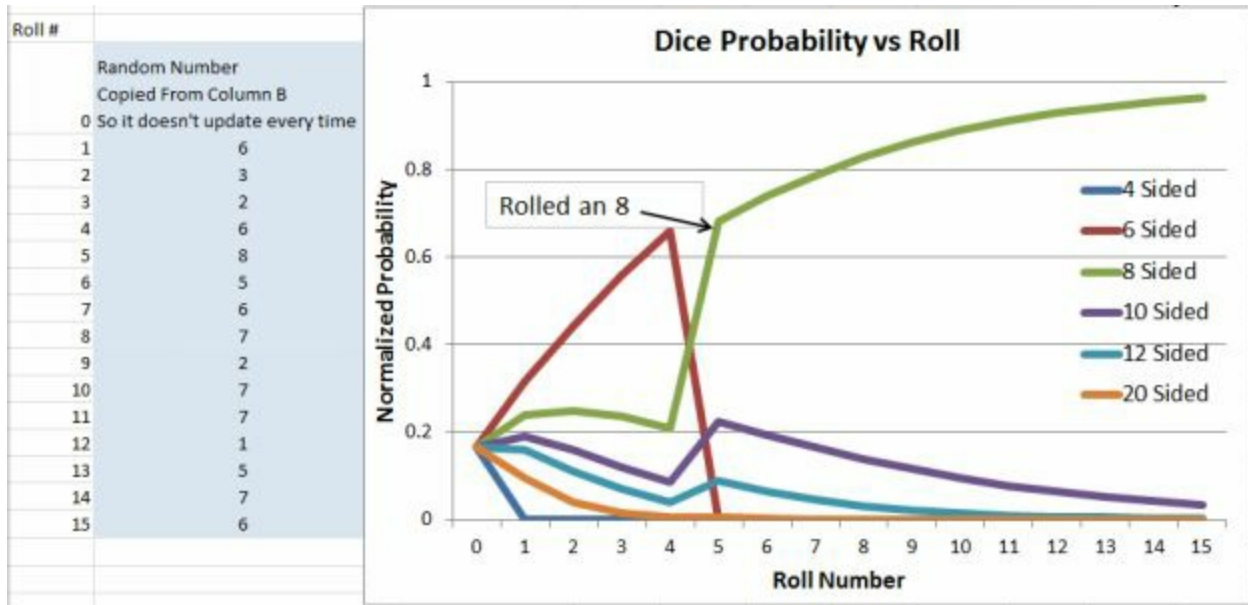
	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
10															
11											Normalized				
12	Dice Possibilities	4	6	8	10	12	20			4	6	8	10	12	20
13	Initial Probability	0.16667	0.16667	0.16667	0.16667	0.16667	0.16667			0.16667	0.16667	0.16667	0.16667	0.16667	0.16667
14	Roll 1	0.000	0.028	0.021	0.017	0.014	0.008		Roll 1	0.000	0.317	0.238	0.190	0.159	0.095
15	Roll 2	0.000	0.053	0.030	0.019	0.013	0.005		Roll 2	0.000	0.442	0.249	0.159	0.110	0.040
16	Roll 3	0.000	0.074	0.031	0.016	0.009	0.002		Roll 3	0.000	0.559	0.236	0.121	0.070	0.015
17	Roll 4	0.000	0.093	0.029	0.012	0.006	0.001		Roll 4	0.000	0.659	0.209	0.085	0.041	0.005
18	Roll 5	0.000	0.000	0.026	0.009	0.003	0.000		Roll 5	0.000	0.000	0.680	0.223	0.090	0.007
19	Roll 6	0.000	0.000	0.085	0.022	0.007	0.000		Roll 6	0.000	0.000	0.739	0.194	0.065	0.003
20	Roll 7	0.000	0.000	0.092	0.019	0.005	0.000		Roll 7	0.000	0.000	0.787	0.165	0.046	0.001
21	Roll 8	0.000	0.000	0.098	0.017	0.004	0.000		Roll 8	0.000	0.000	0.828	0.139	0.032	0.001
22	Roll 9	0.000	0.000	0.104	0.014	0.003	0.000		Roll 9	0.000	0.000	0.862	0.116	0.022	0.000
23	Roll 10	0.000	0.000	0.108	0.012	0.002	0.000		Roll 10	0.000	0.000	0.889	0.095	0.015	0.000
24	Roll 11	0.000	0.000	0.111	0.010	0.001	0.000		Roll 11	0.000	0.000	0.911	0.078	0.011	0.000
25	Roll 12	0.000	0.000	0.114	0.008	0.001	0.000		Roll 12	0.000	0.000	0.929	0.064	0.007	0.000
26	Roll 13	0.000	0.000	0.116	0.006	0.001	0.000		Roll 13	0.000	0.000	0.943	0.052	0.005	0.000
27	Roll 14	0.000	0.000	0.118	0.005	0.000	0.000		Roll 14	0.000	0.000	0.955	0.042	0.003	0.000
28	Roll 15	0.000	0.000	0.119	0.004	0.000	0.000		Roll 15	0.000	0.000	0.964	0.034	0.002	0.000

The equation for normalizing is just the probability found after the IF statement for each step, divided by the total probability for all the dice. Once again, with appropriate use of dollar signs this equation can just be dragged for all the dice and all the rolls

What we end up with is that after the 15th roll, we calculate a 96.4% chance that we had selected the 8 sided dice, and a 3.4% chance we had the 10 sided dice. The 12 and 20 sided dice had a very small percentage, and the 4 and 6 sided dice had a zero percent chance.

This answer of course matches the fact that we input the 8 sided die into Excel's random number generator.

When we plot the percentages after each roll

Roll #	Random Number Copied From Column B So it doesn't update every time
0	
1	6
2	3
3	2
4	6
5	8
6	5
7	6
8	7
9	2
10	7
11	7
12	1
13	5
14	7
15	6

We see that

- After the first roll, the 4 sided die drops to a zero percent chance. This is because the first roll was a 6
- For the first several rolls, the 6 sided die appears to have the highest likelihood, since it is the lowest numbered die that hadn't had a roll exceeding its value. But since there was an 8 rolled on roll #5, the 6 sided die dropped to zero and the 8 sided die became dominant
- The 10 and 12 sided dice experienced decaying likelihood since there was no roll greater than 8. They saw a brief spike in probability when the 6 sided die dropped out since the results would be normalized on a much smaller total probability
- The 20 sided die was the fastest to have its probability decline (except for the 4) since if we had selected a 20 sided die we were very likely to roll a large number soon and knock out most of the small numbered dice. Since that did not occur, the probability we had selected the 20 sided die is very small (but still greater than zero)

Get the Excel File

If you want the Excel file shown, for this or any of the examples, it is available here for free. http://www.fairlynerdy.com/bayes-theorem-examples/

Some things to take note of after this example

- Bayes theorem with multiple data points is basically just repeated multiplication. As a result, the final answer isn't affected by what order the data came in. If we were to roll 1, 5, 10 and calculate our final percentages, we would get the same final answer as if we rolled 10, 5, 1
- In this problem we normalized after every roll. You don't have to normalize each time; you could multiply all the probabilities together for each roll and normalize one time at the very end. The problem with this is that the probability numbers get very small. So small that many computer programs have trouble with them and truncation errors and round off errors can start to make impacts to the results. As a result, it tends to just be easier to normalize each step than to worry about dealing with probabilities that are $1.3 * 10^{-31}$ or $2.2 * 10^{-56}$

Bayes Theorem Terminology

Up until now we've skipped over some of the technical terminology for Bayes theorem. Now it is time to go over it, if for no other reason than so that you will understand it if looking at other sources.

- The initial probability, the probability of each possibility before we see the data, is called the **Prior**
- The normalized answer after computing the probability for each data point is called the **Posterior**
- The total probability that is used to normalize the answer is the **Normalizing Constant**
- The conditional probability, i.e. the probability of each possibility given the data, is called the **Likelihood**

How that looks on the table for the first problem is

Step 1	Prior	Likelihood			Posterior
	Step 2	Step 3		Step 4	Step 5
Die	Initial Probabilities	Chance of Roll		New Probability	Normalized Final Result
4	1/3	1/4		1/12	6/13
6	1/3	1/6		1/18	4/13
8	1/3	1/8		1/24	3/13
		Total Probability		13/72	
		Normalizing Constant			

Looking at it in equation form

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

Posterior   Prior   Likelihood   Normalizing Constant

Example 3 – Is It A Fair Coin

You have a coin that you suspect is not a fair coin. You flip the coin 100 times to determine if it is fair. After 100 flips, calculate the probability that the coin is weighted to come up heads each of these percentages, 0%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%.

Generating The Random Data

The data generation for this problem is set up similar to the last one. I started by inputting a percentage into excel for how frequently the coin will land on heads. Then I generated a random number for each flip in column B using the RAND() function. I copied & pasted special into column C so that it would not be updated each time. And then used an IF statement in Column D which said IF column C is less than The Heads probability (.65) make the cell in column D be 1 for Heads, otherwise make it be 0 for tails

	A	B	C	D
7		Input: Probability That The Coin Will Be Heads		
8		0.65		
9				Count of Number of Heads
10				63
11				
12 Flip #		Random Number Updated Every Time	Random Number Copied From Column B So it doesn't update every time	Column C Converted Into 1 for heads 0 for tails
13	1	0.4032	0.4519	1
14	2	0.4399	0.6133	1
15	3	0.6853	0.9985	1
16	4	0.9852	0.9096	1
17	5	0.6755	0.6004	1
18	6	0.7987	0.2715	0
19	7	0.6922	0.7593	1
20	8	0.2862	0.3592	1
21	9	0.4179	0.2735	0
22	10	0.6885	0.9312	1
23	11	0.4301	0.5135	1
24	12	0.2094	0.2348	0

Of course, if you are flipping an actual coin you can skip all of that and just write down the results from your flips.   What I got as data was 63 heads and 37 tails, which is pretty much what you would expect from a random number generator with the 65% chance of heads as an input.

If you want to follow this example with this exact data, you can grab the EXCEL file from my website, or copy down the 100 1's and 0's that follow which are the Heads & Tails I used

1, 1, 1, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 0, 1, 0

0, 0, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 0, 0, 1, 0, 1, 1, 0

1, 0, 0, 0, 1, 1, 0, 1, 1, 0, 0, 1, 0, 0, 1, 1, 1, 0, 1, 0

0, 0, 1, 1, 1, 1, 0, 1, 0, 1, 1, 1, 1, 0, 1, 0, 1, 1, 1, 1

1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 0, 0, 1, 1, 0, 1

Initial Probability

The first step is to set up what our possible outcomes are. Then add initial probabilities. The possible outcomes are defined, and are percentages from 0% to 100% by 10% steps. So there are 11 initial possibilities.

We will naively assume that all initial possibilities have an equal probability, so a 1 in 11 chance. (More realistically we could change the initial probabilities to more heavily weight the possibilities around 50%, and away from those at 0% or 100%. But that is complication we are avoiding in this book, and since we have 100 flips of the coin the initial probabilities aren't that important anyway. (Something that is known as "Swamping the Prior")

Coin Possibilities	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Initial Probability	0.0909	0.0909	0.0909	0.0909	0.0909	0.0909	0.0909	0.0909	0.0909	0.0909	0.0909

We are representing the initial possibilities as decimals, i.e. 50% is .5 which will be useful when calculating the odds after each flip.

