Probability for Real People

Contents

1 Probability for Real People

1.1 Can we rationally reason?

Some might wonder if this section header even makes sense! A decision maker, such as as CFO, looks at it and knows everyone has a reason, some good and some not so good. But sometimes our reasons are not founded in any data that can be observed either by ourselves or by others. Our reasoning can be founded on falsified data, delusions, unfounded opinions, beliefs with no authoritative grounds. Her questions move from the various relationships in production planning in the previous part of this project to the forecasting of what might happen at all. What would be the rationale for a forecast? So what does our CFO decision maker do? She begins her thought experiment with the weather in her hometown of Albany, New York.

Rationality in the context we have been discussing at least means that we, as decision makers, would tend to act based on the consistency of observed reality with imagined models of the real world and through ideas about the world in which data are collected. We attempt to infer claims about the world based on our beliefs about the world. When confronting ideas, imbued with beliefs, with ovbserved reality we might find ourselves in the position to *update our beliefs*, even those, and sometimes especially those, we so dearly hold.

In our thinking about anything we would venture candidate hypotheses h about the world, say the world of meeting demand, providing services, marshalling resources in specific markets. Of course the whole point is that we do not know which hypothesis h is more plausible, or not. We then collect some data d. When we perform this task, we move from the mental realm of the possibiity of hypotheses, theories, surmises, and model to the realm of observed reality. We may well have to revise our original beliefs about the data.

To implement our maintained hypothesis of rationality, we begin our search for **potential consistencies of the collected data with our hypotheses** that are fed by the data. In our quest we might find that some one of the hypotheses has more ways of being consistent with the data than others. When the data is consistent with a hypothesis, that is, when the hypothesis is reasonable logically, then our belief in that hypothesis strengthens,[1] and becomes more plausible. If the data is less consistent with the hypothesis, our belief in that hypothesis weakens. So far we have performed this set of tasks with conjectures about virus testing and voter alliance in zip codes. Let's switch up our program and consider the following very simplified question about the weather.

The CFO's ask of the analysts is this question about winter consumer behavior in the Albany-area market:

> *We see people carrying snow shovels. Will it snow?*

[1] The core idea of *strengthen* is to take us from a more vulnerable to a less vulnerable place or state. Synonyms for strength include *confirm* and *validate*.

What is the data d? We have recorded a simple observation about the state of the weather so that single piece of data (d = We see people carrying snow shovels). Here is where our beliefs enter. We have two ***hypotheses***, h: either it snows today or it does not.

Let's figure out how to solve this problem? We have three *desiderata*:

1. We should include our experiences with snow in our analysis.
2. We should collect data about carrying snow shovels in January as well.
3. We prefer more consistency of data with hypotheses to less consistency.

Here we go, let's strap ourselves in.

1.1.1 Priors: what we think might happen

Our observation is about the weather: clouds, wind, cold. But we want to know about the snow! That is our objective and we have definite ideas about whether (don't pardon the pun!) it will snow or not. We will identify our beliefs, ever before we make our observations, about snow. The analytical profession and custom is to label these beliefs as *a priori*,[2] and thus the ellipsis *prior*, contentions we hold when we walk into the data story we create with the question of *will it snow?*

Our prior contentions are just the plausibilities of each hypothesis whatever data we eventually collect. After all we have to admit to everyone what we believe to be true as the antecedent to the consequent of observations and the plausibility of snow. This move allows us to learn, to revise, to update our dearly held beliefs. We thus can grow and develop. This is in a phrase a *sine qua non*, a *categorial imperative*, a *virtually unconditioned* requirement for change.

What might we believe about whether it will snow (today)? If you come from Malone, New York, north of the Adirondack mountains, you will have a different belief than if you come from Daytona, Florida, on the matter of how many ways snow might happen in a given month. So let's take as our benchmark Albany, the capital of the state of New York.

