

EMVS: The EM Approach to Bayesian Variable Selection

Caleb Jin

12-15-2017

Contents

1	Foreword	5
2	The EM Approach to Bayesian Variable Selection	7
2.1	Introduction	7
2.2	Stochastic Search Variable Selection (SSVS)	7
2.3	EMVS	9
2.4	Simulation Case Study	13

Chapter 1

Foreword

I am Caleb Jin. After I read this paper, **EMVS: The EM Approach to Bayesian Variable Selection**(Rockova and George, 2014), I write down the nodes of the key idea and R code to realize it.

Chapter 2

The EM Approach to Bayesian Variable Selection

2.1 Introduction

The **EMVS** method is anchored by **EM** algorithm and original stochastic search variable selection (**SSVS**). It is a deterministic alternative to **MCMC** stochastic search and ideally suited for high-dimensional $p > n$ settings. Furthermore, **EMVS** is able to effectively identify the sparse high-probability model and find candidate models within a fraction of the time needed by **SSVS**. This algorithm makes it possible to carry out dynamic posterior exploration for the identification of posterior modes.

The outline of this course project is as follows. In section 2, we build original **SSVS** method, including its prior and posterior distribution for each parameter. In section 3, I derive the author's new idea **EMVS** method in very details that the author omits. In section 4, I perform several simulation case studies from this paper and me, then compare the results of **SSVS** and **EMVS**.

2.2 Stochastic Search Variable Selection (SSVS)

2.2.1 Model Settings

Consider a multiple linear regression model as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (2.1)$$

where $\mathbf{y} = (y_1, \dots, y_n)^\top$ is the n -dimensional response vector, $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]^\top$ is the $n \times p$ design matrix, $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)'$, σ^2 is a scalar, and $\boldsymbol{\epsilon} \sim$

$\mathcal{N}_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$. Both σ^2 and β are considered to be unknown. We assume that $p > n$. From (2.1), the likelihood function is given as

$$\mathbf{y}|\beta, \sigma^2 \sim \mathcal{N}(\mathbf{X}\beta, \sigma^2 \mathbf{I}_n). \quad (2.2)$$

2.2.2 Prior Specification

The cornerstone of **SSVS** is the “spike and slab” Gaussian mixture prior on β ,

$$\beta_j|\sigma^2, \gamma_j \stackrel{ind}{\sim} (1 - \gamma_j)\mathcal{N}(0, \sigma^2 \nu_0) + \gamma_j\mathcal{N}(0, \sigma^2 \nu_1), \quad (2.3)$$

where

$$\gamma_j = \begin{cases} 1 & \beta_j \neq 0 \\ 0 & \beta_j = 0, \end{cases}$$

$j = 1, 2, \dots, p$. $0 \leq \nu_0 < \nu_1$. Hence, $\beta|\sigma^2, \gamma \sim \mathcal{N}_p(\mathbf{0}, \mathbf{D}_{\sigma^2, \gamma})$, where $\mathbf{D}_{\sigma^2, \gamma} = \sigma^2 \text{diag}(\nu_{z_1}, \nu_{z_1}, \dots, \nu_{z_p})$. For the prior on the σ^2 , I follow the GM97 and use inverse gamma prior but different notation,

$$\sigma^2 \sim \mathcal{IG}(\frac{a}{2}, \frac{b}{2}), \quad (2.4)$$

where we will use $a = 1$ and $b = 1$, and σ^2 is also independent of γ . I follow the GM97 and use Bernoulli prior form but different notation,

$$\gamma_j|\theta \stackrel{iid}{\sim} \text{Ber}(\theta), \theta \in [0, 1]. \quad (2.5)$$

Hence,

$$\gamma \propto \theta^{\sum_{j=1}^p \gamma_j} (1 - \theta)^{p - \sum_{j=1}^p \gamma_j} \quad (2.6)$$

We assume

$$\theta \sim \mathcal{B}(c_1, c_2), \quad (2.7)$$

where we will use $c_1 = c_2 = 1$, which leads to the uniform distribution.

