HW3

Ishaan Srivastava (Collaborator: Harshil Desai)

10/30/2020

Question 1

We know that $\pi_{j|-j}(x_{(j)}|x_{-(j)}) = \frac{\pi(x)}{\int \pi(x)dx_{(j)}}$

$\pi(x_1, x_2, x_3) \propto exp(-2.25x_1^2 - 5x_2^2 - x_3^2 + 5x_1x_2 + 4x_2x_3 - 2x_1x_3)$

For each conditional density, we complete the square, treating $x_{(j)}$ as the variable of interest and $x_{-(j)}$ as constant

$\pi(x_1|x_2, x_3) \propto exp(-2.25(x_1^2 - \frac{5x_1x_2-2x_1x_3}{2.25}))$

By completing the square, we see that $(x_1 - \frac{5x_2-2x_3}{4.5})^2 = x_1^2 - \frac{5x_1x_2-2x_1x_3}{2.25} + (\frac{5x_2-2x_3}{4.5})^2$

$\propto x_1^2 - \frac{5x_1x_2-2x_1x_3}{2.25}$, since terms not involving x_1 are simply constants.

Therefore we get $\pi(x_1|x_2, x_3) \propto exp(-2.25(x_1 - \frac{5x_2-2x_3}{4.5})^2)$, meaning that $\pi(x_1|x_2, x_3) \sim N(\frac{5x_2-2x_3}{4.5}, \frac{1}{4.5})$.

The expectation of the distribution comes simply from the completing the square, while we solve $2.25 = \frac{1}{2\sigma^2}$ to get the variance. Both of these stem from the fact that the exponential term in a $N(\mu, \sigma^2)$ distribution is $\frac{-1}{2}(\frac{x-\mu}{\sigma})^2 = \frac{-1}{2\sigma^2}(x-\mu)^2$

Similarly we get, $\pi(x_2|x_1, x_3) \propto exp(-5(x_2^2 - (x_1x_2 + \frac{4x_2x_3}{5}))) \propto exp(-5(x_2 - (\frac{x_1}{2} + \frac{2x_3}{5}))^2)$

Therefore $\pi(x_2|x_1, x_3) \sim N(\frac{x_1}{2} + \frac{2x_3}{5}, \frac{1}{10})$

Also, $\pi(x_3|x_1, x_2) \propto exp(-(x_3^2 - (4x_2x_3 - 2x_1x_3))) \propto exp(-(x_3 - (2x_2 - x_1))^2)$

Therefore $\pi(x_3|x_1, x_2) \sim N(2x_2 - x_1, \frac{1}{2})$

Gibbs sampling algorithm:

Initialise $x_{(1)}^0, x_{(2)}^0, x_{(3)}^0$, where subscript denotes variable of interest and superscript denotes iteration

For t = 1:N

Pick j uniformly at random from {1, 2, 3}

Update $x_{-(j)}^t = x_{-(j)}^{t-1}$

Update $x_{(j)}^t$ by sampling from $\pi_{j|-j}(x_{(j)}|x_{-(j)})$, substituting the appropriate expression from above depending on the value of j

Add $(x_{(1)}^t, x_{(2)}^t, x_{(3)}^t)$ to list of samples

Question 2

$\beta \in R^p, w_i \sim N(x_i^T\beta, 1), y_i|\beta = max(0, w_i)$ where $y_1....y_n$ are independent conditional on β

Following modified version of approach to probit regression in Lecture 13.

Full posterior including data augmentation w_i is $\pi(w, \beta|data)$

$\pi(w, \beta|data) \propto \pi(\beta)\pi(w|\beta)\pi(y|w, \beta)$, where $\pi(\beta)$ is the prior for β. Since we take improper uniform prior for β on R^p, $\pi(w, \beta|data) \propto \prod_{i=1}^n exp(\frac{-1}{2}(w_i - x_i^T\beta)^2) \prod_{i=1}^n (I\{w_i > 0, y_i > 0\} + I\{w_i <= 0, y_i = 0\})$

$\propto exp(\frac{-1}{2}||w_i - x_i^T\beta||^2) \prod_{i=1}^n (I\{w_i > 0, y_i > 0\} + I\{w_i <= 0, y_i = 0\})$

Where X is design matrix with each row i equal to x_i^T.

We now perform block Gibbs sampling with two blocks, namely w and β. To do so, we first find the relevant conditional distributions

$\pi(\beta|w, data) \propto \pi(w, \beta|data) \propto exp(\frac{-1}{2}||w - X\beta||^2) \prod_{i=1}^n (I\{w_i > 0, y_i > 0\} + I\{w_i <= 0, y_i = 0\})$

$\propto exp(\frac{-1}{2}||w - X\beta||^2)$

Since residuals of linear regresion model are orthogonal to column space of X, we get

$exp(\frac{-1}{2}||w_i - x_i^T\beta||^2) = exp(\frac{-1}{2}(\beta - \hat{\beta}_w)(X^TX)(\beta - \hat{\beta}_w))$, where $\hat{\beta}_w = (X^TX)^{-1}X^Tw$, obtained from solving the normal equations.

Recognising the above as the distribution for multivariate normal, we get $\beta|w, data \sim N(\hat{\beta}, (X^TX)^{-1})$

Next, we find conditional distribution for w

$\pi(w|\beta, data) \propto \pi(w, \beta|data) \propto exp(\frac{-1}{2}||w - X^T\beta||^2) \prod_{i=1}^n (I\{w_i > 0, y_i > 0\} + I\{w_i <= 0, y_i = 0\}) \propto \prod_{i=1}^n exp(\frac{-1}{2}(w_i - x_i^T\beta)^2)(I\{w_i > 0, y_i > 0\} + I\{w_i <= 0, y_i = 0\})$

We split this into the two cases, where either $y_i > 0$ or $y_i = 0$. If $y_i > 0$, we have $\pi(w|\beta, y_i > 0) \propto exp(\frac{-1}{2}(w_i - x_i^T\beta)^2)I\{w_i > 0\}$ by conditional independence of all w_i given β and data.