We will refer to some data to form hypotheses and their plausibility.using this weather statistics site. The site reports the average number of days of snowfall in January, when there is at least a 0.25 cm accumulation in a day. It is 10.3 days. These are the number of ways (days) in January, in Albany, NY, that it is true, on average and thus some notion of expected, or believed to be, that it snows. The total number of ways snow could possibly fall in any January (defined by calendar standards) is 31. While the formation of the hypotheses *snowy* and *nice* days is informed by data, we are asking a question about snow because we have yet to observe if will snow. We cannot observe something that has not yet happened. We can thus characterize hypotheses and conjectures as **unobserved data**.

Thus we might conclude that we believe that is it plausible (probable) that snow *can* fall $10.3/31 = 30\%$ of the different ways snow can fall. Note very well we will talk about *priors* as *potentials* and *conjectures* and *hypotheticals*, and thus used the modal verbs *can* or *might*. Thus we believe it might not snow, because it is possible, with plausibility $1 - 0.30 = 0.70$, or, multiplying by 100, 70%, according to the law of total probability of all supposed (hypothesized) events. We only have two such events: *snow* and *not snow*. Probabilities must, by definition, add up to 1 and must, again by definition be a number between 0 and 1.

Nice ideas, nice beliefs, are our as yet to be observed, but projected notions of a snowy day. But how real, how plausible, how rational, that is, how consistent are they with any *observed data*? Is there any *observed data* we can use to help us project which of our unobserved data, our hypotheses, is more or less reasonable?

[2]The *a priori* elements of any argument include just about everything you and I know, including the kitchen sink! We can't help but to have these antecedent thoughts, experiences, shared and not-so-shared histories. They tend to persist in most humans, including us. At least that is what we will maintain. Thus it is a *necessity* to include these beliefs in our discussion. Without their consideration we most plausibly will introduce unsaid and denied bias, let blindspots have the same focus as clearly understood experiences, and produce time and resource consuming blind alleys. But we should hang on here: even blind alleys and blind spots are extremely important bits of knowledge that help us understand what does not work, an *inverse insight* as exposed by Bernard @Lonergan_1970.

Table 1: Priors by hypotheses

hypotheses	priors
snowy day	0.3
nice day	0.7

Table 2: data meets hypotheses

hypotheses	shovels	hands
snow day	0.7	0.3
nice day	0.1	0.9

1.1.2 Likelihoods: thinking about the data

Life in the Northeast United States in January much revolves around the number of snow days, also known as days off from school. A prediction of snow meets with overtime for snow plow drivers, school shut downs, kids at home when they normally are in school. On some snowy days we see people carrying snow shovels, on others we don't. On some nice days we see people with snow shovels, on others we don't. Confusing? Confounding? A bit.

Now we link our observations of shovels with our unobserved, but through about and hypothesized, prediction of snow. We then suppose we observe that people carry snow shovels about 7 of the 10 snowy days in January or about 70%. On nice days we observe that people carry shovels at most 2 days in the 21 nice days or about 10%.

This table records our thinking using data we observe in Januaries about weather conditions.

First of all these probabilities register yet another set of beliefs, this time about whether we see shovels or not, *given*, *conditioned by*, the truth of each hypothesis h. We write the conditional probability $\Pr(d \mid h)$, which you can read as "the probability of d given h". Also here we will follow the convention that this set of results of our assessment of the relationship of shovels to snowy days as a likelihood .[3]

1.1.3 Altogether now

Do we have everything to fulfill our *desiderata*? Let's check where we are now.

1. We should include our experiences with snow in our analysis.

   Yes! We put our best beliefs forward. We even (sometimes this is a courageous analytical step) quantified teh ways in which snow and not snow would occur, we believe, in Albany NY in an average January.[4]

2. We should collect data about carrying snow shovels in January as well.

   Yes we did! Again we elicited yet another opinion, belief, whatever we want to colloguially call it. That belief if what we register and document based on observation of shovels and just hands in the presence of snowy and nice days in a January.