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

- 1)

$$\beta_j|\beta_{-j}, \sigma^2, \gamma, \nu_0, \mathbf{y} \sim \mathcal{N}(\tilde{\beta}_j, \frac{\sigma^2}{\mu_j}).$$

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

- 2)

$$\gamma_j|\gamma_{-j}, \beta, \sigma^2, \mathbf{y} \sim \text{Ber} \left(\frac{\nu_1^{-\frac{1}{2}} \exp \left(-\frac{1}{2\sigma^2 \nu_1} \beta_j^2 \right) \theta}{\nu_0^{-\frac{1}{2}} \exp \left(-\frac{1}{2\sigma^2 \nu_0} \beta_j^2 \right) (1 - \theta) + \nu_1^{-\frac{1}{2}} \exp \left(-\frac{1}{2\sigma^2 \nu_1} \beta_j^2 \right) \theta} \right). \quad (2.8)$$

- 3)

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

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

- 4) For $\theta | \gamma$, we have

$$\begin{aligned} \pi(\theta | \gamma) &\propto \pi(\gamma | \theta) \pi(\theta) \\ &\propto \left[\prod_{j=1}^p \theta^{\gamma_j} (1 - \theta)^{1 - \gamma_j} \right] \theta^{c_1 - 1} (1 - \theta)^{c_2 - 1} \\ &\propto \theta^{\sum_{j=1}^p \gamma_j + c_1 - 1} (1 - \theta)^{p - \sum_{j=1}^p \gamma_j + c_2 - 1}, \end{aligned} \quad (2.9)$$

where we will use $c_1 = c_2 = 1$, which leads to the uniform distribution. We therefore have

$$\theta | \gamma \sim \mathcal{B} \left(\sum_{j=1}^p \gamma_j + c_1, p - \sum_{j=1}^p \gamma_j + c_2 \right). \quad (2.10)$$

2.3 EMVS

Genared toward finding posterior modes of the parameter posterior $\pi(\beta, \sigma^2, \theta | \mathbf{y})$ rather than simulating from the entire model posterior $\pi(\gamma | \mathbf{y})$, the EM algorithm derived here offers potentially enormous computational savings over stochastic search alternatives, especially in problems with a large number p of potential predictors.

EMVS consists of two steps: E-Step, conditional expectation given the observed data and current parameter estimates, and M-Step, entails the maximization of the expected complete data log posterior with respect to $(\beta, \sigma^2, \theta)$.

Define $\phi = (\beta, \sigma^2, \theta)$. EM algorithm maximizes $\pi(\phi | \mathbf{y})$ by iteratively maximizing the objective function

$$Q(\phi | \phi^{(t-1)}, \mathbf{y}) = E_{\gamma | \cdot} [\log \pi(\phi, \gamma | \mathbf{y})],$$

where $E_{\gamma | \cdot}$ denotes $E_{\gamma | \phi^{(t-1)}, \mathbf{y}}$.

According to the class note, it can be written as

$$Q^*(\phi | \phi^{(t-1)}, \mathbf{y}) = E_{\gamma | \cdot} [\log \pi(\phi | \gamma, \mathbf{y})] \quad (2.11)$$

At the $(t-1)$ th iteration, given $\phi^{(t-1)}$, an E-step is first applied, which computes the expectation of the right side of (2.11) to obtain Q . This is followed by M-step, which maximizes Q over $(\beta, \sigma^2, \theta)$ to obtain $(\beta^{(t)}, \sigma^{2(t)}, \theta^{(t)})$.

