# Discussion Section #10
# Due: To be submitted to CatCourses by 11:59pm.

## Instructions:

This lab, we will be doing computational exercises which follow Homework #11 and the concepts of maximum likelihood estimation and least squares regression.

Please follow along with your instructor as you go through an examples of each. For this assignment answer the questions in the following section. When specified you should include a figure and/or you R code.

## Assignment:

1. (4 Points) **Maximum Likelihood Estimation.** (Note: This problem is similar but not identical to Question 2 on Homework #11.) Generate $n$ independent uniformly distributed random numbers $(x_1, x_2, \ldots, x_n)$ on an interval from $[0, a]$ where you specify the value of $a$. (Pick any value of $a$ you want, but not too crazy! Try the range $5 \leq a \leq 50$.)

   (a) (1Points) Explain why the likelihood of the data is just:

   $$P(x_1, x_2, ...x_n) = \left(\frac{1}{a}\right)^n.$$

   (b) (1 Point) For the optimize function in R you need to specify whether to minimize or maximize. What is the lowest values that $a$ can be in order for all the data points you generated to have non-zero probability? Is there an upper bound for what values $a$ can have? (If there's no well-defined upper or lower bound, you can just run the optimize function over different possible ranges and see if you get a different estimate.)

   (c) (1 Point) Modify the R code you received to determine the maximum likelihood choice for $a$ from your likelihood. Plot the likelihood and show the maximum likelihood point on the curve. (You should include the figure and your code.)

   (d) (1Points) Increase your value of $n$, explain what happens to your maximum likelihood estimate. Does it get closer to the true value of $a$? Does this make sense? (Hint: What is the probability density function of the maximum of $n$ uniformly distributed random variables?)

2. (3 Points) **Maximum Likelihood Estimation.** The data file "testDataExp.csv" contains a data set of 50 independent points sampled from an exponential distribution with unknown parameter $\lambda > 0$. Modify the R code you were given to determine the maximum likelihood estimate for $\lambda$

   Remember that the probability density function of a normal distribution is given as follows:

   $$f(x) = \lambda e^{-\lambda x}.$$

   Using what you learned in the previous problem.

(a) (1 Point) Explain in mathematics why the following code in R is the log of the likelihood of your data (data) as a function of $\lambda$ (lambda):

```
like <- function(lambda,data) lambda^(length(data))*exp(-lambda*sum(data))
```

(b) (2 Point) Determine the maximum likelihood estimate for $\lambda$. For full credit you must submit:

- (0.5 Points) R code
- (0.5 Points) a figure showing the likelihood curve and the maximum.
- (1 Point) An explanation of your reasoning.
- (Extra Credit - 1 Point) Explain using Calculus what the maximum likelihood estimate should be in terms of your original data.)

3. (3 Points) **Least Squares Regression.** (Note: This problem is similar but not identical to Question 3 on Homework #11.) Suppose that in a certain chemical process the reaction time $Y$ (hours) is measured in response to the temperature $x$ (degrees Fahrenheit, $^\circ F$). The data are assumed to fit a linear model and you will use R to:

(1) Generate "synthetic" data (i.e., you know the answer!)

(2) Fit a linear model to the data.

(3) Analyze the linear model.

In the chamber in which the reaction takes place according to the simple linear regression model:

$$Y = 5.00 - 0.01x + U$$

where $U$ is the error in the reaction time and is known to be normally distributed with mean $0$ and $\sigma^2 = (0.075)^2$.

Since in simple linear regression we assume that the independent variable (in this case temperature) is known exactly, we can also write:

$$Y_x = 5.00 - 0.01x + U.$$

(a) (1 Point) Determine confidence intervals in the estimate of your values for $\alpha$ and $\beta$.

(b) (1 Point) Increase the number of points you sample, how does this impact the ability to estimate the correct parameters? Do the confidence intervals change? Give another set of estimates for $\alpha$ and $\beta$ for your new choice.

(c) (1 Point) Decrease the number of points you sample (less than 30) and increase the variance in the noise $U$. How does this impact the ability to estimate the correct parameters? Give another set of estimates for $\alpha$ and $\beta$ for your new choice.