Coursework 5: STAT 570

Philip Pham

November 4, 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

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

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

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

Show that the odds ratio is equal to

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

Solution: We have that

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

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

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

The $\mathbb{P}\left(Y = y\right)$ 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

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

so the score function is

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

Thus, the Fisher information is

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

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:

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

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

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

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

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

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

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

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

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

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

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

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

which is the $\operatorname{Beta}\left(a + x_i, b + n_i - x_i\right)$ 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 $p_1 \mid x_1 \sim \operatorname{Beta}\left(97, 105\right)$ and $p_2 \mid x_2 \sim \operatorname{Beta}\left(110, 667\right)$.

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

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

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

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

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

For $p_1 \mid x_1$, the interval is $[0.4226, 0.5380]$. For $p_2 \mid x_2$, the inverval 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: We can reparameterize the beta distribution in terms of two gamma random variables. Let $r_a \sim \operatorname{Gamma}\left(a, 1\right)$ and $r_b \sim \operatorname{Gamma}\left(b, 1\right)$. Let $x = r_a / \left(r_a + r_b\right)$ and $s = r_a + r_b$, so we can invert and get $r_a = xs$ and $r_b = \left(1 - x\right)s$.

Taking the Jacobina and hanging variables, we'll have the density function

$$\begin{aligned} p\left(x, s\right) &= \left(\frac{1}{\Gamma\left(a\right)}\left(xs\right)^{a-1}\exp\left(-xs\right)\right)\left(\frac{1}{\Gamma\left(b\right)}\left(\left(1 - x\right)s\right)^{b-1}\exp\left(-\left(1 - x\right)s\right)\right)s \\ &= \frac{1}{\Gamma\left(a\right)\Gamma\left(b\right)}x^{a-1}\left(1 - x\right)^{b-1}\left(s^{a+b-1}\exp\left(-s\right)\right). \end{aligned}$$

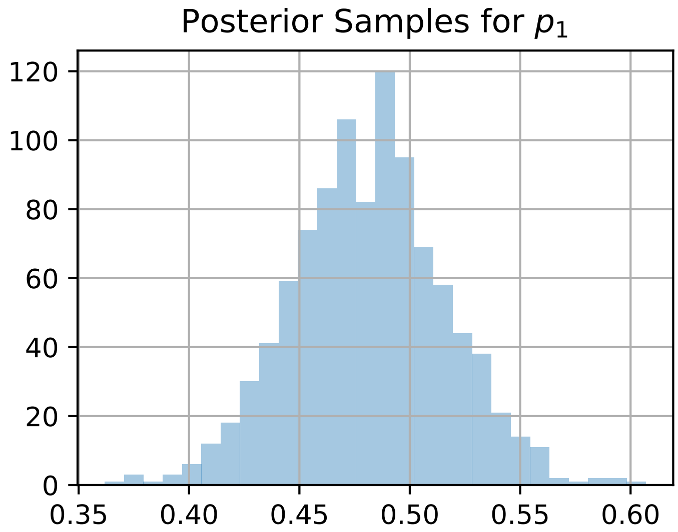

Figure 1: 1,000 samples from the posterior $p_1 \mid x_1$.

We recognize the gamma function and marginalize over s to obtain that $x \sim \text{Beta}\,(a, b)$.

Now, sums of gamma random variables are also gamma random variables, so as $a \to \infty$ and $b \to \infty$ r_a and r_b converge in distribution to the normal distribution.

Thus, we can apply the delta method to get an asymptotic distribution for the beta distribution. Let $h\,(z_1, z_2) = z_1 / \,(z_1 + z_2)$. Then, $x = h\,(z_1, z_2)$, and

$$\mathbb{E}\,[x] = h\,(\mathbb{E}\,[z_1]\,, \mathbb{E}\,[z_2]) = \frac{a}{a+b} \tag{13}$$

$$\begin{aligned}\operatorname{Var}(x) &\approx \begin{pmatrix} \frac{b}{(a+b)^2} & -\frac{a}{(a+b)^2} \end{pmatrix}^\intercal \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \begin{pmatrix} \frac{b}{(a+b)^2} \\ -\frac{a}{(a+b)^2} \end{pmatrix} \\ &= \left(\frac{1}{a+b}\right)\left(\frac{a}{a+b}\right)\left(\frac{b}{a+b}\right). \end{aligned} \tag{14}$$

which results in the same mean, and the variance is asymptotically equivalent to the variance in the previous part.

