

Variable Selection in GLM with Actuarial Applications

Wenyuan Zheng, University of Connecticut

joint work with Brian M. Hartman, University of Connecticut

When pricing an insurance product, companies collect information from potential policyholders to construct a premium structure. Regardless of the type of insurance product, there is always an infinite amount of information that may be gathered from each policyholder (e.g. credit history, medical history, physical exam). It is important to decide which variables are significant to model a fair and equitable pricing structure. Although an informal process of variable selection is often employed, a more formal statistical procedure is recommended to avoid inefficiencies and imprecision. Bayesian variable selection methods offer such a framework and enable the analyst to (1) incorporate prior beliefs about the significance of specific covariates, (2) include information about the relative cost of the covariates, and (3) obtain the probability that each covariate is included in the model. For demonstration purposes, we fit to Generalized Linear Models of claim frequency and severity to data from Singapore automobile insurers. In this talk, we present preliminary results of our investigation, but we intend to expand these procedures to other claim frequency and severity models.