De: **Centre de recherches mathematiques** crm@crm.umontreal.ca
Objet: COLLOQUE DES SCIENCES MATHÉMATIQUES DU QUÉBEC (16/02/2018, Xiao-Li Meng) - statistique
Date: 12 février 2018 10:57
À: activites@CRM.UMontreal.CA

CD

****

COLLOQUE DES SCIENCES MATHÉMATIQUES DU QUÉBEC
http://www.crm.umontreal.ca/Colloques/index.html

****

DATE :
Le vendredi 16 février 2018 / Friday, February 16, 2018

HEURE / TIME :
15 h 30 - 16 h 30 / 3:30 p.m. - 4:30 p.m.

CONFERENCIER(S) / SPEAKER(S) :
Xiao-Li Meng (Harvard University)

TITRE / TITLE :
The Law of Large Populations: The return of the long-ignored N and how it can affect our 2020 vision

LIEU / PLACE :
McGill University, OTTO MAASS 217

RESUME / 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 analyzed 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 election and beyond.

****
Responsables :
Olivier Collin (UQÀM)
Henri Darmon (Université McGill)
Dimitris Koukoulopoulos (Université de Montréal)
Iosif Polterovich (Université de Montréal)
David Stephens (Université McGill)
Hugh Thomas (UQÀM)
Yi Yang (Université McGill)
****