Bayesian Data Analysis: Homework 4

Joey Miller

March 10, 2016

Problem 2.17 Highest posterior densities.

a. Using the transformation $u = \sigma^2$, we have $\sigma = \sqrt{u}$. Given that $p(\sigma) \propto \sigma^{-1}$, $p(u)$ can be found

$$\begin{aligned} p(u) &= p(\sigma)\frac{d\sigma}{du} \\ &= p(\sqrt{u})\frac{1}{2\sqrt{u}} \\ &\propto \frac{1}{\sqrt{u}}\frac{1}{2\sqrt{u}} \\ &\propto \frac{1}{u} = \frac{1}{\sigma^2} \end{aligned}$$

b. A necessary condition for the HDI to be invariant to transformation is that the mode (found through differentiating the log-posterior and setting equal to zero) is invariant to transformation. I will also use the special case $n = 1$. The likelihood is χ^2_1.

$$\begin{aligned} p(\sigma^2|\text{data}) &\propto p(\text{data}|\sigma^2)p(\sigma^2) \\ &\propto \left(\frac{v}{\sigma^2}\right)^{\frac{1}{2}-1} e^{-\frac{v}{2\sigma^2}}\frac{1}{\sigma^2} \\ &\propto \left(\frac{1}{\sigma^2}\right)^{\frac{1}{2}} e^{-\frac{v}{2\sigma^2}} \\ p(\sigma|\text{data}) &\propto p(\text{data}|\sigma)p(\sigma) \\ &\propto \left(\frac{v}{\sigma^2}\right)^{\frac{1}{2}-1} e^{-\frac{v}{2\sigma^2}}\frac{1}{\sigma} \\ &\propto e^{-\frac{v}{2\sigma^2}} \\ \log(p(\sigma^2|\text{data})) &\propto -\frac{1}{2}\log(\sigma^2) - \frac{v}{2\sigma^2} \\ \frac{d}{d\sigma^2}\log(p(\sigma^2|\text{data})) &\Rightarrow \sigma^2 = \frac{v}{2} \\ \log(p(\sigma|\text{data})) &\propto -\frac{v}{2\sigma^2} \\ \frac{d}{d\sigma}\log(p(\sigma|\text{data})) &\Rightarrow \text{no maximum.} \end{aligned}$$

So clearly, $\text{mode}(p(\sigma|\text{data}))^2 \neq \text{mode}(p(\sigma^2|\text{data}))$ and the general case cannot hold.

Problem 2.19 Exponential with gamma prior.

a. Given $y|\theta \sim \text{Exp}(\theta)$ and $\theta \sim \text{Gamma}(\alpha, \beta)$,

$$\begin{aligned} p(\theta|y) &\propto p(y|\theta)p(\theta) \\ &\propto \theta e^{-\theta y}\theta^{\alpha-1}e^{-\beta\theta} \\ &\propto \theta^{\alpha+1-1}e^{-(\beta+y)\theta} \end{aligned}$$

which is the kernel of a Gamma$(\alpha+1, \beta+y)$ distribution. It is easily seen that to generalize this to an n-vector $\vec{y}$, we get a Gamma$(\alpha+n, \beta+\Sigma_{i=1}^n y_i)$ distribution.

b. Given $y|\phi \sim \text{Exp}(1/\phi)$ and $\phi \sim \text{Gamma}(\alpha, \beta)$,

$$\begin{aligned} p(\phi|y) &\propto p(y|\phi)p(\phi) \\ &\propto \frac{1}{\phi}e^{-y/\phi}\phi^{-(\alpha+1)}e^{-\beta/\phi} \\ &\propto \phi^{-(\alpha+1-1)}e^{-(y+\beta)/\phi} \end{aligned}$$

which is the kernel of an Inv-Gamma$(\alpha+1, \beta+y)$ distribution. As before, we can generalize this easily to an n-dimensional vector, $\vec{y}$, where we get an Inv-Gamma$(\alpha+n, \beta+\Sigma_{i=1}^n y_i)$.

