

**BIOMEDICAL DATA SCIENCE
PRESENTS:
BIODS 260C
5/18/23 1:30PM-2:50PM
MSOB X303 (ZOOM LINK BELOW)**

Dan Daniel Erdmann-Pham
Stein Fellow in the
Statistics Department at
Stanford University

Title:

Probing differential expression patterns efficiently and robustly through adaptive linear multi-rank two-sample tests

Abstract:

Two- and K-sample tests are commonly used tools to extract scientific discoveries from data. Naturally, the precise choice of test ought depend on the specifics of the generating mechanisms producing the data: strong parametric assumptions allow for efficient likelihood-based testing, while non-parametric approaches like Mann-Whitney and Kolmogorov-Smirnov-type tests are popular when such prior knowledge is absent. As this talk will argue, practitioners often find themselves in situations of neither full knowledge of all involved distribution nor in full ignorance of them, and therefore are in need of tests that span the spectrum of possible prior knowledge gracefully. It proposes so-called adaptive linear multi-rank statistics as promising candidates for this task, and illustrates their general utility, flexibility (including applications to multiple testing and testing under nuisance alternatives), and computational feasibility on examples from population genetics and single-cell differential expression analysis.

Bio:

I am a statistician working on the rigorous, interpretable, and scalable analysis of data with a specific focus on data arising in biology. Data underpins much modern scientific discovery, which has motivated the development of a rich set of tools to aid its analysis. The field of machine learning in particular has supplied an inventory of quantitative methods ranging from hypothesis testing to function approximation that are available off-the-shelf. However, choosing the most suitable algorithm for a given data set, or indeed whether an algorithm delivering satisfactory performance exists, is often obscured by tacit theoretical assumptions not readily accessible to the user, or a lack of clarity regarding method-specific capabilities and limitations. The broad theme of my work is to bridge such gaps by providing transparent data-analysis schemes for which provable optimality guarantees exist

Zoom link: <https://stanford.zoom.us/j/92124459914?pwd=cFpJYXVLOExUVjMzZkNsYXA0b0RxUT09&from=addon>
Meeting ID: 943 2440 5118
Password: 366430