Midterm take-home exam

Solution (updated and with comments after I graded the exams)

Course: Bayesian Learning

Program: Professional Master in Economics

Instructor: Hedibert Freitas Lopes

Poisson data with Gamma prior for its rate

Poisson model. Let us assume that $y_1, \ldots, y_n$ are a random sample of Poisson counts with rate $\lambda > 0$, i.e. $y_i \sim Poi(\lambda)$ for $i = 1, \ldots, n$. Recall that the Poisson distribution is discrete and take values in $\{0, 1, 2, \ldots\}$ and has probability mass given by

$$Pr(y = k|\lambda) = \frac{\lambda^k e^{-\lambda}}{k!} \qquad k = 0, 1, 2, \ldots.$$

The mean and variance of the Poisson distribution are the same, $E(y|\lambda) = V(y|\lambda) = \lambda$.

Likelihood and MLE. It is easy to show that the likelihood of λ based on observations $y_1, \ldots, y_n$ is

$$L(\lambda|y_1, \ldots, y_n) \equiv p(y_1, \ldots, y_n|\lambda) = \prod_{i=1}^{n} \frac{\lambda^{y_i} e^{-\lambda}}{y_i!} = \frac{\lambda^{n\bar{y}_n} e^{-n\lambda}}{\prod_{i=1}^n y_i!} \propto \lambda^{n\bar{y}_n} e^{-n\lambda}.$$

Notice that the term $\prod_{i=1}^n y_i!$ is totally irrelevant in the likelihood since IT IS NOT a function of λ. In other words, whatever one might want to say about λ can be said without mentioning $\prod_{i=1}^n y_i!$.

The log-likelihood of λ is

$$\mathcal{L}(\lambda) \equiv \log L(\lambda|y_1, \ldots, y_n) = \kappa + n\bar{y}_n \log \lambda - n\lambda,$$

where $\bar{y}_n = (y_1 + \cdots + y_n)/n$ and $\kappa = \sum_{i=1}^n \log y_i!$. It is also easy to check that the maximum likelihood estimator of λ, i.e. $\hat{\lambda}_{mle} = \arg\max_{\lambda>0} \mathcal{L}(\lambda)$, is given by $\bar{y}_n$.

Hint: The likelihood "looks like" a $Gamma(n\hat{y}_n + 1, n)$ (See more details about the Gamma distribution below when we talk about the prior)

Application. As a concrete context, we will consider the `CreditCard` dataset form the `R` package `AER`, which is a cross-section data on the credit history for a sample of applicants for a type of credit card. We will focus on the count variables `reports` that has the number of major derogatory reports. Here $n = 1319$. Check it out by running the following script. However, if you have any difficulty accessing the data, I have added the `reports` variable at the end of this document.

```
install.packages("AER")
library("AER")
data(CreditCard)
reports = CreditCard[,2]
hist(reports)
mean(reports)
```

For the reports data, $\bar{y}_n = 0.4564064$, so $\hat{\lambda}_{mle} = 0.4564064$. The following piece of code plots the likelihood function:

```
ybar = mean(reports)
n = length(reports)
lambdas = seq(0.35,0.55,length=1000)
loglike =  n*ybar*log(lambdas)-n*lambdas
like= exp(loglike-max(loglike))
plot(lambdas,like,xlab="lambda",ylab="Likelihood",type="l")
abline(v=ybar)
```

Prior. Let us assume that the prior on λ is $Gamma(\alpha_0, \beta_0)$ distribution, i.e.

$$p(\lambda) = \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \lambda^{\alpha_0 - 1} e^{-\beta_0 \lambda}, \qquad \lambda > 0,$$

for $\alpha_0, \beta_0 > 0$. The $Gamma(\alpha_0, \beta_0)$ distribution has mean, mode and variance equal to, respectively, α_0/β_0, $(\alpha_0 - 1)/\beta_0$ if $\alpha_0 \geq 1$, and α_0/β_0^2.

Assuming we are fairly agnostic about λ, *a priori*, we will take the values of $\alpha_0 = 1.5$ and $\beta_0 = 1$ as the prior hyperparameters. Prior mean and mode of λ are 1.5 and 0.5, respectively, while the prior standard deviation is $\sqrt{1.5} = 1.224745$. Also, the prior probability that λ falls into $(0, 6)$ is 0.9926168; just use the `R` function `pgamma(6,1.5,1)`. Compare the range of variation of the likelihood (above) with that of the prior density (below). Notice that the prior is way less informative about λ than the likelihood (where the data information is lended and channelled).

```
par(mfrow=c(1,2))
lambdas = seq(0,1,length=1000)
plot(lambdas,dgamma(lambdas,n*ybar+1,n),type="l",xlab="lambda",ylab="Likelihood")
abline(v=ybar,col=2)
lambdas = seq(0,10,length=1000)
plot(lambdas,dgamma(lambdas,1.5,1),type="l",xlab="lambda",ylab="Prior")
abline(v=ybar,col=2)
```

