MA 578 HW3 Solutions

1 (BDA 3.8)

(a)

Let x_i and w_i be the number of bicycles in the residential streets with bike lanes and without bike lanes respectively. Then, conditional on n_i and m_i, the total number of vehicles in streets with and without bike lanes respectively, we can assume $x_i \mid \theta_y \overset{\text{ind}}{\sim} \mathsf{Binom}(n_i, \theta_y)$ and $w_i \mid \theta_z \overset{\text{ind}}{\sim} \mathsf{Binom}(m_i, \theta_z)$, and so $y_i = x_i/n_i$ and $z_i = w_i/m_i$.

(b)

We can take $\theta_y \sim \mathsf{Beta}(\alpha, \beta)$ and $\theta_y \sim \mathsf{Beta}(\alpha, \beta)$ with θ_y and θ_z independent. For an informative prior, say, Jeffreys prior, $\alpha = \beta = 1/2$.

(c) and (d)

Due to conjugacy,

$$\theta_y \mid y, n \sim \mathsf{Beta}\left(\alpha + \sum_{i=1}^{J_y} x_i, \beta + \sum_{i=1}^{J_y} n_i - x_i\right) \quad \text{and} \quad \theta_z \mid z, m \sim \mathsf{Beta}\left(\alpha + \sum_{i=1}^{J_z} w_i, \beta + \sum_{i=1}^{J_z} m_i - w_i\right).$$

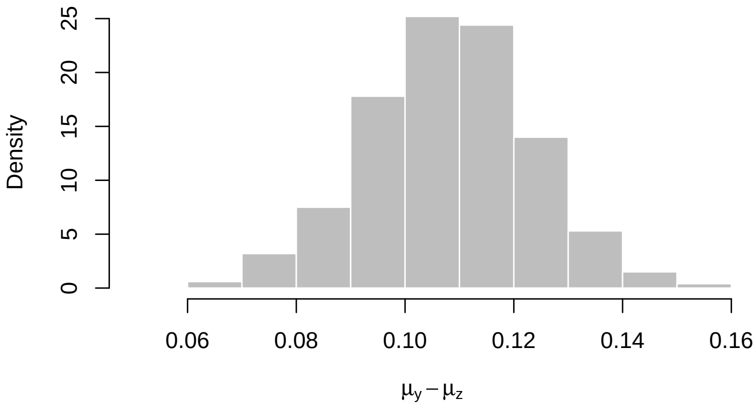
Since $\mu_y = \mathbb{E}[y_i \mid \theta_y] = \theta_y$ and $\mu_z = \mathbb{E}[z_i \mid \theta_z] = \theta_z$, we are interested in $\mu_y - \mu_z = \theta_y - \theta_z$.

```r
bicycle <- read.csv("data/bicycle.csv", comment="#")
bicycle <- bicycle[bicycle$streettype == "residential",]

y <- with(bicycle, bicycles[streettype == "residential" & bikeroute == "yes"])
ny <- with(bicycle, other[streettype == "residential" & bikeroute == "yes"])
ny <- ny + y
z <- with(bicycle, bicycles[streettype == "residential" & bikeroute == "no"])
nz <- with(bicycle, other[streettype == "residential" & bikeroute == "no"])
nz <- nz + z

# Jeffrey's prior
alpha_y <- beta_y <- .5
alpha_z <- beta_z <- .5

ns <- 1000
theta_y_s <- rbeta(ns, alpha_y + sum(y), beta_y + sum(ny - y))
theta_z_s <- rbeta(ns, alpha_z + sum(z), beta_z + sum(nz - z))
hist(theta_y_s - theta_z_s, xlab = expression(mu[y] - mu[z]),
    main = "", col = "gray", border = "white", freq = FALSE)
```




```r
summary(theta_y_s - theta_z_s)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.05687 0.09803 0.10860 0.10802 0.11787 0.15280
```

2 (BDA 3.9)

The joint prior $\mu, \sigma^2 \sim \textsf{N-Inv-}\chi^2(\mu_0, \kappa_0; \nu_0, \tau_0^2)$ is

$$\mathbb{P}(\mu, \sigma^2) \propto (\sigma^2)^{-\frac{\nu_0+1}{2}-1} \exp\left\{ -\frac{1}{2\sigma^2}\left[\nu_0\tau_0^2 + \kappa_0(\mu-\mu_0)^2\right]\right\}.$$

The likelihood is

$$\mathbb{P}(y \mid \mu, \sigma^2) \propto (\sigma^2)^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right\} = (\sigma^2)^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2\sigma^2}\left[s_y^2 + n(\bar{y}-\mu)^2\right]\right\},$$

with $s_y^2 = \sum_{i=1}^{n}(y_i - \bar{y})^2$.

