Instructions: The solutions to this HW will be posted in one week. Please do not write your solutions in red ink as this HW should be self graded or peer graded in red ink. Grades will be awarded for the completion of the problems and having graded the solutions. Correctness or incorrectness of the solutions will not be considered for the overall HW grade.

1 Problem 1

Assume we have one observation X drawn from a Bernoulli distribution with unknown parameter p. p itself follows a beta distribution with shape parameters α and β.

a) Show that the posterior distribution is beta and find its mean and variance (E[p|X] and Var(p|X)).

Hint: The p.d.f. of a beta distribution with parameters α, β is:

\[ f(y) = \frac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha, \beta)} \]  (1)

where \( B(\alpha, \beta) = \frac{\Gamma(\alpha+1)\Gamma(\beta+1)}{\Gamma(\alpha+\beta+1)} \) is the beta function. The mean is \( \frac{\alpha}{\alpha+\beta} \) and the variance is \( \frac{\alpha \beta}{(\alpha+\beta)^2 (\alpha+\beta+1)} \).

b) Find the Maximum a posteriori estimate (MAP) of p.

2 Problem 2

Suppose you sample \( X_1, \ldots, X_n \) (i.i.d.) from the following distribution:

\[ f_\theta(x) = \theta^2 x e^{-\theta x}, \quad 0 < x, \theta < \infty \]  (2)

Estimate \( \theta \) via Maximum Likelihood.

3 Problem 3

The following data set specifies the humidity (in percentage) of the air in 7 different locations. Moreover, it also indicates whether it rained or not in those
locations, where 1 indicates that it rained and 0 otherwise. Suppose we would like to build a linear model that predicts whether it is going to rain or not.

<table>
<thead>
<tr>
<th>Rained (Yes = 1 or No = 0)</th>
<th>1</th>
<th>1</th>
<th>0</th>
<th>1</th>
<th>0</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Humidity</td>
<td>80%</td>
<td>74%</td>
<td>72%</td>
<td>73%</td>
<td>61%</td>
<td>55%</td>
<td>68%</td>
</tr>
</tbody>
</table>

a) State: the predictor, the response and the model. What kind of model is best suited for this? (Hint: Remember what you saw in discussion!)

b) Estimate the parameters of your model via Maximum Likelihood.

4 Problem 4

Suppose you are tasked with predicting the number of shoes you will need to order in order to satisfy the demand for the next month. You have at your disposal the demand for the previous months and how much shoes you ordered. For simplicity, assume that each pair of data (demand and units ordered) are independent across months.

<table>
<thead>
<tr>
<th>demand</th>
<th>21</th>
<th>22</th>
<th>23</th>
<th>28</th>
<th>31</th>
<th>35</th>
<th>43</th>
<th>45</th>
<th>56</th>
<th>57</th>
<th>61</th>
<th>65</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantity Ordered</td>
<td>145</td>
<td>400</td>
<td>413</td>
<td>427</td>
<td>388</td>
<td>569</td>
<td>540</td>
<td>577</td>
<td>750</td>
<td>767</td>
<td>855</td>
<td>898</td>
</tr>
</tbody>
</table>

a) State: the predictors, the response and the linear model.

b) Estimate the parameters of the linear model using least squares.

c) Compute the estimate for the estimation errors of your model. Does it display Heteroscedasticity?

d) Estimate the parameters of the linear model using weighted least squares, with the following weighting function:

\[ w(x) = \begin{cases} 
 1/8, & \text{if } x < 40 \\
 1, & \text{otherwise} 
\end{cases} \quad (3) \]

e) Compute the \( R^2 \) of both models. Which one is better, via this metric?