Applying Equations 13 and 14, we obtain the 90% intervals $(0.4224, 0.5380)$ for p_1 and $(0.1210, 0.1621)$ for p_2, which are virtually identically to the exact calculation in the previous part, which is unsurprising since n_1 and n_2 are quite large.

(f) Simulate samples $p_1(t)$, $p_2(t)$, $t = 1, \ldots, T = 1000$ from the posterior distributions $p_1 \mid x_1$ and $p_2 \mid x_2$. Form histogram representations of the posterior distributions using these samples and obtain sample-based 90% credible intervals.

Solution: The histogram of samples from $p_1 \mid x_1$ and $p_2 \mid x_2$ and be found in Figures 1 and 2, respectively.

The sample 90% interval for p_1 was $(0.4255, 0.5393)$. The sampled 90% intervals for p_2 were $(0.1209, 0.1634)$, which agree with the previous interval calculations.

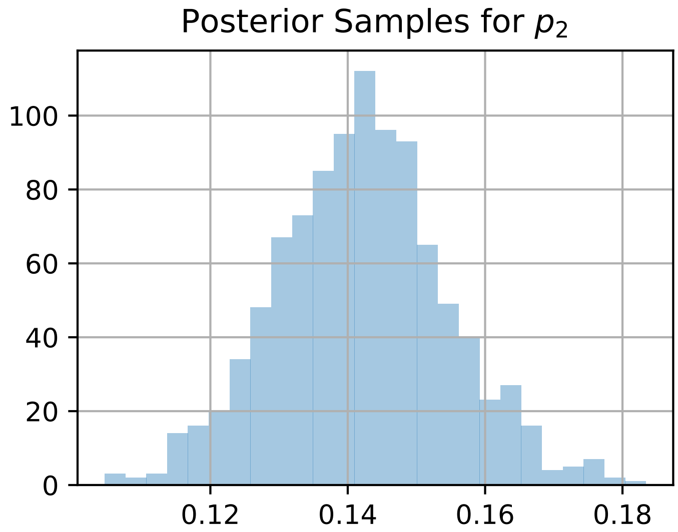

Figure 2: 1,000 samples from the posterior $p_2 \mid x_2$.

(g) Obtain samples from the posterior distribution of $\theta|x_1, x_2$ and form the histogram representation of the posterior. Obtain the posterior median and 90% credible interval for $\theta \mid x_1, x_2$ and compare with the likelihood analysis.

Solution: To get a posterior sample for θ, we draw samples from $p_1 \mid x_1$ and $p_2 \mid x_2$ and calculate θ. The samples can be seen in Figure 3.
The samples 90% credible interval was $(6.3137, 6.4755)$, and the sampled posterior median was 5.6486. The median is very close to the MLE in Equation 9. The 90% credible interval is much smaller however since we make the prior beta assumption for p_1 and p_2.
The computations for the analysis can be found in `case_control.ipynb`.

(h) Suppose the rate of esophageal cancer is 18 in 100,000. Describe how this information may be used to evaluate $q_1 = \mathbb{P}(Y = 1 \mid X = 1)$ and $q_0 = \mathbb{P}(Y = 1 \mid X = 0)$.

