Final Project: Reversible Jump Markov Chain Monte Carlo Algorithm for Model Selection in Linear Regression
Anuj Srivastava

Jaime Frade

May 1, 2009

Contents

List of Figures

Chapter 1

Introduction

1.1 Problem Statement:

The objective of this assignment is to solve for the regression solving for the regression coecients in a standard linear model, yet the selection of predictors is not known. In other words, we are given a large number, say $m = 10$,, of predictors and we have to select an appropriate subset to obtain the optimal model.

This model selection problem is the main focus of this project. We will restrict to a smaller problem where the optimal subset is simply the first n predictors, yet n is unknown. This reduces the possible number of models from the power set of a finite set with $m, 2^m$ elements, to only m possibilities.

Chapter 2

Methodology

2.1 Preliminaries

In this assignment we seek coefficients for the model

$$y = \sum_{i=1}^{n} x_i b_i + \epsilon \tag{2.1}$$

where $n < m$, x_i's are the predictors, y is the response variable, and ϵ is the measurement noise. In this problem, given k independent measurements, denoted by $\mathbf{y}$, $\mathbf{X}$, and ϵ. To solve this problem, the will formulate a Bayesian solution to the joint estimation of $\{n, b_1, \ldots, b_n\}$.

2.2 Definitions

To setup a Bayesian formulation, define a joint posterior density of the type:

$$f(n, \mathbf{b}_n|\mathbf{y}) \propto f(\mathbf{y}, n|\mathbf{b}_n) f(\mathbf{b}_n|n) f(n)$$

Let the notation $X_n = X(:, 1 : n)$ and $\mathbf{b}_n = \{b_1, \ldots, b_n\}$. In this project, will use the following terms

The likelihood function is given by

$$f(\mathbf{y}, n|\mathbf{b}_n) = \left(\frac{1}{\sqrt{2\pi\sigma_0^2}}\right)^k e^{-\frac{1}{2\sigma_0^2}||\mathbf{y}-\mathbf{X_n b_n}||^2} \tag{2.2}$$

The prior on $\mathbf{b}_n$ given n is:

$$f(\mathbf{b}_n|n) = \left(\frac{1}{\sqrt{2\pi\sigma_p^2}}\right)^k e^{-\frac{1}{2\sigma_p^2}||\mathbf{b}_n-\mu_b||^2} \tag{2.3}$$

The prior on n is uniform:

$$f(n) = \frac{1}{m} \tag{2.4}$$

2.3 Sampling from the Posterior

The method of the project will involve using the Reversible Jump Monte Carlo sampling (RJMCMC) algorithm for sampling from the posterior.

Let (n, b_n) be the current samples from the posterior. The following is the algorithm used to apply RJMCMC:

(a) Select a candidate number, $n*$ from the probability $f(n)$.

(b) If $n* \leq n$, generate a random vector $u \sim N(0, \sigma_r I_{n*})$. The candidate coefficient vector is given by:

$$b_{n*} = \begin{bmatrix} \mathbf{b}_n \\ 0 \end{bmatrix} + \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix}, \qquad \mathbf{u} = \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix}$$

Compute the likelihoods:

$$\begin{aligned} h_1(\mathbf{u}) &= \left(\frac{1}{\sqrt{2\pi\sigma_r^2}}\right)^{n*} e^{-\dfrac{1}{2\sigma_r^2}||\mathbf{u}||^2} \\ h_2(\mathbf{u_1}) &= \left(\frac{1}{\sqrt{2\pi\sigma_r^2}}\right)^{n} e^{-\dfrac{1}{2\sigma_r^2}||\mathbf{u_1}||^2} \end{aligned}$$

(c) If $n* < n$, generate a random vector $u_1 \sim N(0, \sigma_r I_{n*})$. The candidate coefficient vector is given by:

$$b_{n*} = \mathbf{b}_n^1 + \mathbf{u}_1, \qquad \mathbf{b}_n = \begin{bmatrix} \mathbf{b}_n^1 \\ \mathbf{b}_n^2 \end{bmatrix}$$

and form $\mathbf{u} = \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{b}_n^2 \end{bmatrix}$

Compute the likelihoods:

$$\begin{aligned} h_2(\mathbf{u}) &= \left(\frac{1}{\sqrt{2\pi\sigma_r^2}}\right)^{n} e^{-\dfrac{1}{2\sigma_r^2}||\mathbf{u}||^2} \\ h_1(\mathbf{u_1}) &= \left(\frac{1}{\sqrt{2\pi\sigma_r^2}}\right)^{n*} e^{-\dfrac{1}{2\sigma_r^2}||\mathbf{u_1}||^2} \end{aligned}$$