The posterior is then

$$\begin{aligned}\mathbb{P}(\mu, \sigma^2 \mid y) &\propto \mathbb{P}(y \mid \mu, \sigma^2)\mathbb{P}(\mu, \sigma^2) \\ &\propto (\sigma^2)^{-\frac{n+\nu_0+1}{2}-1} \exp\left\{ -\frac{1}{2\sigma^2}\left[\nu_0\tau_0^2 + s_y^2 + \kappa_0(\mu-\mu_0)^2 + n(\bar{y}-\mu)^2\right]\right\}.\end{aligned}$$

But, with $\mu_n = (n\bar{y} + \kappa_0\mu_0)/(n+\kappa_0)$,

$$\kappa_0(\mu-\mu_0)^2 + n(\bar{y}-\mu)^2 = (n+\kappa_0)(\mu-\mu_n)^2 + \frac{n\kappa_0}{n+\kappa_0}(\bar{y}-\mu_0)^2,$$

and so

$$\mathbb{P}(\mu, \sigma^2 \mid y) \propto (\sigma^2)^{-\frac{n+\nu_0+1}{2}-1} \exp\left\{ -\frac{1}{2\sigma^2}\left[\nu_0\tau_0^2 + s_y^2 + \frac{n\kappa_0}{n+\kappa_0}(\bar{y}-\mu_0)^2 + (n+\kappa_0)(\mu-\mu_n)^2\right]\right\},$$

which can be seen to be $\textsf{N-Inv-}\chi^2(\mu_n, \kappa_n; \nu_n, \tau_n^2)$ with $\nu_n = \nu_0 + n$, $\kappa_n = \kappa_0 + n$, and

$$\tau_n^2 = \frac{\nu_0\tau_0^2 + s_y^2 + \frac{n\kappa_0}{n+\kappa_0}(\bar{y}-\mu_0)^2}{\nu_0 + n}.$$

3 (BDA 3.12)

(a)

The two most common choices for a non-informative prior on α and β are a flat prior, $\mathbb{P}(\alpha, \beta) \propto 1$, and Jeffreys prior, $\mathbb{P}(\alpha, \beta) \propto |I(\alpha, \beta)|^{1/2}$, where I is Fisher information,

$$I(\alpha, \beta) = \mathbb{E}_{y \mid \alpha, \beta}\left[\frac{\partial^2 \ell}{\partial(\alpha, \beta)^2}\right] = \begin{bmatrix} \sum_i \frac{1}{\alpha+\beta t_i} & \sum_i \frac{t_i}{\alpha+\beta t_i} \\ \sum_i \frac{t_i}{\alpha+\beta t_i} & \sum_i \frac{t_i^2}{\alpha+\beta t_i} \end{bmatrix},$$

that is,

$$\mathbb{P}(\alpha, \beta) \propto \left\{\sum_i \frac{1}{\alpha + \beta t_i} \sum_i \frac{t_i^2}{\alpha + \beta t_i} - \left(\sum_i \frac{t_i}{\alpha + \beta t_i}\right)^2\right\}^{\frac{1}{2}}.$$

Note that we also need to guarantee that $\alpha + \beta t_i > 0$ for every t_i, and so $\alpha + \min_i\{\beta t_i\} > 0$ almost surely under the prior.

