



THE UNIVERSITY OF
CHICAGO

Department of Statistics

MASTER'S THESIS PRESENTATION

JAMES LEINER

Department of Statistics
The University of Chicago

Novel Data Splitting Techniques for Poisson Regression

WEDNESDAY, February 17, 2021, at 10:00 AM
ZOOM Meeting

ABSTRACT

Traditional approaches to data splitting partition available data by allocating individual data points to a selection dataset used by the analyst to inform their choice of model and the remaining data points to be used for conducting inference on the selected model. We propose an alternative approach to data splitting for Poisson regression that exploits the fact that conditioning on the sum of several Poisson-distributed data points allows us to use a logistic model for inference. We demonstrate that this revised method for data splitting tends to better control the false discovery rate during inference, especially in situations where many of the covariates are sparse. We also explore how this approach generalizes to the method of data carving discussed in Fithian et al. (2014).