

Solutions to Homework 4

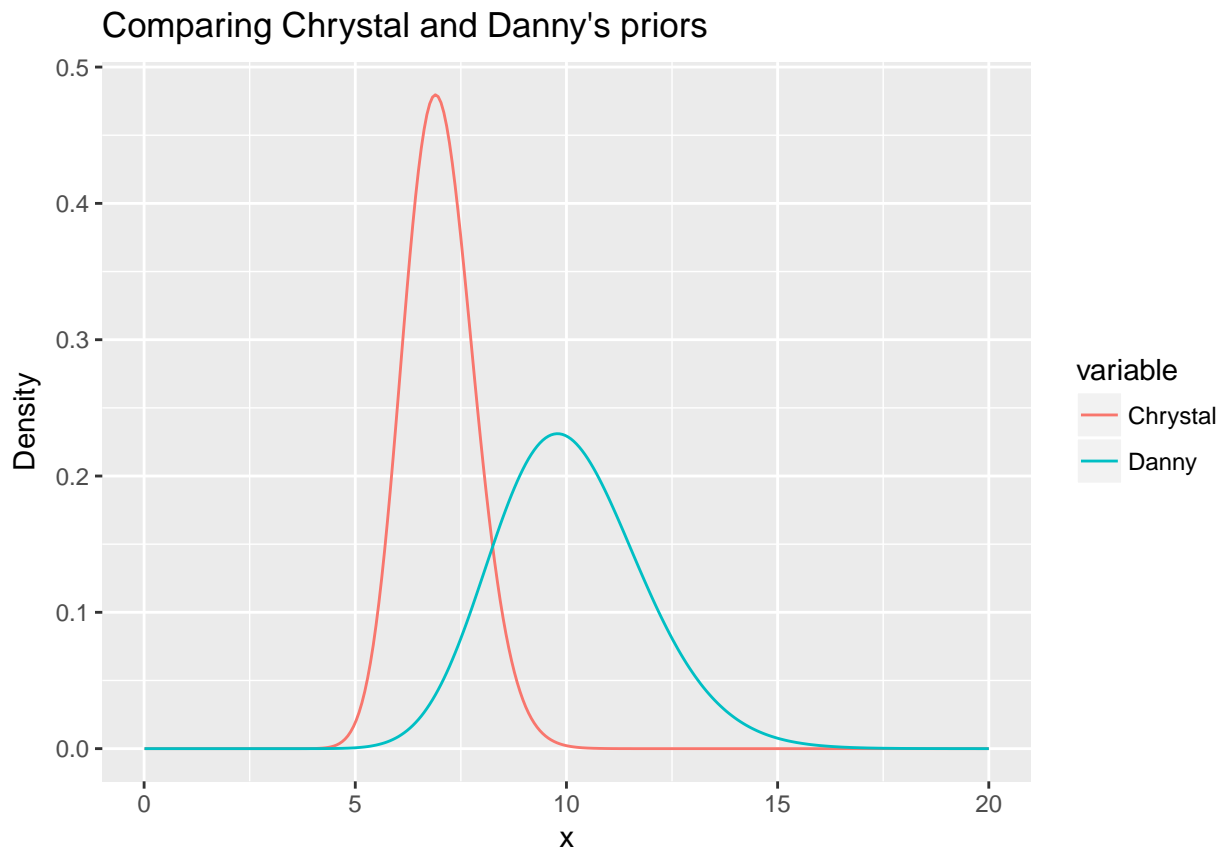
Shashank Sule

10/31/2019

Problem 1

a) Chrystal and Danny's priors are as follows:

```
s <- seq(0.01, 20, length.out = 300)
Chrystal <- sapply(s, dgamma, shape = 70, rate = 10)
Danny <- sapply(s, dgamma, shape = 33.3, rate = 3.3)
dataf <- data.frame(s, Chrystal, Danny)
data <- melt(dataf, id="s")
ggplot(data, aes(x=s, y=value, color=variable)) +
  geom_line() +
  ggtitle("Comparing Chrystal and Danny's priors") +
  labs(x="x", y="Density")
```



b) Danny believes that there are more ER visits on average but Chrystal is more certain

in her belief.

c) Chrystal's 90% interval:

```
qgamma(c(0.05,0.95),70,10)
```

```
## [1] 5.682967 8.430648
```

Danny's 90% interval:

```
qgamma(c(0.05,0.95),33.3,3.3)
```

```
## [1] 7.396876 13.128930
```

d) If Chrystal is less certain about her prior beliefs but believes in the same average, she must increase the variance of her gamma distribution but keep its mean the same. If she scales her α and β by the same amount λ then her new mean would be $\lambda\alpha/\lambda\beta = \alpha/\beta$, i.e her previous mean. But scaling by lambda causes her new variance to be $\lambda\alpha/\lambda^2\beta^2 = (1/\lambda)(\alpha/\beta^2)$ so by scaling, her variance gets scaled by $1/\lambda^2$. As a result, if she wants to increase her variance, she can scale her parameters by $\lambda < 1$ to get a gamma distribution of identical mean but higher variance.