(c)

Since the log-likelihood (up to a constant) is

$$\ell(\alpha, \beta) = \sum_{i=1}^n y_i \log(\alpha + \beta t_i) - n(\alpha + \beta \bar{t}),$$

we need the whole data vector as statistics. The posterior is

$$\mathbb{P}(\alpha, \beta \,|\, y) \propto \prod_i (\alpha + \beta t_i)^{y_i} \exp\{-n(\alpha + \beta \bar{t})\} \mathbb{P}(\alpha, \beta).$$

(d)

Assuming that the prior is improper, posterior impropriety can only arise from the likelihood. Since $\alpha+\beta t_i > 0$ are bounded away from negative values, we just need to be concerned with posterior mass in extremely large values of the parameters. But then $\exp\{-n(\alpha + \beta \bar{t})\}$ goes faster to zero than the power term on y, and so the likelihood goes to zero when $\alpha + \beta t_i$ is large, guaranteeing propriety. In what follows we check propriety numerically.

(e)

```r
airline <- read.csv("data/airline.csv", comment = "#")
base_year <- min(airline$year)
t <- airline$year - base_year
y <- airline$fatal

summary(glm(y ~ t, family = poisson(link = "identity")))
```

```
Call:
glm(formula = y ~ t, family = poisson(link = "identity"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.0292  -0.4847  -0.3624   0.6971   1.1099

Coefficients:
```

```
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  28.0720     3.0027   9.349   <2e-16
t            -0.9493     0.5343  -1.777   0.0756

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 8.3893  on 9  degrees of freedom
Residual deviance: 5.3386  on 8  degrees of freedom
AIC: 59.308

Number of Fisher Scoring iterations: 4
```

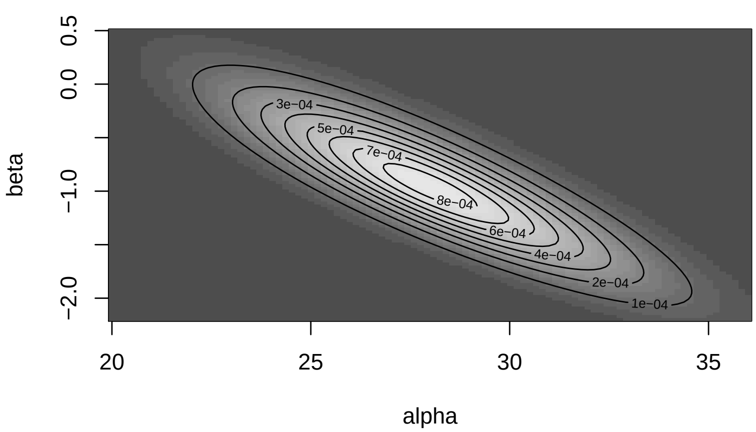
(f)

We use Jeffreys prior, but the results from a flat prior are not much different.

```r
m <- 100
alpha <- seq(20, 36, length = m)
beta <- seq(-2.2, 0.5, length = m)

lprior <- function (a, b) {
  ieta <- 1 / (a + b * t)
  .5 * log(sum(ieta) * sum(t ^ 2 * ieta) - sum(t * ieta) ^ 2)
}
lhood <- function (a, b) {
  eta <- a + b * t
  sum(y * log(eta) - eta)
}

lj <- matrix(nrow = m, ncol = m)
for (ia in 1:m) {
  for (ib in 1:m) {
    a <- alpha[ia]; b <- beta[ib]
    lj[ia, ib] <- lprior(a, b) + lhood(a, b)
  }
}
pab <- exp(lj - log(sum(exp(lj))))
image(alpha, beta, pab, col = gray.colors(28))
contour(alpha, beta, pab, add = TRUE)
```

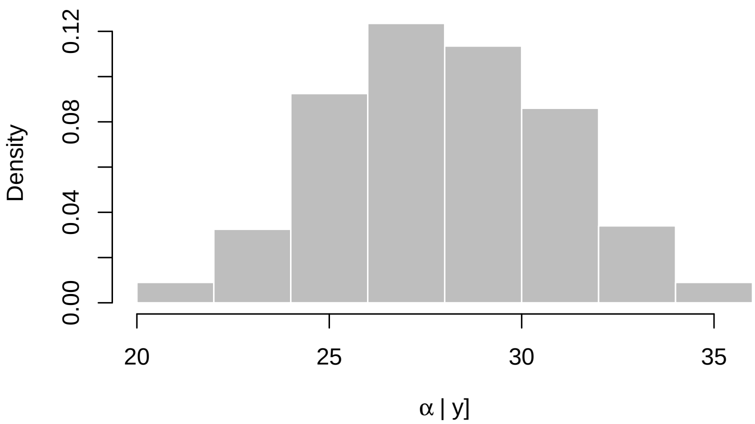


```r
ns <- 1000
is <- arrayInd(sample.int(m * m, ns, replace = TRUE, prob = pab), dim(pab))
alpha_s <- alpha[is[,1]]; beta_s <- beta[is[,2]]

summary(alpha_s)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  20.00   25.82   27.92   27.98   30.02   36.00 
```

