Assignment #3

Yutae Lee (626005947)

2022-09-27

Problem 3.8

Coins: Diaconis and Ylvisaker (1985) suggest that coins spun on a flat surface display long-run frequencies of heads that vary from coin to coin. About 20% of the coins behave symmetrically, whereas the remaining coins tend to give frequencies of 1/3 or 2/3.

a) Based on the observations of Diaconis and Ylvisaker, use an appropriate mixture of beta distributions as a prior distribution for , the long-run frequency of heads for a particular coin. Plot your prior.

We know that 20% have a symmetrical percentage for heads and tails. So we can have first $0.2 * beta(\theta|1,1)$. Then we do not know rather head is $1/3$ or tail is $1/3$, so we should assign 40% for each in order to make them equal. Then we get $0.4 * beta(\theta|1,2) + 0.4 * beta(\theta|2,1)$

In total:

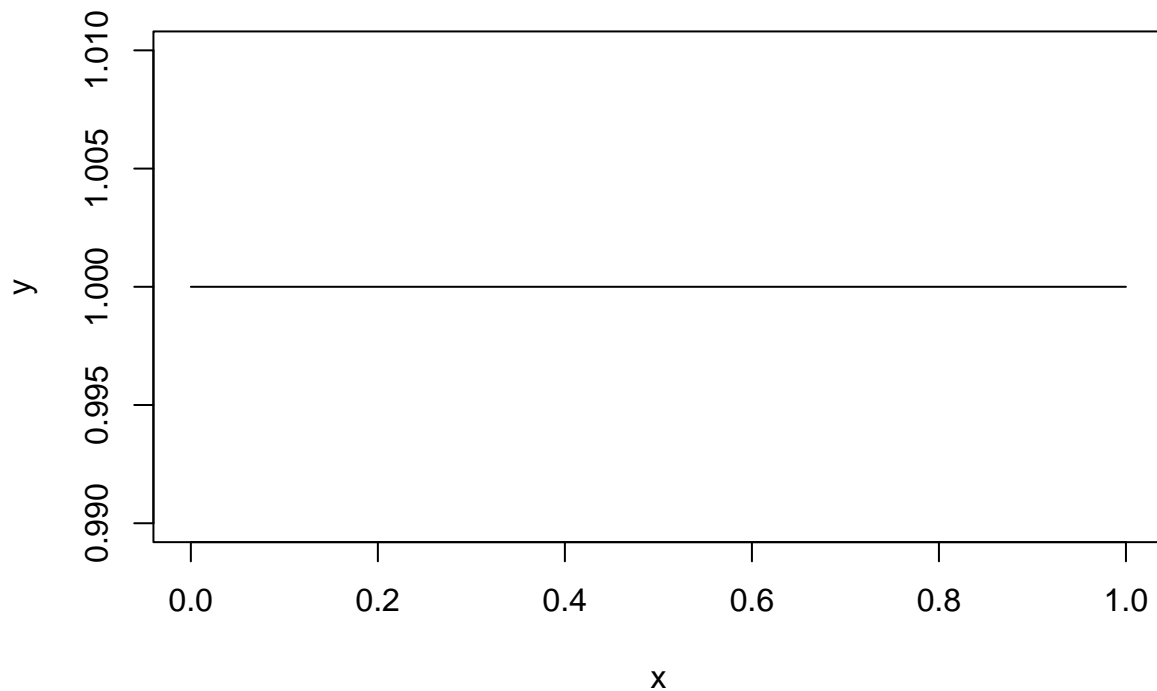$$0.2 * beta(\theta|1,1) + 0.4 * beta(\theta|1,2) + 0.4 * beta(\theta|2,1)$$

which is equal to

$$0.2 * \frac{1}{B(1,1)} + 0.4 * \frac{1-\theta}{B(1,2)} + 0.4 * \frac{\theta}{B(2,1)}$$

where $B(a,b) = \frac{\gamma(a)\gamma(b)}{\gamma(a+b)}$

```
curve(0.2 * dbeta(x,1,1) + 0.4*dbeta(x,1,2) + 0.4*dbeta(x,2,1), ylab = 'y')
```

b) Choose a single coin and spin it at least 50 times. Record the number of heads obtained. Report the year and denomination of the coin.