$$\begin{aligned}
& Q^*(\beta, \sigma^2, \theta | \beta^{(t)}, \sigma^{2(t)}, \theta^{(t)}, \mathbf{y}) \\
&= Q^*(\beta, \sigma^2, \theta | \phi^{(t-1)}, \mathbf{y}) \\
&= C + Q_1(\beta, \sigma^2 | \phi^{(t-1)}, \mathbf{y}) + Q_2(\theta | \phi^{(t-1)}, \mathbf{y}) \\
&= C + E_{\gamma| \cdot} [\log \pi(\beta, \sigma^2 | \phi^{(t-1)}, \mathbf{y})] + E_{\gamma| \cdot} [\log \pi(\theta | \phi^{(t-1)}, \mathbf{y})]
\end{aligned}$$

$$\bullet \quad (1) \quad Q_1(\beta, \sigma^2 | \phi^{(t-1)}, \mathbf{y})$$

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

$$\begin{aligned}
& Q_1(\beta, \sigma^2 | \phi^{(t-1)}, \mathbf{y}) \\
&= E_{\gamma| \cdot} [\log(\pi(\beta, \sigma^2 | \phi^{(t-1)}, \mathbf{y}))] \\
&= -\frac{1}{2}(n+p+a+2) \log(\sigma^2) - \frac{1}{2\sigma^2} (\|y - x\beta\|^2 + b) - \frac{1}{2\sigma^2} \sum_{j=1}^p \beta_j^2 E_{\gamma| \cdot} \left[\frac{1}{\nu_{\gamma_j}}\right]
\end{aligned}$$

$$\bullet \quad (2) \quad Q_2(\theta | \phi^{(t-1)}, \mathbf{y})$$

$$\begin{aligned}
& \pi(\theta | \gamma, \mathbf{y}) \\
&\propto \pi(\gamma | \theta) p(\theta) \\
&\propto \theta^{\sum_{j=1}^p \gamma_j + c_1 - 1} (1 - \theta)^{p - \sum_{j=1}^p \gamma_j + c_2 - 1}
\end{aligned}$$

$$\begin{aligned}
& \log(\pi(\theta | \gamma, \mathbf{y})) \\
&= C + \left(\sum_{j=1}^p \gamma_j + c_1 - 1\right) \log \theta + \left(p - \sum_{j=1}^p \gamma_j + c_2 - 1\right) \log(1 - \theta) \\
&= C + \log \frac{\theta}{1 - \theta} \sum_{j=1}^p \gamma_j + (c_1 - 1) \log \theta + (p + c_2 - 1) \log(1 - \theta)
\end{aligned}$$

$$\begin{aligned}
& Q_2(\theta|\phi^{(t-1)}, \mathbf{y}) \\
&= \log \frac{\theta}{1-\theta} \sum_{j=1}^p E_{\gamma|\cdot}[\gamma_j] + (c_1 - 1) \log \theta + (p + c_2 - 1) \log(1 - \theta)
\end{aligned}$$

2.3.1 The E-Step

The E-step proceeds by computing the conditional expectations $E_{\gamma|\cdot}[\frac{1}{\nu_{\gamma_j}}]$ and $E_{\gamma|\cdot}[\gamma_j]$ for Q_1 and Q_2 , respectively.

- (1) For $E_{\gamma|\cdot}[\gamma_j]$,

$$\begin{aligned}
& \pi(\gamma_j|\theta, \beta_j, \sigma^2, \mathbf{y}) \\
& \propto \pi(\beta_j|\gamma_j, \sigma^2) p(\gamma_j|\theta) \\
& \propto \nu_{\gamma_j}^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2\nu_{\gamma_j}}\beta_j^2\right) \theta^{\gamma_j} (1-\theta)^{1-\gamma_j}
\end{aligned}$$

$$\text{Hence, } \pi(\gamma_j = 1|\theta, \beta_j, \sigma^2, \mathbf{y}) = C\nu_1^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2\nu_1}\beta_j^2\right) \theta \equiv a_i,$$

$$\pi(\gamma_j = 0|\theta, \beta_j, \sigma^2, \mathbf{y}) = C\nu_0^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2\nu_0}\beta_j^2\right) (1-\theta) \equiv b_i.$$

