The Law of Large Populations: The Return of The Long-Ignored "N" and How It Can Affect Our 2020 Vision

Professor Xiao-Li Meng, Harvard University

12 March 2018, Cambridge Judge Business School

Abstract

For over a century now, we statisticians have successfully convinced ourselves and almost everyone else, that in statistical inference the size of the population *N* can be ignored, especially when it is large. Instead, we focused on the size of the sample, *n*, the key driving force for both the Law of Large Numbers and the Central Limit Theorem. We were thus taught that the statistical error (standard error) goes down with *n*, typically at the rate of $1/\sqrt{n}$. However, all these rely on the presumption that our data have perfect quality, in the sense of being equivalent to a probabilistic sample. A largely overlooked statistical identity, a potential counterpart to the Euler identity in mathematics, reveals a Law of Large Populations (LLP), a law that we should be all afraid of. That is, once we lose control over data quality, the systematic error (bias) in the usual estimators, relative to the benchmarking standard error from simple random sampling, goes up with *N* at the rate of $\sqrt{N}$. The coefficient in front of $\sqrt{N}$ can be viewed as a data defect index, which is the simple Pearson correlation between the reporting/recording indicator and the value reported/recorded. Because of the multiplier $\sqrt{N}$, a seemingly tiny correlation, say, 0.005, can have detrimental effect on the quality of inference. Without understanding of this LLP, "big data" can do more harm than good because of the drastically inflated precision assessment hence a gross overconfidence, setting us up to be caught by surprise when the reality unfolds, as we all experienced during the 2016 US presidential election. Data from Cooperative Congressional Election Study (CCES, conducted by Stephen Ansolabehere, Douglas River and others, and analysed by Shiro Kuriwaki), are used to estimate the data defect index for the 2016 US election, with the aim to gain a clearer vision for the 2020 US election and beyond.

Speaker bio

Xiao-Li Meng, Dean of the Harvard University Graduate School of Arts and Sciences (GSAS), Whipple V.N. Jones Professor and former chair of Statistics at Harvard, and President elect of the Institute of Mathematical Statistics, is well known for his depth and breadth in research, his innovation and passion in pedagogy, and his vision and effectiveness in administration, as well as for his engaging and entertaining style as a speaker and writer. Meng has received numerous awards and honours for the more than 150 publications he has authored in at least a dozen theoretical and methodological areas, as well as in areas of pedagogy and professional development; he has delivered more than 400 research presentations and public speeches on these topics, and he is the author of "The XL-Files," a regularly appearing column in the *IMS (Institute of Mathematical Statistics) Bulletin*. His interests range from the theoretical foundations of statistical inferences (for example, the interplay among Bayesian, frequentist, and fiducial perspectives; quantify ignorance via invariance principles; multi-phase and multi-resolution

inferences) to statistical methods and computation (for example, posterior predictive p-value; EM algorithm; Markov chain Monte Carlo; bridge and path sampling) to applications in natural, social, and medical sciences and engineering (for example, complex statistical modelling in astronomy and astrophysics, assessing disparity in mental health services, and quantifying statistical information in genetic studies). Meng received his BS in mathematics from Fudan University in 1982 and his PhD in statistics from Harvard in 1990. He was on the faculty of the University of Chicago from 1991 to 2001 before returning to Harvard as Professor of Statistics, where he was appointed department chair in 2004 and the Whipple V.N. Jones Professor in 2007. He was appointed GSAS Dean on 15 August 2012.

The Economics & Policy subject group

Learn more about the Economics & Policy subject group, and view other seminars in the series, at www.jbs.cam.ac.uk/faculty-research/subject-groups/economics-policy/.