Statistics Seminar

Department of Mathematics and Statistics

DATE:	Thursday, October 24, 2024
TIME:	1:15pm – 2:15pm
LOCATION:	WH 100E
SPEAKER:	Giles Hooker, University of Pennsylvania
TITLE:	Trees and V's: Inference for Ensemble Models

Abstract

This talk discusses uncertainty quantification and inference using ensemble methods. Recent theoretical developments inspired by random forests have cast bagging-type methods as U-statistics when bootstrap samples are replaced by subsamples, resulting in a central limit theorem and hence the potential for inference. However, to carry this out requires estimating a variance for which all proposed estimators exhibit substantial upward bias. In this talk, we convert subsamples without replacement to subsamples with replacement resulting in V-statistics for which we prove a novel central limit theorem. We also show that in this context, the asymptotic variance can be expressed as the variance of a conditional expectation which is approximated by sampling from the empirical distribution and allows for valid bias corrections. We finish by illustrating the use of these tools in combining or comparing statistical models.

Brief Bio

Giles Hooker is Professor of Statistics and Data Science at the University of Pennsylvania. His work has focussed on statistical methods using dynamical systems models, functional data analysis, and statistical aspects of fair and interpretable machine learning. He is the author of “Dynamic Data Analysis: Modeling Data with Differential Equations” and “Functional Data Analysis in R and Matlab”. Much of his work has been inspired by collaborations particularly in ecology, human movement, and citizen science data.

Professor Hooker earned a PhD in Statistics from Stanford University before doing a post-doctoral fellowship at McGill University. Prior to joining Penn, he served as Professor of Statistics and Data Science at Cornell University and Professor of Statistics at UC Berkeley. He also holds a visiting appointment at the Australian National University.