Using the 2014 year coin, I got $y = 28$ heads.

c) Compute your posterior for , based on the information obtained in b).

Because we want to have conjugate with our binomial model, we will have each component each of the posterior as a beta distribution.

The posterior would be as following:

$$p(\theta|y) = \frac{.2 * B(29,22)}{K} * beta(\theta|29,22) + \frac{.4 * B(29,23)}{K} * beta(\theta|29,23) + \frac{.4 * B(30,22)}{K} * beta(\theta|30,22)$$

where $K = 0.2 * B(29,22) + 0.4 * B(29,23) + 0.4 * B(30,22)$

which we can compute using r:

```r
B_func <- function(a,b){
  gamma(a)*gamma(b)/gamma(a+b)
}
K = 0.2 * B_func(29,22) + 0.4 * B_func(29,23) + + 0.4 * B_func(30,22)

print(.2 * B_func(29,22)/K)
```

```
## [1] 0.3333333
```

```
print(.4 * B_func(29,23)/K)
```

```
## [1] 0.2875817
```

```
print(.4 * B_func(30,22)/K)
```

```
## [1] 0.379085
```

So,

$$p(\theta|y = 28) = 0.333 * beta(\theta|29, 22) + 0.288 * beta(\theta|29, 23) + 0.379 * beta(\theta|30, 22)$$

d) Repeat b) and c) for a different coin, but possibly using a prior for that includes some information from the first coin. Your choice of a new prior may be informal, but needs to be justified. How the results from the first experiment influence your prior for the of the second coin may depend on whether or not the two coins have the same denomination, have a similar year, etc. Report the year and denomination of this coin

d)b)

Using coin from 2009, I got y = 18. As informed in the problem, this time I will use prior distribution from last problem

d)c)

$$p(\theta|y) = \frac{.2 * B(47, 54)}{K} * beta(\theta|47, 54) + \frac{.4 * B(47, 55)}{K} * beta(\theta|47, 55) + \frac{.4 * B(48, 54)}{K} * beta(\theta|48, 54)$$

where $K = 0.2 * B(47, 54) + 0.4 * B(47, 55) + 0.4 * B(48, 54)$

which we can compute using r:

```
B_func <- function(a,b){
  gamma(a)*gamma(b)/gamma(a+b)
}
K = 0.2 * B_func(47,54) + 0.4 * B_func(47,55) + + 0.4 * B_func(48,54)

print(.2 * B_func(47,54)/K)
```

```
## [1] 0.3333333
```

```
print(.4 * B_func(47,55)/K)
```

```
## [1] 0.3564356
```

```
print(.4 * B_func(48,54)/K)
```

```
## [1] 0.310231
```

So,

$$p(\theta|y = 18) = 0.333 * beta(\theta|47, 54) + 0.356 * beta(\theta|47, 55) + 0.310 * beta(\theta|48, 54)$$

Problem 3.9

Galenshore distribution: An unknown quantity Y has a $Galenshore(a, \theta)$ **distribution if its density is given by**

$$p(y) = \frac{2}{\gamma(a)} \theta^{2a} y^{2a-1} e^{-\theta^2 y^2}$$

for $y > 0, \theta > 0$ **and** $a > 0$**. Assume for now that a is known. For this density,**

$$E[Y] = \frac{\gamma(a + 1/2)}{\theta \gamma(a)}, E[Y^2] = \frac{a}{\theta^2}$$

.

a) Identify a class of conjugate prior densities for θ**. Plot a few members of this class of densities.**

