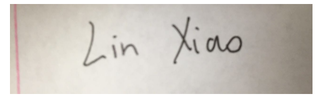


XIAO_Lin_Solutions_HW6

Lin Xiao

November 1, 2015

Note: If not specified as rate, b would be scale parameter of beta.

1. Assume that

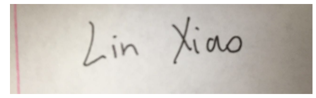
$$X|\lambda \sim Poisson(\lambda)$$
$$\lambda \sim Gamma(a, b)$$

Assume that $\widetilde{X}|\lambda \sim Poisson(\lambda)$ is independent of X. Assume we have a new observation $\widetilde{x}$. Find the posterior predictive distribution, $p(\widetilde{x}|x)$. Assume that a is an integer.
Recall that in class we showed that

$$\begin{aligned} p(\lambda|x) &\propto p(x|\lambda)p(\lambda) \\ &\propto e^{-\lambda}\lambda^x\lambda^{a-1}e^{-\lambda/b} \\ &= \lambda^{x+a+1}e^{-\lambda(1+1/b)} \end{aligned}$$

Thus, $\lambda|x \sim Gamma(x+a, \frac{1}{1+1/b})$, i.e., $\lambda|x \sim Gamma(x+a, \frac{b}{b+1})$. Find and completely derive (showing all steps) $p(\widetilde{x}|x)$.
Solutions:
Because $\widetilde{X}|\lambda \sim Poisson(\lambda)$ is independent of X

$$\begin{aligned} p(\widetilde{x}|x) &= \frac{p(\widetilde{x}, x)}{p(x)} \\ &= \frac{\int p(\widetilde{x}, x, \lambda)d\lambda}{p(x)} \\ &= \int p(\widetilde{x}|x, \lambda)\frac{p(x, \lambda)}{p(x)}d\lambda \\ &= \int p(\widetilde{x}|\lambda)p(\lambda|x)d\lambda \\ &= \int Pois_{\widetilde{x}}(\lambda)Gamma_\lambda(x+a, scale = \frac{b}{b+1})d\lambda \\ &= \frac{1}{\widetilde{x}!}\frac{1}{\Gamma(x+a)\left(\frac{b}{b+1}\right)^{x+a}}\int \lambda^{\widetilde{x}}e^{-\lambda}\lambda^{x+a-1}e^{-\frac{b+1}{b}\lambda}d\lambda \\ &= \frac{1}{\widetilde{x}!}\frac{1}{\Gamma(x+a)\left(\frac{b}{b+1}\right)^{x+a}}\frac{\Gamma(\widetilde{x}+x+a)}{\left(\frac{2b+1}{b}\right)^{\widetilde{x}+x+a}}\int Gamma(\widetilde{x}+x+a, rate = \frac{2b+1}{b})d\lambda \\ &= \frac{b^{\widetilde{x}}}{\widetilde{x}!}\frac{(b+1)^{x+a}}{\Gamma(x+a)}\frac{\Gamma(\widetilde{x}+x+a)}{(2b+1)^{\widetilde{x}+x+a}} \\ &= \frac{(\widetilde{x}+x+a-1)!}{(\widetilde{x})!(x+a-1)!}\left(\frac{b}{2b+1}\right)^{\widetilde{x}}\left(\frac{b+1}{2b+1}\right)^{x+a} \\ &\sim NB(x+a, \frac{b}{2b+1}) \end{aligned}$$



Note: here x+a is the number of failures until the experiment is stopped, p is success probability in each experiment (real)

2. Suppose that X is the number of pregnant women arriving at a particu- lar hospital to deliver their babies during a given month, where λ is the probability of a baby arriving. The model is given by

