STAT 778: Homework 2

Tom Wallace

March 13, 2018

1 Program Organization

Code for **Problem 1** is contained in `hw2_1.c`. It relies upon the `math.h` library and must be compiled with the `-lm` flag. An example compilation command is:

```
gcc hw2_1.c -o hw2_1 -lm
```

Once compiled, execution of `hw2_1` requires one mandatory argument: the name of the data file to be read. An example command is:

```
./hw2_1 HW2_2018.dat
```

Code for **Problem 2** is contained in `hw2_2.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 of this is:

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

Compilation of `hw2_2.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 hw2_2.c
```

This command should create an object file `hw2_2.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 hw2_2.o -o hw2_2 -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:

```
./hw2_2 > output.csv
```

2 Writeup of Problem 2

2.1 Introduction

This document describes a simulation study. The study consisted of generating statistics based on normal random variables and comparing the accuracy of these statistics to their theoretical expectation.

2.2 Methodology

Consider the normal random variable $X \sim N(-0.5, 2)$. For each simulation run, n of such variables were randomly generated. Three different n were used: $n = 50, 100, 200$. For each n, 1000 runs were conducted. On each run, the sample mean and sample variance were computed, as well as their respective standard errors and 95% confidence interval. These statistics were averaged over the 1000 runs conducted for each n. The empirical coverage probability also was computed.

2.3 Results

Simulation results are presented in Table 1. The point estimates closely match the true parameter values. The standard errors of the point estimates decrease as n increases. The empirical coverage probabilities are very near 95%.

Table 1: Simulation results, average of 1000 runs

n	Parameter	True Value	Estimate	SE	95% CI	CP(%)
50	μ	-0.5	-0.487	0.199	(-0.877, -0.097)	95.6
	σ^2	2	1.999	0.404	(1.207, 2.790)	95.0
100	μ	-0.5	-0.496	0.141	(-0.771, -0.220)	94.3
	σ^2	2	1.990	0.283	(1.435, 2.544)	95.1
200	μ	-0.5	-0.498	0.100	(-0.693, -0.302)	94.9
	σ^2	2	1.991	0.200	(1.600, 2.382)	94.4