Tecnológico de Monterrey
IN 4027: Data Science and Statistical Inference
September, 27th, 2019

Midterm Exam

Name:_____
Student Number:_____

1. Please perform the computations described on points a) – c) for the dataset $\boldsymbol{x}$ given below.

$\boldsymbol{x}$ = [2.15, 2.12, 3.78, 1.86, 5.01, 6.01, 4.98, 2.78, 3.61, 1.73, 7.47, 3.11, 7.51, 3.58, 3.2]

a) Compute the $\bar{x}$ and median estimators (5 points).
b) Provide the standard error around both of them using bootstrapping with B=5 (10 points).
c) How higher is the standard error for $\bar{x}$ using bootstrapping vs. using the exact formula for $\hat{\sigma}_{\bar{x}}$? (10 points)

2. Say we are estimating parameter λ from a $Poi(\lambda)$ distribution, using both MLE and Bayesian estimation. Assume that the sample mean of your observation vector $\boldsymbol{x}$ is $\overline{x_0}$. Please come up with a conjugate prior for λ such that the Maximum a Posteriori estimate of your Bayesian estimation $\widehat{\lambda^{MAP}}$ is *a little higher* than the MLE estimator $\widehat{\lambda^{MLE}}$. (20 points)

3. Given the two samples $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ from a $N(0, \sigma^2)$ shown below:

	sample_1	sample_2
0	5.291559	-4.340402
1	1.577549	-2.188433
2	-0.922735	2.411433
3	-1.313747	-1.545780
4	2.271424	7.958170
5	3.360876	8.005331
6	4.541458	-4.984437
7	-1.690501	2.968349
8	4.180382	3.033462
9	0.002797	5.576930
10	5.813225	6.308187
11	-1.246751	-3.201286
12	-3.545230	-2.439179
13	-2.914413	1.236574
14	1.673852	0.763359
15	-4.369372	0.366063
16	-2.352658	-3.706139
17	4.086070	10.681323
18	-2.379780	2.319377
19	-0.424038	-1.496508

a) Compute MLE estimators $\hat{\sigma}_1^{MLE}$ and $\hat{\sigma}_2^{MLE}$ for each sample. (10 points)

b) Compute the standard error around your $\hat{\sigma}_1{}^{MLE}$ and $\hat{\sigma}_2{}^{MLE}$ estimators using the *Observed Fisher Information* bound. (15 points)

c) What distributions do your $\hat{\sigma}_1{}^{MLE}$ and $\hat{\sigma}_2{}^{MLE}$ estimates approximately follow? (10 points)

4. A colleague of yours is having trouble distinguishing if the probability of a defect in his products is higher on one produced on Plant 1 vs. Plant 2. The defect rate on Plant 1 follows a $Be(2,2)$ distribution. He suspects that Plant 2 might have a defect rate that is *higher* and better represented by a $Be(3,2)$ distribution.

a) Describe (like a step-by-step algorithm) the hypothesis testing procedure you would recommend your colleague to follow (15 points).

b) Comment on the behavior of Type I Error vs. Type II error that your colleague will experience if he uses this hypothesis test. (10 points).