Stat 461-561: Solutions Quiz 4

Wednesday 11th April 2007

Exercise 1. Let n observations $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ be modelled through a simple linear regression model. We have for $i = 1, ..., n$

$$y_i = \alpha + \beta x_i + e_i$$

where the errors $e_i \sim \mathcal{N}(0, \sigma^2)$ are independent. We assume that the unknown parameters $(\alpha, \beta, \sigma^2)$ are random and follow the prior distribution

$$p(\alpha, \beta, \sigma^2) = p(\alpha) p(\beta) p(\sigma^2)$$

with

$$\begin{aligned} p(\alpha) &= \mathcal{N}(\alpha; \alpha_0, \sigma_a^2), \\ p(\beta) &= \mathcal{N}(\beta; \beta_0, \sigma_b^2), \\ p(\sigma^2) &= \mathcal{IG}(\sigma^2; a, b) \end{aligned}$$

where the hyperparameters $(\alpha_0, \sigma_a^2, \beta_0, \sigma_b^2, a, b)$ are fixed and kown.

(a) Write down the joint distribution $p(y_{1:n}, \alpha, \beta, \sigma^2 | x_{1:n})$ where $x_{1:n} := (x_1, x_2, \ldots, x_n)$ and $y_{1:n} := (y_{1:n}, y_2, \ldots, y_n)$.

We have

$$\begin{aligned} p(y_{1:n}, \alpha, \beta, \sigma^2 | x_{1:n}) &= p(\alpha) p(\beta) p(\sigma^2) \prod_{i=1}^{n} p(y_i | x_i, \alpha, \beta, \sigma^2) \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \alpha + \beta x_i)^2\right) \\ &\quad \times \frac{1}{(2\pi\sigma_a^2)^{n/2}} \exp\left(-\frac{1}{2\sigma_a^2} (\alpha - \alpha_0)^2\right) \\ &\quad \times \frac{1}{(2\pi\sigma_b^2)^{n/2}} \exp\left(-\frac{1}{2\sigma_b^2} (\beta - \beta_0)^2\right) \\ &\quad \times \frac{b^a}{\Gamma(a)} \frac{\exp(-b/\sigma^2)}{(\sigma^2)^{a+1}} \end{aligned}$$

(b) Write down the full conditional distributions $p(\alpha | x_{1:n}, y_{1:n}, \beta, \sigma^2)$, $p(\beta | x_{1:n}, y_{1:n}, \alpha, \sigma^2)$ and $p(\sigma^2 | x_{1:n}, y_{1:n}, \alpha, \beta)$.

We have

$$\begin{aligned} & p(\alpha | x_{1:n}, y_{1:n}, \beta, \sigma^2) \\ \propto\ & \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \alpha + \beta x_i)^2\right) \exp\left(-\frac{1}{2\sigma_a^2} (\alpha - \alpha_0)^2\right) \\ =\ & \mathcal{N}\left(\alpha; \frac{\sigma_a^2 \sigma^2}{n\sigma_a^2 + \sigma^2} \left(\frac{n(\overline{y} - \beta\overline{x})}{\sigma^2} + \frac{\alpha_0}{\sigma_a^2}\right), \frac{\sigma_a^2 \sigma^2}{n\sigma_a^2 + \sigma^2}\right) \end{aligned}$$

$$\begin{aligned}
& p\left(\beta \mid x_{1:n}, y_{1:n}, \alpha, \sigma^2\right) \\
\propto \; & \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i-\alpha+\beta x_i\right)^2\right)\exp\left(-\frac{1}{2\sigma_b^2}\left(\beta-\beta_0\right)^2\right) \\
= \; & \mathcal{N}\left(\beta; \frac{\sigma^2\sigma_b^2}{\sigma_b^2\sum_{i=1}^{n}x_i^2+\sigma^2}\left(\frac{\sum_{i=1}^{n}x_i\left(y_i-\alpha\right)}{\sigma^2}+\frac{\beta_0}{\sigma_b^2}\right), \frac{\sigma^2\sigma_b^2}{\sigma_b^2\sum_{i=1}^{n}x_i^2+\sigma^2}\right)
\end{aligned}$$