We know that $p(y) = \frac{2}{\gamma(a)} \theta^{2a} y^{2a-1} e^{-\theta^2 y^2}$. Using the fact from the next part, it is evident that the conjugate prior of Galenshore would be Galenshore distribution itself.

```r
gs <- function(a, theta, y){
  p = (2/gamma(a)) * theta^(2*a) * y ^ (2*a - 1) * exp((-(theta^2 * y ^ 2)))
  return(p)
}


y <- seq(0, 10, length.out = 100)

df1 <- data.frame(density = gs(1,1, y), type = "gs(1,1)")
df2 <- data.frame(density = gs(2,1,y), type = "gs(2,1)")
df3 <- data.frame(density = gs(1,2,y), type = "gs(1,2)")
df4 <- data.frame(density = gs(0.5,0.5,y), type = "gs(0.5,0.5)")
df5 <- data.frame(density = gs(10,5,y), type = "gs(10,5)")


dens <- rbind(df1, df2, df3, df4,df5)
yval <- rep(y, 5)

df <- data.frame(yvalues = yval, dens)


library(ggplot2)


## Warning:    'ggplot2' R    4.1.3

g = ggplot(df, aes(x = yval, y = density, group = type, color  = type)) + geom_line() + xlab("y")
g
```
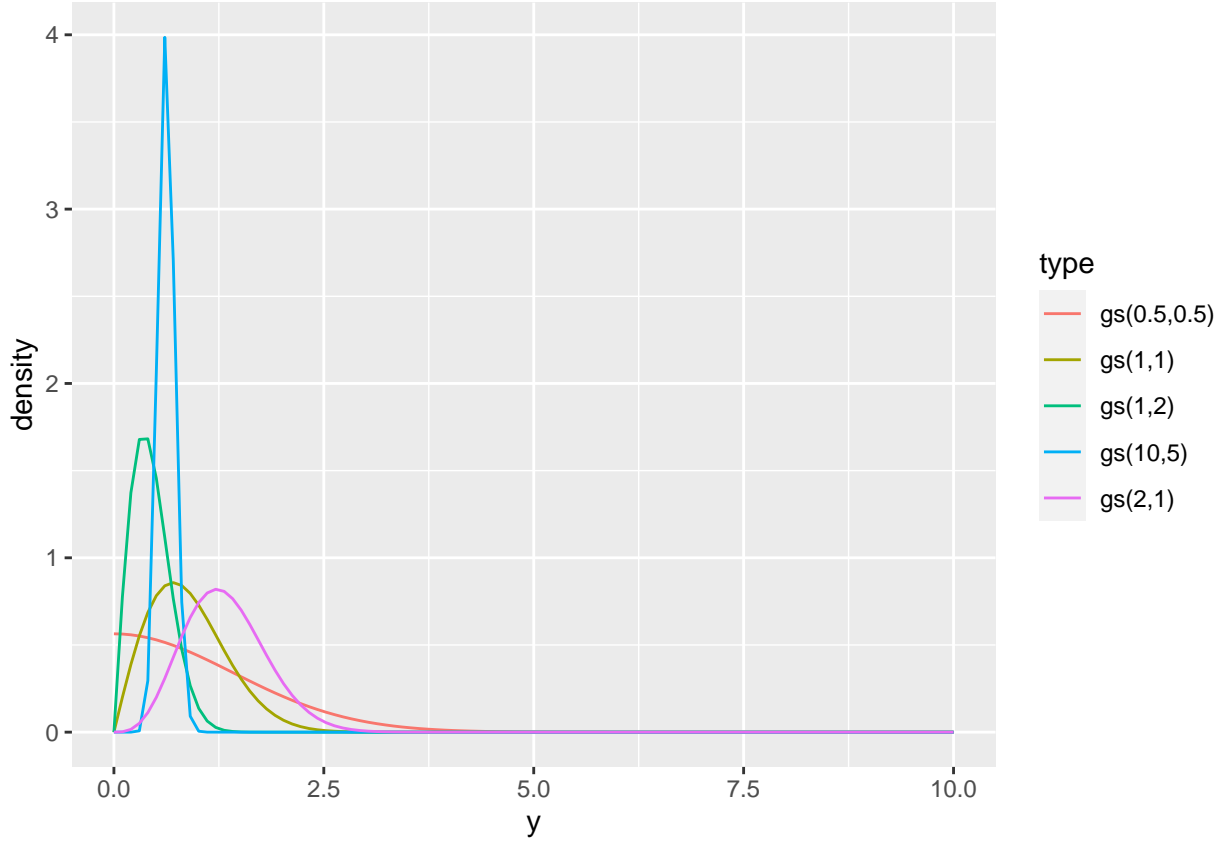
b) Let Y1, . . . , Yn   i.i.d. $Galenshore(a, \theta)$. **Find the posterior distribution of** θ **given Y1, . . . , Yn, using a prior from your conjugate class.**

