Contents

1 Will You Make Money?

> No one can win at roulette unless he steals money from the table while the croupier isn't looking.

— Albert Einstein (possibly)

The development of probability theory is historically linked to attempts to understand games of chance, especially ones in which money was involved (see for example, here. Sometimes betting money on an uncertain outcome falls under the name of gambling; other times it's dignified with the name investment or "smart business decision." But regardless of the label, there are smarter and less smart ways to play money games.

Battle of the Bills

Let's recall a distinction we made earlier in this course about deterministic and stochastic, or random, processes. This time, we'll think about two different bets you make with your friend. In the first bet, you and your friend are debating whether it was Bill Paxton or Bill Pullman in the movie Apollo 13. To make the game interesting, you bet two dollars. You look it up on the internet, and find that it was indeed Paxton. One of you wins. Do you feel the need to check again? Probably not. This particular question, although you may not have known the answer for sure, has only one possible answer.

Now consider another bet, this time for three bucks! You and your friend are walking down the street debating the "merits" of mint chocolate chip vs. cookies and cream as ice cream flavors. You claim that mint chocolate chip is the more popular flavor, and decide to ask the first passer-by which flavor they think is better. Suppose they don't just ignore you, thinking you're a nutcase, and they answer cookies and cream. Are you satisfied with this one answer? Or do you feel the need to ask another pedestrian? And how many?

We might say that the variable "BPA", which stands for "which Bill P. starred in Apollo 13?" has a deterministic answer, but the variable "MCCoCAC," which stands for "mint chocolate chip > cookies and cream?" can take on one of two answers (no ties allowed) depending on whom we ask. Because it is a random or stochastic variable, we have to talk about it using different terms. We might say something like, what proportion of people (in this neighborhood, say) prefer cookies and cream? Or what are the chances that the first person we ask will express that particular preference.

This may all sound like silly bets that are really just games between friends. But people make small and large money bets all the time, in everything from business and life decisions, to recreational games. In this chapter, we explore probability calculations that inform things like advertising, airplane booking, the job market, and march madness.

1.1 Betting on Beer (or Ice Cream)

This section makes reference to Chapter 5 of Naked Statistics by Charles Wheelan

In 1981, Schlitz brewing company, now defunct but at one time the largest beer producer in the US, ran a bold advertising campaign. During the Super Bowl, Schlitz ran a live blind taste test against one of its competitors, Michelob. 100 Michelob drinkers participated in the taste test, which aired LIVE. The advertisement slot itself cost a lot of money. Schlitz could have just run a funny ad involving puppies on the beach, so why take

a risk with a taste test that could conceivably have gone badly. How could Schlitz have been so confident that their beer would be preferred?

> THINK ABOUT IT QUESTION: What information would you need to know to advise the Schlitz brewing company about running such an ad? (Take a few minutes before continuing on, to try to list this information on your own).

As discussed in Wheelan's chapter, some things we would need to know are:

- Actual proportion of Michelob drinkers who would prefer Schlitz in a blind taste test
- Acceptable outcome of live taste test for promoting Schlitz beer
- Intended sample size for taste test
- Rules of mathematical probability

Wheelan adds a lot of context to this particular story, which is part of the fun. In particular, he asserts that Schlitz and Michelob are probably indistinguishable to most beer drinkers. This puts the chances of anyone prefering one beer over the other at 50%. He also points out that the marketing campaign works well even if that is not the observed outcome, because the taste test is conducted with Michelob drinkers. Schlitz executives will be quite happy to be able to say that 40% of Michelob drinkers prefer Schlitz, which sounds (and is) very different from saying 40% of all beer drinkers prefer Schlitz over Michelob.

Wheelan invokes the "law of large numbers" to argue that for a given sample size, and if the actual proportion is 50%, that the results of the live taste test can be almost guaranteed to be satisfactory for Schlitz (at least 40% preferred). The larger the sample size, the greater the probability that the taste test will be a success. We have created a Schlitz simulation for you to explore this for yourself.

