THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學

DEPARTMENT OF APPLIED MATHEMATICS
應 用 數 學 系

# The Hong Kong Polytechnic University
# Department of Applied Mathematics

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

## Second and Higher Order Statistical Methods with High-dimensional Data

**By**

**Prof. Cunhui ZHANG**
**Rutgers University**

**Abstract**

In a number of recent developments, second and higher order methods of statistical inference have been proposed in some once differentiable problems. We consider three specific examples. In the first problem, second order Stein's formula as a special form of twice Gaussian integration by parts corrects the bias of the Lasso and other convex regularized estimators in high-dimensional linear regression via an unbiased estimating equation. With an associated central limit theorem as a special form of the Poincare´ inequality for the once differential function in Stein's formula, the de-biased estimator is proven to achieve asymptotic normality and efficiency in regular statistical inference of linear functionals of the regression coefficient vector. In a related problem, the second order Stein methods justify the use of a scaled Mallow's $C_p$ as a selector of an estimator in the Lasso solution path to achieve the performance of the oracle minimizer of the prediction loss within a regret of smaller order than the minimax convergence rate. In the second problem, bootstrap methods have been developed to approximate the distribution of the component-wise maximum of a sum of independent random vectors in high-dimension. A soft-max function was used to smooth the maximum in Lindeberg's and Slepian's interpolations between the sample and bootstrapped versions of the maximum. More recently, the un-smoothed maximum was directly used in the interpolation to achieve sharper convergence rates and to reduce the sample size requirement for the statistical consistency of bootstrap. This motivates a number of second and higher order Gaussian anti-concentration inequalities. In the third problem, a high-order unbiased statistical expansion is developed in the estimation of a general functional of a high-dimensional mean vector. The formula systematically and explicitly solves the associated de-bias problem of all orders. In particular, the method directly provides optimal convergence rates in the estimation of the absolute and fractional norms of a high-dimensional mean vector and other non-smooth additive functionals in the Gaussian sequence model and under low-moment conditions on the noise based on independent and identically distributed observations.

**Biography**

Cun-Hui Zhang, Distinguished Professor of Statistics at Rutgers University, is a Fellow of the Institute of Mathematical Statistics and a Fellow of American Statistical Association. His research interests include high-dimensional data, machine learning, empirical Bayes, time series, nonparametric methods, multivariate analysis, survival data and biostatistics, functional MRI, closed loop diabetes control, and network tomography.

**Date: 27 September 2022 (Tuesday)**
**Time: 11:00-12:00 (Hong Kong Standard Time GMT +8)**
**Venue: Online Talk via Zoom (Meeting ID: 959 7145 5198)**
**Speaker: Prof. Cunhui Zhang, Rutgers University**
**Host: Prof. Xingqiu Zhao, The Hong Kong Polytechnic University**
**Click to join:**
**https://polyu.zoom.us/j/95971455198?pwd=QmtOMGU2SjJQK1RrU0d1amdRRDFKZz09**

Click to join (Zoom)

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

For enrolment, please send your name and email to wai-yan.moon@polyu.edu.hk on or before 26 September 2022