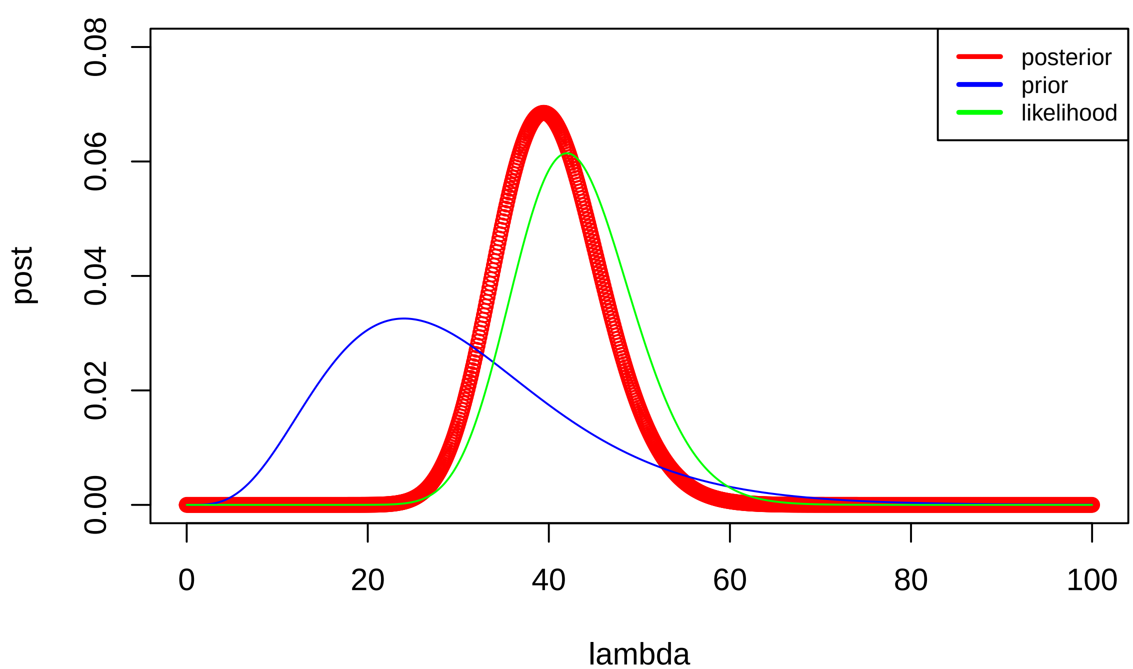
$$X|\lambda \sim Poisson(\lambda)$$

$$\lambda \sim Gamma(a, b)$$

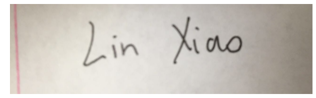
We are also told 42 moms are observed arriving at the particular hospital during December 2007. Using prior study information given, we are told $a = 5$ and $b = 6$. (We found a, b by working backwards from a prior mean of 30 and prior variance of 180). In this problem: do the following:

(a) Plot the likelihood, prior, and posterior distributions as functions of λ in R. (The data has been upload to Sakai).