Likelihood Function

To calculate the odds after the flip, we need to look at each possible outcome for each initial possibility. The possible outcomes are, the flip is heads, or the flip is tails. If the flip is heads, the probability that a given weighting of coin would produce that outcome is equal to its percentage. For instance, if the flip was heads, the probability that the 90% coin would produce that outcome is 90%. If the flip is tails, the probability that a given weighting of coin would produce that outcome is 1 minus the percentage. For instance the probability that a 90% coin would land on tails is 10%

Putting that in Excel, we need an IF statement that checks to see if the flip is heads, and assigns the heads probability multiplied by the old probability if it is heads, or assigns the tails probability multiplied by the old probability if it

is tails

=IF($D14=1, G$12*T13, (1-G$12)*T13)
Flip Result   ←Possibility %↑   Old Probability

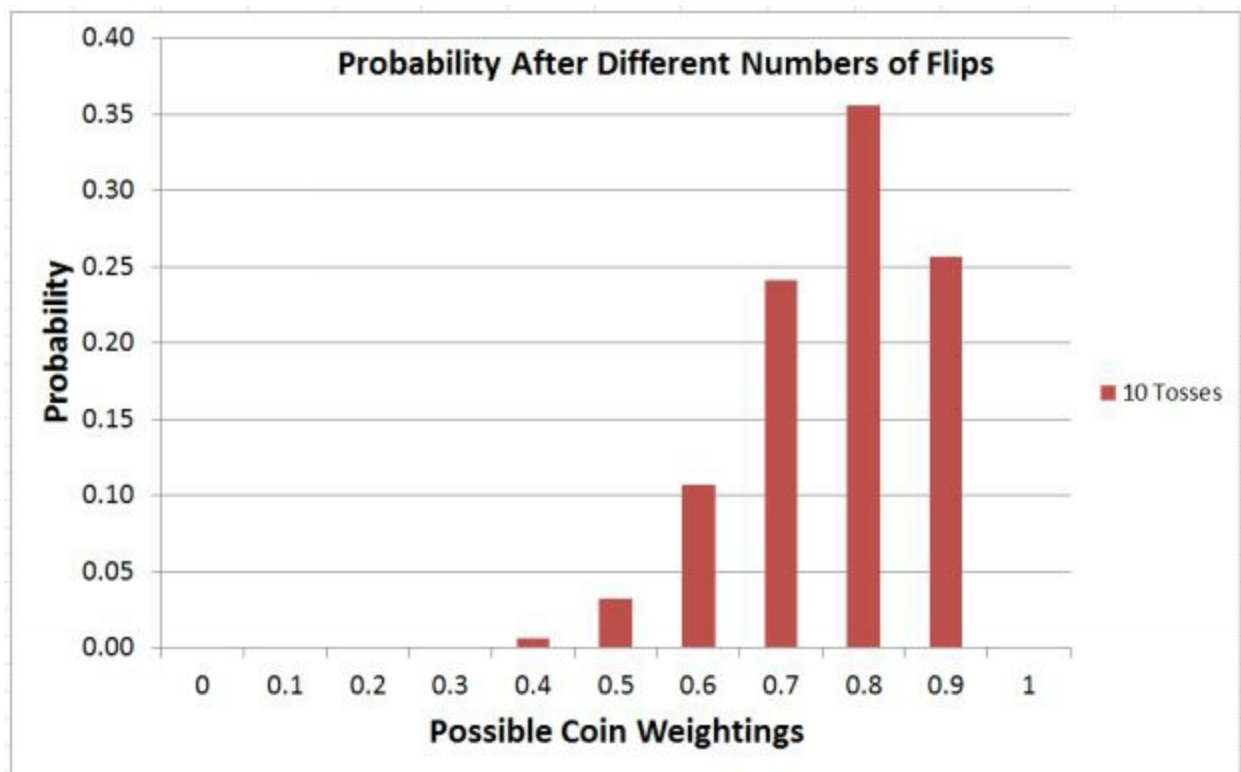|	D	E	F	G	H	I	J	K	L	M	N	O	P
12		Coin Possibilities	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
13	Column C Converted Into 1 for heads 0 for tails	Initial Probability	0.0909	0.0909	0.0909	0.0909	0.0909	0.0909	0.0909	0.0909	0.0909	0.0909	0.0909
14	1	Flip 1	0.00	=IF($D14=	0.02	0.03	0.04	0.05	0.05	0.06	0.07	0.08	0.09
15	1	Flip 2	0.00	0.00	0.01	0.02	0.03	0.05	0.07	0.09	0.12	0.15	0.18
16	1	Flip 3	0.00	0.00	0.00	0.01	0.02	0.03	0.06	0.09	0.13	0.19	0.26
17	1	Flip 4	0.00	0.00	0.00	0.00	0.01	0.02	0.04	0.08	0.14	0.22	0.33
18	1	Flip 5	0.00	0.00	0.00	0.00	0.00	0.01	0.03	0.07	0.13	0.23	0.39
19	0	Flip 6	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.02	0.03	0.03	0.00
20	1	Flip 7	0.00	0.00	0.00	0.00	0.01	0.03	0.08	0.15	0.23	0.23	0.00
21	1	Flip 8	0.00	0.00	0.00	0.00	0.01	0.02	0.07	0.15	0.25	0.28	0.00
22	0	Flip 9	0.00	0.00	0.00	0.00	0.00	0.01	0.03	0.06	0.06	0.04	0.00
23	1	Flip 10	0.00	0.00	0.00	0.00	0.01	0.04	0.10	0.19	0.24	0.16	0.00
24	1	Flip 11	0.00	0.00	0.00	0.00	0.00	0.02	0.08	0.18	0.27	0.19	0.00

Normalizing

Once again we are normalizing every step, which gives us our odds after every step

Normalized	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Initial	0.0909	0.0909	0.0909	0.0909	0.0909	0.0909	0.0909	0.0909	0.0909	0.0909	0.0909
Flip 1	0.00	0.02	0.04	0.05	0.07	0.09	0.11	0.13	0.15	0.16	0.18
Flip 2	0.00	0.00	0.01	0.02	0.04	0.06	0.09	0.13	0.17	0.21	0.26
Flip 3	0.00	0.00	0.00	0.01	0.02	0.04	0.07	0.11	0.17	0.24	0.33
Flip 4	0.00	0.00	0.00	0.00	0.01	0.02	0.05	0.09	0.16	0.26	0.39
Flip 5	0.00	0.00	0.00	0.00	0.00	0.01	0.04	0.08	0.15	0.27	0.45
Flip 6	0.00	0.00	0.00	0.01	0.03	0.07	0.14	0.22	0.29	0.26	0.00
Flip 7	0.00	0.00	0.00	0.00	0.01	0.05	0.11	0.21	0.31	0.31	0.00
Flip 8	0.00	0.00	0.00	0.00	0.01	0.03	0.09	0.19	0.32	0.37	0.00
Flip 9	0.00	0.00	0.00	0.00	0.02	0.07	0.16	0.27	0.30	0.17	0.00
Flip 10	0.00	0.00	0.00	0.00	0.01	0.05	0.13	0.26	0.33	0.21	0.00
Flip 11	0.00	0.00	0.00	0.00	0.01	0.03	0.11	0.24	0.36	0.26	0.00

This ends up being easiest to view as a bar chart. The initial chart is the probability after 10 flips, and then each additional chart adds 20 additional flips.   Since our input data was a coin probability of 65%, and what we

actually got was 63 flips out of 100, what we expect to see in these charts is that the possibilities for 60% and 70% are the most likely
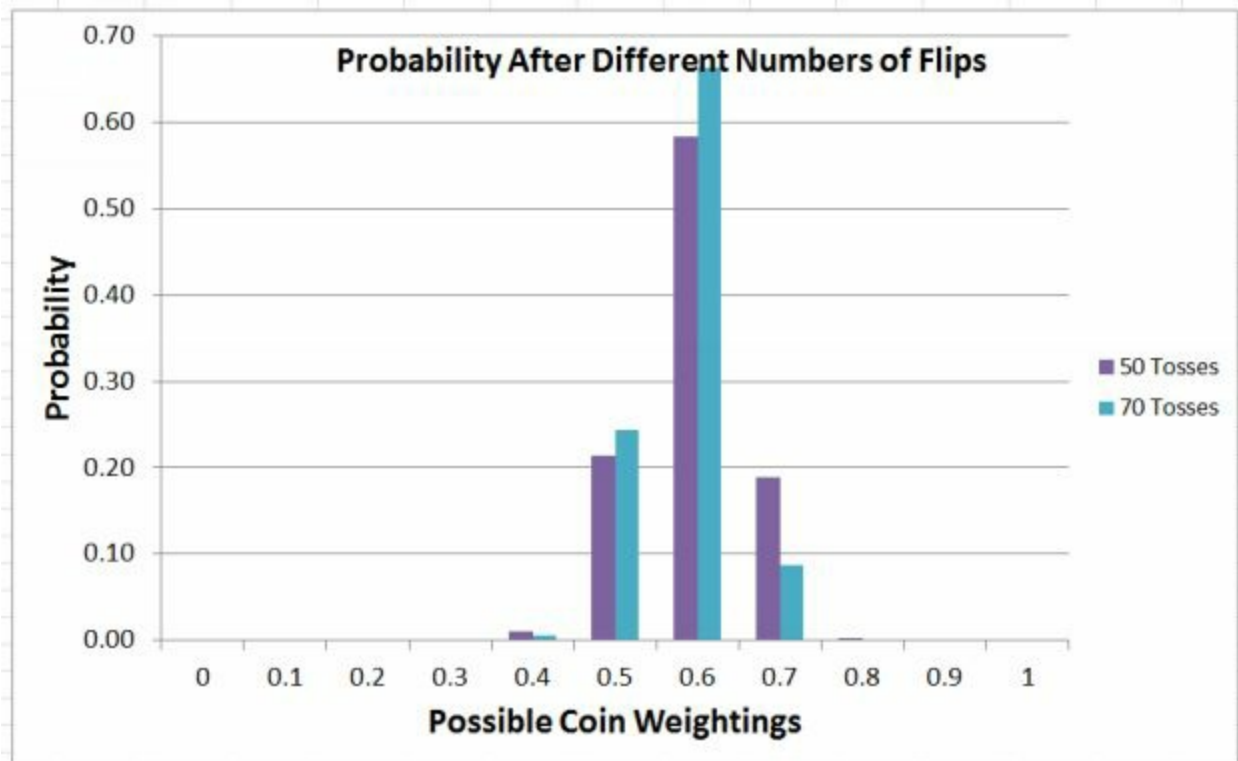
Probability After Different Numbers of Flips

(Chart: bar graph with x-axis "Possible Coin Weightings" ranging 0 to 1, y-axis "Probability" ranging 0.00 to 0.40. Legend: 10 Tosses. Bars at 0.4 ≈ 0.005, 0.5 ≈ 0.03, 0.6 ≈ 0.11, 0.7 ≈ 0.24, 0.8 ≈ 0.355, 0.9 ≈ 0.255.)