we can see that:

$$p(\mathbf{y}) = \prod_{i=1}^{n} \frac{2}{\gamma(a)} \theta^{2a} y_i^{2a-1} e^{-\theta^2 y_i^2}$$

and that:

$$p(\theta|\alpha, \beta) = \frac{2}{\gamma(\alpha)} \beta^{2\alpha} \theta^{2\alpha-1} e^{-\beta^2 \theta^2}$$

Then,

$$\prod_{i=1}^{n} \frac{2}{\gamma(a)} \theta^{2a} y_i^{2a-1} e^{-\theta^2 y_i^2} \frac{2}{\gamma(\alpha)} \beta^{2\alpha} \theta^{2\alpha-1} e^{-\beta^2 \theta^2}$$

$$= \frac{2^n}{\gamma(a)^n} \theta^{2an} * \prod_{i=1}^{n} y_i^{2a-1} * e^{\sum_{i=1}^{n} -\theta^2 y_i^2} * \frac{2}{\gamma(\alpha)} \beta^{2\alpha} \theta^{2\alpha-1} e^{-\beta^2 \theta^2}$$

$$= \frac{2^n \beta^{2\alpha} \theta^{2(na+\alpha)-1} e^{-\theta^2 \beta^2 - \sum_{i=1}^{n} \theta^2 y_i^2}}{\gamma(a)^n} * \prod_{i=1}^{n} y_i^{2a-1}$$

We can see that this $p(\theta|\mathbf{y}) \propto Galenshore(\theta|na + \alpha, \sqrt{\beta^2 + \sum_{i=1}^{n} y_i^2})$

5

c) Write down $p(\theta_a|Y_1,...,Y_n)/p(\theta_b|Y_1,...,Y_n)$ **and simplify. Identify a sufficient statistic.**

we can see that:

$$p(\theta_a|Y_1,...,Y_n)/p(\theta_b|Y_1,...,Y_n) = \frac{\theta_a^{2(na+\alpha)-1}e^{-\theta_a^2\beta^2-\sum_{i=1}^n\theta_a^2y_i^2}}{\theta_b^{2(na+\alpha)-1}e^{-\theta_b^2\beta^2-\sum_{i=1}^n\theta_b^2y_i^2}}$$

$$\frac{\theta_a}{\theta_b}^{2(na+\alpha)-1}*e^{(\theta_b^2-\theta_a^2)(\beta^2+\sum_{i=1}^n y_i^2)}$$

d) Determine $E[\theta|y_1,...,y_n]$.

we can see that $E[Y] = \frac{\gamma(a+1/2)}{\theta\gamma(a)}$

Then,

$$E[\theta|y_1,...,y_n] = \frac{\gamma(na+\alpha+\frac{1}{2})}{\sqrt{\beta^2+\sum_{i=1}^n y_i^2}*\gamma(na+\alpha)}$$

e) Determine the form of the posterior predictive density $p(\tilde{y}|y_1,...,y_n)$.

