MATH 8820 (Fall 2018)
Homework 1

Shirong Zhao

1. (a) **Solution.** From the Lecture notes, we know that the conjugate prior is Gamma distribution. Suppose $\lambda \sim Gamma(a, b)$, then we have

$$
\begin{aligned}
p(\lambda|\boldsymbol{y}) &\propto p(\boldsymbol{y}|\lambda)p(\lambda) \\
&\propto (\prod_{i=1}^{b} \frac{e^{-\lambda}\lambda^{y_i}}{y_i!})\lambda^{a-1}e^{-\lambda b} \\
&\propto (e^{-n\lambda}\lambda^{\sum_{i=1}^{n} y_i})\lambda^{a-1}e^{-\lambda b} \\
&\propto e^{-(n+b)\lambda}\lambda^{\sum_{i=1}^{n} y_i+a-1} \\
&\propto e^{-(n+b)\lambda}\lambda^{n\bar{y}+a-1}
\end{aligned}
$$

where $\bar{y} = \sum_{i=1}^{n} /n$. Therefore the posterior for λ is $Gamma(a + n\bar{y}, b + n)$. Thus the prior and posterior are both Gamma distribution.

$\square$

(b) **Solution.** Check (a) $\square$

(c) **Solution.** For $Gamma(a + n\bar{y}, b + n)$ distribution, the posterior mean is $E(\lambda|\boldsymbol{y}) = \frac{a+n\bar{y}}{n+b}$. Variance is $Var(\lambda|\boldsymbol{y}) = \frac{a+n\bar{y}}{(n+b)^2}$. $\square$

(d) **Solution.** The log likelihood function is:

$$
\begin{aligned}
LLF(\lambda|\boldsymbol{y}) &= log(\prod_{i=1}^{b} \frac{e^{-\lambda}\lambda^{y_i}}{y_i!}) \\
&= -n\lambda + (\sum_{i=1}^{n} y_i)log\lambda - \sum_{i=1}^{n} log(y_i!)
\end{aligned}
$$

Taking derivative with respect to λ, we get $0 = -n + \lambda^{-1}\sum_{i=1}^{n} y_i$, hence $\hat{\lambda}_{MLE} = \bar{y}$. Notice that $\frac{a+n\bar{y}}{n+b} = \frac{n}{n+b}\bar{y} + \frac{b}{n+b}\frac{a}{b} = \frac{n}{n+b}\hat{\lambda}_{MLE} + \frac{b}{n+b}\frac{a}{b}$. where $\frac{a}{b}$ is the mean of prior distribution ($Gamma(a, b)$) of λ. That is to say, the posterior mean is a weighted mean of the MLE and prior mean

□

(e) **Solution.** See the appendix codes. □

(f) **Solution.** The data I am going to use is from http://users.stat.ufl.edu/ winner/data/brit_rail_acc.dat

and explanations about the variables are from http://users.stat.ufl.edu/ winner/data/brit_rail_acc.txt. The count data is British Rail Accidents from 1946 to 2003. The response data is count data, and is interger, i.e., the y variable could only take value $0, 1, 2, ...,$. Moreover, the number of rail accidents from year to year should be independent. The number of this years' accidents could not be used to predict the next year's number. Hence this data is good to use here. □

(g) **Solution.** In our data, we have that $n = 58$, and $\bar{y} = 3.362$, if in prior, we take $a = b = 1$, then the posterior mean is $\frac{1+58*3.362}{1+58} = 3.322034$. Also the equal-tailed is $[2.873213, 3.802950]$, the HPD interval is $[2.8624576, 3.7911364]$.

Hence, we conclude that the λ has 95% probability that will land on $[2.8624576, 3.7911364]$. The number of rail accidents is a Poisson distribution with parameter equal to 3.322 □