Hence,

$$\begin{aligned}
& E_{\gamma|\cdot}[\gamma_j] = \pi(\gamma_j = 1|\theta, \beta_j, \sigma^2, \mathbf{y}) \\
&= \frac{\pi(\gamma_j = 1, \Omega)}{\sum_{\gamma_j} \pi(\gamma_j, \Omega)} \\
&= \frac{C\nu_1^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2\nu_1}\beta_j^2\right) \theta}{C\nu_1^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2\nu_1}\beta_j^2\right) \theta + C\nu_0^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2\nu_0}\beta_j^2\right) (1-\theta)} \\
&= \frac{\nu_1^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2\nu_1}\beta_j^2\right) \theta}{\nu_1^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2\nu_1}\beta_j^2\right) \theta + \nu_0^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2\nu_0}\beta_j^2\right) (1-\theta)} \\
&= p_i^*
\end{aligned}$$

Hence,

$$\begin{aligned}
& Q_2(\theta|\boldsymbol{\phi}^{(t-1)}, \mathbf{y}) \\
&= \log \frac{\theta}{1-\theta} \sum_{j=1}^p p_i^* + (c_1 - 1) \log \theta + (p + c_2 - 1) \log(1 - \theta) \quad (2.12)
\end{aligned}$$

- (2) For $E_{\gamma|\cdot}[\frac{1}{\nu_{\gamma_j}}]$,

$$\begin{aligned}
& E_{\gamma|\cdot}[\frac{1}{\nu_{\gamma_j}}] \\
&= p(\gamma_j = 1|\cdot) \frac{1}{\nu_{\gamma_j=1}} + p(\gamma_j = 0|\cdot) \frac{1}{\nu_{\gamma_j=0}} \\
&= \frac{p_i^*}{\nu_1} + \frac{1 - p_i^*}{\nu_0} \equiv d_i^*
\end{aligned}$$

Hence,

$$\begin{aligned}
& Q_1(\boldsymbol{\beta}, \sigma^2|\boldsymbol{\phi}^{(t-1)}, \mathbf{y}) \\
&= -\frac{1}{2}(n + p + a + 2) \log(\sigma^2) - \frac{1}{2\sigma^2} (\|\mathbf{y} - x\boldsymbol{\beta}\|^2 + b) - \frac{1}{2\sigma^2} \sum_{j=1}^p \beta_j^2 d_i^* \\
&= -\frac{1}{2}(n + p + a + 2) \log(\sigma^2) - \frac{1}{2\sigma^2} (\|\mathbf{y} - x\boldsymbol{\beta}\|^2 + b) - \\
&\quad \frac{1}{2\sigma^2} \|\mathbf{D}^{*1/2} \boldsymbol{\beta}\|^2, \quad (2.13)
\end{aligned}$$

where $\mathbf{D}^* = \text{diag}\{d_i^*\}_{i=1}^{p^*}$

2.3.2 The M-Step

- (1) Maximize Q_1

$\boldsymbol{\beta}$ values that maximizes Q_1 in Eq.(2.13) is equivalent to

$$\boldsymbol{\beta}^{(t)} = \arg \min_{\boldsymbol{\beta} \rightarrow \mathcal{R}^p} \|\mathbf{y} - x\boldsymbol{\beta}\|^2 + \|\mathbf{D}^{*1/2} \boldsymbol{\beta}\|^2.$$

This is quickly obtained by the well-known solution to the ridge regression problem:

$$\boldsymbol{\beta}^{(t)} = (\mathbf{X}^\top \mathbf{X} + \mathbf{D}^*)^{-1} \mathbf{X}^\top \mathbf{y}$$

Given $\boldsymbol{\beta}^{(t)}$, solve the solution of $\frac{\partial Q_1}{\partial \sigma^2}$, we can easily obtain

$$\sigma^{2(t)} = \frac{\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^{(t)}\|^2 + \|\mathbf{D}^{*1/2} \boldsymbol{\beta}^{(t)}\|^2 + b}{n + p + a + 2}.$$

