

Subgroup Identification Using Randomized Clinical Trial Data

Jared Foster

Department of Biostatistics
University of Michigan

Wednesday, April 10, 2013

3:30-4:30pm, **New PI 4th Floor Room (4002)**

Light refreshments provided

Abstract

We consider the situation of a randomized clinical trial with a moderate number of baseline covariates. While the overall treatment effect may be modest, it is plausible that there is a subgroup of individuals, defined by their baseline covariate values, that have an enhanced treatment effect. The challenge is to identify a subgroup from the randomized trial data. Such subgroups may be used to aid in future treatment decisions, so it is desirable that they be simple, and depend on only a limited number of covariates. We consider two methods which use randomized clinical trial data to identify such subgroups. The first method, called "Virtual Twins", is algorithmic in nature, and involves two stages. In stage 1, a Random Forest is fit, and predicted subject-specific treatment effects are obtained. In the second stage, these treatment effects are used as the outcome in a single regression tree. An alternative approach is based on the model $y = h(x) + (T - \pi)g(x) + \varepsilon$, where h and g are unknown functions and π is the treatment randomization probability. In the first stage, the unknown functions h and g are estimated using non-parametric regression. The second stage involves selecting the region A which maximizes the expected response to treatment under a "treat-if-in- A " treatment assignment regime. We restrict the form of A to simple regions of the covariate space of up to three dimensions. In this talk I will illustrate the methods on real data and discuss properties of the methods.

Biographical Note

I am a PhD candidate in the Department of Biostatistics at the University of Michigan, working under the supervision of Professors Jeremy MG Taylor and Bin Nan. My doctoral research has been in the area of statistical methods for personalized medicine. Specifically, I have worked to develop

¹ The PI Biostatistics Seminar Series is held on Tuesdays at New York State Psychiatric Institute. If you are interested in receiving regular announcements for our seminars in the future, or if you need further information, please contact Jina James (jamesji@nyspi.columbia.edu, (212) 543-5589).

methods which use randomized clinical trial data to identify simple subgroups of enhanced treatment effect, which may potentially be used to aid in future treatment decisions. I have a great appreciation for statistical methods which are motivated by real-world problems, and will thus see wide use in practice. As a result, I hope that, during my career, I will have many opportunities for collaboration, especially with non-statisticians. I believe this type of collaboration provides an excellent opportunity to learn the science of other fields, and can often lead to the development of very useful new statistical methods.