BIOSTATISTICS SEMINAR

A Scalable L0-Based Sparse Regression Method for Large-Scale Competing Risks Data

This paper develops a scalable L0-based simultaneous variable selection and parameter estimation method for the popular Fine-Gray (1999) proportional subdistribution hazards (PSH) model for large competing risks time-to-event data via the broken adaptive ridge method (BAR). We first establish that the BAR estimator, defined as the limit of an L0-based iteratively reweighted L2-regularization algorithm, is selection consistent and has an oracle property for the PSH model. To further make it scalable to large data, we develop two novel high performance algorithms to overcome its computational bottlenecks: one concerns fitting multiple L2-regularizations required by the BAR estimator and the other pertains to the computation of the log-pseudo likelihood and its derivatives for the PSH model. Specifically, we derive 1) a cyclic coordinate-wise BAR (cycBAR) algorithm which effectively avoids fitting multiple reweigh-ted L2-regularizations, and 2) a two-way linear scan algorithm that reduces the computation costs of the log-pseudo likelihood and its derivatives for the PSH model from the order of O(n^2) to O(n). In comparison with the original BAR algorithm and other current regularization methods for the PSH model, our scalable BAR method has yielded over 1,000-fold speedups in computation time, while showing comparable selection and estimation performance in numerical studies. An illustration of the scalability of various regularization methods for large competing risks data is given using a United States Renal Data System data.

Gang Li

PhD, Professor of Biostatistics and Biomathematics, Fielding SPH, UCLA

Thursday March 28, 2019 **3:30 pm - 4:30 pm**

Blue Cross and Blue Shield of North Carolina Foundation Auditorium

UNC GILLINGS SCHOOL OF GLOBAL PUBLIC HEALTH