Homework 3

November 6, 2019

1 Problem 6.8-1

Let Y be the sum of the observations of a random sample from a Poisson distribution with mean θ. Let the prior pdf of θ be gamma with parameters α and β.

1.1 Part a

Find the posterior pdf of θ given that $Y = y$.

By definition of Bayes, which is

$p(\theta|y) = \dfrac{p(y|\theta)p(\theta)}{p(y)}$

Let $\lambda = \theta$

Therefore the prior is,

$g(\theta) = \dfrac{\theta^{\alpha-1}e^{-\theta/\beta}}{\beta^\alpha\Gamma(\alpha)}$

And the likelihood is,

$L(\theta|Y) = (\prod \dfrac{1}{y_i!})\theta^y e^{-n\theta}$

Making the denominator,

$p(y|\theta)p(\theta) = \dfrac{\theta^{\alpha-1}e^{-\theta/\beta}}{\beta^\alpha\Gamma(\alpha)}(\prod \dfrac{1}{y_i!})\theta^y e^{-n\theta}$

$= \dfrac{\prod \dfrac{1}{y_i!}}{\beta^\alpha\Gamma(\alpha)}\theta^{\alpha-1}e^{-\theta/\beta}\theta^y e^{-n\theta}$

$= \dfrac{\prod \dfrac{1}{y_i!}}{\beta^\alpha\Gamma(\alpha)}\theta^{\alpha+y-1}e^{-\theta(\dfrac{1}{\beta}+n)}$

Above I am missing an exponent of n for e

$\therefore k(\theta|y) \propto \theta^{\alpha+y-1}e^{-n\theta(n+1/\beta)}$

1.2 Part b

If the loss functions is $[w(y) - \theta]^2$, find the Bayesian point estimate $w(y)$.

The expected value of Gamma distribution is the mean, which is the product of $\alpha\theta$

$w(y) = E(\theta|y) = \dfrac{(\alpha + y)}{n + 1/\beta}$

1.3 Part c

Show that $w(y)$ found in (b) is a weighted average of the maximum likelihood estimate of y/n and the prior mean $\alpha\beta$ with respective weights of $\frac{n}{n+1/\beta}$

$w(y) = (\frac{y}{n})(\frac{n}{n+1/\beta}) + (\alpha\beta)(\frac{1/\beta}{n+1/\beta})$

Therefore, the averages are weighted by $\frac{y}{n}$ and $\alpha\beta$

2 Problem 6.8-3

In Example 6.8-2, take $n = 30, \alpha = 15, \beta = 5$.

2.1 Part a

Using the squared error loss, compute the expected loss (risk function) associated with the Bayes estimator $w(Y)$.

Example 6.8-2 has $w(y) = \frac{\alpha + y}{\alpha + \beta + n}$
$E(Y) = \mu$
And since $\mu = np$, and $p = \theta$
Therefore, $E(Y) = 30\theta$
As for $Var(Y)$,
$Var(Y) = \sigma^2 = npq$
As well as,
$Var(Y) = E(Y^2) - (E(Y))^2$
Recall, $q = 1 - p$
Therefore, $Var(Y) = 30\theta(1-\theta)$
Expected loss is equal to the expected value of $(\theta - w(y))^2$
$E[(\theta - w(y))^2] = E((\theta - \frac{\alpha + Y}{\alpha + \beta + n})^2)$
Plug in values,
$= E((\theta - \frac{15 + Y}{15 + 5 + 30})^2)$
$= E((\theta - \frac{15 + Y}{50})^2)$
Expand,
$= E(\theta^2 - 2\theta\frac{15 + Y}{50} + (\frac{15 + Y}{50})^2)$
$= \theta^2 - \theta\frac{15 + E(Y)}{25} + E(\frac{225 + 30Y + Y^2}{2500})$
$= \theta^2 - \theta\frac{15 + E(Y)}{25} + \frac{225 + 30E(Y) + E(Y^2)}{2500}$
Let $E(Y^2) = Var(Y) + (E(Y))^2$
$\frac{500\theta^2}{500} - \frac{300\theta}{500} - \frac{600\theta^2}{500} + \frac{186\theta}{500} + \frac{174\theta^2}{500} + \frac{45}{500}$
$= \frac{74\theta^2 - 114\theta + 45}{500}$

2.2 Part b

The risk function associated with the usual estimator Y/n is $\theta(1-\theta)/30$. Find those values of θ for which the risk function in part (a) is less than $\theta(1-\theta)/30$. In particular, if the prior mean $\alpha/(\alpha+\beta) = 3/4$ is a reasonable guess, then the risk function in part (a) is better of the two (i.e., is smaller in a neighborhood of $\theta = 3/4$) for what values of θ?