- (2) Maximize Q_2

$$\begin{aligned}
& \frac{\partial Q_2}{\partial \theta} \\
&= \frac{1-\theta}{\theta} \frac{1-\theta+\theta}{(1-\theta)^2} \sum_{j=1}^p p_j^* + \frac{a-1}{\theta} - \frac{p+b-1}{1-\theta} \\
&= \frac{\sum_{j=1}^p p_j^*}{\theta(1-\theta)} + \frac{a-1}{\theta} - \frac{p+b-1}{1-\theta} \\
&= \frac{\sum_{j=1}^p p_j^* - (a+b+p-2)\theta + a-1}{\theta(1-\theta)} \\
&\equiv 0
\end{aligned}$$

Hence, we have $\theta^{(t)} = \frac{\sum_{j=1}^p p_j^* + a - 1}{a + b + p - 2}$.

2.4 Simulation Case Study

According to this paper, the author simulate a dataset consisting of $n = 100$ observations and $p = 1000$ predictors. Predictor values for each observation were sampled from $\mathcal{N}_p(\mathbf{0}, \mathbf{\Sigma})$ where $\mathbf{\Sigma} = (\rho_{ij})_{i,j=1}^p$ with $\rho_{ij} = 0.6^{|i-j|}$. Response values were then generated according to the linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{\beta} = (3, 2, 1, 0, 0, \dots, 0)^\top$ and $\boldsymbol{\epsilon} \sim \mathcal{N}_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ with $\sigma^2 = 3$.

The R code is shown bellow

```

library(mvtnorm)
library(ggplot2)
library(reshape2)
n <- 100
p <- 1000
beta <- c(3,2,1,rep(0,p-3))
Sigma <- matrix(NA,p,p)
for (i in 1:p) {
  for (j in 1:p) {
    Sigma[i,j] = 0.6^(abs(i-j))
  }
}
set.seed(2144)
x <- matrix(rmvnorm(n,mean = rep(0,p), sigma = Sigma),nrow = n)
y <- x%*%beta + rnorm(n,0,sqrt(3))
## prior and initial values
MC = 20
nu0 = 0.5

```

```

nu1 <- 1000
hat.theta = rep(1,MC)
hat.theta[1] = 0.5
hat.gamma <- rep(1, p)
hat.beta = matrix(NA,nrow = MC, ncol = p)
hat.beta[1,] = rep(1,p)
hat.sig2 <- rep(NA,MC)
hat.sig2[1] <- 1#mean((y - x %%% hat.beta[1,])^2)

for (i in 2:MC) {
  a <- dnorm(hat.beta[i-1,],0,sd = sqrt(hat.sig2[i-1]*nu1))*(hat.theta[i-1])
  b <- dnorm(hat.beta[i-1,],0,sd = sqrt(hat.sig2[i-1]*nu0))*(1-hat.theta[i-1])
  p.star = a/(a+b)
  D.star <- diag((1-p.star)/nu0+p.star/nu1)
  D.star.5 <- diag(sqrt((1-p.star)/nu0+p.star/nu1))
  hat.beta[i,] <- solve(t(x)%%x+D.star)%%t(x)%%y
  hat.sig2[i] <- (t(y-x%%hat.beta[i,])%%(y-x%%hat.beta[i,]) +
    t(D.star.5%%hat.beta[i,])%%(D.star.5%%hat.beta[i,]) + 1)/(n+p+1)
  hat.theta[i] <- sum(p.star)/(p)
  # hat.theta[i] = 0.5

  #####
  ## uncomment if you want to see the convergence ##

  # plot(beta,hat.beta[i,], ylim = c(-0.5,3))
  # abline(h = 0, col = 4)
  # abline(a=0,b=1,lty=2)
  # print(i)
}
data1 <- data.frame(beta=beta,hat.beta=hat.beta[MC,])
fig1 <- ggplot(data1,aes(beta,hat.beta))+
  geom_point(color = 'red',size = 2,alpha = 0.7)+
  ylab(expression(hat(beta)))+
  xlab(expression(beta))+
  ylim(c(-0.3,3)) +
  ggtitle(label = 'MAP Estimate versus True Coefficients')+
  geom_abline(intercept = 0, slope = 1,lty = 2)+
  theme(axis.title.x = element_text(size=12),
        axis.title.y = element_text(size=12),
        axis.text.x = element_text(size=12),
        axis.text.y = element_text(size=12))
# print hat.theta and hat.sigma
print(round(c(hat.theta[MC],sqrt(hat.sig2[MC])),4))