In his book, Wheelan claims that (a) for 10 blind tast testers, the probability of a happy outcome is 0.83 and (b) for 100 blind taste testers, the probability is 0.98. If you don't want to take this assertion at face value, you might try convincing yourself by opening the simulation, running 100 simulated experiments of sample size 10 or 100, and inspecting the proportion of those experiments that led to a favorable outcome. You should see values around .83 and .98 for sample sizes of 10 and 100, respectively.

For a moment, let's pull back the curtain on the Schlitz simulation and see how it works. The following code walks through the process of repeatedly surveying 10 people, recording the proportion who preferred Schlitz (under the assumption that each person has a 50% chance of preferring Schlitz), and calculating the proportion of those 10-person surveys that led to an acceptable outcome. If we collect 10,000 samples of 10 people and calculate the proportion of those 10 person samples where at least 4/10 people preferred Schlitz, we can estimate the probability of an acceptable outcome very accurately:

```
nIter = 10000 #set some number of repeated experiments to run
SampSize = 10 #set the sample size
Prob = .5 #set the probability of preferring Schlitz
Acceptable = .4 #set an acceptable proportion of Schlitz preferrers

results = vector(length=nIter) #create a vector of length nIter

for(i in 1:nIter){ #repeat the following process nIter times
   #Choose SampSize values from the set (0,1) with replacement
   #where the probability of drawing a 1 is equal to Prob
   #save the results in a vector called Samp
   Samp = sample(c(0,1), size=SampSize, prob=c(1-Prob, Prob), replace=TRUE)

   #In the ith location of "results", calculate the proportion of 1s in Samp
   results[i] = sum(Samp)/SampSize
}

#Calculate the proportion of random experiments that were "acceptable"
sum(results>=Acceptable)/nIter
```

```
## [1] 0.8329
```

Feel free to copy this code over into your own script in R Studio and play with the parameters to see what happens. If you decrease nIter to 1000 and re-run the simulation a few times, you might see that there is more variation in the estimated probability; however, if you increase nIter to 100000, you are more likely to observe values very close to .83 every time.

How do statisticians solve problems like this?

In this book, I have tried to emphasize conceptual understanding through simulation and discussion. In the example above, you can, for example, run a bunch of simulations of the experiment and (very accurately) estimate the probability of an acceptable outcome. But, you'll get slightly different answers each time you run the simulation. If this bothers you, read on.

Mathematical statistics does have precise answers that depend on properties of continuous distributions like the normal distribution and the binomial distribution. The Schlitz commercial is exactly the kind of scenario that is explained using a binomial distribution (more on that later). If we ran the simulation (always with samples of 10), taking more and more observations (i.e., 10 samples of 10, 100 samples of 10, etc.) and checked our success rate (defined by at least 4/10 preferring Schlitz), we would see that indeed this proportion does converge. This is plotted in Figure 1. The x-axis is the number of samples, but **the x-axis is shown using logarithmic scales**. We need to use this scale, or else all of the points at smaller values would be bunched together.

```r
set.seed(1234567)
sample_size <- 10
min_success <- 4
prob_success <- 0.5

lognobs <- seq(2,12,by=0.25)
n_obs <- exp(lognobs)
obs_prop <- c()
for (n in n_obs) {
   tmp <- rbinom(n, sample_size, prob_success)
   obs_prop <- c(obs_prop, length(which(tmp >= min_success))/length(tmp))
}
plot(n_obs, obs_prop,
     xlab="number of observations",
     ylab="proportion of successes",
     log="x", type="l")
abline(h=sum(dbinom(min_success:sample_size,sample_size,prob_success)), col=2, lty=2, lwd=2)
```

```r
#
# plot(n_obs, obs_prop,
#     xlab="number of observations",
#     ylab="proportion of successes", type="l")
# abline(h=sum(dbinom(min_success:sample_size,sample_size,0.65)), col=2, lty=2, lwd=2)
```

So we see that there's some convergence: If we run more and more experiments, we find that the proportion of "successful" experiments converges to a stable value. But how can you calculate that value precisely? To find the empirical (i.e., exact) answer, instead of using a simulation to estimate it, it might be helpful to consider a smaller sample size, say 2. Now that we've reduced our scope, we have some hope of writing all possible outcomes of this experiment and their probabilities.