$p(\tilde{y}|y_1,...,y_n)$ is actually $\propto p(\tilde{y})$ which is equal to

$$p(\tilde{y}) = \frac{2}{\gamma(a)}\theta^{2a}\tilde{y}^{2a-1}e^{-\theta^2\tilde{y}^2}$$

Also from b) we know that $p(\theta|y) \propto Galenshore(\theta|\alpha,\beta)$ which is equal to

$$\frac{2}{\gamma(na+\alpha)}*(\sum_{i=1}^n y_i^2+\beta^2)^{2(na+\alpha)}*\theta^{2(na+\alpha)-1}e^{-(\sum_{i=1}^n y_i^2+\beta^2)^2\theta^2}$$

so the posterior is $\int_\Theta p(\tilde{y}|\theta)p(\theta|y)d\theta$

which is equal to

$$\int_\Theta (\frac{2}{\gamma(a)}\theta^{2a}\tilde{y}^{2a-1}e^{-\theta^2\tilde{y}^2})*(\frac{2}{\gamma(na+\alpha)}*(\sum_{i=1}^n y_i^2+\beta^2)^{2(na+\alpha)}*\theta^{2(na+\alpha)-1}e^{-(\sum_{i=1}^n y_i^2+\beta^2)^2\theta^2})d\theta$$

$$= \frac{4}{\gamma(a)\gamma(na+\alpha)}*\int_\Theta \theta^{2(n+1)a+2\alpha-1}e^{-\theta^2(\tilde{y}^2+(\sum_{i=1}^n y_i^2+\beta^2)^2)}\tilde{y}^{2a-1}*(\sum_{i=1}^n y_i^2+\beta^2)^{2na+2\alpha}$$

$$= \frac{4\tilde{y}^{2a-1}(\sum_{i=1}^n y_i^2+\beta^2)^{2na+2\alpha}}{\gamma(na+\alpha)}*\int_\Theta \theta^{2((n+1)a+\alpha)-1}e^{-\theta^2(\tilde{y}^2+(\sum_{i=1}^2 y_i^2+\beta^2)^2)}d\theta$$

From this we can say that the integral is actually similar to $Galenshore(\theta|(n+1)a+\alpha, \sqrt{\tilde{y}^2+(\sum_{i=1}^n y_i^2+\beta^2)^2})$.

Now we get,

$$\int_\Theta \frac{2(\tilde{y}^2+(\sum_{i=1}^n y_i^2+\beta^2)^2)^{(n+1)a+\alpha}*\theta^{2((n+1)a+\alpha)-1}e^{-\theta^2(\tilde{y}^2+\sum_{i=1}^n y_i^2+\beta^2)^2}}{\gamma(a(n+1)+\alpha)}d\theta = 1$$

we can now simply the same terms to get

$$p(\tilde{y}|y) = \frac{4\tilde{y}^{2a-1}(\sum_{i=1}^{n} y_i^2 + \beta^2)^{2na+2\alpha}}{\gamma(na+\alpha)} * \frac{\gamma(a(n+1)+\alpha)}{2(\tilde{y}+(\sum_{i=1}^{n} y_i^2 + \beta^2)^2)^{a(n+1)+\alpha}}$$

$$= 2 * \frac{\gamma(a(n+1)+\alpha) * (\tilde{y}^{2a-1} * (\sum_{i=1}^{n} y_i^2) + \beta^2)^{2(na+\alpha)}}{\gamma(a)\gamma(na+\alpha) * (\tilde{y}^2 + (\sum_{i=1}^{n} y_i^2 + \beta^2)^2)^{(n+1)a+\alpha}}$$

is our answer.

Problem 3.14

Unit information prior: Let Y1, . . . , Yn i.i.d. $p(y|\theta)$. Having observed the values Y1 = y1, . . . , Yn = yn, the log likelihood is given by $l(\theta|y) = \sum log p(y_i|\theta)$, and the value $\hat{\theta}$ of θ that maximizes $l(\theta|y)$ is called the maximum likelihood estimator. The negative of the curvature of the log likelihood, $J(\theta) = \frac{\partial^2 l(\theta|y)}{\partial \theta^2}$, describes the precision of the MLE $\hat{\theta}$ and is called the observed Fisher information. For situations in which it is difficult to quantify prior information in terms of a probability distribution, some have suggested that the "prior" distribution be based on the likelihood, for example, by centering the prior distribution around the MLE $\hat{\theta}$. To deal with the fact that the MLE is not really prior information,the curvature of the prior is chosen so that it has only "one nth" as muchinformation as the likelihood, so that $\frac{\partial^2 log p(\theta)}{\partial \theta^2} = \frac{J(\theta)}{n}$. Such a prior is called a unit information prior (Kass and Wasserman, 1995; Kass and Raftery, 1995), as it has as much information as the average amount of information from a single observation. The unit information prior is not really a prior distribution, as it is computed from the observed data. However, it can be roughly viewed as the prior information of someone with weak but accurate prior information.