COMPLETE THIS PROBLEM LATER

3 Problem 7.1-1

A random sample of size 16 from the normal distribution $N(\mu, 25)$ yielded $\bar{x} = 73.8$. Find a 95% confidence interval for μ.

$z_{\alpha/2} = z_0.025 = 1.96$ according the table in the appendix.

Definition of CI,

$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

Substitute values,

$73.8 \pm 1.96\frac{\sqrt{25}}{\sqrt{16}}$

Therefore the CI for μ = (71.35,76.25)

4 Problem 7.1-4

Let X equal the weight in grams of a "52-gram" snack pack of candies. Assume that the distribution of X is $N(\mu, 4)$. A random sample of $n = 10$ observations of X yielded the following data:

[1]:
```python
import pandas as pd
import numpy as np
from scipy import stats

Problem7_1_4data = pd.read_csv('E7_1-04.txt', sep=" ", header=None)
Problem7_1_4data.rename(columns={0:"Snack_pack_weight"},inplace=True)
print(Problem7_1_4data)
```

```
   Snack_pack_weight
0              55.95
1              56.54
2              57.58
3              55.13
4              57.48
5              56.06
6              59.93
7              58.30
8              52.57
9              58.46
```

4.1 Part a

Give a point estimate for μ

[2]:
```
x_bar = Problem7_1_4data.mean() #mean
print("The sample mean is",x_bar.values[0])
```

```
The sample mean is 56.8
```

4.2 Part b

For the endpoints for a 95% confidence interval for μ

[3]:
```
data_std = np.sqrt(4) #std
data_n = len(Problem7_1_4data) #number of elements
z = 1.96 # Conf interval of 95%

lower_data_ci = (x_bar.values[0] - (z * (data_std/np.sqrt(data_n)))).round(4)
upper_data_ci = (x_bar.values[0] + (z * (data_std/np.sqrt(data_n)))).round(4)
print("Lower confidence interval is",lower_data_ci)
print("Upper confidence interval is",upper_data_ci)
```

```
Lower confidence interval is 55.5604
Upper confidence interval is 58.0396
```

4.3 Part c

On the basis of these very limited data, what is the probability that an individual snack pack selected at random is filled with less than 52 grams of candy?

If we assume that $\bar{x} = \mu$ then we can calculate z score using the following formula,

$z = \dfrac{x - \mu}{\sigma}$

Then find the probability using,

$P(X < 52) \approx P(Z < z)$

[4]:
```
z_score = (52-56.8)/np.sqrt(4)
data_prob = (stats.norm.cdf(z_score)*100).round(2)
print("The probability of less than 52 grams of candy is ",data_prob,"%")
```

```
The probability of less than 52 grams of candy is  0.82 %
```

5 Problem 7.1-5

As a clue to the amount of organic waste in Lake Macatawa, a count was made of the number of bacteria colonies in 100 milliliters of water. The number of colonies, in hundreds, for $n = 30$ samples of water from the east basin yielded the following,

[6]:
```
Problem7_1_5data = pd.read_csv('E7_1-05.txt', sep=" ", header=None)
Problem7_1_5data.rename(columns={0:
↪"bacteria_colonies_in_100mL_water"},inplace=True)
print(Problem7_1_5data)
```

```
    bacteria_colonies_in_100mL_water
0                                 93
1                                140
2                                  8
3                                120
4                                  3
5                                120
6                                 33
7                                 70
8                                 91
9                                 61
10                                 7
11                               100
12                                19
13                                98
14                               110
15                                23
16                                14
17                                94
18                                57
19                                 9
20                                66
21                                53
22                                28
23                                76
24                                58
25                                 9
26                                73
27                                49
28                                37
29                                92
```

Using Student's t distribution, with degree of freedom being 29 and $1 - (1 - .90)/2 = 0.95$. Table has $t_{\alpha/2} = 1$.

[20]:
```
x_bar = Problem7_1_5data.mean().round(4).values[0]
n = len(Problem7_1_5data)
std = Problem7_1_5data.std().round(4).values[0]
t = 1.699

lower_ci = (x_bar-t*std/np.sqrt(n)).round(4)
upper_ci = (x_bar+t*std/np.sqrt(n)).round(4)

print("The lower confidence interval is",lower_ci)
print("The upper confidence interval is",upper_ci)
```

```
The lower confidence interval is 48.0762
The upper confidence interval is 72.6572
```