RUTGERS
School of Arts and Sciences

Department of
Statistics and Biostatistics

DEPARTMENT OF STATISTICS AND BIOSTATISTICS

Kengo Kato
Department of Statistical Sciences
Cornell University

Randomized incomplete U-statistics in high dimensions

September 26, 2018
3:20 – 4:20pm
Light refreshments will be served
110 Frelinghuysen Road
Hill Center, Room 552

Abstract: In this talk, I will discuss inference for the mean vector of a high-dimensional U-statistic. In the era of Big Data, the dimension of the U-statistic and the sample size of the observations tend to be both large, and the computation of the U-statistic is prohibitively demanding. Data-dependent inferential procedures such as the empirical bootstrap for U-statistics is even more computationally expensive. To overcome such computational bottleneck, we introduce randomized incomplete U-statistics with sparse weights whose computational cost can be made independent of the order of the U-statistic. We derive non-asymptotic Gaussian approximation error bounds for the randomized incomplete U-statistics in high dimensions, namely in cases where the dimension is possibly much larger than the sample size, for both non-degenerate and degenerate kernels. In addition, we propose generic bootstrap methods for the incomplete U-statistics that are computationally much less-demanding than existing bootstrap methods and establish finite sample validity of the proposed bootstrap methods. Our methods are illustrated on the application to nonparametric testing for the pairwise independence of a high-dimensional random vector under weaker assumptions than those appearing in the literature. This talk is based on a joint work with Xiaohui Chen (UIUC).

Bio: Kengo Kato is an Associate Professor of Department of Statistical Science, Cornell University. Prior to joining Cornell University, he was an Associate Professor of Graduate School of Economics, University of Tokyo. He received his Ph.D. in 2009 from University of Tokyo. He has broad interests in Mathematical Statistics and Econometrics, and has been working on high dimensional statistics, empirical process theory, quantile regression, and measurement error problems.