

On the use of the hurdle LASSO for modeling excess zero count data

Yang-Li Liao, Ph.D.

Department of Statistics, Feng Chia University, Taiwan

Abstract

Dealing with excess zeros often poses challenges in model fitting when using traditional Poisson or other count regression models. To address these issues, hurdle (HP) and zero-inflated Poisson (ZIP) models offer alternative approaches. Specifically, the HP model separates the data into two distinct components: the hurdle and the non-zero count components, enabling separate modeling and estimation. This structure is particularly well-suited for the integration of shrinkage techniques, such as LASSO, to facilitate effective variable selection. The proposed hurdle LASSO (HP LASSO) not only produces a selection path and yields accurate results, but also demonstrates applicability to ZIP data under certain conditions in simulation studies. Moreover, the HP LASSO demonstrates practical efficacy in real-life scenarios, as evidenced by applications to the horseshoe crabs and traffic incidents datasets.