Half Title

Solution Manual
Introduction to Bayesian Inference:
A GUIded tour using R

Title Page

Solution Manual
Introduction to Bayesian Inference:
A GUIded tour using R

by Andrés Ramírez-Hassan, PhD. Statistical Science.

To my parents, Nancy and Orlando.

Contents

Foreword

Preface

Symbols

Symbol Description

$\neg$	Negation symbol.	$\mathcal{R}$	The Real set.
$\propto$	Proportional symbol.	$\emptyset$	Empty set.
$\perp$	Independence symbol.	$\mathbb{1}$	Indicator function.

Part I

Foundations: Theory, simulation methods and programming

1

Basic formal concepts

1.1 Solutions of Exercises

1. *The court case: the blue or green cap*

   A cab was involved in a hit and run accident at night. There are two cab companies in the town: blue and green. The former has 150 cabs, and the latter 850 cabs. A witness said that a blue cab was involved in the accident; the court tested his/her reliability under the same circumstances, and got that 80% of the times the witness correctly identified the color of the cab. *What is the probability that the color of the cab involved in the accident was blue given that the witness said it was blue?*

   Answer

   Set WB and WG equal to the events that the witness said the cab was blue and green, respectively. Set B and G equal to the events that the cabs are blue and green, respectively. We need to calculate $P(B|WB)$, then:

   $$P(B|WB) = \frac{P(B, WB)}{P(WB)} \tag{1.1}$$
   $$= \frac{P(WB|B) \times P(B)}{P(WB|B) \times P(B) + (1 - P(WB|B)) \times (1 - P(B))}$$
   $$= \frac{0.8 \times 0.15}{0.8 \times 0.15 + 0.2 \times 0.85}$$
   $$= 0.41$$

2. *The Monty Hall problem*

   What is the probability of winning a car in the *Monty Hall problem* switching the decision if there are four doors, where there are three goats and one car? Solve this problem analytically and computationally. What if there are n doors, $n - 1$ goats and one car?

   Answer

   Let's name P_i the event *contestant picks door No. i*, H_i the event *host picks door No. i*, and C_i the event *car is behind door No. i*. Let's assume that the

contestant picked door number 1, and the host picked door number 3, then the contestant is interested in the probability of the event $P(C_i|H_3, P_1), i = 2$ or 4. Then, $P(H_3|C_3, P_1) = 0$, $P(H_3|C_2, P_1) = P(H_3|C_4, P_1) = 1/2$ and $P(H_3|C_1, P_1) = 1/3$. Then,

$$
\begin{aligned}
P(C_i|H_3, P_1) &= \frac{P(C_i, H_3, P_1)}{P(H_3, P_1)} \\
&= \frac{P(H_3|C_i, P_1)P(C_i|P_1)P(P_1)}{P(H_3|P_1) \times P(P_1)} \\
&= \frac{P(H_3|C_i, P_1)P(C_i)}{P(H_3|P_1)} \\
&= \frac{1/2 \times 1/4}{1/3} \\
&= \frac{3}{8},
\end{aligned}
\tag{1.2}
$$

where the third equation uses the fact that C_i and P_i are independent events, and $P(H_3|P_1) = 1/3$ due to this depending just on P_1 (not on C_i).

Therefore, changing the initial decision increases the probability of getting the car from 1/4 to 3/8!

Let's check the case with n doors, and assume that the contestant picks the door No. 1, the car is behind the door No. n, and the host, who knows what is behind each door, opens any of the remaining $n - 2$ doors, where there is a goat. The contestant is interested in the probability of the event:

$$
\begin{aligned}
P(C_n|(H_2 \cup \ldots \cup H_{n-1}) \cap P_1) &= \frac{P((H_2 \cup H_3 \cup \ldots \cup H_{n-1})|C_n \cap P_1) P(C_n|P_1)P(P_1)}{P((H_2 \cup H_3 \cup \ldots \cup H_{n-1})|P_1) P(P_1)} \\
&= \frac{[P(H_2|C_n \cap P_1) + \ldots + P(H_{n-1}|C_n \cap P_1)] P(C_n)}{P(H_2|P_1) + P(H_3|P_1) + \ldots + P(H_{n-1}|P_1)} \\
&= \frac{1 \times \left(\frac{1}{n}\right)}{\frac{1}{n-1} + \frac{1}{n-1} + \ldots + \frac{1}{n-1}} \\
&= \left(\frac{1}{n}\right)\left(\frac{n-1}{n-2}\right).
\end{aligned}
\tag{1.3}
$$

In general, the probability of winning the car changing the pick is $\frac{1}{n}\frac{n-1}{n-2}$, while the probability of winning given no change is $\frac{1}{n}$. Given that $\frac{1}{n}\frac{n-1}{n-2} > \frac{1}{n}$ for all $n \geq 3$, where the difference between both probabilities is $\frac{1}{n(n-2)}$. We observe that as the number of doors increases, the difference between the two probabilities becomes zero.

Let's see a code for the general setting,

R code. The Monty Hall Problem

```r
set.seed(0101) # Set simulation seed
S <- 100000 # Simulations
Game <- function(opt = 3){
# opt: number of options. opt > 2
opts <- 1:opt
car <- sample(opts, 1) # car location
guess1 <- sample(opts, 1) # Initial guess
if(opt == 3 && car != guess1) {
 host <- opts[-c(car, guess1)]
 } else {
 host <- sample(opts[-c(car, guess1)], 1)
}
win1 <- guess1 == car # Win given no change
if(opt == 3) {
 guess2 <- opts[-c(host, guess1)]
 } else {
 guess2 <- sample(opts[-c(host, guess1)], 1)
 }
win2 <- guess2 == car # Win given change
return(c(win1, win2))
}
#Win probabilities
Prob <- rowMeans(replicate(S, Game(opt = 4)))
#Winning probabilities no changing door
Prob[1]
0.25151
#Winning probabilities changing door
Prob[2]
0.37267
```

3. Solve the health insurance example using a Gamma prior in the rate parametrization, that is, $\pi(\lambda) = \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \lambda^{\alpha_0 - 1} \exp\{-\lambda\beta_0\}$.

Answer

First, we get the posterior distribution,

$$\pi(\lambda|\mathbf{y}) = \left(\frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \lambda^{\alpha_0 - 1} e^{-\lambda\beta_0} \right) \left(\prod_{i=1}^{N} \frac{\lambda^{y_i} e^{-\lambda}}{y_i!} \right) \tag{1.4}$$

$$= \left(\frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \lambda^{\alpha_0 - 1} e^{-\lambda \beta_0} \right) \left(\frac{\lambda^{\sum_{i=1}^{N} y_i} e^{-N\lambda}}{\prod_{i=1}^{N} y_i!} \right)$$

$$= \left(\frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \frac{1}{\prod_{i=1}^{N} y_i!} \right) \lambda^{\sum_{i=1}^{N} y_i + \alpha_0 - 1} e^{-\lambda(\beta_0 + N)}$$

$$\propto \lambda^{\sum_{i=1}^{N} y_i + \alpha_0 - 1} e^{-\lambda(\beta_0 + N)}. \tag{1.5}$$

The last expression is the kernel of a Gamma distribution with parameters $\alpha_n = \sum_{i=1}^{N} y_i + \alpha_0$ and $\beta_n = \beta_0 + N$.

Given that $\int_0^{\infty} \pi(\lambda|\mathbf{y}) \, d\lambda = 1$, then the constant of proportionality in the last expression is $\Gamma(\alpha_n)/\beta_n^{\alpha_n}$. Therefore the posterior density function $\pi(\lambda|\mathbf{y})$ is $G(\alpha_n, \beta_n)$.

The posterior mean is

$$E[\lambda|\mathbf{y}] = \frac{\alpha_n}{\beta_n} \tag{1.6}$$

$$= \frac{\sum_{i=1}^{N} y_i + \alpha_0}{\beta_0 + N}$$

$$= \left(\frac{N}{\beta_0 + N} \right) \bar{y} + \left(\frac{\beta_0}{\beta_0 + N} \right) \frac{\alpha_0}{\beta_0}$$

$$= w\bar{y} + (1 - w) E[\lambda],$$

where $w = \frac{N}{\beta_0 + N}$, $\bar{y}$ is the sample mean, and $E[\lambda] = \frac{\alpha_0}{\beta_0}$.

The posterior predictive distribution is given by

$$\pi(Y_0|\mathbf{y}) = \int_0^{\infty} \frac{\lambda^{y_0} e^{-\lambda}}{y_0!} \pi(\lambda|\mathbf{y}) \, d\lambda \tag{1.7}$$

$$= \int_0^{\infty} \left(\frac{\lambda^{y_0} e^{-\lambda}}{y_0!} \right) \left(\frac{\beta_n^{\alpha_n}}{\Gamma(\alpha_n)} \lambda^{\alpha_n - 1} e^{-\lambda \beta_n} \right) d\lambda$$

$$= \frac{\beta_n^{\alpha_n}}{\Gamma(\alpha_n) y_0!} \int_0^{\infty} \lambda^{y_0 + \alpha_n - 1} e^{-\lambda(1 + \beta_n)} d\lambda$$

$$= \frac{\beta_n^{\alpha_n}}{\Gamma(\alpha_n) y_0!} \frac{\Gamma(y_0 + \alpha_n)}{(1 + \beta_n)^{y_0 + \alpha_n}}$$

$$= \frac{\Gamma(y_0 + \alpha_n)}{\Gamma(\alpha_n) y_0!} \left(\frac{1}{1 + \beta_n} \right)^{y_0} \left(\frac{\beta_n}{1 + \beta_n} \right)^{\alpha_n}$$

$$= \frac{(y_0 + \alpha_n - 1)!}{(\alpha_n - 1)! y_0!} \left(\frac{1}{1 + \beta_n} \right)^{y_0} \left(\frac{\beta_n}{1 + \beta_n} \right)^{\alpha_n}$$

$$= \binom{y_0 + \alpha_n - 1}{y_0} \left(\frac{1}{1 + \beta_n} \right)^{y_0} \left(\frac{\beta_n}{1 + \beta_n} \right)^{\alpha_n}.$$

Therefore $Y_0|y \sim NB(\alpha_n, p_n)$ where $p_n = \frac{1}{1+\beta_n}$.

