Summary of Research

Yoshiko Hayashi

I have mainly conducted theoretical research on outlier problems using the Student *t*-distribution based on Bayesian models. This is generally referred to as heavy tail modeling, in which the model automatically rejects outliers without a priori identification of outliers.

There are two main methods of analysis in this area of research, the first one using credence based on extreme value theory and the second one assuming log-regularly varying as the error term.

(1) Objective Robust FDR (under submission)

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.

(2) Conditions for Robustness and Limitation on Bayesian Student-t Linear Regression Modeling

This study uses methods based on extreme value theory to derive sufficient conditions for a robust model when heavy-tailed modelling is applied to linear regression analysis and to present intervals where the robust model does not work. Whether the model itself can recognize the observed values as outliers is quite important in heavy-tailed modeling to linear regression analysis. This study shows intervals where the model can function as a robust model using the hat matrix. Furthermore, in these intervals, it shows the sufficient conditions for a regression model based on the t-distribution under Jeffreys prior to be a robust model.

(3) Theoretical properties of Bayesian Student-t linear regression (Statistics and Probability Letters).

In this study, we use heavy-tailed modeling in linear regression analysis assuming log-regularly varying in the error term. We derived sufficient conditions and efficiency for the robust model in the presence of multiple outliers. The results are consistent with methods based on extreme value theory.