



The University of Chicago
Department of Statistics

Seminar Series

MARK VAN DER LAAN

Department of Statistics
University of California, Berkeley

Targeted Maximum Likelihood Learning of Scientific Questions

MONDAY, April 16, 2007 at 4:00 PM
133 Eckhart Hall, 5734 S. University Avenue

Refreshments following the seminar in Eckhart 110.

ABSTRACT

The choice of a statistical model and method must be based on careful consideration of the scientific question of interest in order to provide robust tests of the null hypothesis and to minimize bias in the parameter estimate. For this purpose we developed a new generally applicable targeted maximum likelihood estimation methodology.

As an example, I will distinguish between scientific questions concerned with prediction of an outcome based on a set of input variables versus scientific questions in which the goal is to estimate the variable importance or causal effect of one particular variable/treatment. I will show the limitations of fitting regression models for the purpose of learning about a causal effect or variable importance, and present the alternative targeted maximum likelihood approach. Both observational studies and randomized trials will be used to illustrate the advantages of the targeted approach. I will present results from data analyses in which the targeted approach is used to 1) analyze the importance of each of a set of HIV mutations for protease inhibitor resistance and 2) estimate the causal effect of interventions to improve adherence to antiretroviral drugs.

The differences between prediction and causal effect estimation are further highlighted by the additional assumptions needed for the estimation of the causal effect of an intervention in an observational study. Beyond the familiar “no unmeasured confounding” assumption, causal effect estimation also requires an experimental treatment assignment assumption, violation of which can cause severe bias and increased variance in a causal effect estimate. To address this problem, I will show that estimation of the causal effect of a “realistic” intervention (similar to the parameter one estimates in an intention-to-treat analysis) provides an important generalization which can always be fully identified from the data. Targeted estimators of this realistic parameter are also available.

Finally, I will discuss the advantages of applying targeted maximum likelihood estimation in the context of a randomized trial. Like standard approaches, the targeted approach relies only on the randomization assumption. However, the targeted approach yields an improved estimate of the causal effect of a treatment in a randomized trial relative to the commonly used marginal estimate of the treatment effect.

A reference for this presentation which can be downloaded is given by:

Mark J. van der Laan and Daniel Rubin, “Targeted Maximum Likelihood Learning” (October 2006). U.C. Berkeley Division of Biostatistics Working Paper Series. Working Paper 213. <http://www.bepress.com/ucbbiostat/paper213> published in the International Journal of Biostatistics.

Please send email to Mathias Drton (drton@galton.uchicago.edu) for further information. Information about building access for persons with disabilities may be obtained in advance by calling Karen Gonzalez (Department Administrator and Assistant to Chair) at 773.702.8335 or by email (karen@galton.uchicago.edu).