```

```
## [1] 0.0031 0.0404
```

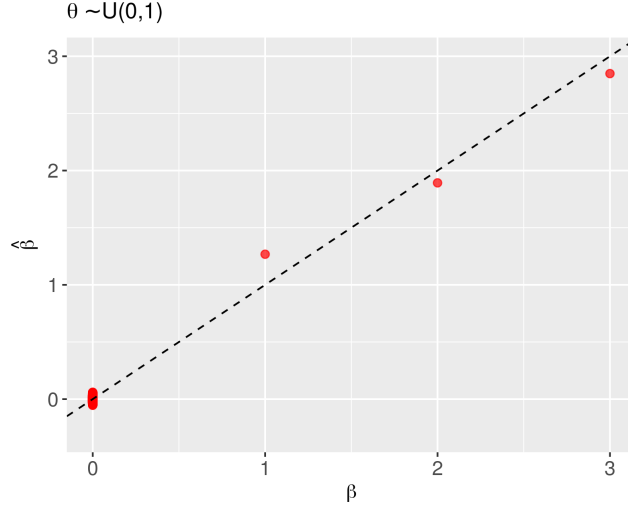


Figure 2.1: Modal estimates of the regression coefficients

Beginning with an illustration of the **EM** algorithm from the section 3, we apply it to the simulated data using the spike-and-slab prior (2.4) with a single value $\nu_0 = 0.5, \nu_1 = 1000$, and the $\theta \sim U(0, 1)$. The starting values for the **EM** algorithm were set to $\sigma^{2(0)} = 1$ and $\beta^{(0)} = \mathbf{1}_p$. After only four iterations, the algorithm obtained the modal coefficient estimates $\hat{\beta}$ depicted in Figure 2.1, displaying the estimated β and true β . In my case, the associated modal estimates of $\hat{\theta}$ and $\hat{\sigma}^2$ were 0.003 and 0.040, respectively; which are very similar to results of the paper.

For comparison, I applied the same formulation except with the Bernoulli prior (2.6) under fixed $\theta = 0.5$ Figure 2.2. Note the inferiority of the estimates near zero due to the lack of adaptivity of the Bernoulli prior in determining the degree of underlying sparsity.

2.4.1 The effect of different ν_0 and $\beta^{(0)}$ values on variable selection

Rather than fixing ν_0 as 0.5, I vary it from 0 to 0.5 to see its effect on the variable selection. I imitate the author's method to consider the grid of ν_0 values $V = \{0.01 + k \times 0.01 : k = 0, \dots, 50\}$ again with $\nu_1 = 1000$ fixed and $\beta^{(0)} = \mathbf{1}_p$. Figure 2.3 shows the modal estimates of regression coefficients obtained for each $\nu_0 \in V$. We can discover that only when $\nu_0 > 0.15$, we can get a good estimation for the β . I also consider the effect different initial values of β on the variable selection. Fix $\nu_0 = 0.5$ and $\nu_1 = 1000$, I set the grid of $\beta^{(0)} = \mathbf{c}_p$ where \mathbf{c} is a p -dimension vector with each value $c_i = \{-5 + 0.2 \times k : k = 0, \dots, 50\}, i = 1, 2, \dots, p$. Figure

