STAT 778: Midterm Exam

Tom Wallace

March 22, 2018

Preface: Program Organization and Compilation

Source code is contained in `midterm.c`. The program requires the GNU Scientific Library (GSL), an open-source numerical library. It can be obtained from `www.gnu.org/software/gsl`; or, it can installed from most standard Linux package managers. An example command to achieve the latter is:

```
sudo apt-get install gsl-bin libgsl-dev
```

Compilation of `midterm.c` is best achieved in two steps. First, use the below command to compile the program but not link it. You may need to change the argument passed to the `-I` flag to wherever the `gsl` header files live on your computer.

```
gcc -I/usr/include -c midterm.c
```

This command should create an object file `midterm.o`. Link this object file to relevant libraries with the following command. You may need to change the argument passed to the `-L` flag to wherever `libgsl` lives on your computer.

```
gcc -L/usr/lib midterm.o -o midterm -lgsl -lgslcblas -lm
```

Once successfully compiled, the program can be executed. It does not require any arguments. Output is comma-separated text printed to `stdout`. You likely want to pipe this output to a text file, as per the following command:

```
./midterm > output.csv
```

Introduction

This study seeks to compare the performance of the two-sample t-test and the Wilcoxon rank-sum test. It uses simulation to do so. Data is randomly generated under different scenarios. For each scenario, the two methods are used to test the null hypothesis of no difference of means against the simple alternate hypothesis. The goal is to ascertain which method performs better by various criteria.

The remainder of this document is organized into two sections. The **Methods** section provides more detail on how the two methods were implemented and how their performance was compared. The **Simulation Study** section presents output data and results.

Methods

Tests for Difference of Means

The study compares the performance of two different tests of means. The first is **Welch's t-test**. This test assumes that the two populations are independent (i.e. unpaired), that they have normal distributions, and that they may have unequal variances. Although Welch's t-test is possible to perform with different population sizes, this study only examined the case of equal population sizes. The test statistic t is calculated as:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{1}$$

with $\bar{X}_i$, s_i^2, and n_i denoting the sample mean, sample variance, and sample size of group i. The degrees of freedom for the t test statistic are calculated by the Welch-Sattherthwaite equation:

$$df \approx \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{s_1^4}{n_1^2(n_1-1)} + \frac{s_2^4}{n_2^2(n_2-1)}} \tag{2}$$

The second is the **Wilcoxon rank-sum test** (sometimes also called the Mann-Whitney U test). This test—or at least this paper's implementation of it—assumes independent (i.e. unpaired) samples, but makes no parametric assumptions nor any assumptions regarding common variance. Observations from the two groups are pooled, and then sorted and ranked in ascending order (i.e. the smallest observation is 1, the second-smallest is 2, and so on).

- Type I error
- Power
- Robustness to sample size
- Robustness to false parametric assumptions

- Robustness to outliers