3. We prefer more consistency of data with hypotheses to less consistency.

   Not yet! We will impose our definition of rationality here.

Let's start out with one of the rules of probability theory. The rule in question is the one that talks about the probability that *two* things are true. In our example, we will calculate the probability that today is snowy

[3] For Pierre Simon @Laplace_1902, likelihood also has the idea of $\Pr(h \mid d)$. Let's stick to our knitting, and tolerance for ambiguity, with using the rows of this table as our entries for likelihood.

[4] we really need to think further about our notions of an average or centrally located anything. This means more consideration later, including deviations from these locations measured by scale.

(i.e., hypothesis h is true) *and* people carry shovels (i.e., data d is observed). The ***joint probability*** of the hypothesis and the data is written $\Pr(d, h) - \Pr(d \wedge h$, and you can calculate it by multiplying the prior $\Pr(h)$ by the likelihood $\Pr(d \mid h)$. The conjunction is a *both-and* statement. We express conjunctions using the wedge $\wedge$ symbol. Logically, when the statement that both d and h is true, then the plausibility, now grown into probability is:

$$\Pr(d \wedge h) = \Pr(d \mid h) \Pr(h)$$

When we divide both sides by $\Pr(h)$ we get the definition, some say derivation, of condition probability. If we count $\#()$ the ways $d \wedge h$ are true and the ways that h are true then

$$\#(d \mid h) = \frac{\#(d \wedge h)}{\#(h)}$$

Then the number of ways the data d are true, given h is true, equals the total number of ways that d and h per each way that h is true. We have thus normed our approach to understanding a conditional statement like if h, then d. Even more so, when we combine the law of conditional probability with the law of total probability we get Bayes Theorem. This allows us to recognize the dialectical principle that, yes, we recognize $h = snowy$, but we also know that every cloud has its silver lining and that there is a non-snowy day and thus a

$$not\ h = \neg h = nice$$

lurking in our analysis.

Here it in in all its glory.

$$\Pr(h \mid d) = \frac{\Pr(d \mid h)\ \Pr(h)}{\Pr(d \mid h)\ \Pr(h) + \Pr(d \mid \neg h)\ \Pr(\neg h)} \tag{1}$$

$$== \frac{\Pr(d \wedge h)}{\Pr(d \mid h)\ \Pr(h) + \Pr(d \mid \neg h)\ \Pr(\neg h)} \tag{2}$$

The numerator is the same as the conjunction both d and \$h. The denominator is the probability that either both d and h or both d and h are true. While the build up to this point is both instructive, and thus may at first be *confusing*, it is useful as it will highlight the roles these probabilities perform in the drama that is our analysis.

We had better get back to the data or get lost in the weeds of the maths. So, what is the probability it is true that today is a snowy day *and* we observed people to bring a shovel?

Let's see what we already have. Our prior tells us that the probability of a snowy day in any January is about 30%. Thus $\Pr(h) = 0.30$. The probability that we observe people carrying shovels is true given it is a snowy day is 70%. So the probability that both of these things are true is calculated by multiplying the two to get 0.21. We can make this

$$l \Pr(snowy,\ shovels) = \Pr(shovels \mid snowy) \times \Pr(snowy) \tag{3}$$

$$= 0.70 \times 0.30 \tag{4}$$

$$= 0.21 \tag{5}$$

This is an interesting result, something odds makers intuitively know when punters put skin in the game. There will be a 21% chance of a snowy day when we see shovels in people's hands. However, there are of

Table 3: Both data and hypotheses

hypotheses	shovels	hands	sum
snow day	0.21	0.09	0.3
nice day	0.07	0.63	0.7
sum	0.28	0.72	1.0

Table 4: both data and hypotheses in days in January

hypotheses	shovels	hands	sum
snowy day	6.51	2.79	9.3
nice day	2.17	19.53	21.7
sum	8.68	22.32	31.0

course *four* possible pairings of hypotheses and data that could happen. We then repeatthis calculation for all four possibilities. We then have the following table.

