

Coursework 5: STAT 570

Philip Pham

November 2, 2018

1. Consider the data given in Table 1, which are a simplified version of those reported in Breslow and Day (1980). These data arose from a case-control study that was carried out to investigate the relationship between esophageal cancer and various risk factors. Disease status is denoted Y with $Y = 0$ and $Y = 1$ corresponding to without/with disease and alcohol consumption is represented by X with $X = 0$ and $X = 1$ denoting less than 80g and greater than or equal to 80g on average per day. Let the probabilities of high alcohol consumption in the cases and controls be denoted

$$p_1 = \mathbb{P}(X = 1 \mid Y = 1) \text{ and } p_2 = \mathbb{P}(X = 1 \mid Y = 0), \quad (1)$$

respectively. Further, let X_1 be the number exposed from n_1 cases and X_2 be the number exposed from n_2 controls. Suppose $X_i \mid p_i \sim \text{Binomial}(n_i, p_i)$ in the case ($i = 1$) and control ($i = 2$) groups.

- (a) Of particular interest in studies such as this is the odds ratio defined by

$$\theta = \frac{\mathbb{P}(Y = 1 \mid X = 1) / \mathbb{P}(Y = 0 \mid X = 1)}{\mathbb{P}(Y = 1 \mid X = 0) / \mathbb{P}(Y = 0 \mid X = 0)}. \quad (2)$$

Show that the odds ratio is equal to

$$\theta = \frac{\mathbb{P}(X = 1 \mid Y = 1) / \mathbb{P}(X = 0 \mid Y = 1)}{\mathbb{P}(X = 1 \mid Y = 0) / \mathbb{P}(X = 0 \mid Y = 0)} = \frac{p_1 / (1 - p_1)}{p_2 / (1 - p_2)}. \quad (3)$$

Solution: We have that

$$\mathbb{P}(Y = y \mid X = x) = \frac{\mathbb{P}(X = x \mid Y = y) \mathbb{P}(Y = y)}{\mathbb{P}(X = x)} \quad (4)$$

by Bayes' rule. Applying Equation 4 to Equation 2, we get

$$\theta = \frac{[\mathbb{P}(X = 1 \mid Y = 1) \mathbb{P}(Y = 1)] / [\mathbb{P}(X = 0 \mid Y = 1) \mathbb{P}(Y = 0)]}{[\mathbb{P}(X = 0 \mid Y = 1) \mathbb{P}(Y = 1)] / [\mathbb{P}(X = 0 \mid Y = 0) \mathbb{P}(Y = 0)]}. \quad (5)$$

The $\mathbb{P}(Y = y)$ factors cancel and we obtain the first part of Equation 3.

Using Equation 1, we substitute to obtain the second part of Equation 3.

	$X = 0$	$X = 1$	
$Y = 1$	104	96	200
$Y = 0$	666	109	775

Table 1: Case-control data: $Y = 1$ corresponds to the event of esophageal cancer, and $X = 1$ exposure to greater than 80g of alcohol per day. There are 200 cases and 775 controls.

- (b) Obtain the MLE and a 90% confidence interval for θ , for the data of Table 1.

Solution: The likelihood and log-likelihood functions are

$$\begin{aligned} L(p_1, p_2) &= \binom{n_1}{x_1} p_1^{x_1} (1-p_1)^{n_1-x_1} + \binom{n_2}{x_2} p_2^{x_2} (1-p_2)^{n_2-x_2} \\ l(p_1, p_2) &= \log L(p_1, p_2) \\ &= \sum_{i=1}^2 \left[\log \binom{n_i}{x_i} + x_i \log p_i + (n_i - x_i) \log (1-p_i) \right], \end{aligned} \quad (6)$$

so the score function is

$$S(p_1, p_2) = \nabla \log L(p_1, p_2) = \begin{pmatrix} \frac{x_1 - n_1 p_1}{p_1(1-p_1)} \\ \frac{x_2 - n_2 p_2}{p_2(1-p_2)} \end{pmatrix} \quad (7)$$