Hence conditional distribution is $N(x_i^T\beta, 1)$, truncated to the positive half-line. Using formula for truncated normal from lecture 13, we know that $w_i \sim x_i^T\beta + \Phi^{-1}((1 - U)\Phi(-x_i^T\beta + U)$ when $y_i > 0$, where $U \sim Uniform(0, 1)$

Similarly, if $y_i = 0$, we have $\pi(w|\beta, y_i > 0) \propto exp(\frac{-1}{2}(w_i - x_i^T\beta)^2)I\{w_i <= 0\}$, which is $N(x_i^T\beta,\ 1)$, truncated to the non-positive half-line. Hence we know that $w_i \sim x_i^T\beta + \Phi^{-1}(U\Phi(-x_i^T\beta)$ when $y_i > 0$, where $U \sim Uniform(0, 1)$

Hence we now write out the full Gibbs sampling algorithm:

Initialise $\beta^{(0)}, w^{(0)}$, where superscript denotes iteration.

For t = 1:N

a) Pick j uniformly at random from {0, 1}

b i) If j = 0, Update $\beta^{(t)} = \beta^{(t-1)}$ (Note that this convention is arbitrary, and would work the same way if we updated β for j = 0 and vice versa)

ii) Sample $U \sim Uniform(0, 1)$

iii) If $y_i > 0$, sample $w_i \sim x_i^T\beta^{(t-1)} + \Phi^{-1}((1 - U)\Phi(-x_i^T\beta^{(t-1)} + U)$

If $y_i = 0$, sample $w_i \sim x_i^T\beta^{(t-1)} + \Phi^{-1}(U\Phi(-x_i^T\beta^{(t-1)}))$

c) If j = 1, $w^{(t)} = w^{(t-1)}$

Sample $\beta^{(t)} \sim N((X^TX)^{-1}X^Tw^{(t-1)}, (X^TX)^{-1})$

Question 3

Note: This solution references Page 71 of Bayesian Data Analysis

First we write down the likelihood and priors. $y_i|\beta, \sigma^2 \sim N_n(X\beta, \sigma^2 I_n)$, $\beta|\sigma \sim N_p(\beta_0, \sigma^2\ \Sigma_0)$, $\sigma^{-2} \sim Gamma(a, b)$. Note that $x_1, x_2...x_n, y_1, y_2...y_n \in R^{\ p}$ and the parameter of interest is $\theta = (\beta, \sigma)$, meaning $\theta \in R^{\ p+1}$

Thus $\pi(\beta, \sigma|y) \sim \pi(y|\beta, \sigma^2) \cdot \pi(\beta|\sigma) \cdot \pi(\sigma)$ After applying the appropriate change of density formula to get $\pi(\sigma)$ from the prior, we get

$\pi(\beta, \sigma|y) \sim \sigma^{-n} exp(-\frac{1}{2\sigma^2}(y - X\beta)^T(y - X\beta)) \cdot \sigma^{-p} exp(-\frac{1}{2\sigma^2}(\beta - \beta_0)^T \Sigma_0^{-1}(\beta - \beta_0)) \cdot \sigma^{-2a-1} exp(-b\sigma^{-2})$

Combining exponents, leveraging the fact that residuals of linear regression on X are orthognal to the column space of X, and rearranging terms, we get that

$\pi(\beta, \sigma|y) \propto \sigma^{-n-p-2a-1} exp(-\frac{1}{2\sigma^2}(||y - X\hat{\beta}||^2 + 2b)) \cdot exp(-\frac{1}{2\sigma^2}((\beta - \hat{\beta})^T X^T X(\beta - \hat{\beta}) + (\beta - \beta_0)^T \Sigma_0(\beta - \beta_0)))$

Use identity for product of multivariate normals. We define the following variables for convenience: $\mu_n = (X^T X + \Sigma_0^{-1})^{-1}(X^T X\hat{\beta} + \Sigma_0^{-1}\beta_0)$ and $\Lambda^{-1} = (X'X + \Sigma_0^{-1})$, since this is the convention the textbook follows. Then we get that:

$\pi(\beta, \sigma|y) \propto \sigma^{-n-p-2a-1} exp(-\frac{1}{2\sigma^2}(||y - X\hat{\beta}||^2 + 2b)) exp(-\frac{1}{2\sigma^2}(\beta - \mu_n)^T \Lambda^{-1}(\beta - \mu_n))$

Note the use of $\hat{\beta}$ instead of β and β_0 since we use properties of least squares regression, as mentioned before, to manipulate terms. Since our term of interest is $\pi(\beta|y)$, we integrate $\pi(\beta, \sigma|y)$ to eliminate σ and obtain desired term.

$\int_0^\infty \sigma^{-n-p-2a-1} exp(-\frac{1}{2\sigma^2}(||y - X\hat{\beta}||^2 + 2b + (\beta - \mu_n)^T \Lambda^{-1}(\beta - \mu_n)) d\sigma$

In this form, the integral is not easily tractable. Hence, we apply a change of variable.

