Combi Seminar

Wednesday, 12.7.22 | Foege Auditorium | 1:30
remote viewing option: **https://depts.washington.edu/gsrestrc/remote.htm**

Dr. Daniela Witten

University of Washington
https://www.danielawitten.com/

"Single-cell RNA-sequencing data analysis without double-dipping"

When analyzing single-cell RNA-sequencing data, we often wish to perform unsupervised learning of latent structure among the cells, and then to test for association between this latent structure and gene expression. For example, we might cluster the cells into cell types, and then test whether gene expression differs between the clusters. Or we might estimate a low-dimensional subspace representing a cellular developmental trajectory, and then test whether gene expression is correlated with this trajectory. However, a classical statistical test of the association between gene expression and the latent structure will not control the Type 1 error, since the latent structure was estimated on the same data used for hypothesis testing. Furthermore, a straightforward sample splitting approach does not fix the problem.

In this talk, I will discuss two solutions to this problem. The first involves selective inference, and the second involves "count splitting", a simple variant of sample splitting that does control the Type 1 error.

This is joint work with PhD student Anna Neufeld, PhD alumni Lucy Gao (now at U. British Columbia) and Yiqun Chen (now at Stanford), and collaborators Jacob Bien (USC) and Alexis Battle and Joshua Popp (Johns Hopkins).