Paper Review

Md Kamrul Hasan Khan

May 8, 2017

The Bayesian Elastic Net
Qing Li * *and Nan Lin* †

In 2010 Li and Lin proposed a Bayesian analysis of the Elastic Net problem and solve the problem using a Gibbs sampler. Model hierachy is given below:

$y|\beta,\sigma^2 \sim \mathcal{N}_{p+1}(X\beta,\sigma^2 I_n)$

$\beta|\sigma^2 \sim exp\{-\lambda_1||\beta||_1 - \lambda_2||\beta||_2^2\}$

$\sigma^2 \sim \frac{1}{\sigma^2}$

The above model is difficult to solve directly through a Gibbs sampler because the absolute values $|\beta_j|$'s in the prior would yield unfamiliar full conditional distributions. Then introduing a latent variable τ, they propose another hierarcical model:

$y|\beta,\sigma^2 \sim \mathcal{N}(\mathbf{X}\beta,\sigma^2\mathbf{I}_n)$

$\beta|\tau,\sigma^2 \sim \prod_{j=1}^p \mathcal{N}\Big(0, \Big(\frac{\lambda_2}{\sigma^2}\frac{\tau_j}{\tau_j-1}\Big)^{-1}\Big)$

$\tau|\sigma^2 \sim \prod_{j=1}^p \mathcal{TG}\Big(\frac{1}{2}, \frac{8\lambda_2\sigma^2}{\lambda_1^2}.(1,\infty)\Big)$

$\sigma^2 \sim \frac{1}{\sigma^2}$

Now the mdoel becomes computationally easier to solve since the full conditional distribtutions now are

$\beta|y,\sigma^2,\tau \sim \mathcal{N}_p(A^{-1}X^Ty, \sigma^2A^{-1})$

where $A = X^TX + \lambda_2 diag(\frac{\tau_1}{\tau_1-1},...,\frac{\tau_p}{\tau_p-1})$

$(\boldsymbol{\tau}-\mathbf{1_p})|\boldsymbol{y},\sigma^2,\beta \sim \prod_{j=1}^p GIG\Big(\lambda=\frac{1}{2}, \psi=\frac{\lambda_1}{4\lambda_2\sigma^2}, \chi=\frac{\lambda_2\beta_j^2}{\sigma^2}\Big)$

$\sigma^2|y,\beta,\beta_0,\tau \sim \frac{1}{\sigma^2}^{\frac{n}{2}+p+1}\Big\{\Gamma U\Big(\frac{1}{2},\frac{\lambda_1^2}{8\sigma^2\lambda_2}\Big)\Big\}^{-p} \times exp\Big[-\frac{1}{2\sigma^2}\Big\{||y-X\beta||_2^2 + \lambda_2\sum_{j=1}^p \frac{\tau_j}{\tau_j-1}\beta_j^2 + \frac{\lambda_1^2}{4\lambda_2}\sum_{j=1}^p \tau_j\Big\}\Big]$

Sampling from the full conditional distribution

It is straight forward to sample from $\beta|y,\sigma^2,\tau$.

For τ instead of sampling from generalized inverse Gaussian, they sample it in the following way:

$\frac{1}{(\boldsymbol{\tau}-\mathbf{1_p})} \sim \prod_{j=1}^p \mathcal{IG}(\mu=\frac{\sqrt{(\lambda_1)}}{(2\lambda_2)|\beta_j|}, \lambda=\frac{\lambda_1}{4\lambda_2\sigma^2})$

where μ is the mean and λ is the shape parameter.

To sample $\sigma^2|\boldsymbol{Y},\boldsymbol{\beta},\boldsymbol{\tau}$they suggested acceptance-rejectaion algorithm. Their description is given below:

Denote the function of σ^2 on the right-hand side of the prosterior distribution as $f(\sigma^2)$. Then by the definition of incomplete gamma functions,

$$f(\sigma^2) \leq \Gamma\Big(\frac{1}{2}\Big)^{-p}\Big(\frac{1}{\sigma^2}\Big)^{a+1} exp\Big\{\frac{1}{\sigma^2}b\Big\} = \frac{\Gamma(a)\Gamma\Big(\frac{1}{2}\Big)^{-p}}{b^a}h(\sigma^2)$$

where $h(\cdot)$ is the pdf for inverse-gamma (a, b) and

$$a = \frac{n}{2} + p, b = \frac{1}{2}\Bigg[\boldsymbol{(Y - X\beta)^T(Y - X\beta)} + \lambda_2 \textstyle\sum_{j=1}^p \frac{\tau_j}{\tau_j - 1}\beta_j^2 + \frac{\lambda_1^2}{4\lambda_2^2}\sum_{j=1}^p \tau_j\Bigg]$$

To ger σ^2 from $f(\sigma^2)$, we first generate a candidate Z from with h and a u from uniform(0,1) and then accept Z if $u \leq \Gamma\Big(\frac{1}{2}\Big)^p b^a f(Z)/\Gamma(a)h(Z)$ or equivalentlty, if $log(u) \leq plog(\Gamma(\frac{1}{2})) - plog\Gamma_U(\frac{1}{2}, \frac{{\lambda_1}^2}{8Z\lambda_2})$

Varibale Selection

In this paper authors described two criterion for variable selection. (i) Credible interval criterion, (ii) Scaled neighborhood criterion.

(i) Credible interval criterion: A predictor $\boldsymbol{x_j}$ is excluded if the credible interval of β_j covers 0 and is retained otherwise.

(ii) Scaled neighborhood criterion: Consider the posterior probability in $\Big[-\sqrt{var(\beta_j|\boldsymbol{y})}, \sqrt{var(\beta_j|\boldsymbol{y})}\Big]$. A predictor is excluded if the posterior probability exceeds a certain probability threshold and retained otherwise.

They suggested to take the level credible interval as 0.5 and the probablity of threshold in the scaled neighborhood criterion as 0.5.

My Suggestion

In this method authors suggested to use acceptance-rejectaion algorithm to sample from $\sigma^2|\boldsymbol{Y, \beta, \tau}$. That means all σ^2 will not be accepted. But instead of using acceptance-rejectaion algorithm we can use Slice sampling to sample from $\sigma^2|\boldsymbol{Y, \beta, \tau}$. The description is given below:

$u|\sigma^2 \sim unif(0, h(\sigma^2))$ where $\sigma^2 = \Big\{\Gamma_U\Big(\frac{1}{2}, \frac{{\lambda_1}^2}{8\sigma^2\lambda_2}\Big)\Big\}$

$\sigma^2|u \sim \mathcal{TIG}(a, b)I_{\{0,c\}}(\sigma^2)$

where, $a = \frac{n}{2} + p$,

$$b = \frac{1}{2}\Bigg[\boldsymbol{(Y - X\beta)^T(Y - X\beta)} + \lambda_2 \textstyle\sum_{j=1}^p \frac{\tau_j}{\tau_j - 1}\beta_j^2 + \frac{\lambda_1^2}{4\lambda_2^2}\sum_{j=1}^p \tau_j\Bigg] \text{ and}$$

$$c = \frac{{\lambda_1}^2}{8\lambda_2 F^{-1}\left(\frac{1 - exp\left\{\frac{-log(u)}{p}\right\}}{\sqrt{\pi}}\right)}$$, where F^{-1} is the inverse cdf of Gamma Distribution.

Future work

Using MCMC diagnosis I will check which method, acceptance-rejectaion algorithm or Slice sampling, works better here.