別表５
(3)

主　論　文　要　旨　No.1

報告番号	甲　乙　第　　　　号	氏　名	髙畑　圭佑

主 論 文 題 名：

Identification and estimation of the joint distribution of potential outcomes in causal inference

（内容の要旨）

In this thesis, I study the identification and estimation of the joint distribution of potential outcomes in causal inference. In a standard causal analysis, a parameter of interest is often an average treatment effect (ATE), or in recent years, heterogeneous treatment effect (HTE), which is a conditional treatment effect given a subset of observed covariates. For identifying these parameters, unconfoundedness plays a primal role. However, if one wants to identify, for example, the correlation between potential outcomes or quantile of individual treatment effects, unconfoundedness is not sufficient, because they are a functional of the joint distribution of potential outcomes, not of the (conditional) expectations of each potential outcome. Therefore, for valid inference of these parameters, we need to study the identification of the joint distribution of potential outcomes. In this thesis, I study this issue and propose a Bayesian estimation method for several different setups.

In Chapter 1, as a preliminary, I review basic concepts and definitions regarding to identification theory. Identification problem originally arose when researchers tried to apply statistical methods to the estimation of a supply-demand curve in the 1920s. However, it took almost thirty years to be mathematically formalized and integrated as one of the fundamental topics in econometrics. The first successful attempt to systematize identification theory was given in association with simultaneous equation models. Then, for motivating further analyses to be developed later chapters, I illustrate the basic results of identification for simultaneous equation models. More generally, identification is characterized by a probability density function. I define the identification of a model as the injectivity of the density function with respect to the parameter. Identification in a Bayesian approach is also discussed. These days, identification is discussed in various kinds of models, including missing data analysis, causal inference, nonparametric instrumental variable models, and measurement error models. I provide concise reviews for the literature.

In Chapter 2, I consider the identification of the joint distribution of the potential outcomes under various parametric specifications. The key factor of the identification is the non-normality of the distribution of the observed variables, with which we can obtain the information of higher-order moments that are not determined only by mean and variance. In particular, I show the identification of the joint distribution of the potential outcomes when it is specified by a normal mixture. Because any continuous distribution is well-approximated by a finite mixture distribution, our result may cover a wide class of distributions. The

主　論　文　要　旨

identification results derived are useful for estimating quantile treatment effects, causal mediation effects, and heterogeneous treatment effects, which cannot be estimated even if the unconfoundedness assumption is satisfied.

In Chapter 3, I provide sufficient conditions for the identification of the heterogeneous treatment effects, which is defined as the conditional expectation for the differences of potential outcomes given the untreated outcome, under the nonignorable treatment condition and availability of the information on the marginal distribution of the untreated outcome. The identified function is useful both to identify the average treatment effects (ATE) and to determine the treatment assignment policy. The identification holds in the following two general setups prevalent in applied studies: (i) a randomized controlled trial with one-sided noncompliance, or (ii) an observational study with nonignorable assignment in which the information on the marginal distribution of the untreated outcome or its sample moments are available. I propose a Bayesian estimation method for HTE and ATE and examine its properties through simulation studies. I also apply the proposed method to the dataset obtained by the National Job Training Partnership Act Study, which has been analyzed by various researchers. I show a novel finding about women participants, along with consistent results to existing studies.

In Chapter 4, I propose a nonparametric identification condition for the joint distribution of potential outcomes. The key assumptions are two-fold. One is the existence of a control variable, which is employed in various literature in econometric research. The other is the moving support condition, which requires the support of one of the potential outcomes to vary according to the control variable. I demonstrate that the latter assumption is approximately satisfied in the wide class of distributions. Based on these assumptions, the identification problem is shown to reduce to the uniqueness of the solution to a Volterra integral equation of the second-kind and is immediately solved due to the well-known uniqueness result of Volterra integral equations. For estimation, I employ a nonparametric Bayesian density estimation with the data augmentation technique to deal with the missing outcomes. I show the results of numerical experiments that encourage our identification result based on the proposed algorithm.

Finally, I conclude the thesis in Chapter 5.