To use empirical Bayes, we have the following setting

$$\left[\hat{\alpha}_0 \hat{\beta}_0 \right] = \arg \max_{\alpha_0, \beta_0} \ln(p(\mathbf{y})),$$

where

$$p(y) = \int_0^\infty \left(\frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \lambda^{\alpha_0 - 1} e^{-\lambda \beta_0} \right) \left(\prod_{i=1}^N \frac{\lambda^{y_0} e^{-\lambda}}{y_0!} \right) d\lambda \qquad (1.8)$$

$$= \left(\frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0) \prod_{i=1}^N y_i!} \right) \int_0^\infty \lambda^{\sum_{i=1}^N y_i + \alpha_0 - 1} e^{-\lambda(\beta_0 + N)} d\lambda$$

$$= \left(\frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0) \prod_{i=1}^N y_i!} \right) \left(\frac{\Gamma\left(\sum_{i=1}^N y_i + \alpha_0 \right)}{(\beta_0 + N)^{\sum_{i=1}^N y_i + \alpha_0}} \right)$$

$$= \frac{\Gamma\left(\sum_{i=1}^N y_i + \alpha_0 \right)}{\Gamma(\alpha_0) \prod_{i=1}^N y_i!} \left(\frac{1}{\beta_0 + N} \right)^{\sum_{i=1}^N y_i} \left(\frac{\beta_0}{\beta_0 + N} \right)^{\alpha_0}.$$

R code. Health insurance, predictive distribution using vague hyperparameters

```
1  set.seed(010101)
2  y <- c(0, 3, 2, 1, 0) # Data
3  N <- length(y)
4
5  # Predictive distribution
6  ProbBo <- function(y, a0, b0){
7    N <- length(y)
8    #sample size
9    aN <- a0 + sum(y)
10   # Posterior shape parameter
11   bN <- b0 + N
12   # Posterior scale parameter
13   p <- 1 / (bN + 1)
14   # Probability negative binomial density
15   Pr <- 1 - pnbinom(0, size = aN, prob = (1 - p))
16   # Probability of visiting the Doctor
17   # Observe that in R there is a slightly
18   # different parametrization.
19   return(Pr)
20 }
21
22 # Using a vague prior:
23 a0 <- 0.001 # Prior shape parameter
24 b0 <- 0.001 # Prior scale parameter
25 PriMeanV <- a0 / b0 # Prior mean
26 PriVarV <- a0 / b0^2 # Prior variance
27 Pp <- ProbBo(y, a0 = 0.001, b0 = 0.001)
28 # This setting is vague prior information.
29 Pp
30 0.67
```

R code. Health insurance, predictive distribution using empirical Bayes

```r
# Using Emprirical Bayes
LogMgLik <- function(theta, y){
  N <- length(y)
  #sample size
  a0 <- theta[1]
  # prior shape hyperparameter
  b0 <- theta[2]
  # prior scale hyperparameter
  aN <- sum(y) + a0
  # posterior shape parameter
  if(a0 <= 0 || b0 <= 0){
    #Avoiding negative values
    lnp <- -Inf
  }else{lnp <- lgamma(aN) - sum(y)*log(b0+N) + a0*log(b0/(b0
    +N)) - lgamma(a0)}
    # log marginal likelihood
  return(-lnp)
}
theta0 <- c(0.01, 0.01)
# Initial values
control <- list(maxit = 1000)
# Number of iterations in optimization
EmpBay <- optim(theta0, LogMgLik, method = "BFGS", control =
    control, hessian = TRUE, y = y)
# Optimization
EmpBay$convergence
# Checking convergence
EmpBay$value # Maximum
a0EB <- EmpBay$par[1]
# Prior shape using empirical Bayes
a0EB
128.383
b0EB <- EmpBay$par[2]
# Prior scale using empirical Bayes
b0EB
106.801
PriMeanEB <- a0EB / b0EB
# Prior mean
PriVarEB <- a0EB / b0EB^2
# Prior variance
PpEB <- ProbBo(y, a0 = a0EB, b0 = b0EB)
# This setting is using empirical Bayes.
PpEB
0.69
```

4. Suppose that you are analyzing to buy a car insurance next year. To make

a better decision you want to know *what is the probability that you have a car claim next year?* You have the records of your car claims in the last 15 years, $\mathbf{y} = \{0, 1, 0, 1, 0, 1, 1, 0, 0, 1, 0, 0, 1, 1, 0\}$.

Assume that this is a random sample from a data generating process (statistical model) that is Bernoulli, $Y_i \sim Ber(p)$, and your probabilistic prior beliefs about p are well described by a beta distribution with parameters α_0 and β_0, $p \sim B(\alpha_0, \beta_0)$, then, you are interested in calculating the probability of a claim the next year $P(Y_0 = 1|\mathbf{y})$.

Solve this using an empirical Bayes approach and a non-informative approach where $\alpha_0 = \beta_0 = 1$ (uniform distribution).

Answer

The posterior distribution is given by

$$\pi\left(p|\mathbf{y}\right) = \left[\frac{\Gamma\left(\alpha_0 + \beta_0\right)}{\Gamma\left(\alpha_0\right)\Gamma\left(\beta_0\right)}p^{\alpha_0 - 1}\left(1 - p\right)^{\beta_0 - 1}\right]\left[\prod_{i=1}^{N}p^{y_i}\left(1 - p\right)^{1 - y_i}\right] \quad (1.9)$$

$$= \frac{\Gamma\left(\alpha_0 + \beta_0\right)}{\Gamma\left(\alpha_0\right)\Gamma\left(\beta_0\right)}p^{\sum_{i=1}^{N}y_i + \alpha_0 - 1}\left(1 - p\right)^{\beta_0 + N - \sum_{i=1}^{N}y_i - 1}$$

$$\propto p^{\sum_{i=1}^{N}y_i + \alpha_0 - 1}\left(1 - p\right)^{\beta_0 + N - \sum_{i=1}^{N}y_i - 1}.$$

The last expression is the kernel of a Beta distribution with parameters $\alpha_n = \sum_{i=1}^{N}y_i + \alpha_0$ and $\beta_n = \beta_0 + N - \sum_{i=1}^{N}y_i$. Thus, the posterior mean is

$$\begin{aligned}
E\left[p|\mathbf{y}\right] &= \frac{\alpha_n}{\alpha_n + \beta_n} \\
&= \frac{\sum_{i=1}^{N}y_i + \alpha_0}{\alpha_0 + \beta_0 + N} \\
&= \frac{N\bar{y}}{\alpha_0 + \beta_0 + N} + \frac{\alpha_0}{\alpha_0 + \beta_0 + N} \\
&= \frac{N}{\alpha_0 + \beta_0 + N}\left(\bar{y}\right) + \frac{\alpha_0 + \beta_0}{\alpha_0 + \beta_0 + N}\left(\frac{\alpha_0}{\alpha_0 + \beta_0}\right) \\
&= w\left(\bar{y}\right) + (1 - w)E\left[p\right],
\end{aligned} \quad (1.10)$$

where $w = \frac{N}{\alpha_0 + \beta_0 + N}$, $\bar{y}$ is the sample mean, and $E\left[p\right] = \frac{\alpha_0}{\alpha_0 + \beta_0}$ is the prior mean.

The posterior predictive distribution of claim the next year is given by

$$\begin{aligned}
\pi(Y_0 = 1|\mathbf{y}) &= \int_0^1 P(Y_0 = 1|\mathbf{y}, p)\pi\,(p|\mathbf{y})\,dp \\
&= \int_0^1 p \times \pi\,(p|\mathbf{y})\,dp \\
&= \mathbb{E}\,[p|\mathbf{y}] \\
&= \frac{\alpha_n}{\alpha_n + \beta_n}.
\end{aligned}$$

(1.11)

To use empirical Bayes, we have the following setting

$$\left[\hat{\alpha}_0 \;\; \hat{\beta}_0\right] = \arg\max_{\alpha_0, \beta_0} \ln(p(\mathbf{y})),$$

where

$$\begin{aligned}
p(\mathbf{y}) &= \int_0^1 \left[\frac{\Gamma\,(\alpha_0 + \beta_0)}{\Gamma\,(\alpha_0)\,\Gamma\,(\beta_0)} p^{\alpha_0 - 1}\,(1-p)^{\beta_0 - 1}\right]\left[\prod_{i=1}^N (1-p)^{1-y_i}\right]dp \quad (1.12) \\
&= \frac{\Gamma\,(\alpha_0 + \beta_0)}{\Gamma\,(\alpha_0)\,\Gamma\,(\beta_0)} \int_0^1 p^{\sum_{i=1}^N y_i + \alpha_0 - 1}\,(1-p)^{\beta_0 + N - \sum_{i=1}^N y_i - 1}\,dp \\
&= \frac{\Gamma\,(\alpha_0 + \beta_0)}{\Gamma\,(\alpha_0)\,\Gamma\,(\beta_0)} \frac{\Gamma\left(\sum_{i=1}^N y_i + \alpha_0\right)\Gamma\left(\beta_0 + N - \sum_{i=1}^N y_i\right)}{\Gamma\,(\alpha_0 + \beta_0 + N)}.
\end{aligned}$$

R code. Car claim, predictive distribution using vague hyperparameters

```r
set.seed(010101)
y <- c(0, 1, 0, 1, 0, 1, 1, 0, 0, 1, 0, 0, 1, 1, 0)
# Data
N <- length(y)
#require(TailRank)
# Predictive distribution
ProbBo <- function(y, a0, b0){
  N <- length(y)
  #sample size
  aN <- a0 + sum(y)
  # Posterior shape parameter
  bN <- b0 + N - sum(y)
  # Posterior scale parameter
  pr <- aN / (aN + bN)
  # Probability of a claim the next year
  return(pr)
}
# Using a vague prior:
a0 <- 1 # Prior shape parameter
b0 <- 1 # Prior scale parameter
PriMeanV <- a0 / (a0 + b0)
# Prior mean
PriVarV <- (a0*b0) / (((a0+b0)^2)*(a0+b0+1))
# Prior variance
Pp <- ProbBo(y, a0 = 1, b0 = 1)
# This setting is defining vague prior information.
# The probability of a claim
Pp
0.47
```

