Behavioral Finance

Problems with Probability

Duke UNIVERSITY

Problems with Probability

In this Section, we will examine our innate ability (or otherwise) to make reasonable inferences based on probabilistic information. Using a series of examples and mini case studies, we will categorize some common and predictable traps that many of us fall into when thinking about correlations between events, and making assumptions about their causal relationships.

We will learn that:

(1) We are surprisingly bad at forecasting *correlation* (the extent to which future events are related to one another).

(2) We are inclined to see strong correlations where none exists, as well as ignoring correlations where a strong relationship *might* exist.

(3) When we look at correlated events, we are also inclined to assume *causality*: that one outcome *causes* another, where in fact the relationship is either a coincidence, or both factors are being driven by some other cause altogether, or even that the causal relationship is in the opposite direction!

Are you good with probability problems?

Q1: Two cards are picked at random from a deck (without replacement). What is the probability that the first card is red and the second is black?

Q2: Suppose I turn one card over in a deck, and observe that it is the 7 of Diamonds. I then draw another card, without replacement. What is the probability that this second card is also a diamond?

Q3: Two cards are picked from a deck without replacement. What is the probability that the second card is a queen?

Q4: Suppose you toss a coin 3 times, and each time it comes up heads. What is the probability that the 4^{th} toss of the same coin will also be heads?

Please answer these questions for yourself before moving on.

Probability Solutions

Q1: Two cards are picked at random from a deck (without replacement). What is the probability that the first card is red and the second is black?

Half of the cards in a deck are red, so the probability that the first card is red is: ½

Now we've removed a (red) card, so there are 51 cards left, of which 26 are black. So the probability that the second card is black is: 26/51

Hence the probability that the first card is red and the second is black is:

½ * 26/51 = 0.255 (i.e., slightly higher than ¼)

Q2: Suppose I turn one card over in a deck, and observe that it is the 7 of Diamonds. I then draw another card, without replacement. What is the probability that this second card is also a diamond?

There are 51 cards remaining after the 7 of Diamonds has been removed. Of these, 12 are diamonds. So the probability is 12/51, or about 24%

Probability Solutions

Q3: Two cards are picked from a deck without replacement. What is the probability that the second card is a queen?

The probability that *any* card picked at random is a queen is 1/13. Nonetheless, if you got this wrong, you are in very good company. People often mistakenly assume that there is some significance in the fact that we are asking about the *second* card. But since we haven't been given any information about the *first* card, it doesn't make any difference!

Q4: Suppose you toss a coin 3 times, and each time it comes up heads. What is the probability that the 4^{th} toss of the same coin will also be heads?

Coin tosses are independent events. Unless the coin is mis-weighted, the probability that the 4^{th} toss is heads is still ½.

The Juror's Fallacy

Base Rates and False Positives

You are a member of a jury. A taxi driver is accused of having run down a pedestrian on a stormy night and having fled the scene of the accident. The prosecutor, in asking for a conviction, bases his whole case on a single witness, a lady who saw the accident from her window a little way away. The witness testifies that she saw the pedestrian struck by a blue taxi and then saw that taxi drive away from the scene. The accused works for a taxi company whose taxis are all blue.

During the trial, the following emerges:

1. There are only two taxi companies in this town. The whole fleet of one company is green; the other has only blue cabs. Eighty-five percent (85%) of all taxis on the road that night were green, and only fifteen percent were blue.

2. The single witness has undergone a number of vision tests in conditions similar to those of the night of the accident. She has been shown to be able to identify the two colors correctly about 80% of the time; i.e., out of all of the blue & green cabs she saw during the vision test, she got the color right 4 out of 5 times.

How likely is it that she actually saw a blue taxi that night?
Please make a note of your answer before moving on.

The Juror's Fallacy

Base Rates and False Positives

Let's try to answer this question. We'll start by assuming that there are 1,000 cabs in this town. From the information on the previous page, we know that 850 of the cabs are green, and 150 are blue.