$$\begin{aligned}
& p\left(\sigma^2 \mid x_{1:n}, y_{1:n}, \alpha, \beta\right) \\
\propto \; & \frac{1}{\left(\sigma^2\right)^{n/2}}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i-\alpha+\beta x_i\right)^2\right)\frac{\exp\left(-b/\sigma^2\right)}{\left(\sigma^2\right)^{a+1}} \\
= \; & \mathcal{IG}\left(a+n/2, b+\frac{1}{2}\sum_{i=1}^{n}\left(y_i-\alpha+\beta x_i\right)\right)
\end{aligned}$$

(c) In words, explain how you would implement the Gibbs sampler to obtain (approximate) samples from $p\left(\alpha, \beta, \sigma^2 \mid x_{1:n}, y_{1:n}\right)$.

See lecture notes.

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

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

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

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

where

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

Second, the hyperparameter is assigned an improper prior distribution

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

(a) Write the expression of the joint posterior distribution $p\left(\theta_1, \theta_2, ..., \theta_n, \phi \mid y_1, ..., y_n\right)$ up to a normalizing constant.

We have

$$p\left(\theta_i \mid \phi\right) = \frac{\phi}{\Gamma\left(1\right)}\exp\left(-\phi\theta_i\right)$$

so

$$p\left(\theta_1, \theta_2, ..., \theta_n, \phi \mid y_1, ..., y_n\right) \propto \phi^n\left(\prod_{i=1}^{n}\theta_i^{y_i}\right)\exp\left(-\left(1+\phi\right)\sum_{i=1}^{n}\theta_i\right).$$

(b) We introduce $z = (1+\phi)^{-1}$. Derive the (improper) prior $p\left(z\right)$.

We have

$$\phi = \frac{1-z}{z} \Rightarrow \left|\frac{d\phi}{dz}\right| = \frac{1}{z^2}$$

so

$$p(z) = p(\phi)\left|\frac{d\phi}{dz}\right| \propto \frac{1_{(0,1)}(z)}{z^2}.$$

(c) Establish the expression of the joint posterior distribution $p(\theta_1, \theta_2, ..., \theta_n, z| y_1, ..., y_n)$ up to a normalizing constant.

We have

$$p(\theta_i| z) = \frac{\frac{1-z}{z}}{\Gamma(1)} \exp\left(-\frac{1-z}{z}\theta_i\right)$$

so

$$\begin{aligned} p(\theta_1, \theta_2, ..., \theta_n, z| y_1, ..., y_n) &\propto \left(\frac{1-z}{z}\right)^n \left(\prod_{i=1}^n \theta_i^{y_i}\right) \exp\left(-\frac{1}{z}\sum_{i=1}^n \theta_i\right)\frac{1}{z^2} \\ &\propto \left(\prod_{i=1}^n \theta_i^{y_i}\right)\frac{(1-z)^n}{z^{n+2}} \exp\left(-\frac{1}{z}\sum_{i=1}^n \theta_i\right). \end{aligned}$$

(d) Let $\overline{y} = \frac{1}{n}\sum_{i=1}^n y_i$. Show that if $n\overline{y} > 1$ then

$$p(z| y_1, ..., y_n) = \mathcal{B}eta(z; n\overline{y} - 1, n+1).$$

Thus we have

$$\begin{aligned} p(z| y_1, ..., y_n) &\propto \int \cdots \int p(\theta_1, \theta_2, ..., \theta_n, z| y_1, ..., y_n)\, d\theta_1 ... d\theta_n \\ &\propto \frac{(1-z)^n}{z^{n+2}} \prod_{i=1}^n \left(\int \theta_i^{y_i} \exp\left(-\frac{1}{z}\theta_i\right) d\theta_i\right) \end{aligned}$$

but

$$\begin{aligned} \int \theta_i^{y_i} \exp\left(-\frac{1}{z}\theta_i\right) d\theta_i &= \Gamma(y_i + 1)\, z^{y_i+1} \\ &= y_i! z^{y_i+1} \end{aligned}$$

so

$$\begin{aligned} p(z| y_1, ..., y_n) &\propto \frac{(1-z)^n}{z^{n+2}} \prod_{i=1}^n z^{y_i+1} \\ &\propto \frac{(1-z)^n}{z^{n+2}} z^{n\overline{y}+n} = (1-z)^n z^{n\overline{y}-2}. \end{aligned}$$

Hence the result follows if $n\overline{y} > 1$.