R code. Car claim, predictive distribution using empirical Bayes

```r
1  # Using Emprirical Bayes
2  LogMgLik <- function(theta, y){
3    N <- length(y)
4    #sample size
5    a0 <- theta[1]
6    # prior shape hyperparameter
7    b0 <- theta[2]
8    # prior scale hyperparameter
9    aN <- sum(y) + a0
10   # posterior shape parameter
11   if(a0 <= 0 || b0 <= 0){
12     #Avoiding negative values
13     lnp <- -Inf
14   }else{lnp <- lgamma(a0+b0) + lgamma(aN) + lgamma(b0+N-sum(
       y)) -lgamma(a0) - lgamma(b0) - lgamma(a0+b0+N)}
15   # log marginal likelihood
16   return(-lnp)
17 }
18 theta0 <- c(0.1, 0.1)
19 # Initial values
20 control <- list(maxit = 1000)
21 # Number of iterations in optimization
22 EmpBay <- optim(theta0, LogMgLik, method = "BFGS", control =
       control, hessian = TRUE, y = y)
23 # Optimization
24 EmpBay$convergence
25 # Checking convergence
26 EmpBay$value # Maximum
27 a0EB <- EmpBay$par[1]
28 # Prior shape using empirical Bayes
29 b0EB <- EmpBay$par[2]
30 # Prior scale using empirical Bayes
31 PriMeanEB <- a0EB /(a0EB + b0EB)
32 # Prior mean
33 PriVarEB <- (a0EB*b0EB)/(((a0EB+b0EB)^2)*(a0EB+b0EB+1))
34 # Prior variance
35 PpEB <- ProbBo(y, a0 = a0EB, b0 = b0EB)
36 # This setting is using empirical Bayes.
37 PpEB
38 0.47
```
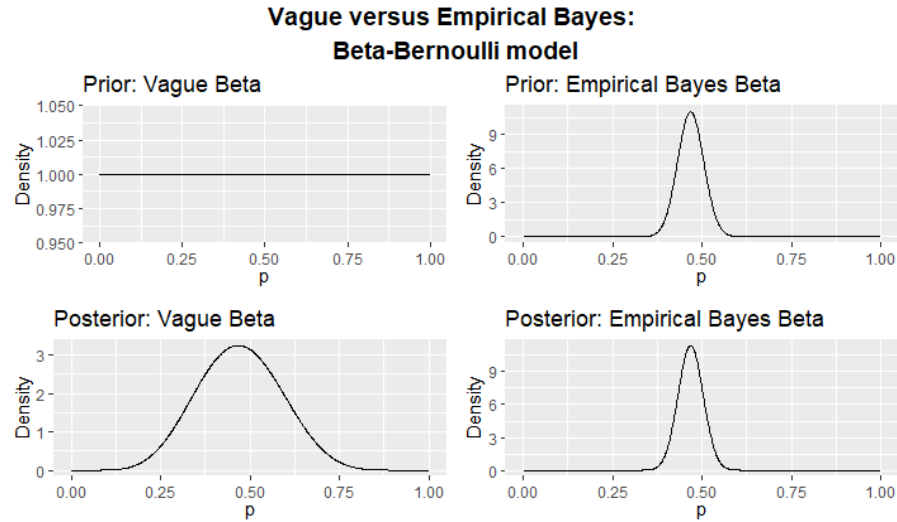
R code. Car claim, density plots

```r
1  # Density figures
2  lambda <- seq(0.001, 1, 0.001)
3  # Values of lambda
4  VaguePrior <- dbeta(lambda, shape1 = a0, shape2 = b0)
5  EBPrior <- dbeta(lambda, shape1 = a0EB, shape2 = b0EB)
6  PosteriorV <- dbeta(lambda, shape1 = a0 + sum(y), shape2 =
       b0 + N - sum(y))
7  PosteriorEB <- dbeta(lambda, shape1 = a0EB + sum(y), shape2
       = b0EB + N - sum(y))
8  # Likelihood function
9  Likelihood <- function(theta, y){
10   LogL <- dbinom(y, 1, theta, log = TRUE)
11   #  LogL <- dbern(y, theta)
12   Lik <- prod(exp(LogL))
13   return(Lik)
14 }
15 Liks <- sapply(lambda, function(par) {Likelihood(par, y = y)
       })
16 Sc <- max(PosteriorEB)/max(Liks)
17 #Scale for displaying in figure
18 LiksScale <- Liks * Sc
19 data <- data.frame(cbind(lambda, VaguePrior, EBPrior,
       PosteriorV, PosteriorEB, LiksScale))
20 #Data frame
21 require(ggplot2)
22 # Cool figures
23 require(latex2exp)
24 # LaTeX equations in figures
25 require(ggpubr)
26 # Multiple figures in one page
27 fig1 <- ggplot(data = data, aes(lambda, VaguePrior)) +
28   geom_line() +
29   xlab(TeX("$p$")) + ylab("Density") + ggtitle("Prior: Vague
       Beta")
30 fig2 <- ggplot(data = data, aes(lambda, EBPrior)) +
31   geom_line() +
32   xlab(TeX("$p$")) + ylab("Density") +
33   ggtitle("Prior: Empirical Bayes Beta")
34 fig3 <- ggplot(data = data, aes(lambda, PosteriorV)) +
35   geom_line() +
36   xlab(TeX("$p$")) + ylab("Density") +
37   ggtitle("Posterior: Vague Beta")
38 fig4 <- ggplot(data = data, aes(lambda, PosteriorEB)) +
39   geom_line() +
40   xlab(TeX("$p$")) + ylab("Density") +
41   ggtitle("Posterior: Empirical Bayes Beta")
42 FIG <- ggarrange(fig1, fig2, fig3, fig4,
43 ncol = 2, nrow = 2)
44 annotate_figure(FIG,
45 top = text_grob("Vague versus Empirical Bayes: Beta-
       Bernoulli model", color = "black", face = "bold", size =
       14))
```

Vague versus Empirical Bayes:
Beta-Bernoulli model

FIGURE 1.1
Vague versus Empirical Bayes: Bernoulli-Beta model.

R code. Car claim, prior, likelihood and posterior density plots

```
1  # Prior, likelihood and posterior:
2  #Empirical Bayes Binonial-Beta model
3  dataNew <- data.frame(cbind(rep(lambda, 3),
4  c(EBPrior, PosteriorEB, LiksScale),
5  rep(1:3, each = 1000)))
6  #Data frame
7
8  colnames(dataNew) <- c("Lambda", "Density", "Factor")
9  dataNew$Factor <- factor(dataNew$Factor, levels=c("1", "3",
10 "2"), labels=c("Prior", "Likelihood", "Posterior"))
11
12 ggplot(data = dataNew, aes_string(x = "Lambda",
13 y = "Density", group = "Factor")) +
14 geom_line(aes(color = Factor)) +
15 xlab(TeX("$\\lambda$")) + ylab("Density") +
16 ggtitle("Prior, likelihood and posterior: Empirical Bayes
17  Poisson-Gamma model") +
18 guides(color=guide_legend(title="Information")) +
19 scale_color_manual(values = c("red", "yellow", "blue"))
```

Prior, likelihood and posterior: Empirical Bayes Poisson-Gamma model

FIGURE 1.2
Prior, likelihood and posterior: Bernoulli-Beta model.

R code. Car claim, predictive probabilities plots

```r
# Predictive distributions
require(TailRank)
PredDen <- function(y, y0, a0, b0){
  N <- length(y)
  aN <- a0 + sum(y) # Posterior shape parameter
  bN <- b0 + N - sum(y) # Posterior scale parameter
  Pr <- aN/(aN+bN)
  Probs <- dbinom(y0, 1, prob = Pr)
  return(Probs)
}
y0 <- 0:1
PredVague <- PredDen(y = y, y0 = y0, a0 = a0, b0 = b0)
PredEB <- PredDen(y = y, y0 = y0, a0 = a0EB, b0 = b0EB)
dataPred <- as.data.frame(cbind(y0, PredVague, PredEB))
colnames(dataPred) <- c("y0", "PredictiveVague",
"PredictiveEB")
ggplot(data = dataPred) +
  geom_point(aes(y0, PredictiveVague, color = "red")) +
  xlab(TeX("$y_0$")) + ylab("Density") +
  ggtitle("Predictive density: Vague and Empirical Bayes
    priors") + geom_point(aes(y0, PredictiveEB, color = "
    yellow")) +
  guides(color = guide_legend(title="Prior")) +
  scale_color_manual(labels = c("Vague", "Empirical Bayes"),
    values = c("red", "yellow")) +
  scale_x_continuous(breaks=seq(0,1,by=1))
```

FIGURE 1.3
Predictive probabilities: Bernoulli-Beta model.

R code. Car claim, Bayesian model average

```r
1  # Posterior odds: Vague vs Empirical Bayes
2  PO12 <- exp(-LogMgLik(c(a0EB, b0EB), y = y))/exp(-LogMgLik(c
       (a0, b0), y = y))
3  PostProMEM <- PO12/(1 + PO12)
4  # Posterior model probability Empirical Bayes
5  PostProMEM
6  0.757
7  PostProbMV <- 1 - PostProMEM
8  # Posterior model probability vague prior
9  PostProbMV
10 0.242
11 # Bayesian model average (BMA)
12 PostMeanEB <- (a0EB + sum(y)) / (a0EB + b0EB + N)
13 # Posterior mean Empirical Bayes
14 PostMeanV <- (a0 + sum(y)) / (a0 + b0 + N)
15 # Posterior mean vague priors
16 BMAmean <- PostProMEM * PostMeanEB + PostProbMV * PostMeanV
17 # BMA posterior mean
18 PostVarEB <- (a0EB + sum(y))*(b0EB + N - sum(y)) / ((a0EB +
       b0EB + N)^2)*(a0EB + b0EB + N -1)
19 # Posterior variance Empirical Bayes
20 PostVarV <- (a0 + sum(y))*(b0 + N - sum(y)) / ((a0 + b0 + N)
       ^2)*(a0 + b0 + N -1)
21 # Posterior variance vague prior
22 BMAVar <- PostProMEM * PostVarEB + PostProbMV * PostVarV +
       PostProMEM * (PostMeanEB - BMAmean)^2 + PostProbMV * (
       PostMeanV - BMAmean)^2
23 # BMA posterior variance
24 # BMA: Predictive
25 BMAPred <- PostProMEM * PredEB + PostProbMV * PredVague
26 dataPredBMA <- as.data.frame(cbind(y0, BMAPred))
27 colnames(dataPredBMA) <- c("y0", "PredictiveBMA")
28 ggplot(data = dataPredBMA) +
29   geom_point(aes(y0, PredictiveBMA, color = "red")) +
30   xlab(TeX("$y_0$")) + ylab("Density") +
31   ggtitle("Predictive density: BMA") +
32   guides(color = guide_legend(title="BMA")) +
33   scale_color_manual(labels = c("Probability"), values = c("
       red")) + scale_x_continuous(breaks=seq(0,1,by=1))
```

R code. Car claim, Bayesian updating plots

```r
# Bayesian updating
BayUp <- function(y, lambda, a0, b0){
 N <- length(y)
 aN <- a0 + sum(y)
 # Posterior shape parameter
 bN <- b0 + N - sum(y)
 # Posterior scale parameter
 p <- dbeta(lambda, shape1 = aN, shape2 = bN)
 # Posterior density
 return(list(Post = p, a0New = aN, b0New = bN))
}
PostUp <- NULL
 for(i in 1:N){
 if(i == 1){
  PostUpi <- BayUp(y[i], lambda, a0 = 1, b0 = 1)}
  else{
  PostUpi <- BayUp(y[i], lambda,
  a0 = PostUpi$a0New, b0 = PostUpi$b0New)
 }
 PostUp <- cbind(PostUp, PostUpi$Post)
}
DataUp <- data.frame(cbind(rep(lambda, 15), c(PostUp), rep
     (1:15, each = 1000)))   #Data frame
colnames(DataUp) <- c("Lambda", "Density", "Factor")
DataUp$Factor <- factor(DataUp$Factor, levels=c("1","2","3",
     "4","5","6","7","8","9","10","11","12","13","14","15"),
     labels=c("Iter_1","Iter_2","Iter_3","Iter_4","Iter_5","
     Iter_6","Iter_7","Iter_8","Iter_9","Iter_10","Iter_11","
     Iter_12","Iter_13","Iter_14","Iter_15"))
ggplot(data = DataUp, aes_string(x = "Lambda",
 y = "Density", group = "Factor")) +
 geom_line(aes(color = Factor)) +
 xlab(TeX("$p$")) + ylab("Density") +
 ggtitle("Bayesian updating:
 Beta-Binomial model with vague prior") +
 guides(color=guide_legend(title="Update"))
```

5. Show that given the loss function, $L(\theta, a) = |\theta - a|$, then the optimal decision rule minimizing the risk function, $a^*(\mathbf{y})$, is the median.

