1. Using a prior of

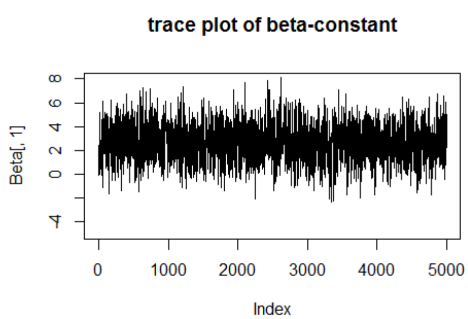
```
b0 <- matrix(c(0,0,0,0),nrow=4,ncol=1)
S0 <- matrix(0,nrow=4,ncol=4)
diag(S0) <- 5
```

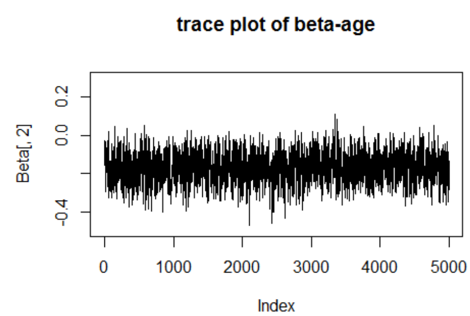
We can implement the Probit Model and have the following result:

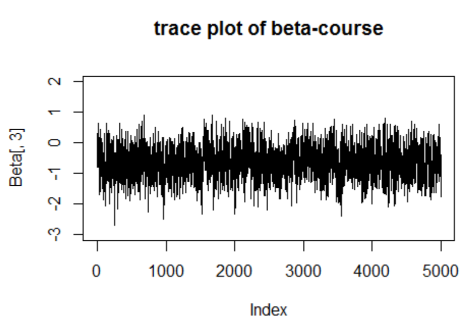
	mean	Median	Standard Deviation	95%-lower	95%-upper
1	2.62	2.59	1.45	-0.18	5.51
age	-0.17	-0.17	0.07	-0.32	-0.03
Driving courses?	-0.63	-0.64	0.52	-1.63	0.4
Like stats?	0.71	0.71	0.52	-0.31	1.74

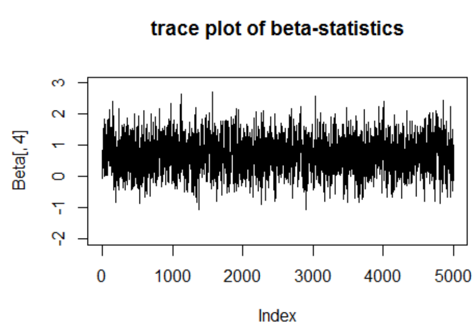
We can see that larger age, having taken driving courses, and hates statistics tends to lead to lower risk in accidents in the past 6 months.

The trace plot of the betas are copied below which shows the convergence.









2. Using the Probit Regression Gibbs Sampling results, we can simulate the results of the two persons for each iteration. Comparing the expectation of simulation results, we can have the following table.

	Person 1	Person 2
1	1	1
age	17	18
Driving courses	1	1
Like stats?	0	1
Probability of having an accident?	21.4%	37.8%

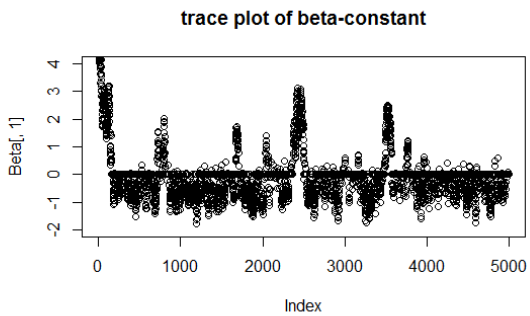
Thus, we should pick ***person 1*** to park the car.