So what are all of the possible outcomes of 2 independent taste tests? There are two possibilities for each taste tester (they can choose either Schlitz or Michelob), so if we represent each possible outcome as (taste-tester 1's choice, taste-tester 2's choice), we get 4 possibilities (in many probability textbooks, possible outcomes of
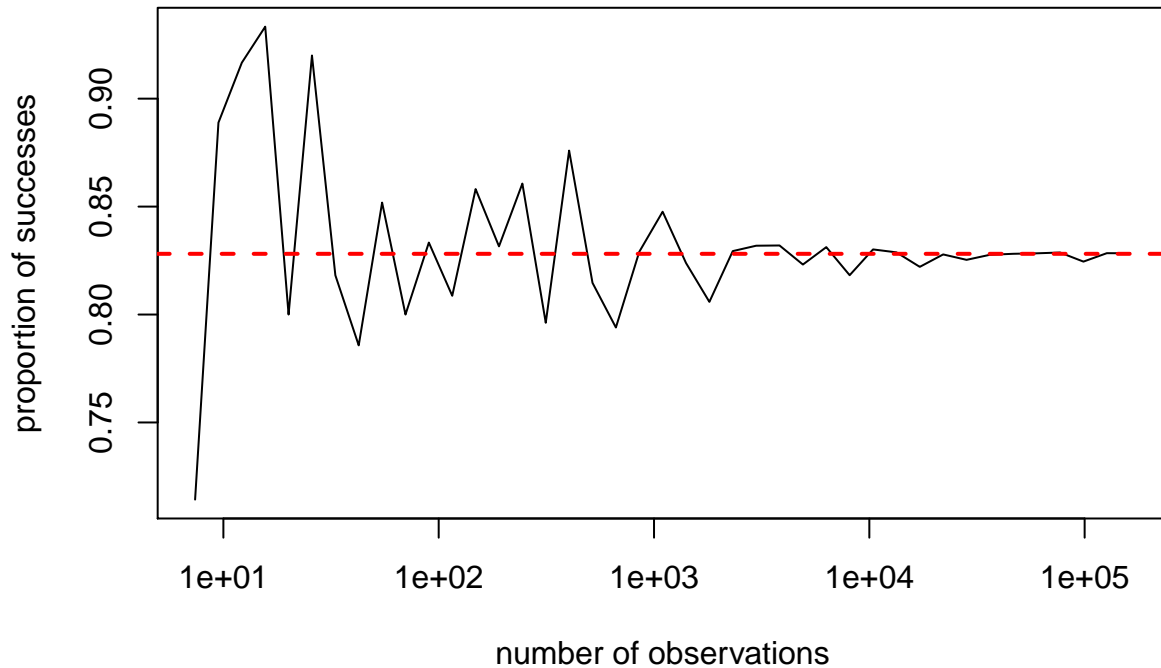
Figure 1: Convergence of successful tasting proportions

a random experiment are represented as sets, which are enclosed in curly brackets): {(Michelob, Michelob), (Michelob, Schlitz), (Schlitz, Michelob), (Schlitz, Schlitz)}. This is also called the **sample space** of the random experiment.

At this point, it might be useful to look back at Section **??**. If we run this experiment with two taste-testers, the outcomes listed above are both disjoint (only one can occur) and complete (one of them *MUST* occur). Therefore, we know that their probabilities must add up to 1. And, because each taste test is independent from all other taste tests (the first person's beer choice does not have any impact on the second person's beer choice), we already know how to calculate the probability of any outcome! For example, the probability of (Michelob, Schlitz) is:

```
P(Michelob, Schlitz) = P(taster 1 prefers Michelob) * P(taster 2 prefers Schlitz)
```

If we assume that every taste tester is equally likely to prefer either beer, then this is simply (.5)*(.5)=.25. In other words, there is a 25% chance that the outcome of the experiment is that the first taste-tester prefers Michelob and the second taste-tester prefers Schlitz. In fact, assuming a 50% chance of preferring either beer, we'll get the same probability for any of the four outcomes.