Answer

$\int_{\Theta} |\theta - a| \pi(\theta|\mathbf{y}) d\theta = \int_{-\infty}^{a} (a - \theta) \pi(\theta|\mathbf{y}) d\theta + \int_{a}^{\infty} (\theta - a) \pi(\theta|\mathbf{y}) d\theta$. Differentiating with respect to a, and equating to zero,

$$\int_{-\infty}^{a} \pi(\theta|\mathbf{y}) d\theta = \int_{a}^{\infty} \pi(\theta|\mathbf{y}) d\theta, \qquad (1.13)$$

FIGURE 1.4
Predictive probabilities: Bernoulli-Beta Bayesian model average.

FIGURE 1.5
Predictive probabilities: Bernoulli-Beta Bayesian model updating.

then,

$$2 \int_{-\infty}^{a} \pi(\theta|\mathbf{y})d\theta = \int_{-\infty}^{\infty} \pi(\theta|\mathbf{y})d\theta = 1, \tag{1.14}$$

that is, $a^*(\mathbf{y})$ is the median.

2

Conceptual differences of the Bayesian and Frequentist approaches

2.1   Solutions of Exercises

1. **Jeffreys-Lindley's paradox**

   The **Jeffreys-Lindley's paradox** [5, 7] is an apparent disagreement between the Bayesian and Frequentist frameworks to a hypothesis testing situation.

   In particular, assume that in a city 49,581 boys and 48,870 girls have been born in 20 years. Assume that the male births is distributed Binomial with probability θ. We want to test the null hypothesis H_0. $\theta = 0.5$ versus H_1. $\theta \neq 0.5$.

   - Show that the posterior model probability for the model under the null is approximately 0.95. Assume $\pi(H_0) = \pi(H_1) = 0.5$, and $\pi(\theta)$ equal to $U(0, 1)$ under H_1.

   - Show that the p-value for this hypothesis test is equal to 0.023 using the normal approximation, $Y \sim N(N \times \theta, N \times \theta \times (1 - \theta))$.

   Answer

   - The marginal likelihood under the null hypothesis is $p(y|H_0) = \binom{n}{y} \theta^y (1 - \theta)^{n-y} \approx 1.95 \times 10^{-4}$ given $\theta = 0.5$ under H_0, $N = 49,581 + 48,870$ and $y = 49,581$. On the other hand, the marginal likelihood under the alternative hypothesis is

$$p(y|H_1) = \int_0^1 \binom{n}{y} \theta^y (1-\theta)^{n-y} d\theta$$

$$= \binom{n}{y} B(y+1, n-k+1)$$

$$= \frac{\Gamma(N+1)}{\Gamma(y+1)\Gamma(n-y+1)} \frac{\Gamma(y+1)\Gamma(N-y+1)}{\Gamma(N+2)}$$

$$= \frac{N!}{(N+1)!}$$

$$= \frac{1}{N+1}$$

$$\approx 1.016 \times 10^{-5}.$$

Then, $PO_{01} = \frac{1.95 \times 10^{-4}}{1.016 \times 10^{-5}} = 19.19$, this implies that the posterior model probability under the null hypothesis is $\pi(H_0|y) = \frac{19.19}{1+19.19} = 0.95$.

- Under the null hypothesis,

$$p = 2 \int_{49,581}^{\infty} (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2}(y-\mu)^2\right\} dy$$

$$= 0.0235,$$

where $\mu = N \times \theta = 49,225.5$, and $\sigma^2 = N \times \theta \times (1-\theta) = 24,612.75$ under the null hypothesis ($\theta = 0.5$).

Observe that the posterior model probability supports the null hypothesis, whereas the p-value implies rejection of the null hypothesis using a 5% significance level.

Observe that actually this is not a paradox, as we are answering two different questions. The Bayes factor is comparing two models ($\theta = 0.5$ versus $\theta \sim U(0,1)$), whereas the *p*-value is checking the compatibility between $\theta = 0.5$ and the sample information. Despite that $\theta = 0.5$ is not compatible with sample information, it is better than the models assuming $\theta \sim U(0,1)$ as most of these values of θ are far away from the sample mean. Thus, the model under the null is a bad description of the data, but it is better than the model under the alternative hypothesis.[1]

---

[1]Observe that there are at least another two issues in this example. First, the prior under the alternative is non-informative, this implies problems for Bayes factors, and second, the prior under the alternative is positive at $\theta = 0.5$, which is the null ([6] propose non-local prior densities in Bayesian hypothesis tests to tackle these issues).

2. We want to test H_0. $\mu = \mu_0$ vs H_1. $\mu \neq \mu_0$ given $y_i \overset{iid}{\sim} N(\mu, \sigma^2)$.

   Assume $\pi(H_0) = \pi(H_1) = 0.5$, and $\pi(\mu, \sigma) \propto 1/\sigma$ under the alternative hypothesis.

   Show that

   $$p(\mathbf{y}|\mathcal{M}_1) = \frac{\pi^{-N/2}}{2}\Gamma(N/2)2^{N/2}\left(\frac{1}{\alpha_n\hat{\sigma}^2}\right)^{N/2}\left(\frac{N}{\alpha_n\hat{\sigma}^2}\right)^{-1/2}\frac{\Gamma(1/2)\Gamma(\alpha_n/2)}{\Gamma((\alpha_N+1)/2)} \quad \text{and}$$

   $$p(\mathbf{y}|\mathcal{M}_0) = (2\pi)^{-N/2}\left[\frac{2}{\Gamma(N/2)}\left(\frac{N}{2}\frac{\sum_{i=1}^{N}(y_i - \mu_0)^2}{N}\right)^{N/2}\right]^{-1}. \text{ Then,}$$

   $$PO_{01} = \frac{p(\mathbf{y}|\mathcal{M}_0)}{p(\mathbf{y}|\mathcal{M}_1)}$$

   $$= \frac{\Gamma((\alpha_n + 1)/2)}{\Gamma(1/2)\Gamma(\alpha_N/2)}(\alpha_n\hat{\sigma}^2/N)^{-1/2}\left[1 + \frac{(\mu_0 - \bar{y})^2}{\alpha_n\hat{\sigma}^2/N}\right]^{-\left(\frac{\alpha_n+1}{2}\right)},$$

   where $\alpha_N = N - 1$ and $\hat{\sigma}^2 = \frac{\sum_{i=1}^{N}(y_i - \bar{y})^2}{N-1}$.

   Find the relationship between the posterior odds and the classical test statistic for the null hypothesis.

   Answer

   $$p(\mathbf{y}|\mathcal{M}_1) = \int_{-\infty}^{\infty}\int_{0}^{\infty}(2\pi)^{-N/2}\sigma^{-N}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - \mu)^2\right\}\frac{1}{\sigma}d\sigma d\mu$$

   $$= (2\pi)^{-N/2}\int_{-\infty}^{\infty}\int_{0}^{\infty}\sigma^{-(N+1)}\exp\left\{-\frac{N}{2\sigma^2}\frac{\sum_{i=1}^{N}(y_i - \mu)^2}{N}\right\}d\sigma d\mu$$

   $$= (2\pi)^{-N/2}\frac{\Gamma(N/2)}{2}2^{N/2}\int_{-\infty}^{\infty}\left[\sum_{i=1}^{N}(y_i - \mu)^2\right]^{-N/2}d\mu$$

   $$= (2\pi)^{-N/2}\frac{\Gamma(N/2)}{2}2^{N/2}\int_{-\infty}^{\infty}\left[\sum_{i=1}^{N}[(y_i - \bar{y}) - (\mu - \bar{y})]^2\right]^{-N/2}d\mu$$

   $$= (2\pi)^{-N/2}\frac{\Gamma(N/2)}{2}2^{N/2}\int_{-\infty}^{\infty}\left[\alpha_n\hat{\sigma}^2 + N(\mu - \bar{y})^2\right]^{-N/2}d\mu$$

   $$= (2\pi)^{-N/2}\frac{\Gamma(N/2)}{2}2^{N/2}\left(\frac{\alpha_n\hat{\sigma}^2}{\alpha_n\hat{\sigma}^2}\right)^{-N/2}\int_{-\infty}^{\infty}\left[\alpha_n\hat{\sigma}^2 + N(\mu - \bar{y})^2\right]^{-N/2}d\mu$$

   $$= (2\pi)^{-N/2}\frac{\Gamma(N/2)}{2}2^{N/2}\left(\alpha_n\hat{\sigma}^2\right)^{-N/2}\int_{-\infty}^{\infty}\left[1 + \frac{N(\mu - \bar{y})^2}{\alpha_n\hat{\sigma}^2}\right]^{-N/2}d\mu$$

   $$= \frac{\pi^{-N/2}}{2}\Gamma(N/2)2^{N/2}\left(\frac{1}{\alpha_n\hat{\sigma}^2}\right)^{N/2}\left(\frac{N}{\alpha_n\hat{\sigma}^2}\right)^{-1/2}\frac{\Gamma(1/2)\Gamma(\alpha_n/2)}{\Gamma((\alpha_N+1)/2)}.$$

   The third line takes into account that the integral in the second line is the kernel of an inverted-gamma distribution, and the last line takes into account that the integral in the previous line is the kernel of a student's t distribution [12].

$$p(\mathbf{y}|\mathcal{M}_0) = \int_0^\infty (2\pi)^{-N/2}\sigma^{-N} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^N (y_i - \mu_0)^2\right\} \frac{1}{\sigma} d\sigma$$

$$= (2\pi)^{-N/2} \int_0^\infty \sigma^{-(N+1)} \exp\left\{-\frac{N}{2\sigma^2}\frac{\sum_{i=1}^N (y_i - \mu_0)^2}{N}\right\} d\sigma$$

$$= (2\pi)^{-N/2}\left[\frac{2}{\Gamma(N/2)}\left(\frac{N}{2}\frac{\sum_{i=1}^N (y_i - \mu_0)^2}{N}\right)^{N/2}\right]^{-1}.$$

The third line takes into account that the integral in the second line is the kernel of an inverted-gamma distribution [12].

Given these results is easy to get PO_{01}.

In addition,

$$PO_{01} = \frac{\Gamma((\alpha_n + 1)/2)}{\Gamma(1/2)\Gamma(\alpha_N/2)}(\alpha_n\hat{\sigma}^2/N)^{-1/2}\left[1 + \frac{(\mu_0 - \bar{y})^2}{\alpha_n\hat{\sigma}^2/N}\right]^{-\left(\frac{\alpha_n+1}{2}\right)}$$

$$= \frac{\Gamma((\alpha_n + 1)/2)}{\Gamma(1/2)\Gamma(\alpha_N/2)}(\alpha_n\hat{\sigma}^2/N)^{-1/2}\left[1 + \frac{1}{\alpha_n}\left(\frac{\mu_0 - \bar{y}}{\hat{\sigma}/\sqrt{N}}\right)^2\right]^{-\left(\frac{\alpha_n+1}{2}\right)}$$

$$= \frac{\Gamma((\alpha_n + 1)/2)}{\Gamma(1/2)\Gamma(\alpha_N/2)}(\alpha_n\hat{\sigma}^2/N)^{-1/2}\left[1 + \frac{1}{\alpha_n}t^2\right]^{-\left(\frac{\alpha_n+1}{2}\right)},$$

where $t = \frac{\bar{y}-\mu_0}{\hat{\sigma}/\sqrt{N}}$ is the classical statistical test. Then, as t increases then the PO_{01} decreases, both indicating support against the null hypothesis H_0. $\mu = \mu_0$. However, there are other terms affecting the posterior odds, then, there is no necessary agreement between the classical test statistic and the posterior odds.