After 10 flips, the 80% coin has the highest probability. This is because there were 8 heads in the first 10 flips. The 0% coin and the 100% have a zero probability, since there was at least 1 head and 1 tail. The 10%, 20%, 30% coin have a very small probability.
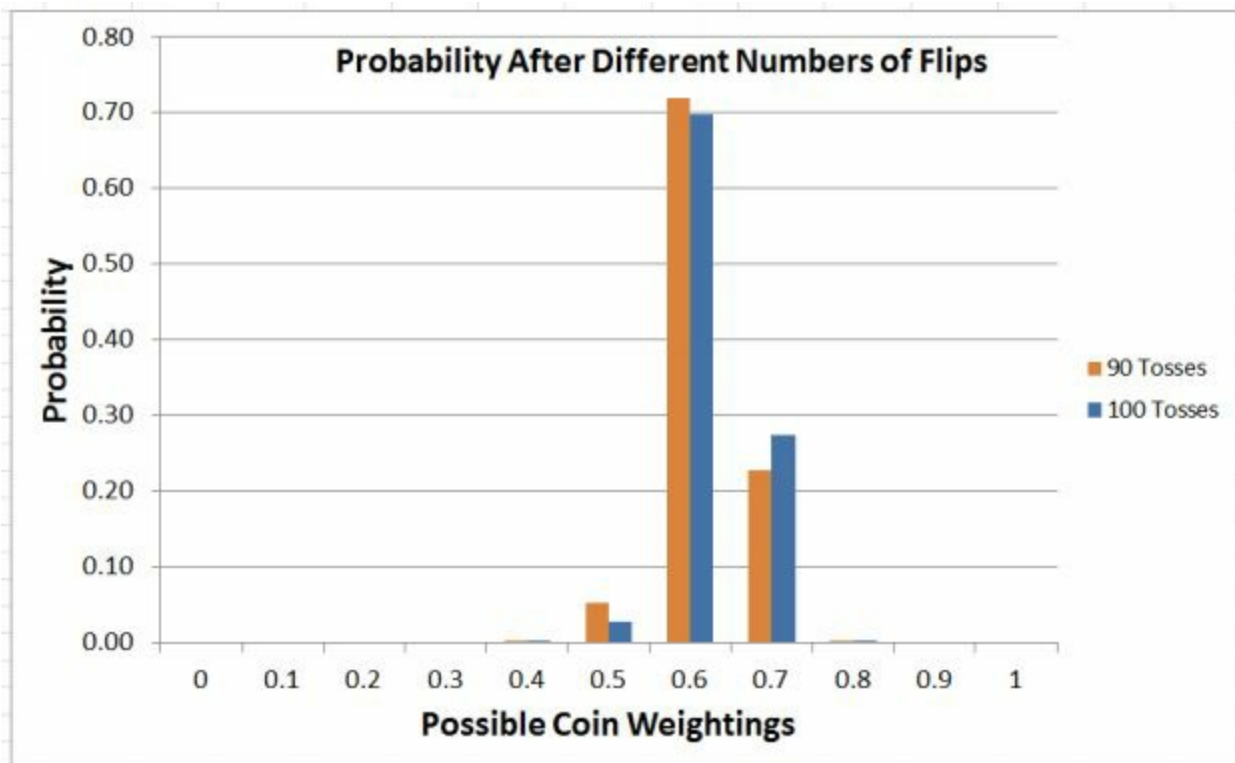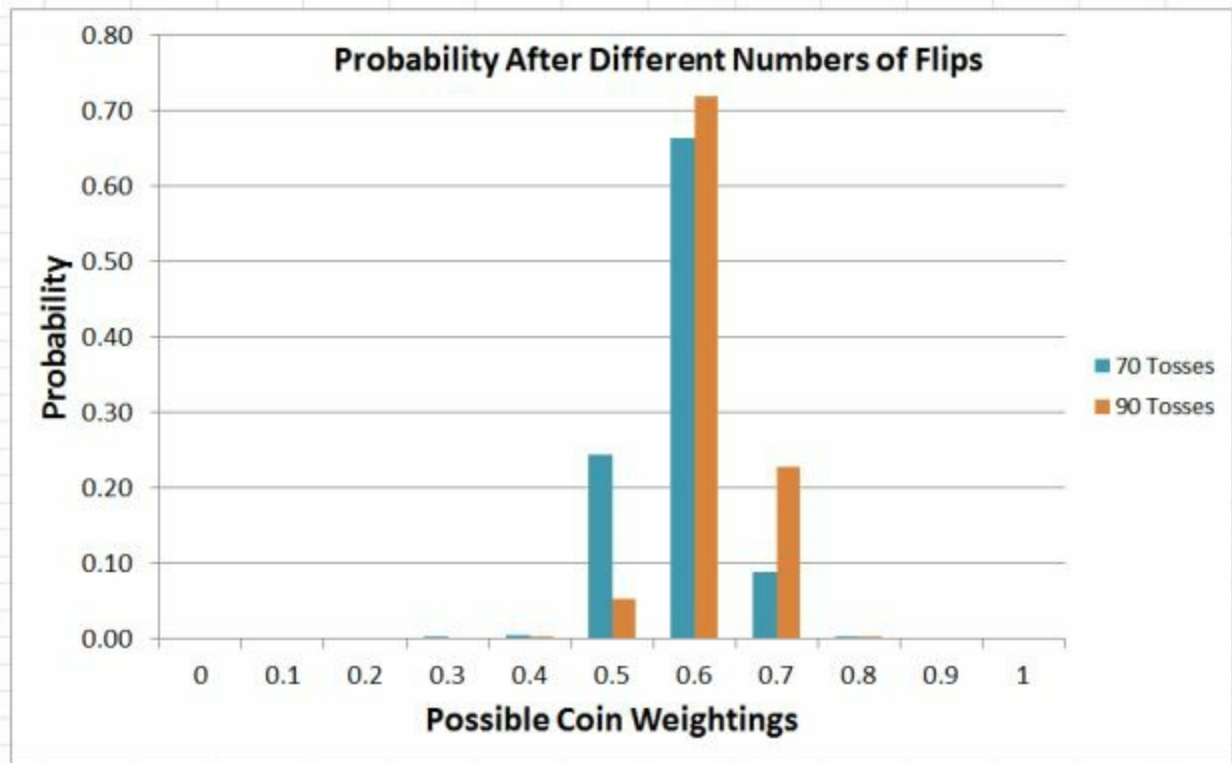
Probability After Different Numbers of Flips
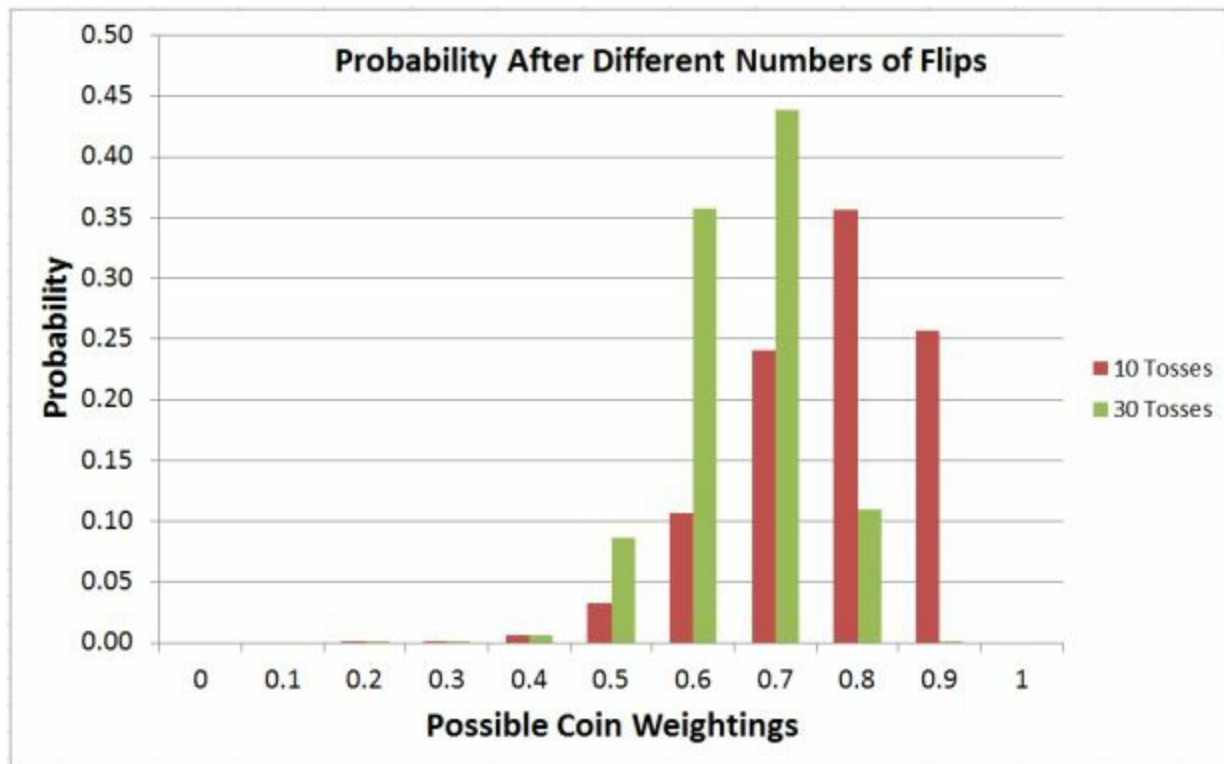
After 30 flips, the 70% coin has the highest probability, with the 60% close behind. This reflects 20 heads in the first 30 flips (66%)

Probability After Different Numbers of Flips
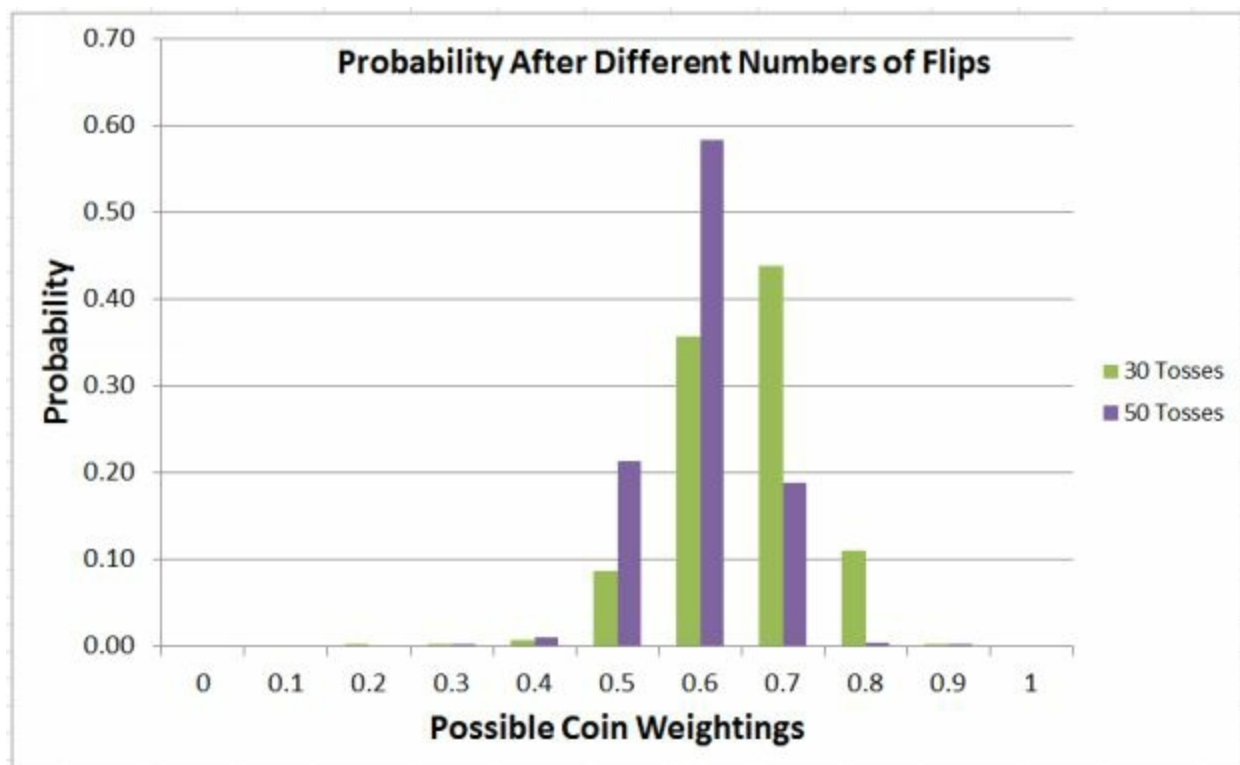
In subsequent flips, there is continued movement towards the 60% probability number, and the 50% and 70% probability numbers have reasonably high percentages as well.

Probability After Different Numbers of Flips

Probability After Different Numbers of Flips

(70 Tosses / 90 Tosses)

Probability After Different Numbers of Flips

(90 Tosses / 100 Tosses)

Finally, after 100 flips, we end up with an approximate 70% chance that we have a .6 weighted coin, a 27% chance we have a .7 weighted coin, and a 3%

chance that we have a .5 weighted coin.

What does the 70% chance for the .6 weighted coin actually mean? It is not a 70% chance that the coin is exactly weighted 60 percent. Since we had step sizes of 10% between each coin, we are estimating there is a 70% chance this coin would land heads between 55 and 65 percent of the time. The decision at the beginning to use 11 initial possibilities, with a step size of 10% was completely arbitrary. We could have used 101 initial possibilities with a step size of 1%. If we had done that, we would have gotten a result that had the highest probability at 63% (since we had 63 heads) and a smoother drop off in the chart.

If we consider a fair coin to be one that comes up between 45% and 55%, we are estimating that there is only a 3% chance that this is a fair coin.

Some things to take note of after this example

Another interesting point from this example is

- We saw convergence slower than we saw on the dice example from problem 2. This is because there is less difference in the probabilities for the coin flipping problem than for the dice problem. i.e. There was less distinction between a coin flipping 60% vs. 70% than there was for a die rolling an 8 vs a 10

Example 4 – More Dice, But With Errors In The Data Stream

The setup for this example is the same as problem 2, except that we are going to assume there could be errors in our data. Our friend is pulling dice out of a bag again, and the dice could have 4, 6, 8, 10, 12, or 20 sides. He is going to roll the die 80 times and write down the results.

The only problem is that this time he is sloppy at writing down the results. Perhaps he is a bad typist and his fingers keep slipping on the keyboard. As a result, 95% of the time the data will be entered correctly, but the other 5% of the time the data we are working with will just be a random number between 1 and 20.

Data Generation

As input I entered the number of sides, as well as the odds that the data had an error. In this case, the die rolled is an 8 sided die, and there is a 5% chance of an error

Input: Actual Dice Sides
8
Probability of an error writing down the number
0.05

To generate the input, I used a combination of these functions

- IF()
- RAND()
- RANDBETWEEN()

The equation is

= IF(RAND() > Probability of Error, RANDBETWEEN(1, # of Sides on Die), RANDBETWEEN(1,20))

Which says, if the random number is greater than the probability of an error (.05 in this example) then there was not an error on this roll, so generate a random integer between 1 and the number of sides on the die, otherwise generate a random integer between 1 and 20

Once again, the data is copied over to another column so it doesn't update every EXCEL calculation

=IF(RAND()>B11,RANDBETWEEN(1,B9),RANDBETWEEN(1,20))

	A	B	C
8		Input: Actual Dice Sides	
9		8	
10		Probability of an error writing down the number	
11		0.05	
12			
13	Roll #		
14	0	Random Number Updated Every Time	Random Number Copied From Column B So it doesn't update every time
15	1	8	5
16	2	5	6
17	3	11	1
18	4	8	7
19	5	4	2
20	6	5	2
21	7	3	6
22	8	1	3
23	9	8	3
24	10	8	6
25	11	7	4
26	12	15	5
27	13	6	1

The EXCEL file with the data and solution is available here, and the 80 die rolls that I used are

5, 6, 1, 7, 2, 2, 6, 3, 3, 6

4, 5, 1, 7, 7, 6, 7, 3, 4, 1

7, 4, 3, 3, 2, 7, 2, 7, 3, 6

8, 4, 1, 2, 6, 6, 1, 4, 3, 1

2, 4, 7, 6, 3, 8, 4, 5, 8, 7

4, 8, 3, 7, 4, 1, 6, 5, 8, 7

3, 7, 17, 5, 6, 1, 4, 8, 7, 3

2, 3, 5, 7, 6, 3, 4, 19, 1, 4

(Side-note, what do you think your error rate for manually copying those numbers down for your own use is? 1-2% ?)