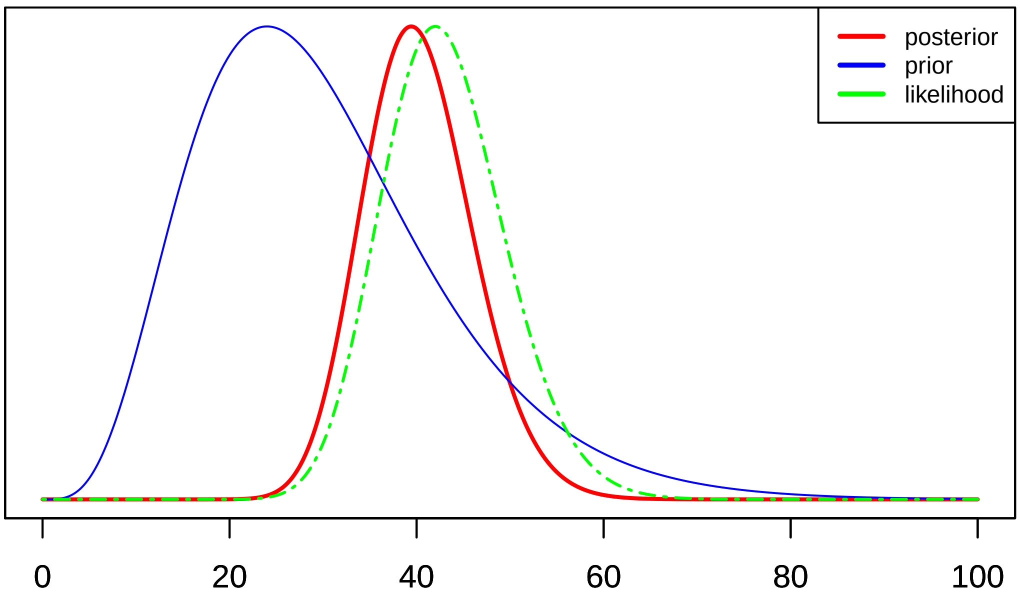
```
lambda <- seq(0,100,length.out = 1000)
lhd <- lambda^42*exp(-lambda)/factorial(42)
prior <- dgamma(lambda, shape = 5, scale = 6)
post <- dgamma(lambda, shape = 5+42, scale = 6/7)

# Regular plot

plot(lambda, post, ylim = c(0,0.08), col = "red")
lines(lambda, prior, col = "blue")
lines(lambda, lhd, type = "l", col = "green")
legend("topright", c("posterior","prior","likelihood"),lty=c(1,1,1),
       lwd=c(2.5,2.5,2.5), col=c("red", "blue", "green"),cex=0.75)
```




```
# Kernel Plot

par(yaxt="n",  ann=FALSE)
plot(lambda, post,
```



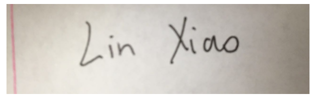

```r
     type="l", col="red",lty=1, lwd=2)
par(new = T)
plot(lambda, prior,type="l",col="blue")
par(new = T)
plot(lambda, lhd, col="green",type="l",lty=6, lwd=1.5)
legend("topright", c("posterior","prior","likelihood"),lty=c(1,1,1),
       lwd=c(2.5,2.5,2.5), col=c("red", "blue", "green"),cex=0.75)
```