Thus, the Fisher information is

$$I(p_1, p_2) = \text{mathbb{E}}[S(p_1, p_2) S(p_1, p_2)^\top] = \begin{pmatrix} \frac{n_1}{p_1(1-p_1)} & 0 \\ 0 & \frac{n_2}{p_2(1-p_2)} \end{pmatrix}. \quad (8)$$

From Equation 7, we can solve $S(\hat{p}_1, \hat{p}_2) = \mathbf{0}$ to get the MLEs $\hat{p}_1 = x_1/n_1$ and $\hat{p}_2 = x_2/n_2$. Since the MLE is invariant to reparameterization, we have the MLE for θ :

$$\hat{\theta} = \frac{\hat{p}_1 / (1 - \hat{p}_1)}{\hat{p}_2 / (1 - \hat{p}_2)} = \frac{1992}{1417} \approx 5.640. \quad (9)$$

We estimate the confidence interval for $\log \hat{\theta}$ which works since \log is a monotonic transform. Using the delta method and Equation 8, we have that

$$\begin{aligned} \text{Var}(\log \hat{\theta}) &\approx (\nabla \log \hat{\theta})^\top (I(\hat{p}_1, \hat{p}_2))^{-1} (\nabla \log \hat{\theta}) \\ &= \begin{pmatrix} \frac{1}{\hat{p}_1(1-\hat{p}_1)} & \frac{1}{\hat{p}_2(1-\hat{p}_2)} \end{pmatrix} \begin{pmatrix} \frac{\hat{p}_1(1-\hat{p}_1)}{n_1} & 0 \\ 0 & \frac{\hat{p}_2(1-\hat{p}_2)}{n_2} \end{pmatrix} \begin{pmatrix} \frac{1}{\hat{p}_1(1-\hat{p}_1)} \\ \frac{1}{\hat{p}_2(1-\hat{p}_2)} \end{pmatrix} \\ &= \frac{1}{n_1 \hat{p}_1 (1 - \hat{p}_1)} + \frac{1}{n_2 \hat{p}_2 (1 - \hat{p}_2)} \\ &= \frac{1}{n_1 \hat{p}_1} + \frac{1}{n_1 (1 - \hat{p}_1)} + \frac{1}{n_2 \hat{p}_2} + \frac{1}{n_2 (1 - \hat{p}_2)}. \end{aligned} \quad (10)$$

Numerically, this is $\text{Var}(\log \hat{\theta}) \approx 0.0307$.

The 90% confidence interval for $\log \hat{\theta}$ is approximately

$$\left(\log \hat{\theta} - \Phi^{-1}(0.95) \sqrt{\text{Var}(\log \hat{\theta})}, \log \hat{\theta} + \Phi^{-1}(0.95) \sqrt{\text{Var}(\log \hat{\theta})} \right), \quad (11)$$

which is about (1.441, 2.018). Taking the exponent of both sides, we have a 90% confidence interval for $\hat{\theta}$ of $(4.228, 7.524)$.

- (c) We now consider a Bayesian analysis. Assume that the prior distribution for p_i is the beta distribution $\text{Beta}(a, b)$ for $i = 1, 2$. Show that the posterior distribution $p_i \mid x_i$ is given by the beta distribution $\text{Beta}(a + x_i, b + n_i - x_i)$, $i = 1, 2$.

Solution: From Equation 6, we have that the posterior:

$$\begin{aligned} p(p_i | X_i = x_i) &\propto \mathbb{P}(X_i = x_i | p_i) p(p_i) \\ &\propto p_i^{x_i+a-1} (1-p_i)^{n_i-x_i+b-1}. \end{aligned}$$

Integration from 0 to 1, we have the beta function, so

$$p(p_i | X_i = x_i) = \frac{\Gamma(a+x_i+b+n_i-x_i)}{\Gamma(a+x_i)\Gamma(b+n_i-x_i)} p_i^{a+x_i-1} (1-p_i)^{b+n_i-x_i-1}, \quad (12)$$

which is the Beta($a+x_i, b+n_i-x_i$) distribution.

- (d) Consider the case $a = b = 1$. Obtain expressions for the posterior mean, mode, and standard deviation. Evaluate these posterior summaries for the data of Table 1. Report 90% posterior credible intervals for p_1 and p_2 .