Questions. Your job is to answer the following questions:

a) Show that the posterior of λ, i.e. $p(\lambda|y_1, \ldots, y_n) \propto L(\lambda|y_1, \ldots, y_n)p(\lambda|\alpha_0, \beta_0)$, follows a Gamma distribution with parameters α_1 and β_1, where

$$\alpha_1 = \alpha_0 + n\bar{y}_n \quad \text{and} \quad \beta_1 = \beta_0 + n.$$

The solution is pretty straightforward once you collect the two main components I have provided above, i.e. prior density and likelihood function. Again, we only need the prior and the likelihood up to normalizing constants:

$$\begin{aligned} p(\lambda||y_1, \ldots, y_n) &\propto p(\lambda)L(\lambda|y_1, \ldots, y_n) \propto \left(\lambda^{\alpha_0 - 1} e^{-\beta_0 \lambda}\right)\left(\lambda^{n\bar{y}_n} e^{-n\lambda}\right) \\ &\propto \lambda^{(\alpha_0 + n\bar{y}_n - 1)} \exp\{-(\beta_0 + n)\lambda\}. \end{aligned}$$

Since the kernel of the density of a $Gamma(a, b)$ is $p(x) \propto a^{x-1}e^{-bx}$, it follows that $\lambda|y_1, \ldots, y_n \sim Gamma(\alpha_0 + n\bar{y}_n, \beta_0 + n)$. With $\alpha_0 = 1.5$, $\beta_0 = 1$, $n = 1319$ and $n\bar{y}_n = 602$, it follows that $\alpha_1 = 1.5 + 602 = 603.5$ and $\beta_1 = 1320$.

b) Compare prior and posterior means, modes and standard deviations:

b1) $E(\lambda)$ and $E(\lambda|y_1,\ldots,y_n)$,

$$E(\lambda)=\frac{\alpha_0}{\beta_0}=\frac{1.5}{1}=1.5 \quad \text{and} \quad E(\lambda|y_1,\ldots,y_n)=\frac{\alpha_1}{\beta_1}=\frac{603.5}{1320}=0.457197.$$

b2) $Mode(\lambda)$ and $Mode(\lambda|y_1,\ldots,y_n)$, and

$$Mode(\lambda)=\frac{\alpha_0-1}{\beta_0}=\frac{0.5}{1}=0.5 \quad \text{and} \quad Mode(\lambda|y_1,\ldots,y_n)=\frac{\alpha_1-1}{\beta_1}=\frac{602.5}{1320}=0.456.$$

Notice that the prior is more skewed (mean lower than mode) than the posterior (mean and mode about the same).

b3) $\sqrt{var(\lambda)}$ and $\sqrt{var(\lambda|y_1,\ldots,y_n)}$.

$$var(\lambda)=\frac{\alpha_0}{\beta_0^2}=\frac{1.5}{1^2}=1.5 \quad \text{and} \quad var(\lambda|y_1,\ldots,y_n)=\frac{\alpha_1}{\beta_1^2}=\frac{603.5}{1320^2}=0.0003463613,$$

so $\sqrt{var(\lambda)}=1.2247$ and $\sqrt{var(\lambda|y_1,\ldots,y_n)}=0.0186$.

Needless to say that all these derivations can be performed in closed form.

c) Let us now pretend that we only know how to evaluate pointwise both prior density $p(\lambda|\alpha_0,\beta_0)$ and likelihood function $L(\lambda|y_1,\ldots,y_n)$, already given above. Besides, we also know how to sample from the prior $p(\lambda|\alpha_0,\beta_0)$. Use these abilities and a SIR algorithm to produce $N=10,000$ draws from the posterior $p(\lambda|y_1,\ldots,y_n)$. Use these $N=10,000$ draws to approximate the quantities obtained in exact form in b1), b2) and b3). Comment your findings abundantly.

Solution. Since the kernel of the likelihood, $\lambda^{n\bar{y}_n}e^{-n\lambda}$, *looks just like* a $Gamma(n\hat{y}_n+1,n)$, we will use this Gamma as proposal for your SIR algorithm. More precisely

$$q(\lambda)\equiv Gamma(n\hat{y}_n+1,n).$$

Many of you have used the prior as proposal. Nothing wrong with that, but remember that the likelihood is pretty concentrated while the prior is fairly uninformative (flat!). In other words, your SIR might need way more draws than mine since its proposal is not as good as mine in mimicking the posterior distribution.