c. Setting the coefficient of variation of a Gamma(α, β) distribution to $1/2$, we get

$$\frac{\text{SD}}{\text{mean}} = \frac{\sqrt{\alpha/\beta^2}}{\alpha/\beta} = \frac{1}{\sqrt{\alpha}} = \frac{1}{2}$$

which implies $\alpha = 4$. To reduce the coefficient of variation to $1/10$, simply solve $1/\sqrt{\alpha_{post}} = 1/10$, so $\alpha_{post} = 100$. As pointed out before, $\alpha_{post} = \alpha + n$, so $n = 96$.

d. Doing the same to an Inv-Gamma(α, β) distribution, we get

$$\frac{\text{SD}}{\text{mean}} = \frac{\sqrt{\frac{\beta^2}{(\alpha-1)^2(\alpha-2)}}}{\frac{\beta}{\alpha-1}} = \frac{1}{\sqrt{\alpha-2}} = \frac{1}{2}$$

which implies $\alpha = 6$. Analogously to before, $1/\sqrt{\alpha_{post}-2} = 1/10$, so $\alpha_{post} = 102$. Again, $\alpha_{post} = \alpha + n$, so $n = 96$, as before.

Problem 3.1 Multinomial distribution with Dirichlet prior (parameters $\alpha_1, \ldots, \alpha_J$).

a. To find the marginal distribution for $\theta_1/(\theta_1+\theta_2)$, the transformation $a = \theta_1/(\theta_1+\theta_2)$, $b = (\theta_1+\theta_2)$ is helpful. Inverting, we get $\theta_1 = ab$ and $\theta_2 = b - ab$. The Jacobian,

$$|J| = \begin{vmatrix} \beta & \alpha \\ -\beta & 1-\alpha \end{vmatrix} = \beta.$$

Recall the multinomial-Dirichlet posterior after integrating out (as demonstrated in class) $\theta_3, \ldots, \theta_J$ which I will make the substitution into and simplify:

$$\begin{aligned} p(\vec{\theta}|\text{data}) &\propto \theta_1^{y_1+\alpha_1-1}\theta_2^{y_2+\alpha_2-1}(1-(\theta_1+\theta_2))^{\Sigma y_i+\Sigma\alpha_i-1} \\ p(a,b|\text{data}) &\propto (ab)^{y_1+\alpha_1-1}(b-ab)^{y_2+\alpha_2-1}(1-b)^{\Sigma y_i+\Sigma\alpha_i-1}b \\ &\propto a^{y_1+\alpha_1-1}(1-a)^{y_2+\alpha_2-1}b^{y_1+y_2+\alpha_1+\alpha_2-1}(1-b)^{\Sigma_{i=3}(y_i+\alpha_i)-1} \end{aligned}$$

which is a product of two independent Beta distributions. The portion with a in it is a Beta$(y_1+\alpha_1, y_2+\alpha_2)$, which is the distribution of $\theta_1/(\theta_1+\theta_2)$.

b. Considering y_1 as an observation from a binomial distribution with probabilioty a and number of observations $n = y_1 + y_2$, we also arrive at the Beta$(y_1+\alpha_1, n-y_1+\alpha_2) = \text{Beta}(y_1+\alpha_1, y_2+\alpha_2)$ distribution.

