Statistics Former Student Network (SFSN)
Texas A&M University
Webinar Series

MINSUK SHIN

Assistant Professor, Department of Statistics
University of South Carolina

SCALABLE AND FAST BOOTSTRAP FOR GENERAL MODELS INCLUDING DEEP LEARNING

ABSTRACT:

It has been believed that the virtue of using statistical procedures is on uncertainty quantification in statistical decisions, and the bootstrap method has been commonly used for this purpose. However, nowadays as the size of data massively increases and statistical models become more complicated, the implementation of bootstrapping turns out to be practically challenging due to its repetitive nature in computation. To overcome this issue, we propose a novel computational procedure called Generative Bootstrap Sampler (GBS), which constructs a generator function of bootstrap evaluations, and this function transforms the weights on the observed data points to the bootstrap distribution. The GBS is implemented by one single optimization, without repeatedly evaluating the optimizer of bootstrapped loss function as in standard bootstrapping procedures. As a result, the GBS is capable of reducing computational time of bootstrapping by hundreds of folds when the data size is massive. We show that the bootstrapped distribution evaluated by the GBS is asymptotically equivalent to the conventional counterpart and empirically they are indistinguishable. We examine the proposed idea to bootstrap various models such as linear regression, logistic regression, Cox proportional hazard model, and Gaussian process regression model, quantile regression, etc. The results show that the GBS procedure is not only accelerating the computational speed, but it also attains a high level of accuracy to the target bootstrap distribution. Additionally, we apply this idea to accelerate the computation of other repetitive procedures such as bootstrapped cross-validation, tuning parameter selection, and permutation test.

BIO: Minsuk Shin is an Assistant Professor of Statistics at the University of South Carolina. He is a former student receiving his PhD in 2017 under Drs. Valen Johnson and Anirban Bhattacharya. His research focuses on developing high-dimensional and semi-parametric Bayesian methods motivated by improving practical performance in real world applications and studying theoretical properties of such procedures. His current and previous work focuses on procedures for high-dimensional Bayesian model selection and developing continuous shrinkage priors on functions in nonparametric settings. He has also worked on a scalable stochastic search algorithm to explore the model space of high-dimensional linear models. In general, he is broadly interested in obtaining theoretically-grounded solutions to applied biomedical and machine learning problems. One of his fundamental research goals is to develop general purpose tools for inferences, predictions, and hypothesis testing in high-dimensional settings, while also developing scalable computing algorithms for implementing such tools. He has recently started a new research project for scalable uncertainty quantification via generative processes. The main idea of this is to circumvent a computational bottleneck in traditional frameworks, like posterior computation via MCMC, by constructing a generator of parameter samples from the corresponding posterior distribution.

Thursday, June 25, 2020
1:00 PM - 2:00 PM, CST
Online webinar only. No meeting room.

CLICK HERE TO REGISTER FOR THE WEBINAR
After registering, you will receive a confirmation email containing information about joining the meeting.
Want to Join the SFSN Listserv? Subscribe Here