OSVALDO ESPIN GARCIA, Western University

Using genetic algorithms in the design of two-phase studies

The two-phase study is a cost-effective way to leverage available information in phase 1 of the study to strategically select a most informative subset in phase 2. Expensive information is then collected in the phase 2 subset only, reducing the overall cost. Last, information from both study phases is jointly analyzed by performing statistical inference. Two-phase studies provide a desirable trade-off by economically and strategically using limited resources such as budget without compromising statistical performance by leveraging missing-by-design data methods. A main challenge lies in identifying such an informative subset, which can rely on both outcome and (inexpensive) phase 1 covariates. Genetic algorithms (GAs) are stochastic optimization techniques that mimic nature's evolutionary process. Often used in discrete optimization, GAs offer wide flexibility and ease of implementation. However, these advantages also come with some obstacles, for instance lack of a unique solution and unclear converge criteria are two of the main critiques of these approaches. In this talk, I will present my work on using a GA to identify an informative sample for two-phase fine-mapping studies. I will discuss some of the mathematical and computational challenges found as well as potential future work.