Now we can ask: which of the four possible outcomes will meet our requirement that at least 40% of respondents prefer Schlitz? Again, probability textbooks often use the word **event** to describe some subset of the sample space for a random experiment. In this case, we could describe the event that at least 40% of respondents prefer Schlitz as the set: Acceptable = {(Michelob, Schlitz), (Schlitz, Michelob), (Schlitz, Schlitz)}. If the probability of each outcome is 0.25, then because the outcomes are disjoint, the probability of "Acceptable" is

```
P(Acceptable) = P(Michelob, Schlitz) + P(Schlitz, Michelob) + P(Schlitz, Schlitz)
              =0.25 + 0.25 + 0.25 = 0.75
```

> Excercise: How would the probability of an acceptable outcome change if we believed that each
> test-taster only had a 40% chance of preferring Schlitz?

After all this work, you might still be thinking: well that's all fine and good, but it's a lot of work to write out the sample space and list of acceptable outcomes for a sample size of 10 or even 100. The truth is: you're

right. For larger sample sizes, we need to employ some new techniques and R functions (which are only briefly covered in this text). But, for the sake of completeness, let's briefly examine two ways to conceptualize the problem and calculate the empirical probabilities in R. Don't worry if it doesn't make perfect sense yet; we won't make you do this by hand!

First, if taste-testers are equally likely to prefer Schlitz or Michelob, we can calculate the probability of an acceptable outcome (i.e., at least 40% preferring Schlitz) among 10 taste-testers as follows: Since taste-testers are equally likely to choose either beer, we know that all possible outcomes of this experiment are equally likely. Therefore, the probability of an acceptable outcome reduces down to the number of acceptable outcomes divided by the total number of possible outcomes. The total number of possible outcomes is 2^{10} for a 10 person sample size (to convince yourself: think about how many "types of people" you would observe by asking 10 independent, dichotomous questions). To figure out the number of acceptable outcomes, we can utilize the choose() function. If you've never seen it before: choose(n,k) (often written $\binom{n}{k}$) is the number of possible ways to choose k items out of a group of n total items. In this context, choose(n=10,k=4) can be thought of as the number of unique groups of 4 taste-testers among a total pool of 10 taste-testers (i.e., number of ways that exactly 4/10 taste-testers could prefer Schlitz). Take a look at how we might compute this in R:

```
#calculate choose(n=10,k=4) in R
choose(n=10, k=4)
```

```
## [1] 210
```

```
#calculate choose(n=10,k= all the numbers between 4 and 10)
choose(n=10, k=4:10)
```

```
## [1] 210 252 210 120  45  10   1
```

```
#add up all of the values above using sum() and then divide by 2^10
sum(choose(n=10, k=4:10))/2^10
```

```
## [1] 0.828125
```

> Excercise: Can you modify the above code to calculate the empirical probability of an acceptable outcome for a sample size of 100 (again assuming preference for Schlitz and Michelob are equally likely)?

Finally, if we want to account for different probabilities of preferring Schlitz or Michelob, it's helpful to add one more tool to the toolbox: the binomial probability distribution. The **probability mass function** of the binomial distribution (which can be calculated in R using: dbinom(x,n,p)) gives the probability of x "successes" in n independent random trials, where each random trial has probability of success=p. For example, dbinom(4,10,.5) could be thought of as the probability that exactly 4 out of 10 people prefer Schlitz if the probability of any individual preferring Schlitz is .5. Using this function, we can now repeat the calculation above using some slightly different code:

```
#calculate dbinom(4,10,.5)
dbinom(4,10,.5)
```

```
## [1] 0.2050781
```

```
#add up the probabilities of 4,5,6,7,8,9, or 10 people preferring Schlitz
sum(dbinom(4:10,10,.5))
```

```
## [1] 0.828125
```

> Excercise 1: Can you explain why choose(n=10, k=4)/2^k is equal to dbinom(4,10,.5)?

> Excercise 2: Can you modify the code above to calculate the probability of an acceptable outcome if each taste-tester only has a 40% chance of preferring Schlitz?

All of these calculations might feel a little overwhelming and confusing at this point. If that's the case, don't fear! Instead, revel in the fact that you just got (approximately) the same number using three different conceptualizations of the same problem. The point is: there are many ways to answer probabilistic questions, and simulation can be a powerful tool to side-step advanced probability calculations.

