Radu's Rides: Will the Future Mature?

Contributing Editor Radu Craiu writes: At the end of a year that felt like a decennium, it might be wise to remind ourselves that we are constantly torn between, on the one hand, the excitement of answering questions about the future *("What will it look like, and where do I fit in?")* and, on the other hand, the deflating realization that our imaginary trajectories tend to go off the rails *("Where did I go wrong?")*. Leaving aside the individual-level tribulations, one can ask, *"Where are Statistics and Data Science heading?"* and, *"What significant changes can we expect?"*

This is the kind of fun stuff that gets one in the doghouse, but let's try anyway to conjure up some futuristic visions, and break them down by activity.

Teaching

If you have been immersed in one of the disciplines at the core of Data Science, you may have been caught up in the maelstrom of creativity that has engulfed Statistics or Computer Science, but you may have missed the incredibly high demand from *other* disciplines for these core fields. Data Science is no longer desired only by its traditional dancing partners—Medicine, Astronomy, Finance, etc.—but also by Psychology, Political Science, English, History, Archeology and Philosophy. Their students will also want to dip their *curricula vitae* into DS's potent mix of good hiring chances, decent salaries and promising start-up benefits. In order to train them and influence their careers, we will need to adapt our teaching to cater to people without a ton of mathematical training, skills or abilities. The use of software will become ubiquitous, complementing the powerful principles we have been peddling for years, e.g. randomization in experimentation, restraint in modelisation, and skepticism in inference, to name a few of the ones that make us popular at parties. If thousands are already flooding Statistics and CS programs[1], tens of thousands will do so in the not-so-distant future. Our future selves will need to train those who will teach them the fundamentals of our discipline: we must rely on competent teachers to maximize benefits and minimize the potentially large damage that an improperly trained Bayesian, or any other sub-species of Data Scientist, can inflict on the world. Looking under the asymptotic hood is not for everyone and users of our methods should be allowed to safely avoid such curiosity, as long as they understand where the playground ends and the monsters roam. The high school curriculum will also have to change to account for the fact that, while less than 5% of graduates will end up calculating the roots of a polynomial of degree seven, more than 80% will need to understand odds-ratios and use regression in their daily interactions with the roulette that life sometimes becomes. (Disclaimer: these numbers are pure fiction and any resemblance to God's or other inventive souls' data is purely coincidental). This brings up a question that has long haunted Statisticians with dreams of a political career: how can we sell to the masses what Statistics can do, without scarring or scaring them off? A realistic, carefully thought-over answer to that question can only benefit the future of all.

Data

This is a category that currently does not exist separately, so its introduction is already significant. Our ability to collect data will explode. Imagine for a second that the Oculus will stay on as the dividing lines

....

1   See 'Growing Pains and Gains in Statistics, the Toronto Way' (2018). *IMS Bulletin*, 47(6).

between virtual and real become even more blurry. Our cybernetic overlords will not only have access to our shopping carts, they'll be able to shop our whole lives. Personalized treatments, recommendations and customizable offers will be on the table. They will be available at different resolution levels, with granular resolution levels available only to top-level, or top-paying, customers. Methods for data anonymization, along with ethical issues related to their use, will only gain relevance, triggering changes in how we perceive privacy, how we legislate data collection and how we regulate information flow. The battlegrounds might move or expand from an individual's right to privacy, to entire countries' or continents'. Arms treaties will be emulated by data collaboration treaties, with rogues as clearly defined as friends. Companies will emerge or crumble according to their ability to harness the power of information. In 20–30 years, most of the population in the West will have 90% of their life documented on the web and stored in some data repository; the battles these people will fight for their right to be forgotten, not to mention forgiven, will reach previously unseen levels. Friends and foes will bow to a new God.