Solution: We can apply Bayes' rule since we know $\mathbb{P}(Y = 1) = 9/50,000 = 0.00018$ and $\mathbb{P}(Y = 0) = 49,991/50,000 = 0.99982$.
Thus, we have that

$$\begin{aligned}
\mathbb{P}(Y = 1 \mid X = 1) &= \frac{\mathbb{P}(X = 1 \mid Y = 1)\,\mathbb{P}(Y = 1)}{\mathbb{P}(X = 1 \mid Y = 0)\,\mathbb{P}(Y = 0) + \mathbb{P}(X = 1 \mid Y = 1)\,\mathbb{P}(Y = 1)} \\
&= \frac{p_1 \mathbb{P}(Y = 1)}{p_2 \mathbb{P}(Y = 0) + p_1 \mathbb{P}(Y = 1)} \\
\mathbb{P}(Y = 1 \mid X = 0) &= \frac{\mathbb{P}(X = 0 \mid Y = 1)\,\mathbb{P}(Y = 1)}{\mathbb{P}(X = 0 \mid Y = 0)\,\mathbb{P}(Y = 0) + \mathbb{P}(X = 0 \mid Y = 1)\,\mathbb{P}(Y = 1)} \\
&= \frac{(1 - p_1)\,\mathbb{P}(Y = 1)}{(1 - p_2)\,\mathbb{P}(Y = 0) + (1 - p_1)\,\mathbb{P}(Y = 1)}.
\end{aligned}$$

We can either substitute the the MLE $\hat{p}_i$ for p_i or integrate over the posteriors $p_1 \mid x_1$ and $p_2 \mid x_2$.
For example, the MLE estimates are $\hat{q}_1 \approx 0.0006140$ and $\hat{q}_2 \approx 0.0001089$.

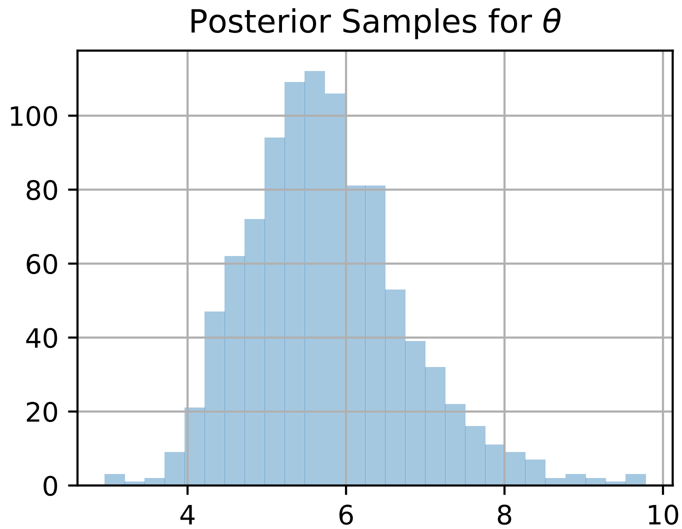

Figure 3: 1,000 samples from the posterior $\theta \mid x_2$.

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

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

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

   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

$$
\frac{p\left(\hat{\theta} \mid M_0\right)}{p\left(\hat{\theta} \mid M_1\right)} = \frac{1}{\sqrt{1-r}} \exp\left(-\frac{Z^2}{2} r\right), \tag{15}
$$

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

Solution: We have that

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

after completing the square. Substituting into the left-hand side of Equation 15, we obtain

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

as desired.

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

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

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

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

Solution: We have

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

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

$$\theta \mid \theta_1, \theta_2 \sim \mathcal{N}\left(\frac{V_2 W \theta_1 + V_1 W \theta_2}{V_1 V_2 + V_1 W + V_2 W}, \frac{V_1 V_2 W}{V_1 V_2 + V_1 W + V_2 W}\right). \tag{16}$$

(e) Derive the Bayes factor

$$\frac{p(\theta_1, \theta_2 \mid M_0)}{p(\theta_1, \theta_2 \mid M_1)}, \tag{17}$$

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

$$\begin{aligned} p(\theta_1, \theta_2 \mid M_0) &= p(\theta_1, \theta_2 \mid \theta = 0) \\ &= \frac{1}{2\pi\sqrt{V_1 V_2}} \exp\left(-\frac{1}{2}\begin{pmatrix}\theta_1 & \theta_2\end{pmatrix}\begin{pmatrix}\frac{1}{V_1} & 0 \\ 0 & \frac{1}{V_2}\end{pmatrix}\begin{pmatrix}\theta_1 \\ \theta_2\end{pmatrix}\right). \end{aligned} \tag{18}$$

Under M_1, we have that

$$p(\theta_1, \theta_2 \mid M_1) = \int_{-\infty}^{\infty} p(\theta_1, \theta_2 \mid \theta)\, p(\theta)\ \mathrm{d}\theta. \tag{19}$$

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

$$\theta_1, \theta_2 \mid M_1 \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} V_1 + W & W \\ W & V_2 + W \end{pmatrix}\right) \tag{20}$$

by conjugacy of the multivariate normal distribution.

