Approaches for Bayesian Variable Selection (SSVS)

Shiqiang Jin

4-17-2017

Contents

Chapter 1

Foreword

I am Caleb Jin. After I read this paper, **Approaches for Bayesian Variable Selection (SSVS)**(George and McCulloch, 1997) and (George and McCulloch, 1993), I write down the nodes of the key idea and R code to realize it.

Chapter 2

Approaches for Bayesian Variable Selection (SSVS)

Consider a high dimensional linear regression model as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{2.1}$$

where $\mathbf{y} = (y_1, y_2, \ldots, y_n)^\top$ is the n-dimensional response vector, $\mathbf{X} = [1, \mathbf{M}] = [\mathbf{x}_1, \ldots, \mathbf{x}_p]$ is the $n \times p$ design matrix, and $\boldsymbol{\epsilon} \sim \mathcal{N}_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$. Note that as $p > n$, $\mathbf{X}$ is not full rank.

From (2.1), the likelihood function is given as

$$\mathbf{y}|\boldsymbol{\beta}, \sigma^2 \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta}, \sigma^2 I_n).$$

We define the prior as follows:

$$\beta_j|\sigma^2, \gamma_j \overset{ind}{\sim} (1-\gamma_j)\mathcal{N}(0, \sigma^2\nu_0) + \gamma_j\mathcal{N}(0, \sigma^2\nu_1),$$

where ν_0 and ν_1 will be chosen to be small and large, respectively. Note that the likelihood is independent of $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_p)$. Assume

$$\sigma^2 \sim \mathcal{IG}(\frac{a}{2}, \frac{b}{2}),$$

which is also independent of $\boldsymbol{\gamma}$. We consider

$$\gamma_j \overset{iid}{\sim} Ber(\omega).$$

To make our model robust to the choice of ω, we will assign the following prior on ω.

$$w \sim \mathcal{B}(c_1, c_2),$$

where we will use $c_1 = c_2 = 1$, which leads to the uniform distribution. Recall the density function of beta distribution is proportional $\pi(w) \propto w^{c_1-1}(1-w)^{c_2-1}$.

It is easy to show that the **full conditionals** are as follows:

- 1)

$$\beta_j|\boldsymbol{\beta}_{-j}, \sigma^2, \boldsymbol{\gamma}, \mathbf{y} \sim \mathcal{N}(\tilde{\beta}_j, \frac{\sigma^2}{\mu_j}).$$

  where $\mu_j = x_j^\top x_j + \frac{1}{\nu_{\gamma_j}}$ and $\tilde{\beta}_j = \mu_j^{-1} x_j^\top(\mathbf{y} - \mathbf{X}_{-j}\boldsymbol{\beta}_{-j})$.

- 2)

$$\gamma_j|\boldsymbol{\gamma}_{-j}, \boldsymbol{\beta}, \sigma^2, \mathbf{y} \sim Ber\left(\frac{\nu_1^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2\nu_1}\beta_j^2\right)\omega}{\nu_0^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2\nu_0}\beta_j^2\right)(1-\omega) + \nu_1^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2\nu_1}\beta_j^2\right)\omega}\right).$$

- 3)

$$\sigma^2|\boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{y} \sim \mathcal{IG}(a^*, b^*),$$

  where $a^* = \frac{1}{2}(n+p+a)$ and $b^* = \frac{1}{2}\left(\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 + \sum_{j=1}^p \frac{\beta_j^2}{\nu_{\gamma_j}} + b\right)$.

- 4)

$$w|\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma^2, \mathbf{y} \sim \mathcal{B}\left(\sum_{j=1}^p \gamma_j + c_1, p - \sum_{j=1}^p \gamma_j + c_2\right).$$

To speed up, we consider the following conditionals:

- 1')

$$\boldsymbol{\beta}|\sigma^2, \boldsymbol{\gamma}, \mathbf{y} \sim \mathcal{N}(\tilde{\boldsymbol{\beta}}, \sigma^2(\mathbf{X}^\top\mathbf{X} + \mathbf{V}_{\boldsymbol{\gamma}}^{-1})^{-1}),$$

  where $\mathbf{V}_{\boldsymbol{\gamma}} = \mathrm{diag}(v_{\gamma_j})_{j=0}^p$ and $\tilde{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X} + \mathbf{V}_{\boldsymbol{\gamma}}^{-1})^{-1}\mathbf{X}^\top\mathbf{y}$.