There are two possible scenarios, given that the witness identified the cab as blue:

- The cab **was** blue, and the witness identified it correctly
- The cab was **green**, and the witness **mistakenly** said it was blue

We know that the witness only gets the cab color right 80% of the time.

- She will incorrectly report **170** of the *green* cabs as *blue* (20% x 850). This is called a "false positive."
- She will correctly report **120** of the *blue* cabs as blue (80% x 150).

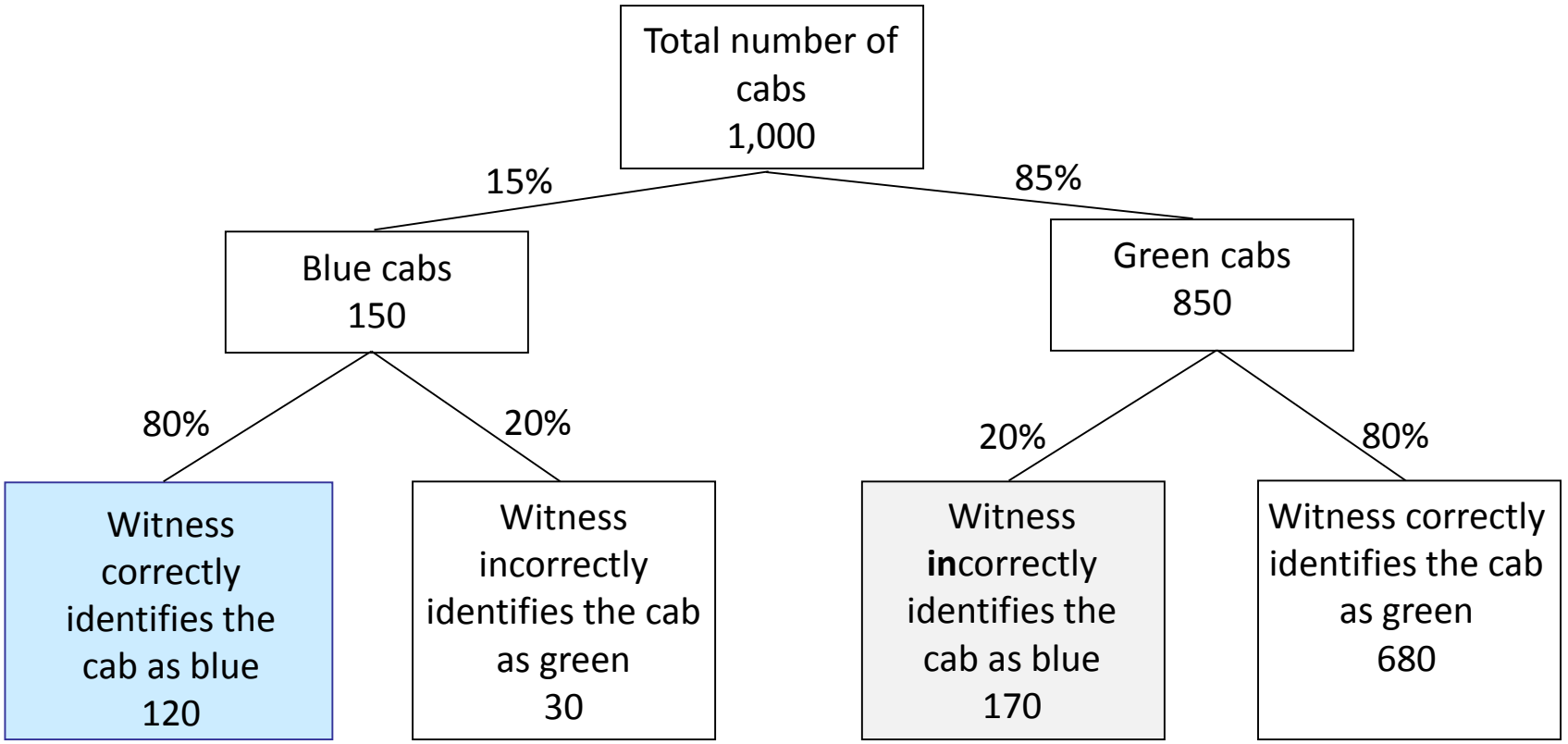
Hence, of the 1,000 cabs on the road that night, she will report **290** as blue (170 + 120). But only **120** of these cabs are *truly* blue.

