

## Prediction with Abstention in Classification Problems

**Leying Guan, PhD**  
**Assistant Professor**  
**Department of Biostatistics**  
**Yale University**

### ABSTRACT

We consider the multi-class classification problem when the training data and the out-of-sample test data may have different distributions and propose a method called BCOPS (balanced and conformal optimized prediction sets). BCOPS constructs a prediction set  $C(x)$  as a subset of class labels, possibly empty. It tries to optimize the out-of-sample performance, aiming to include the correct class as often as possible, but also detecting outliers  $x$ , for which the method returns no prediction (corresponding to  $C(x)$  equal to the empty set). The proposed method combines supervised-learning algorithms with the method of conformal prediction to minimize a misclassification loss averaged over the out-of-sample distribution. The constructed prediction sets have a finite-sample coverage guarantee without distributional assumptions. We also propose a method to estimate the outlier detection rate of a given method. We prove asymptotic consistency and optimality of our proposals under suitable assumptions and illustrate our methods on real data examples.

**12:00 Noon, Tuesday, September 10, 2019**  
**47 College Street, Room 106B**  
**11:45 AM - Lunch served outside Room 106B**