a) Let Y1, . . . , Yn i.i.d. $binary(\theta)$. Obtain the MLE $\hat{\theta}$ and $J(\hat{\theta})/n$

We know that the y_i is binary, so

$$\prod_{i=1}^{n} p(y_i|\theta) = \theta^{\sum_{i=1}^{n} y_i}(1-\theta)^{n-\sum_{i=1}^{n} y_i}$$

we can input log in, so

$$log(\prod_{i=1}^{n} p(y_i|\theta)) = log(\theta^{\sum_{i=1}^{n} y_i}(1-\theta)^{n-\sum_{i=1}^{n} y_i}) = \sum_{i=1}^{n} y_i log(\theta) + (n - \sum_{i=1}^{n} y_i)log(1-\theta)$$

Taking the derivative and set it equal to zero I get,

$$\frac{\partial^2}{\partial \theta^2}(\sum_{i=1}^{n} y_i log(\theta) + (n - \sum_{i=1}^{n} y_i)log(1-\theta)) = 0$$

$$\frac{\sum_{i=1}^{n} y_i}{\theta} = \frac{n - \sum_{i=1}^{n} y_i}{1-\theta}$$

$$1 = \theta * \frac{n - \sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} y_i} + \theta$$

$$\frac{1}{\theta} = \frac{n - \sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} y_i} + 1 = \frac{n}{\sum_{i=1}^{n} y_i}$$

This indicates that $\hat{\theta} = \frac{\sum_{i=1}^n y_i}{n}$

Now get $J(\hat{\theta})$,

$$-\frac{\partial^2}{\partial\theta^2}log(p(\theta|y)) = -\frac{\partial^2}{\partial\theta^2}log(\theta^y(1-\theta)^{n-y}) = -\frac{\partial^2}{\partial\theta^2}(ylog(\theta) + (n-y)log(1-\theta))$$

$$= -\frac{\partial}{\partial\theta}(\frac{y}{\theta} - \frac{n-y}{1-\theta}) = -(-\frac{y}{\theta^2} - \frac{n-y}{(1-\theta)^2}) = \frac{y}{\theta} + \frac{n-y}{(1-\theta)^2}$$

Then,

$$\frac{J(\hat{\theta})}{n} = \frac{y}{n\theta^2} + \frac{n-y}{n(1-\theta)^2}$$

b) Find a probability density $p_U(\theta)$ such that $logp_U(theta) = l(\theta|y)/n + c$, where c *is a constant that does not depend on θ. Compute the information $-\partial^2 logp_U(\theta)/\partial\theta^2$ of this density.**

To get,

$$log(p_U(\theta)) = \frac{1}{n}(y * log(\theta) + (n-y)log(1-\theta) + C$$

we need to have $p_U(\theta) = \hat{C}\theta^{\frac{y}{n}}(1-\theta)^{\frac{n-y}{n}}$

this is in fact $beta(\theta|\frac{y}{n} + 1, \frac{n-y}{n} + 1)$.