So she will correctly report a blue cab only **41%** of the time (120/290 = 0.41).

Still confused? See the graphic on the next slide.

The Juror's Fallacy

Base Rates and False Positives



With 1,000 cabs on the road, only 150 of them are blue, and the witness will correctly identify the color of *either* cab only 80% of the time.

The witness's 20% false positive rate, along with the fact that far more of the cabs were green than blue that night, results in a high probability that the cab was green, but that she *misreported* it as blue.

The Juror's Fallacy

Base Rates and False Positives

There are two significant sources of confusion here:

- Ignoring the *base rate probability* of a blue cab (15%). This is sufficiently low that a relatively small possibility of color error on the part of the witness leads to a significant probability that she is mistaken about what color cab she saw on the night in question

- Failing to take into account the 20% *false positive* rate. Since the witness incorrectly identifies 20% of the green cabs as blue, and since most of the cabs on the road that night were green, it is more likely that she saw a green cab and *thought* it was blue, than that she correctly identified a blue cab.

Any time you have a *low* base rate, and a non-trivial false positive rate, the potential for error becomes surprisingly high.

Correlation / Causation error

See the Lecture: Correlation and Causation Error

"Don't confuse brains and a bull market!"

- Young investors who are doing well and making lots of money tend to ignore the *base rate* of profit & loss success (the S&P500, perhaps, if they are investing in US equities)
- If *all* equity prices are going up, then you would have to work hard *not* to make money.
- Your ability to make money in this environment is a result of strong correlation with the overall market – it is not necessarily caused by your investing brilliance.
- One of the single biggest mistakes made by young investors is to mistake luck for skill. More on this later, when we examine over-confidence and the confirmation bias.

A brief diversion on subprime mortgages
Invisible Correlation error

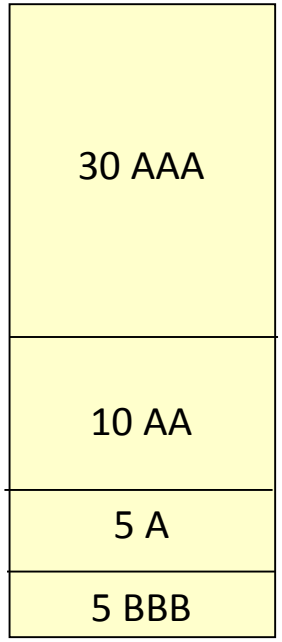
How do mortgages work?

- Mary borrows $200,000 from the bank to buy a house
  - Each month, Mary pays $1,000 to the bank
  - Some of this payment is interest on the loan
  - Some of this payment reduces the size of the loan
- After 30 years, Mary has paid down the entire loan, and the house is hers*
- Until 2001, the vast majority of mortgage loans went to "prime" borrowers: individuals with *proven* savings, *proven* monthly income and a *proven* credit history
  - Historically, average default rates on prime mortgages are less than 3%
  - If a borrower defaults over multiple months, the lender can foreclose (i.e., repossess the house), and sell the house to pay off the loan balance
  - Provided the house value has not fallen, the sale will be enough for the lender to recover the total value of the loan
  - Historically, while there have been house price declines in various regions of the US at different times, a US-wide house price decline (in *nominal* terms) has not been seen since the Great Depression

*In the vast majority of cases, borrowers pay down the loan in less than 30 years; either because they move house and close out the loan, or interest rates go down and they refinance into a lower rate mortgage

