Freie Universität Berlin
Chair of Statistics
Location: Berlin
Summer Term 2017
Lecture: Einführung in die Bayes-Statistik
Examiner: Dr. Florian Meinfelder

Program leave-one-out posterior predictive checking in R

Johannes Brinkmann (), jojo-brinkmann@gmx.de
Carlo Michaelis (5043128), carlo.michaelis@gmail.com
Max Reinhardt (), max_reinhardt@me.com
Adrian Rolf (), adrian.rolf@gmx.de

Master Statistics
August 15, 2017

Contents

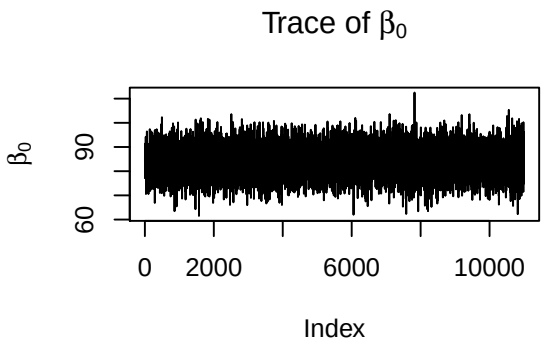
List of Figures

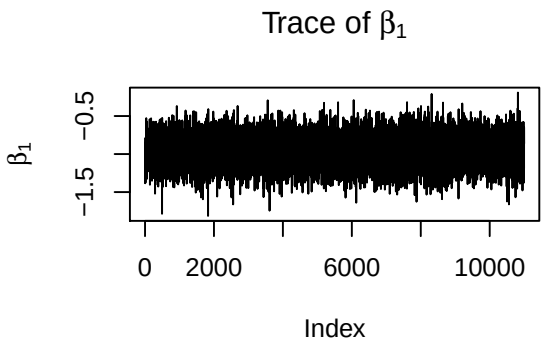
List of Tables

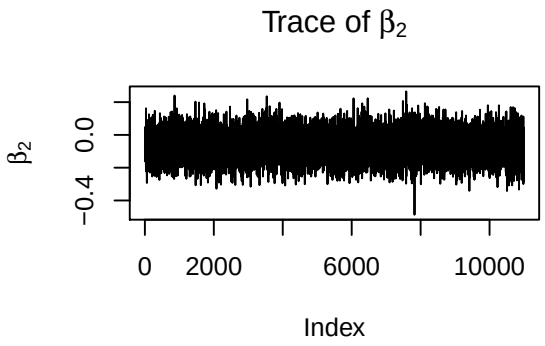
1 Introduction

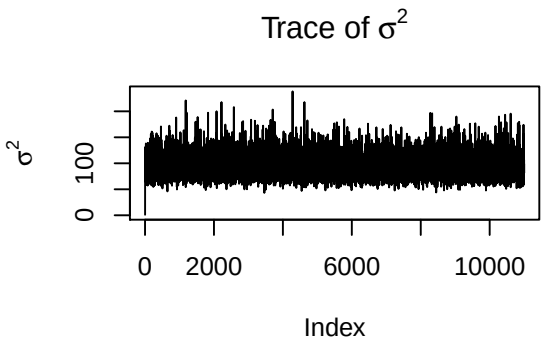
2 Code

```
# Swiss data
dat <- swiss

# Sample Size
n = nrow(dat)

# Response variable
Y = dat$Fertility

# Design matrix
X = matrix(c(rep(1,n), dat$Education, dat$Agriculture), nrow=n)
p = ncol(X)

# Number of samples
b <- 1000  # Burn in
R <- 10000 # Random draws to evaluate
B <- R + b

# Variables to store the samples in
beta = matrix(NA, nrow = B, ncol = p)
sigma = c(1, rep(NA, B))

# The Gibbs Sampler
for(i in 1:B){

  # LSE of beta
  V = solve(t(X)%*%X) # (X^T X)^-1
  beta_hat = V%*%t(X)%*%Y # (X^T X)^-1 X^T Y

  # LSE of sigma
  sigma_hat = t(Y-X%*%beta_hat)%*%(Y-X%*%beta_hat)/(n-p)

  # Sample beta from the full conditional
  beta[i,] = rmvnorm(1,beta_hat,sigma[i]*V)

  # Sample sigma from the full conditional
  sigma[i+1] = 1/rgamma(1,(n-p)/2,(n-p)*sigma_hat/2)
}

# Plot traces
par(mfrow=c(2,2))
plot(beta[,1],type='l',ylab=expression(beta[0]),main=expression("Trace of "*beta[0]))
plot(beta[,2],type='l',ylab=expression(beta[1]),main=expression("Trace of "*beta[1]))
plot(beta[,3],type='l',ylab=expression(beta[2]),main=expression("Trace of "*beta[2]))
plot(sigma,type='l',ylab=expression(sigma^2),main=expression("Trace of "*sigma^2))
```

