（別紙様式５）

変革を駆動する先端物理・数学プログラム（FoPM）
国外連携機関長期研修　報告書

氏　　名	吉田淳一郎
所属部局	数理科学　研究科　　　　数理科学　専攻
受入先	Université d'Évry Val d'Essonne
日程	西暦 2025 年　2 月　12 日　～　西暦　2025 年　3 月　9 日

Itinerary

In this visit, I collaborated with Professor Arnaud Gloter at University of Évry Val d'Essone. In the first week 2/12~2/16, my supervisor, Professor Nakahiro Yoshida, also participated in our joint research. Some of our discussions were held at the hotel where Professor Nakahiro Yoshida and I stayed, and the other discussions took place at the laboratory of Professor A. Gloter in the university. Through those meetings, I proposed a new method described below, drafted a paper and fixed it with Professor A. Gloter.

Main topic

The main topic of our research is "differential privacy". (The mathematical definition is given in the table below.) This statistical concept is necessary to protect data used in estimation when results of the estimation are released to public. Without such a concept, someone might derive private information from published statistics including published estimators and models. Therefore, there is plenty of demand and research on "differential privacy" to evaluate the risk of personal data being identified from published statistics.

Definition

Let T be a published statistic and X a private data. Let $p(|x)$ be the conditional density of T given $X = x$. Then T is said to be ε-differentially private if the following inequality holds for any t, x, x', where x and x' only differs one component.

$$p(t|x)/p(t|x') \leq e^{\varepsilon}.$$

That is, ε indicates the level of sensitivity of the published statistic T to different values of private data.

Results

In the first half of our research, we discussed some existing method that gives a differentially private estimator called "private point estimator". It is easy to prove differential privacy and efficiency of the private point estimator. However, there are two problems as follows.

1. The differential privacy level ε cannot be reduced to the optimal rate without sacrificing efficiency of the estimator.
2. It takes an enormous amount of computation time to calculate the estimator if there is a lot of data.

Therefore, in the second half of our research, I proposed a new method to improve estimators, such as the above private point estimator, by using some classical algorithm called "multi-step estimation". (We refrain from going into detail in this report because it includes unpublished information.) Thorough this algorithm, Problem 1 is solved except for the term of a logarithm, and Problem 2 is more relaxed. Moreover, in this algorithm, if we adopt another initial estimator that is easier to calculate than the private point estimator, then computation time will be greatly reduced.

After returning to Japan, we have kept in contact with each other. The above results are going to be summarized in a paper and published in the future as our joint research.

Future work: combination of two joint studies

As further research, we aim to apply our proposed method to "non-identifiable models". In the context of parametric estimation, "non-indentifiability" means that the true value of the parameter is not uniquely determined. There are plenty of examples of non-identifiable models, and one of the most important examples is neural network used in machine learning. In many cases, machine learning is performed with private data such as personal location data. Thus, mathematical theory of data protection, including "differential privacy", needs to be developed in this area.

Contrary to this demand, such mathematical theory is not sufficiently developed yet due to the difficulty of specifically analyzing the asymptotic behavior of estimators for non-identifiable model. In the first place, it has been difficult to specifically calculate the asymptotic behavior of estimators even if we do not consider any data protection. In joint research with Professor Nakahiro Yoshida in Japan, however, we had proposed the asymptotic theory of estimators for non-identifiable models by using penalization. Therefore, by applying that theory to the differentially private method we proposed during this visit, it is expected to ensure not only differential privacy but also efficiency of estimators even for non-identifiable models, which will have a huge impact on practical world with machine learning.

Experiences

What I gained most from this visit (except for mathematics) was motivation to improve my own English and other languages including French. I struggled to convey my ideas correctly to Professor A. Gloter throughout those mathematical discussions. Although my English had gradually started to adapt to such communication through this visit, there is still much room for improvement. Furthermore, in France, there were many opportunities to need to speak French even in Paris. Considering that I will visit France repeatedly in the future, I should learn at least a minimum of French. In order to eliminate any barriers due to language, I will continue to work on learning other languages.

Living in Paris also gave me other various experiences such as local culture, delicious food, beautiful atmosphere of magnificent cities, surprise at high prices, convenient transportation (though not as reliable as in Japan) and so on. I also learned the current situation of academia in French university. There were fewer Ph. D students compared to Japan. The educational and support system in Japan for mathematical Ph. D students may be more extensive than in France.

Finally, Professor A. Gloter was a truly sincere host as well as an excellent scholar. I would like to thank him and hope to be an educator or host as him in the future.