A brief diversion on subprime mortgages

Invisible Correlation error



How do Mortgage-Backed Securities (MBS) work? A simple example

- Take 50 mortgages: one from each US state
- Pool them together as a mortgage-backed security and sell "tranches" to investors
  - Investors pay a fixed amount today in return for borrowers' cashflows each month
  - The higher the rating on the tranche, the higher the price: the AAA tranche owner pays more upfront (per mortgage) than the BBB tranche owner
  - Any borrower defaults hit the lower tranches first

- Each month, borrowers make their monthly mortgage payments
  - If one borrower in the pool defaults, the BBB tranche owner gets a smaller cash flow (if a borrower fails to pay for multiple months, the lender can take the house and sell it, after which he delivers the proceeds as a lump sum to the relevant MBS investors)
  - If 5 borrowers default, the BBB tranche owner receives nothing (that was the risk he took when he paid less up-front)
  - If 10 borrowers default, both the BBB and A tranches get nothing
  - If 25 borrowers default, even the supposedly super-safe AAA tranche starts receiving fewer cash flows.

A brief diversion on subprime mortgages
Invisible Correlation error

- In 2002 – 2006, the proportion of mortgage loans made to "subprime" borrowers (no job and/or no savings and/or no credit history) increased from around 5% to 35% of all US mortgages

- For the first time in history, more than 1 trillion dollars in subprime mortgages were packaged into MBS and sold to investors

- What went wrong?

  - Investors assumed that subprime mortgage default rates would be similar to prime default rates. In fact, they were about **20 times higher**

  - Investors assumed that house prices would continue to rise, so that even if borrowers defaulted, each loan would be paid off in full by selling the house. House prices across the country started to fall in late 2005

  - Investors believed that, even if house prices declined in some parts of the US, this would not be a country-wide phenomenon, so geographically diversified pools of mortgages would be safer. By 2009, house prices across the US had declined by 50% on average, with at least some degree of decline in every US state.

  - Investors significantly underestimated correlations between (1) house price decreases across the US, and (2) subprime mortgage default rates across the US

"Frankly, when I took out the loan I sincerely believed we'd all be blown to hell before it came due."

The Sniper:
Confusion of the Inverse

In the Fall of 2002, a huge man-hunt took place for the sniper who killed a number of people in Washington, D.C., and surrounding areas.

Early in the search, the police arrested a man who owned a white van, a number of rifles, and a manual for snipers. It was thought at the time that there was one sniper* and that he owned all these items, so for the sake of this story let's assume that these facts are true.

Now, suppose that there are around 4 million people in the DC area, and also suppose that fewer than a dozen of these people own all three items. Think about the following two questions:

(a) What is the probability that an innocent man would own all these items?
(b) What is the probability that a man who owned all these items would be innocent?

Please come up with your best guess for these two probabilities before moving on.

* In fact, it turned out that there were two men working together

The Sniper:
Confusion of the Inverse

In the Fall of 2002, a huge man-hunt took place for the sniper who killed a number of people in Washington, D.C., and surrounding areas.

Early in the search, the police arrested a man who owned a white van, a number of rifles, and a manual for snipers. It was thought at the time that there was one sniper and that he owned all these items.

Now suppose that there are around 4 million people in the DC area, and also suppose that fewer than a dozen of these people own all three items. Think about the following two questions:

(a) What is the probability that an innocent man would own all these items
(b) What is the probability that a man who owned all these items would be innocent

We have 4 million innocent people in the Washington area (and one guilty sniper).
10 of them (including the guilty one) own all three of the items mentioned above. So:

(a) is 9 / 4,000,000 (0.0002%)
(b) is 9 / 10 (90%)

Did you correctly estimate just how different these two probabilities are? Most people do not even come close…

This is an example of what is sometimes called “Confusion of the Inverse”. A similar and tragically common error is to confuse the statement: “many terrorists come from middle eastern countries” with the statement: “many middle easterners are terrorists.” Can you come up with a good example from the financial markets?