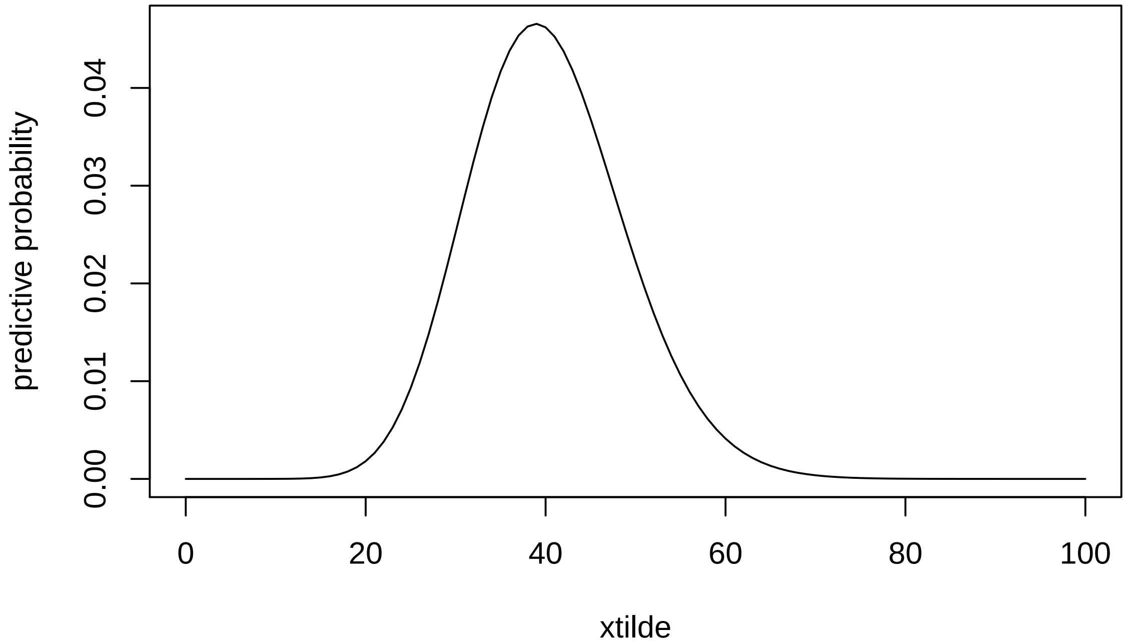


We can see from both plot that the prior is not weak, it affects the shape of the posterior, and the posterior seems from the plots to be a combination of prior and likelihood.

(b) Plot the posterior predictive distribution where the number of pregnant women arriving falls between [0,100], integer valued.

```r
# The number of pregnant women arriving is set as xtilde
xtilde <- seq(0, 100, by = 1)
pred <- dnbinom(xtilde, size = 47, prob = 1-6/13)
par(mfrow=c(1,1))
plot(xtilde, pred, type = "l", ylab="predictive probability")
```





For the dnbinom function, size parameter is number of successful trials, while prob is probability of success in each trial. It is inconsistent with what we have defined for negative binomial, so here just assign 1-p to prob parameter.

(c) Find the posterior predictive probability that the number of pregnant women arrive is between 40 and 45 (inclusive). Do this for homework.

```r
pnbinom(45, size = 47, prob = 1-6/13)-pnbinom(39, size = 47, prob = 1-6/13)
```

```
## [1] 0.2533062
```

So the posterior predictive probability is 0.2533062

3. This problem comes from the Bayesian lecture on the lasso. Specifically, the lasso estimate can be viewed as the mode of the posterior distribution of β

$$\widehat{\beta_L} = \underset{\beta}{argmax} p(\beta|y, \sigma^2, \tau)$$

when

$$p(\beta|\tau) = (\tau/2)^p exp(-\tau||\beta||_1)$$

and the likelihood on

$$p(y|\beta, \sigma^2) = N(y|X\beta, \sigma^2 I_n)$$

For any fixed values $\sigma^2 > 0, \tau > 0$, the posterior mode of β is the lasso estimate with penalty $\lambda = 2\tau\sigma^2$. Prove this result empirically (so prove this using plots in R). If you choose to wish to show a formal mathematical proof this is fine too, but it's much harder.
Proof:
The prior for β is

$$p(\beta|\tau) = (\tau/2)^p exp(-\tau||\beta||_1)$$

We plug $\lambda = 2\tau\sigma^2$ into the function above and get:

$$p(\beta|\sigma^2) = \left(\frac{\tau}{4\sigma^2}\right)^p exp\left(-\frac{\lambda}{2\sigma^2}||\beta||_1\right)$$

Lin Xiao

And we have the likelihhod on y:

$$p(y|\beta,\sigma^2) = N(y|X\beta,\sigma^2 I_n)$$

$$p(y|\beta,\sigma^2) = \frac{1}{\sqrt{(2\pi)^n|\sigma^2 I_n|}} exp[-(Y - X\beta)^T(\sigma^2 I_n)^{-1}(Y - X\beta)]$$

So the posterior distribution is:

$$\begin{aligned} p(\beta|y,\sigma^2,\tau) &\propto p(\beta|\sigma^2)p(y|\beta,\sigma^2) \\ &\propto exp[-(\lambda||\beta||_1 + ||Y - X\beta||_2)] \end{aligned}$$

Since for lasso estimate

$$\widehat{\beta_L} = argmin_{\beta} \lambda||\beta||_1 + ||Y - X\beta||_2$$

Then we complete our proof by having

$$\widehat{\beta_L} = argmax_{\beta}[p(\beta|y,\sigma^2,\tau)]$$