3. **Math test continues**

   Using the setting of the **Example: Math test** in subsection 2.6.1 in the book, test H_0. $\mu = \mu_0$ vs H_1. $\mu \neq \mu_0$ where $\mu_0 = \{100, 100.5, 101, 101.5, 102\}$.

   - What is the p-value for these hypothesis tests?
   - Find the posterior model probability of the null model for each μ_0.

   Answer

R code. Example: Math test

```r
1  N <- 50 # Sample size
2  y_bar <- 102 # Sample mean
3  s2 <- 10 # Sample variance
4  alpha <- N - 1
5  serror <- (s2/N)^0.5
6  y.H0 <- c(100, 100.5, 101, 101.5, 102)
7  test <- (y.H0 - y_bar)/serror
8  pval <- 2*pt(test, alpha)
9  pval
10 0.0000459 0.0015431 0.0299338 0.2690040 1
11 # p-values
12 PO01 <- (gamma(N/2)*((N-1)*serror^2)^(-0.5)*(1+test^2/alpha)
       ^(-N/2))/(gamma(1/2)*gamma((N-1)/2))
13 PO01/(1+PO01)
14 0.0001705 0.0050345 0.0725330 0.3210223 0.4702050
15 # Posterior model probability of the null hypothesis.
```

3

Cornerstone models: Conjugate families

3.1 Solutions of Exercises

1. Write in the canonical form the distribution of the Bernoulli example, and find the mean and variance of the sufficient statistic.

   Answer

   Given $p(\mathbf{y}|\theta) = (1-\theta)^N \exp\left\{\sum_{i=1}^{N} y_i \log\left(\frac{\theta}{1-\theta}\right)\right\}$ where $\eta = \log\frac{\theta}{1+\theta}$ which implies $\theta = \frac{\exp(\eta)}{1-\exp(\eta)}$, then $p(\mathbf{y}|\theta) = \exp\left\{\sum_{i=1}^{N} y_i\eta - N\log(1+\exp(\eta))\right\}$. Thus $B(\eta) = N\log(1+\exp(\eta))$, $\nabla(B(\eta)) = N\frac{\exp(\eta)}{1+\exp(\eta)} = N\theta$ and $\nabla^2(B(\eta)) = N\left\{\frac{\exp(\eta)(1+\exp(\eta))}{(1+\exp(\eta))^2} - \frac{\exp(\eta)\exp(\eta)}{(1+\exp(\eta))^2}\right\} = N\theta(1-\theta)$.

2. Given a random sample $\mathbf{y} = [y_1, y_2, \ldots, y_N]^\top$ from N *binomial experiments* each having known size n_i and same unknown probability θ. Show that $p(\mathbf{y}|\theta)$ is in the exponential family, and find the posterior distribution, the marginal likelihood and the predictive distribution of the binomial-beta model assuming the number of trials is known.

   Answer

   The density function is

   $$
   \begin{aligned}
   p(\mathbf{y}|\theta) &= \prod_{i=1}^{N} \binom{n_i}{y_i} \theta^{y_i}(1-\theta)^{n_i-y_i} \\
   &= \prod_{i=1}^{N} \binom{n_i}{y_i} \theta^{\sum_{i=1}^{N} y_i}(1-\theta)^{\sum_{i=1}^{N} n_i - \sum_{i=1}^{N} y_i} \\
   &= \prod_{i=1}^{N} \binom{n_i}{y_i} \exp\left\{\sum_{i=1}^{N} y_i \log\left(\frac{\theta}{1-\theta}\right) + \sum_{i=1}^{N} n_i \log(1-\theta)\right\} \\
   &= \prod_{i=1}^{N} \binom{n_i}{y_i} (1-\theta)^{\sum_{i=1}^{N} n_i} \exp\left\{\sum_{i=1}^{N} y_i \log\left(\frac{\theta}{1-\theta}\right)\right\},
   \end{aligned}
   $$

   Observe that $\sum_{i=1}^{N} n_i$ is the total sample size of Bernoulli experiments.

Using Theorem 1 in Chapter 4, the prior distribution is

$$\pi(\theta) \propto (1-\theta)^{B_0} \exp\left\{a_0 \log\left(\frac{\theta}{1-\theta}\right)\right\}$$
$$= \theta^{a_0}(1-\theta)^{B_0-a_0}$$
$$= \theta^{\alpha_0-1}(1-\theta)^{\beta_0-1},$$

where $\alpha_0 = a_0 + 1$ and $\beta_0 = B_0 - a_0 + 1$. This is the kernel of a beta distribution. Thus, the posterior distribution is

$$\pi(\theta|\mathbf{y}) \propto \theta^{\alpha_0-1}(1-\theta)^{\beta_0-1} \times \theta^{\sum_{i=1}^{N} y_i}(1-\theta)^{\sum_{i=1}^{N} n_i - \sum_{i=1}^{N} y_i}$$
$$= \theta^{\alpha_0 + \sum_{i=1}^{N} y_i - 1}(1-\theta)^{\beta_0 + \sum_{i=1}^{N} n_i - \sum_{i=1}^{N} y_i - 1}$$
$$= \theta^{\alpha_n-1}(1-\theta)^{\beta_n-1},$$

where $\alpha_n = \alpha_0 + \sum_{i=1}^{N} y_i$ and $\beta_n = \beta_0 + \sum_{i=1}^{N} n_i - \sum_{i=1}^{N} y_i$.
The marginal likelihood is

$$p(\mathbf{y}) = \int_0^1 \frac{\theta^{\alpha_0-1}(1-\theta)^{\beta_0-1}}{B(\alpha_0,\beta_0)} \times \prod_{i=1}^{N}\binom{n_i}{y_i}\theta^{\sum_{i=1}^{N} y_i}(1-\theta)^{\sum_{i=1}^{N} n_i - \sum_{i=1}^{N} y_i} d\theta$$
$$= \frac{\prod_{i=1}^{N}\binom{n_i}{y_i}}{B(\alpha_0,\beta_0)} \int_0^1 \theta^{\alpha_0 + \sum_{i=1}^{N} y_i - 1}(1-\theta)^{\beta_0 \sum_{i=1}^{N} n_i - \sum_{i=1}^{N} y_i - 1} d\theta$$
$$= \frac{\prod_{i=1}^{N}\binom{n_i}{y_i}B(\alpha_n,\beta_n)}{B(\alpha_0,\beta_0)}.$$

The third line due to having the kernel of a Beta distribution.

Finally, the predictive distribution is

$$p(Y_0|\mathbf{y}) = \int_0^1 \binom{n_{y_0}}{y_0}\theta^{y_0}(1-\theta)^{n_{y_0}-y_0}\frac{\theta^{\alpha_n-1}(1-\theta)^{\beta_n-1}}{B(\alpha_n,\beta_n)}d\theta$$
$$= \frac{\binom{n_{y_0}}{y_0}}{B(\alpha_n,\beta_n)}\int_0^1 \theta^{\alpha_n+y_0-1}(1-\theta)^{\beta_n+n_{y_0}-y_0-1}d\theta$$
$$= \binom{n_{y_0}}{y_0}\frac{B(\alpha_n+y_0,\beta_n+n_{y_0}-y_0)}{B(\alpha_n,\beta_n)},$$

where n_{y_0} is the known size associated with y_0, and the last line due to having the kernel of a beta distribution. The predictive is a *beta-binomial distribution*.

3. Given a random sample $\mathbf{y} = [y_1, y_2, \ldots, y_N]^\top$ from a *exponential distribution*. Show that $p(\mathbf{y}|\lambda)$ is in the exponential family, and find the posterior distribution, marginal likelihood and predictive distribution of the exponential-gamma model.

Answer

We see that the exponential distribution belongs to the exponential family as $p(\mathbf{y}|\lambda) = \prod_{i=1}^{N} \lambda \exp(-\lambda y_i) = \lambda^N \exp(-\lambda \sum_{i=1}^{N} y_i)$.

Using the gamma distribution in the rate parametrization, we see that $\pi(\lambda|\mathbf{y}) \propto \lambda^{\alpha_0-1} \exp(-\lambda \beta_0) \times \lambda^N \exp(-\lambda \sum_{i=1}^{N} y_i) = \lambda^{\alpha_0+N-1} \exp(-\lambda(\beta_0 + \sum_{i=1}^{N} y_i))$. This is the kernel of a gamma distribution, that is, $\lambda|\mathbf{y} \sim G(\alpha_n, \beta_n)$ where $\alpha_n = \alpha_0 + N$ and $\beta_n = \beta_0 + \sum_{i=1}^{N} y_i$.

The marginal likelihood is

$$
\begin{aligned}
p(\mathbf{y}) &= \int_0^\infty \lambda^N \exp\left\{-\lambda \sum_{i=1}^{N}\right\} \lambda^{\alpha_0-1} \exp\left\{-\beta_0\lambda\right\} \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} d\lambda \\
&= \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \int_0^\infty \lambda^{\alpha_0+N-1} \exp\left\{-\lambda\left(\beta_0 + \sum_{i=1}^{N}\right)\right\} d\lambda \\
&= \frac{\beta_0^{\alpha_0}\Gamma(\alpha_n)}{\Gamma(\alpha_0)\beta_n^{\alpha_n}}.
\end{aligned}
$$

Finally, the predictive distribution is

$$
\begin{aligned}
p(Y_0|\mathbf{y}) &= \int_0^\infty \lambda \exp\left\{-\lambda y_0\right\} \lambda^{\alpha_n-1} \exp\left\{-\beta_n\lambda\right\} \frac{\beta_n^{\alpha_n}}{\Gamma(\alpha_n)} d\lambda \\
&= \frac{\beta_n^{\alpha_n}}{\Gamma(\alpha_n)} \int_0^\infty \lambda^{\alpha_n+1-1} \exp\left\{-\lambda(\beta_n + y_0)\right\} d\lambda \\
&= \frac{\beta_n^{\alpha_n}}{\Gamma(\alpha_n)} \times \frac{\Gamma(\alpha_n + 1)}{(\beta_n + y_0)^{\alpha_n+1}} \\
&= \frac{\alpha_n \beta_n^{\alpha_n}}{(\beta_n + y_0)^{\alpha_n+1}}.
\end{aligned}
$$

This is a *Lomax distribution.*

4. Given $\mathbf{y} \sim N_N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, that is, a *multivariate normal distribution* show that $p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is in the exponential family.