What If You Ignored The Possibility of An Error?

At first thought, we could solve this problem just like we did in example 2. But when you do that, the plot of probabilities becomes what is shown below. (The easiest way in the EXCEL file is to leave the error rate as .05 when generating the roll data, but then set it back to zero since the Bayes equations are referencing it)
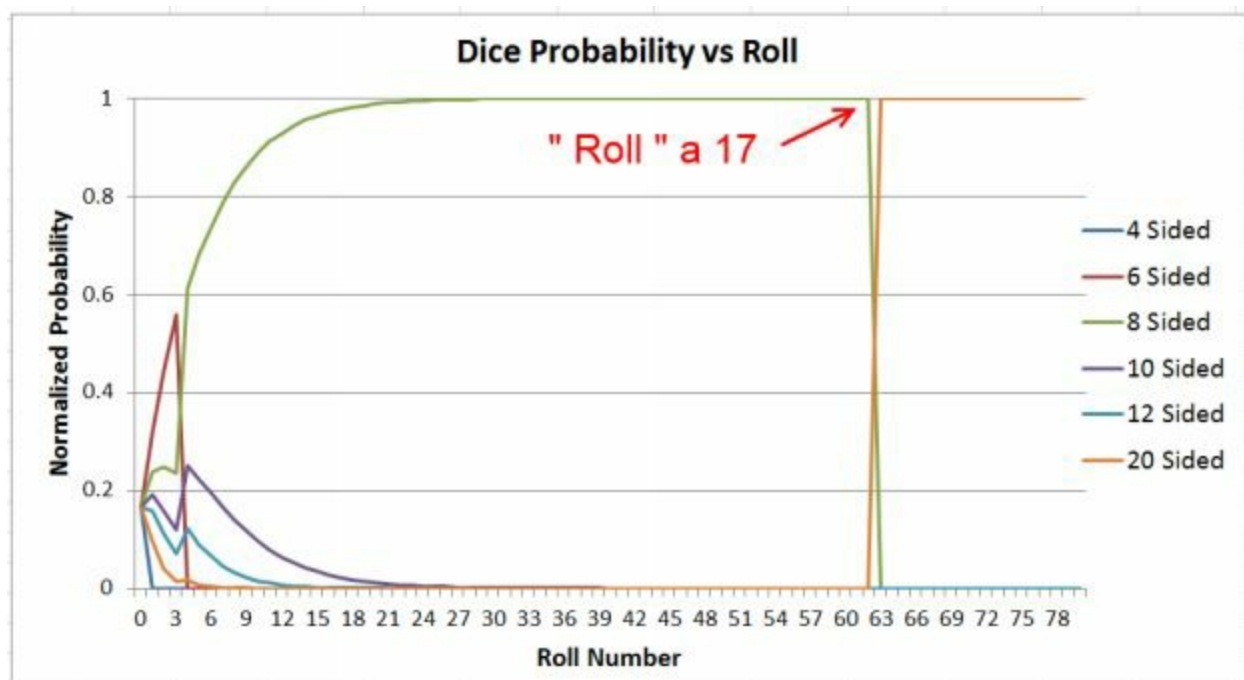
What occurs here is that the 8 sided die becomes dominant within the first 20 rolls, which are all less than 8. (statistically, with a 5% chance of error, there could have been an error in the first 20 rolls, but even if there is an error, that error has an 8 in 20 chance of being less than or equal to 8)   By roll number 63 the 8 sided die is extremely dominant, and the odds of the next closest die,

which is the 10, are less than 1 in 1 million.

However the odds of getting an error are only 1 in 20, and the odds that error will be greater than 8 are 60%. So 3% of the rolls will be an error greater than 8, which is what happens on roll #63 where a 17 is "rolled".

Since the equation from example 2 is set up to set the odds of a die equal to zero whenever a number greater than it is rolled, the odds for everything except the 20 sided die go to zero, which makes the normalized odds for the 20 sided die 100%

Now if a person was scrupulously watching the data, they might spot that error and cut the 17 from the dataset. But they might not. A more robust way is to bake some possible error into the conditional probability equations. The high level thought process behind this is

Never set any probability to zero unless you are positive that it cannot occur and you want to cut it from any future consideration.

If you do set the conditional probabilities to zero, then you could have a scenario where you roll one thousand numbers between 1 and 8, but the one thousand and first number gets written down as a 9 and you would strike the 8 from the possibilities. If instead of setting the conditional probabilities to zero, you just set it to a really small number, then the initial 1000 rolls would overwhelm the 1001st roll which is an error.

So how should the equations be set up?

Initial Probabilities

Once again we are setting the initial probabilities for each die to be 1 in 6

Dice Possibilities	4	6	8	10	12	20
Initial Probability	0.166667	0.166667	0.166667	0.166667	0.166667	0.166667

Likelihood Function

The conditional probabilities are what are different from example 2. But we

are still setting up the equation as "IF the roll is greater than the number on the die, use one probability, otherwise use a different probability". So all we need to do is calculate what the two probabilities are, baking our error rate into the equation. The two probabilities that we need to calculate are

- If the recorded number is greater than the number on the die, what are the odds you would have gotten that recorded number if you had selected that die from the bag
- If the recorded number is less than the number on the die, what are the odds you would have gotten that recorded number if you had selected that die from the bag

I am going to explain the equations using a 6 sided die as an example. And the same equations would be applied separately for the other dice.

If the recorded number is greater than 6, for instance if it is a 7

If you had a 6 sided die, the odds of recording a 7 are the probability of recording an error, divided by 20. That probability gets multiplied by the probability from the previous die roll, or initial probability. This is because the only way you can get a number greater than the die is if it is an error, and if the recorded number is an error the odds that it will be any given number are 1 in 20

If the recorded number is less than 6, for instance if it is a 2

If you had a 6 sided die, the odds of recording a 2 are One minus the probability of an error (i.e. the probability of correctly recording the die roll) divided by the number of sides on the die (6 in this case) PLUS the probability of recording an error divided by 20. That total probability gets multiplied by the probability from the previous die roll, or initial probability.

These odds are because there are two ways of recording a 2 for this 6 sided die. The first is that you roll a 2, and correctly record it. In that case the probability a 2 is written down is 1/6 * the probability of correctly recording the die roll. The second way of recording a 2 is to roll any number, and then incorrectly record it and randomly write down a 2.

This could be confusing, but you could actually take the 6 sided die, roll a 2,

and then ignore the roll but still randomly write down a 2. I.e. one way to think about it is that when you make an error, 5% of the time in this problem, you don't even bother looking at the die roll and just write down a random number.

A different way of arriving at the conditional probability equation

If the conditional probability equation with the error doesn't make sense, here is a different way of thinking about it.

- Start with the conditional probability for when there is no error. i.e. a 0 chance if the roll is greater than the number on the die, and 1 divided by the number on the die otherwise
- Then subtract the probability of writing down an error from the 1 divided by the die number. i.e. if you write down an error 5% of the time, the probability of not writing down the error is .95, which is then multiplied by the 1 divided by the die number
- Then, for both parts of the if statement, add in the probability that you would write down an error divided by 20 (since if you write down an error, there is a 1 in 20 chance you would write down any given number)
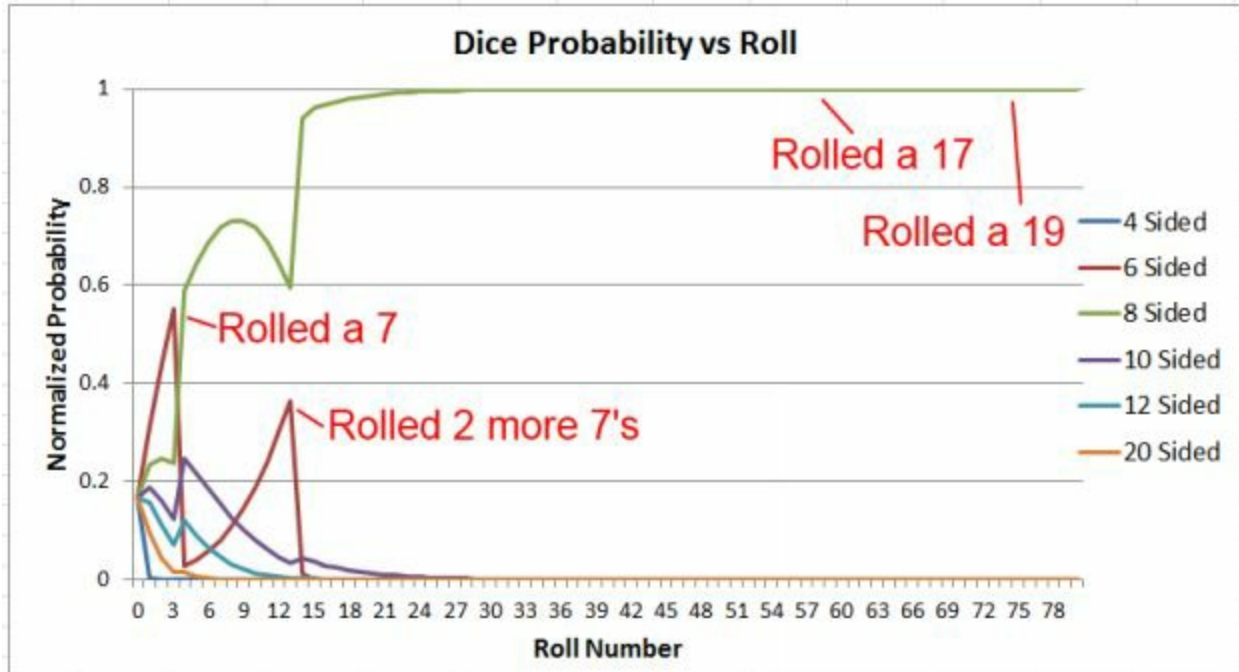
The equation in Excel

The way the equation looks in Excel


```
=IF($C15>F$13, $B$11*1/20*N14, ($B$11*1/20+(1-$B$11)/F$13)*N14)
```
Die | # on | Odds of | Previous | Odds of | # on | Previous
Roll | Die | Error | Probability | Error | Die | Probability

The Results

When you drag that conditional probability equation into all of the dice, and all of the rolls, and normalize as usual, the results are

Dice Probability vs Roll

In this example, the 6 sided die initially is the favorite, because the first 3 rolls are 5, 6, 1. Then a 7 is rolled and the 8 sided die is the favorite. However in this example, the 7 doesn't make the probability of having a 6 sided die completely equal to 0, because that 7 could be an error.

For the next nine rolls, they are all less than or equal to 6, and the 6 sided die starts to rise in probability again. However then there are two 7's rolled in quick succession and the probability of having the 6 sided die becomes very small.

Later on, there is a 17 and a 19 rolled that are clearly an error. Without the error checking the 20 sided die became the only possibility. But with the error checking the 8 sided die is still the most likely.

Example 4A - What if you have a really high error rate?

