THE UNIVERSITY OF CHICAGO | DEPARTMENT OF STATISTICS

Statistics Colloquium

Matteo Sesia
Department of Data Sciences and Operations
USC Marshall School of Business

“Conformal Inference for Open-Set and Imbalanced Classification”

Monday December 01, 2025, at 11:30 AM
Jones 303, 5747 S. Ellis Avenue
Pre-Seminar refreshments will be served in Jones 303 at 11:00 am

Abstract

This talk presents a conformal prediction method for classification in highly imbalanced and open-set settings, where there are many possible classes and not all may be represented in the data. Existing approaches require a finite, known label space and typically involve random sample splitting, which works well when there is a sufficient number of observations from each class. Consequently, they have two limitations: (i) they fail to provide adequate coverage when encountering new labels at test time, and (ii) they may become overly conservative when predicting previously seen labels. To obtain valid prediction sets in the presence of unseen labels, we compute and integrate into our predictions a new family of conformal p-values that can test whether a new data point belongs to a previously unseen class. We study these p-values theoretically, establishing their optimality, and uncover an intriguing connection with the classical Good–Turing estimator for the probability of observing a new species. To make more efficient use of imbalanced data, we also develop a selective sample splitting algorithm that partitions training and calibration data based on label frequency, leading to more informative predictions. Despite breaking exchangeability, this allows maintaining finite-sample guarantees through suitable re-weighting. With both simulated and real data, we demonstrate our method leads to prediction sets with valid coverage even in challenging open-set scenarios with infinite numbers of possible labels, and produces more informative predictions under extreme class imbalance.

Information about building access for persons with disabilities may be obtained in advance by calling the Department Office, at 773-702-8333. If you wish to subscribe to our email list, please use the form located on the bottom of the following webpage: https://stat.uchicago.edu/events/statistics-colloquium/