- 2')

$$\boldsymbol{\gamma}|\boldsymbol{\beta}, \sigma^2, \mathbf{y} \sim \prod_{j=0}^p Ber\left(\frac{\nu_1^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2\nu_1}\beta_j^2\right)\omega}{\nu_0^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2\nu_0}\beta_j^2\right)(1-\omega) + \nu_1^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2\nu_1}\beta_j^2\right)\omega}\right).$$

- 3')

$$\sigma^2|\boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{y} \sim \mathcal{IG}(a^*, b^*).$$

  where $a^* = \frac{1}{2}(n+p+a)$ and $b^* = \frac{1}{2}\left(\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 + \sum_{j=1}^p \frac{\beta_j^2}{\nu_{\gamma_j}} + b\right)$.

- 4')

$$w|\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma^2, \mathbf{y} \sim \mathcal{B}\left(\sum_{j=1}^p \gamma_j + c_1, p - \sum_{j=1}^p \gamma_j + c_2\right).$$

2.1 Rcode

```
library(invgamma)
p <- 100
n = 100
power <- numeric()
e<-rnorm(n, mean = 0, sd = sqrt(2)) # error
X<-matrix(data = rnorm(n * p, 0, 1), nrow = n, ncol = p)
Beta <- c(1, 2, rep(0,(p - 3)), 3)# exclude intercept beta0
y <- X %*% Beta + e # true model
true.gamma<-as.numeric(Beta != 0)
#setup for initial values#####
hat.beta <- as.numeric(solve(t(X) %*% (X) + diag(1, p)) %*% t(X) %*% y) #p-dim vector
hat.gamma <- rep(1, p)
hat.sig2 <- mean((y - X %*% hat.beta)^2)
#setup for priors ############
w <- 0.5
v0 <- 0.001
v1 <- 1000
v01 <- c(v0, v1)
a0 <- 1
b0 <- 1
###############################
MC.size <- 2000 + 3000
hat.BETA <- matrix(0, MC.size, p) # to store beta for each iteration
hat.Gamma <- matrix(0, MC.size, p) # to store z for each iteration
hat.Sig2 <- rep(0, MC.size) # to store variance for each iteration
for (goh in 1:MC.size) {
  # Gibbs sampling
  # 1) for beta_j
  for (j in 1:p) {
    mu_j <- t(X[, j]) %*% X[, j] + 1/v01[(hat.gamma[j] + 1)]
    y.star <- y - X[, -j] %*% hat.beta[-j]
    tilde.beta.j <- as.numeric((1/mu_j) * t(X[, j]) %*% y.star)
    var.beta <- as.numeric(hat.sig2/mu_j)
    hat.beta[j] <- rnorm(1, tilde.beta.j, sqrt(var.beta))  #sampling from beta_j|others
  }
  # 2) gamma_j
  p.j <- dnorm(hat.beta, 0, sqrt(v1 * hat.sig2)) * w
  q.j <- dnorm(hat.beta, 0, sqrt(v0 * hat.sig2)) * (1 - w)
  prob.j <- p.j/(p.j + q.j)
  hat.gamma <- rbinom(p, 1, prob.j)
  hat.Gamma[goh, ] <- hat.gamma
  hat.BETA[goh, ] <- hat.beta
  # 3) sig2
```

```r
  a.star <- 1/2 * (n + p + a0)
  v.z_j <- hat.gamma * v1 + (1 - hat.gamma) * v0
  b.star <- 1/2 * (sum((y - X %*% hat.beta)^2) + sum(hat.beta^2/v.z_j) + b0)
  hat.sig2 <- rinvgamma(1, shape = a.star, rate = b.star)
  par(mfrow=c(1,1))
  plot(hat.gamma, main = paste("rep:", goh))
  points(true.gamma, col = 2, pch = "*")
}
colMeans(hat.Gamma)>0.5
```