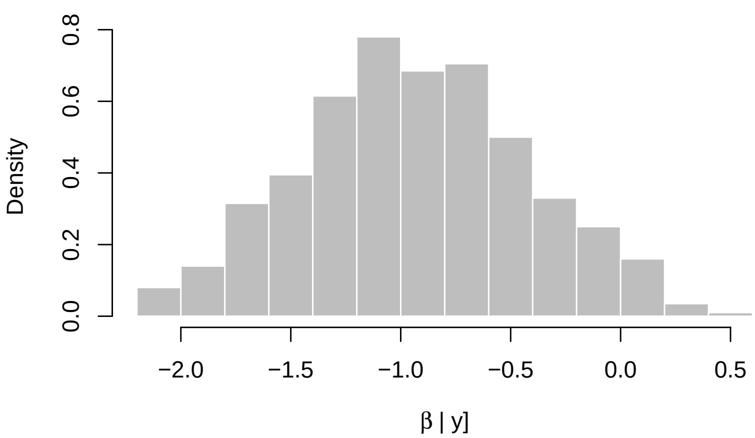
```r
hist(alpha_s, col = "gray", border = "white", freq = FALSE,
     xlab = expression(alpha~"| y]"), main = "")
```




```r
summary(beta_s)
```

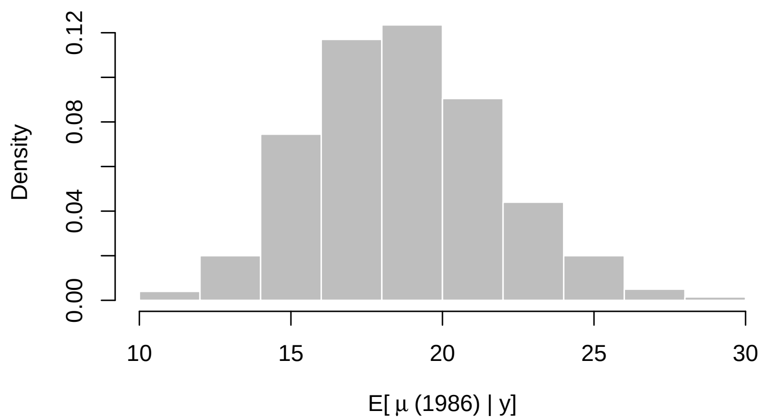
```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
-2.2000 -1.2727 -0.9455 -0.9283 -0.5909  0.5000 
```

```r
hist(beta_s, col = "gray", border = "white", freq = FALSE,
     xlab = expression(beta~"| y]"), main = "")
```



(g)

```r
ttilde <- 1986 - base_year
hist(alpha_s + beta_s * ttilde,
     col = "gray", border = "white", freq = FALSE,
     xlab = expression('E['~mu~'(1986) | y]'), main = "")
```




```r
summary(alpha_s + beta_s * ttilde)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  10.60   16.55   18.55   18.70   20.65   29.20 
```

(h)

```r
ytilde_s <- rpois(ns, alpha_s + beta_s * ttilde)
quantile(ytilde_s, c(.025, .975))
```

```
 2.5% 97.5% 
    9    30 
```

4

First let us recall that if $\sigma^2 \sim \mathsf{Inv}\text{-}\chi^2(\nu_0, \sigma_0^2)$ then

$$\mathbb{P}(\sigma^2) = \frac{(\nu_0\sigma_0^2/2)^{\nu_0/2}}{\Gamma(\nu_0/2)}(\sigma^2)^{-\left(\frac{\nu_0}{2}+1\right)} \exp\left\{ -\frac{\nu_0\sigma_0^2}{2\sigma^2}\right\}.$$

(a)