Solution: For $a = b = 1$, we have that $p_1 | x_1 \sim \text{Beta}(97, 105)$ and $p_2 | x_2 \sim \text{Beta}(110, 667)$.

For the posterior means, we have that $\mathbb{E}[p_1 | x_1] = 97/202$ and $\mathbb{E}[p_2 | x_2] = 110/777$.

The mode of a Beta(α, β) distributed random variable is $\frac{\alpha-1}{\alpha+\beta-2}$. So, for the posterior modes, we have that $\text{mode}(p_1 | x_1) = 12/25$ and $\text{mode}(p_2 | x_2) = 109/775$.

The variance of a Beta(α, β) distributed random variable is $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$. For $p_1 | x_1$ and $p_2 | x_2$, we have standard errors:

$$\begin{aligned} \sigma_{p_1|x_1} &= \frac{1}{202} \sqrt{\frac{10185}{203}} \approx 0.0351 \\ \sigma_{p_2|x_2} &= \frac{1}{777} \sqrt{\frac{36685}{389}} \approx 0.0125. \end{aligned}$$

For the 90% credible interval, I choose l and u such that $\mathbb{P}([l, u]) = 0.9$, $\mathbb{P}((-\infty, l)) = 0.05$ and $\mathbb{P}((u, \infty)) = 0.05$. This is called the *equal-tailed interval*.

For $p_1 | x_1$, the interval is $[0.4226, 0.5380]$. For $p_2 | x_2$, the interval is $[0.1215, 0.1626]$ This is computed numerically with `scipy.stats.beta.interval` in `case_control.ipynb`.

- (e) Obtain the asymptotic form of the posterior distribution and obtain 90% credible intervals for p_1 and p_2 . Compare this interval with the exact calculation of the previous part.

Solution:

2. (a) Consider the likelihood, $\hat{\theta} | \theta \sim \mathcal{N}(\theta, V)$ and the prior $\theta \sim \mathcal{N}(0, W)$ with V and W known. Show that $\theta | \hat{\theta} \sim \mathcal{N}(r\hat{\theta}, rV)$, where $r = W/(V+W)$.

Solution: This result follows from the conjugacy of the normal distribution with

itself:

$$\begin{aligned}
p(\theta | \hat{\theta}) &\propto p(\hat{\theta} | \theta) p(\theta) \\
&\propto \exp\left(-\frac{1}{2V}(\hat{\theta} - \theta)^2 - \frac{1}{2W}\theta^2\right) \\
&\propto \exp\left(-\frac{V+W}{2(VW)}\left(\frac{W}{V+W}\hat{\theta}^2 - 2\frac{W}{V+W}\hat{\theta}\theta + \theta^2\right)\right) \\
&\propto \exp\left(-\frac{V+W}{2(VW)}\left(\theta - \frac{W}{V+W}\hat{\theta}\right)^2\right) = \exp\left(-\frac{1}{2(rV)}(\theta - r\hat{\theta})^2\right)
\end{aligned}$$

after completing the square. We recognize this distribution as being part of the normal family, which gives us the result.

- (b) Suppose we wish to compare the models $M_0: \theta = 0$ versus $M_1: \theta \neq 0$. Show that the Bayes factor is given by

$$\frac{p(\hat{\theta} | M_0)}{p(\hat{\theta} | M_1)} = \frac{1}{\sqrt{1-r}} \exp\left(-\frac{Z^2}{2}r\right), \quad (13)$$

where $Z = \hat{\theta}/\sqrt{V}$.

Solution: We have that

$$\begin{aligned}
p(\hat{\theta} | M_0) &= p(\hat{\theta} | \theta = 0) = \frac{1}{\sqrt{2\pi V}} \exp\left(-\frac{1}{2V}\hat{\theta}^2\right) \\
p(\hat{\theta} | M_1) &= \int_{-\infty}^{\infty} p(\hat{\theta} | \theta) p(\theta) d\theta \\
&= \frac{1}{\sqrt{2\pi(V+W)}} \exp\left(-\frac{1}{2(V+W)}\hat{\theta}^2\right)
\end{aligned}$$

after completing the square. Substituting into the left-hand side of Equation 13, we obtain

$$\frac{p(\hat{\theta} | M_0)}{p(\hat{\theta} | M_1)} = \sqrt{\frac{V+W}{V}} \exp\left(-\frac{1}{2} \cdot \frac{W}{V+W} \cdot \frac{\hat{\theta}^2}{V}\right) = \frac{1}{\sqrt{1-r}} \exp\left(-\frac{Z^2}{2}r\right)$$

as desired.

