

STAT 788 - Homework 2

Daniel Hartig

March 14, 2018

1 How to run

There are two source files for the two parts of the homework problem. The first is `partial_like_hw2.c`. This program takes two arguments at the command line, the two arguments `b1` and `b2` which are the parameter values to test log-likelihood against. The homework assignment is compiled by

```
cc -o hw2a partial_like_hw2.c -lm.
```

The data file `HW2_2018.dat` must be in the same folder; then this is executed by

```
./hw2a 0.5 -0.5
```

with the result being that log-likelihood = -684.5.

The second file is `normal_est_hw2.c`. This program takes four arguments, `truemean` and `truevar`, which are the parameters of the normal distribution to sample from; `n` which is the number of normal random variables to pull in each sample; and `iter` which is the number of sample estimates to make. Part a. of the homework is completed by the function `getNormRand`; part b. by the function `estGaussParams`; and part c. is reported and printed to the screen by the `main` function. The main function is compiled by

```
gcc -o hw2b normal_est_hw2.c -lm
```

and executed by

```
./hw2b -0.5 2.0 50 1000
```

substituting 50 for the number of random variables to be generated in each sample.

2 Problem 2

2.1 Methodology

The functionality of Part a. of the homework is accomplished by generating two random normal variables using the Box-Muller transform. Using C, we generate two approximately uniform random variables on $(0, 1)$ using the `rand()` function and use them to generate two random normal variables. This process is repeated until we have generated n random variable. For our three test runs, n will take the values 50, 100, and 200.

The functionality of Part b. of the homework is accomplished by a function that generates parameter estimates for a normal distribution. The mean of n normally distributed random variables X_i is estimated by

$$\hat{\mu} = \frac{\sum X_i}{n}.$$

The variance of this distribution is estimated by

$$\hat{\sigma}^2 = \frac{\sum (X_i - \bar{X})^2}{n - 1}.$$

From the mean and variance, other attributes like the standard deviation and sample standard error and confidence interval can be calculated.

For Part c. we estimate the parameters of a sample of size n , repeating 1000 times and analyzing the result. The sample average and sample standard deviation of the 1000 mean estimates and 1000 variance estimated are both calculated the same way as above, with the standard deviation being the square root of the variance.

The sample means are normally distributed, so the standard error of each mean estimator is

$$SE_{\hat{\mu}} = \frac{s}{\sqrt{n}} = \frac{\sqrt{s^2}}{\sqrt{n}}.$$

A 95% confidence interval is generated using z -values of $z_{0.05} = -1.960$ and $z_{0.95} = 1.960$ to obtain

$$(\hat{\mu} + z_{0.05} \frac{s}{\sqrt{n}}, \hat{\mu} + z_{0.95} \frac{s}{\sqrt{n}}).$$

The variances, on the other hand, are χ^2 distributed with $n - 1$ degrees of freedom. The variance of the variance estimators is

$$\text{Var}(\hat{\sigma}^2) = \frac{2s^4}{n - 1}$$

so the standard error of a variance estimator is then,

$$SE_{s^2} = \sqrt{\frac{2\hat{s}^4}{n - 1}} = \hat{s}^2 \cdot \sqrt{\frac{2}{n - 1}}.$$

The 95% confidence interval according to the χ^2 distribution with $n - 1$ degrees of freedom is

$$\left(\frac{(n - 1)\hat{s}^2}{\chi_{0.025, n-1}^2}, \frac{(n - 1)\hat{s}^2}{\chi_{0.975, n-1}^2} \right).$$

The empirical coverage probability is determined by dividing the number of samples where the true value of the parameter is within the confidence interval by the number of iterations.

2.2 Results

n	Parameter	True Value	Estimate	Std Dev	Avg Std Err	Cover Prob
50	μ	-0.5	-508	0.195	0.200	0.950
	s^2	2.0	2.024	0.446	0.409	0.926
100	μ	-0.5	-.504	0.139	0.140	0.955
	s^2	2.0	1.983	0.306	0.282	0.937
200	μ	-0.5	-0.500	0.098	0.100	0.951
	s^2	2.0	1.995	0.211	0.200	0.945

As the size of each sample within the 1000 iterations increases, the estimate of the true parameter stays roughly the same. Even in the case with the least samples ($n = 50$), there are 50,000 normal variables created, so the estimates of the true parameters are already quite close to their true values.

The standard deviations and standard errors decrease with the square root of the sample size (n), as expected. As sample size increases, both tend towards zero. Both standard deviations and standard errors decrease proportional to each other, so the coverage probability should remain constant. This is observed on the table.