

Radu's Ride: Is More Less?

Our contributing columnist Radu Craiu, University of Toronto, writes:

In this age of excess, the question in the title will sound sacrilegious. Has the permissive attitude towards a bloated pantry, shoe rack or refrigerator spilled over into (data) science? While the evidence is piling up, the irony of our lived reality is that whatever one columnist says will be drowned in the noise, so here goes nothing.

The first item of evidence concerns ChatGPT, which seems to haunt everybody's Wells-ian nightmares—who needs a War of the (outer) Worlds when we are building our own from scratch? There is also the other side of the fear-mongering equation, that we need ChatGPT to improve on our lives (which I take as meaning that we could produce more derivative stuff of questionable quality, but I am not here to nitpick). I hate to disappoint this latter camp, but ChatGPT has left me feeling as helpless as I was before meeting it. For a number of reasons that, when all is said and done, boil down to my very human nature, I was late preparing a slide deck for a class I was scheduled to give on computational statistics. The material was ready, but it was scribbled down in a way that goes back to the code of Hammurabi: by hand. So, I figured I could get ChatGPT to give me a head start on those demanding LaTeX slides that take so much time to write when you type with two fingers. I will spare you the details of which topics I needed help with, since I suspect that ChatGPT is equally useless for many others, but I can confirm that what I got back was a bunch of general, mostly useless drivel that one might expect from someone who knows someone who has a friend who played roulette in Monte Carlo. I am happy for all those who found ChatGPT a menace to society as we know it, since it means that their life got better for a brief moment,

enough to wonder whether that relief they're feeling is the siren song of doom. Alas, these people seem to be all elsewhere.

Panicking that I would be left behind in my level of panic, I started to pay more attention to what was said around me, and the most worrisome message I could get is that ChatGPT is awesome at writing grant introductions or could take a lifeless letter and pour some Drake slang into it. I am not sure this is enough to wake up the neighbors.

My second item of evidence should have been the first, because it is not only epistemological but also historical. At the core of classical statistics lies the dictum that “less is more.” One might step back and ponder whether in our discipline's DNA it is also, non-equivalently, inscribed that “more is less.” This might explain our reluctance to juggle the data science juggernaut in which *more is more*: more data, more parameters, more attention. We talk to our students about the merits of large data, bringing more and more information about our reasonable models, but we eye suspiciously those who play in bigger backyards with increasingly complex models that are harder and harder to interpret—yet, just like toys that sparkle, they seem to mesmerize and fascinate beyond any reasonable doubts. Facing parameter spaces that grow with the data size, our asymptotics are often caught in some sort of statistical purgatory in which they neither kick in nor are completely abandoned, thus leaving them at the mercy of higher powers, also known as machine learners. Our one-sided, valiant attempts to create structures and foundations within the noise seem to be drowned in a cavalcade of

advances dressed in loud enthusiasm and shiny success.

The story is not as simple as a column's writer would like, though. Clearly, we have become very good at observing the universe, whether it is at the macro-cosmic level of outer galaxies or the microscopic level of disposable income. An Astrostatistics conference left me dizzy with the number of terabytes of data that have been collected already, and the even larger number that will be available in the near future. The odd pairing between our thirst for data and our limited abilities to handle it notwithstanding, one must justify all this effort with an analysis, and one that yields results, no less. While galaxies eons of light years away are sampled thoroughly, my colleague who is an expert in Demography (to be read as mortality) is grieving the lack of reliable data on migration, child mortality and human trafficking in much closer parts of the universe. My statistical brain is challenged by both sides, but I feel like I have a bit of a slow start in at least one of them.

By the time you read this in that famous January mood, you will likely be already riddled with guilt about all those New Year resolutions you have broken or are about to break... but leave all regrets behind and know that we are all, more or less, in the same boat!

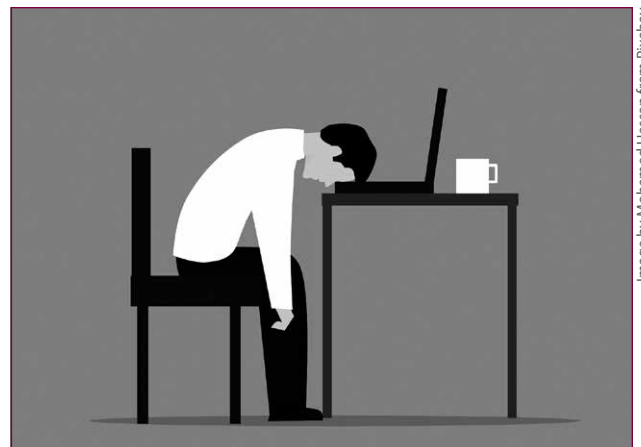


Image by Mohamed Hassan from Pixabay