- (c) Suppose we have a prior probability $\pi_1 = \mathbb{P}(M_1)$ of model M_1 being true. Write down an expression for the posterior probability $\mathbb{P}(M_1 | \hat{\theta})$ in terms of the Bayes factor.

Solution: Let K be the Bayes factor. By applying Bayes' rule, we have that

$$\begin{aligned}
\mathbb{P}(M_1 | \hat{\theta}) &= \frac{\mathbb{P}(\hat{\theta} | M_1) \mathbb{P}(M_1)}{\mathbb{P}(\hat{\theta} | M_0) \mathbb{P}(M_0) + \mathbb{P}(\hat{\theta} | M_1) \mathbb{P}(M_1)} \\
&= \frac{K^{-1} \mathbb{P}(\hat{\theta} | M_0) \pi_1}{\mathbb{P}(\hat{\theta} | M_0) (1 - \pi_1) + K^{-1} \mathbb{P}(\hat{\theta} | M_0) \pi_1} \\
&= \frac{K^{-1} \pi_1}{(1 - \pi_1) + K^{-1} \pi_1} = \frac{\pi_1}{K(1 - \pi_1) + \pi_1}.
\end{aligned}$$

- (d) Now suppose we have summaries from two studies, θ_j , V_j , $j = 1, 2$. Assuming, $\theta_j \mid \theta \sim \mathcal{N}(\theta, V_j)$ and the prior $\theta \sim \mathcal{N}(0, W)$, derive the posterior $p(\theta \mid \theta_1, \theta_2)$.

Solution: We have

$$\begin{aligned} p(\theta \mid \theta_1, \theta_2) &\propto p(\theta_2 \mid \theta_1, \theta) p(\theta_1 \mid \theta) p(\theta) = p(\theta_2 \mid \theta) p(\theta_1 \mid \theta) p(\theta) \\ &\propto \exp\left(-\frac{1}{2V_2}(\theta_2 - \theta)^2\right) \exp\left(-\frac{V_1 + W}{2(V_1W)}\left(\theta - \frac{W}{V_1 + W}\theta_1\right)^2\right) \\ &\propto \exp\left(-\frac{V_1V_2 + V_1W + V_2W}{2(V_1V_2W)}\left(\theta - \frac{V_2W\theta_1 + V_1W\theta_2}{V_1V_2 + V_1W + V_2W}\right)^2\right), \end{aligned}$$

after repeatedly completing the square and dropping factors that don't depend on θ .

Thus, we have that

$$\theta \mid \theta_1, \theta_2 \sim \mathcal{N}\left(\frac{V_2W\theta_1 + V_1W\theta_2}{V_1V_2 + V_1W + V_2W}, \frac{V_1V_2W}{V_1V_2 + V_1W + V_2W}\right). \quad (14)$$

- (e) Derive the Bayes factor

$$\frac{p(\theta_1, \theta_2 \mid M_0)}{p(\theta_1, \theta_2 \mid M_1)}, \quad (15)$$

again comparing the models M_0 : $\theta = 0$ versus M_1 : $\theta \neq 0$.