$\sigma = v(||y - X\hat{\beta}||^2 + 2b + (\beta - \mu_n)^T \Delta^{-1}(\beta - \mu_n)^{1/2}$

Consequently, we get that $d\sigma = \frac{1}{2}(||y - X\hat{\beta}||^2 + 2b + (\beta - \mu_n)^T \Lambda^{-1}(\beta - \mu_n)^{-1/2} dv$

For convenience, let $g = (||y - X\hat{\beta}||^2 + 2b + (\beta - \mu_n)^T \Lambda^{-1}(\beta - \mu_n)$. Simplifying our expression, we get $\pi(\beta|y) \sim g^{\frac{-n-p-2a-1}{2}} g^{1/2} \int_0^\infty exp(-\frac{1}{v^2}) dv$

$\propto g^{(-n-p-2a)/2}$

Resubstituting, we get $\pi(\beta|y) \propto [1 + (\beta - \mu_n)^T \frac{\Lambda^{-1}}{||y - X\hat{\beta}||^2 + 2b}(\beta - \mu_n)]^{-(n+2a+p)/2}$

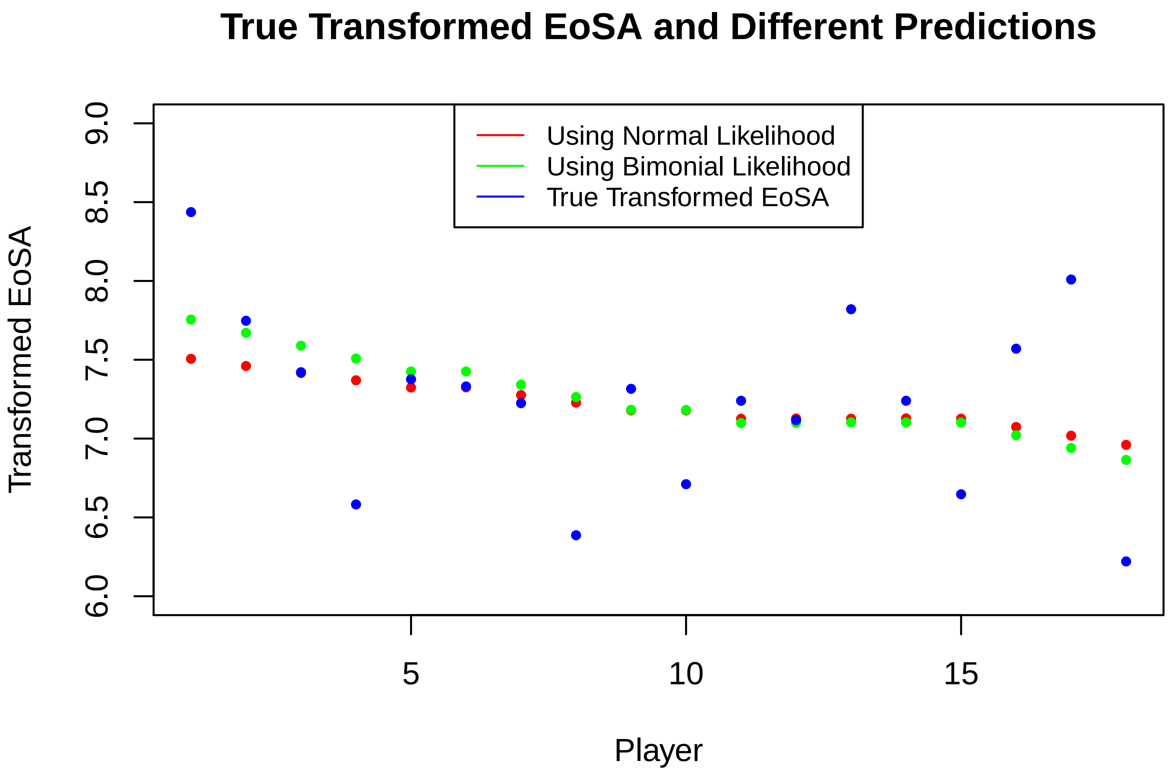
As seen in lecture, this density resembles multivariate t-distribution with $dof = n + 2a$. To get this in the form a multivariate t density, we rewrite as:

$\pi(\beta|y) \propto [1 + \frac{1}{n+2a}(\beta - \mu_n)^T \frac{\Lambda^{-1}(n+2a)}{||y - X\hat{\beta}||^2 + 2b}(\beta - \mu_n)]^{(-n-2a-p)/2}$

Hence we get that $\pi(\beta|y) \sim t_{n+2a}(\mu_n, (\frac{\Lambda^{-1}(n+2a)}{||y - X\hat{\beta}||^2 + 2b})^{-1})$

As an aside, we remember that the multivariate t-distribution differs from the univariate case in that it has two parameters instead of one. Hence the final answer is: $\pi(\beta|y) \sim t_{n+2a}(X^T X + \Sigma_0^{-1})^{-1}(X^T X\hat{\beta} + \Sigma_0^{-1}\beta_0), (\frac{||y - X\hat{\beta}||^2 + 2b}{n+2a})((X'X + \Sigma_0^{-1})^{-1})$

