Introduction to Bayesian Data Analysis

Tutorial 8 - Solutions

(1) (a) $Var[y_{i,j}|\mu, \tau^2]$ - extra element of variability in first sampling a group.

(b) $Cov[y_{i_1,j}, y_{i_2,j}|\theta_j, \sigma^2] = 0$ - - given the group membership, we assume data are independent. $Cov[y_{i_1,j}, y_{i_2,j}|\mu, \tau^2] > 0$ - - not knowing the group membership, then $y_{i_2,j}$ is informative about $y_{i_1,j}$ because they come from the same subpopulation.

(c) $Var[y_{i,j}|\theta_i, \sigma^2] = \sigma^2$

$Var[y_{i,j}|\mu, \tau^2] = E[Var[y_{i,j}|\theta_j, \mu, \tau^2]\big|\mu, \tau^2] + Var[E[y_{i,j}|\theta_j, \mu, \tau^2]\big|\mu, \tau^2] = E[\sigma^2\big|\mu, \tau^2] + Var[\theta_j\big|\mu, \tau^2] = \sigma^2 + \tau^2 > Var[y_{i,j}|\theta_i, \sigma^2]$

$Var[\bar{y}_{\cdot,j}|\theta_i, \sigma^2] = \sigma^2/n_j$

$Var[\bar{y}_{\cdot,j}|\mu, \tau^2] = E[Var[\bar{y}_{\cdot,j}|\theta_j, \mu, \tau^2]\big|\mu, \tau^2] + Var[E[\bar{y}_{\cdot,j}|\theta_j, \mu, \tau^2]\big|\mu, \tau^2] = E[\sigma^2/n_j\big|\mu, \tau^2] + Var[\theta_j\big|\mu, \tau^2] = \sigma^2/n_j + \tau^2 > Var[\bar{y}_{\cdot,j}|\theta_i, \sigma^2]$

$Cov[y_{i_1,j}, y_{i_2,j}|\theta_j, \sigma^2] = 0$

$Cov[y_{i_1,j}, y_{i_2,j}|\mu, \tau^2] = E[Cov[y_{i_1,j}, y_{i_2,j}|\theta_j, \mu, \tau^2]\big|\mu, \tau^2] +$
$Cov[E[y_{i_1,j}|\theta_j, \mu, \tau^2]E[y_{i_2,j}|\theta_j, \mu, \tau^2]\big|\mu, \tau^2] = E[0\big|\mu, \tau^2] + Cov[\theta_j \times \theta_j\big|\mu, \tau^2] = \tau^2 (> 0)$

Our answers in (c) are in line with our intuition in parts (a) and (b)

(d)

$$\begin{aligned}
p(\mu|\theta_1,...,\theta_m,\sigma^2,\tau^2,\mathbf{y}_1,...,\mathbf{y}_m) &= \frac{p(\mu)\times p(\theta_1,...,\theta_m,\sigma^2,\tau^2,\mathbf{y}_1,...,\mathbf{y}_m|\mu)}{p(\theta_1,...,\theta_m,\sigma^2,\tau^2,\mathbf{y}_1,...,\mathbf{y}_m)} \\
&= \frac{p(\mu)p(\mathbf{y}_1,...,\mathbf{y}_m|\theta_1,...,\theta_m,\sigma^2,\tau^2,\mu)p(\theta_1,...,\theta_m,\sigma^2,\tau^2|\mu)}{p(\mathbf{y}_1,...,\mathbf{y}_m|\theta_1,...,\theta_m,\sigma^2,\tau^2)p(\theta_1,...,\theta_m,\sigma^2,\tau^2)} \\
&= \frac{p(\mu)p(\mathbf{y}_1,...,\mathbf{y}_m|\theta_1,...,\theta_m,\sigma^2)p(\theta_1,...,\theta_m|\sigma^2,\tau^2,\mu)p(\sigma^2,\tau^2|\mu)}{p(\mathbf{y}_1,...,\mathbf{y}_m|\theta_1,...,\theta_m,\sigma^2)p(\theta_1,...,\theta_m|\sigma^2,\tau^2,\mu)p(\sigma^2,\tau^2)} \\
&= \frac{p(\mu)p(\theta_1,...,\theta_m|\tau^2,\mu)p(\sigma^2|\tau^2,\mu)p(\tau^2|\mu)}{p(\theta_1,...,\theta_m|\tau^2)p(\sigma^2|\tau^2)p(\tau^2)} \\
&= \frac{p(\mu)p(\theta_1,...,\theta_m|\tau^2,\mu)p(\tau^2|\mu)}{p(\theta_1,...,\theta_m|\tau^2)p(\tau^2)} \\
&= p(\mu|\theta_1,...,\theta_m,\tau^2)
\end{aligned}$$

where the second to last line assumes independence between the (hyper)parameters, τ^2, σ^2 and μ.

The result in (d) means that posterior inference on μ does not depend directly on the data $\mathbf{y}_1,...,\mathbf{y}_m$, but rather on the posterior draws of the parameters $\theta_1,...,\theta_m$ and τ^2. The result is analagous to a one-sample normal model, where $\theta_1,...,\theta_m$ are i.i.d samples from a normal population, and μ and τ^2 are the unknown population mean and variance.