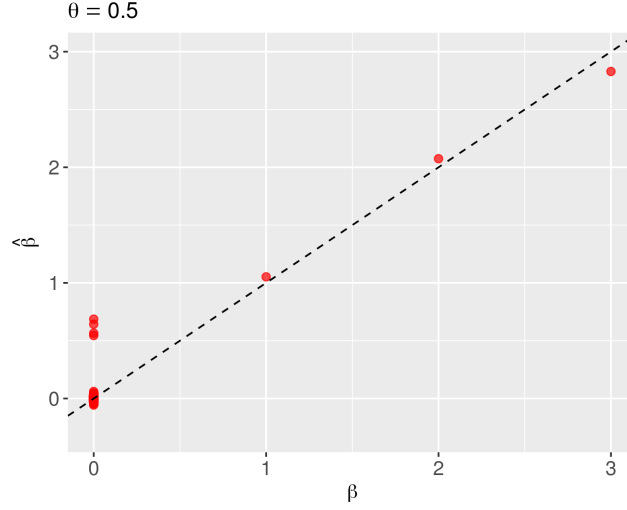


Figure 2.2: Modal estimates of the regression coefficients

2.4 depicts modal estimates of regression coefficients with different initial values of $\beta^{(0)}$. We discover only when $|\beta^{(0)}| < 2$, we can get good estimation for the β .

I also consider the effect different initial values of β on the variable selection. Fix $\nu_0 = 0.5$ and $\nu_1 = 1000$, I set the grid of $\beta^{(0)} = \mathbf{c}_p$ where \mathbf{c} is a p -dimension vector with each value $c_i = \{-5 + 0.2 \times k : k = 0, \dots, 50\}$, $i = 1, 2, \dots, p$. Figure 2.4 depicts modal estimates of regression coefficients with different initial values of $\beta^{(0)}$. We discover only when $|\beta^{(0)}| < 2$, we can get good estimation for the β .

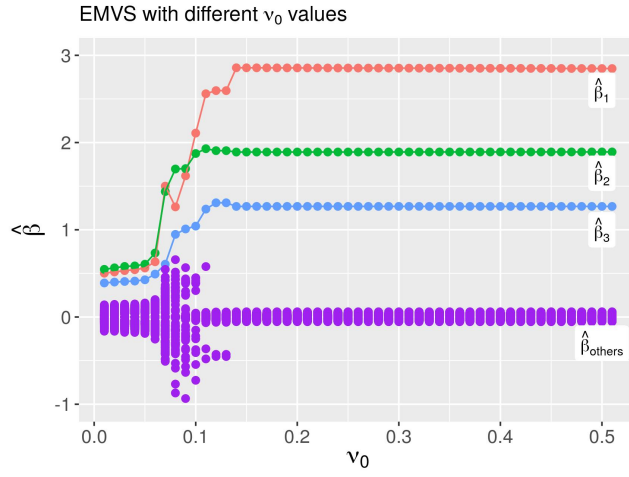


Figure 2.3: Modal estimates of the regression coefficients

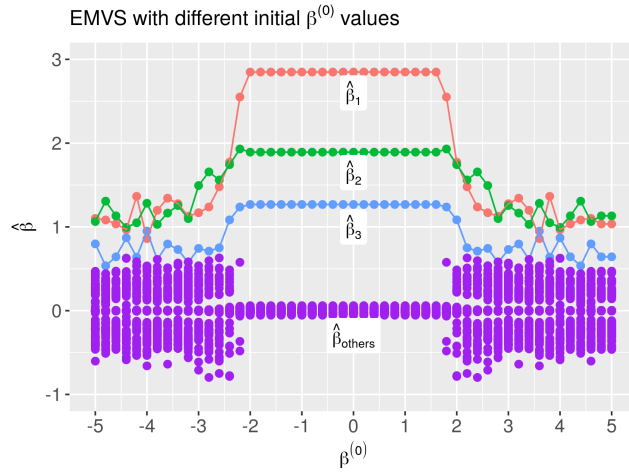


Figure 2.4: Modal estimates of the regression coefficients

Bibliography

Rockova, V. and George, E. (2014). Emvs: The em approach to bayesian variable selection. *Journal of the American Statistical Association*, 109.