STAT 254 Final Project Proposal

Spencer Slaton and Leon Weingartner

April 3, 2023

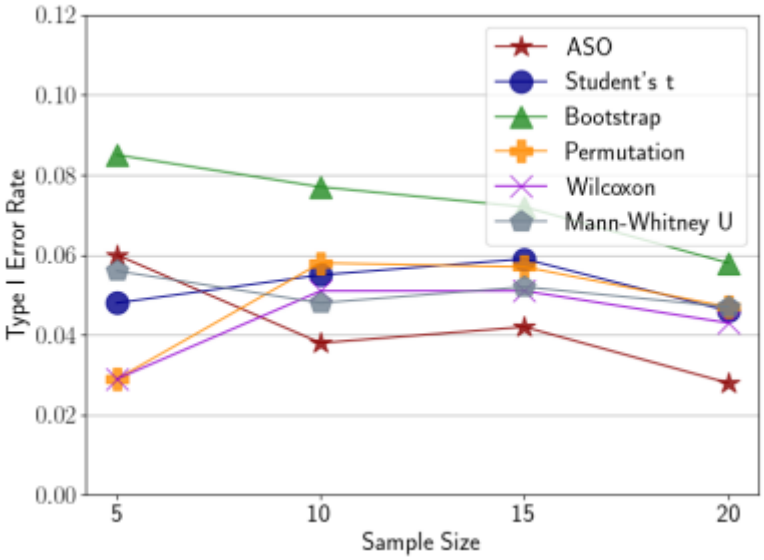
1 OVERVIEW

We propose replicating and extending the results from the 2022 paper EASY AND MEANINGFUL STATISTICAL SIGNIFICANCE TESTING IN THE AGE OF NEURAL NETWORKS by Ulmer, D. et al. The paper examines the relatively novel and assumption free significant test *Almost Stochastic Order* (ASO) and how it compares to other significance tests. ASO is used to compare the variability of two random variables. In model selection, we want to identify significant differences between models instead of just relying on the error criteria to determine which model is "best". We will replicate the figures from the paper on our own dataset and evaluate the performance of ASO on classical ML models in addition to neural networks. This topic is important because significance testing is often overlooked in model selection, especially for models like neural networks where parametric tests are not possible.

2 DETAILS

Our goal is to compare two models via significance testing to see if there is a difference. For interpretability, we will use well-known image classification datasets such as MNIST, CIFAR-10, etc. Ideally, we pick datasets with significant signal to more easily showcase the difference between appropriate and poor models.

We will replicate plots such as the following on a variety of different experimental set-ups.



We can compare models that are known to perform the same and plot the type I error rates for the test. Similarly, we can compare models that are known to perform

differently and plot the type II error rates for the test. The error will be examined over a variety of scoring criteria, covering both regression and classification problems. We would also like to compare the ASO test to parametric tests in the case where the model is a simple linear model.