(d) Compute the acceptance-rejection function:

$$\begin{aligned} \rho &= \min\left(1, \frac{f(n^*, \mathbf{b}_{n^*}|\mathbf{y})h_2}{f(n, \mathbf{b}_n|\mathbf{y})h_1}\right) = \min\left(1, \frac{f(\mathbf{y}|n^*, \mathbf{b}_{n^*})f(\mathbf{b}_{n^*}|n^*)h_2}{f(\mathbf{y}|n^*, \mathbf{b}_n)f(\mathbf{b}_n|n)h_1}\right) \\ &= \min\left(1, \frac{e^{(E_1-E_2)}(2\pi\sigma_p^2)^{\frac{n-n^*}{2}} e^{-\frac{1}{2\sigma_p^2}(||\mathbf{b_{n^*}}-\mu_\mathbf{b}||^2 - ||b_n-\mu_b||^2)} h_2}{h_1}\right) \end{aligned}$$

Where $E_1 = \frac{1}{2\sigma_0^2}||\mathbf{y} - \mathbf{X}_{n^*}\mathbf{b}_{n^*}||^2$ and $E_2 = \frac{1}{2\sigma_0^2}||\mathbf{y} - \mathbf{X}_n\mathbf{b}_n||^2$

(e) If $U \sim U[0,1]$ is less than ρ then set $(n, \mathbf{b}_n) = (n^*, \mathbf{b}_{n^*})$. Else, return to Step (a).

Chapter 3

Matlab Code

```
clear all
clc
N=100000;
m=10;
k=10;
sigma0=.2;
sigmap=.3;
sigmar=.2;

%initial conditions
n0=ceil(rand*m);
mub=ones(1,n0)*2;
b=normrnd(2,sigmap,n0,1);
X=5*randn(k,m);
y=X(:,1:n0)*b+sigma0*randn(k,1);
tic
for i=1:N
nstar=ceil(rand*m);
mustar=ones(1,nstar)*2;
        if (nstar<n0)
            u1=normrnd(0,sigma0,nstar,1);
            b1=b(1:length(u1));
            b2=b(length(u1)+1:end);
            bnstar=b1+u1;
            u=[u1;b2];
            Unorm = sum(u(:).^2);
            h2 = ((1/sqrt(2*pi*sigmar^2))^n0)*exp((-1/(2*sigmar^2))*(Unorm));
            U1norm = sum(u1(:).^2);
            h1 = ((1/sqrt(2*pi*sigmar^2))^nstar)*exp((-1/(2*sigmar^2))*(U1norm));

        else
            u=normrnd(0,sigma0,nstar,1);
            u1=u(1:n0);
            u2=u(n0+1:end);
            bn=[b;zeros(length(u2),1)];
```

```
            bnstar=bn+u;
            Unorm = sum(u(:).^2);
            h1 = ((1/sqrt(2*pi*sigmar^2))^nstar)*exp((-1/(2*sigmar^2))*(Unorm));
            U1norm = sum(u1(:).^2);
            h2 = ((1/sqrt(2*pi*sigmar^2))^n0)*exp((-1/(2*sigmar^2))*(U1norm));
        end
    Xstar=X(1:nstar,:);
    Xn=X(1:n0,:);
    y1=y'-bnstar'*Xstar;
    y2=y'-b'*Xn;
    y1 = sum(y1(:).^2);
    y2 = sum(y2(:).^2);
    E1=y1/(2*sigma0^2);
    E2=y2/(2*sigma0^2);
    E3 = bnstar-mustar';
    E3 = sum(E3(:).^2);
    E4 = b-mub';
    E4 = sum(E4(:).^2)';
    rho1=(exp(-(E1-E2)))*((2*pi*sigmap^2)^((n0-nstar)/2))*(exp((-1/(2*sigmap^2))*(E3-E4)))
    rho=min(rho1,1);

    U=rand;
        if(U<rho)
            n(i+1)=nstar;
        else
            n(i+1)=n0;
        end
end
toc
n;
figure(1)
hist(n)
h=hist(n);
idx=find(h==max(h))
b
```

Chapter 4

Results

4.1 Experimental Results

Simulated a dataset with the following code below. For this data $(\mathbf{y}, \mathbf{X})$ implement the RJMCMC algorithm (with $\sigma_r = 0.2$) to a sample and using $N = 100,000$ samples from the posterior.

```
m=10;
k=10;
sigma0=.2;
sigmap=.3;
sigmar=.2;
n0=ceil(rand*m);
mub=ones(1,n0)*2;
b=normrnd(2,sigmap,n0,1);
X=5*randn(k,m);
y=X(:,1:n0)*b+sigma0*randn(k,1);
```

4.1.1 Histogram plots of several different runs

Show a histogram of n values visited by Markov chain. This estimates the posterior probabilty, $f(n|\mathbf{y})$.