Research

This column has mentioned, maybe *ad nauseam,* the incredible volume of output from Data Science research. If anything, these trends seem to accelerate and overtake (or take over) the field of Statistics. The research community will have to adapt in order to handle the sheer volume and resulting pressure. We will move away from the solitary researcher model towards building teams or the type of consortia we currently see in the study of genetics, medicine and public health. Theoretical Statisticians will be seen as the cult inside Data Science, made up of people who obsess over

minutiae and from time to time emerge from obscurity to play pivotal roles for very short times… so, nothing much will change there. I suspect that computation will become even more enmeshed with inference, and the pressure of applications will diminish the unhealthy reliance on unrealistic or uncheckable model assumptions. A new Box-ian phrase will emerge: "All models are regression-based, but only some are linear". Closed form estimators will be all but nonexistent, forcing us to evaluate

uncertainty in novel ways. Data will often reach the analyst after they have been heavily privatized, thus requiring essential new methods to recover as much of the original truth as possible. Most importantly, in a world in which attention span will be measured in seconds, data-based decisions will need to be made in milliseconds. Even after accounting for a tremendous increase in computing power, this constraint will spur interest in approximation methods and computational tricks that can short-circuit

the processing of terabytes of data in the blink of an eye.

Disciplines tend to take on a life of their own. The future DS ecosystem will split or expand according to tensions or priorities that are impossible to predict. But if I had a penny for every time I said "data" in the last 25 years, I would bet all of that money on DS staying at the center of human endeavours for the foreseeable future.

IMS members' work on COVID-19

A couple more IMS members wrote to share what they have been working on, around COVID-19. If you'd like a mention in the next issue, please contact ims@imstat.org (send a paragraph about your work, and a link to the paper, or location where interested readers can find out more). **Our next deadline is December 1, then February 1.**

Ioannis Kontoyiannis
Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, UK

Ioannis writes, "In collaboration with Jussi Taipale (a colleague from Biochemistry here in Cambridge) and Sten Linarsson (a molecular biologist at the Karolinska Institute in Stockholm), we propose a non-pharmaceutical intervention that would allow rapid control of the Covid-19 pandemic. The intervention is based on: (1) randomly testing every individual, (2) repeatedly, and (3) isolation of infected individuals. Using both a deterministic SIR model, and randomized, continuous-time Reed–Forest epidemic models on general random networks, we rigorously determine the necessary rate of testing for the pandemic to be quickly suppressed. Our results imply that the reproduction number of Covid-19 can be brought well below 1, even with rapid antigen tests having very low sensitivity, with limited quarantine compliance, and with perfectly realistic testing rates. Moreover, we show that this approach is robust to failure—any rate of testing will significantly reduce the size of the pandemic, improving both public health and economic conditions.

The research is available as a preprint (https://www.dpmms.cam.ac.uk/~ik355/PAPERS/CovidTesting.pdf) and as a seminar talk: https://www.youtube.com/watch?v=xEeEaCB2-qY

Klaus Krickeberg
Retired Professor of Statistics

IMS Fellow Klaus Krickeberg, together with Bernhelm Booß-Bavnbek, wrote an article called "Dynamics and Control of Covid-19: Comments by Two Mathematicians" that appeared in the September 2020 issue of the *Newsletter of the European Mathematical Society*. In the article, they pose the question: "Why are the dynamics and control of Covid-19 most interesting for mathematicians, and why are mathematicians urgently needed for controlling the pandemic?"

In the first part of the article, they examine the historical and country-specific contexts and responses, including the public health element. In the second part, the focus is "…not 'country-oriented' but 'problem-oriented'. From a given problem [they] go 'top-down' to its solutions and their applications in concrete situations." They organised this by the mathematical methods that play a role in their solution, and examine the particular need for mathematics in the development of a vaccine and the strategy for applying it "without losing sight of basic ethical principles."

You can read the article on pages 29–37 of the EMS Newsletter, issue 117: https://www.ems-ph.org/journals/newsletter/pdf/2020-09-117.pdf