Solution: (θ_1, θ_2) have a bivariate normal distribution. Under M_0 , we have that

$$\begin{aligned} p(\theta_1, \theta_2 \mid M_0) &= p(\theta_1, \theta_2 \mid \theta = 0) \\ &= \frac{1}{2\pi\sqrt{V_1V_2}} \exp\left(-\frac{1}{2}\begin{pmatrix}\theta_1 & \theta_2\end{pmatrix}\begin{pmatrix}\frac{1}{V_1} & 0 \\ 0 & \frac{1}{V_2}\end{pmatrix}\begin{pmatrix}\theta_1 \\ \theta_2\end{pmatrix}\right). \end{aligned} \quad (16)$$

Under M_1 , we have that

$$p(\theta_1, \theta_2 \mid M_1) = \int_{-\infty}^{\infty} p(\theta_1, \theta_2 \mid \theta) p(\theta) d\theta. \quad (17)$$

We can consider θ as having the improper prior $\mathcal{N}\left(\mathbf{0}, \begin{pmatrix} W & W \\ W & W \end{pmatrix}\right)$, which results in

$$\theta_1, \theta_2 \mid M_1 \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} V_1 + W & W \\ W & V_2 + W \end{pmatrix}\right) \quad (18)$$

by conjugacy of the multivariate normal distribution.

The Bayes factor can then be computed:

$$\sqrt{\frac{V_1V_2 + V_1W + V_2W}{V_1V_2}} \exp\left(-\frac{1}{2}\begin{pmatrix}\theta_1 & \theta_2\end{pmatrix}\Lambda\begin{pmatrix}\theta_1 \\ \theta_2\end{pmatrix}\right), \quad (19)$$

where

$$\Lambda = \begin{pmatrix}\frac{1}{V_1} & 0 \\ 0 & \frac{1}{V_2}\end{pmatrix} + \frac{1}{V_1V_2 + V_1W + V_2W} \begin{pmatrix}V_2 + W & -W \\ -W & V_1 + W\end{pmatrix}. \quad (20)$$

- (f) We will show these results can be used in the context of a genome-wide association study on Type II diabetes, reported by Frayling et al. (2007, Science). Two sets of data were independently collected, resulting in two log odds ratios $\hat{\theta}_j$, $j = 1, 2$, for each SNP.

For SNP rs9939609 point estimates (95% confidence intervals) were 1.27 (1.16, 1.37) and 1.15 (1.09, 1.23). Suppose we have a normal prior for the odds ratio that has a 95% range (0.67, 1.50).

- i. Find W from this interval, and then calculate the posterior median and 95% intervals for θ based on (i) the first dataset only, (ii) both of the populations.

Solution:

- ii.
- iii.

3. We will carry out a Bayesian analysis of the lung cancer and radon data, that were examined in lectures, using INLA. These data are available on the class website. The likelihood is $Y_i | \beta \sim \text{Poisson}(E_i \exp(\beta_0 + \beta_1 x_i))$ independently distributed, where $\beta = (\beta_0 \ \beta_1)^\top$, Y_i and E_i are observed and expected counts of lung cancer incidence in Minnesota in 1998–2002, and x_i is a measure of residential radon in county i , $i = 1, \dots, n$.

- (a) Analyze these data using the default prior specifications in INLA. Produce figures of the INLA approximations to the marginal distributions of β_0 and β_1 , along with the posterior means, posterior standard deviations, and 2.5%, 50%, 97.5% quantiles.

Solution:

- (b) For a more informative prior specification we may reparameterize the model as independently distributed

$$Y_i | \theta \sim \text{Poisson}(E_i \theta_0 \theta_1^{x_i - \bar{x}_i}), \quad (21)$$

where $\theta = (\theta_0 \ \theta_1)^\top$ and

$$\theta_0 = \mathbb{E} \left[\frac{Y}{E} \mid x = \bar{x} \right] = \exp(\beta_0 + \beta_1 \bar{x}) \quad (22)$$

where is the expected standardized mortality ratio in an area with average radon. The parameter $\theta_1 = \exp(\beta_1)$ is the relative risk associated with a one-unit increase in radon. For θ_0 we assume a lognormal prior with 2.5% and 97.5% quantiles of 0.67 and 1.5 to give $\mu = 0$ and $\sigma = 0.21$. For θ_1 we again take a lognormal prior and assume the relative risk associated with a one-unit increase in radon is between 0.8 and 1.2 with probability 0.95, to give $\mu = -0.02$ and $\sigma = 0.10$. By converting these into normal priors in INLA, rerun your analysis, and report the same summaries.