2.2 Appendix

- For $\beta_j|\boldsymbol{\beta}_{-j},\sigma^2,\boldsymbol{\gamma},\mathbf{y}$ $(j=1,2,\dots,p)$, we have

$$\begin{aligned}
\pi(\beta_j|\boldsymbol{\beta}_{-j},\sigma^2,\boldsymbol{\gamma},\mathbf{y}) &\propto f(\mathbf{y}|\boldsymbol{\beta},\sigma^2)\pi(\boldsymbol{\beta}|\boldsymbol{\gamma},\sigma^2)\\
&= f(\mathbf{y}|\boldsymbol{\beta}_{\boldsymbol{\gamma}},\sigma^2)\prod_{k=1}^{p}\pi(\beta_k|Z_k,\sigma^2)\\
&\propto f(\mathbf{y}|\boldsymbol{\beta},\sigma^2)\pi(\beta_j|\gamma_j,\sigma^2)\\
&\propto \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\|^2\right)\exp\left(-\frac{1}{2\sigma^2\nu_{\gamma_j}}\beta_j^2\right)\\
&= \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{y}-\mathbf{X}_{-j}\boldsymbol{\beta}_{-j}-x_j\beta_j\|^2\right)\exp\left(-\frac{1}{2\sigma^2\nu_{\gamma_j}}\beta_j^2\right)\\
&= \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{y}^*-x_j\beta_j\|^2\right)\exp\left(-\frac{1}{2\sigma^2\nu_{\gamma_j}}\beta_j^2\right)\\
&= \exp\left[-\frac{1}{2\sigma^2}\left(\mathbf{y}^{*\top}\mathbf{y}^*-2\beta_j x_j^\top\mathbf{y}^*+\beta_j x_j^\top x_j\beta_j+\frac{1}{\nu_{\gamma_j}}\beta_j^2\right)\right]\\
&\propto \exp\left[-\frac{1}{2\sigma^2}\left(\beta_j^2(x_j^\top x_j+\frac{1}{\nu_{\gamma_j}})-2\beta_j x_j^\top\mathbf{y}^*\right)\right]\\
&= \exp\left[-\frac{1}{2\sigma^2}\left(\beta_j^2(x_j^\top x_j+\frac{1}{\nu_{\gamma_j}})-2\beta_j(x_j^\top x_j+\frac{1}{\nu_{\gamma_j}})(x_j^\top x_j+\frac{1}{\nu_{\gamma_j}})^{-1}x_j^\top\mathbf{y}^*\right)\right]\\
&= \exp\left[-\frac{a}{2\sigma^2}\left(\beta_j^2-2a\beta_j\tilde{\beta}_j\right)\right]\\
&\propto \exp\left[-\frac{a}{2\sigma^2}\left(\beta_j-\tilde{\beta}_j\right)^2\right]
\end{aligned}$$

where $\mathbf{y}^*=\mathbf{y}-\mathbf{X}_{-j}\boldsymbol{\beta}_{-j}$, $\quad \mu_j=x_j^\top x_j+\frac{1}{\nu_{\gamma_j}}$, $\quad \tilde{\beta}_j=\mu_j^{-1}x_j^\top\mathbf{y}^*$. Hence,

$$\beta_j|\boldsymbol{\beta}_{-j},\sigma^2,\boldsymbol{\gamma},\mathbf{y}\sim\mathcal{N}(\tilde{\beta}_j,\frac{\sigma^2}{\mu_j}).$$

