Introduction to Semiparametric Theory
Prof. Eric Tchetgen Tchetgen
Summer 2021

This course covers sound and principled approaches to inference in large, i.e. non and semiparametric, models and large sample modern tools to analyze the properties of the inferential procedures. Applications of this theory will focus primarily on causality and missing data problems.

Instructor: Eric Tchetgen Tchetgen; **email:** ett@wharton.upenn.edu

Class notes, homeworks and exam.

There is no required textbook although some parts of the course will draw heavily from specific books and articles. I will be mentioning the main source for each part as I present it. You will benefit from complementing the class notes with readings of the specific sections from these sources.

Recommended bibliography: we will be drawing mostly from the following two books

1. *van der Vaart, A. (2000). Asymptotic Statistics. Cambridge University Press.*

2. *Tsiatis, A. (2006). Semiparametric theory and missing data. Springer*

Other bibliography

1. For semiparametric efficiency theory

   (a) Kosorok, M. (2008). Introduction to Empirical Processes and Semiparametric Inference. Springer Series in Statistics.

   (b) Bolthausen, E., Perkins, E. and van der Vaart, A. (2002) Semiparametric Statistics in Lectures on Probability and Statistics, Ecole d'Ete de Probabilites de Saint Flour XXiX -1999. Springer.

   (c) Bickel, P., Klaassen, C.. Ritov, Y., Wellner, J. (1993). Efficient and Adaptive Estimation for Semiparametric Models. Springer

2. **For asymptotic theory**

(a) Wasserman, L. (2007). All of non-parametric statistics. First edition. Springer.

(b) Gyorfi L, Kohler M, Kryzak A, Walk H (2002). A distribution-free theory of Nonparametric Regression.Springer.

(c) Tsybakov, A (2008) An introduction to nonparametric estimation. First edition. Springer.

(d) Pagan,A., Ullah, A. (1999). Nonparametric Econometrics. Cambridge University Press

(e) Härdle, W., Müller, M., Sperlich, S. and Werwatz, A.(2004). Nonparametric and semiparametric models, first edition. Springer.

For the material on semiparametric efficiency theory I will borrow primarily from van der Vaart (2000) and Tsiatis (2008). The second is not quite rigorous but it is conceptual and easier to read than the first. I strongly recommend that you complement the reading of the course notes with this book. The first book is perhaps the most comprehensive (and in my opinion, the best) book written on pretty much all topics that involve asymptotic theory. Several chapters of this book cover rigorously the theory of asymptotic efficiency in parametric and semiparametric models, but it is a hard book to read, as it is too terse and technical. I will lecture at a technical level in between the two books.

For the material on M-estimation I will borrow primarily from van der Vaart (2000). A rigorous and general treatment of the asymptotic distribution of M- estimators requires empirical processes theory. This is an advanced and quite technical area of probability theory that is beyond this course. I will nevertheless invoke results from this theory so that our treatment of the topic is as general and rigorous as possible.

Syllabus

1. **Introduction to semiparametric and parametric models. Introduction to Hilbert space theory.**

2. **Efficiency theory in parametric models.**

   (a) The Cramer Rao bound, Regular parametric models, Regular estimators,

(b) Hajek's representation theorem,

(c) Asymptotically linear estimators, Characterization of the influence functions of regular asymptotically linear estimators, The efficient influence function, The efficient score.

3. **Elements of efficiency theory in semiparametric models**.

   (a) Regular Parametric Submodels, Regular Parameters,

   (b) The Semiparametric Variance bound,

   (c) Pathwise differentiable parameters, Gradients, The Tangent Space.,

   (d) The efficient influence function, Asymptotic Efficiency in Semi Parametric Models The semiparametric efficient score.

4. **Applications: we will cover semiparametric efficient inference in prominent examples:** Classical functionals in the non-parametric model; Location shift model; Parameters of models defined by conditional moment restrictions; Parameters of partially parametric regression models under identity, log and logit link functions; Proportional hazards model; Missing data inference under MAR; causal inference parameters under unconfoundedness. Semiparametric instrumental variable models. Semiparametric proximal causal Inference. Missing data inference under MNAR. Semiparametric sensitivity analysis to selection bias in missing data and causal inference problems.

5. **Elements of asymptotic theory for M-estimators.**

   (a) Estimating equations.

   (b) Consistency of M-estimators. Glivenko-Cantelli Theorem.

   (c) The empirical process and the drift parts of the asymptotic expansion of estimating functions. Donsker classes and stochastic equicontinuity.

   (d) The geometry of M-estimators with nuisance parameters.

   (e) Double robustness.