Problem 2

a) The 95% credible interval for λ is

```
Y <- c(8,6,6,9,8,9,7)
```

```
n <- 7
```

```
qgamma(c(0.025,0.975), shape = 70+sum(Y), rate = 10+n)
```

```
## [1] 6.013237 8.568716
```

b) From the credible interval we see that the left endpoint is almost 6 so we can already infer that the probability $p(\lambda \leq 6 \mid \mathbf{Y}) \approx 0.025$.

```
pgamma(6, shape = 70+sum(Y), rate = 10+n)
```

```
## [1] 0.02368644
```

Almost! So the hospital administrator is only right about 2.4% of the time.

c) I interpret this question as asking to predict the number of total ER visits during 10-11pm for another week. I assume that the number of visits in a week is modeled by a poisson process so the poisson parameters scales from the per-day parameter λ to 7λ . Then predictive distribution using monte carlo sampling is as follows:

```
m <- 10000
```

```
ERVisits <- numeric(m)
```

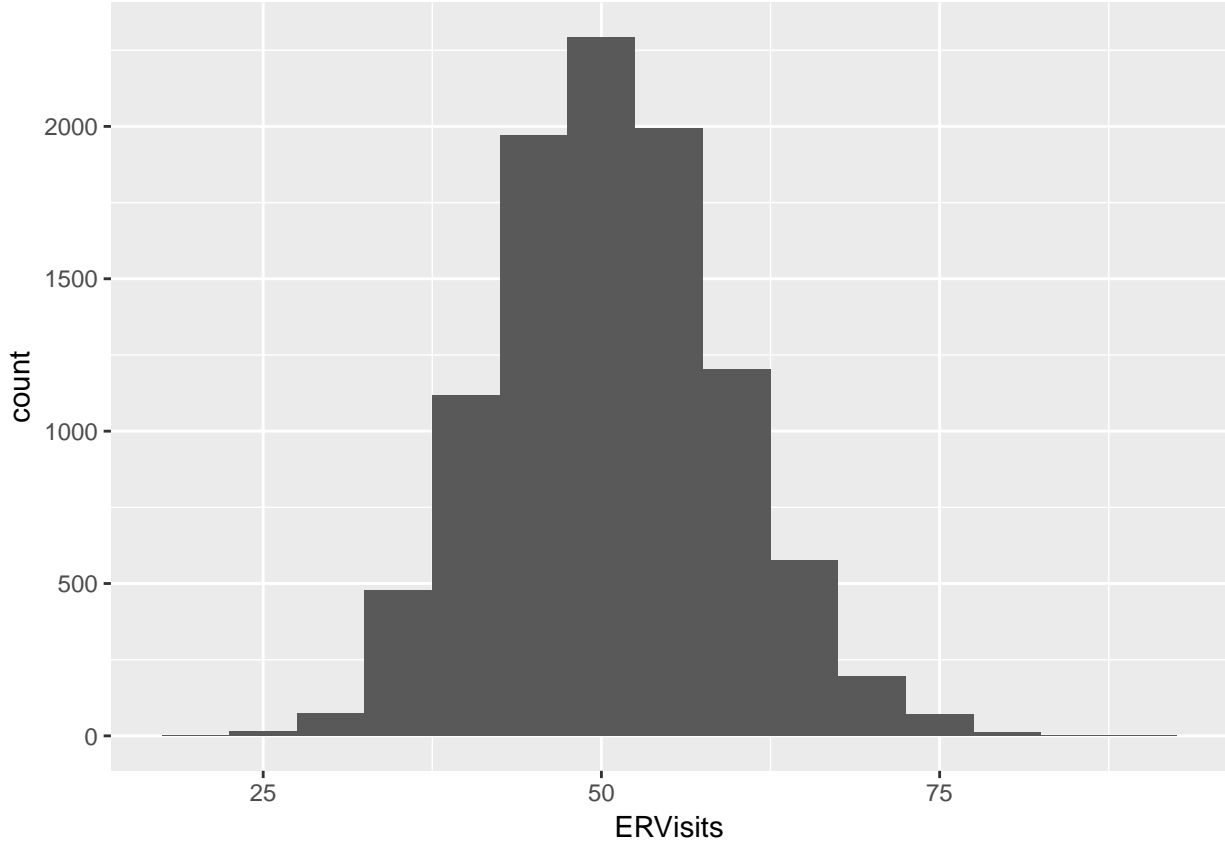
```
for(i in 1:m){
```

```

  lambdax <- rgamma(1, shape = 70+sum(Y), rate = 10+n)
  ERVisits[i] <- rpois(1, lambda = 7*lambdax)
}

ggplot(as.data.frame(ERVisits), aes(x=ERVisits))+
  geom_histogram(binwidth = 5)