Here is our implementation of the SIR algorithm. Notice that the weights are now proportional to the prior. This is because our proposal is, in fact, the likelihood (normalized to become a Gamma distribution). In fact, our SIR-based approximation to $E(\lambda|y_1,\ldots,y_n)=0.457197$ is 0.457344, while the approximation to $\sqrt{var(\lambda|y_1,\ldots,y_n)}=0.01861079$ is 0.01848491. As you can see, the approximations are pretty good up to three decimal digits.

```
set.seed(12345)
M = 100000
N = 10000
n = 1319
ybar = 602/n
alpha0 = 1.5
beta0 = 1
```

```
lambda.draw = rgamma(M,n*ybar+1,n)
w = dgamma(lambda.draw,alpha0,beta0)
lambda.post = sample(lambda.draw,replace=TRUE,size=N,prob=w)
hist(lambda.post)
c(mean(lambda.post),sqrt(var(lambda.post)))
```

Some of you have messed up the proper derivation of the weights, which is crucial for the success of the SIR algorithm. As a matter of fact, the ratio `target/proposal` appears in ALL MCMC algorithms as well, so carefully deriving it is extremely important.

d) The posterior predictive for a new count y_{n+1} is obtained as follows:

$$\begin{aligned}
Pr(y_{n+1}=k|y_1,\ldots,y_n,\alpha_0,\beta_0) &= \int_0^\infty Pr(y_{n+1}=k|\lambda)p(\lambda|y_1,\ldots,y_n)d\lambda \\
&= \int_0^\infty \frac{\lambda^k e^{-\lambda}}{k!}\frac{(\beta_0+n)^{\alpha_0+n\bar{y}_n}}{\Gamma(\alpha_0+n\bar{y}_n)}\lambda^{\alpha_0+n\bar{y}_n-1}e^{-(\beta_0+n)\lambda}d\lambda \\
&= \frac{1}{k!}\frac{(\beta_0+n)^{\alpha_0+n\bar{y}_n}}{\Gamma(\alpha_0+n\bar{y}_n)}\int_0^\infty \lambda^{\alpha_0+n\bar{y}_n+k-1}e^{-(\beta_0+n+1)\lambda}d\lambda \\
&= \frac{1}{k!}\frac{(\beta_0+n)^{\alpha_0+n\bar{y}_n}}{\Gamma(\alpha_0+n\bar{y}_n)}\underbrace{\int_0^\infty \lambda^{\alpha_0+n\bar{y}_n+k-1}e^{-(\beta_0+n+1)\lambda}d\lambda}_{\text{Kernel of } Gamma(\alpha_0+n\bar{y}_n+k,\beta_0+n+1)} \\
&= \frac{1}{k!}\times\frac{(\beta_0+n)^{\alpha_0+n\bar{y}_n}}{\Gamma(\alpha_0+n\bar{y}_n)}\times\frac{\Gamma(\alpha_0+n\bar{y}_n+k)}{(\beta_0+n+1)^{\alpha_0+n\bar{y}_n+k}}
\end{aligned}$$

which is a function of $(n,\bar{y}_n,\alpha_0,\beta_0)$, i.e. $(n,\bar{y}_n)$ is sufficient statistic for λ. Show that

$$Pr(y_{n+1}=k|n,\bar{y}_n,\alpha_0,\beta_0)=\frac{1}{k!}\times\frac{(\beta_0+n)^{\alpha_0+n\bar{y}_n}}{(\beta_0+n+1)^{\alpha_0+n\bar{y}_n+k}}\times\frac{\Gamma(\alpha_0+n\bar{y}_n+k)}{\Gamma(\alpha_0+n\bar{y}_n)}$$

for $k=0,1,\ldots$. Recall that $n=1319$, $n\bar{y}_n=602$, $\alpha_0=1.5$ and $\beta_0=1$, so

$$Pr(y_{1320}|n,\bar{y}_n,\alpha_0,\beta_0)=\frac{1}{k!}\times\frac{(1320)^{603.5}}{(1321)^{603.5+k}}\times\frac{\Gamma(603.5+k)}{\Gamma(603.5)},\qquad k=0,1,2,\ldots.$$

Based on the following piece of `R` code, we can see that the posterior predictive is almost all concentrated in values below 5.

```
k = 0:5
term1 = 1/factorial(k)
term2 = (1320/1321)^(603.5)/(1321)^k
term3 = exp(lgamma(603.5+k)-lgamma(603.5))
postpred = term1*term2*term3
plot(k-0.1,postpred,type="h",lwd=2,xlab=expression(y[n+1]),
        ylab="Probability",xlim=c(-0.5,5.5))
lines(k+0.1,dpois(k,602/1319),type="h",col=2,lwd=2)
legend("topright",legend=c("MLE","Bayes"),col=2:1,bty="n",lwd=2,lty=1)
```