Answer

$$p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-N/2}|\boldsymbol{\Sigma}|^{-1/2} \exp\left\{-\frac{1}{2}\left(\mathbf{y}-\boldsymbol{\mu}\right)^{\top}\boldsymbol{\Sigma}^{-1}\left(\mathbf{y}-\boldsymbol{\mu}\right)\right\}$$

$$= (2\pi)^{-N/2} \exp\left\{-\frac{1}{2}\left(\mathbf{y}^{\top}\boldsymbol{\Sigma}^{-1}\mathbf{y} - 2\mathbf{y}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \boldsymbol{\mu}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \log(|\boldsymbol{\Sigma}|)\right)\right\}$$

$$= (2\pi)^{-N/2} \exp\left\{-\frac{1}{2}\left(tr\left\{\mathbf{y}^{\top}\boldsymbol{\Sigma}^{-1}\mathbf{y}\right\} - 2\mathbf{y}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \boldsymbol{\mu}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \log(|\boldsymbol{\Sigma}|)\right)\right\}$$

$$= (2\pi)^{-N/2} \exp\left\{-\frac{1}{2}\left(vec\left(\mathbf{y}\mathbf{y}^{\top}\right)^{\top}vec\left(\boldsymbol{\Sigma}^{-1}\right) - 2\mathbf{y}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \boldsymbol{\mu}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \log(|\boldsymbol{\Sigma}|)\right)\right\},$$

where tr and vec are the trace and vectorization operators, respectively.

Then, $h(\mathbf{y}) = (2\pi)^{-N/2}$, $\eta(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \left[\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \ \ vec\left(\boldsymbol{\Sigma}^{-1}\right)\right]$, $T(\mathbf{y}) = \left[\mathbf{y} \ \ -\frac{1}{2}vec(\mathbf{y}\mathbf{y}^{\top})\right]$ and $C(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \exp\left\{-\frac{1}{2N}\left(\boldsymbol{\mu}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \log(|\boldsymbol{\Sigma}|)\right)\right\}$.

5. Find the marginal likelihood in the normal/inverse-Wishart model.

Answer

$$p(\mathbf{Y}) = \int_{\mathcal{R}^p}\int_{\mathcal{S}}(2\pi)^{-pN/2}|\boldsymbol{\Sigma}|^{-N/2}\exp\left\{-\frac{1}{2}tr[(\mathbf{S}+N(\boldsymbol{\mu}-\hat{\boldsymbol{\mu}})(\boldsymbol{\mu}-\hat{\boldsymbol{\mu}})^{\top})\boldsymbol{\Sigma}^{-1}]\right\}$$

$$\times (2\pi)^{-p/2}\beta_0^{p/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\left\{-\frac{\beta_0}{2}tr[(\boldsymbol{\mu}-\boldsymbol{\mu}_0)(\boldsymbol{\mu}-\boldsymbol{\mu}_0)^{\top}\boldsymbol{\Sigma}^{-1}]\right\}$$

$$\times |\boldsymbol{\Sigma}|^{-(\alpha_0+p+1)/2}\frac{2^{-\alpha_0 p/2}|\boldsymbol{\Psi}_0|^{\alpha_0/2}}{\Gamma_p(\alpha_0/2)}\exp\left\{-\frac{1}{2}tr(\boldsymbol{\Psi}_0\boldsymbol{\Sigma}^{-1})\right\}d\boldsymbol{\Sigma}d\boldsymbol{\mu}$$

$$= \frac{(2\pi)^{-frac12(pN+p)}|\boldsymbol{\Psi}_0|^{\alpha_0/2}\beta_0^{p/2}2^{-\alpha_0 p/2}}{\Gamma_p(\alpha_0/2)}\int_{\mathcal{R}^p}\int_{\mathcal{S}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}(N+1+\alpha_0+p+1)}$$

$$\times \exp\left\{-\frac{1}{2}tr[(\mathbf{S}+N(\boldsymbol{\mu}-\hat{\boldsymbol{\mu}})(\boldsymbol{\mu}-\hat{\boldsymbol{\mu}})^{\top}+\beta_0(\boldsymbol{\mu}-\boldsymbol{\mu}_0)(\boldsymbol{\mu}-\boldsymbol{\mu}_0)^{\top}+\boldsymbol{\Psi}_0)\boldsymbol{\Sigma}^{-1}]\right\}d\boldsymbol{\Sigma}d\boldsymbol{\mu}.$$

We have in the integral the kernel of an Inverse-Wishart distribution, then

$$p(\mathbf{Y}) = \frac{\Gamma_p\left(\frac{N+1+\alpha_0}{2}\right)|\mathbf{\Psi}_0|^{\alpha_0/2}\beta_0^{p/2}}{\Gamma_p(\alpha_0/2)\pi^{p(N+1)/2}}$$

$$\times \int_{\mathcal{R}^p} |\mathbf{S} + \mathbf{\Psi}_0 + (N+\beta_0)(\boldsymbol{\mu}-\boldsymbol{\mu}_n)(\boldsymbol{\mu}-\boldsymbol{\mu}_n)^\top$$

$$+ N\beta_0/(N+\beta_0)(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu}_0)(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu}_0)^\top|d\boldsymbol{\mu}$$

$$= \frac{\Gamma_p\left(\frac{N+1+\alpha_0}{2}\right)|\mathbf{\Psi}_0|^{\alpha_0/2}\beta_0^{p/2}}{\Gamma_p(\alpha_0/2)\pi^{p(N+1)/2}}$$

$$\times \int_{\mathcal{R}^p} |\mathbf{\Psi}_n||1 + \beta_n(\boldsymbol{\mu}-\boldsymbol{\mu}_n)\mathbf{\Psi}_n^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_n)^\top|^{-\frac{1}{2}(\alpha_n+1)}d\boldsymbol{\mu}$$

$$= \frac{\Gamma_p\left(\frac{\alpha_n+1}{2}\right)|\mathbf{\Psi}_0|^{\alpha_0/2}\beta_0^{p/2}}{\Gamma_p(\alpha_0/2)\pi^{p(N+1)/2}}|\mathbf{\Psi}_n|^{-\frac{1}{2}(\alpha_n+1)}$$

$$\times \int_{\mathcal{R}^p} [1 + \beta_n(\boldsymbol{\mu}-\boldsymbol{\mu}_n)^\top\mathbf{\Psi}_n^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_n)]^{-\frac{1}{2}(\alpha_n+1)}d\boldsymbol{\mu}.$$

The last equality uses the definition of $\mathbf{\Psi}_n$, β_n and α_n, and the Sylvester's determinant theorem. Observe that we have the kernel of a multivariate t distribution [8]. Then,

$$p(\mathbf{Y}) = \frac{\Gamma_p\left(\frac{\alpha_n+1}{2}\right)|\mathbf{\Psi}_0|^{\alpha_0/2}\beta_0^{p/2}}{\Gamma_p(\alpha_0/2)\pi^{p(N+1)/2}}|\mathbf{\Psi}_n|^{-\frac{1}{2}(\alpha_n+1)}$$

$$\times \int_{\mathcal{R}^p} \left[1 + \frac{1}{\alpha_n+1-p}(\boldsymbol{\mu}-\boldsymbol{\mu}_n)^\top\left(\frac{\mathbf{\Psi}_n}{\beta_n(\alpha_n+1-p)}\right)^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_n)\right]^{-\frac{1}{2}(\alpha_n+1-p+p)} d\boldsymbol{\mu}$$

$$= \frac{\Gamma_p\left(\frac{\alpha_n+1}{2}\right)\Gamma_p\left(\frac{\alpha_n+1-p}{2}\right)|\mathbf{\Psi}_0|^{\alpha_0/2}\beta_0^{p/2}(\alpha_n+1-p)^{p/2}\pi^{p/2}|\mathbf{\Psi}_n|^{-\frac{1}{2}(\alpha_n+1)}}{\Gamma_p(\alpha_0/2)\pi^{p(N+1)/2}\Gamma_p\left(\frac{\alpha_n+1-p+p}{2}\right)\left(\frac{\mathbf{\Psi}_n}{\alpha_n+1-p}\right)^{-1/2}}$$

$$= \frac{\Gamma_p\left(\frac{v_n}{2}\right)}{\Gamma_p\left(\frac{\alpha_0}{2}\right)}\frac{|\mathbf{\Psi}_0|^{\alpha_0/2}}{|\mathbf{\Psi}_n|^{\alpha_n/2}}\left(\frac{\beta_0}{\beta_n}\right)^{p/2}(2\pi)^{-Np/2},$$

where $v_n = \alpha_n + 1 - p$.

6. Find the posterior predictive distribution in the normal/inverse-Wishart model, and show that $\mathbf{Y}_0|\mathbf{Y} \sim T_{N_0,M}(\alpha_n - M + 1, \mathbf{X}_0\mathbf{B}_n, \mathbf{I}_{N_0} + \mathbf{X}_0\mathbf{V}_n\mathbf{X}_0^\top, \mathbf{\Psi}_n)$ in the multivariate regression linear model.

   Answer

$$p(\mathbf{Y}_0|\mathbf{Y}) \propto \int_{\mathcal{R}^p} \int_{\mathcal{S}} |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{-\frac{1}{2}tr[(\mathbf{y}_0 - \boldsymbol{\mu})(\mathbf{y}_0 - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}]\right\}$$

$$\times |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{-\frac{\beta_n}{2}tr[(\boldsymbol{\mu} - \boldsymbol{\mu}_n)(\boldsymbol{\mu} - \boldsymbol{\mu}_n)^\top \boldsymbol{\Sigma}^{-1}]\right\}$$

$$\times |\boldsymbol{\Sigma}|^{-(\alpha_n + p + 1)/2} \exp\left\{-\frac{1}{2}tr(\boldsymbol{\Psi}_n \boldsymbol{\Sigma}^{-1})\right\} d\boldsymbol{\Sigma} d\boldsymbol{\mu}$$

$$\propto \int_{\mathcal{R}^p} |(\mathbf{y}_0 - \boldsymbol{\mu})(\mathbf{y}_0 - \boldsymbol{\mu})^\top + (\boldsymbol{\mu} - \boldsymbol{\mu}_n)(\boldsymbol{\mu} - \boldsymbol{\mu}_n)^\top + \boldsymbol{\Psi}_n|^{-(\alpha_n + 2)/2} d\boldsymbol{\mu}.$$

The last equality uses that there is the kernel of an Inverse Wishart distribution.