Please take a look at the information on the following slide, which was part of an education campaign, a few years ago, from what is now called Partnership for Drug-Free Kids.
Ask yourself why we have included this public service announcement in the course materials. Once you've answered that question, please move on to the subsequent slide.

MARIJUANA AND YOUR TEEN'S MENTAL HEALTH

Depression. Suicidal Thoughts. Schizophrenia.

If you have outdated perceptions about marijuana, you might be putting your teen at risk. New research is giving us better insight into the serious consequences of teen marijuana use, especially how it impacts mental health.

Did you know that young people who use marijuana weekly have double the risk of depression later in life?[1] And that teens aged 12 to 17 who smoke marijuana weekly are three times more likely than non-users to have suicidal thoughts?[2]

And if that's not bad enough, marijuana use in some teens has been linked to increased risk for schizophrenia in later years.[3]

Today's teens are smoking a more potent drug[4] and starting use at increasingly younger ages during crucial brain development years.[5] Still think marijuana's no big deal?

Remember, you are the most important influence in your teen's life when it comes to drugs,[6] so tell your teen the facts about marijuana. Teens who learn about the risks from their parents are less likely to smoke marijuana or use other drugs than teens who don't.

Let your teens know you don't want them using marijuana. Their mental health may depend on it.

Signed,

PARENTS | THE ANTI-DRUG | **1-800-788-2800**

www.theantidrug.com

[1]Patton, GC et al. Cannabis use and mental health in young people: cohort study. *British Medical Journal*, 325: 1195-1198, 2002. [2]Greenblatt, J. Adolescent self-reported behaviors and their association with marijuana use, Substance Abuse and Mental Health Services Administration (SAMHSA), 1998. [3]Arseneault, L et al. Cannabis use in adolescence and risk for adult psychosis: longitudinal prospective study. *British Medical Journal*, 325: 1212-1213, 2002. Feen, N et al. Cannabis use and age at onset of schizophrenia. *The American Journal of Psychiatry,* 161:501-506, 2004. [4]*Marijuana Potency Monitoring Project. Report No. 83,* University of Mississippi, 2003. [5]SAMHSA. Trends in Initiation of Substance Use, 2003. [6]SAMHSA. Parental Disapproval of Youths' Substance Abuse, 2002.

More Confusion of the Inverse

WEIGHT WATCHERS – DO YOU KNOW THE RISKS?

Obesity. Diabetes. Heart Disease.

If you have outdated perceptions about attending Weight Watchers, you might be putting yourself at risk. New research is giving us better insight into the serious consequences of attending Weight Watchers, and especially how it impacts your physical health.

Did you know that people who attend Weight Watchers weekly have double the risk of obesity?

And if that's not bad enough, attending Weight Watchers has been linked to diabetes, heart disorders, and strokes in later years.

Today's population has more access to Weight Watchers than ever before, as they continue to open new offices domestically and worldwide.

Still think Weight Watchers is no big deal?

No doubt you have realized that this example is a parody of the (genuine) public service announcement on the previous slide. While the anti-drug slide's intent is laudable, its argument confuses correlation for causation, and may even fall into the "confusion of the inverse" category.

Perhaps, in parallel to our weight-watchers parody above, it is the teenagers who already *have* mental health issues who are most likely to self-medicate with marijuana…

Ironically, the academic papers that the organization cites at the bottom of the anti-drug announcement also note that, while they find evidence of *correlation* between drug use and later mental health issues, *causation* has not been determined.

What have we learned?

In this section, we have reviewed a number of common traps that arise from correlation and causation errors:

- Mistaking *correlation* for *causation*
  - Ice cream and shark attacks
  - My investing brilliance during a bull market
- Ignoring base rates and the impact of false positives
  - Juror's Fallacy
- Invisible correlation
  - The nationwide house price collapse leading to the credit crisis
- "Confusion of the Inverse"
  - The Sniper
  - The anti-drug campaign announcement and the Weight Watchers parody