Ch 2 Heavy tail error model

Thursday, February 21, 2019 10:13 PM

The full model:

$$p(y \mid \beta, \omega, \Lambda) \sim N(X\beta, (\omega\Lambda)^{-1})$$

$$\Lambda = \mathrm{Diag}(\lambda_1, \dots, \lambda_n)$$
$$\lambda_i \sim \mathrm{Gamma}\left(\tfrac{h}{2}, \tfrac{h}{2}\right) \quad \text{where } h \text{ is a fixed hyperparameter.}$$
$$(\beta \mid \omega) \sim N(m, (\omega K)^{-1})$$
$$\omega \sim \mathrm{Gamma}\left(\tfrac{d}{2}, \tfrac{\eta}{2}\right)$$

A) Under this model, what is the implied conditional distribution $p(y_i \mid X, \beta, \omega)$? Notice that λ_i has been marginalized out. This should look familiar.

The conditional distribution of y_i

$$P(y_i \mid X, \beta, \omega) = P(y_i \mid X, \beta, \omega, \lambda_i) P(\lambda_i)$$

$$\propto \int_0^\infty \sqrt{\omega\lambda_i}\, \exp\left(-\frac{\omega\lambda_i}{2}(y_i - X_i^T\beta)^2\right) \lambda_i^{h/2 - 1} \exp\left(-\frac{\lambda_i h}{2}\right) d\lambda_i$$

$$\propto \int_0^\infty \underbrace{\lambda_i^{\frac{h+1}{2}-1} \exp\left(-\frac{\lambda_i}{2}\left(\omega(y_i - X_i^T\beta)^2 + h\right)\right) d\lambda_i}_{\text{Kernel of } \mathrm{gamma}\left(\frac{h+1}{2},\, \frac{1}{2}(\omega(y_i - X_i\beta)^2 + h)\right)}$$

$$\propto \left(\frac{1}{2}\left(\omega(y_i - X_i^T\beta)^2 + h\right)\right)^{-\frac{(h+1)}{2}}$$

$$\propto \left(1 + \frac{1}{h}\frac{(y_i - X_i^T\beta)^2}{(\sqrt{\omega^{-1}})^2}\right)^{-\frac{(h+1)}{2}}$$

$$\sim T_h(X_i^T\beta, \sqrt{\omega^{-1}})$$, a t dist'n with h degrees of freedom, centered at $X_i^T\beta$, with $\sqrt{\omega^{-1}}$ scale

B) What is the conditional posterior distribution $p(\lambda_i \mid y_i, \beta, \omega)$?

To find the conditional posterior, I will first find the full posterior. We have

$$p(\lambda_i, \beta, \omega \mid y_i) \propto p(y_i \mid \beta, \omega, \lambda_i)\, p(\beta \mid \omega)\, p(\omega)\, p(\lambda_i)$$

$$\propto \sqrt{\frac{\omega\lambda_i}{2\pi}} \exp\left(-\frac{\omega\lambda_i}{2}(y_i - X_i^T\beta)^2\right) \sqrt{\frac{\omega K}{2\pi}} \exp\left(-\frac{\omega K}{2}(\beta - m)^2\right) \omega^{\frac{d}{2}-1} \exp\left(-\frac{\omega\eta}{2}\right) \lambda_i^{\frac{h}{2}-1} \exp\left(-\frac{\lambda_i h}{2}\right) \quad (1)$$

To find $p(\lambda_i \mid y_i, \beta, \omega)$, we can set y_i, β, and ω to constants and only work with the pieces of the posterior that involve λ_i. Thus, (1) will simplify to

$$p(\lambda_i \mid y_i, \beta, \omega) \propto \sqrt{\frac{\omega\lambda_i}{2\pi}} \exp\left(-\frac{\omega\lambda_i}{2}(y_i - X_i^T\beta)^2\right) \lambda_i^{\frac{h}{2}-1} \exp\left(-\frac{\lambda_i h}{2}\right)$$

$$\propto \lambda^{\frac{h+1}{2}-1} \exp\left(-\lambda_i\left(\frac{1}{2}\left(\omega(y_i - X_i^T\beta)^2 + h\right)\right)\right)$$

$$\sim \mathrm{gamma}\left(\frac{h+1}{2}, \frac{1}{2}\left(\omega(y_i - X_i^T\beta)^2 + h\right)\right)$$

C) Code up a gibbs sampler that repeatedly cycles through

c) Code up a gibbs sampler that repeatedly cycles through

$$p(\beta \mid y, \omega, \Lambda)$$

$$p(\omega \mid y, \Lambda)$$

$$p(\lambda_i \mid y, \beta, \omega)$$

The first two should follow identically from your previous results, except that we are explicitly conditioning on Λ, which is a random variable rather than a fixed hyperparameter.

$$p(\lambda_i, \beta, \omega \mid y_i) \propto \sqrt{\tfrac{\omega\lambda_i}{2\pi}} \exp\left(-\tfrac{\omega\lambda_i}{2}(y_i - x_i^T\beta)^2\right) \sqrt{\tfrac{\omega K}{2\pi}} \exp\left(-\tfrac{\omega K}{2}(\beta - m)^2\right) \omega^{\frac{d}{2}-1} \exp\left(-\tfrac{\omega \eta}{2}\right) \lambda_i^{\frac{h}{2}-1} \exp\left(-\tfrac{\lambda_i h}{2}\right) \quad (1)$$

