A Bayesian Estimator of Sample Size

Dehua Bi[1], **Yuan Ji[1]**

[1]*Department of Public Health Sciences, The University of Chicago, IL*

ABSTRACT

We consider a Bayesian estimator of sample size (BESS) and an application to oncology dose optimization clinical trials. BESS is built upon three pillars, Sample size, Evidence from observed data, and Confidence in posterior inference. It uses a simple logic of "given the evidence from data, a specific sample size can achieve a degree of confidence in the posterior inference." The key distinction between BESS and standard sample size estimation (SSE) is that SSE, typically based on Frequentist inference, specifies the true parameters values in its calculation while BESS assumes possible outcome from the observed data. As a result, the calibration of the sample size is not based on type I or type II error rates, but on posterior probabilities. We demonstrate that BESS leads to a more interpretable statement for investigators, and can easily accommodates prior information as well as sample size re-estimation. We explore its performance in comparison to the standard SSE and demonstrate its usage through a case study of oncology optimization trial. BESS can be applied to general hypothesis tests. An R tool is available at https://ccte.uchicago.edu/BESS.

Keywords: Clinical trial; Confidence; Evidence; Hypothesis testing; Posterior inference; Priors; Type I error.

Sampling from the Random Linear Model via Stochastic Localization Up to the AMP Threshold

Han Cui, Zhiyuan Yu, **Jingbo Liu**

Department of Statistics, University of Illinois

ABSTRACT

Recently, Approximate Message Passing (AMP) has been integrated with stochastic localization (diffusion model) by providing a computationally efficient estimator of the posterior mean. Existing (rigorous) analysis typically proves the success of sampling for sufficiently small noise, but determining the exact threshold involves several challenges. In this work, we focus on sampling from the posterior in the linear inverse problem, with an i.i.d. random design matrix, and show that the threshold for sampling coincides with that of posterior mean estimation. We give a proof for the convergence in smoothed KL divergence whenever the noise variance is below the computation threshold for mean estimation introduced by (Barbier et al., 2020). We also show convergence in the Wasserstein distance under the same threshold assuming a dimension-free bound on the operator norm of the posterior covariance matrix, a condition strongly suggested by recent breakthroughs on operator norm bounds in similar replica symmetric systems. A key step in our analysis is to show that phase transition does not occur along the sampling and interpolation paths when the noise variance is below the computation threshold for mean estimation. We also discuss a new method for rigorously proving the consistency of an emerging Thouless-Anderson-Palmer (TAP) approach for mean estimation, which is believed to offer a more robust estimation than the AMP approach. (Based on arXiv:2407.10763 and arXiv:2506.20768)

Keywords: Posterior sampling; Bayesian estimation; Approximate message passing; M-estimation

Back to Sessions List

Bayesian Smoothing and Feature Selection via Variational Automatic Relevance Determination

Zihe Liu, Diptarka Saha, **Feng Liang**

Department of Statistics, University of Illinois at Urbana-Champaign

ABSTRACT

This study introduces Variational Automatic Relevance Determination (VARD), a novel approach for fitting sparse additive regression models in high-dimensional settings. VARD stands out by independently assessing the smoothness of each feature while precisely determining whether its contribution to the response is zero, linear, or nonlinear. Additionally, we present an efficient coordinate descent algorithm for implementing VARD. Empirical evaluations on both simulated and real-world datasets demonstrate VARD's superior performance compared to alternative variable selection methods for additive models.

Keywords: Variational inference; Additive model; Smoothing; Feature selection

Prediction Interval Transfer Learning for Linear Regression Using an Empirical Bayes Approach

Anand Dixit[1], Weining Shen, **Min Zhang[3]**, Dabao Zhang[3]

[1]*Department of Statistics, Purdue University, West Lafayette, Indiana, USA*

[2]*Department of Statistics, University of California, Irvine, California, USA*

[3]*Department of Epidemiology and Biostatistics, University of California, Irvine, California, USA*

ABSTRACT

Current research in transfer learning focuses on improving the predictive performance for small datasets by leveraging information from larger, but potentially biased datasets. However, these methods do not provide prediction intervals, and as a result, one has to either rely solely on the small dataset or combine it with the possibly biased dataset to obtain prediction intervals using traditional linear regression methods. In this project, we propose a new approach, namely the Empirical Bayes approach for prediction interval transfer learning, to calculate prediction intervals within transfer learning for linear regression tasks. We have showed that the Gibbs sampler associated with our method is geometrically ergodic, which allows for the quantification of Monte Carlo uncertainty associated with its predicted values. In addition, the efficiency of our proposed approach is demonstrated through simulation studies and an application to a real-world dataset.

Keywords: Empirical Bayes; Prediction interval; Transfer learning