5^{th} Homework

Thomas Saltos

Exercise 1

a) Considering that the measurements of the random variable x are independent, we get the likelihood function:

$$L(\theta) = \prod_{i=1}^{N} p(x_i;\theta) = \prod_{i=1}^{N} \theta^2 x_i \exp\left(-\theta x_i\right),\ x_i \geq 0$$

The log-likelihood function is the following:

$$lnL(\theta) = \sum_{i=1}^{N} \ln\left(\theta^2 x_i \exp(-\theta x_i)\right) = 2N\theta + \sum_{i=1}^{N} lnx_i - \theta \sum_{i=1}^{N} x_i$$

The maximum likelihood estimate can be obtained by taking the gradient of the log-likelihood function and equating to 0:

$$\frac{\partial lnL(\theta)}{\partial\theta} = 0 \Leftrightarrow \sum_{i=1}^{N}\left(\frac{2}{\theta_{ML}} - x_i\right) = 0 \Leftrightarrow \theta_{ML} = \frac{2N}{\sum_{i=1}^{N} x_i}$$

b) The maximum likelihood estimate for the given values is the following:

$$\widehat{\theta_{ML}} = \frac{2*5}{2+2.2+2.7+2.4+2.6} = \frac{10}{11.9}$$

The mean of the random variable x is: $\hat{\mu} = \frac{2}{\widehat{\theta_{ML}}} = 2.38$

The formula of Erlang distribution that explains these data is:

$$\widehat{p(x_\iota)} = \widehat{\theta_{ML}^2} x_i \exp(-\theta x_i)\,, i = 1, \dots, 5$$

c) $\widehat{p(2.1)} = 0.84^2 * 2.1e^{-0.84*2.1} \cong 0.25$
$\widehat{p(2.3)} = 0.84^2 * 2.3e^{-0.84*2.3} \cong 0.23$
$\widehat{p(2.9)} = 0.84^2 * 2.9e^{-0.84*2.9} \cong 0.17$

Exercise 2

a) The MAP estimate of θ:

$$\theta_{MAP} = argnaxp(x|\theta)p(\theta) = argmax\ln\left(p(x|\theta)p(\theta)\right)$$

It is $\ln\left(p(x|\theta)p(\theta)\right) = \sum_{i=1}^{N} \theta^2 x_i \exp(-\theta x_i) + \frac{1}{2}\ln(2\pi) - \ln\sigma_0 - \frac{1}{2}(\theta - \theta_0)^2$
Taking the gradient of the above wrt θ and equating to 0 we have:

$$\frac{\partial \ln(p(x|\theta)p(\theta))}{\partial\theta} = 0 \Rightarrow \frac{2N}{\theta_{MAP}} - \sum_{i=1}^{N} x_i - \frac{(\theta_{MAP}-\theta_0)}{\sigma_0^2} = 0 \Rightarrow$$

$$\theta_{MAP}^2 - (\theta_0 - \sigma_0^2 \sum_{i=1}^{N} x_i)\theta_{MAP} - 2N\sigma_0^2 = 0 \Rightarrow$$

$$\theta_{MAP} = \frac{1}{2}\left((\theta_0 - \sigma_0^2 \sum_{i=1}^{N} x_i) + \sqrt{\left(\theta_0 - \sigma_0^2 \sum_{i=1}^{N} x_i\right)^2 + 8N\sigma_0} \right)$$

The other solution of θ_{MAP} has been rejected since it leads to a negative estimation of θ which is not appropriate for the Erlang distribution.

b)

i. $N \to +\infty$ If we divide the following equation by N, we get:

$$\frac{2N}{\theta_{MAP}} - \sum_{i=1}^{N} x_i - \frac{(\theta_{MAP}-\theta_0)}{\sigma_0^2} = 0 \Rightarrow \frac{2}{\theta_{MAP}} - \frac{\sum_{i=1}^{N} x_i}{N} - \frac{(\theta_{MAP}-\theta_0)}{N\sigma_0^2} = 0$$

It can be noticed that as $N \to +\infty \Rightarrow \frac{(\theta_{MAP}-\theta_0)}{N\sigma_0^2} \to 0$. In this case $\theta_{MAP} \to \theta_{ML}$.

ii. For large values of σ_0^2 the term $\frac{(\theta_{MAP}-\theta_0)}{N\sigma_0^2} \to 0$ and again $\theta_{MAP} \to \theta_{ML}$
This case indicates that the prior knowledge regarding θ_0 is not reliable.

