Stat 461-561: Quiz 4

Wednesday 9th April 2008

- **Exercise 1**. Consider the following linear regression model where

$$y = X\beta + \varepsilon$$

with $y = (y_1, ..., y_n)^{\mathrm{T}} \in \mathbb{R}^n$, $\beta = (\beta_1, ..., \beta_p)^{\mathrm{T}} \in \mathbb{R}^p$, X is a known matrix of appropriate dimension and $\varepsilon \sim \mathcal{N}\left(0, \sigma^2 I_n\right)$ where I_n is the identity matrix of dimension $n \times n$. We follow a full Bayesian approach and propose the following normal-inverse Gamma prior

$$\begin{aligned} p\left(\beta, \sigma^2\right) &= p\left(\beta \mid \sigma^2\right) p\left(\sigma^2\right) \\ &= \mathcal{IG}\left(\sigma^2; a, b\right) \mathcal{N}\left(\beta; 0, \sigma^2 \delta^2 I_p\right). \end{aligned}$$

where a, b, δ^2 are fixed hyperparameters.

Question 1.1: **[1 point]** Write down the joint distribution $p\left(y, \beta, \sigma^2\right)$.

We have

$$\begin{aligned} p\left(y, \beta, \sigma^2\right) &= p\left(y \mid \beta, \sigma^2\right) p\left(\beta \mid \sigma^2\right) p\left(\sigma^2\right) \\ &= \frac{1}{\left(2\pi\sigma^2\right)^{n/2}} \exp\left(-\frac{(y - X\beta)^{\mathrm{T}} (y - X\beta)}{2\sigma^2}\right) \\ &\quad \times \frac{1}{\left(2\pi\delta^2\sigma^2\right)^{p/2}} \exp\left(-\frac{\beta^{\mathrm{T}}\beta}{2\delta^2\sigma^2}\right) \\ &\quad \times \frac{b^a}{\Gamma(a)} \frac{\exp\left(-b/\sigma^2\right)}{\left(\sigma^2\right)^{a+1}} \end{aligned}$$

Question 1.2: **[3 points]** Show that the posterior distribution $p\left(\beta, \sigma^2 \middle| y\right)$ is also a normal inverse-Gamma distribution and determine its exact expression. [Hint: First establish the expression of $p\left(\beta \mid y, \sigma^2\right)$ then the expression of the marginal distribution $p\left(\sigma^2 \middle| y\right)$.]

We have

$$p\left(\beta \mid y, \sigma^2\right) \propto \exp\left(-\frac{(y - X\beta)^{\mathrm{T}} (y - X\beta)}{2\sigma^2}\right) \exp\left(-\frac{\beta^{\mathrm{T}}\beta}{2\delta^2\sigma^2}\right)$$

and

$$\begin{aligned}
& (y - X\beta)^{\mathrm{T}} (y - X\beta) + \frac{\beta^{\mathrm{T}}\beta}{\delta^2} \\
= & \; y^{\mathrm{T}} y - 2y^{\mathrm{T}} X\beta + \beta^{\mathrm{T}} \left(X^{\mathrm{T}} X\right) \beta + \frac{\beta^{\mathrm{T}}\beta}{\delta^2} \\
= & \; (\beta - m)^{\mathrm{T}} \Sigma^{-1} (\beta - m) - m^{\mathrm{T}} \Sigma^{-1} m + y^{\mathrm{T}} y
\end{aligned}$$

where

$$\begin{aligned}
\Sigma^{-1} &= X^{\mathrm{T}} X + \delta^{-2} I, \\
m &= \Sigma X^{\mathrm{T}} y.
\end{aligned}$$

So

$$p\left(\beta \middle| y, \sigma^2\right) = \mathcal{N}\left(\beta; m, \sigma^2 \Sigma\right).$$

It follows that

$$\begin{aligned}
p\left(y, \sigma^2\right) &= \frac{1}{\left(2\pi\sigma^2\right)^{n/2}} \frac{|\Sigma|^{p/2}}{\left(\delta^2\right)^{p/2}} \exp\left(-\frac{y^{\mathrm{T}} y - m^{\mathrm{T}} \Sigma^{-1} m}{2\sigma^2}\right) \\
& \quad \times \frac{b^a}{\Gamma(a)} \frac{\exp\left(-b/\sigma^2\right)}{\left(\sigma^2\right)^{a+1}}
\end{aligned}$$

so

$$p\left(\sigma^2 \middle| y\right) = \mathcal{IG}\left(\sigma^2; a + \frac{n}{2}, b + \frac{y^{\mathrm{T}} y - m^{\mathrm{T}} \Sigma^{-1} m}{2}\right).$$

Question 1.3: **[1 point]** Establish the expression of the marginal likelihood $p(y)$.