The Bayes factor can then be computed:

$$\sqrt{\frac{V_1 V_2 + V_1 W + V_2 W}{V_1 V_2}} \exp\left(-\frac{1}{2}\begin{pmatrix}\theta_1 & \theta_2\end{pmatrix}\Lambda\begin{pmatrix}\theta_1 \\ \theta_2\end{pmatrix}\right), \tag{21}$$

where

$$\Lambda = \begin{pmatrix}\frac{1}{V_1} & 0 \\ 0 & \frac{1}{V_2}\end{pmatrix} + \frac{1}{V_1 V_2 + V_1 W + V_2 W}\begin{pmatrix} V_2 + W & -W \\ -W & V_1 + W \end{pmatrix}. \tag{22}$$

(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 of the odds ratios (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 log odds ratio that has a 95% range $(\log(2/3), \log(3/2))$.

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: $\theta/\sqrt{W}$ will be a standard normal random variable, and we want $\mathbb{P}(\theta \leq \log(3/2)) = 0.975$, so we solve $\log(3/2)/\sqrt{W} = \Phi^{-1}(0.975) \implies W \approx 0.04279675$.

Similarly, we estimate $V_1 \approx 0.00180178$ and $V_2 \approx 0.000950254$.

Part 2a gives the posterior for one dataset $\mathcal{N}(0,)$, so $\theta \mid \theta_1 \sim \mathcal{N}(0.22936, 0.001729)$.

This has median 0.22936 and 95% interval $(0.14786, 0.310858)$.

Equation 16 gives us the posterior for both datasets. $\theta \mid \theta_1, \theta_2 \sim \mathcal{N}(0.1715400.0006132)$. This has median 0.171540 and 95% interval $(0.123005, 0.2201)$.

The analysis can be found at `genome_association.ipynb`.

ii. Calculate the Bayes factor based on the first dataset only, and then based on both datasets.

Solution: Simply using the formulas from Parts and , for the dataset only, we get 1.22993×10^{-6}. For both datasets, we obtain 4.40892×10^{-12}. In both cases, the data heavily favors M_1.

The analysis can be found at `genome_association.ipynb`.

iii. With a prior of $\pi_1 = 1/5000$, calculate the probabilities, $\mathbb{P}(M_1 \mid \theta_1)$ and $\mathbb{P}(M_1 \mid \theta_1, \theta_2)$.

Solution: We apply the result in Part 2c combined with the Bayes factor in the previous part.

We obtain $\mathbb{P}(M_1 \mid \theta_1) = 0.993889$ and $\mathbb{P}(M_1 \mid \theta_1, \theta_2) = 0.99999998$, so the M_1 is heavily favored.

The analysis can be found at `genome_association.ipynb`.

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 $Y_i \mid \beta \sim \text{Poisson}(E_i \exp(\beta_0 + \beta_1 x_i))$ independently distributed, where $\beta = \begin{pmatrix}\beta_0 & \beta_1\end{pmatrix}^\intercal$, 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, $i = 1, \ldots, n$.

(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: The results are shown in Table 2.

	β_0	β_1
Posterior mean	0.17	-0.04
Posterior standard error	0.03	0.01
2.5% quantile	0.12	-0.05
50% quantile	0.17	-0.04
97.5% quantile	0.22	-0.03

Table 2: Results for default prior.

The marginals and joint distribution can be found in Figures 4, 5, 6. They are almost identical to those of Figure 3.6 in the textbook, which was fit with MCMC.

Details of the analysis can be found in `lung_cancer_radon.ipynb`.