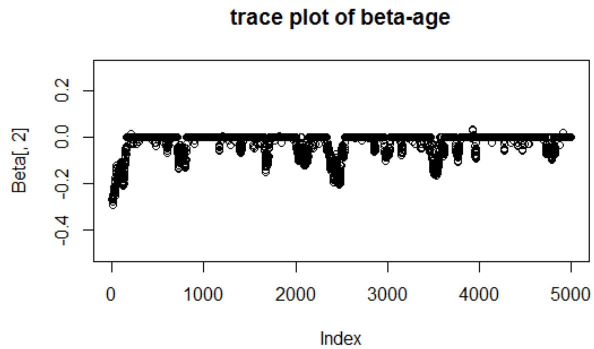
3. Implementing SSVS in the Model. S

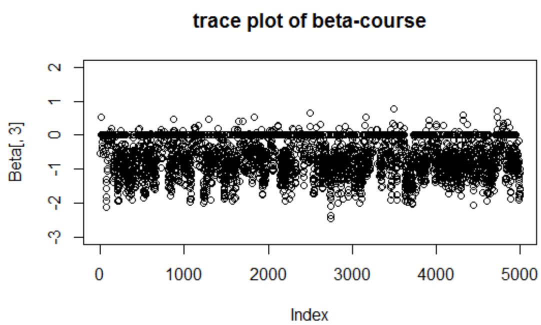
	mean	Median	Standard Deviation	95%-lower	95%-upper	Probability of Exclude this Predictor?
1	0.77	0	1.57	-1.16	4.2	0.32
age	-0.06	0	0.09	-0.25	0	0.58
Driving courses?	-0.36	0	0.5	-1.47	0.05	0.52
Like stats?	0.29	0	0.47	0	1.5	0.6
s1	-0.01	0	0.09	-0.27	0.1	0.88
s2	0.01	0	0.07	-0.02	0.26	0.9
s3	0	0	0.03	0	0.08	0.94

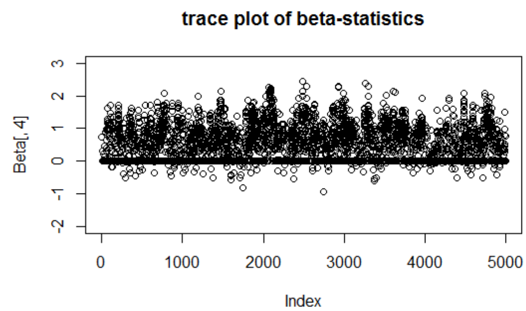
From the SSVS result we can see s1, s2, s3 has very high probability of being excluded from the Probit regression. Compared with the result using MCMC in problem 1, the 95% CI is more centered here.

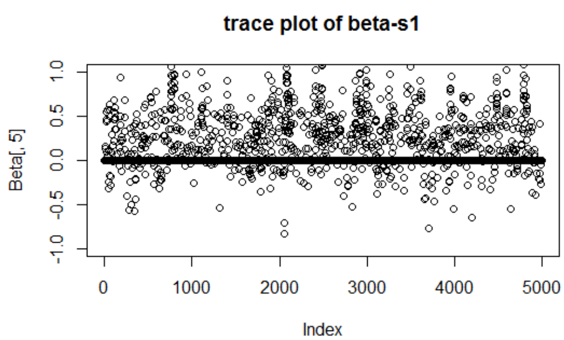
We can also take a look at the trace plot of several betas:

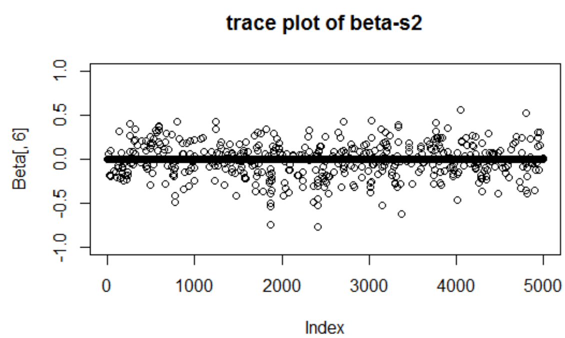

trace plot of beta-constant
Beta[, 1]
Index

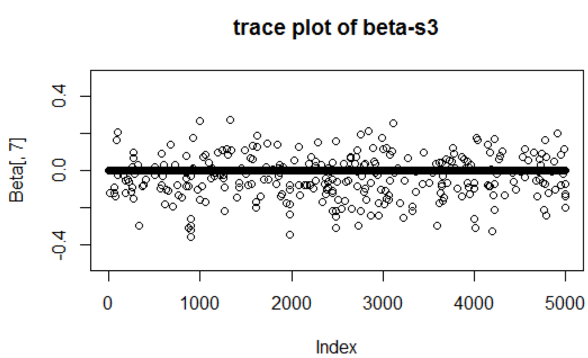

trace plot of beta-age
Beta[, 2]
Index


trace plot of beta-course
Beta[, 3]
Index


trace plot of beta-statistics
Beta[, 4]
Index


trace plot of beta-s1
Beta[, 5]
Index


trace plot of beta-s2
Beta[, 6]
Index


trace plot of beta-s3
Beta[, 7]
Index

######Code

```
data <- read.csv("C:\\Users\\Dai\\Desktop\\Bayesian Statistics\\Lab
Assignments\\12\\data.csv",header=T)
#data$age<- (data$age - mean(data$age))/sqrt(var(data$age))
library(mvtnorm)
X <- cbind(matrix(1,nrow=35,ncol=1),unlist(data[,1]),unlist(data[,2]),unlist(data[,3]))
y <- data[,4]

##Prior
b0 <- matrix(c(0,0,0,0),nrow=4,ncol=1)
S0 <- matrix(0,nrow=4,ncol=4)
diag(S0) <- 5

#Initials
beta <- b0
Sigma <- S0
n <- nrow(X)
z <- rep(0,n)

#Gibbs
T=5000
Beta <- matrix(nrow=T,ncol=4)
for (t in 1:T)
{
  #latent variable z
  eta<- X%*%beta
  z[y==0]<- qnorm(runif(sum(1-y),0,pnorm(0,eta[y==0],1)),eta[y==0],1)
  z[y==1]<- qnorm(runif(sum(y),pnorm(0,eta[y==1],1),1),eta[y==1],1)

  #conditional distribution
  Sigma.hat <- solve(solve(Sigma)+t(X)%*%X)
  beta.hat <- Sigma.hat%*%(solve(Sigma)%*%b0+t(X)%*%z)
  beta <- c(rmvnorm(1,beta.hat,Sigma.hat))
  Beta[t,] <- beta
}

#output
table<- matrix(0,4,5)
for(i in 1:4){
  table[i,]<- c(mean(Beta[,i]),median(Beta[,i]),sqrt(var(Beta[,i])),quantile(Beta[,i],c(0.025,0.975)))
}
table <- round(table*100)/100
write(t(table),file="betas.txt",ncol=5)

#plot
plot(Beta[,1],main="trace plot of beta-constant",type='l',ylim=c(-5,8))
plot(Beta[,2],main="trace plot of beta-age",type='l',ylim=c(-0.5,0.3))
plot(Beta[,3],main="trace plot of beta-course",type='l',ylim=c(-3,2))
plot(Beta[,4],main="trace plot of beta-statistics",type='l',ylim=c(-2,3))
```

```
######## Problem 2  #######
Input=t(matrix(c(1,17,1,0,1,18,1,1),nrow=4,ncol=2))
Output=pnorm(Input%*%t(Beta))
prob=rowMeans(Output)
print(prob)


########### Problem 3###########

library(mvtnorm)
data <- read.csv("C:\\Users\\Dai\\Desktop\\Bayesian Statistics\\Lab
Assignments\\12\\data.csv",header=T)
library(mvtnorm)
X <- cbind(matrix(1,nrow=35,ncol=1),unlist(data[,1]),unlist(data[,2]),unlist(data[,3]))
s1=rnorm(dim(X)[1],0,1)
s2=rnorm(dim(X)[1],0,1.5)
s3=rnorm(dim(X)[1],0,2)
X <-
cbind(X,matrix(s1,nrow=dim(X)[1],ncol=1),matrix(s2,nrow=dim(X)[1],ncol=1),matrix(s3,nrow=dim(X)[1],
ncol=1))
y <- data[,4]

# Prior
p<- ncol(X)
p0<- rep(0.5,p)
b0<- rep(0,p)
s0<- rep(2,p)

# SSVS
#use mle estimate for beta's starting value
mle<- glm(y ~ -1+X, family=binomial("probit"))
beta.mle<- mle$coef
beta<- beta.mle

n<- nrow(X)
z<- rep(0,n)
T<- 5000

Beta <- matrix(nrow=T,ncol=p)
count <- matrix(0,ncol=p)

for(t in 1:T){
  #latent variable
  eta<- X%*%beta # linear predictor
  z[y==0]<- qnorm(runif(sum(1-y),0,pnorm(0,eta[y==0],1)),eta[y==0],1)
  z[y==1]<- qnorm(runif(sum(y),pnorm(0,eta[y==1],1),1),eta[y==1],1)
```

```
  for(j in 1:p){

    V<- 1/(s0[j]^{-2} + sum(X[,j]^2))
    E<- V*sum(X[,j]*(z-X[,-j]%*%beta[-j]))
    p.hat<- 1/(1 + p0[j]/(1-p0[j]) * dnorm(0,E,sqrt(V))/dnorm(0,b0[j],s0[j]) )
    m<- rbinom(1,1,p.hat)
    if (m==0) {count[j]=count[j]+1}
    rbinom(1,1,p.hat)
    beta[j]<- m*rnorm(1,E,sqrt(V))
  }

  # output results to a file for later processing
  Beta[t,] <- matrix(beta,ncol=7)
}

plot(Beta[,1],main="trace plot of beta-constant",ylim=c(-2,4))
plot(Beta[,2],main="trace plot of beta-age",,ylim=c(-0.5,0.3))
plot(Beta[,3],main="trace plot of beta-course",,ylim=c(-3,2))
plot(Beta[,4],main="trace plot of beta-statistics",,ylim=c(-2,3))
plot(Beta[,5],main="trace plot of beta-s1",,ylim=c(-1,1))
plot(Beta[,6],main="trace plot of beta-s2",,ylim=c(-1,1))
plot(Beta[,7],main="trace plot of beta-s3",,ylim=c(-0.5,0.5))

#output
table<- matrix(0,7,6)
for(i in 1:7){
  table[i,]<- c(mean(Beta[,i]),median(Beta[,i]),
            sqrt(var(Beta[,i])),quantile(Beta[,i],c(0.025,0.975)),count[i]/T)
}
table <- round(table*100)/100
write(t(table),file="problem3.txt",ncol=6)
```