Question 4

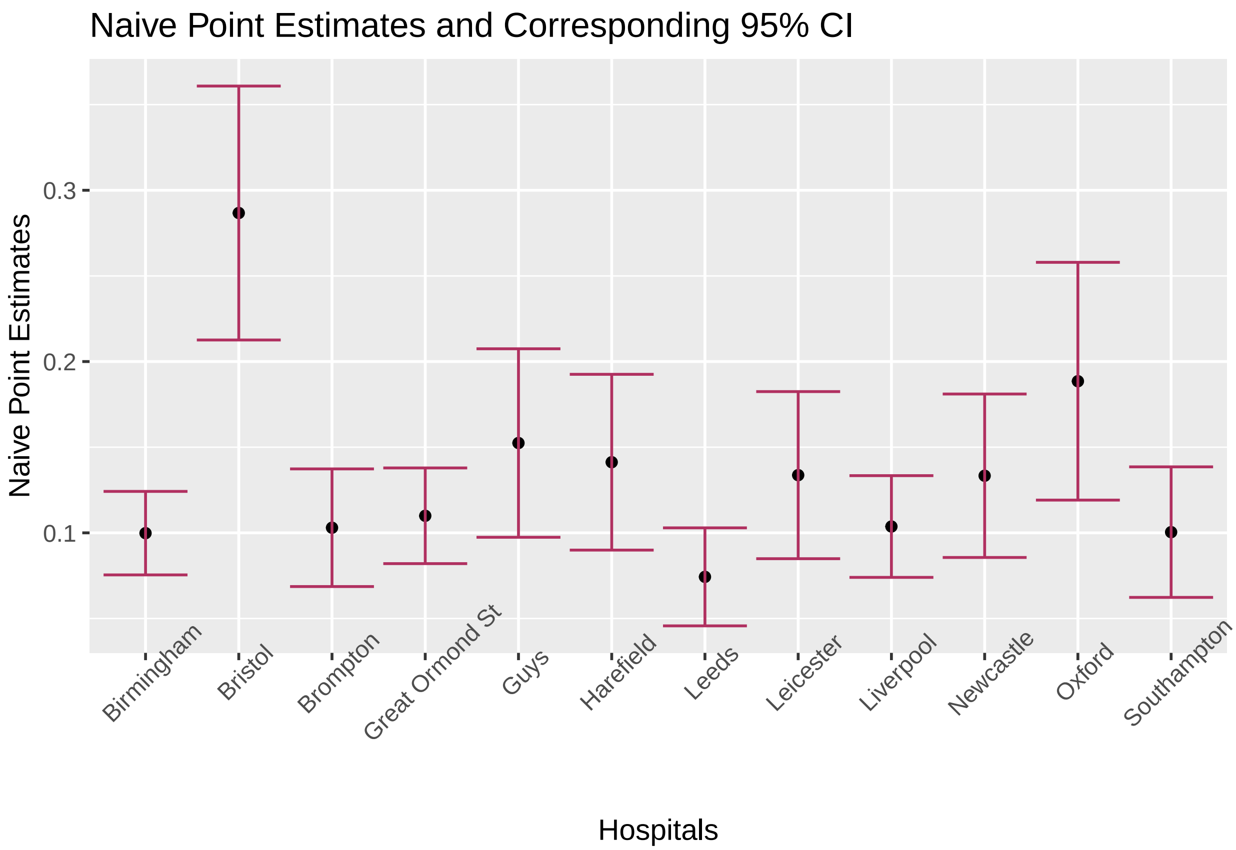


Please refer to the appendix if you would like to refer to code used. To generate my predictions, I used $\mu_0 = 0 \ \eta_0 = 100, \alpha_0 = \beta_0 = 0.01$ as my non-informative values for the parameters for my prior distributions. For each player, I took the average of the untransformed end of season average, and then transformed it to come up with a point estimate for the transformed end of season average. We see that for the first few players, the binomial likelihood results in greater estimates for transformed end of season average than the normal likelihood. In the mid range, the predictions are nearly identical, while for the last fewer players using the normal likelihood results in greater estimates for the transformed end of season average. In all cases, there us good agreement between the predicted values, but these estimates are not always good estimates for the true transformed end of season averages.

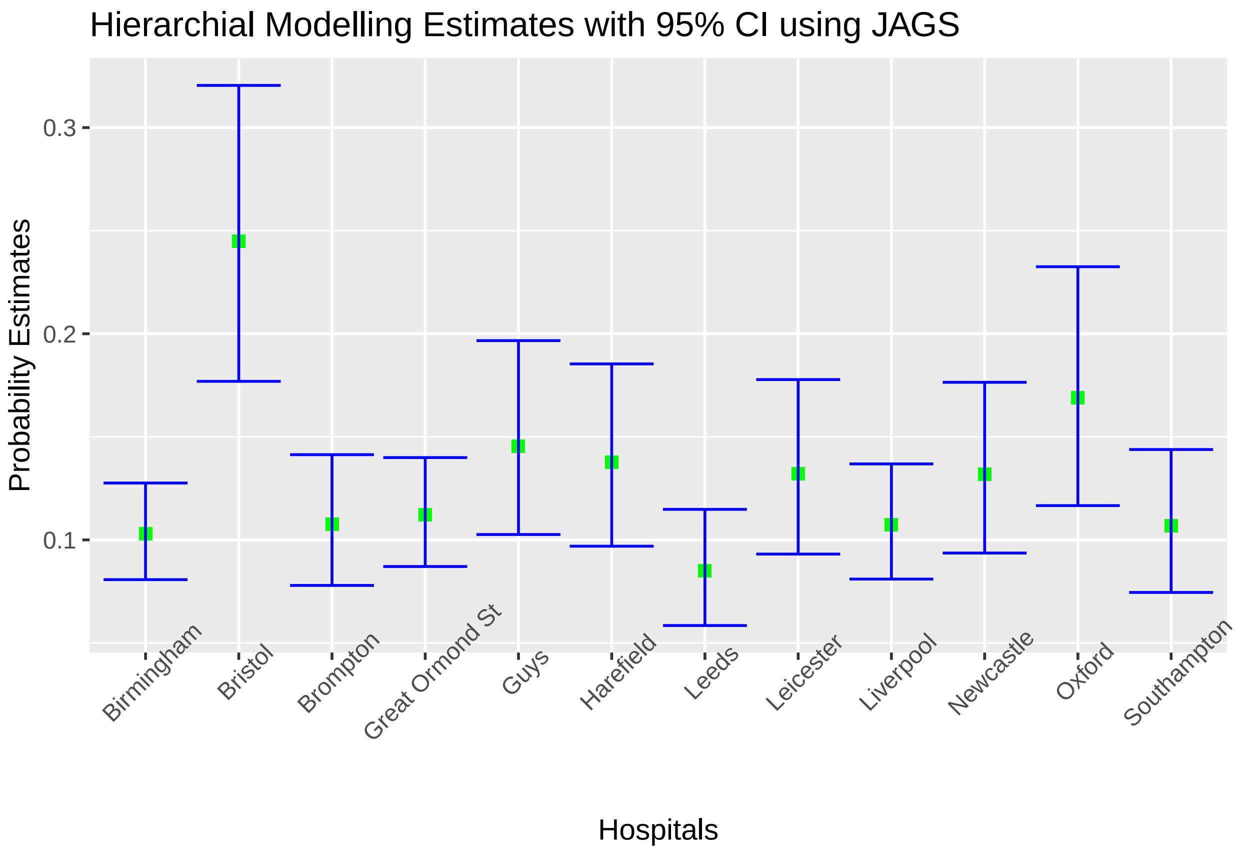
Question 5

Recall that for the $Binomial(n, p)$, the variance of the average is $\frac{p(1-p)}{n}$.

