Cornell CIS
Statistical Science

STATISTICS SEMINAR

Wednesday, January 31, 2017
4:15 pm
G01 Biotechnology

KENGO KATO
Graduate School of Economics,
The University of Tokyo

CLT and bootstrap for sample averages and incomplete U-statistics in high dimensions

In modern statistical problems, there is a great interest in tractable distributional approximations to basic statistics such as sample averages and U-statistics where the dimensionality is possibly much larger than the sample size.

In this talk, I will discuss some recent developments on Gaussian and bootstrap approximations in high dimensions. Specifically, in the first part, we will consider distributional approximations to the sum of independent random vectors, and show that under mild regularity conditions the Gaussian approximation holds uniformly on the hyperrectangles even when the dimension of the random vectors is much larger than the sample size. Building upon the Gaussian approximation result, we also establish finite sample validity of some bootstrap methods in high dimensions.

In the second part, we will consider incomplete U-statistics. In general the computation of (standard) U-statistics tends to be prohibitively demanding even when the sample size and the dimension of the kernel are moderately large. To overcome such computational bottleneck, we shall employ randomized incomplete U-statistics with sparse weights whose computational cost can be made independent of the order of the kernel. Then we derive non-asymptotic Gaussian approximation error bounds for the incomplete U-statistics in high dimensions. In addition, we propose novel and generic bootstrap methods for the incomplete U-statistics that take computational considerations into account, and establish finite sample validity of the proposed bootstrap methods.

Refreshments will be served following the seminar in 1181 Comstock Hall.