Solution: For the priors, we have two independent normals

$$\begin{aligned} \log \theta_0 &\sim \mathcal{N}(0, 0.21^2) \\ \log \theta_1 &\sim \mathcal{N}(-0.02, 0.1^2). \end{aligned}$$

We can rewrite

$$E_i \theta_0 \theta_1^{x_i - \bar{x}} = E_i \exp(\log \theta_0 + (x_i - \bar{x}) \log \theta_1), \quad (23)$$

so after centering the x_i , we can specify priors on the intercept and coefficients as usual.

4. Consider the simple linear regression model $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, with $\epsilon_i \mid \sigma^2 \sim_{\text{iid}} \mathcal{N}(0, \sigma^2)$, $i = 1, \dots, n$. Suppose the prior distribution is of the form

$$\pi(\beta_0, \beta_1, \sigma^2) = \pi(\beta_0, \beta_1) \pi(\sigma^{-2}). \quad (24)$$

The prior for $(\beta_0 \ \beta_1)^\top$ is

$$\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} m_0 \\ m_1 \end{pmatrix}, \begin{pmatrix} v_{00} & v_{01} \\ v_{01} & v_{11} \end{pmatrix}\right) \quad (25)$$

and the prior for σ^{-2} is Gamma(a, b). In this exercise the conditional distribution required for Gibbs sampling will be derived.

- (a) Write down the form of the posterior distribution (up to proportionality) and derive the conditional distributions $p(\beta_0 \mid \beta_1, \sigma^2, \mathbf{y})$, $p(\beta_1 \mid \beta_0, \sigma^2, \mathbf{y})$ and $p(\sigma^2 \mid \beta_0, \beta_1, \mathbf{y})$. Hence, give details of the Gibbs sampling algorithm.

Solution: Using the normal likelihood and prior, we have the posterior

$$\begin{aligned} p(\beta_0, \beta_1, \sigma^2 \mid \mathbf{y}) &\propto p(\mathbf{y} \mid \beta_0, \beta_1, \sigma^2) \pi(\beta_0, \beta_1, \sigma^2) \\ &\propto \pi(\beta_0, \beta_1) \pi(\sigma^{-2}) \left(\frac{\sigma^{-2}}{2\pi}\right)^{n/2} \prod_{i=1}^n \exp\left(-\frac{1}{2\sigma^2} (y_i - \beta_0 + \beta_1 x_i)^2\right), \end{aligned}$$

where

$$\begin{aligned} \pi(\beta_0, \beta_1) &\propto \exp\left(-\frac{1}{2} (\beta - m)^\top V^{-1} (\beta - m)\right), \text{ where } V = \begin{pmatrix} v_{00} & v_{01} \\ v_{01} & v_{11} \end{pmatrix} \\ \pi(\sigma^{-2}) &\propto (\sigma^{-2})^{a-1} \exp(-b\sigma^{-2}). \end{aligned}$$

For $p(\sigma^2 \mid \beta_0, \beta_1, \mathbf{y})$, we can drop all the factors without σ^2 , so we have

$$\begin{aligned} p(\sigma^2 \mid \beta_0, \beta_1, \mathbf{y}) &\propto (\sigma^{-2})^{a+n/2-1} \exp(-b\sigma^{-2}) \prod_{i=1}^n \exp\left(-\frac{1}{2\sigma^2} (y_i - \beta_0 + \beta_1 x_i)^2\right) \\ &\propto (\sigma^{-2})^{a+n/2-1} \exp\left(-\sigma^{-2} \left(b + \frac{1}{2} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2\right)\right), \end{aligned}$$

$$\text{so } \boxed{(\sigma^{-2} \mid \beta_0, \beta_1, \mathbf{y}) \sim \text{Gamma}\left(a + \frac{n}{2}, b + \frac{1}{2} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2\right)}.$$

The conditional posteriors for β_0 and β_1 can be obtained from marginalizing $p(\beta \mid \sigma^2, \mathbf{y})$ from the next part. We'll get

$$\begin{aligned} (\beta_0 \mid \beta_1, \sigma^2, \mathbf{y}) &\sim \mathcal{N}(m_1^*, V_{11}^*) \\ (\beta_1 \mid \beta_0, \sigma^2, \mathbf{y}) &\sim \mathcal{N}(m_2^*, V_{22}^*). \end{aligned}$$