Problem 3.7 Independent Poisson distributions and a Binomial distribution. The joint distribution, $p(b, v)$ is straightforward to calculate because it is only a product of two poisson distributions. The binomial model, $p(b|n)$ is dependent upon the sample size $n = b + v$, which is itself a poisson random variable with parameter $\theta = \theta_b + \theta_v$. We also have that $p = \theta_b/(\theta_b + \theta_v)$.

$$\begin{aligned}
p(b,c) = p_{\text{pois}}(b)p_{\text{pois}}(v) &= \frac{\theta_b^b \theta_v^v e^{-(\theta_b+\theta_v)}}{b!v!} \\
p(b,c) = p_{\text{bin}}(b|n)p_{\text{pois}}(n) &= \binom{n}{v} p^b (1-p)^{n-v} \frac{(\theta_b+\theta_v)^n e^{-(\theta_b+\theta_v)}}{n!} \\
&= \binom{b+v}{v} p^b (1-p)^b \frac{(\theta_b+\theta_v)^{b+v} e^{-(\theta_b+\theta_v)}}{(b+v)!} \\
&= \frac{(b+v)!}{b!v!} \left(\frac{\theta_b}{\theta_b+\theta_v}\right)^b \left(\frac{\theta_v}{\theta_b+\theta_v}\right)^v \frac{(\theta_b+\theta_v)^{b+v} e^{-(\theta_b+\theta_v)}}{(b+v)!} \\
&= \frac{1}{b!v!} \frac{\theta_b^b}{(\theta_b+\theta_v)^b} \frac{\theta_v^v}{(\theta_b+\theta_v)^v} \frac{(\theta_b+\theta_v)^{b+v} e^{-(\theta_b+\theta_v)}}{1} \\
&= \frac{\theta_b^b \theta_v^v e^{-(\theta_b+\theta_v)}}{b!v!}
\end{aligned}$$

Problem 3.9 Conjugate normal model where $y \sim \mathrm{N}(\mu, \sigma^2)$ and $(\mu, \sigma^2) \sim \text{N-Inv-}\chi^2(\mu_0, \sigma^2/k_0; v_0, \sigma_0^2)$ with a dataset of size n.

$$\begin{aligned}
p(\mu, \sigma^2|\vec{y}) &\propto p(\vec{y}|\mu, \sigma^2)p(\mu|\sigma^2/k_0)p(\sigma^2) \\
&\propto \Pi_{i=1}^n \left(\frac{1}{(\sigma^2)^{1/2}} \exp\left(\frac{(y_i - u)^2}{2\sigma^2}\right)\right) \left(\frac{k_0}{\sigma^2}\right)^{1/2} (\sigma^2)^{-(v_0/2+1)} \exp\left(-\frac{v_0\sigma_0^2 + k_0(\mu_0 - \mu)^2}{2\sigma^2}\right) \\
&\propto \frac{1}{(\sigma^2)^{n/2}} \exp\left(\frac{(n-1)s^2 + n(\bar{y} - \mu)^2 + v_0\sigma_0^2 + k_0(\mu_0 - \mu)^2}{2\sigma^2}\right) \left(\frac{k_0}{\sigma^2}\right)^{1/2} (\sigma^2)^{-(v_0/2+1)} \\
&\propto \frac{1}{(\sigma^2)^{\frac{n+v_0+3}{2}}} \exp\left(\frac{(n-1)s^2 + n(\bar{y} - \mu)^2 + v_0\sigma_0^2 + k_0(\mu_0 - \mu)^2}{2\sigma^2}\right) \\
&\propto \frac{1}{(\sigma^2)^{\frac{n+v_0+3}{2}}} \exp\left(\frac{(n-1)s^2 + v_0\sigma_0^2 + n\bar{y}^2 + k_0\mu_0^2 - 2n\bar{y}\mu + n\mu^2 - 2k_0\mu_0\mu + k_0\mu^2}{2\sigma^2}\right) \\
&\propto \frac{1}{(\sigma^2)^{\frac{n+v_0+3}{2}}} \exp\left(\frac{(n-1)s^2 + v_0\sigma_0^2 + \frac{nk_0(\bar{y}-\mu_0)^2}{n+k_0} + (n+k_0)(\mu - \frac{n\bar{y}+\mu_0 k_0}{n+k_0})^2}{2\sigma^2}\right)
\end{aligned}$$

where s^2 is the sample standard deviation and $\bar{y}$ is the mean. In the last step, I completed the square in μ. Matching up to the joint prior distribution, we have

$$\mu, \sigma^2 \sim \text{N-Inv-scaled-}\chi^2\left(\frac{n\bar{y} + \mu_0 k_0}{n + k_0}, \frac{\sigma_n^2}{n + k_0}; n + v_0, \frac{(n-1)s^2 + v_0\sigma_0^2 + \frac{nk_0(\bar{y}-\mu_0)^2}{n+k_0}}{n + v_0}\right)$$

