by Jyotishka Datta, Dept. of Statistics, Virginia Tech

Paradoxes

Readers of this column are surely familiar with Zeno and his paradoxes, for example, the one involving Achilles and the tortoise. If you are not, think of a regular morning, and you have to walk to the kitchen counter to grab a cup of coffee, the cup of coffee being your 'limit', but in order to get there you first have to walk half of the distance, then half of whatever is remaining, then half of that, and so on ad infinitum. In Aristotle's words, "in a race, the quickest runner can never overtake the slowest, since the pursuer must first reach the point whence the pursued started, so that the slower must always hold a lead." An analogous version was coined by ancient Chinese philosophers, "If from a stick a foot long you every day take the half of it, in a myriad ages it will not be exhausted." But surely Achilles will beat the tortoise and you, your cup of coffee, and given sufficient information, will be able to solve exactly when and where it happens. Still, many years back, when I was a little kid, with no idea of infinite series (or of infinity for that matter), this was the problem that piqued my interest in Mathematics, one that has never let me go. One can argue that this classic tale extends a little bit into the territory of philosophy, as philosophers argue that the central tenet of Zeno's paradox is not just the mathematical solution but the impossibility of ever reaching a 'final resolution'. And, anamalies and paradoxes play a role in scientific revolutions too, as Thomas Kuhn argued in his 'Structures of Scientific Revolutions', that the occasional paradigm shifts happen in science when accumulated paradoxes or anomalies outweigh the current paradigm of 'normal' science, despite the initial and often dogmatic resistance.

So, I thought, why not talk about a few of my favourite paradoxes in Statistics and Probability that have fascinated me ever since I learned about them? I should note here that these are not paradoxes in the strictest sense, as in, they don't expose logical inconsistencies in a paradigm but they do shed light on the slippery slopes and remind us to be cautious, and not depend too much on our intuitions, especially when hidden structures or infinite objects are at play.

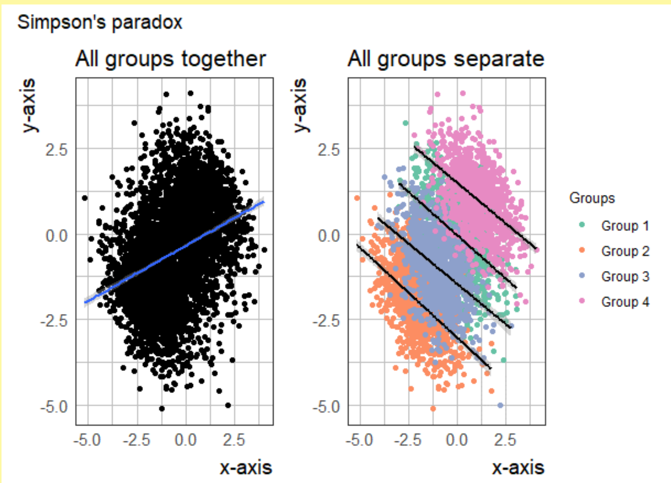

Figure 1

The first of this is perhaps one of the most well-known and yet often overlooked phenomena: the Simpson's paradox, that happens when a trend observed within individual groups reverses when the groups are combined (see Figure 1 for a simulated example inspired by [1]). A student of statistics usually learns this in their undergraduate curriculum and learns about the famous examples: the case of gender discrimination at Berkeley and the case of kidney stones. We also realize soon that it is not as paradoxical as the name suggests, it is rather a problem of ignoring a latent variable, a problem pervasive but well-understood, attracting comments that Simpson's 'isn't really a paradox at all' [2]. However, it is more commonplace than we tend to think, as we learned during the Covid-19 polycrisis that Simpson's paradox can strike whenever we look at grouped data and can still take us by surprise. One of the most striking examples of this is due to Prof. Jeff Morris who showed that for Israeli vaccine data, failure to stratify by age, which is a key factor in both vaccination status and risk of severe diseases, leads to a misleading picture of vaccine effectiveness [3]. The example or the analysis is much more nuanced than I can discuss here, and I would encourage readers to read the excellent blog by Prof. Morris [3] for a detailed discussion. This naturally makes one wonder, how often do we see a misleading statement in research papers or media headlines because of an omitted variable bias?

The second paradox is one that I first read in one of the most popular probability textbooks, "A First Course in Probability", by Sheldon M. Ross. This is also known as the Ross-Littlewood paradox or the ping-pong ball problem. A simplified version is as follows:

Imagine you have an bottomless vase and an infinite bag of balls, labeled with the natural numbers 1, 2, 3, and so on. Starting at 1 minute to noon, every $(1/2)^n$ minute you add ten more balls to the vase, and then you remove one ball. The process is as follows: At minute 1, you add balls 1 to 10 and remove ball numbered 10. At minute 1/2, you add balls 11 to 20 and remove ball 20. At minute 1/4, you add balls 21 to 30 and remove ball 30. ... and so on. The question posed is: how many balls are left in the vase at noon? The answer is clearly infinity since all balls numbered anything other than 10n would be there. But, if you change the way the balls are drawn: at minute 1 to noon, you add 1-10 and draw 1, at $1/2$ minute you add 11-20 and draw 2, and so on, ad infinitum. But, now, paradoxically the vase must be empty at noon because for any ball number 'n', there's an epoch, viz. $(1/2)^{(n-1)}$, when it must have been drawn. What if you draw a randomly selected ball at every epoch? Using Boole's inequality, Ross showed that with probability 1, the vase will be empty at noon .

Apart from the apparent simplicity of this problem statement, another reason I like it is that different philosophers, mathematicians or logicians, favor different solutions that are completely different, and it led to a long discussion on forums like StackExchange [4] and other places. One of these views is that this is an ill-posed problem, as James Meyer writes, it "is an infinite unending process that is not completed" [5], or, to hark back to Zeno's paradox, we have to complete an infinite number of steps before noon, and 'noon' is never reached.

My wrist-watch says it's 11 in the morning, and as I make plans for lunch in this lovely weather, I know Achilles will eventually beat the tortoise and noon will always be reached.

After all, we live in a paradoxical world.

References:

1 https://www.r-bloggers.com/2020/11/simpsons-paradox-and-misleading-statistical-inference/

2 Coronavirus vaccines work. But this statistical illusion makes people think they don't. - The Washington Post

3 Israeli data: How can efficacy vs. severe disease be strong when 60% of hospitalized are vaccinated?

4 https://stats.stackexchange.com/questions/315502/at-each-step-of-a-limiting-infinite-process-put-10-balls-in-an-urn-and-remove-o

5 https://www.jamesrmeyer.com/paradoxes/balls-in-the-urn-paradox

---

Meet our author, Jyotishka Datta!

Introducing our new feature column, 'Datta's Data Dives'. A relaxed way to dive into the intricacies of the data world.