

**The Hong Kong Polytechnic University  
Department of Applied Mathematics**

**AMA Distinguished Seminar Series in Data Science and Machine Learning**

**Analysis of Large Data with Subsampling**

**By**

**Prof. Zhezhen JIN  
Columbia University**

**Abstract**

Analysis of large data is challenging due to its size and computational issues. For the analysis of large-scale data, subsampling methods and divide-and-conquer procedures are appealing because they ease computational burden. However, it is challenging to preserve the validity of the resulting estimation and inference. In this talk, we will discuss a perturbation subsampling approach based on independent and identically distributed stochastic weights for the analysis of large data. We justify the method based on optimizing convex objective functions by establishing asymptotic consistency and normality for the resulting estimators. Simulation studies and real data analysis will also be used to illustrate the finite sample performance of the method.

**Biography**

Zhezhen Jin is Professor of Biostatistics in the Department of Biostatistics in Mailman School of Public Health at Columbia University. He received his Ph.D. degree in Statistics from Columbia University in 1998. After 1998-2000 two years of postdoctoral studies at Harvard School of Public Health, he returned to Columbia as a faculty member in the Department of Biostatistics in 2000. He has been conducting statistical and biostatistical methodological research on resampling methods, survival analysis, nonparametric and semiparametric methods, smoothing methods, and statistical computing. He has also been collaborating with clinical investigators to address statistical issues in neurology, cardiology, oncology, transplantation, psychiatry, pathology and alternative medicine. He was a co-founding editor of the Contemporary Clinical Trials Communication. He has served as an associate editor for several statistical journals including Journal of American Statistical Association, Statistica Sinica, Lifetime Data Analysis, and is on the editorial board for Kidney International, the Journal of the International Society for Nephrology. He is a Fellow of the American Statistical Association. He is currently President of the International Chinese Statistical Association (ICSA).

**Date: 8 April 2022 (Thursday)**

**Time: 9:00-10:00 (Hong Kong Standard Time GMT +8)**

**Venue: Online Talk via Zoom (Meeting ID: 966 6314 4407)**

**Speaker: Prof. Zhezhen Jin, Columbia University**

**Host: Prof. Xingqiu Zhao, The Hong Kong Polytechnic University**

**Click to join:**

**<https://polyu.zoom.us/j/96663144407?pwd=Mzh6TEpXUU1sRCtyM3ZCYUozV245dz09>**



[Click to join \(Zoom\)](#)

**\*\*\* ALL ARE WELCOME \*\*\***

For enrolment, please send your name and email to [wai-yan.moon@polyu.edu.hk](mailto:wai-yan.moon@polyu.edu.hk) on or before 7 April 2022