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

$$Y_i \mid \theta \sim \text{Poisson}\left(E_i \theta_0 \theta_1^{x_i - \bar{x}_i}\right), \tag{23}$$

where $\theta = \begin{pmatrix}\theta_0 & \theta_1\end{pmatrix}^\intercal$ and

$$\theta_0 = \mathbb{E}\left[\frac{Y}{E} \mid x = \bar{x}\right] = \exp(\beta_0 + \beta_1 \bar{x}) \tag{24}$$

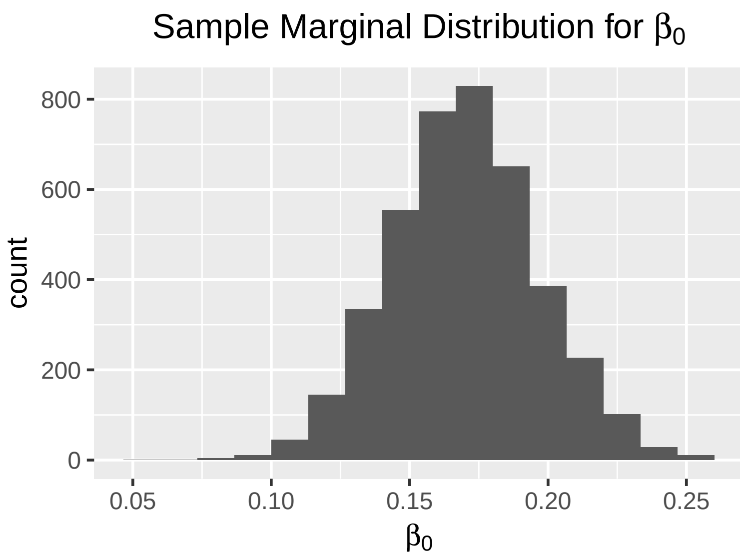

Figure 4: Marginal distribution for β_0 fitted with INLA.

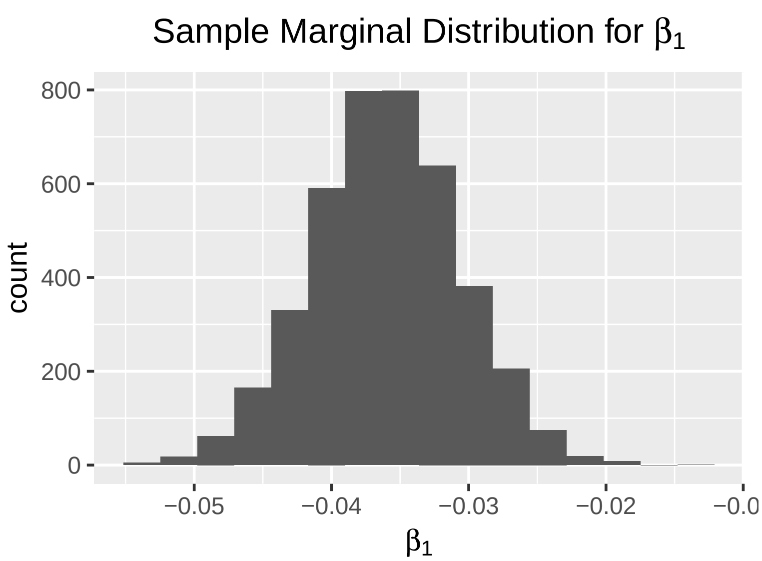

Figure 5: Marginal distribution for β_1 fitted with INLA.

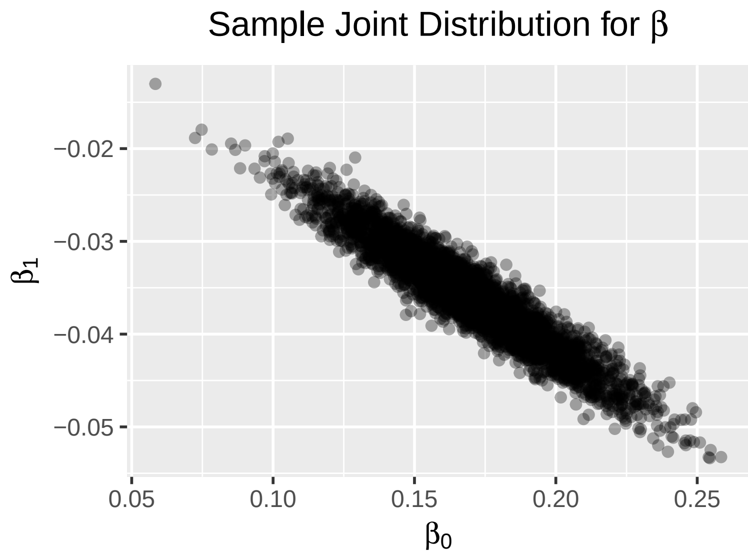

Figure 6: Joint distribution for β fitted with INLA.