Taking into account that

$$(\mathbf{y}_0 - \boldsymbol{\mu})(\mathbf{y}_0 - \boldsymbol{\mu})^\top + (\boldsymbol{\mu} - \boldsymbol{\mu}_n)(\boldsymbol{\mu} - \boldsymbol{\mu}_n)^\top = (1 + \beta_n)\left(\boldsymbol{\mu} - \frac{(\mathbf{y}_0 + \beta_n \boldsymbol{\mu}_n)}{1 + \beta_n}\right)\left(\boldsymbol{\mu} - \frac{(\mathbf{y}_0 + \beta_n \boldsymbol{\mu}_n)}{1 + \beta_n}\right)^\top$$

$$+ \frac{\beta_n}{1 + \beta_n}(\mathbf{y}_0 - \boldsymbol{\mu}_n)(\mathbf{y}_0 - \boldsymbol{\mu}_n)^\top.$$

Then,

$$p(\mathbf{Y}_0|\mathbf{Y}) \propto \int_{\mathcal{R}^p} |(\mathbf{y}_0 - \boldsymbol{\mu})(\mathbf{y}_0 - \boldsymbol{\mu})^\top + (\boldsymbol{\mu} - \boldsymbol{\mu}_n)(\boldsymbol{\mu} - \boldsymbol{\mu}_n)^\top + \boldsymbol{\Psi}_n|^{-(\alpha_n + 2)/2} d\boldsymbol{\mu}$$

$$= \int_{\mathcal{R}^p} \left| (1 + \beta_n)\left(\boldsymbol{\mu} - \frac{(\mathbf{y}_0 + \beta_n \boldsymbol{\mu}_n)}{1 + \beta_n}\right)\left(\boldsymbol{\mu} - \frac{(\mathbf{y}_0 + \beta_n \boldsymbol{\mu}_n)}{1 + \beta_n}\right)^\top \right.$$

$$\left. + \frac{\beta_n}{1 + \beta_n}(\mathbf{y}_0 - \boldsymbol{\mu}_n)(\mathbf{y}_0 - \boldsymbol{\mu}_n)^\top + \boldsymbol{\Psi}_n \right|^{-(\alpha_n + 2)/2} d\boldsymbol{\mu}$$

$$= \int_{\mathcal{R}^p} \left[\left| \underbrace{\boldsymbol{\Psi}_n + \frac{\beta_n}{1 + \beta_n}(\mathbf{y}_0 - \boldsymbol{\mu}_n)(\mathbf{y}_0 - \boldsymbol{\mu}_n)^\top}_{\boldsymbol{\Lambda}_n} \right| \right.$$

$$\left. \left| 1 + (1 + \beta_n)\left(\boldsymbol{\mu} - \frac{(\mathbf{y}_0 + \beta_n \boldsymbol{\mu}_n)}{1 + \beta_n}\right)^\top \frac{1}{\alpha_n + 2 - p}\left(\frac{\boldsymbol{\Lambda}_n}{\alpha_n + 2 - p}\right)^{-1}\left(\boldsymbol{\mu} - \frac{(\mathbf{y}_0 + \beta_n \boldsymbol{\mu}_n)}{1 + \beta_n}\right) \right| \right]^{-(\alpha_n + 2 - p + p)/2} d\boldsymbol{\mu}$$

$$\propto \left| \boldsymbol{\Psi}_n + \frac{\beta_n}{1 + \beta_n}(\mathbf{y}_0 - \boldsymbol{\mu}_n)(\mathbf{y}_0 - \boldsymbol{\mu}_n)^\top \right|^{-(\alpha_n + 2)/2}$$

$$\times \left| \boldsymbol{\Psi}_n + \frac{\beta_n}{1 + \beta_n}(\mathbf{y}_0 - \boldsymbol{\mu}_n)(\mathbf{y}_0 - \boldsymbol{\mu}_n)^\top \right|^{1/2}$$

$$= \left| \boldsymbol{\Psi}_n + \frac{\beta_n}{1 + \beta_n}(\mathbf{y}_0 - \boldsymbol{\mu}_n)(\mathbf{y}_0 - \boldsymbol{\mu}_n)^\top \right|^{-(\alpha_n + 1)/2}$$

$$\propto \left[1 + (\mathbf{y}_0 - \boldsymbol{\mu}_n)^\top \frac{1}{\alpha_n + 1 - p}\left(\frac{\boldsymbol{\Psi}_n(1 + \beta_n)}{(\alpha_n + 1 - p)\beta_n}\right)^{-1}(\mathbf{y}_0 - \boldsymbol{\mu}_n) \right]^{-(\alpha_n + 1 - p + p)}.$$

The second equality and last line use the Sylvester's determinant theorem, and the second equality uses that there is the kernel of a multivariate t distribution.

Then, we have that the predictive distribution is a multivariate t distribution centered at $\boldsymbol{\mu}_n$, $\alpha_n + 1 - p$ degrees of freedom, and scale matrix $\frac{\boldsymbol{\Psi}_n(1+\beta_n)}{(\alpha_n+1-p)\beta_n}$.

To show the second statement, let's start by the definition of the predictive density to show that $\mathbf{Y}_0|\mathbf{Y} \sim T_{N_0,M}(\alpha_n - M + 1, \mathbf{X}_0\mathbf{B}_n, \mathbf{I}_{N_0} + \mathbf{X}_0\mathbf{V}_n\mathbf{X}_0^\top, \boldsymbol{\Psi}_n)$.

$$
\begin{aligned}
\pi(\mathbf{Y}_0|\mathbf{Y}) \propto &\int_{\mathcal{S}}\int_{\mathcal{B}} \left\{ |\boldsymbol{\Sigma}|^{-N_0/2}\exp\left\{-\frac{1}{2}tr[(\mathbf{Y}_0 - \mathbf{X}_0\mathbf{B})^\top(\mathbf{Y}_0 - \mathbf{X}_0\mathbf{B})\boldsymbol{\Sigma}^{-1}]\right\} \right. \\
&\times |\boldsymbol{\Sigma}|^{-K/2}\exp\left\{-\frac{1}{2}tr[(\mathbf{B} - \mathbf{B}_n)^\top\mathbf{V}_n^{-1}(\mathbf{B} - \mathbf{B}_n)\boldsymbol{\Sigma}^{-1}]\right\} \\
&\times \left. |\boldsymbol{\Sigma}|^{-(\alpha_n+M+1)/2}\exp\left\{-\frac{1}{2}tr[\boldsymbol{\Psi}_n\boldsymbol{\Sigma}^{-1}]\right\}\right\} d\mathbf{B}d\boldsymbol{\Sigma} \\
= &\int_{\mathcal{S}}\int_{\mathcal{B}} \left\{ |\boldsymbol{\Sigma}|^{-(N_0+K+\alpha_n+M+1)/2}\exp\left\{-\frac{1}{2}tr\left[\left((\mathbf{Y}_0 - \mathbf{X}_0\mathbf{B})^\top(\mathbf{Y}_0 - \mathbf{X}_0\mathbf{B})\right.\right.\right.\right. \\
&\left.\left.\left.\left. +(\mathbf{B} - \mathbf{B}_n)^\top\mathbf{V}_n^{-1}(\mathbf{B} - \mathbf{B}_n) + \boldsymbol{\Psi}_n\right)\boldsymbol{\Sigma}^{-1}\right]\right\}\right\} d\mathbf{B}d\boldsymbol{\Sigma}.
\end{aligned}
$$

Setting $\mathbf{M} = (\mathbf{X}_0^\top\mathbf{X}_0 + \mathbf{V}_n^{-1})$, and $\mathbf{B}_* = \mathbf{M}^{-1}(\mathbf{V}_n\mathbf{B}_n + \mathbf{X}_0^\top\mathbf{Y}_0)$, we have that $(\mathbf{B} - \mathbf{B}_*)^\top\mathbf{M}(\mathbf{B} - \mathbf{B}_*) + \mathbf{B}_n^\top\mathbf{V}_n^{-1}\mathbf{B}_n + \mathbf{Y}_0^\top\mathbf{Y}_0 - \mathbf{B}_*^\top\mathbf{M}\mathbf{B}_* = (\mathbf{Y}_0 - \mathbf{X}_0\mathbf{B})^\top(\mathbf{Y}_0 - \mathbf{X}_0\mathbf{B}) + (\mathbf{B} - \mathbf{B}_n)^\top\mathbf{V}_n^{-1}(\mathbf{B} - \mathbf{B}_n)$. Then,

$$
\begin{aligned}
\pi(\mathbf{Y}_0|\mathbf{Y}) \propto &\int_{\mathcal{S}} |\boldsymbol{\Sigma}|^{-(N_0+K+\alpha_n+M+1)/2} \\
&\times \exp\left\{-\frac{1}{2}tr[(\boldsymbol{\Psi}_n + \mathbf{B}_n^\top\mathbf{V}_n^{-1}\mathbf{B}_n + \mathbf{Y}_0^\top\mathbf{Y}_0 - \mathbf{B}_*^\top\mathbf{M}\mathbf{B}_*)\boldsymbol{\Sigma}^{-1}]\right\} \\
&\times \int_{\mathcal{B}}\exp\left\{-\frac{1}{2}tr[(\mathbf{B} - \mathbf{B}_*)^\top\mathbf{M}(\mathbf{B} - \mathbf{B}_*)\boldsymbol{\Sigma}^{-1}]\right\} d\mathbf{B}d\boldsymbol{\Sigma}.
\end{aligned}
$$

The latter is the kernel of a matrix normal distribution, thus

$$
\begin{aligned}
\pi(\mathbf{Y}_0|\mathbf{Y}) \propto &\int_{\mathcal{S}} |\boldsymbol{\Sigma}|^{-(N_0+\alpha_n+M+1)/2} \\
&\times \exp\left\{-\frac{1}{2}tr[(\boldsymbol{\Psi}_n + \mathbf{B}_n^\top\mathbf{V}_n^{-1}\mathbf{B}_n + \mathbf{Y}_0^\top\mathbf{Y}_0 - \mathbf{B}_*^\top\mathbf{M}\mathbf{B}_*)\boldsymbol{\Sigma}^{-1}]\right\} d\boldsymbol{\Sigma}
\end{aligned}
$$

This is the kernel of an inverse-Wishart distribution, then

$$
\pi(\mathbf{Y}_0|\mathbf{Y}) \propto \left|\boldsymbol{\Psi}_n + \mathbf{B}_n^\top\mathbf{V}_n^{-1}\mathbf{B}_n + \mathbf{Y}_0^\top\mathbf{Y}_0 - \mathbf{B}_*^\top\mathbf{M}\mathbf{B}_*\right|^{-(N_0+\alpha_n)/2}.
$$

Setting $\mathbf{C}^{-1} = \mathbf{I}_{N_0} + \mathbf{X}_0 \mathbf{V}_n \mathbf{X}_0^\top$ such that $\mathbf{C} = \mathbf{I}_{N_0} - \mathbf{X}_0 (\mathbf{X}_0^\top \mathbf{X}_0 + \mathbf{V}_n^{-1})^{-1} \mathbf{X}_0^\top$ (see footnote 4 in Chapter 4), then $\mathbf{B}_n^\top \mathbf{V}_n^{-1} \mathbf{B}_n + \mathbf{Y}_0^\top \mathbf{Y}_0 - \mathbf{B}_*^\top \mathbf{M} \mathbf{B}_* = (\mathbf{Y}_0 - \mathbf{X}_0 \mathbf{B}_n)^\top \mathbf{C} (\mathbf{Y}_0 - \mathbf{X}_0 \mathbf{B}_n)$. This is done following exactly same procedure as deducing the predictive distribution in the linear regression model in the book. Thus,

$$\pi(\mathbf{Y}_0|\mathbf{Y}) \propto \left| \boldsymbol{\Psi}_n + (\mathbf{Y}_0 - \mathbf{X}_0 \mathbf{B}_n)^\top \mathbf{C} (\mathbf{Y}_0 - \mathbf{X}_0 \mathbf{B}_n) \right|^{-(N_0 + \alpha_n)/2}$$
$$\propto \left| \mathbf{I}_{N_0} + \mathbf{C} (\mathbf{Y}_0 - \mathbf{X}_0 \mathbf{B}_n) \boldsymbol{\Psi}^{-1} (\mathbf{Y}_0 - \mathbf{X}_0 \mathbf{B}_n)^\top \right|^{-(\alpha_n + 1 - M + N_0 + M - 1)/2}.$$

The second proportionality follows from the Sylvester's theorem. Observe that this is the kernel of a matrix t distribution with $\alpha_n + 1 - M$ degrees of freedom, location $\mathbf{X}_0 \mathbf{B}_n$ and scale matrices $\boldsymbol{\Psi}_n$ and $\mathbf{C}^{-1} = \mathbf{I}_{N_0} + \mathbf{X}_0 \mathbf{V}_n \mathbf{X}_0^\top$.