For the previous example we used a fairly low error rate, only 5%. At that level of error, a person could likely manually review data and spot odds numbers, which is basically what we were doing with this chart when we determined that it was weird for the 8 sided die to take such a hit
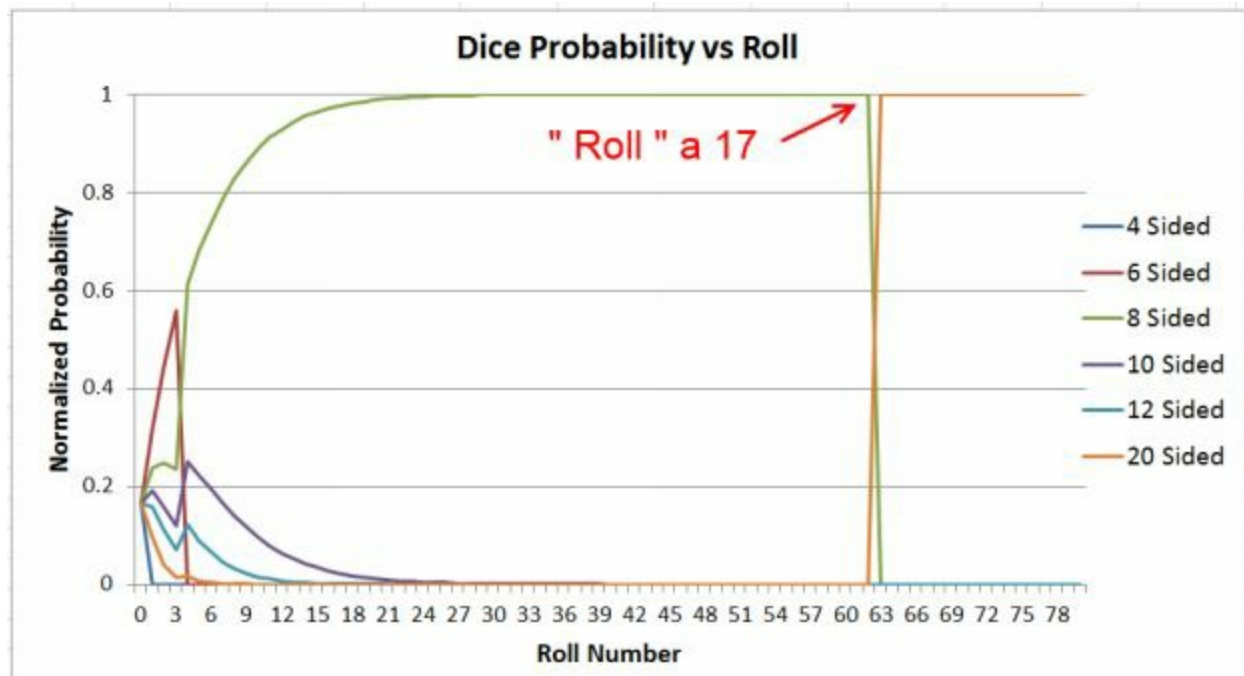
But what if you had a much higher error rate? This example is the same as the last one, except instead of a 5% error rate, we are assuming a 75% error rate. Since we have such a high error rate, we also need more data, so we have 250 rolls of the dice instead of 80. Once again we are starting with an 8 sided die as the die actually drawn from the bag

The data that I am using is

9, 8, 16, 3, 8, 4, 3, 6, 3, 18, 9, 17, 18, 5, 13, 8, 7, 6, 4, 2
11, 17, 17, 5, 14, 12, 6, 12, 6, 6, 3, 6, 6, 13, 13, 5, 4, 2, 6, 4
2, 3, 1, 8, 16, 4, 1, 11, 8, 7, 14, 16, 5, 10, 2, 3, 8, 2, 1, 3
10, 6, 16, 6, 10, 17, 1, 18, 19, 4, 5, 7, 3, 4, 2, 4, 20, 14, 20, 10
16, 6, 1, 7, 5, 1, 15, 8, 7, 1, 12, 10, 13, 1, 6, 9, 11, 13, 7, 15
10, 14, 6, 3, 5, 6, 1, 6, 2, 13, 20, 5, 18, 1, 12, 10, 10, 4, 5, 16
5, 14, 3, 4, 4, 2, 11, 12, 18, 4, 8, 1, 5, 19, 14, 7, 14, 17, 16, 8

1, 20, 18, 4, 5, 1, 7, 4, 5, 13, 8, 17, 3, 7, 6, 16, 18, 2, 11, 6
10, 19, 13, 2, 7, 12, 11, 18, 12, 4, 10, 8, 3, 10, 14, 18, 16, 3, 15, 11
16, 14, 7, 7, 13, 12, 11, 2, 6, 19, 7, 3, 19, 1, 5, 19, 18, 3, 18, 7
4, 4, 11, 14, 12, 5, 1, 12, 8, 19, 5, 19, 14, 16, 18, 9, 7, 3, 18, 2
5, 8, 1, 15, 11, 14, 19, 6, 4, 19, 3, 15, 11, 6, 18, 17, 13, 17, 9, 7
7, 15, 3, 15, 16, 16, 20, 2, 2, 5

Here it is not so easy to tell which items in the data are in error

Applying the same equations from Example 4 to this dataset, the plot below gets generated
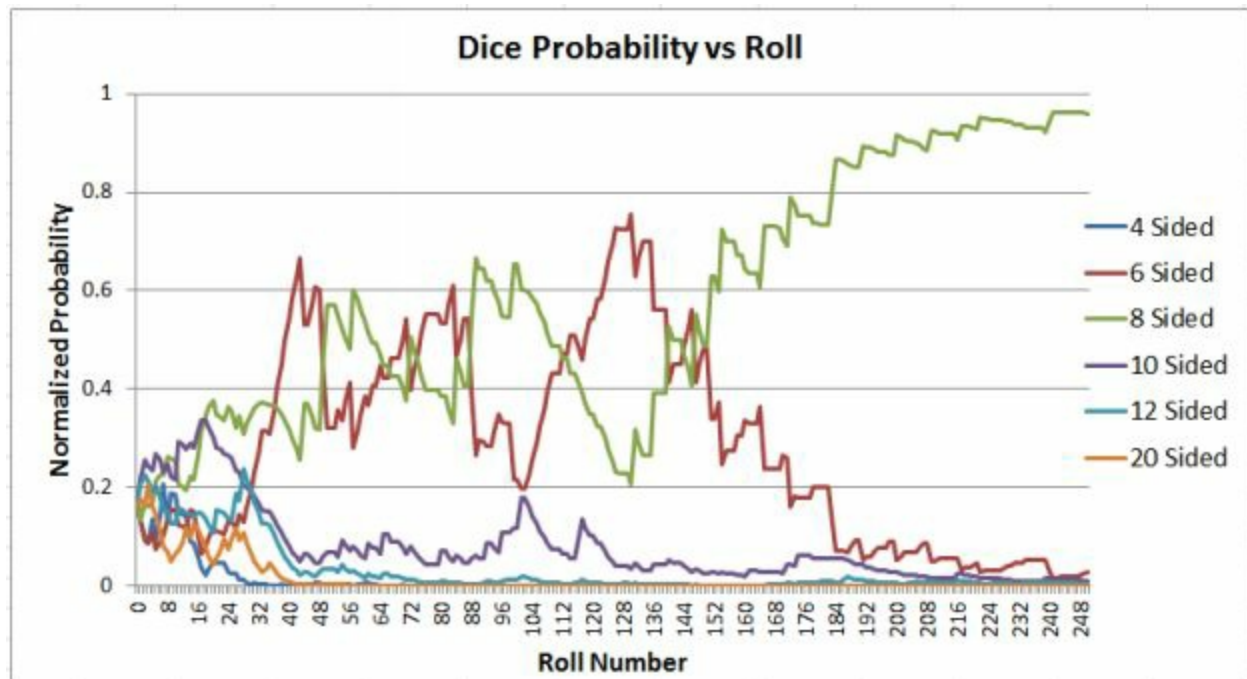
Once again the 8 sided die becomes the clear most likely die that was drawn from the bag. With this high error rate, it took many more rolls, and the result was a lot choppier. However even after an only moderate number of rolls, the result was most likely either the 6 or the 8, with the 10 being a smaller possibility.

If you have an even higher error rate, for instance 99%, you could still get the correct answer out of it. However the higher the error rate, the more rolls you will need. For this 75% error rate we are getting 1 good data point out of every 4. If we had a 99% error rate we would get 1 good data point out of every 100, so we would likely need at least 25 times as much data as we have now to get a dominant die.

What if you don't know the probability of an error?

Part of my solution in Examples 4 and 4A could be considered cheating. After all, I input the probability of the error, so therefore I knew exactly how likely there was to be an error and could bake that into the likelihood equations. What if you didn't know the likelihood of the error? After all, if you are working with actual data from actual measurements, you would have at best a rough guess as to what your error rate was.

For these plots I'm going to use the same dice rolls as I used in examples 4 & 4A. Those charts were generated with the 5% error rate and the 75% error rate respectively. However I am going to vary the error rate that is built into the conditional probability equations to be something different.
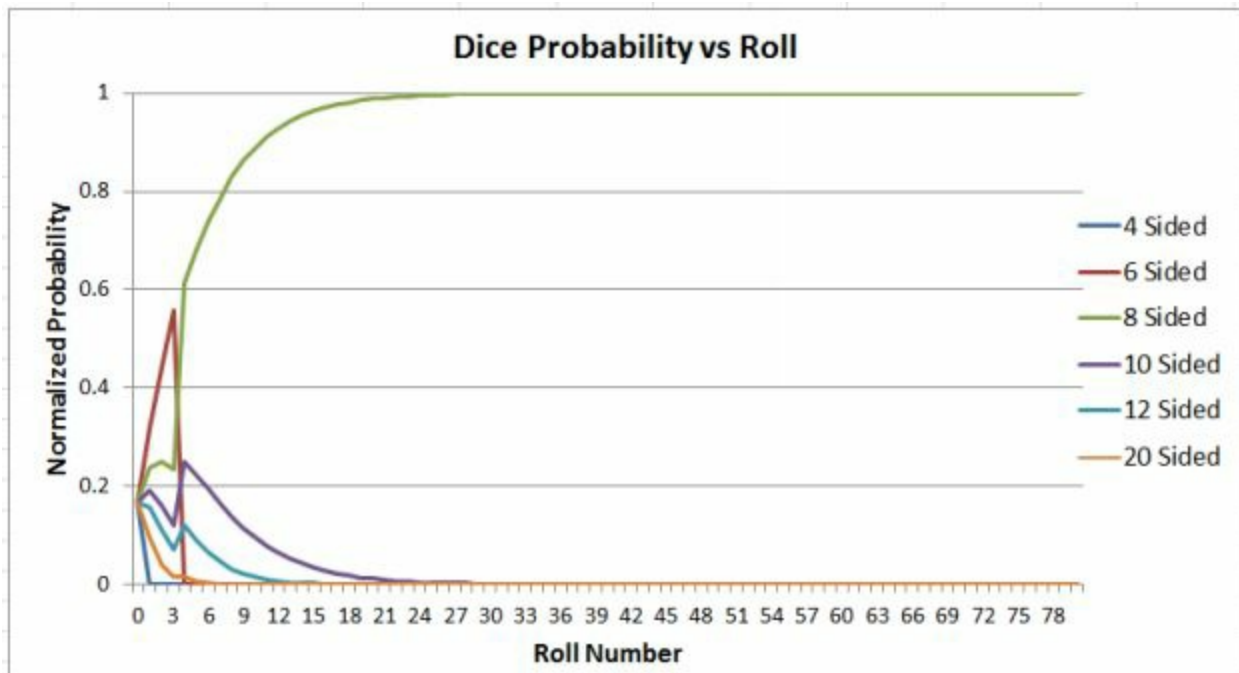
For the 5% error rate data, this was the baseline plot, which we generated knowing there was a 5% error rate
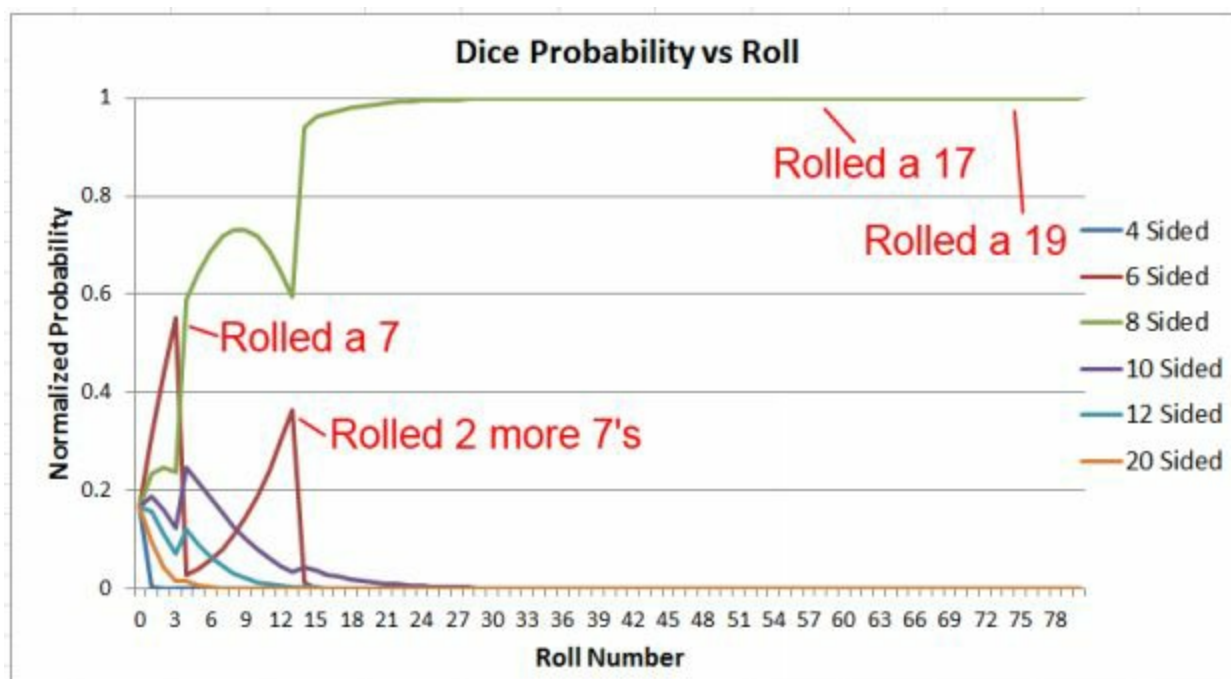
If instead we mistakenly assume there is a 1% error rate the plot changes to be
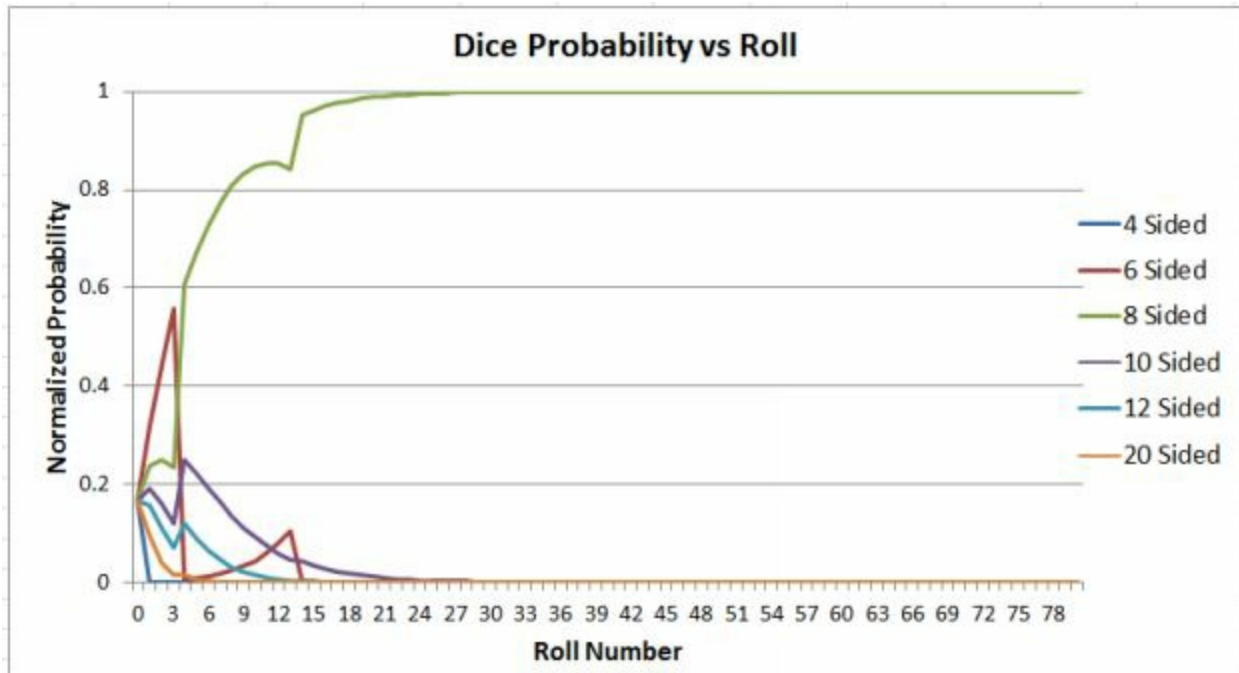
This is very similar to the baseline plot, except that it is converging faster. In fact we can knock the guessed error rate down even farther, and get substantially the same result

