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) Bayesian Analysis on Limiting the Student-t Linear Regression Model (under submission).

This study uses methods based on extreme value theory to derive sufficient conditions for a robust model when heavy tail 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 tail modeling to linear regression analysis. This study shows intervals where the model can function as a robust model using 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.

(2) Theoretical properties of Bayesian student-t linear regression (Statistics and Probability Letters).

In this study, we use heavy tail modeling to 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.

(3) Bayesian analysis for mixtures of discrete distributions with a non-parametric component (Journal of Applied Statistics)

The analysis in this study assumes a mixture modeling of outlier and non-outlier distributions, which are often used in outlier problems. The same distributions are generally assumed for mixed distributions for outliers and non-outliers. In this study, constraints on distributions are relaxed by assuming a non-parametric distribution for outliers. The results show that better results are obtained when evaluated with FDR and FNDR.