- (b) Another blocked Gibbs sampling algorithm would simulate from the distributions $p(\beta \mid \sigma^2, \mathbf{y})$ and $p(\check{\mathbf{I}}\check{\mathbf{C}}^{-2} \mid \beta, \mathbf{y})$. Derive the distributions

$$(\beta \mid \sigma^2, \mathbf{y}) \sim \mathcal{N}(m^*, V^*) \quad (26)$$

$$(\sigma^2 \mid \beta, \mathbf{y}) \sim \text{Gamma}\left(a + \frac{n}{2}, b + \frac{1}{2}(\mathbf{y} - X\beta)^\top (\mathbf{y} - X\beta)\right), \quad (27)$$

where

$$\begin{aligned} m^* &= W\hat{\beta} + (I_2 - W)m \\ V^* &= W \text{Var}(\hat{\beta}) \end{aligned}$$

and $W = (X^\top X + V^{-1}\sigma^2)^{-1} X^\top X$ and $\hat{\beta}$ is the MLE of β .

Solution: Let X be the a $n \times 2$ matrix with all 1s in the first column and $(x_1 \cdots x_n)^\top$ as the second column.

Then, Equation 27 follows from the previous part after rewriting the term

$$\frac{1}{2} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 = \frac{1}{2} (\mathbf{y} - X\beta)^\top (\mathbf{y} - X\beta). \quad (28)$$

As before, we complete the square to derive the posterior for β :

$$\begin{aligned} &(\mathbf{y} - X\beta)^\top \sigma^{-2} I_n (\mathbf{y} - X\beta) + (\beta - m)^\top V^{-1} (\beta - m) \\ &= \beta^\top \sigma^{-2} X^\top X \beta + \beta^\top V^{-1} \beta - 2\beta^\top (\sigma^{-2} X^\top \mathbf{y} + V^{-1} m) + \sigma^{-2} \mathbf{y}^\top \mathbf{y} - m^\top V^{-1} m \\ &= \beta^\top (\sigma^{-2} X^\top X + V^{-1}) \beta \\ &\quad - 2\beta^\top (\sigma^{-2} X^\top X + V^{-1}) (\sigma^{-2} X^\top \mathbf{y} + V^{-1} m) + C, \end{aligned}$$

where we have collapsed the terms that don't depend on β into C .

Recall that $\text{Var}(\hat{\beta}) = \sigma^2 (X^\top X)^{-1}$, so we have that

$$(\sigma^{-2} X^\top X + V^{-1})^{-1} = (X^\top X + \sigma^2 V^{-1})^{-1} (X^\top X) \sigma^2 (X^\top X)^{-1} = V^*,$$

so continuing the process of completing the square:

$$\begin{aligned} &\beta^\top (V^*)^{-1} \beta - 2\beta^\top (V^*)^{-1} W (X^\top X)^{-1} (X^\top \mathbf{y} + \sigma^2 V^{-1} m) + C \\ &= \beta^\top (V^*)^{-1} \beta - 2\beta^\top (V^*)^{-1} (W\hat{\beta} + W\sigma^2 (X^\top X)^{-1} V^{-1} m) + C \\ &= \beta^\top (V^*)^{-1} \beta - 2\beta^\top (V^*)^{-1} \left(W\hat{\beta} + \sigma^2 \left(V (X^\top X + \sigma^2 V^{-1}) \right)^{-1} m \right) + C \\ &= \beta^\top (V^*)^{-1} \beta - 2\beta^\top (V^*)^{-1} \left(W\hat{\beta} + \sigma^2 (V X^\top X + \sigma^2 I)^{-1} m \right) + C \\ &= \beta^\top (V^*)^{-1} \beta - 2\beta^\top (V^*)^{-1} (W\hat{\beta} + (I_2 - W)m) + C \\ &= (\beta - m^*)^\top (V^*)^{-1} (\beta - m^*) + C', \end{aligned}$$

where we have applied the Woodbury matrix identity.

Thus, all the factors that contain β can be written as a quadratic form which gives us the result in Equation 26.