- For $\pi(\gamma_j|\boldsymbol{\gamma}_{-j},\boldsymbol{\beta},\sigma^2,\mathbf{y})$ $(j=1,2,\ldots,p)$ we have

$$\begin{aligned}
\pi(\gamma_j|\boldsymbol{\gamma}_{-j},\boldsymbol{\beta},\sigma^2,\mathbf{y}) &\propto \pi(\boldsymbol{\gamma},\boldsymbol{\beta}|\sigma^2)\\
&= \pi(\boldsymbol{\beta}|\sigma^2,\boldsymbol{\gamma})\pi(\boldsymbol{\gamma})\\
&= \prod_{k=1}^{p}\pi(\beta_k|\sigma^2,z_k)\prod_{i=1}^{p}\pi(z_i)\\
&\propto \pi(\beta_j|\sigma^2,z_j)\pi(\gamma_j)\\
&\propto \nu_{\gamma_j}^{-\frac{1}{2}}\exp\left(-\frac{1}{2\sigma^2\nu_{\gamma_j}}\beta_j^2\right)\omega^{\gamma_j}(1-\omega)^{1-\gamma_j}.
\end{aligned}$$

Note that

$$\begin{aligned}
\pi(\gamma_j=0|\boldsymbol{\gamma}_{-j},\boldsymbol{\beta},\sigma^2,\mathbf{y}) &= C\nu_0^{-\frac{1}{2}}\exp\left(-\frac{1}{2\sigma^2\nu_0}\beta_j^2\right)(1-\omega)\,;\\
\pi(\gamma_j=1|\boldsymbol{\gamma}_{-j},\boldsymbol{\beta},\sigma^2,\mathbf{y}) &= C\nu_1^{-\frac{1}{2}}\exp\left(-\frac{1}{2\sigma^2\nu_1}\beta_j^2\right)\omega.
\end{aligned}$$

This implies that

$$\begin{aligned}
\pi(\gamma_j=1|\boldsymbol{\gamma}_{-j},\boldsymbol{\beta},\sigma^2,\mathbf{y}) &= \frac{P(\gamma_j=1,\Omega)}{\sum_{\gamma_j}P(\gamma_j,\Omega)}\\
&= \frac{C\nu_1^{-\frac{1}{2}}\exp\left(-\frac{1}{2\sigma^2\nu_1}\beta_j^2\right)\omega}{C\nu_0^{-\frac{1}{2}}\exp\left(-\frac{1}{2\sigma^2\nu_0}\beta_j^2\right)(1-\omega)+C\nu_1^{-\frac{1}{2}}\exp\left(-\frac{1}{2\sigma^2\nu_1}\beta_j^2\right)\omega}\\
&= \frac{\nu_1^{-\frac{1}{2}}\exp\left(-\frac{1}{2\sigma^2\nu_1}\beta_j^2\right)\omega}{\nu_0^{-\frac{1}{2}}\exp\left(-\frac{1}{2\sigma^2\nu_0}\beta_j^2\right)(1-\omega)+\nu_1^{-\frac{1}{2}}\exp\left(-\frac{1}{2\sigma^2\nu_1}\beta_j^2\right)\omega}.
\end{aligned}$$

Hence,

$$\gamma_j|\boldsymbol{\gamma}_{-j},\boldsymbol{\beta},\sigma^2,\mathbf{y}\sim Ber\left(\frac{\nu_1^{-\frac{1}{2}}\exp\left(-\frac{1}{2\sigma^2\nu_1}\beta_j^2\right)\omega}{\nu_0^{-\frac{1}{2}}\exp\left(-\frac{1}{2\sigma^2\nu_0}\beta_j^2\right)(1-\omega)+\nu_1^{-\frac{1}{2}}\exp\left(-\frac{1}{2\sigma^2\nu_1}\beta_j^2\right)\omega}\right).$$