Problem 3.10 To show that the posterior ratio $(s_1^2/\sigma_1^2)/(s_2^2/\sigma_2^2)$ is distributed as an F random variable with (n_1-1) and (n_2-2) degrees of freedom, I will show that s_j^2/σ_j^2 are distributed independently as $\chi^2_{n_j-1}$, which is equivalent to an F distribution with the corresponding degrees of freedom. First, we have that $y_{j1}, \ldots, y_{jn_j}|\mu_j, \sigma_j^2 \sim \text{iid } \mathrm{N}(\mu_j, \sigma_j^2)$ and $p(\mu_j, \sigma_j^2) \propto \sigma_j^{-2}$, where (μ_j, σ_j^2) are also independent. The posterior, $\sigma_j^2|\vec{y}$ is found as follows,

$$\begin{aligned}
p(\sigma_j^2|\vec{y}) &= \int_{\mu_j} p(\mu_j, \sigma_j^2|\vec{y}) \\
&\propto \int_{\mu_j} p(\vec{y}_j|\mu_j, \sigma_j^2)p(\mu_j, \sigma_j^2) \\
&\propto \int_{\mu_j} \frac{1}{\sigma_j^{n+2}} \exp\left(-\frac{(n_j-1)s_j^2 + n_j(\bar{y}_j - \mu_j)^2}{2\sigma_j^2}\right) \\
&\propto \frac{1}{\sigma_j^{n+1/2}} \exp\left(-\frac{(n_j-1)s^2}{2\sigma_j^2}\right)
\end{aligned}$$

which is the kernel of a scaled Inv-$\chi^2(n_j - 1, s_j^2)$. Now note that the distribution σ_j^2/s_j^2 must be an Inv-$\chi^2_{n_j-1}$ distribution implying that its reciprocal, s_j^2/σ_j^2 is distributed $\chi^2_{n_j-1}$ (and still independent for all j), the desired result.

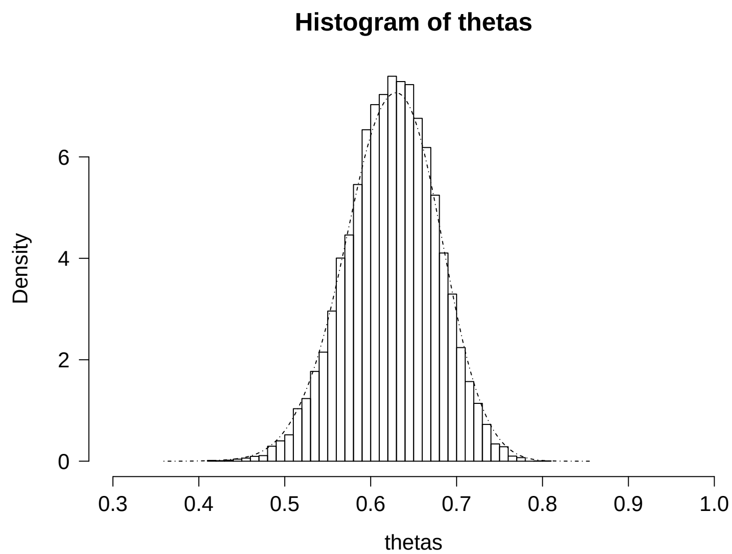
Part 2 Genetic linkage model with data $Y = (125, 18, 20, 34)$ with Beta$(1/2, 1/2)$ prior for θ.