```



Problem 3

- a) Note that $P(\mathbf{Y} \mid N, \theta, \lambda) = P(\mathbf{Y} \mid N, \theta \mid \lambda) = \text{Binom}(N, \theta)$. Furthermore, $P(N, \theta, \lambda) = P(N \mid \theta, \lambda) * P(\theta, \lambda) \propto e^{-\lambda/\theta} \frac{(\lambda/\theta)^N}{N!} \frac{1}{\lambda}$. Thus,

$$\begin{aligned}
 \pi(N, \theta, \lambda \mid \mathbf{Y}) &\propto L(N, \theta, \lambda) P(N, \theta, \lambda) \\
 &\propto \prod_{1 \leq i \leq 5} \binom{N}{y_i} \theta^{y_i} (1 - \theta)^{N - y_i} e^{-\lambda/\theta} \frac{(\lambda/\theta)^N}{N!} \frac{1}{\lambda} \\
 &\propto \prod_{1 \leq i \leq 5} \binom{N}{y_i} \theta^{\sum y_i} (1 - \theta)^{5N - \sum y_i} \frac{\lambda^{N-1} e^{-\lambda/\theta}}{\theta^N N!}
 \end{aligned}$$

b) If we regard all expressions not containing λ as constants, we have that

$$\pi(\lambda \mid N, \theta, \mathbf{Y}) \propto \lambda^{N-1} e^{(-\frac{1}{\theta})\lambda} \sim \text{Gamma}(N, 1/\theta)$$

c) Since we know that $\pi(\lambda \mid N, \theta, \mathbf{Y}) \sim \text{Gamma}(N, 1/\lambda)$ we have that $\int \pi(\lambda \mid N, \theta, \mathbf{Y}) d\lambda = \frac{\Gamma(N)}{(1/\theta)^N} = \frac{(N-1)!}{(1/\theta)^N} = (N-1)! \theta^N$ so

$$\begin{aligned} f(N, \theta \mid \mathbf{Y}) &= \int \pi(N, \theta, \lambda \mid \mathbf{Y}) d\lambda \propto \prod_{1 \leq i \leq 5} \binom{N}{y_i} \theta^{\sum y_i} (1-\theta)^{5N - \sum y_i} \frac{1}{\theta^N N!} \int \lambda^{N-1} e^{(-\frac{1}{\theta})\lambda} \\ &= \prod_{1 \leq i \leq 5} \binom{N}{y_i} \theta^{\sum y_i} (1-\theta)^{5N - \sum y_i} \frac{1}{\theta^N N!} (N-1)! \theta^N \end{aligned}$$

Cancelling the θ^N terms and simplifying $(N-1)!/N!$ as $1/N$ we get that

$$f(N, \theta \mid \mathbf{Y}) \propto \frac{1}{N} \prod_{1 \leq i \leq 5} \binom{N}{y_i} \theta^{\sum y_i} (1-\theta)^{5N - \sum y_i}$$

The right hand side is the kernel of the distribution.

d) To get $f(\theta \mid N, \mathbf{Y})$ we find the part of the kernel dependent on θ :

$$f(\theta \mid N, \mathbf{Y}) \propto \theta^{\sum y_i} (1-\theta)^{5N - \sum y_i} = \theta^{(\sum y_i + 1) - 1} (1-\theta)^{(5N - \sum y_i + 1) - 1} \sim \text{Beta}\left(\sum y_i + 1, 5N - \sum y_i + 1\right)$$

Problem 4

- Comparing Figure 1 with Figure 2, we see that the density of the MC samples is higher around the second mode, 0 while the MCMC samples have little density around that mode. This means that MC yielded more samples around 0 while the Gibbs samples were almost exclusively clustered around either -3 or 3. This is reflected in Figures 3 and 4, where the data is more evenly spread for MC while it clusters around 3 and -3 for the Gibbs samples.
- The traceplots in Figures 5 and 6 reveal the trace left by the data points from Figures 3 and 4. While the samples in figure 5 seem to be spreading evenly across the possible values, the samples in figure 6 are mostly stuck around -3 and 3. Thus the traceplots reveal insufficient mixing for Gibbs sampling as the samples are not fully exploring the parameter space (they are not exploring the region around 0 where they should have a locally large density had they been sampled from the posterior).
- The ACF plot for the MC simulations reveals that there is little correlation between the samples while the one for Gibbs sampling reveals a high correlation, or stickiness. Thus,

the issue is also that the samples seem to be more correlated in the Gibbs algorithm. This is proved further from the evidence of the effective sample size. For MC, the effective sample size is the number of samples, which means that the MC samples were almost perfectly independent. On the other hand, the Gibbs effective sampling size was higher than 1000 which reveals that a higher number of draws would have been required to achieve independence.

- d) Two strategies can be implemented: first we can choose a higher thinning or burn-in rate for the Gibbs sampling. Second, we may choose to run multiple chains and pick the end value of each chain to shake off the correlation.