This plot is guessing a .0001% error rate. It looks almost exactly like the plot would if we assumed no errors and then actually did not have any errors. This is because by rolls 63 and 78, when we are getting our error rolls of 17 and

19, the probability of the other numbers are so low that even assuming a very low error rate doesn't change the results.

It is however possible to guess an error rate so low that the results change on this plot. Here is the plot using the ridiculously low error rate of 1e-25
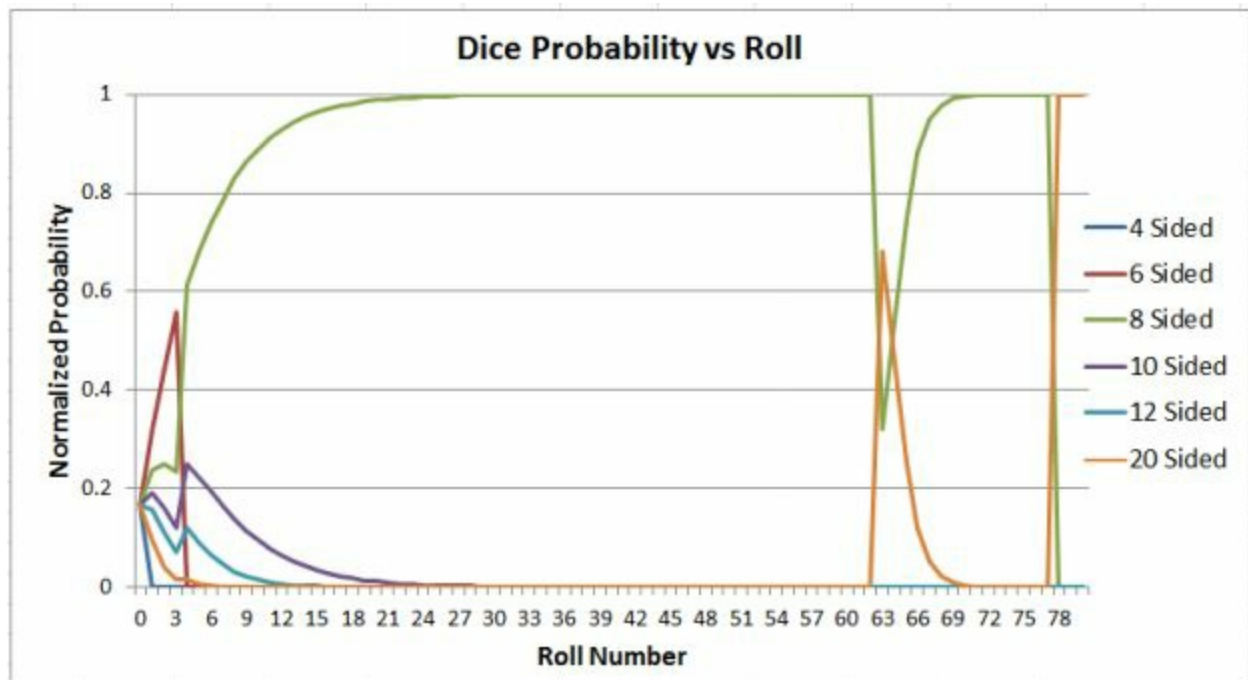

So clearly if we set the error rate too low the results change, and end up being the same as if we had not included an error rate at all. In this case "too low" is "ridiculously low", but that would not be true for every problem. What happens if we overestimate our error rate?

Here the error rate is set to 10%, when it was generated at 5%.
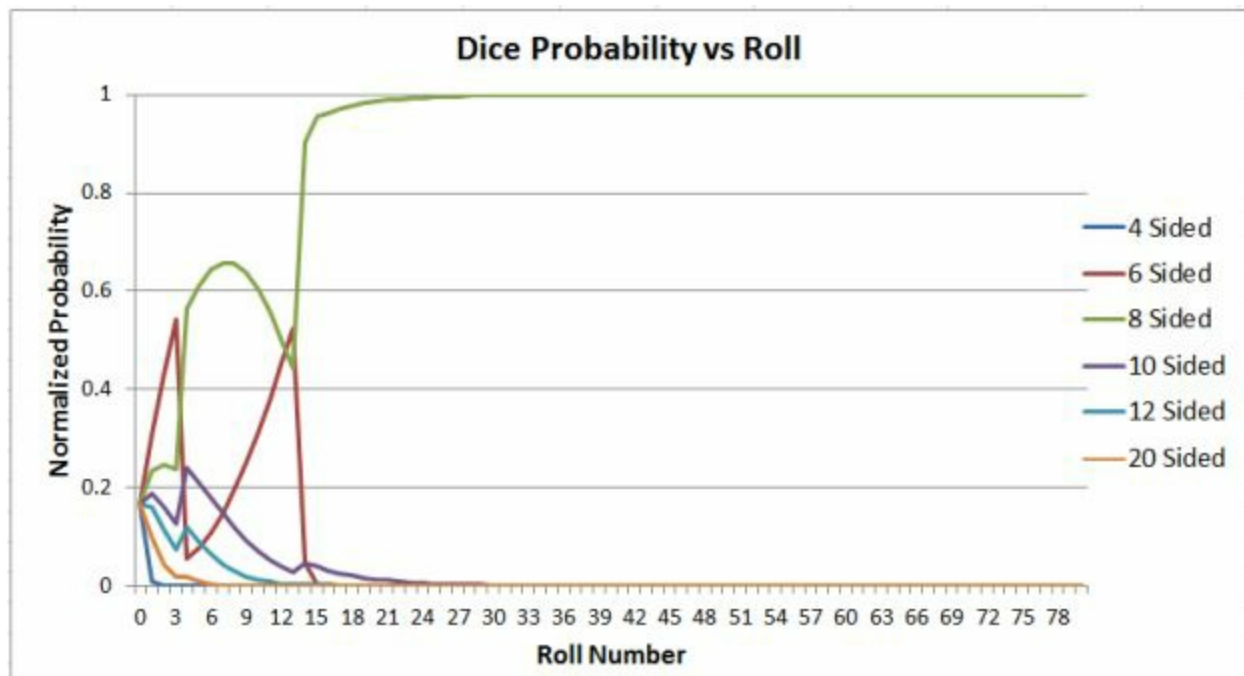
We start to see a slower convergence towards the 8 sided die, and a greater likelihood of the 6 sided die until it got knocked out of the running by multiple 7's right in a row.

Here we are assuming the error rate is 50%.
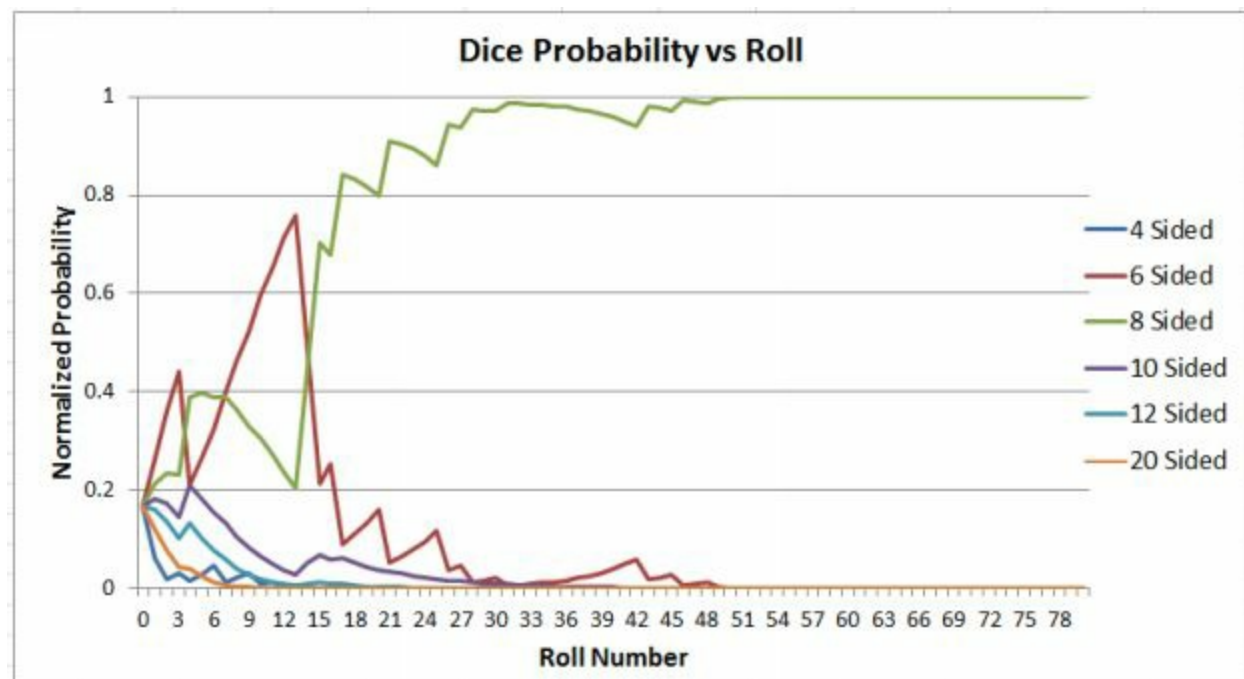
And we start to see it take quite a bit longer to converge, although it does

converge on the correct solution.

