

1 Raven paradox

The Revan paradox questions what we see as evidence.

Consider the statement: *All ravens are black*. Using implication, we can express the same statement as: *If something is a raven, then it is black*.

$$\text{raven} \implies \text{black}$$

If we ever see a raven that is not black, we have shown this statement to be false. But, no amount of black ravens can prove this statement to be true. That said, the more black ravens we observe, the more confidence we have in this statement being true.

Now consider the truth table:

p	q	$\neg p$	$\neg p \vee q$	$p \implies q$
T	T	F	T	T
T	F	F	F	F
F	T	T	T	T
F	F	T	T	T

It appears that $p \implies q$ is identical to $\neg p \vee q$. So to get confidence in *All ravens are black* statement we can also make observations of things that are

$$\neg \text{raven} \vee \text{black}$$

So we observe a *green apple*, and gain confidence in that *all ravens are black*.

2 Selection/Survivorship bias

This is a logical error of only considering individuals or things that made it past some selection event. A good example is when analysts during World War II were examining damage characteristics of airplanes. They noted that enemy fire concentrated the damage mostly on wings and tail of the aircraft. It seemed obvious to put additional armor on wings and tail.

Statistician Abraham Wald noted that the study only considered airplanes that survived (returned to base). The damaged parts represented the most resilient parts of the airplane, where the airplane can take damage and still survive. The parts which had no damage (engines and cockpit) were the most vulnerable, as those airplanes did not return. In other words, the engines and cockpit should gain extra protection, not the wings and tail.

Selection bias, a more generalized survivorship bias, show up quite a bit in all sorts of studies.

Students who have a double major tend to have higher GPA than those with a single major. It seems obvious that to raise GPAs we should encourage all students, especially those with low GPAs to have a double major.

From past data, the best investments seem to be to purchase a stock when the company is on the verge of bankruptcy, and then wait a few years for it to recover. Most such analysis does not consider all the companies that no longer exist.

Similarly, it seems many day-traders are doing well. Perhaps we should all be doing day-trading. This misses the fact that most folks who start end up badly—and only those that are doing well are still doing it.

3 Luck

It seems silly to take advice from lottery winners about how to play lotteries to win. They often let the machine pick few random numbers—there is no strategy or skill involved.

A coin flip is considered random. If we play the coin-flip-game against 10 random people, it seems unlikely that we would win all 10 times.

Consider we setup a tournament to pick the best coin-flip-winner, we start with 1024 participants, and they begin playing, tournament style. On the first toss, we are left with 512 participants, keeping only those who got heads. On 2nd toss we are left with 256 participants, again, keeping only those who got heads. Those folks got 2 heads in a row!

As this process winds down, we are left with a single winner, who won 10 coin flips in a row!

This tournament winner might give away advice and publish books on how to be good at coin-flipping game, as evidenced by their history of winning so many times (obviously there is skill involved).

This is often observed in financial markets, management, and business. Just because someone won, should not immediately imply that it was due to their skill.

4 Fairness

Machine learning models generalize. That is how they work and what they do. That means they take a few example instances, and create generalized rules to apply to everyone. That can be good and bad, depending on perspective.

When you are applying for insurance, you want to be generalized. If there is 1 in 1000 chance that folks of your age group might require expensive medical treatment in the next year, you want your rates to be determined by that. Not the fact that *your* family history (or results of a genetic test) might indicates that *your* chances are are 1 in 100.

For the insurance company (if you were the owner of the insurance company), it makes sense to charge higher rates for higher risk (in other words, insurance company would like to utilize your individual circumstances to differentiate you from the low-risk group, and perhaps put you in a high-risk group to charge you higher rates).

Similarly, when applying for a loan, and the algorithm determines that folks of your age group, educational background, job history, etc., are generally bad at repaying the loan. At

this point, you do not want to be generalized. You are not the group, and actions of the group are not yours. Of course you will repay the loan.

For the bank (if you were the owner of the bank), it makes sense to charge higher rates and ignore the individual circumstances that differentiates you from the group.

Some of these are lot more emotionally charged—some states use a computer program (implementing a proprietary and not auditable model) to determine whether a prisoner should be paroled.

5 Confusing Cause

Imagine a computer program that predicts crime-hot-spots, and concentrates police patrols in those hot-spot areas.

As expected, the police make more arrests in those hot-spot areas. Program accuracy is confirmed. Everyone involved gets promotions, etc.

But the cause of more arrests is likely the presence of the police patrols. Areas with less police (areas the model did not deem as hot-spots) will get less arrests simply because there are less patrols there.

6 Proxy metrics

Computer models an optimize anything—given a well defined metric, we can use all sorts of methods (even random walk) to optimize the given metric.

How does one measure customer satisfaction? Or “danger to society”? Or intelligence?

The simplest method is to use some proxy metric that is easily measured. For customer satisfaction, it could be number of purchases, or inverse frequency of help-desk calls, etc. For “danger to society” it could be number of arrests, etc. For intelligence, we could measure performance at some task, such as playing chess, or video games.

Proxy measures are often flawed. Some in more ways than one. Just because someone was arrested for something stupid, does not mean they are a danger to society (protesting for peace outside of a gov building might result in an arrest). Measuring intelligence by measuring performance on any one task leads the machine to optimize performance at that task: that’s how we got chess playing programs that are amazing at playing chess and nothing else.

7 Right now is not special

Suppose the average arrival time between busses is 10 minutes. You show up at a bus stop, how long would you have to wait on average for the next bus? Less than 5 minutes? 5-minutes exactly? Or more than 5 minutes?

The answer turns out to be more than 5-minutes. Time is sliced up into intervals between busses, where the average length of these intervals is 10 minutes. But longer intervals take

more time, and your chances of showing up during one of those longer intervals is higher than during the shorter duration ones. If you show up during a longer interval, your wait time on average is longer than 5 minutes.

8 Doomsday argument

Everything that has a beginning has an end (said a character in The Matrix).

The New York Public Library Main Branch, the Stephen A. Schwarzman Building on 5th avenue, was constructed in 1911. When will that building close?

This question could have been asked in the past. It can be asked in the future. Let us consider the current time as a random time in Library Building lifetime.

To have 95% confidence, this question would have to be asked in the middle 95% of the lifetime of that building. That means today, 2019 could be between 2.5% or 97.5% of the lifetime of that building. In other words, with 95% confidence, it will last until 2022 (building is in the last 2.5% of its life), but not beyond year 4320 (building is in the first 2.5% of its life).

Some folks have applied this logic to the human race to come up with the Doomsday argument (wikipedia for it).