The conditional prior on σ^2 is

$$\begin{aligned}\mathbb{P}(\sigma^2 \,|\, \mu) &\propto (\sigma^2)^{-\frac{1}{2}} \exp\left\{ -\frac{\kappa_0}{2\sigma^2}(\mu-\mu_0)^2\right\}(\sigma^2)^{-\left(\frac{\nu_0}{2}+1\right)} \exp\left\{ -\frac{\nu_0\sigma_0^2}{2\sigma^2}\right\} \\ &= (\sigma^2)^{-\left(\frac{\nu_0+1}{2}+1\right)} \exp\left\{ -\frac{\kappa_0(\mu-\mu_0)^2+\nu_0\sigma_0^2}{2\sigma^2}\right\},\end{aligned}$$

which, with $\nu_\mu = \nu_0 + 1$ and $\nu_\mu\sigma_\mu^2 = \kappa_0(\mu-\mu_0)^2 + \nu_0\sigma_0^2$, can be seen to be $\mathsf{Inv}\text{-}\chi^2(\nu_\mu, \sigma_\mu^2)$.

(b)

The conditional joint is

$$\begin{aligned}\mathbb{P}(X, \sigma^2 \,|\, \mu) &= \mathbb{P}(X \,|\, \mu, \sigma^2)\mathbb{P}(\sigma^2 \,|\, \mu) \\ &\propto (\sigma^2)^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2\sigma^2}(X-\mu\mathbb{1}_n)^\top(X-\mu\mathbb{1}_n)\right\}(\sigma^2)^{-\frac{\nu_\mu}{2}+1} \exp\left\{ -\frac{\nu_\mu\sigma_\mu^2}{2\sigma^2}\right\} \\ &= (\sigma^2)^{-\frac{n+\nu_\mu}{2}+1} \exp\left\{ -\frac{1}{2\sigma^2}\Big[(X-\mu\mathbb{1}_n)^\top(X-\mu\mathbb{1}_n)+\nu_\mu\sigma_\mu^2\Big]\right\}.\end{aligned}$$

We can recognize the kernel of $\mathsf{Inv}\text{-}\chi^2(\nu_{X,\mu}, \sigma^2_{X,\mu})$ with $\nu_{X,\mu} = n + \nu_\mu$ and $\nu_{X,\mu}\sigma^2_{X,\mu} = (X - \mu\mathbb{1}_n)^\top(X - \mu\mathbb{1}_n) + \nu_\mu\sigma^2_\mu$ and so, after marginalizing out σ^2 we are left with the inverse of the normalizing constant:

$$\mathbb{P}(X \mid \mu) \propto \Gamma(\nu_{X,\mu}/2)\left[\frac{\nu_{X,\mu}\sigma^2_{X,\mu}}{2}\right]^{-\nu_{X,\mu}/2} \propto \left[1 + \frac{(X - \mu\mathbb{1}_n)^\top(X - \mu\mathbb{1}_n)}{\nu_\mu\sigma^2_\mu}\right]^{-\frac{n+\nu_\mu}{2}}.$$

(c)

From lecture we know that

$$\mathbb{P}(X, \mu, \sigma^2) \propto (\sigma^2)^{-\left(\frac{\nu_n+1}{2}+1\right)} \exp\left\{ -\frac{1}{2\sigma^2}\left[\nu_n\sigma^2_n + \kappa_n(\mu - \mu_n)^2\right]\right\}$$

which resembles a $\mathsf{Inv}\text{-}\chi^2(\tilde{\nu}, \tilde{\sigma}^2)$ with $\tilde{\nu} = \nu_n + 1$ and $\tilde{\nu}\tilde{\sigma}^2 = \nu_n\sigma^2_n + \kappa_n(\mu - \mu_n)^2$. After marginalizing σ^2 again, we have

$$\mathbb{P}(\mu \mid X) \propto \Gamma(\tilde{\nu}/2)\left[\frac{\tilde{\nu}\tilde{\sigma}^2}{2}\right]^{-\tilde{\nu}/2} \propto \left[1 + \frac{\kappa_n(\mu - \mu_n)^2}{\nu_n\sigma^2_n}\right]^{-\frac{\nu_n+1}{2}}.$$