Just to put this into perspective, we have for the 31 days in a January this table.

We have four logical possibilities for the interaction of observed data and unobserved hypotheses. We arrange these possibilities in two stacked rows. We recall that visualizatiton is everything, even in tables! Here is the first row.

1. Snowy and shovels

$$
\begin{aligned}
l\Pr(\textit{snowy}, \textit{shovels}) &= \Pr(\textit{shovels} \mid \textit{snowy}) \times \Pr(\textit{snowy}) && (6)\\
&= 0.70 \times 0.30 && (7)\\
&= 0.21 && (8)
\end{aligned}
$$

2. Snowy and just hands

$$
\begin{aligned}
l\Pr(\textit{snowy}, \textit{hands}) &= \Pr(\textit{hands} \mid \textit{snowy}) \times \Pr(\textit{snowy}) && (9)\\
&= 0.30 \times 0.30 && (10)\\
&= 0.09 && (11)
\end{aligned}
$$

In this row the prior probability about snow is 0.30.

Here is the second row with its separate calculations.

1. Nice and shovels

$$
\begin{aligned}
l\Pr(\textit{nice}, \textit{shovels}) &= \Pr(\textit{shovels} \mid \textit{nice}) \times \Pr(\textit{nice}) && (12)\\
&= 0.10 \times 0.70 && (13)\\
&= 0.07 && (14)
\end{aligned}
$$

2. Nice and just hands

$$
\begin{aligned}
l\Pr(\textit{nice}, \textit{hands}) &= \Pr(\textit{hands} \mid \textit{nice}) \times \Pr(\textit{nice}) && (15)\\
&= 0.90 \times 0.70 && (16)\\
&= 0.63 && (17)
\end{aligned}
$$

In this row the prior probability about nice days is 0.70.

An insightful exercise is to carry these calculations from the number of ways snow with and without shovels occurs given we think we know something about snow. The same with the number of ways a nice day might occur with and without shovels, given what we think about nice days.

Let's put one calculation together with a not so surprising requirement. When we conjoin snow with shovels, how many possible ways can these logical statements occur? It is just the 31 days.

We now have all of the derived information to carry our investigation further. We also total the rows and, of course, the columns. We will see why very soon.

The row sums just tell us as a check that we got all of the ways in which snow occurs in 31 days. What is brand new are the column sums. They add up the ways that data occurs across the two ways we hypothesize that data can occur: snow, no snow (nice day). They tell us the probability of carrying a shovel or not, across the two hypotheses. Another way of thinking about the $p(d)$ column sums is that they are the expectation of finding snow or hands in the data. The consistency of all of these calculations is that column sums equal row sums, 100%. All regular, all present and correct, probability-wise.

1.1.4 Updating beliefs

The table lays out each of the four logically possible combinations of data and hypotheses. So what happens to our beliefs when they confront data? In the problem, we are told that we really see shovels, just like the picture from Albany, NY at the turn of the 20th century. Is surprising? Not necessarily in Albany and in January, so you might expect this behavior out of habit during a rough Winter. The point is that whatever our beliefs have been about shovel behavior, we should still subject them to the possibility of accomodating the fact of seeing shovels in hands in Albany in January, a winter month in the Northern Hemisphere.

We should recall this formula about the probability of seeing both an hypothesis and data:

$$\Pr(h \mid d) = \frac{\Pr(d \wedge h)}{\Pr(d)} = \frac{\Pr(d \mid h)\Pr(h)}{\Pr(d)}$$

Now we can trawl through about our intuitions and some arithmetic. We worked out that the joint probability of *both* ***snowy day*** *and* ***shovel*** is 21%, a rate reasonable given the circumstances. In our formula, this is the product of the likelihood $\Pr(d = \textit{shovels} \mid h = \textit{snow}) = 0.70$ and the prior probability we registered that snow might occur $\Pr(h = \textit{snow}) = 0.30$.