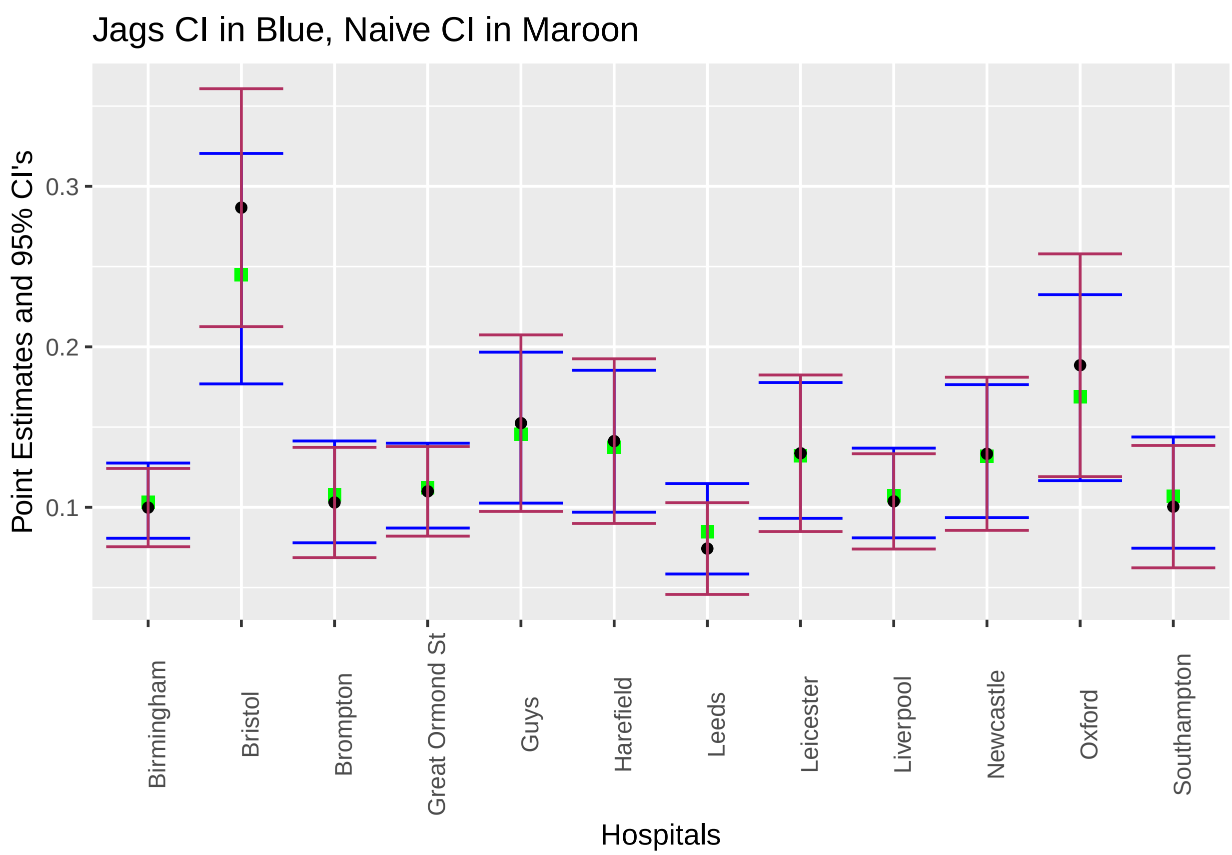
In this question, we denote θ_i as the probability of a heart operation resulting in death at hospital $i \forall i \in \{1, 2, ...12\}$, where m_i denotes the number of operations and X_i denotes the number of deaths in the i-th hospital. Modelling X_i as $\text{Bin}(m_i, \theta_i)$ independently across i gives the simple point estimate $\hat{\theta}_i = \frac{X_i}{m_i}$. We construct 95% C.I. using the formula $[\hat{\theta}_i - 1.96\sqrt{\frac{\hat{\theta}_i}{m_i(1-\theta_i)}}, \hat{\theta}_i + 1.96\sqrt{\frac{\hat{\theta}_i}{m_i(1-\theta_i)}}]$. Note that we use the normal approximation here.



Using Jags with the model specified in Question 4 we have the following estimates and CI:



For convenience, we plot both estimates and confidence intervals together:



The naive point estimates are represented as black dots (CI in maroon), while the hierarchial estimates are plotted as green squares (CI in blue). In most cases the two CI are fairly similar in width, with the hierarchial CIs slightly narrower on average. The point estimates are also in close agreement in most cases, except in the case of Bristol hospital. If I had to choose, I'd probably go with the hierarchial model approach since it has slightly narrower CI, but the approach should not make a big difference here regardless.

Question 6

Note that this answer references Page 41 of Bayesian Data Analysis.

We are given that $Y_{i1} = \mu + \delta + \epsilon_{i1}, Y_{i2} = \mu - \delta + \epsilon_{i2}, \epsilon_{ij} \sim^{iid} N(0, \sigma^2)$

Additonally, our priors are $\mu \sim N(50, 625), \delta \sim N(0, 625), \sigma^{-2} \sim Gamma(0.5, 50)$

Hence we express likelihoods as $Y_{i1}|\mu, \delta, \sigma \sim N(\mu + \delta, \sigma^2), Y_{i2}|\mu, \delta, \sigma \sim N(\mu - \delta, \sigma^2)$, meaning that our posterior is $(\prod_{i=1}^{31} exp(-\frac{1}{2\sigma^2}(Y_{i1} - \mu - \delta)^2))(\prod_{i=1}^{22}(exp(-\frac{1}{2\sigma^2}(Y_{i2} - \mu + \delta)^2)))(exp(-\frac{1}{2*625}(\mu - 50)^2))(exp(-\frac{1}{2*625}(\delta)^2)\sigma^{-2(.5)-1})(exp(50\sigma^{-2}))$

Next, we find our conditional distributions, remebering that they are proportional to our full posterior.

