



THE UNIVERSITY OF CHICAGO

Department of Statistics
STATISTICS COLLOQUIUM

STEPHEN BATES

Departments of Statistics and EECS
UC Berkeley

Distribution-Free, Risk-Controlling Prediction Sets

MONDAY, April 12, 2021 at 4:00 PM

Via Zoom (session information will be e-mailed to subscribers)

ABSTRACT

To enable valid statistical inference in prediction tasks, we show how to generate set-valued predictions with black-box models that control various notions of statistical error. Our approach guarantees that the expected loss on future test points falls below a user-specified level, for any predictive model and underlying distribution. Building on conformal prediction, we use a holdout set to calibrate the size of the prediction sets, generalizing the approach to control error notions such as the false rejection rate. We demonstrate our procedure in four large-scale problems: (1) multi-label classification, where each observation has multiple associated labels; (2) classification problems where the labels have a hierarchical structure; (3) image segmentation, where we wish to predict a set of pixels containing an object of interest; and (4) protein structure prediction.