Using probability to make a "good" bet

The Schlitz example feels a bit contrived, at least to me, because in the scenario, the Schlitz executives don't seem to really care about the truth. They just care about what will play well to Super Bowl audiences. Nevertheless, the beer taste test is just one example of a procedure in which we sample from a population to make inferences about the whole population.

You saw–using the simulation—that whenever you collect data from a sample, you get slightly different results. In this case, you observe a **sampling distribution** in the observed proportion of Schlitz-preference. It had that bell-shaped curve . I want to show you that the same simulation can be used to help you resolve your bet with your friend about whether most people prefer mint chocolate chip or cookies and cream as flavors of ice cream.

Let's now return to the ice cream bet that you made with your friend. If you're willing to make a bet, that probably means don't believe that the flavors are equally preferred. You believe that the *true* value is greater than 50% in favor of mint chocolate chip. By true value, I mean the answer you would get if you could literally ask everyone in the world this question or, to save time, if you could consult the all-knowing-one and just ask them. This conviction is important, because if both you and your friend believe the proportion really is 50/50, then you are just betting on a coin toss. Which is cool, if you want to do that.

Suppose you believe that the true value is 65%. Just about two out of three people prefer mint chocolate chip. Your friend thinks the edge goes slightly to cookies and cream, but not enough to notice, and what's the point of betting on a coin toss. You want to make a point, though, so you're willing to go out on a limb. You say, "Friend, I will give you 3 to 1 odds on this bet. If you lose, you pay me $3, and if you win, I will pay you $9." It works. Your friend becomes interested.

Now that the stakes have been raised, you and your friend start negotiating terms. You both agree that people in Washington Square, for the purposes of this bet, represent an **unbiased sample**. You are not, for example, offering to poll people at the I-heart-Mint festival. You have limited time, but you agree to ask 10 people (you can each pick 5, just to make it fair), and everyone has to pick one preference (no ties). If at least 6 out of 10 prefer mint chocolate chip over cookies and cream you win; 5 or fewer and your friend wins. These are the terms.

Assuming you're right about the true proportion, will you make money?

By now you know that you are still not guaranteed to observe proportions of 6.5/10 because (a) you can only observe 6 or 7, not 6.5, in a sample size of 10 (no ties!) and (b) because there is variance in the sampling distribution. You want to know the probability of winning the bet, so you can decide whether you gave your friend good odds or maybe you were too impulsive.

If you use the simulation tool and take 10-person samples one-at-a-time, you might find something like this: 6, 6, 7, . . . so far so good!. . . 5, 8, 5, . . . uh oh, you would have lost 2 out of 6. If you use the Run 100 times feature of the simulation, you might get 70%, 72%, 78%, 75%, 80%, . . . so it looks like your chances are maybe around 75%. You reason as follows: I have a 75% chance of winning $3 and a 25% chance of losing $9.

> Excercise: Can you calculate the exact probability using dbinom() in R?

The weighted average of these two outcomes is called the **expected value** of the bet. That is, each outcome has a probability and a return (including possible loss) in dollars. You "weigh" each return by the probability of it occurring. This is done by multiplication.

```
Expected Value = P(outcome 1) * Return(outcome 1) + P(outcome 2) * Return(outcome 2)
              = 0.75 * 3 - 0.25 * 9 = 0
```

What does it mean that your expected return for your ice cream bet is 0? For one, it means that you are just as likely to win money as to lose money. In other words, you have proposed a fair bet. Which is the nice thing to do, since after all this is a friendly bet. Your friend originally thought that the two flavors were equally preferred. Given the odds you offered, what was the expected value from your friend's point of view? (Answer in the footnote).[1].

> Exercise: What would the expected value from your and your friend's perspective have been if the odds given were 2 to 1 instead of 3 to 1 on a $3 bet?

If the bet described above took place, there would be one outcome. Either you win $3 from your friend or they win $9 from you. The expected value of 0 will never actually occur. But, as we often like to do, we can imagine 100 (or one million) alternative universes to ours. Assuming the probability calculations above, you win the bet in roughly 75% of them and lost the bet in 25% of them. This is what we meant by the probability in the first place. The expected value can be thought of as your average earning across all of these multiple universes.