We have

$$\begin{aligned}
p(y) &= \int p\left(y, \sigma^2\right) d\sigma^2 \\
&= \frac{1}{(2\pi)^{n/2}} \frac{|\Sigma|^{p/2}}{\left(\delta^2\right)^{p/2}} \frac{b^a}{\Gamma(a)} \frac{\Gamma\left(a + \frac{n}{2}\right)}{\left(b + \frac{y^{\mathrm{T}} y - m^{\mathrm{T}} \Sigma^{-1} m}{2}\right)^{a + \frac{n}{2}}}.
\end{aligned}$$

Question 1.4: [**1 point**] Assume you have two candidate Bayesian linear models and that these models have the same hyperparameters a, b, δ^2 but $\beta \in \mathbb{R}$ for the first model and $\beta \in \mathbb{R}^2$ for the second model. What is the limiting value of the Bayes factor $p\left(y \middle| \text{model } 1\right) / p\left(y \middle| \text{model } 2\right)$ as $\delta^2 \to \infty$? [You do not need to detail all the calculations]

As $\delta^2 \to \infty$ we have an improper prior and Lindley's paradox tells us we always pick the simplest model and

$$\lim_{\delta^2 \to \infty} \frac{p\left(y \middle| \text{model } 1\right)}{p\left(y \middle| \text{model } 2\right)} = \infty.$$

You can use the expression of the marginal likelihood to establish it rigourously.

• **Exercise 2**. Given parameters $\theta_1, \theta_2, ..., \theta_n$, the observations $y_1, y_2, ..., y_n$ are assumed independently distributed as Poisson random variables

$$p\left(y_i \middle| \theta_i\right) = \frac{\theta_i^{y_i} \exp\left(-\theta_i\right)}{y_i!}.$$

The prior distribution for the vector $(\theta_1, \theta_2, ..., \theta_n)$ is constructed as follows. First, given an hyperparameter ϕ, we have independent and identically distributed parameters $(\theta_1, \theta_2, ..., \theta_n)$, that is

$$p\left(\theta_1, \theta_2, ..., \theta_n \middle| \phi\right) = \prod_{i=1}^{n} p\left(\theta_i \middle| \phi\right)$$

where

$$p\left(\theta_i \middle| \phi\right) = \mathcal{G}\left(\theta_i; 1, \phi\right).$$

Question 2.1: **[1 point]** Show that $p\left(\theta_i \middle| y_i, \phi\right)$ is a Gamma distribution and establish its parameters. Deduce the expression of $p\left(\theta_1, \theta_2, ..., \theta_n \middle| y_1, y_2, ..., y_n, \phi\right)$.

We have

$$p\left(\theta_1, \theta_2, ..., \theta_n \middle| y_1, y_2, ..., y_n, \phi\right) = \prod_{i=1}^{n} p\left(\theta_i \middle| y_i, \phi\right)$$

and

$$\begin{aligned} p\left(\theta_i \middle| y_i, \phi\right) &\propto p\left(y_i \middle| \theta_i\right) p\left(\theta_i \middle| \phi\right) \\ &\propto \theta_i^{y_i} \exp\left(-\theta_i\right) \exp\left(-\phi\theta_i\right) \\ &\propto \theta_i^{y_i} \exp\left(-\left(1+\phi\right)\theta_i\right) \end{aligned}$$

so

$$p\left(\theta_i \middle| y_i, \phi\right) = \mathcal{G}\left(\theta_i; y_i + 1, \phi + 1\right).$$

Question 2.2: **[1 point]** Propose an empirical Bayes estimate of ϕ.

We have

$$\begin{aligned} \mathbb{E}\left(y_i\right) &= \mathbb{E}\left(\mathbb{E}\left(y_i \middle| \theta_i\right)\right) = \mathbb{E}\left(\theta_i\right) \\ &= \frac{1}{\phi} \end{aligned}$$

so one potential estimate is

$$\widehat{\phi} = \left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)^{-1}.$$

Question 2.3: **[2 points]** Now consider the case where the hyperparameter is assigned an improper prior distribution

$$p\left(\phi\right) \propto 1_{(0,\infty)}\left(\phi\right).$$

Detail a Gibbs sampling algorithm to sample from $p\left(\theta_1, \theta_2, ..., \theta_n, \phi \middle| y_1, ..., y_n\right)$. What is the main advantage of this full Bayesian approach over the empirical Bayes approach?

To sample from $p\left(\theta_1, \theta_2, ..., \theta_n, \phi \middle| y_1, ..., y_n\right)$, we sample iteratively from $p\left(\theta_1, \theta_2, ..., \theta_n \middle| y_1, ..., y_n, \phi\right)$ which is a product of Gamma and from

$$\begin{aligned}
& p\left(\phi \middle| y_1, ..., y_n, \theta_1, \theta_2, ..., \theta_n\right) \\
= \;& p\left(\phi \middle| \theta_1, \theta_2, ..., \theta_n\right) \\
\propto \;& \prod_{i=1}^{n} \mathcal{G}\left(\theta_i; 1, \phi\right) \\
\propto \;& \phi^n \exp\left(-\phi \sum_{i=1}^{n} \theta_i\right) \\
= \;& \mathcal{G}\left(\phi; n+1, \sum_{i=1}^{n} \theta_i\right).
\end{aligned}$$