Relative to the product of the likelihood of shovels given a nice day and the chance that snow might occur is the the joint probability of *both* ***nice day*** *and* ***shovel*** at 10%, or $\Pr(d = \textit{shovels} \mid h = \textit{nice})\Pr(h = \textit{nice}) = 0.10 \times 0.70 = 0.07$, again a reasonable idea, since we plausibly wouldn't see much shovel handling on that nice day in January..

Both of these estimates are consistent with actually seeing shovels in people's hands. But what are the chances of just seeing shovels at all? This is an ***either or*** question. We see shovels 21% of the time on snowy days or we see shovels 7% of the total days in January on nice days. We then add them up to get 28% of the time we see shovels in all of January, whether it snows or not.

So back to the question: if we do see shovels in the hands of those folk, will it snow? The hypothesis is $h = \textit{snow}$ and the data is $d = \textit{shovels}$. The joint probability of both snow and shovels is $\Pr(d, h) = 0.21$. But just focusing on the data we just observed, namely that we see shovels, we now know that the chances of seeing shovels on any day in January in Albany, NY is $\Pr(d) = 0.27$. Out of all of the ways that shovels can be seen in January then we would anticipate that the probability of snow, upon seeing shovels, must be $\Pr(h \mid d) = \Pr(d, h)/\Pr(d) = 0.21/0.28 = 0.75$.

What is the chance of a nice day given we see shovels? It would be again likelihood times prior or $0.10 \times 0.7 = 0.07$ divided by the probability of seeing shovels any day in January 28%. We then calculate $0.07/0.28 = 0.25$. We now have the posterior distribution of the two hypotheses, snow or nice, in the face of data, shovels. So what are the odds in favor of snow when we see shovels?

Table 5: Unobserved belief tempered by observed data = posteriors.

hypotheses	shovels	hands	priors	posterior shovels	posterior hands
snow day	0.7	0.3	0.3	0.75	0.125
nice day	0.1	0.9	0.7	0.25	0.875
sum	0.8	0.2	1.0	1.00	1.000

$$OR(h \mid d) = \frac{\Pr(h = snow \mid d = shovels)}{\Pr(h = nice \mid d = shovels)} \tag{18}$$

$$= \frac{0.75}{0.25} \tag{19}$$

$$= 3 \tag{20}$$

We can read this as: when we see people with shovels in January in Albany, NY, then it is 3 times more plausible to have a snowy day than a nice day. The ratio of two posteriors gives us some notion of the plausible divergence in likely outcomes of snowy versus nice days. Again we must append the circumstances of time and place: in a January and in Albany, NY.

Here is table that summarizes all of our work to date.

1.2 Enough with the shovels!

Let's apply the probability analysis we developed so far to understanding the plausibility of the demand for pies at one of the smaller restaurants the vegan pie maker Make-A-Pie LLC vends to. Here is a day by day count of the number of pies sold (the *count* variable) and the weather (clear coded as C or rainy coded as R). The original data only has the date, count, and the weather. The rest is analysis based on this data.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
2														
3		date	count	clear	weekday	day_of_week		day_index	day					
4		2021-12-27	11	R	2	Monday		1	Sunday					
5		2021-12-28	32	C	3	Tuesday		2	Monday					
6		2021-12-29	21	C	4	Wednesday		3	Tuesday					
7		2021-12-30	19	C	5	Thursday		4	Wednesday					
8		2021-12-31	13	C	6	Friday		5	Thursday					
9		2022-01-01	2	R	7	Saturday		6	Friday					
10		2022-01-02	1	R	1	Sunday		7	Saturday					
11		2022-01-03	10	C	2	Monday								
12		2022-01-04	9	C	3	Tuesday			E4:					
13		2022-01-05	3	R	4	Wednesday			=WEEKDAY(B4)					
14		2022-01-06	6	C	5	Thursday			F4:					
15		2022-01-07	19	C	6	Friday			=INDEX(day, MATCH(E4, day_index, 0),)					
16		2022-01-08	25	C	7	Saturday								
17		2022-01-09	26	C	1	Sunday								
18		2022-01-10	10	C	2	Monday								
19		2022-01-11	2	C	3	Tuesday								
20		2022-01-12	1	C	4	Wednesday								
21		2022-01-13	22	R	5	Thursday								
22		2022-01-14	21	C	6	Friday								
23														