You should have noticed that the range of observed proportions is the true value "plus or minus" some amount of variation due to sampling. Furthermore, the variation in proportion got smaller as the sample size got larger. The result of this is that you had a 75% of observing a majority when the sample size was 10. But if the sample size had been larger, say 100, then the chance was much higher. This reduction in the error due to increased sample size is related to a statistical concept called **power.** For the example above, your test was "more than half prefer mint chocolate chip." Your assumed true value for the proportion was 0.65. With a sample of 10, your power was 0.75. Note that your power came from two sources. One was (a belief) that the difference you wanted to detect, the **effect** of mint chocolate chip preference, was reasonably large. 0.65 is not the same as 0.52 or 0.91. It would be easier to detect a majority in the latter case (high power) and harder in the former case (low power). But for small and large effects, a large enough sample size can increase your power.

1.2   Betting on your Future

Consider the diagram on page 83 of the Naked Statistics chapter that you read at the beginning of this module. This diagram shows the drug approval process from investment to pay off. At each node of the graph, there are two possibilities: first, you may or may not develop a drug that cures a particular disease. Next, even if the drug works, it may or may not get approved. And even if it gets approved, it may or may not make it to the market. At each branch, there are two possibilities with estimated probabilities (note: how do we estimate these probabilities in practice?). All of the branches lead to 5 possible outcomes, with different pay-offs. Using what we know about all of the possibilities, all of their probabilities, and the potential payoffs, we can estimate the expected value of the investment.

As before, we can think of expected value as a weighted average of all of the possible pay-offs, where the weights are precisely the probabilities of each. (Weighted averages are not always weighted by probabilities. For example, a course grade may be determined as 50% exams, 30% papers, and 20% homework. But weights, and probabilities, must always add up to 100%, or probability = 1, if you include all of the components.)
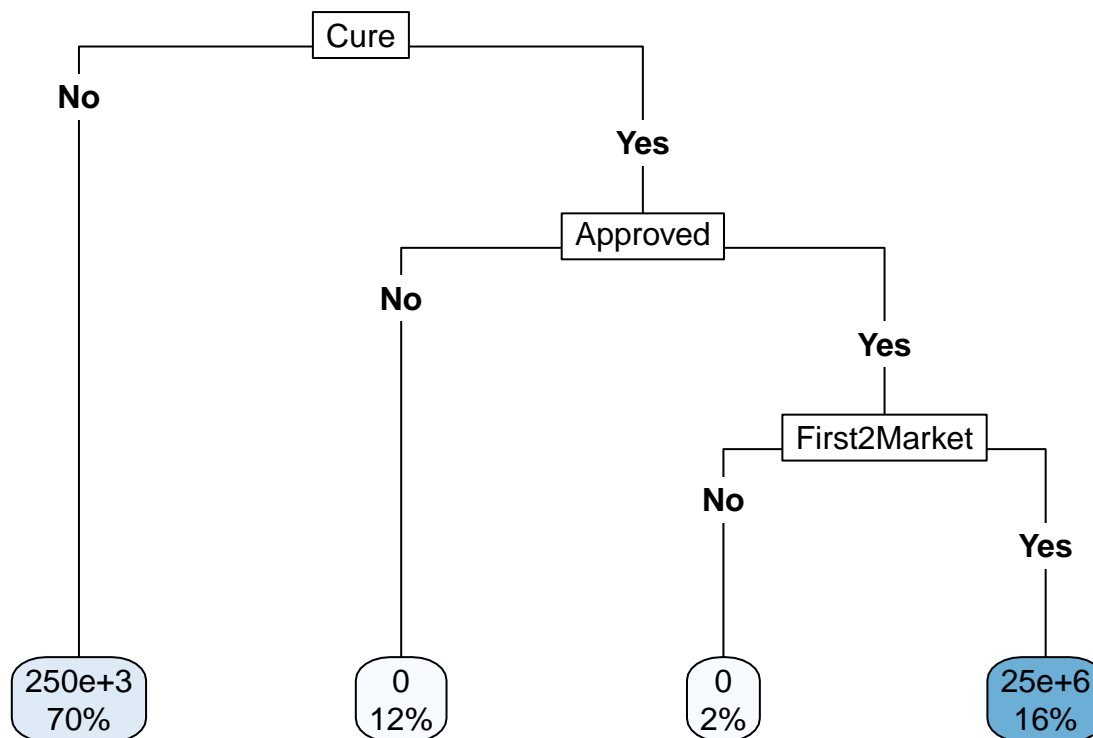
Since the five possibilities are. . . . . . .   the expected value is.