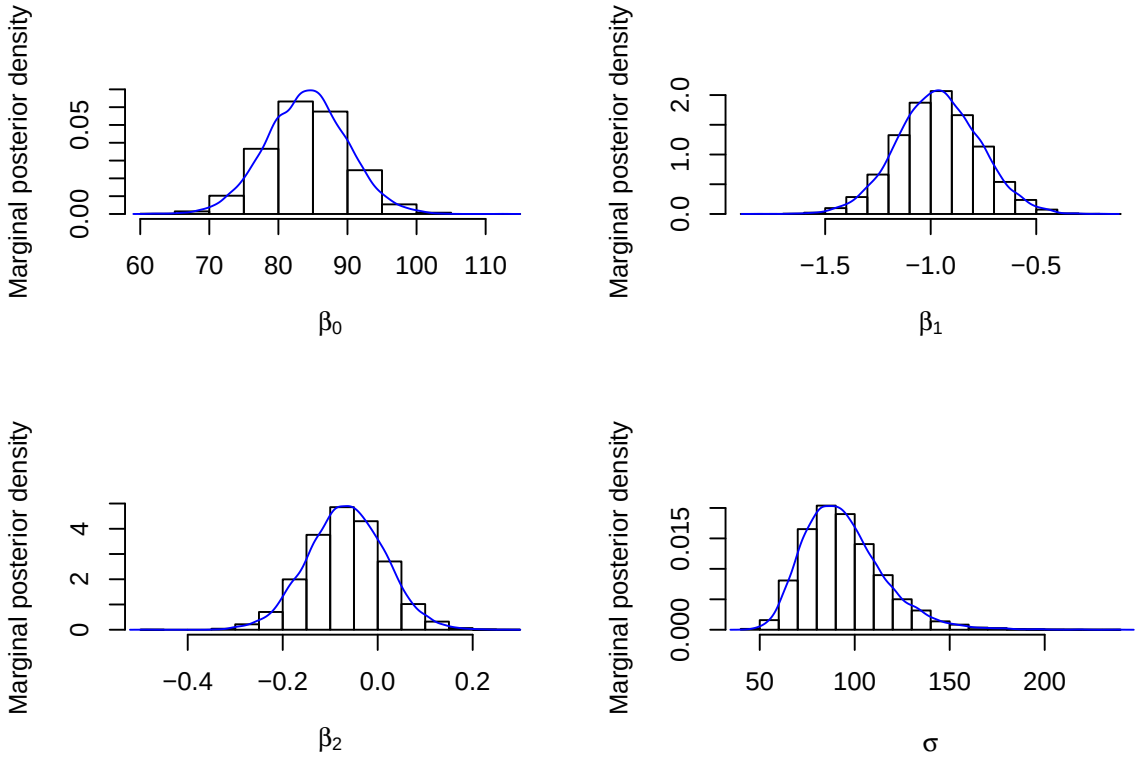









```r
# Marginal posterior densities

drawHistDensity <- function(para, para_name) {
  # para      : Parameter (e.b. Beta, Sigma)
  # para_name: Title of plot

  # Estimate density for parameter values
  density <- density(para)

  # Draw histogram and add estimated density line
  hist(para, freq = FALSE, ylim = c(0,max(density$y)), xlab = para_name,
       ylab="Marginal posterior density", main = NULL)
  lines(density, col="blue")
}

par(mfrow=c(2,2))
drawHistDensity(beta[-(1:b),1], expression(beta[0]))
drawHistDensity(beta[-(1:b),2], expression(beta[1]))
drawHistDensity(beta[-(1:b),3], expression(beta[2]))
drawHistDensity(sigma[-(1:b)], expression(sigma))
```




```
# Posterior means
cat(paste("Posterior means:\nbeta_0:", mean(beta[,1]), "\nbeta_1:", mean(beta[,2]), "\nbeta_2:", m
```

```
## Posterior means:
## beta_0: 84.0670194365396
## beta_1: -0.963491135031477
## beta_2: -0.0660612348837858
```

```
# Compare with frequentist linear regression
lm(dat$Fertility ~ dat$Education + dat$Agriculture)
```

```
##
## Call:
## lm(formula = dat$Fertility ~ dat$Education + dat$Agriculture)
##
## Coefficients:
##     (Intercept)    dat$Education  dat$Agriculture
##        84.08005         -0.96276         -0.06648
```

3 References