



THE UNIVERSITY OF
CHICAGO

DEPARTMENT OF STATISTICS

Statistics Colloquium

ALESSANDRO RINALDO

Department of Statistics & Data Science
Carnegie Mellon University

“High-dimensional CLTs: near-optimal rates and modern inference for OLS regression”

MONDAY, February 6, 2023, at 4:30 PM
Jones 303, 5747 S. Ellis Avenue

ABSTRACT

The central limit theorem (CLT) is a fundamental result in probability and statistics, stating that the average of many independent variables is approximately Gaussian. The CLT underpins numerous and widely-used data analysis methods for estimation, hypothesis testing, constructing confidence intervals, and uncertainty assessment. However, the accuracy of the CLT approximation may degrade significantly in high-dimensional data problems. To address this challenge, a growing body of literature has recently emerged aimed at developing CLT bounds to support valid statistical inference in high dimensions. In this talk, I will introduce a novel and near-rate-optimal CLT for hyper-rectangles that holds under minimal conditions. As an application, I will examine ordinary least squares regression in high-dimensional and model-free settings commonly encountered in data science. I will present bounds for the Gaussian approximation error of the ordinary least squares estimator, yielding practical confidence sets with guaranteed coverage and accuracy. Our results highlight the dependence on the dimensionality and other characteristics of the data-generating distribution, enabling high-dimensional and efficient inference with off-the-shelf methods.

Joint work with Heejong Bong, Arun Kumar Kuchibhotla, and Larry Wasserman.