a. The moving acceptance rate was approximately 0.5328.

b. The posterior mean, μ_{post}, and variance, σ^2_{post} is computed by evaluating the integrals:

$$\mu_{\text{post}} = E[\theta_{\text{post}}] = \frac{\int_0^1 \theta(2+\theta)^{125}(1-\theta)^{37.5}\theta^{33.5}}{\int_0^1 (2+\theta)^{125}(1-\theta)^{37.5}\theta^{33.5}} \approx 0.624177$$

$$\sigma^2_{\text{post}} = E[(\theta_{\text{post}} - \mu_{\text{post}})^2] = \frac{\int_0^1 (\theta-\mu_{\text{post}})^2(2+\theta)^{125}(1-\theta)^{37.5}\theta^{33.5}}{\int_0^1 (2+\theta)^{125}(1-\theta)^{37.5}\theta^{33.5}} \approx .00261699$$



c. The approximate posterior mean is 0.6245139 and the approximate posterior variance is 0.002649265, found by calling *mean()* and *var()* on the vector *thetas* in R.

Appendix: R code

```
par(mfrow=c(2,1),las=1.0)

y <- c(125,18,20,34)
n <- sum(y)
a <- 1/2
b <- 1/2

# likelihood function .......
f<-function(theta){
  (2+theta)^y[1]*(1-theta)^(y[2]+y[3])*theta^y[4]
  }

# Using numerical method to find the posterior density of theta
t1<-seq(0.005,.995,0.005)
post<-f(t1)*dbeta(t1, shape1 = a, shape2 = b)
post<-post/(sum(post*0.005))
```

```
# plot(t1, post, type='l')

l<-500
m<-20000
theta    <- rep(NA,l+m)
theta[1] <- 0.1

N<-1
for (k in 2:(l+m)){
      thetanew <- rbeta(n = 1, shape1 = 14,shape2 = 9)
      theta[k]<-theta[k-1]
      u<-runif(1)
      r=min(1, f(thetanew)/f(theta[k-1]))
      if (u < r) {
           theta[k] <- thetanew
           N<-N+1
           }
      }
# Part a:
# ---------------- The proportion of moving: ------
N/(l+m)
# 0.532878

plot(theta,type='l')
thetas<-theta[(l+1):(l+m)]

# Histogram of the density of theta
hist(thetas,probability=T,nclass=40,xlim=c(.3, 1))
# ---- plot the real posterior density---------------
#  plot(t1,post,type='l')
lines(t1,post,lty=1)

# plot Kernel density of theta:
 summary(thetas)
 ipd<-summary(thetas)[5]-summary(thetas)[2]
 ipd
 lines(density(thetas,width=ipd),lty=2)

# Part b:
par(mfrow=c(1,1),las=1.0)

y <- c(125,18,20,34)
n <- sum(y)
a <- 1/2
b <- 1/2

l<-500
m<-20000
theta    <- rep(NA,l+m)
Z        <- rep(NA,l+m)
theta[1] <- 0.8
Z[1]     <- rbinom(1,y[1],theta[1])

for (k in 2:(l+m)){
        theta[k] <- rbeta(1,y[1]-Z[k-1]+y[4]+a, y[2]+y[3]+b);
        Z[k]     <- rbinom(1,y[1],2/(theta[k]+2))
      }
```

```
plot(theta,type='l')
thetas <- theta[(l+1):(l+m)]
Zs <- Z[(l+1):(l+m)]

# Histogram of the density of theta
# hist(thetas,probability=T)
 hist(thetas,probability=T,nclass=40,xlim=c(.3, 1))

# plot Kernel density of theta:
 summary(thetas)
 ipd<-summary(thetas)[5]-summary(thetas)[2]
 ipd
 lines(density(thetas,width=ipd),lty=4)

# Mean and Variance
mean(thetas) # 0.6245139
var(thetas) # 0.002649265
```