

A Brief Introduction to My Research

Yi Li

Department of Biostatistics
University of Michigan, Ann Arbor
Banff Workshop

July, 2022

Main Research Areas



- Risk prediction and assessment
- Data science
- High dimensional inference
- Survival analysis

DOI: 10.1111/biom.13587

BIOMETRIC METHODOLOGY

Biometrics WILEY

Debiased lasso for generalized linear models with a diverging number of covariates

Lu Xia¹ | Bin Nan²  | Yi Li³ 

¹ Department of Biostatistics, University of Washington, Seattle, Washington, USA

² Department of Statistics, University of California, Irvine, California, USA

³ Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, USA
(Email: yili@umich.edu)

Abstract

Modeling and drawing inference on the joint associations between single-nucleotide polymorphisms and a disease has sparked interest in genome-wide associations studies. In the motivating Boston Lung Cancer Survival Cohort (BLCSC) data, the presence of a large number of single nucleotide polymorphisms of interest, though smaller than the sample size, challenges inference on

- the de-biased lasso approach, which assumes sparsity on the inverse information matrix, cannot yield confidence intervals with satisfactory coverage probabilities for high dimensional generalized linear models
- under “large n , diverging p ”, we proposed an alternative de-biased lasso approach by inverting the Hessian matrix without imposing the matrix sparsity assumption
- established the asymptotic distributions, laying the theoretical ground for drawing inference
- examined the joint effects of SNPs on the overall risk of lung cancer
- the results differed by smoking status, suggesting molecular mechanisms of lung cancer differ between smokers and non-smokers

Estimation and Inference for High Dimensional Generalized Linear Models: A Splitting and Smoothing Approach

Zhe Fei

*Department of Biostatistics
UCLA
Los Angeles, California, 90025*

FEIZ@UCLA.EDU

Yi Li

*Department of Biostatistics
University of Michigan
Ann Arbor, Michigan, 48109*

YILI@UMICH.EDU

Editor: Boaz Nadler


Abstract

The focus of modern biomedical studies has gradually shifted to explanation and estimation of joint effects of high dimensional predictors on disease risks. Quantifying uncertainty in these estimates may provide valuable insight into prevention strategies or treatment decisions for both patients and physicians. High dimensional inference, including confidence

- proposed a “splitting and smoothing” approach, which splits samples into two parts, performs variable selection using one part and conducts partial regression with the other part
- averaging the estimates over multiple random splits, we obtained the smoothed estimates, which are numerically stable
- showed the estimates are consistent, asymptotically normal, and constructed confidence intervals with proper coverage probabilities for all predictors



Inference for High-Dimensional Censored Quantile Regression

Zhe Fei^a, Qi Zheng^b, Hyokyoung G. Hong^c, and Yi Li^d 

^aDepartment of Biostatistics, University of California, Los Angeles, CA; ^bDepartment of Bioinformatics and Biostatistics, University of Louisville, KY; ^cDepartment of Statistics and Probability, Michigan State University, East Lansing, MI; ^dDepartment of Biostatistics, University of Michigan, Ann Arbor, MI

ABSTRACT

With the availability of high-dimensional genetic biomarkers, it is of interest to identify heterogeneous effects of these predictors on patients' survival, along with proper statistical inference. Censored quantile regression has emerged as a powerful tool for detecting heterogeneous effects of covariates on survival outcomes. To our knowledge, there is little work available to draw inferences on the effects of high-dimensional predictors for censored quantile regression (CQR). This article proposes a novel procedure to draw inference on all predictors within the framework of global CQR, which investigates covariate-response associations over an interval of quantile levels, instead of a few discrete values. The proposed estimator combines a sequence of low-dimensional model estimates that are based on multi-sample splittings and variable selection. We show that, under some regularity conditions, the estimator is consistent and asymptotically follows a Gaussian process indexed by the quantile level. Simulation studies indicate that our

ARTICLE HISTORY

Received December 2019
Accepted July 2021

KEYWORDS

Conditional quantiles;
Fused-HDCQR;
High-dimensional predictors;
Statistical inference; Survival
analysis

- proposed a new procedure to draw inference on all high dimensional predictors within the framework of global CQR
- combined a sequence of low-dimensional model estimates that are based on multi-sample splittings and variable selection
- analyzed the heterogeneous effects of SNPs residing in lung cancer pathways on patients' survival and showed different roles of SNPs in the high- and low-risk groups, providing more information as to how SNPs affect lung cancer survival

INDIVIDUALIZED RISK ASSESSMENT OF PREOPERATIVE OPIOID USE BY INTERPRETABLE NEURAL NETWORK REGRESSION

Yuming Sun
Department of Biostatistics
University of Michigan, Ann Arbor
ysun@umich.edu

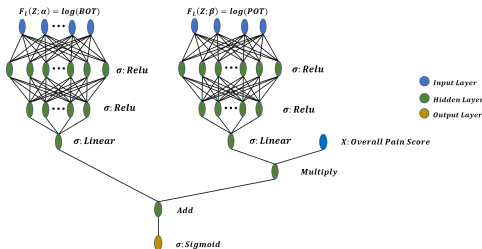
Jian Kang
Department of Biostatistics
University of Michigan, Ann Arbor
jiankang@umich.edu

Chad Brummett
Department of Anesthesiology
University of Michigan, Ann Arbor
chbrummett@umich.edu

Yili
Department of Biostatistics
University of Michigan, Ann Arbor
yili@umich.edu

ABSTRACT

Preoperative opioid use has been reported to be associated with higher preoperative opioid demand, worse postoperative outcomes, and increased postoperative healthcare utilization and expenditures. Understanding the risk of preoperative opioid use helps establish patient-centered pain management. In the field of machine learning, deep neural network (DNN) has emerged as a powerful means for risk assessment because of its superb prediction power; however, the blackbox algorithms may make the results less interpretable than statistical models. Bridging the gap between the statistical and machine learning fields, we propose a novel Interpretable Neural Network Regression (INNER), which combines the strengths of statistical and DNN models. We use the proposed INNER to



- in press at Annals of Applied Statistics
- proposed a novel Interpretable Neural Network Regression (INNER), which combines the strengths of statistical and DNN models
- used the proposed INNER to conduct individualized risk assessment of pre-operative opioid use, leading to more straightforward interpretations of the tendency to use opioids than deep learning

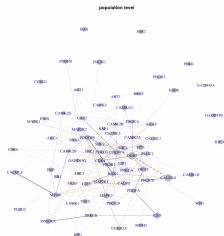
High-Dimensional Gaussian Graphical Regression Models with Covariates

Jingfei Zhang^a and Yi Li^b

^a Department of Management Science, University of Miami, Coral Gables, FL ^b Department of Biostatistics, University of Michigan, Ann Arbor, MI

ABSTRACT

Though Gaussian graphical models have been widely used in many scientific fields, relatively limited progress has been made to link graph structures to external covariates. We propose a Gaussian graphical regressic model, which regresses both the mean and the precision matrix of a Gaussian graphical model on covariates. In the context of co-expression quantitative trait locus (QTL) studies, our method can determine how genetic variants and clinical conditions modulate the subject-level network structures, and recover both the population-level and subject-level gene networks. Our framework encourages sparsity of covariate effects on both the mean and the precision matrix. In particular for the precision matrix, we stipulate simultaneous sparsity, that is, group



- proposed a Gaussian graphical regression model, regressing both the mean and precision matrix on covariates
- used simultaneous sparsity, i.e., group sparsity and element-wise sparsity, on effective covariates and their effects on network edges, respectively

look forward to applications in climate changes