Those were examples where the actual error rate was low, 5%. What about when the actual error rate is high, 75% ? This is the plot of the baseline problem, where the data was generated with a 75% error rate, and the conditional probabilities were correctly done with the 75% error rate
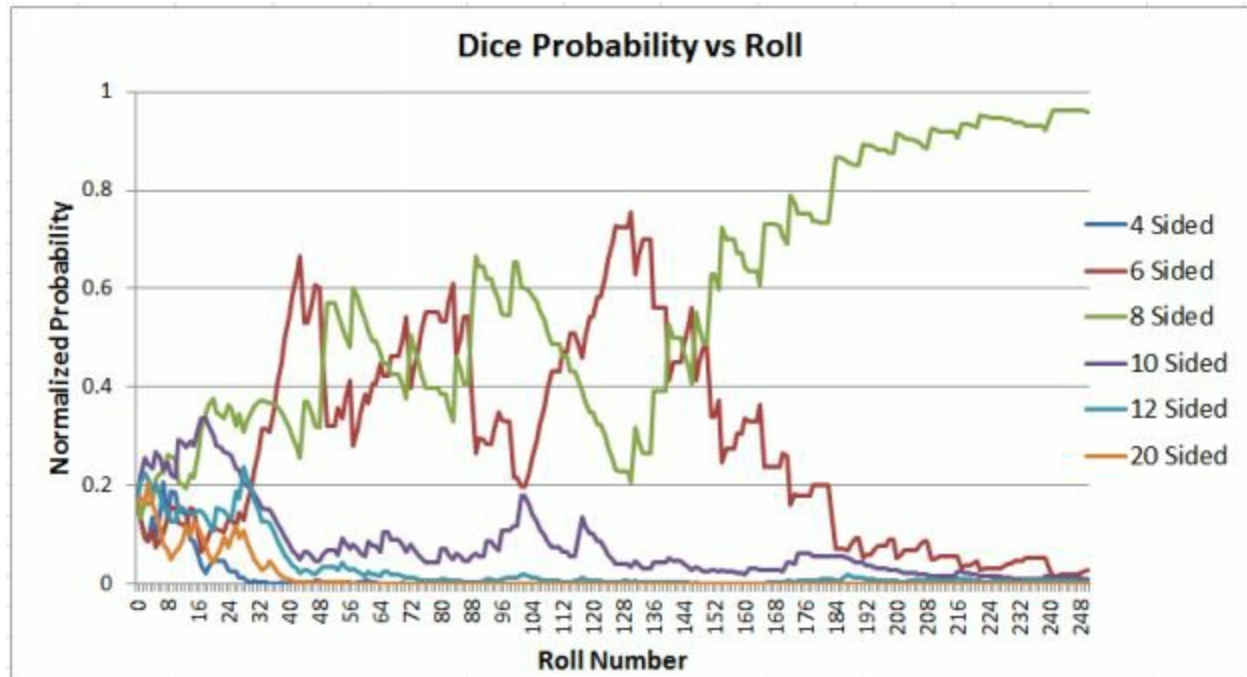
If instead, we generate the data with the 75% error rate, but guess that the error rate is 5% our result is

Dice Probability vs Roll
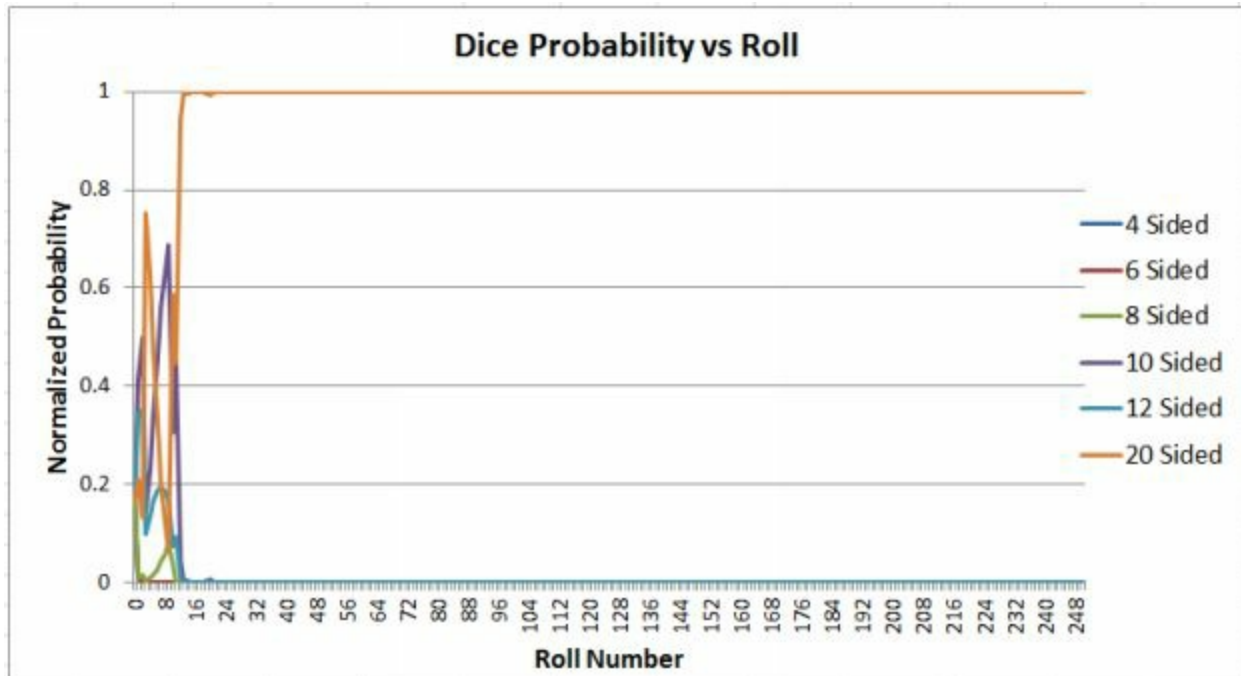
That result is wrong. It calculated the 20 sided die as the one selected, which is not correct. Even if we guess a 50% error rate on the data that was generated with the 75% error rate, it still trends to the wrong answer of the 20 sided die

Dice Probability vs Roll
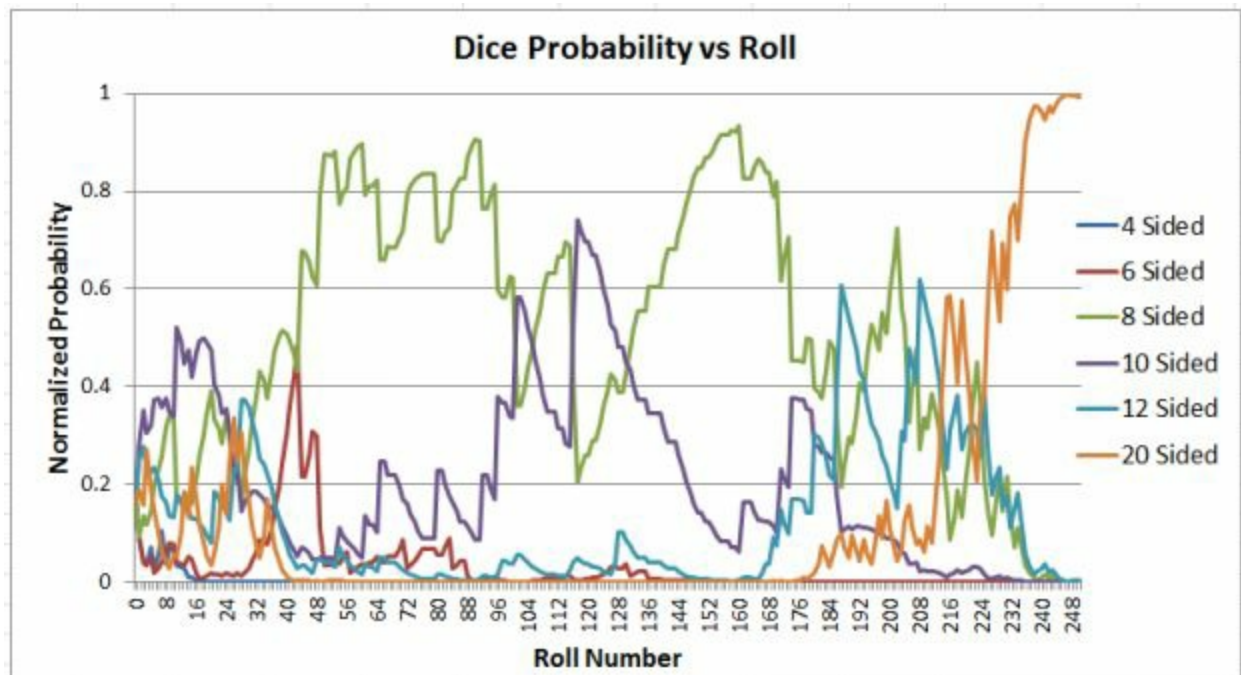
If we guess high, and guess an error rate of 90% we get the following chart
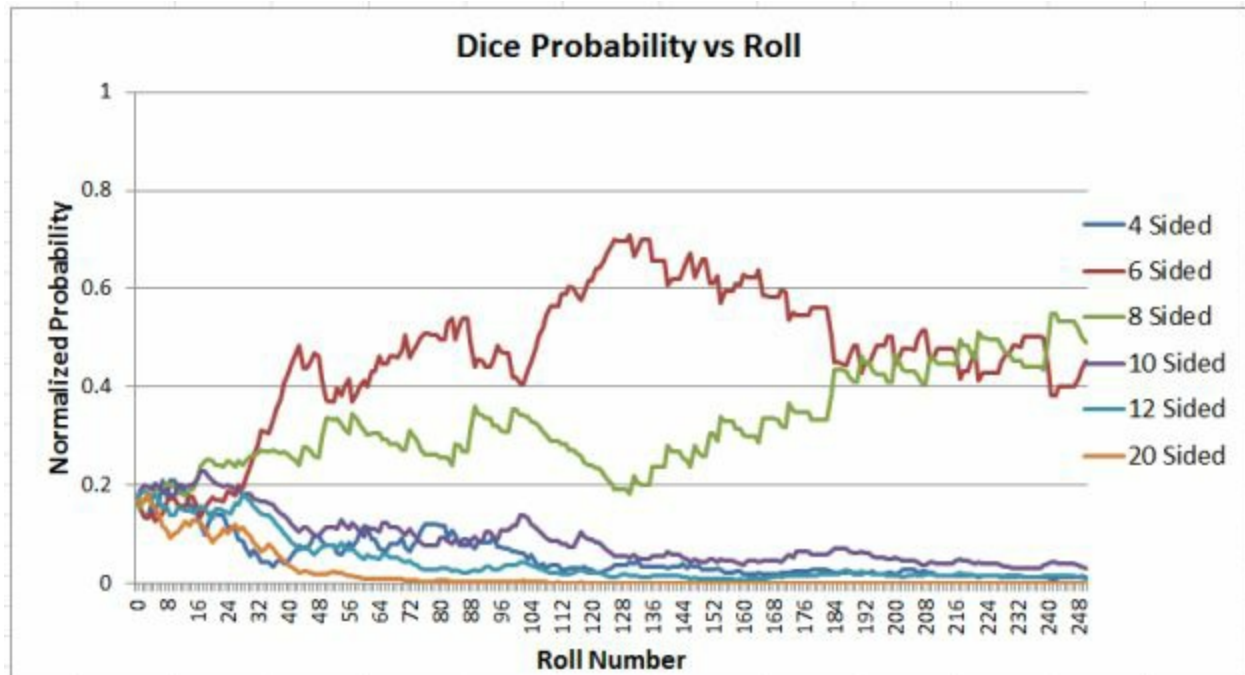
This shows the 6 and 8 sided die as having the highest probability, but it has not yet converged as having one die have the dominant results yet.

Big Picture Takeaway

So what is the big take-away from all these charts of the sensitivity around different error rates?

- If your estimate of the error rate is higher than the actual error rate, the results will converge slower but will still converge to the correct result
- If you estimate the error rate too low there is a risk that the results will not be correct
- The smaller the actual error rate is, the more wiggle room you have in guessing the error rate
- The higher the actual error rate is, the more data you need

Example 5 – The German Tank Problem

It's time to take a break from dice and coins and look at a few other applications of Bayes Theorem. One of the most famous applications of it is the German Tank Problem. In this problem you are trying to estimate how many tanks have been produced, based off of the serial numbers of captured tanks. Bayes theorem was used in World War 2 by the Allies to do exactly that, and ended up with results that were substantially lower for total number of tanks produced than conventional intelligence estimates (i.e. spies) were reporting. After the war, records indicated that the statistical estimates using Bayes Theorem were also substantially more correct.

Problem

You are analyzing serial numbers pulled off of wrecked or captured tanks. Use those numbers to estimate how many tanks have been produced. You know this about the tank serial numbers

- They start at 1
- They are sequential without gaps

You have found these serial numbers, 30, 70, 140, 125

Setting Up The Problem

Of all the examples in this book, this is the one where there is the most opportunity for differences in the problem set up. The biggest question to ask is, what is the maximum number of tanks in the estimate?

For this problem, I am going to choose a maximum of 1,000 tanks. But a reasonable person could choose a different number like 500 tanks, or 2,000 tanks and get different results. Because I am solving this problem in Excel, I am going to analyze for every 20 tanks, which means I have 50 initial possibilities for numbers of tanks. You could analyze for every single number in Excel with some difficulty, or do it easily in a programming language, but making this assumption will give more or less the same answer and substantially reduce the size of the Excel sheet.