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

$$\log\theta_0 \sim \mathcal{N}\left(0, 0.21^2\right)$$
$$\log\theta_1 \sim \mathcal{N}\left(-0.02, 0.1^2\right).$$

We can rewrite

$$E_i\theta_0\theta_1^{x_i-\bar{x}_i} = E_i\exp\left(\log\theta_0 + (x_i - \bar{x})\log\theta_1\right), \tag{25}$$

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

Results can be found in Table 3.

The marginals and joint distribution can be found in Figures 7, 8, 9, which was fit with MCMC. They are almost identical to those of Figure 3.6 in the textbook.

Details of the analysis can be found in `lung_cancer_radon.ipynb`.

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

$$\pi\left(\beta_0, \beta_1, \sigma^2\right) = \pi\left(\beta_0, \beta_1\right)\pi\left(\sigma^{-2}\right). \tag{26}$$

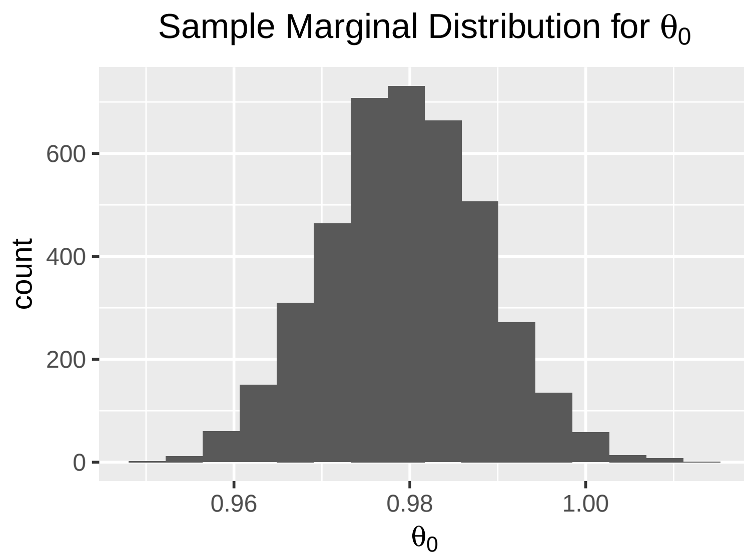

Figure 7: Marginal distribution for θ_0 fitted with INLA.

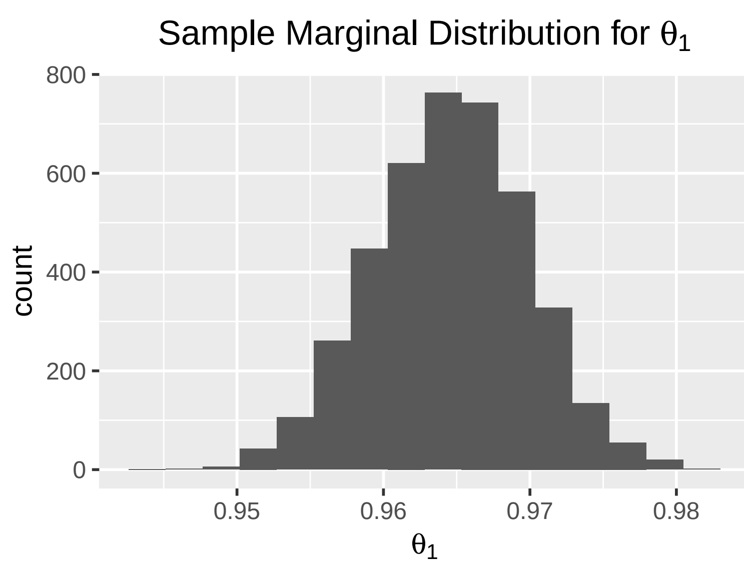

Figure 8: Marginal distribution for θ_1 fitted with INLA.