(2) (a) This is the R-code to implement the Gibbs sampler:

```
mu0<-75 ; g02<-100
del0.sens<-c(rep(-4,4),rep(-2,4),rep(0,4),rep(2,4),rep(4,4)) ;
t02.sens<-rep(c(10,50,100,500),5)
s20<-100; nu0<-2
##### starting values
ybarA<-75.2
sA<-7.3
ybarB<-77.5
sB<-8.1
nA<-nB<-16
mu<- (ybarA + ybarB )/2
del<- (ybarA- ybarB )/2
##### Gibbs sampler
MU<-DEL<-S2<-array(NA,c(20,5000))
for (i in 1:20)
{
del0<-del0.sens[i]
t02<-t02.sens[i]
set.seed(1)
for(s in 1:5000)
{
 ##update s2
  s2<-1/rgamma(1,(nu0+nA+nB)/2,
        (nu0*s20+(nA-1)*sA^2+nA*((mu+del)-ybarA)^2+(nB-1)*sB^2+nB*((mu-del)-
 ##update mu
  var.mu<-  1/(1/g02+ (nA+nB)/s2 )
  mean.mu<- var.mu*( mu0/g02 + nA*(ybarA-del)/s2 + nB*(ybarB+del)/s2 )
  mu<-rnorm(1,mean.mu,sqrt(var.mu))
  ##update del
  var.del<-  1/(1/t02+ (nA+nB)/s2 )
  mean.del<- var.del*( del0/t02 + nA*(ybarA-mu)/s2-  nB*(ybarB-mu)/s2 )
  del<-rnorm(1,mean.del,sqrt(var.del))
  ##save parameter values
  MU[i,s]<-mu ; DEL[i,s]<-del ; S2[i,s]<-s2
  }
}
> apply(DEL,1,function(x) mean(x<0))
```

(i) Values for $Pr(\delta_0 < 0|\mathbf{Y})$ are more sensitive across δ_0 values for lower τ_0^2 values, that is, as we are firmer on our prior beliefs for the value of δ.

τ_0^2	δ_0				
	-4	-2	0	2	4
10	0.90	0.84	0.77	0.68	0.58
50	0.82	0.80	0.78	0.77	0.75
100	0.80	0.79	0.79	0.78	0.77
500	0.79	0.79	0.79	0.79	0.78

Table 1: $Pr(\delta_0 < 0|\mathbf{Y})$

(ii) All 95% confidence intervals for δ contain zero, and the intervals are of similiar width for different combinations of δ_0 and τ_0^2.

τ_0^2	δ_0				
	-4	-2	0	2	4
10	(-4.2 , 0.91)	(-3.8 , 1.2)	(-3.4 , 1.6)	(-3.1 , 2.0)	(-2.7 , 2.4)
50	(-4.0 , 1.5)	(-3.9 , 1.6)	(-3.8 , 1.7)	(-3.7 , 1.1)	(-3.6 , 1.8)
100	(-4.0, 1.6)	(-3.9 , 1.6)	(-3.9 , 1.7)	(-3.8 , 1.7)	(-3.8 , 1.7)
500	(-4.0, 1.6)	(-3.9 , 1.7)	(-3.9 , 1.7)	(-3.9 , 1.7)	(-3.9 , 1.7)

Table 2: 95% posterior confidence intervals for δ

(iii) Prior correlation between θ_A and θ_B:

$$\begin{aligned}
Cov(\theta_A, \theta_B) &= Cov(\mu + \delta, \mu - \delta) \\
&= E[(\mu + \delta)(\mu - \delta)] - E[(\mu + \delta)]E[(\mu - \delta)] \\
&= E[\mu^2 - \delta^2] - (\mu_0 - \delta_0)(\mu_0 + \delta_0) \\
&= \mu_0^2 + \gamma_0^2 - \delta_0^2 - \tau_0^2 - \mu_0^2 - \delta_0^2 \\
&= \gamma_0^2 - \tau_0^2
\end{aligned}$$

$$Cor(\theta_A, \theta_B) = \frac{Cov(\theta_A, \theta_B)}{SD(\theta_A) \times SD(\theta_B)} = \frac{\gamma_0^2 - \tau_0^2}{\gamma_0^2 + \tau_0^2}$$

The posterior correlations are a lot smaller than the prior correlations and decrease in size as τ_0^2 increases, that is, as we assume higher prior variance on δ (see Table 3).

τ_0^2	Prior	δ_0				
		-4	-2	0	2	4
10	0.82	0.0914	0.0943	0.0944	0.0917	0.0862
50	0.33	0.0179	0.0181	0.0182	0.0181	0.0178
100	0.00	0.0074	0.075	0.0076	0.0076	0.0075
500	-0.67	-0.0012	-0.0011	-0.0011	-0.0011	-0.0011

Table 3: Posterior correlation between θ_A and θ_B

(b) For all prior opinions, $Pr(\delta_0 < 0|\mathbf{Y}) > 50\%$, which indicates that $\theta_B > \theta_A$, even if the prior belief on δ is greater than zero, and plots of the posterior density of δ are less diffuse than plots of the prior density on δ, so we are firmer in our beliefs on the value of δ and hence the relationship between θ_A and θ_B after incorporating the data into our inference.