	A	B	C	D	E	F	G	H	I	J	K
1											
2			Example 5								
3			You are analyzing serial numbers pulled off of wrecked or captured tanks. Use those numbers to estimate how many tanks have been produced								
4			You know this about the tank serial numbers								
5			- They start at 1 - They are sequential without gaps								
6			You have found these serial numbers, 30, 70, 140, 125								
7											
8											
9	Tank Number	Serial Number	Total Number of Tanks	20	40	60	80	100	120	140	160
10	0		Initial Probability	1	1	1	1	1	1	1	1
11	1	30	Tank 1	0.00000	0.00050	0.00033	0.00025	0.00020	0.00017	0.00014	0.00013
12	2	70	Tank 2	0.00000	0.00000	0.00000	0.00089	0.00057	0.00040	0.00029	0.00022
13	3	140	Tank 3	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00055	0.00037
14	4	125	Tank 4	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00180	0.00106

I am assuming that all possibilities of number of tanks are equal (i.e. in the prior there is the same probability of having 50 tanks as having 500 tanks). Once again, this is something that a reasonable person could make a different assumption on, and get different results. (Note, there are more columns in the Excel file than shown in the picture, it wasn't possible to show all the way up to 1000 and still make the picture legible)

Likelihood Function

The conditional probability is a lot like the conditional probability for the dice problem.

- If the serial number observed is greater than the max serial number for that possibility, the probability of having that number of tanks that number is 0
- If the serial number observed is less than the max serial number for that possibility, the probability of observing that number is 1 divided by that number of tanks, multiplied by the probability for the previous step

A	B	C	D	E	F	G	H	I	J	K
8										
Tank Number	Serial Number	Total Number of Tanks	20	40	60	80	100	120	140	160
10 0		Initial Probability	1	1	1	1	1	1	1	1
11 1	30	Tank 1	L1>D$9, 0,	0.00050	0.00033	0.00025	0.00020	0.00017	0.00014	0.00013
12 2	70	Tank 2	0.00000	0.00000	0.00000	0.00089	0.00057	0.00040	0.00029	0.00022
13 3	140	Tank 3	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00055	0.00037
14 4	125	Tank 4	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00180	0.00106

After multiplying these conditional probabilities through the 4 tanks observed and normalizing, these are the results
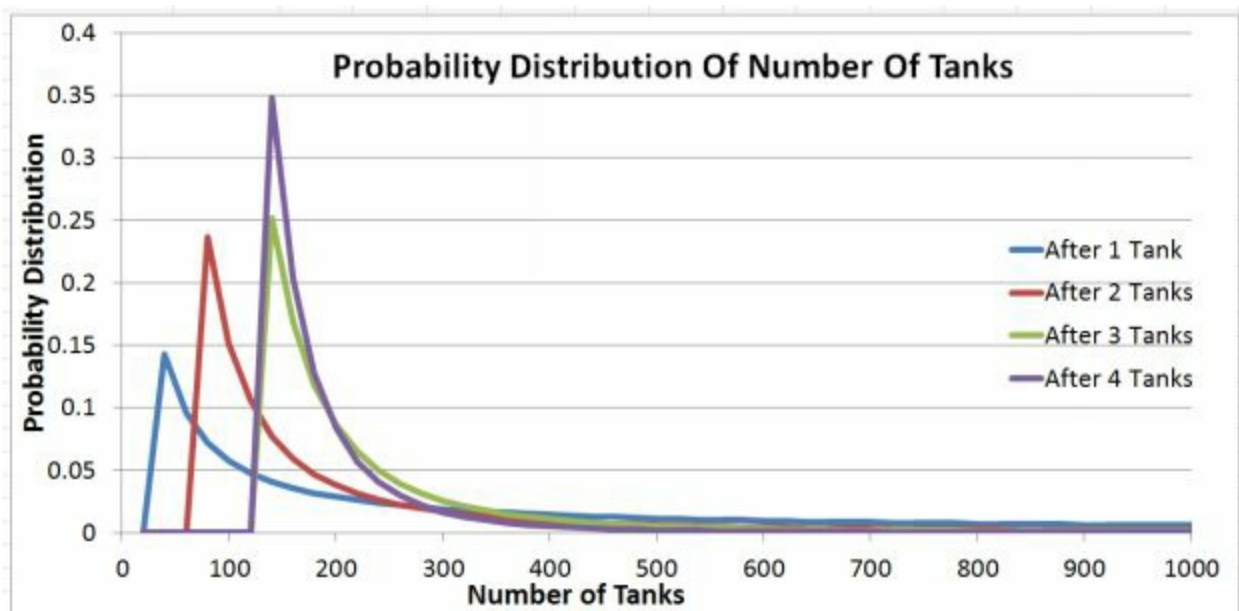
There is a big spike in probability at the maximum serial number observed. After that there is an asymptotic decay towards the greater number of tanks. After 4 tanks, the maximum number observed is 140, so the single most likely answer is that there are 140 tanks.   The location of that spike changes as the maximum observed serial number changes for different observations.

A person might then estimate that there are 140 tanks, but even though that number is the most likely answer, it is not the best estimate since it is almost definitely under-predicting the number of tanks.

If we take a weighted average of the number of tanks (i.e. multiply 140 tanks by its final probability, 160 tanks by its final probability, 180 by its

probability, etc., and sum that result) we get a total of 193.3 tanks, which is a good estimate for total number of tanks.

If we had assumed 2,000 tanks at the beginning, the weighted average would be 195.4 tanks, which is substantially the same result

Example 6 – Drug Testing

This problem will show how Bayes Theorem can give results that might be surprising.

You are testing for a certain drug. You know that .5% of the population uses this drug. You have a test that will produce 99% true positive results for users, and 98% true negative results for non-users. You randomly test 1 person from the population and get a positive result. What is the probability that they are actually a user of the drug?

Follow Up Parts To The Question

If the results from multiple tests are independent of the other tests (i.e. you aren't likely to get an incorrect result over and over again on the same person because they have, for instance, weird blood), then if you randomly test 1 person and get 1 positive result, then 1 negative result, what is the probability that they are actually a user of the drug? How about 2 positive results?

Initial Probability

Like all of the Bayes problems, we will start with initial probabilities. In this case, for our random individual, the initial probability that they are a drug user is .5%, and the probability that they are not a drug user is 99.5%.

The next step is conditional probability. There are two possibilities, the subject uses the drug or they do not.

- If the subject tested uses the drug, the chance the test will report a positive (true positive) is .99, and the chance the test will report a negative (false negative) is .01.
- If the subject does not use the drug, the chance the test will report a positive (false positive) is .02 and the chance it will report a negative (true negative) is .98.

Putting this in Excel is simple, we need an IF statement to check if the test reported a positive or a negative, and then we need to multiply the probability from the previous step by the conditional probability for both the drug user and

non-user possibilities.

Likelihood Function For Drug User Possibility

=IF($B18=1,$D$11*H17,(1-$D$11)*H17)
Test Result   .99   Prior   1-0.99   Prior

	B	C	D	E	F	G	H	I
		True Positive	0.99					
		True Negative	0.98					
			Non-Normalized				**Normalized**	
		Person's Probabilities	Drug User	Non-User			Drug User	Non-User
	Test Result 1 = Positive Result 0 = Negative Result	Initial Probability	0.005	0.995			0.0050	0.9950
	1	Test 1	=IF($B18=1,$D$	0.019900		Test 1	0.1992	0.8008

Likelihood Function For Non-User Possibility

=IF($B18=1, (1-$D$12)*I17, D12*I17)
Test Result   1 - 0.98   Prior   0.98   Prior

	B	C	D	E	F	G	H	I
			IF(logical_test, [value_if_true], [value_if_false])					
		True Positive	0.99					
		True Negative	0.98					
			Non-Normalized				**Normalized**	
		Person's Probabilities	Drug User	Non-User			Drug User	Non-User
	Test Result 1 = Positive Result 0 = Negative Result	Initial Probability	0.005	0.995			0.0050	0.9950
	1	Test 1	0.004950	, (1-D12)*I17, D:		Test 1	0.1992	0.8008

Once the resulting probabilities are normalized, we see that the probability that this random individual who tested positive was actually a drug user is only 19.9%. This result surprises many people, because after all, the 99% accurate test should mean that the subject tested has a 99% chance to be a drug user if the test comes back positive, right? This turns out not to be the case. Since the initial probability was so small, only .5%, even a large increase in that probability wasn't enough to make it a really high probability

Most people's intuition fails to account for the initial probability. Even if the conditional probability is really high, a very low initial probability can result in a low final probability. Most people's intuition is tuned around the initial probability being 50 / 50. If that is the case, the result ends up being a 98% chance that the positive test result means that the subject was a user of that drug, as shown below

| | True Positive | 0.99 | | | | | |
| | True Negative | 0.98 | | | | | |

		Non-Normalized				Normalized	
	Person's Probabilities	Drug User	Non-User			Drug User	Non-User
Test Result							
1 = Positive Result							
0 = Negative Result	Initial Probability	0.5	0.5			0.5000	0.5000
1	Test 1	0.495000	0.010000		Test 1	0.9802	0.0198

Follow Up Questions

Going back to the initial problem, where we had an initial probability of .5% that a randomly tested person was a user of the drug, what are the odds if we get either 1 positive and 1 negative, or 2 positives ?

With 1 positive and one negative, the probabilities end up being pretty close to the initial probability, there is a 99.75% chance that the subject is not a user of the drug. (If the true positive and true negative reliability percentages were the same, then a True & False result would exactly cancel each other out, but since they are slightly different we don't end up with the exact initial probabilities)

| | True Positive | 0.99 | | | | | |
| | True Negative | 0.98 | | | | | |

		Non-Normalized				Normalized	
	Person's Probabilities	Drug User	Non-User			Drug User	Non-User
Test Result							
1 = Positive Result							
0 = Negative Result	Initial Probability	0.005	0.995			0.0050	0.9950
1	Test 1	0.004950	0.019900		Test 1	0.1992	0.8008
0	Test 2	0.001992	0.784789		Test 2	0.0025	0.9975

If we have 2 positive results, we end up with a 92.5% chance that the subject is a user of the drug

| | True Positive | 0.99 | | | | | |
| | True Negative | 0.98 | | | | | |

		Non-Normalized				Normalized	
	Person's Probabilities	Drug User	Non-User			Drug User	Non-User
Test Result							
1 = Positive Result							
0 = Negative Result	Initial Probability	0.005	0.995			0.0050	0.9950
1	Test 1	0.004950	0.019900		Test 1	0.1992	0.8008
1	Test 2	0.197203	0.016016		Test 2	0.9249	0.0751

For this problem, the most important numbers end up being the .5% initial chance of the randomly tested subject being a user of the drug, and the 98% change that a non-user will test negative. Since out of the general population only 5 people in 1000 actually use the drug, if you were to administer the test to 1000 people you would (approximately) get the 5 actual users showing up as positive results, and you would also get (approximately) 20 non-users showing up as positive results. So of the 25 people who tested positive, only 5 of them, or 20% would actually be users of the drug. This is why we got the 19.92% number after 1 test using Bayes theorem. (The slight difference between 20% and 19.92% is because we would get slightly less than 5 positives in 1000 due to the 99% true positive rate, and we would get slightly less than 20 non-users showing positive results from multiplying 2% by 995 non-users.)

If we were to change the 98% True Negative rate to something much higher, like 99.99%, so that instead of 20 false positives in 1,000 tests we got 1 false positive in 10,000 tests, then the chance that a random person is a user of the drug after 1 positive test becomes much higher, around 98%

More Books

If you liked this book, you may be interested in checking out some of my other books such as

- [Machine Learning With Random Forests and Decision Trees](#) – This is a detailed walk through of how the machine learning algorithms Random Forest and Decision Tree work. It uses some Python code, but is more focused on examples and analogies to give you an intuitive understanding
- [Excel Pivot Tables: Crunch Large Amounts of Data](#): This goes through how to use the pivot tables in Excel to quickly filter and sort data. Pivot tables are more powerful than data filters, and much, much more powerful than trying to crunch the data manually, and they are not that difficult to use.
- [Hypothesis Testing Examples](#) – This goes through examples of calculating statistical significance using Z tests and T-tests

Another good book on Bayes Theorem that is more technical than this one, and deals with some problems that need more complicated priors or more complicated probability functions is [Think Bayes](#).

If you'd like some humorous examples of Bayes Theorem in action, check out

https://xkcd.com/1132/

https://what-if.xkcd.com/65/

And if you want to get any of the Excel files shown in this book, they can be downloaded at http://www.fairlynerdy.com/bayes-theorem-examples/ for free

Thank You

Before you go, I'd like to say thank you for purchasing my eBook.   I know you have a lot of options online to learn this kind of information.   So a big thank you for downloading this book and reading all the way to the end.

If you like this book, then I need your help.   **Please take a moment to leave a review for this book on Amazon.**   It really does make a difference, and will help me continue to write quality eBooks on Math, Statistics, and Computer Science.

If you want to keep up to date on any new eBooks, examples, or cheat sheets you can find us on Facebook at

https://www.facebook.com/FairlyNerdy

or interact with us at our home page

http://www.FairlyNerdy.com