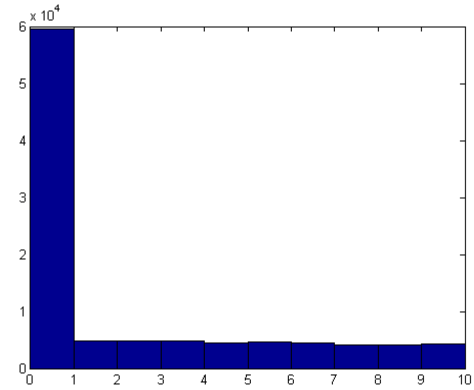

Figure 4.1: Histogram: n=1

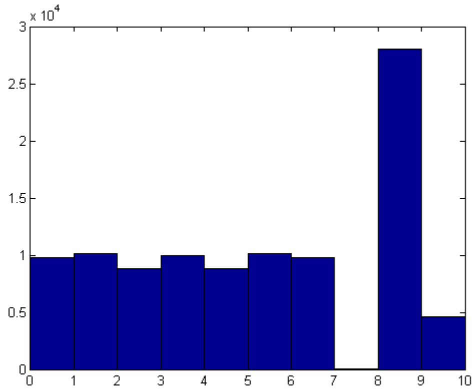

Figure 4.2: Histogram: n=9

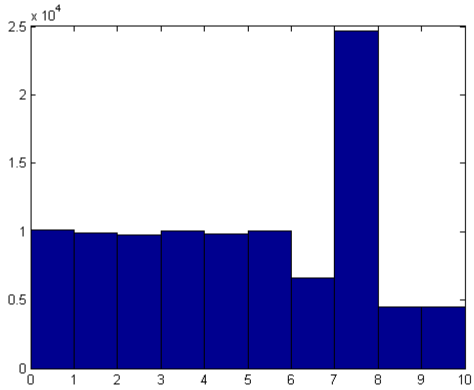

Figure 4.3: Histogram: n=8

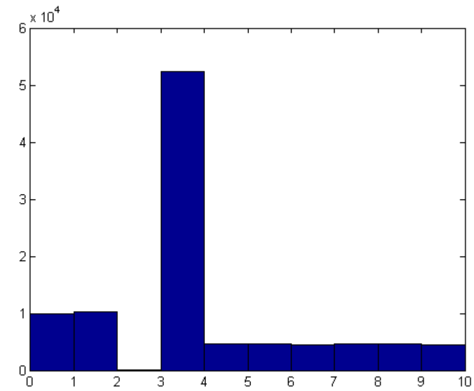

Figure 4.4: Histogram: n=4

4.1.2 Computation Times

$n = 1$	481.608973 seconds
$n = 9$	668.774289 seconds.
$n = 8$	651.934132 seconds.
$n = 4$	533.207485 seconds.

4.1.3 Table of Conditional Mean Estimates of $\mathbf{b}_n$

For the n value that is visited most often, find the conditional mean estimate of $\mathbf{b}_n$ under the density, $f(\mathbf{b}_n|\mathbf{y})$

n	b	n	b
1	1.6652	9	2.4377 1.9074 1.8166 1.9085 2.1502 2.2462 1.8852 2.1858 2.1320
n	b	n	b
8	1.5613 1.8213 1.9352 2.0828 1.7574 2.0421 2.1864 1.7970	4	2.0281 2.4511 2.0766 1.6940

Chapter 5

Conclusion

In this paper, the main focus was on using the RJMCMC for model selection in solving for the regression coefficients in a standard linear model, however, the selection of predictors is not known. We were given a large number of predictors and were required to select an appropriate subset to obtain the optimal model. In this problem, we focused on the first n predictors, a subset, of a total number of predictors, m.

It was found that by using RJMCMC algorithm to perform the required selection implicitly by simulating from the posterior distribution of $\mathbf{b}_n$, under the density, $f(\mathbf{b}_n|\mathbf{y})$. Although the computations times were relatively high in comparison to other programs/homeworks, the method was able to calculate the coefficients for the model

$$y = \sum_{i=1}^{n} x_i b_i + \epsilon \tag{5.1}$$