We note from the symmetry of the likelihood and prior that μ and δ will have the same family of conditional posterior. Hence it is sufficient to show detailed calulcations and manipulations for one. I do so for mu and then state the final result for δ, but the reverse could also have been done. With that clarified, we proceed with the conditional distribution for μ: $\pi(\mu|\delta, \sigma, Y) \propto exp(-\frac{1}{2\sigma^2}\sum_{i=1}^{31}((Y_{i1} - \delta) - \mu)^2 + \sum_{i=1}^{22}((Y_{i2} + \delta) - \mu)^2))exp(-\frac{1}{2*625}(\mu - 50)^2)$

Rather than look at each component of the scores separately, we combine some of them them where $s = Y + \mu$. This allows us to rewrite the conditional distribution as $\pi(\mu|\delta, \sigma, Y) \propto exp(-\frac{1}{2\sigma^2}\sum_{i=1}^{53}(s_i - \mu)^2)exp(-\frac{1}{2*625}(\mu - 50)^2)$

We then employ the "completing the square method", which gives us:

$$\pi(\mu|\delta, \sigma, Y) \sim N(\frac{\frac{\sum_{i=1}^{53} s_i/53}{\sigma^2/53} + \frac{50}{625}}{\frac{1}{\sigma^2/53} + \frac{1}{625}}, \frac{1}{\frac{1}{\sigma^2/53} + \frac{1}{625}})$$

We can now subsitute the expression for each student score back in. Note that the original substution was made so that we could apply the formula in BDA about completing the square. Hence we get the conditional distribution of μ to be : $\pi(\mu|\delta, \sigma, Y) \sim N(\frac{\frac{\sum_{i=1}^{31}(Y_{i1}-\delta) + \sum_{i=1}^{22}(Y_{i2}+\delta)}{\sigma^2} + \frac{50}{625}}{\frac{1}{\sigma^2/53} + \frac{1}{625}}, \frac{1}{\frac{1}{\sigma^2/53} + \frac{1}{625}})$

As mentioned before, we can apply the same approach to get the conditional distribution of δ. $\pi(\mu|\delta, \sigma, Y) \propto exp(-\frac{1}{2\sigma^2}\sum_{i=1}^{31}((Y_{i1} - \mu) - \delta)^2 + \sum_{i=1}^{22}((\mu - Y_{i2}) - \delta)^2)exp(-\frac{1}{2*625}(\delta)^2)$

Here we once again employ a substitution, mildly changing our definition of s to become $s = Y - \mu$ for school 1 and $s = \mu - Y$ for school 2. This allows us to simplify the expression and complete the square method. $\pi(\delta|\mu, \sigma, Y) \propto exp(-\frac{1}{2\sigma^2}\sum_{i=1}^{53}(\alpha_i - \delta)^2)exp(-\frac{1}{2*625}(\delta)^2)$

On completing the square, we obtain the simplified conditional distribution of δ

$$\pi(\delta|\mu, \sigma, Y) \sim N(\frac{\frac{\sum_{i=1}^{53} s_i/53}{\sigma^2/53} + \frac{0}{625}}{\frac{1}{\sigma^2/53} + \frac{1}{625}}, \frac{1}{\frac{1}{\sigma^2/53} + \frac{1}{625}})$$

We re-substitute the value of s_i for each student score, giving us $\pi(\delta|\mu, \sigma, Y) \sim N(\frac{\frac{\sum_{i=1}^{31}(Y_{i1}-\mu) + \sum_{i=1}^{22}(\mu - Y_{i2})}{\sigma^2}}{\frac{1}{\sigma^2/53} + \frac{1}{625}}, \frac{1}{\frac{1}{\sigma^2/53} + \frac{1}{625}})$

Note that since σ^2 has a Gamma prior, the conditional distribution will also be different. Specifically, we have:

$$\pi(\sigma^{-2}|\delta, \mu, Y) \propto \sigma^{-31-22} \cdot exp(-\frac{1}{2\sigma^2}(\sum_1^{31}(Y_{i1} - \mu - \delta)^2 + \sum_1^{22}(Y_{i2} - \mu + \delta)^2)) \cdot \sigma^{-2(0.5)-1}exp(50\sigma^{-2})$$

Combining exponential terms and removing unnecessary constants, we get

$$\pi(\sigma^{-2}|\delta, \mu, Y) \propto exp(-\frac{1}{\sigma^2}\frac{(\sum_1^{31}(Y_{i1}-\mu-\delta)^2 + \sum_1^{22}(Y_{i2}-\mu+\delta)^2 + 100)}{2})\sigma^{-55}$$

Looking at this expression, we recognise the distribution to be a gamma distrbution. Since gamma is a conjugate prior for distributions involving exponential likelihoods (eg. Poisson), this makes sense. The paramters are $\pi(\sigma^{-2}|\delta, \mu, Y) \sim Gamma(a = 27, b = \frac{(\sum_{i=1}^{31}(Y_{i1}-\mu-\delta)^2 + \sum_{i=1}^{22}(Y_{i2}-\mu+\delta)^2 + 100)}{2})$

We now perform Gibbs sampling, where each parameter is a block. The results are also compared to jags to verify our answers

The jags estimates are provided below, calculated by taking the average of all the samples.

```
##       mu
## 48.64943
```

```
##      del
## 2.156359
```

```
##    sigma
## 10.86538
```

Using my Gibbs sampler we find the following point estimates:

```
##       mu
## 48.65887
```

```
##      del
## 2.162441
```

```
##    sigma
## 10.86661
```

We see good agreement between values here. In order to find the probability $\delta > 0$, we use both the Gibbs sampler and jags. The first output is from jags, and the second is from Gibbs.

```
## [1] 0.923174
```

```
## [1] 0.923226
```

As before, there is good agreement. Finally in order to find probability that a student from the first school has a higher score than a student from the second we will use the likelihood functions for Y_{i1} and Y_{i2} for each sample and calculate the proportion of interest. As before, the jags output is first, followed by the Gibbs one.

```
## [1] 0.609692
```

```
## [1] 0.609868
```

As a final confirmation, the probabilities are quite similar to one another.