Where $\hat{C} = \frac{\gamma(3)}{\gamma(\frac{y}{n}+1)\gamma(\frac{n-y}{n}+1)}$

taking the log of $\hat{C}$ we can get $C = \gamma(3) - \gamma(\frac{y}{n} + 1)\gamma(\frac{n-y}{n} + 1)$

now getting the $-\partial^2 logp_U(\theta)/\partial\theta^2$,

$$-\frac{\partial^2}{\partial\theta^2}log(\theta^{\frac{y}{n}}(1-\theta)^{\frac{n-y}{n}}) = -\frac{\partial^2}{\partial\theta^2}(\frac{y}{n}log(\theta) + \frac{n-y}{n}log(1-\theta))$$

$$= -\frac{\partial}{\partial\theta}(\frac{y}{n\theta} - \frac{n-y}{n(1-\theta)})$$

$$= \frac{y}{n\theta^2} + \frac{n-y}{n(1-\theta)^2}$$

c) Obtain a probability density for θ that is proportional to $p_U(\theta) \times p(y_1, ..., y_n|\theta)$. Can this be considered a posterior distribution for θ?

Because beta distribution and the binomial are a conjugate pairs, we can say that

$p_U(\theta)p(y_1, ..., y_n|\theta) \propto beta(\theta|\frac{y}{n} + 1 + y, n - y + \frac{n-y}{n} + 1$

Above could be considered as prior, however when we are dealing with pure bayesian this would not be considered as a valid prior. Thus we could no consider posterior distribution for theta.

d) Repeat a), b) and c) but with $p(y|\theta)$ being the Poisson distribution.

a)

This time we are using poisson as y_i

$$\prod_{i=1}^{n} p(y_i|\theta) = \prod_{i=1}^{n} \frac{e^{-\theta}\theta^{y_i}}{y_i!}$$

taking the log

$$log(\prod_{i=1}^{n} p(y_i|\theta)) = log(\prod_{i=1}^{n} \frac{e^{-\theta}\theta^{y_i}}{y_i!}) = log(e^{-n\theta}\theta^{\sum_{i=1}^{n} y_i}\prod_{i=1}^{n} \frac{1}{y_i!}) = -n\theta + log(\theta)\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} log(y_i!)$$

taking the derivative and set it equal to 0 I get,

$$\frac{\partial}{\partial\theta}(-n\theta + log(\theta)\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} log(y_i!)) = 0$$

$$\frac{\sum_{i=1}^{n} y_i}{\theta} = n$$

$$\hat{\theta} = \frac{\sum_{i=1}^{n} y_i}{n}$$

Doing the same process as a) we can get $J(\theta)$

$$-\frac{\partial^2}{\partial\theta^2}log(p(\theta|y_1,...,y_n)) = -\frac{\partial^2}{\partial\theta^2}(-n\theta + log(\theta)\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} log(y_i!)) = \frac{\partial}{\partial\theta}(-n + \frac{\sum_{i=1}^{n} y_i}{\theta}) = \frac{\sum_{i=1}^{n} y_i}{\theta^2}$$

b)

Same process as b) above to get

$$log(p_U(\theta)) = \frac{1}{n}(-n\theta + log(\theta)\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} log(y_i!)) + C$$

we need

$$p_U(\theta) = \hat{C} * (e^{-n\theta}\theta^{\sum_{i=1}^{n} y_i}\prod_{i=1}^{n} \frac{1}{y_i!})^{1/n}$$

$$= \hat{C}e^{-\theta}\theta^{\bar{y}}(\prod_{i=1}^{n} \frac{1}{y_i!})^{(1/n)} = \hat{C}e^{-\theta}\theta^{\bar{y}}$$

which is $gamma(\theta|\alpha = \bar{y} + 1, \beta = 1)$ distribution, where
$\hat{C} = \frac{1}{\gamma(\bar{y}+1)}$ then $C = -\gamma(\bar{y} + 1)$
we can get

$$-\frac{\partial^2}{\partial\theta^2}log(e^{-\theta}\theta^{\bar{y}}) = \frac{\bar{y}}{\theta^2}$$

c)

because Gamma and Poisson distribution are a conjugate pair, we can say that

$p_U(\theta)p(y_1, ..., y_n|\theta) \propto gamma(\theta|\bar{y}n + \bar{y}, n + 1).$

Above could be considered as prior, however when we are dealing with pure bayesian this would not be considered as a valid prior. Thus we could no consider posterior distribution for theta