This value represents the average amount of money we would expect to make, if we invested in a lot of drugs, that all went through this same process independently.

Question to think about: even though the expected average pay-off is $4,225,000, which is more than 4 times the original investment, would you want to make this investment for any single drug? Why or why not?

Now that we've made this calculation empirically, it might be helpful to simulate it in R! https://a3sr. shinyapps.io/Drug

---

[1] 0.5 * 9 - 0.5 * 3 = 3, so three dollars

Cure
- No: 250e+3, 70%
- Yes → Approved
  - No: 0, 12%
  - Yes → First2Market
    - No: 0, 2%
    - Yes: 25e+6, 16%

In the last two weeks, we've explored the concept of expected value. As you may have noticed, expected value can be very useful for making decisions, but it does not tell the whole story. In the college majors example, expected career earnings for each possible major were only useful in describing what happens on average, for a lot of people going through the job market. However, we still had a lot of uncertainty about what would happen to any single individual who pursued a career in acting or accounting. Similarly, we would not have been as confident running the Schlitz commercial with a sample of 10.

This brings us to one of the most important concepts in probability and statistics: the "law of large numbers". Random variables are by definition, random, so we cannot predict any single outcome – we can only say how often we expect a particular outcome to occur if we observe a lot of data. Thus, you might see how expected value is particularly useful in situations where the same process gets repeated many times. In the airplane overbooking example, there may be a reasonable probability that airlines will lose money on any given flight. But airlines can make probability calculations to ensure that they earn money than they lose overall.

The usefulness of the law of large numbers is particularly evident in gambling. Many people go to casinos and play the same games over and over again, which each (theoretically) have consistent probabilities of winning or losing. One person may win and another may lose, but if a lot of people play a game of roulette (for example), casinos can estimate approximately what proportion will win and how many will lose. And, the more games that are played, the more confident the casino can be about their predicted proportion of wins/losses.

As you've probably heard, casinos are designed such that the house always has the advantage. In other words, every casino game is designed such that the casino always has a slightly better chance at winning than the player. That doesn't mean that players can't win individual games, it just means that, assuming enough people play the game, the casino will almost surely win more games than all the individual players.

This construct tempts us to believe that we can cheat the system. There is a long and interesting history of people trying to figure out betting strategies that will give them back the advantage. The Martingale betting strategy is one example. Please read the following article to learn more about it! https://www.roulettesites.org/strategies/martingale/

1.3   Betting on Basketball

As a final example of how we can use probabilistic thinking to inform betting decisions, let's consider March Madness. March Madness is a basketball tournament, where 64 teams compete. In the first round, 32 games are played. The winners of those games then play each other in the 16 second round games. This continues until a single team is named victorious. Every year, people make bets on which teams they think will win each match-up. Data from regular season games can be used to estimate the probability of any team winning a particular match-up, and this information is readily available to fans. Yet, despite the wealth of data that is available to make game predictions, no one has ever correctly predicted all of the sequential game winners in the tournament. In fact, only one person has ever predicted the first two rounds perfectly (this happened in 2019!). How is this possible? In order to learn more about the complexities of choosing a March Madness bracket, please read the following article: https://www.scientificamerican.com/article/how-much-math-do-you-need-to-win-your-march-madness-pool/

Now that you know a little more about March Madness, let's just consider the probability of predicting the entire first round correctly. To start, we will (incorrectly) assume that every team has a 50% chance of winning their first round game. In this case, we have a 50% chance of guessing the correct winning team. Given that there are 32 games in the first round, how could we calculate the probability of predicting them all correctly? [walk through calculation, drawing a tree diagram, similar to those in the first two weeks of the module]