From here, we can see that if we were to find the conditional posterior of β and ω, we can simply disregard the prior for λ_i since both the $p(\beta \mid \omega)$ and $p(\omega)$ are independent of λ_i. Thus, the conditional posterior $p(\beta, \omega \mid y, \Lambda)$ is:

$$p(\beta, \omega \mid y, \Lambda) \propto \frac{\omega^{\frac{d}{2}-1}}{|(\omega\Lambda)^{-1}|^{1/2} \, |(\omega K)^{-1}|^{1/2}} \exp\left(-\tfrac{\omega}{2}(\beta - A^{-1}b)^T A (\beta - A^{-1}b)\right) \exp\left(-\tfrac{\omega}{2}(\eta - b^T A^{-1} b + y^T \Lambda y + m^T K m)\right)$$

where $A = (X^T X + K)$
$b = (y\Lambda x + m^T K)$

Thus, $p(\beta \mid y, \omega, \Lambda) \propto \exp\left(-\tfrac{\omega}{2}(\beta - A^{-1}b)^T A (\beta - A^{-1}b)\right)$

$$\sim MVN(A^{-1}b, (\omega A)^{-1})$$

$$p(\omega \mid y, \beta, \Lambda) \propto \omega^{\frac{d+n+p}{2}-1} \exp\left[-\tfrac{\omega}{2}\left((\beta - A^{-1}b)^T A (\beta - A^{-1}b) + (\eta + b^T A^{-1} b + y^T \Lambda y + m^T K m)\right)\right]$$

$$\sim Gam\left(\tfrac{d+n+p}{2}, \tfrac{1}{2}\left((\beta - A^{-1}b)^T A (\beta - A^{-1}b) + (\eta + b^T A^{-1} b + y^T \Lambda y + m^T K m)\right)\right)$$

$$p(\lambda_i \mid y, \beta, \omega) \propto \lambda^{\frac{h+1}{2}-1} \exp\left(-\lambda_i \left(\tfrac{1}{2}(\omega(y_i - x_i^T\beta)^2 + h)\right)\right)$$

$$\sim gamma\left(\tfrac{h+1}{2}, \tfrac{1}{2}(\omega(y_i - x_i^T\beta)^2 + h)\right)$$

The Gibbs sampler algorithm is:

1.) Start with some $(\beta, \omega, \Lambda)^{(0)}$

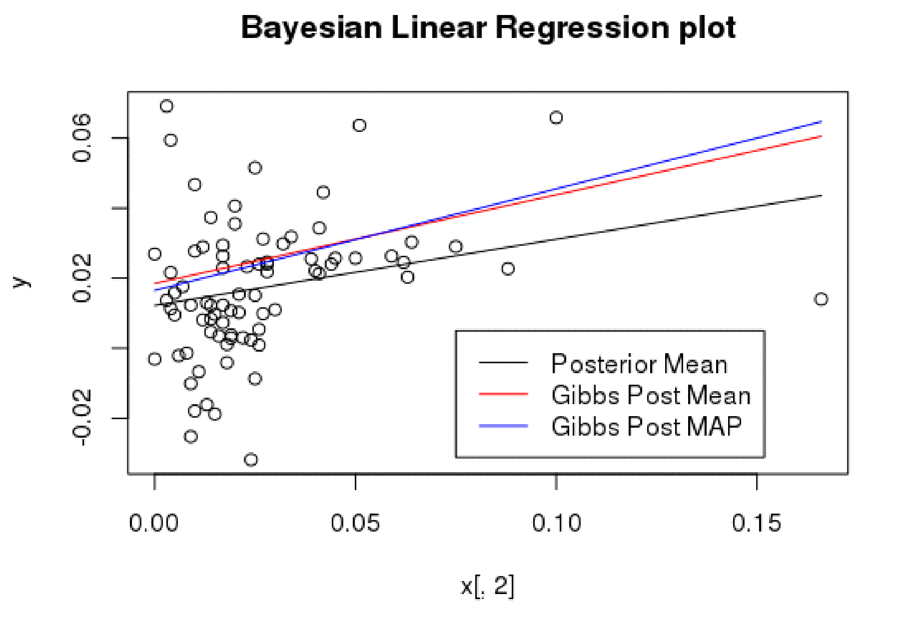
2.) At each iteration t, for each $j = 1, \dots, p$, sample $(\beta^{(t)}, \omega^{(t)}, \Lambda^{(t)})$ from

$$p(\beta^{(t)} \mid \omega^{(t-1)}, \Lambda^{(t-1)})$$

$$p(\omega^{t} \mid \beta^{(t)}, \Lambda^{(t-1)})$$

$$p(\lambda_i^{(t)} \mid \beta^{(t)}, \omega^{(t)}, \lambda_1^{(t)}, \dots, \lambda_{i-1}^{(t)}, \lambda_{i+1}^{(t-1)}, \dots, \lambda_p^{(t-1)})$$

We can discard the first thousand draws as the burn in and draw an additional 3000 samples.



The hyper parameters I used for this simulation were:

$$d = 2$$
$$\eta = 2$$
$$h = 2$$
$$m = (0.02, 0.1)$$
$$K = \begin{bmatrix} 0.01 & 0 \\ 0 & 0.01 \end{bmatrix}$$

After burning 1000 samples, I thinned the remaining 3000 samples by a factor of 3. To compute the red fitted line, I took the average of my sampled betas. As you can see from the figure above, the red line has a slightly higher slope and intercept than the original Bayesian linear model. Considering the true fit of the model should have a very steep slope, the averaged gibbs result is a little bit better. However, it still is influenced by the outliers a lot. The MAP estimates for beta again produce a slightly better result but it still has the same problem. In general, the MAP estimate was slightly better in all cases. Tuning the hyper parameters showed that the model is very sensitive to changes in m. This makes sense as m holds our prior beliefs about what β_0 and β_1 should be.