Appendix

```r
library(rjags)
# Sampling for Question 4. Following code from lecture. Define DataFrame
baseball = data.frame(
players = c("Clemente", "F Robinson", "F Howard", "Johnstone",
            "Berry", "Spencer", "Kessinger", "L Alvarado",
            "Santo", "Swoboda", "Unser", "Williams", "Scott",
            "Petrocelli", "E Rodriguez", "Campaneris", "Munson", "Alvis"),
hits = c(18, 17, 16, 15, 14, 14, 13, 12, 11,
         11, 10, 10, 10, 10, 10, 9, 8, 7),
atbats = rep(45, 18),
EoSaverage = c(.346, .298, .276, .222, .273, .270,
               .263, .210, .269, .230, .264, .256,
               .303, .264, .226, .286, .316, .200),
stringsAsFactors = FALSE
)

# Apply the transformation: 2arcsin(sqrt(p))
baseball$norm.data = 2*sqrt(45)*asin(sqrt(baseball$hits/45))
baseball$true.mean = 2*sqrt(45)*asin(sqrt(baseball$EoSaverage))

# Use jags to sample
n = nrow(baseball)
jags.model <- "
model{
  for(i in 1:n) {
  logit[i] ~ dnorm(mu, theta_precision)
  theta[i] = exp(logit[i])/(1+exp(logit[i]))
  y[i] ~ dbin(theta[i], 45)
}
mu ~ dnorm(mu0, mu_precision)
theta_precision ~ dgamma(alpha, beta)
}
"
jm = jags.model(textConnection(jags.model),
                data = list(y = baseball$hits, n = n, mu0 = 0,
                mu_precision = 1/(100^2), alpha = 0.01, beta = 0.01))
cs = coda.samples(jm, c("theta", "mu", "theta_precision"), 400000)
s = as.data.frame(cs[[1]])

# Compare to lecture output
jags.model <- "
model{
for(i in 1:n){
y[i] ~ dnorm(theta[i], 1)
theta[i] ~ dnorm(mu, prec.theta)
}
mu ~ dnorm(prior.mu.mean, prior.mu.prec)
prec.theta ~ dgamma(prior.prec.alpha, prior.prec.beta)
}
"
jm = jags.model(textConnection(jags.model),
```

```
        data = list(y = baseball$norm.data, n = nrow(baseball),
        prior.mu.mean = 0, prior.mu.prec = 1/(100^2),
        prior.prec.alpha = 0.01, prior.prec.beta = 0.01))
cs = coda.samples(jm, c("theta", "mu", "prec.theta"), 400000) #JAGS is very fast too
s_class = as.data.frame(cs[[1]])
# Plotting for Question 4
column_means = colMeans(s)[2:19]
transformed_jags = 2*sqrt(45)*asin(sqrt(column_means))
normal_approx = colMeans(s_class)[3:20]
plot(1:n, normal_approx , ylim=c(6, 9),col='red',
     pch = 16, cex = 0.7, xlab = "Player", ylab = "Transformed EoSA",
     main = "True Transformed EoSA and Different Predictions")
points(1:n, transformed_jags, col='green', pch = 16, cex = 0.7)
points(1:n, baseball$true.mean, col = "blue", pch = 16, cex = 0.7)
legend("top", legend=c("Using Normal Likelihood",
      "Using Bimonial Likelihood", "True Transformed EoSA"),
       col=c("red", "green", "blue"), lty=1, cex=0.8)

# End of Question 4
#Code for Question 5
#Creating DataFrame
library(ggplot2)
library(matrixStats)

hospitals = c("Bristol", "Leicester", "Leeds", "Oxford", "Guys",
        "Liverpool", "Southampton", "Great Ormond St",
        "Newcastle", "Harefield", "Birmingham", "Brompton")
deaths = c(41, 25, 24, 23, 25, 42, 24, 53, 26, 25, 58, 31)
operations = c(143, 187, 323, 122, 164, 405, 239, 482, 195, 177, 581, 301)
theta_estimates = deaths/operations

lower = theta_estimates - 1.96 * sqrt(theta_estimates * (1 - theta_estimates)/operations)
upper = theta_estimates + 1.96 * sqrt(theta_estimates * (1 - theta_estimates)/operations)
data = data.frame(hospitals, operations, deaths, theta_estimates, lower, upper)

#Referencing code from
#https://www.datanovia.com/en/blog/ggplot-axis-ticks-set-and-rotate-text-labels/
ggplot(data, aes(x = hospitals, y = theta_estimates)) + geom_point() + theme(axis.text.x=element_text(a
  ylab('Naive Point Estimates') + xlab("Hospitals") + ggtitle("Naive Point Estimates and Corresponding 
  geom_errorbar(aes(ymin = lower, ymax= upper), colour = "maroon")
n = nrow(data)
jags.model <- "
model{
  for(i in 1:n) {
  logit[i] ~ dnorm(mu, theta_precision)
  theta[i] = exp(logit[i])/(1+exp(logit[i]))
  y[i] ~ dbin(theta[i], m[i])
}
mu ~ dnorm(mu0, mu_precision)
theta_precision ~ dgamma(alpha, beta)
}
"
jm = jags.model(textConnection(jags.model),
```

```
                data = list(y = deaths, n = n, m = operations, mu0 = 0,
                mu_precision = 1/(100^2), alpha = 0.01, beta = 0.01))
cs = coda.samples(jm, c("theta", "mu", "theta_precision"), 400000)
