

**University of California, Irvine
Statistics Seminar**

U-learning for Prediction Inference via Combinatory Multi-Subsampling: With Applications to LASSO and Neural Networks

**Zhe Fei
Assistant Professor
UC Riverside**

**4 p.m., Thursday, April 11, 2024
6011 DBH**

Epigenetic aging clocks play a pivotal role in estimating an individual's biological age through the examination of DNA methylation patterns at numerous CpG (Cytosine-phosphate-Guanine) sites within their genome. However, making valid inferences on predicted epigenetic ages, or more broadly, on predictions derived from high-dimensional predictors, presents challenges. We introduce a new U-learning approach via combinatorial multi-subsampling for making ensemble predictions and constructing prediction intervals for continuous outcomes when traditional asymptotic methods are not applicable. More specifically, our approach conceptualizes the ensemble estimators within the framework of generalized U-statistics and invokes the Hajek projection for deriving prediction variances and intervals with valid conditional coverage probabilities. We applied our approach with two commonly used predictive algorithms, Lasso and deep neural networks (DNNs), and illustrated the validity of inferences with extensive numeric examples. We applied these methods to predict the DNA methylation age (DNAmAge) of patients with various health conditions, aiming to accurately characterize the aging process and potentially guide anti-aging interventions.