THE UNIVERSITY OF CHICAGO | DEPARTMENT OF STATISTICS

Statistics Colloquium

DANIELA WITTEN
Department of Biostatistics, School of Public Health
University of Washington

"Double dipping: problems and solutions, with application to single-cell RNA-sequencing data"

MONDAY, OCTOBER 10, 2022, at 4:30 PM
Jones 303, 5747 S. Ellis Avenue
Refreshments before the seminar at 4:00 PM in Jones 304.

ABSTRACT

In contemporary applications, it is common to collect very large data sets with the vaguely-defined goal of *hypothesis generation.* Once a dataset is used to generate a hypothesis, we might wish to *test* that hypothesis on the same set of data. However, this type of "double dipping" violates a cardinal rule of statistical hypothesis testing: namely, that we must decide what hypothesis to test before looking at the data. When this rule is violated, then standard statistical hypothesis tests (such as t-tests and z-tests) fail to control the selective Type 1 error --- that is, the probability of rejecting the null hypothesis, provided that the null hypothesis holds, and given that we decided to test this null hypothesis.

While double dipping is pervasive throughout many application areas, in this talk I'll focus on the analysis of single-cell RNA-sequencing data. In the first part of my talk, I'll apply ideas from selective inference to enable valid hypothesis testing after hierarchical clustering or k-means clustering. In the second part of my talk, I'll introduce count splitting, an approach to overcome issues associated with double dipping in the context of latent variable estimation for count-valued data.

This work was conducted in collaboration with UW PhD students Lucy Gao (Biostat PhD 2020), Yiqun Chen (Biostat PhD 2022), and Anna Neufeld (Stat PhD ongoing), as well as Jacob Bien (USC) and Alexis Battle and Joshua Popp (Hopkins).

Information about building access for persons with disabilities may be obtained in advance by calling Shannon Jordan, Department Secretary, at 773-702-8333. If you wish to subscribe to our email list, please visit the following website: https://lists.uchicago.edu/web/info/statseminars.