The `reports` variable from the `CreditCard` data

```
reports = c(
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,7,0,3,0,1,0,1,0,0,0,0,0,0,
0,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,
2,0,0,0,3,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,1,0,0,3,0,0,0,0,1,
2,0,0,4,2,0,1,1,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,3,1,1,0,0,0,
4,0,0,0,1,0,0,0,0,0,5,0,0,1,0,0,0,0,0,0,0,0,0,0,2,0,0,0,0,1,3,
2,0,1,5,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,5,0,0,0,0,0,1,1,0,
1,3,0,0,3,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,6,1,0,
0,1,0,0,1,0,0,0,0,0,0,0,0,7,2,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,
1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,2,0,0,11,1,0,0,0,0,0,0,
0,1,0,0,1,4,0,0,0,0,0,0,0,4,0,0,1,0,0,0,0,2,0,0,0,1,0,2,0,0,
2,0,0,0,0,0,4,0,0,0,0,0,0,1,0,0,5,0,0,0,0,0,2,0,0,0,0,2,0,2,3,0,
1,0,0,5,0,0,1,0,0,0,0,0,0,0,1,2,0,0,2,0,0,0,2,3,0,0,0,1,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,4,3,0,0,0,0,2,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,4,0,0,1,0,0,3,0,0,1,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,2,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,1,1,1,0,0,7,0,0,0,0,1,0,0,1,0,0,2,0,0,0,2,0,
0,0,0,0,0,0,0,0,1,0,0,0,0,1,0,0,0,0,2,0,0,0,5,0,0,0,0,2,0,0,0,
0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,1,0,1,0,0,0,0,0,0,0,0,0,2,0,0,
1,2,4,0,0,0,0,0,0,0,1,0,3,2,0,0,0,0,0,0,0,0,0,0,0,0,3,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,4,0,0,0,0,0,0,1,0,0,1,0,1,0,0,0,0,0,0,3,0,0,
2,7,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,1,2,0,0,0,0,0,0,0,0,0,0,1,0,
0,0,4,0,0,0,4,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,7,0,0,0,0,0,0,6,
0,1,2,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,1,0,1,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,1,0,0,0,1,0,1,1,0,0,0,3,2,0,0,0,0,0,0,0,0,0,0,
0,0,1,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,1,1,3,2,0,7,0,
0,0,1,0,0,0,0,0,1,0,0,0,1,0,0,0,0,0,9,0,0,0,0,1,0,0,0,0,0,0,0,
0,0,1,0,1,0,0,0,5,0,0,0,1,0,0,1,1,0,0,0,0,0,0,0,0,1,0,0,1,2,0,
0,1,0,0,0,0,3,0,0,0,0,0,0,0,0,1,0,1,0,0,0,0,0,0,0,0,2,0,0,12,0,0,
0,2,1,0,1,0,0,0,0,0,1,0,0,0,0,0,1,0,0,0,3,0,0,1,0,0,0,0,0,0,0,
0,2,0,0,0,1,0,2,0,0,4,0,0,0,0,2,0,0,0,2,0,0,11,0,0,0,0,0,0,11,
0,0,0,0,1,0,0,0,0,2,0,0,0,1,0,0,0,0,14,1,0,1,0,0,0,0,1,0,0,0,
3,0,0,0,0,0,0,2,0,0,0,0,0,0,0,0,0,0,5,1,1,1,0,0,0,0,0,0,5,0,
0,0,1,0,0,0,0,0,0,0,0,1,1,0,0,0,0,0,3,0,0,0,0,0,2,0,0,2,0,0,
0,0,0,0,0,0,0,0,5,0,0,0,0,0,10,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,
0,0,0,1,1,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,1,0,3,3,0,0,0,0,0,
1,0,0,0,4,0,0,2,2,0,0,0,1,0,0,0,0,0,0,0,4,1,0,2,0,3,0,0,0,1,
0,1,1,0,0,0,0,0,4,0,0,0,1,0,0,0,0,1,0,0,0,0,0,0,0,3,0,0,0,0,
0,0,1,2,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,2,0,1,0,0,1,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,4,2,0,0,0,0,0,0,9,0,6,0,0,0,0,0,0,1,
0,0,0,2,1,0,0,0,0,0,1,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,1,0,0,0,
0,0,0,0,0,4,0,1,0,0,0,0,0,0,0,0,2,1,0,0,0,0,1,0,0,0,0,0,0,0,
0,0,1,6,0,0,0,6,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,2,1,11,0,0,
0,0,0,0,0,0,1,0,0,0,0,0,0,0,1,0,0,0,0,1,0,0,0,1,0,5,0,0,0)
```