	θ_0	θ_1
Posterior mean	0.98	0.96
Posterior standard error	0.01	0.01
2.5% quantile	0.96	0.95
50% quantile	0.98	0.96
97.5% quantile	1.00	0.97

Table 3: The results of fitting θ with INLA.

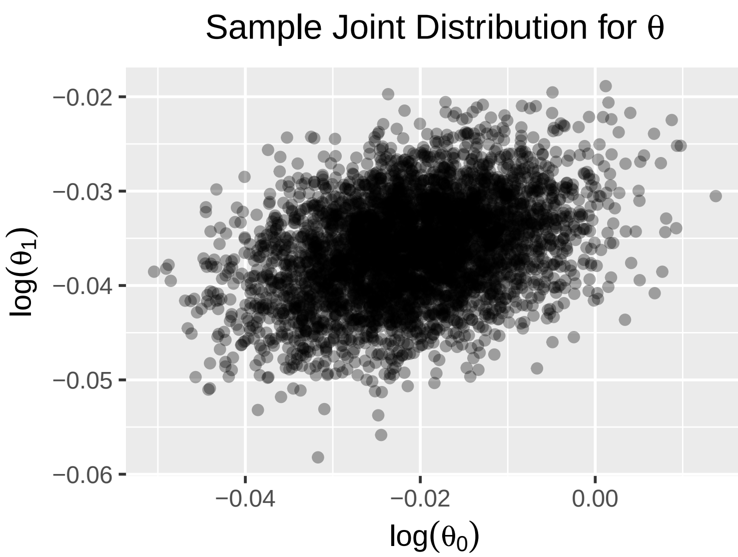

Figure 9: Joint distribution for θ fitted with INLA.

The prior for $\begin{pmatrix}\beta_0 & \beta_1\end{pmatrix}^\intercal$ is

$$\begin{pmatrix}\beta_0\\ \beta_1\end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix}m_0\\ m_1\end{pmatrix}, \begin{pmatrix}v_{00} & v_{01}\\ v_{01} & v_{11}\end{pmatrix}\right) \tag{27}$$

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 $p\left(\beta_0 \mid \beta_1, \sigma^2, \mathbf{y}\right)$, $p\left(\beta_1 \mid \beta_0, \sigma^2, \mathbf{y}\right)$ and $p\left(\sigma^2 \mid \beta_0, \beta_1, \mathbf{y}\right)$. Hence, give details of the Gibbs sampling algorithm.

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

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

where

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

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

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

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

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

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

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

$$\left(\beta \mid \sigma^2, \mathbf{y}\right) \sim \mathcal{N}(m^*, V^*) \tag{28}$$
$$\left(\sigma^{-2} \mid \beta, \mathbf{y}\right) \sim \text{Gamma}\left(a + \frac{n}{2}, b + \frac{1}{2}(\mathbf{y} - X\beta)^\intercal(\mathbf{y} - X\beta)\right), \tag{29}$$

where

$$m^* = W\hat{\beta} + (I_2 - W)m$$
$$V^* = W \operatorname{Var}\left(\hat{\beta}\right)$$

and $W = \left(X^\intercal X + V^{-1}\sigma^2\right)^{-1} X^\intercal X$ 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 $\begin{pmatrix} x_1 \cdots x_n \end{pmatrix}^\intercal$ as the second column.

Then, Equation 29 follows from the previous part after rewriting the term

$$\frac{1}{2}\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 = \frac{1}{2}(\mathbf{y} - X\beta)^\intercal(\mathbf{y} - X\beta). \tag{30}$$

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

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

where we have collapsed the terms that don't depend on β into C.
Recall that $\text{Var}\left(\hat{\beta}\right) = \sigma^2 \left(X^\intercal X\right)^{-1}$, so we have that

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

so continuing the process of completing the square:

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

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 28.