BIOSTATISTICS SEMINAR

David Page, PhD
Chair of Biostatistics
and Bioinformatics,
Duke University

Probably Approximately Correct Causal Discovery

This talk begins by reviewing some recent applications of machine learning to electronic health records. It then presents and empirically evaluates some novel machine learning algorithms for associating drugs with adverse or fortuitous events they may cause. The general approach is to add to machine learning algorithms a patient-specific baseline risk, as in self-controlled case series. Nevertheless, while these algorithms perform well empirically, unlike much work in causal discovery or causal inference, they do not provide conditions under which provably correct causal discoveries are guaranteed. Much of the rest of machine learning outside of causal discovery is satisfied with probably approximately correct (PAC) models or discoveries, and with algorithms that yield high empirical accuracy rather than perfect predictions. Therefore, analogous to PAC-learning, this talk proposes a theoretical model of "PAC causal discovery (PACC-discovery)." Intuitively, PACC-discovery requires an inference or learning algorithm that can, with high probability, accurately distinguish between two competing causal models, or randomized Turing machines, that differ only by a causal association in question. As with PAC-learning, the inference or learning algorithm receives time and data polynomial in the complexity of the models and the desired accuracy.

Thursday February 20, 2020   **3:30 pm - 4:30 pm**
Blue Cross and Blue Shield of North Carolina Foundation Auditorium

UNC | GILLINGS SCHOOL OF GLOBAL PUBLIC HEALTH