iii. However, the small variance increases the importance of the prior knowledge regarding θ_0 and $\theta_{MAP} \to \theta_0$.

c) $\theta_{MAP} = \frac{1}{2}\left((\theta_0 - \sigma_0^2 \sum_{i=1}^{N} x_i) + \sqrt{\left(\theta_0 - \sigma_0^2 \sum_{i=1}^{N} x_i\right)^2 + 8N\sigma_0} \right)$

If we substitute the given data in the above equation, we get:

$$\widehat{\theta_{MAP}} = 0.855$$

The mean of the random variable x is: $\hat{\mu} = \frac{2}{\widehat{\theta_{MAP}}} = 2.33$

The formula of Erlang distribution that explains these data is:

$$\widehat{p(x_i)} = \widehat{\theta_{MAP}^2} x_i \exp(-\theta x_i), i = 1,2,3,4,5$$

d) $\widehat{p(2.1)} = 0.855^2 * 2.1e^{(-0.855*2.1)} \cong 0.25$
$\widehat{p(2.3)} = 0.855^2 * 2.3e^{(-0.855*2.3)} \cong 0.23$
$\widehat{p(2.9)} = 0.855^2 * 2.9e^{(-0.855*2.9)} \cong 0.17$

e) The results of this exercise are similar to the results of exercise 1, since $\widehat{\theta_{MAP}}$ is really close to $\widehat{\theta_{ML}}$.

Exercise 3

We take the derivative of the Lagrangian function and equating to 0:

$$\frac{\partial L(\mu)}{\partial \mu} = 0 \Rightarrow -2\sum_{n=0}^{N} x_n - \mu + 2\lambda(\mu - \mu_0) = 0 \Rightarrow 2\lambda(\mu - \mu_0) = 2\sum_{n=0}^{N} x_n - \mu \Rightarrow \widehat{\mu_{RR}} = \frac{\sum_{n=0}^{N} x_n + \lambda\mu_0}{N+\lambda}$$

Exercise 4

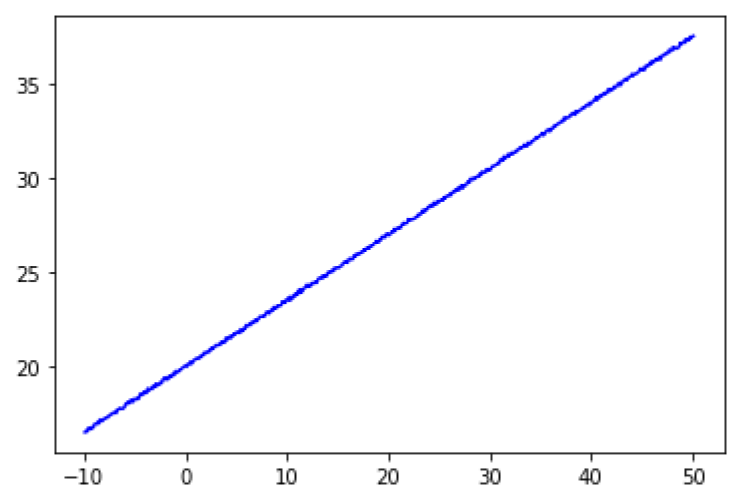
i.

```
import scipy.io as sio
import numpy as np
import numpy.matlib
import matplotlib.pyplot as plt
import matplotlib.image as mpimg
import scipy.signal as scsig

mat = sio.loadmat('HW5.mat')

training=mat.get('Data')

plt.plot(training.T[0],training.T[1],'b')
plt.show()
```



```
N = len(training)
th = training.T[0]
R = training.T[1]
m = np.mean(training,axis=0)
S = 1/(N)*((training - np.matlib.repmat(m,N,1)).T).dot(training - np.matlib.repmat(m,N,1))
print('mean=',m)
print('S=',S)
```

```
mean= [20.00158101 26.99674366]

S= [[300.96444327 105.36107986] [105.36107986 36.88748843]]
```

ii.

```
def Eyx(x,mean,S):
  return mean[1] + (S[0,1]/np.sqrt(S[0,0]*S[1,1]))*np.sqrt(S[1,1]/S[0,0])*(x-mean[0])

testing = mat.get('Data_test')
R_est = Eyx(testing[:,0],m,S)
mse_err = np.mean((R_est - testing[:,1])**2,axis=0)
mse_err
```

```
2.514438744054635e-05
```

We know that the joint pdf of the random variables θ and R is normal distribution, so we use the MSE criterion. Then we estimate the values of the testing dataset and we find the mean square error. As the mean square error indicates the performance of the regressor is good.