Research Plan

Yoshiko Hayashi

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

(1) Robust Bayesian CART (joint research with Dr. Shonosuke Sugasawa, Keio University)

We are considering applying this method to Bayesian CART. 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. Our study aims to derive sufficient conditions to solve this problem by applying heavy tail modeling with the Student t-distribution to Bayesian CART.

(2) Objective Robust FDR

The objective of this study is to develop a Robust Bayesian analysis method for large-scale data based on heavy tail modeling. When applying Bayesian analysis to gene expression data, the threshold of false discovery rate (FDR) is arbitrarily determined, as is the significance level of frequentist. In this study, to solve the outlier problem in large data sets, the Student t-distribution will be used to measure the p values for the differences of individual genes. After defining the Bayesian FDR threshold using these statistics, we estimate the null ratio using Storey's q-value method. The estimated value is then used to objectively set the cut point of the FDR.

(3)Empirical Local Bayes Correction for Bayesian modeling

The James-Stein estimator has attracted as a 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 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.