別表５
(3)

主　論　文　要　旨

No.1

報告番号	甲　乙　第　　　　　号	氏　名	中北　誠

主論文題名：ESSAYS ON BAYESIAN COMPUTATION WITH ANCILLARITY-SUFFICIENCY INTERWEAVING STRATEGY
（ベイズ計算における ANCILLARITY-SUFFICIENCY INTERWEAVING STRATEGY の応用に関する研究）

（内容の要旨）

Researchers have developed more complex models for more realistic data analysis. In general, model complexity tends to increase computational burdens in terms of both computing time and memory/storage usage. As for Bayesian statistics in particular, the model complexity makes statistical inference with the posterior distribution almost intractable and impractical. To tackle this problem, numerous computational methods have been developed since the late 20th century. Among them, the most prominent ones are Markov chain Monte Carlo (MCMC) methods such as the Gibbs sampler (Geman and Geman (1984); Gelfand and Smith (1990)), the Metropolis-Hastings (MH) method (Metropolis et al. (1953); Hastings (1970)) and the data augmentation (Tanner and Wong (1987)). Since the adoption of MCMC in late 1980s, computational Bayesian statistics has attracted more attention for its ability to deal with highly complex problems that were previously unsolvable.

Even after the extraordinary progress of computers in the last several decades, however, naive implementation of MCMC methods is still insufficient to handle the increasing complexity of statistical models. For a high-dimensional complex model, random series of model parameters drawn from the posterior distribution with MCMC often exhibits strong positive autocorrelation. Since such high autocorrelation causes slow convergence to the true posterior distribution, acceptable precision of the posterior statistics cannot be achieved in practice. Although it may be possible to solve this problem by generating a huge sample from the posterior distribution through gigantic computer processing power (e.g., supercomputers), this is neither practical nor eco-friendly.

Numerous studies have been conducted to improve the sampling efficiency of MCMC. Among them, the ancillarity-sufficiency interweaving strategy (ASIS)

主　論　文　要　旨

proposed by Yu and Meng (2011) is an easy-to-implement and widely applicable sampling algorithm for improving the sampling efficiency of MCMC. In principle, ASIS samples random series of parameters and latent variables alternately from the posterior distribution with centralized parametrization (CP) or from the one with non-centralized parametrization (NCP). Whether CP outperforms NCP or not depends on models and data sets and we cannot precisely say which is better in general. ASIS tries to solve this problem by combining two types of parametrization in one sampling cycle of MCMC. In our doctoral dissertation, we will focus on applications of ASIS to several models used in applied econometric analysis and demonstrate the efficacy of ASIS in Bayesian computation. The organization of our dissertation is as follows.

In Chapter 2, the formal definition of ASIS and an illustrative example of its application will be presented. As the example, we estimated a panel data regression model with wage data used by Vella and Verbeek (1998), which is a balanced panel data set, to check whether ASIS can improve the sampling efficiency of random series from the posterior distribution and the precision of Monte Carlo evaluation of the posterior means and variances. As a result, we found that ASIS could help improve MCMC efficiency in panel data regression analysis.

In Chapter 3, we propose a hierarchical Bayesian model of evaluating horse ability and jockey skills in horse racing and estimate it with ASIS. In the proposed method, we aim to estimate unobservable individual effects of horses and jockeys simultaneously along with regression coefficients for explanatory variables such as horse age, racetrack conditions and others in the regression model. The data used in this study are records on 1800-m races (excluding steeplechases) held by the Japan Racing Association from 2016 to 2018, including 22,183 runs with 4,063 horses and 143 jockeys. Since the number of entries varies by racehorses and jockeys, unlike the example in Chapter 2, it is an unbalanced panel data set. We apply the hierarchical Bayesian model to stably estimate such a large amount of individual effects. Since some racehorses and jockeys have extremely small numbers of runs, it is difficult to make stable estimation of the individual effects with conventional sampling methods. Hence, we use the Gibbs sampling coupled with ASIS for

主　論　文　要　旨

Bayesian estimation of the model and choose the best model with the widely applicable information criterion (WAIC) as a model selection criterion. As a result, we found a large difference in the ability among horses and jockeys. Additionally, we observed a strong relationship between the individual effects and the race records for both horses and jockeys.

In Chapter 4, we apply ASIS to complex stochastic volatility (SV) models with high-frequency intraday time series data of stock returns. Since intraday financial data tend to have time-dependent characteristics such as volatility clustering and intraday seasonality, it is crucial to properly capture them. Our modeling strategy is two-fold. First, we model the intraday seasonality of return volatility as a Bernstein polynomial and estimate it along with the stochastic volatility simultaneously. Second, we incorporate skewness and excess kurtosis of stock returns into the SV model by assuming that the error term follows a family of generalized hyperbolic distributions including variance-gamma and Student's t distributions. Furthermore, we developed an efficient MCMC sampling algorithm for Bayesian inference of the proposed model. To improve efficiency of MCMC implementation, we apply ASIS and generalized Gibbs sampling. As a demonstration of our new method, we estimated intraday SV models with 1-minute return data of a stock price index (TOPIX), and conducted model selection among various specifications with WAIC. The result shows that the SV model with the skew variance-gamma error is the best among the candidates.

In Chapter 5, we provide the summary of this dissertation and remarks on the future prospects of further research.