(e) Show that the posterior means of the θ_i, $\mathbb{E}[\theta_i| y_1, ..., y_n]$, are shrunk by a factor equal to $(n\overline{y} - 1)/(n\overline{y} + n)$ relative to the conditional means $\mathbb{E}[\theta_i| y_1, ..., y_n, z]$. (Hint: use the fact that $\mathbb{E}[\theta_i| y_1, ..., y_n] = \mathbb{E}[\mathbb{E}[\theta_i| y_1, ..., y_n, z]]$).

We have

$$\begin{aligned} p\left(\theta_i \middle| y_1, ..., y_n, z\right) &= p\left(\theta_i \middle| y_i, z\right) \\ &\propto \exp\left(-\frac{1-z}{z}\theta_i\right)\theta_i^{y_i}\exp\left(-\theta_i\right) \\ &\propto \theta_i^{y_i}\exp\left(-\frac{1}{z}\theta_i\right) \\ &= \mathcal{G}\left(\theta_i; y_i+1, z^{-1}\right). \end{aligned}$$

So we have

$$\mathbb{E}\left[\theta_i \middle| y_1, ..., y_n, z\right] = \left(y_i+1\right)z.$$

We also have because of (d) that

$$\mathbb{E}\left[z \middle| y_1, ..., y_n\right] = \frac{n\overline{y}-1}{n\overline{y}+n}$$

so

$$\begin{aligned} \mathbb{E}\left[\theta_i \middle| y_1, ..., y_n\right] &= \mathbb{E}\left[\mathbb{E}\left[\theta_i \middle| y_1, ..., y_n, z\right]\right] \\ &= \left(y_i+1\right)\frac{n\overline{y}-1}{n\overline{y}+n}. \end{aligned}$$

(f) Is the posterior distribution proper if $n\overline{y} \leq 1$?

If $n\overline{y} \leq 1$, then we have either $y_i = 0$ for all i or there exists a single index j such that $y_j = 1$.

Using results similar to (d), we obtain

$$\begin{aligned} \int_0^1 p\left(z \middle| y_1, ..., y_n\right) &\propto \int_0^1 \int \cdots \int p\left(\theta_1, \theta_2, ..., \theta_n, z \middle| y_1, ..., y_n\right) d\theta_1 ... d\theta_n \\ &\propto \int_0^1 \frac{(1-z)^n}{z^{n+2}} \prod_{i=1}^n \left(\int \theta_i^{y_i}\exp\left(-\frac{1}{z}\theta_i\right)d\theta_i\right) \end{aligned}$$

Now if all $y_i = 0$ then

$$\int \theta_i^{y_i}\exp\left(-\frac{1}{z}\theta_i\right)d\theta_i = z$$

and

$$\begin{aligned} \int_0^1 p\left(z \middle| y_1, ..., y_n\right) &\propto \int_0^1 \frac{(1-z)^n}{z^{n+2}} z^n dz \\ &\propto \int_0^1 \frac{(1-z)^n}{z^2} dz. \end{aligned}$$

This last integral is divergent. Similarly if there is exists a single index j such that $y_j = 1$ then

$$\begin{aligned} \int_0^1 p\left(z \middle| y_1, ..., y_n\right) &\propto \int_0^1 \frac{(1-z)^n}{z^{n+2}} z^{n+1} dz \\ &\propto \int_0^1 \frac{(1-z)^n}{z} dz \end{aligned}$$

which is also divergent.