Figure 1: Daily Pie demand and the weather.

After carefully verifying and validating the data and its ssources, the next step in spreadsheet engineering is always to create named ranges. Here they are the names of the columns.

To derive the days of the week in a potential analysis, we can create a table of the day number, with 1 as Sunday. Then the day of the week would be an INDEX() lookup of the date's corresponding day of the week

in the table. The result is the day name. With this we can possibly sort out all Saturdays and Sundays to isolate activity on weekdays and weekends.

For our purposes we would like to know the odds of selling more than 10 pies when the weather is clear. We can answer this query by making a 2x2 contingency table of the conditional data counts of pie demand. This panel details the computations.

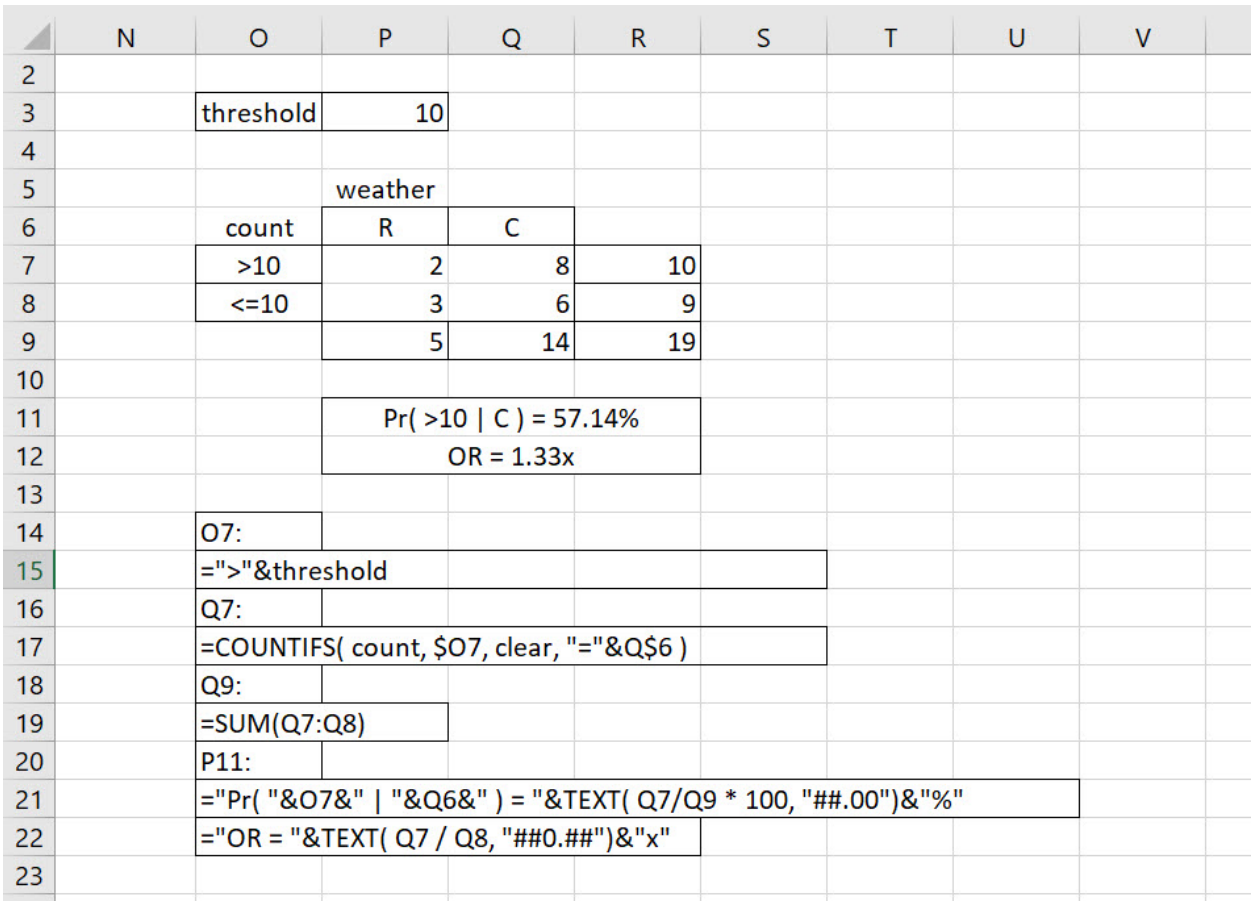

Figure 2: What are the odds?

The key is the associative counting of pies by *both* a count greater than a threshold of 10 *and* clear weather. We accomplish this, for example in cell Q7, with =COUNTIFS(count, $O7, clear, "="&Q$6). We can read this statement as count the pies (count) greater than 10 ($O7) *and* with weather (clear) which is indeed clear (Q$6). This yields 8 pies. When the weather is clear we sell a total of 14 pies.

The probability of selling more than 10 pies when the weather is clear is about $8/14 \times 100 = 57\%$. We sold 8 pies greater than the threshold and 6 pies less than the threshold when the weather was clear. The ratio of $8 : 6 = 1.33$ reports the odds of selling greater than 10 pies. We interpret this number with the phrase: it is 1.33x more likely to sell greater than 10 pies, than not, when the weather is clear.[5]

1.3 What did we accomplish?

We have travelled through the complete model of probabilistic reasoning.