- For $\sigma^2|\boldsymbol{\beta},\boldsymbol{\gamma},\mathbf{y}$, we have

$$\begin{aligned}
\pi(\sigma^2|\boldsymbol{\beta},\boldsymbol{\gamma},\mathbf{y}) &\propto f(\mathbf{y}|\boldsymbol{\beta},\sigma^2,\boldsymbol{\gamma})\pi(\boldsymbol{\beta},\sigma^2|\boldsymbol{\gamma}) \\
&= f(\mathbf{y}|\boldsymbol{\beta},\sigma^2,\boldsymbol{\gamma})\pi(\boldsymbol{\beta}|\sigma^2,\boldsymbol{\gamma})\pi(\sigma^2|\boldsymbol{\gamma}) \\
&= f(\mathbf{y}|\boldsymbol{\beta},\sigma^2)\pi(\boldsymbol{\beta}|\sigma^2,\boldsymbol{\gamma})\pi(\sigma^2) \\
&\propto (\sigma^2)^{-\frac{n}{2}}\exp\left(-\frac{1}{2\sigma^2}\|\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\|^2\right)\times\prod_{j=1}^{p}\left[(\sigma^2\nu_{\gamma_j})^{-\frac{1}{2}}\exp\left(-\frac{1}{2\sigma^2\nu_{\gamma_j}}\beta_j^2\right)\right] \\
&\quad \times(\sigma^2)^{-\frac{a}{2}-1}\exp\left(-\frac{b}{2\sigma^2}\right) \\
&\propto (\sigma^2)^{-\frac{n}{2}}\exp\left(-\frac{1}{2\sigma^2}\|\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\|^2\right)\times(\sigma^2)^{-\frac{p}{2}}\exp\left(-\frac{1}{2\sigma^2}\sum_{j=1}^{p}\frac{\beta_j^2}{\nu_{\gamma_j}}\right) \\
&\quad \times(\sigma^2)^{-\frac{a}{2}-1}\exp\left(-\frac{b}{2\sigma^2}\right) \\
&= (\sigma^2)^{-\frac{1}{2}(n+p+a)-1}\exp\left(-\frac{\frac{1}{2}\left(\|\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\|^2+\sum_{j=1}^{p}\frac{\beta_j^2}{\nu_{\gamma_j}}+b\right)}{\sigma^2}\right) \\
&= (\sigma^2)^{-a^*-1}\exp(-\frac{b^*}{\sigma^2}),
\end{aligned}$$

where $a^*=\frac{1}{2}(n+p+a)$ and $b^*=\frac{1}{2}\left(\|\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\|^2+\sum_{j=1}^{p}\frac{\beta_j^2}{\nu_{\gamma_j}}+b\right)$.

Therefore,

$$\sigma^2|\boldsymbol{\beta},\boldsymbol{\gamma},\mathbf{y}\sim\mathcal{IG}(a^*,b^*).$$

- For $w|\boldsymbol{\beta},\boldsymbol{\gamma},\sigma^2,\mathbf{y}$, we have

$$\begin{aligned}
\pi(w|\boldsymbol{\beta},\boldsymbol{\gamma},\sigma^2,\mathbf{y}) &\propto \pi(\boldsymbol{\gamma}|w)\pi(w) \\
&\propto \left[\prod_{j=1}^{p}w^{\gamma_j}(1-w)^{1-\gamma_j}\right]w^{c_1-1}(1-w)^{c_2-1} \\
&\propto w^{\sum_{j=1}^{p}\gamma_j+c_1-1}(1-w)^{p-\sum_{j=1}^{p}\gamma_j+c_2-1}.
\end{aligned}$$

We therefore have

$$w|\boldsymbol{\beta},\boldsymbol{\gamma},\sigma^2,\mathbf{y}\sim\mathcal{B}\left(\sum_{j=1}^{p}\gamma_j+c_1,p-\sum_{j=1}^{p}\gamma_j+c_2\right).$$

Bibliography

George, E. I. and McCulloch, R. E. (1993). Variable selection via gibbs sampling. *Journal of the American Statistical Association*, pages 881–889.

George, E. I. and McCulloch, R. E. (1997). Approaches for bayesian variable selection. *Statistica Sinica*, 7(2):339–373.