Research Plan

Yoshiko Hayashi

I would like to apply heavy-tailed modeling to the Bayesian model using the Student *t*-distribution. The main issues are as follows.

(1) Robust Bayesian CART

In statistical inference, removing the influence of outliers is one of important problems. Numerous studies have been conducted on robustness to outliers. However, theoretical research based on Bayesian statistics remains extremely scarce. Within the field of statistical inference, decision trees have recently become frequently employed as predictive models in applied fields such as marketing. This study considers constructing a robust model for a Bayesian regression tree. In regression tree analysis, it is known that results do not perform well when the data are biased. Bootstrap-based methods, such as random forests, are used as a solution to this problem. On the other hand, these methods are computationally heavy.

The representative algorithm for generating regression trees is called CART (Classification and Regression Trees), proposed by Breiman (1984 Chapman and Hall). Chipman et al. (2000, Stat. Comput.) provide a Bayesian CART using hierarchical Bayesian models. Rockva and Van der Pa (2020, Ann. Stat.) theoretically demonstrate that, in the absence of outliers, the trees in Bayesian CART grow to a finite size without overgrowing. However, the case where outliers are included in the regression tree has not yet been studied. This study aims to derive sufficient conditions for resolving this issue by applying heavy-tailed modelling using the Student *t*-distribution to Bayesian CART.

(2) Empirical Local Bayes Correction for Bayesian Modeling

The James-Stein estimator has attracted as an estimator that yields better estimates than the maximum likelihood estimator. In contrast, however, it leads to a mixture distribution with means that are not considered close enough for the problem being handled with large data sets. Therefore, it is not appropriate to apply the James-Stein estimator. There is a local empirical Bayes correction proposed by Efron that yields better estimates for problems involving such a large data set. In this study, we adapt the local empirical Bayes correction to Bayesian modeling and propose a local empirical Bayes correction that is robust to outliers using a hierarchical Bayesian model based on heavy tail modeling.