s = as.data.frame(cs[[1]])
jags_estimates = s[, 2:13]
names(jags_estimates) = hospitals
jags_means = colMeans(jags_estimates)
CI = colQuantiles(as.matrix(jags_estimates),probs=c(.025,.975))
CI_jags = data.frame(CI)
names(CI_jags)=c('Lower','Upper')
jags_data = data.frame(jags_means, CI_jags)
jags_data$hospitals = row.names(jags_data)
ggplot(jags_data, aes(x=hospitals)) + geom_point(aes(y = jags_means), color = "green", pch = 15, cex = 
  geom_errorbar(aes(ymin = Lower, ymax = Upper), color = "blue") +
  theme(axis.text.x=element_text(angle = 45)) +
  ggtitle("Hierarchial Modelling Estimates with 95% CI using JAGS") +
  ylab('Probability Estimates') + xlab("Hospitals")
merged_data = merge(data ,jags_data)
ggplot(merged_data, aes(x = hospitals)) + geom_point(aes(y = jags_means), color = "green", pch = 15, ce
  theme(axis.text.x=element_text(angle=90))+
  ggtitle('Jags CI in Blue, Naive CI in Maroon')+
  ylab('Point Estimates and 95% CI\'s')+
  geom_point(aes(y = theta_estimates), col="black")+
  geom_errorbar(aes(ymin= lower, ymax= upper), color ="maroon") + xlab("Hospitals")

#Code for Question 6

school1 = c(52.11, 57.65, 66.44, 44.68, 40.57, 35.04, 50.71, 66.17, 39.43, 46.17, 58.76, 47.97, 39.18, 

school2 = c(57.57, 42.40, 41.41, 55.22, 43.90, 53.04, 49.00, 62.45, 53.78, 49.08, 40.25, 43.08, 52.43, 

theta = rep(0, (3))

#Initialize the theta vector:
theta[(1)] = 50
theta[(2)] = 0
theta[(3)] = 10
number_it = 500000
path = matrix(0, number_it, (3))
path[1,] = theta
#Update block depending on index
for(t in 2:number_it)
{
    j = sample(3)[1]

    if(j == 1)

    {
        delta_cur = theta[2]
        sigma_cur = theta[3]
        var = 1/( 1/((sigma_cur^2)/53) + 1/625)
        mu = ((sum(school1-delta_cur)+sum(school2+delta_cur))/(sigma_cur^2) + 50/625)*var
        sd = sqrt(var)
```

```
        path[t,1] = rnorm(1,mean=mu,sd=sd)
        path[t,2] = delta_cur
        path[t,3] = sigma_cur
    }

    if(j == 2)

    {
        mu_cur = theta[1]
        sigma_cur = theta[3]
        var = 1/( 1/((sigma_cur^2)/53) + 1/625)
        mu = ( (sum(school1-mu_cur)+sum(mu_cur-school2)) /(sigma_cur^2))*var
        sd = sqrt(var)
        path[t,1] = mu_cur
        path[t,2] = rnorm(1,mean = mu,sd = sd)
        path[t,3] = sigma_cur
    }

    if(j == 3)

    {
        mu_cur = theta[1]
        delta_cur = theta[2]
        alpha = 27
        beta = ( sum((school1-mu_cur-delta_cur)^2) + sum((school2-mu_cur+delta_cur)^2) + 100)/2
        path[t,1] = mu_cur
        path[t,2] = delta_cur
        path[t,3] = 1/sqrt(rgamma(1, alpha, beta))
    }
    theta = path[t,]
}
library(rjags)
jags.model <- "
model{
for (i in 1:n1){
y1[i] ~ dnorm(mu+del,prec.sigma)
}
for (i in 1:n2){
y2[i] ~ dnorm(mu-del,prec.sigma)
}
mu ~ dnorm(prior.mu.mean, prior.mu.prec)
del ~ dnorm(prior.del.mean,prior.del.prec)
eps ~ dnorm(prior.eps.mean,prec.sigma)
prec.sigma ~ dgamma(prior.prec.alpha, prior.prec.beta)
}
"
jm = jags.model(textConnection(jags.model), data = list(y1 = school1, y2=school2,
n2=length(school2), n1=length(school1), prior.mu.mean = 50, prior.mu.prec = 1/625,
prior.del.mean = 0, prior.del.prec=1/625, prior.eps.mean=0, prior.prec.alpha = .5,
prior.prec.beta = 50))

cs = coda.samples(jm, c("mu", "del", "prec.sigma"), 500000)
```

```r
s_jags = as.data.frame(cs[[1]])
s_jags$sigma = 1/sqrt(s_jags$prec.sigma)
colMeans(s_jags)[2]
colMeans(s_jags)[1]
colMeans(s_jags)[4]
gibbs = as.data.frame(path)
names(gibbs) = c('mu','del','sigma')
colMeans(gibbs)[1]
colMeans(gibbs)[2]
colMeans(gibbs)[3]
jags_del0 = mean(s_jags$del>0)
gibbs_del0 = mean(gibbs$del>0)
print(jags_del0)
print(gibbs_del0)
y1_jags=c()
y2_jags=c()
for (i in 1:nrow(s_jags)){
  y1_jags[i] = rnorm(1,s_jags[i,2]+s_jags[i,1],s_jags[i,4])
  y2_jags[i] = rnorm(1,s_jags[i,2]-s_jags[i,1],s_jags[i,4])
}
jags_prob = mean(y1_jags>y2_jags)
print(jags_prob)
y1_me=c()
y2_me=c()
for (i in 1:nrow(path)){
  y1_me[i] = rnorm(1,path[i,1]+path[i,2],path[i,3])
  y2_me[i] = rnorm(1,path[i,1]-path[i,2],path[i,3])
}
gibbs_prob = mean(y1_me>y2_me)
print(gibbs_prob)

labs = knitr::all_labels()
labs = labs[!labs %in% c("setup", "toc", "getlabels", "allcode")]
```