[5]Some folks might want to aggrandize this statement by saying that is is 33% more likely to sell more than 10 pies when clear. We might caution ourselves as we might be falling into a distortion by magnifying an apparent difference. The probability of selling more than 10 pies when clear is $8/14 = 4/7$. on the other hand the complementary probability of selling up to 10 pies when clear is $6/14 = 3/7$. We get to sell only 1 pie in crossing the 10 pie threshold given this data set. An inflated way of expressing the results of this analysis? Isocrates would agree.

1. We started with a question. The question at least bifurcates into the dialectical ***is it?*** or ***is it not?***.
2. We then began to think about beliefs inherent in the question for each of the hypotheses buried in the question.
3. We then collected data that is relevant to attempting an answer to the question relative to each hypothesis.
4. Then we conditioned the data with the hypotheses inside the question. It is always about the question!
5. Finally we derived plausible answers to the question.
6. We then illustrated the example with data collected, threshold set, table constructed, odds computed.

What is next? We continue to use this recurring scheme of heuristic thinking, sometimes using algorithms to count more efficiently, applied to questions of ever greater complexity. In the end our goal will be to learn, and learning is inference.

1.4 Some exercise is in order

Exercise 1.1. Start with a question for analysis using a indicative-interrogative statement format, for example "We observe X. Will Y occur?" Based on this statement identify the unobserved data of the hypothesis and the observed data. Use binary hypotheses and observations.

Exercise 1.2. Rework the Albany NY example using your hometown or city. Develop initial distribution of hypotheses, distributions of data given a hypothesis, joint distributions of hypotheses and data. Find the probability that a particular hypothesis might occur given a specific piece of data.

Exercise 1.3. Rework the spreadsheet example using various thresholds. Use the Data Table sensitivity analysis to calculate results. Plot the Data Table. Lead a discussion with some skeptical people about your findings.