7. Show that $\delta_n = \delta_0 + (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) + (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^\top ((\mathbf{X}^\top \mathbf{X})^{-1} + \mathbf{B}_0)^{-1} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ in the linear regression model, and that $\boldsymbol{\Psi}_n = \boldsymbol{\Psi}_0 + \mathbf{S} + (\hat{\mathbf{B}} - \mathbf{B}_0)^\top \mathbf{V}_n (\hat{\mathbf{B}} - \mathbf{B}_0)$ in the linear multivariate regression model.

Answer

Taking into account that

$$\delta^* = \delta_0 + \mathbf{y}^\top \mathbf{y} + \boldsymbol{\beta}_0^\top \mathbf{B}_0^{-1} \boldsymbol{\beta}_0 - \boldsymbol{\beta}_n^\top \mathbf{B}_n^{-1} \boldsymbol{\beta}_n$$
$$= \delta_0 + \mathbf{y}^\top \mathbf{y} + \boldsymbol{\beta}_0^\top \mathbf{B}_0^{-1} \boldsymbol{\beta}_0 - (\mathbf{B}_0^{-1} \boldsymbol{\beta}_0 + \mathbf{X}^\top \mathbf{X}\hat{\boldsymbol{\beta}})^\top \mathbf{B}_n (\mathbf{B}_0^{-1} \boldsymbol{\beta}_0 + \mathbf{X}^\top \mathbf{X}\hat{\boldsymbol{\beta}})$$
$$= \delta_0 + \mathbf{y}^\top \mathbf{y} - \hat{\boldsymbol{\beta}}^\top \mathbf{X}^\top \mathbf{X} \mathbf{B}_n \mathbf{X}^\top \mathbf{X}\hat{\boldsymbol{\beta}} - 2\hat{\boldsymbol{\beta}}^\top \mathbf{X}^\top \mathbf{X} \mathbf{B}_n \mathbf{B}_0^{-1} \boldsymbol{\beta}_0 + \boldsymbol{\beta}_0^\top (\mathbf{B}_0^{-1} - \mathbf{B}_0^{-1} \mathbf{B}_n \mathbf{B}_0^{-1}) \boldsymbol{\beta}_0$$
$$- \hat{\boldsymbol{\beta}}^\top \mathbf{X}^\top \mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}}^\top \mathbf{X}^\top \mathbf{X}\hat{\boldsymbol{\beta}}$$
$$= \delta_0 + \mathbf{y}^\top \mathbf{y} - \hat{\boldsymbol{\beta}}^\top \mathbf{X}^\top \mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}}^\top (\mathbf{X}^\top \mathbf{X} - \mathbf{X}^\top \mathbf{X} \mathbf{B}_n \mathbf{X}^\top \mathbf{X})\hat{\boldsymbol{\beta}}$$
$$- 2\hat{\boldsymbol{\beta}}^\top \mathbf{X}^\top \mathbf{X} \mathbf{B}_n \mathbf{B}_0^{-1} \boldsymbol{\beta}_0 + \boldsymbol{\beta}_0^\top (\mathbf{B}_0^{-1} - \mathbf{B}_0^{-1} \mathbf{B}_n \mathbf{B}_0^{-1}) \boldsymbol{\beta}_0.$$

Observe that

$$(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{y}^\top \mathbf{y} - 2\hat{\boldsymbol{\beta}} \mathbf{X}^\top \mathbf{y} + \hat{\boldsymbol{\beta}}^\top \mathbf{X}^\top \mathbf{X}\hat{\boldsymbol{\beta}}$$
$$= \mathbf{y}^\top \mathbf{y} - 2\hat{\boldsymbol{\beta}}^\top \mathbf{X}^\top (\mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\mu}}) + \hat{\boldsymbol{\beta}}^\top \mathbf{X}^\top \mathbf{X}\hat{\boldsymbol{\beta}}$$
$$= \mathbf{y}^\top \mathbf{y} - \hat{\boldsymbol{\beta}}^\top \mathbf{X}^\top \mathbf{X}\hat{\boldsymbol{\beta}},$$

where $\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\mu}}$, and $\mathbf{X}^\top \hat{\boldsymbol{\mu}} = 0$.

The following matrix identities are useful [11]:

$$(\mathbf{D} + \mathbf{E})^{-1} = \mathbf{D}^{-1} - \mathbf{D}^{-1} (\mathbf{D}^{-1} + \mathbf{E}^{-1})^{-1} \mathbf{D}^{-1},$$

and

$$(\mathbf{D} + \mathbf{E})^{-1} = \mathbf{D}^{-1} (\mathbf{E}^{-1} + \mathbf{D}^{-1}) \mathbf{E}^{-1}.$$

Using these identities,

$$
\begin{aligned}
[(\mathbf{X}^{\top}\mathbf{X})^{-1} + \mathbf{B}_0]^{-1} &= \mathbf{X}^{\top}\mathbf{X} - \mathbf{X}^{\top}\mathbf{X}(\mathbf{X}^{\top}\mathbf{X} + \mathbf{B}_0^{-1})^{-1}\mathbf{X}^{\top}\mathbf{X} \\
&= \mathbf{B}_0^{-1} - \mathbf{B}_0^{-1}(\mathbf{X}^{\top}\mathbf{X} + \mathbf{B}_0^{-1})^{-1}\mathbf{B}_0^{-1} \\
&= \mathbf{X}^{\top}\mathbf{X}(\mathbf{X}^{\top}\mathbf{X} + \mathbf{B}_0^{-1})^{-1}\mathbf{B}_0^{-1}.
\end{aligned}
$$

Then,

$$
\begin{aligned}
\delta^* &= \delta_0 + (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^{\top}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) + \hat{\boldsymbol{\beta}}^{\top}[(\mathbf{X}^{\top}\mathbf{X})^{-1} + \mathbf{B}_0]^{-1}\hat{\boldsymbol{\beta}} \\
&\quad - 2\hat{\boldsymbol{\beta}}[(\mathbf{X}^{\top}\mathbf{X})^{-1} + \mathbf{B}_0]^{-1}\boldsymbol{\beta}_0 + \boldsymbol{\beta}_0^{\top}[(\mathbf{X}^{\top}\mathbf{X})^{-1} + \mathbf{B}_0]^{-1}\boldsymbol{\beta}_0 \\
&= \delta_0 + (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^{\top}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\
&\quad + (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^{\top}[(\mathbf{X}^{\top}\mathbf{X})^{-1} + \mathbf{B}_0]^{-1}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0).
\end{aligned}
$$

In a similar way for the second part,

$$
\begin{aligned}
(\mathbf{V}_0 + (\mathbf{X}^{\top}\mathbf{X})^{-1})^{-1} &= \mathbf{V}_0^{-1} - \mathbf{V}_0^{-1}(\mathbf{V}_0^{-1} + \mathbf{X}^{\top}\mathbf{X})^{-1}\mathbf{V}_0^{-1} \\
&= \mathbf{X}^{\top}\mathbf{X} - \mathbf{X}^{\top}\mathbf{X}(\mathbf{V}_0^{-1} + \mathbf{X}^{\top}\mathbf{X})^{-1}\mathbf{X}^{\top}\mathbf{X} \\
&= \mathbf{X}^{\top}\mathbf{X}((\mathbf{X}^{\top}\mathbf{X})^{-1} + \mathbf{V}_0)^{-1}\mathbf{V}_0^{-1},
\end{aligned}
$$

we use these results and some algebra to show that $\mathbf{B}_0^{\top}\mathbf{V}_0^{-1}\mathbf{B}_0 + \widehat{\mathbf{B}}^{\top}\mathbf{X}^{\top}\mathbf{X}\widehat{\mathbf{B}} - \mathbf{B}_n^{\top}\mathbf{V}_n^{-1}\mathbf{B}_n = (\hat{\mathbf{B}} - \mathbf{B}_0)^{\top}\mathbf{V}_n(\hat{\mathbf{B}} - \mathbf{B}_0)$ taking into account that $\mathbf{V}_n = (\mathbf{V}_0^{-1} + \mathbf{X}^{\top}\mathbf{X})^{-1}$ and $\hat{\mathbf{B}} = (\mathbf{X}^{\top}\mathbf{X})^{-1}\mathbf{X}^{\top}\mathbf{Y}$.

8. Show that in the linear regression model $\boldsymbol{\beta}_n^{\top}(\mathbf{B}_n^{-1} - \mathbf{B}_n^{-1}\mathbf{M}^{-1}\mathbf{B}_n^{-1})\boldsymbol{\beta}_n = \boldsymbol{\beta}_{**}^{\top}\mathbf{C}\boldsymbol{\beta}_{**}$ and $\beta_{**} = \mathbf{X}_0\beta_n$.

Answer

Taking into account that $(\mathbf{A} + \mathbf{B})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}(\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1}\mathbf{A}^{-1}$ [11], then we observe that $(\mathbf{B}_n^{-1} - \mathbf{B}_n^{-1}\mathbf{M}^{-1}\mathbf{B}_n^{-1}) = (\mathbf{B}_n + (\mathbf{X}_0^{\top}\mathbf{X}_0)^{-1})^{-1}$, where $(\mathbf{B}_n + (\mathbf{X}_0^{\top}\mathbf{X}_0)^{-1})^{-1} = \mathbf{X}_0^{\top}\mathbf{X}_0 - \mathbf{X}_0^{\top}\mathbf{X}_0(\mathbf{B}_n^{-1} + \mathbf{X}_0^{\top}\mathbf{X}_0)^{-1}\mathbf{X}_0^{\top}\mathbf{X}_0 = \mathbf{X}_0^{\top}\mathbf{X}_0 - \mathbf{X}_0^{\top}\mathbf{X}_0\mathbf{M}^{-1}\mathbf{X}_0^{\top}\mathbf{X}_0$, thus

$$
\begin{aligned}
\boldsymbol{\beta}_n^{\top}(\mathbf{B}_n^{-1} - \mathbf{B}_n^{-1}\mathbf{M}^{-1}\mathbf{B}_n^{-1})\boldsymbol{\beta}_n &= \boldsymbol{\beta}_n^{\top}(\mathbf{X}_0^{\top}\mathbf{X}_0 - \mathbf{X}_0^{\top}\mathbf{X}_0\mathbf{M}^{-1}\mathbf{X}_0^{\top}\mathbf{X}_0)\boldsymbol{\beta}_n \\
&= \boldsymbol{\beta}_n^{\top}\mathbf{X}_0^{\top}(\mathbf{I}_{N_0} - \mathbf{X}_0\mathbf{M}^{-1}\mathbf{X}_0^{\top})\mathbf{X}_0\boldsymbol{\beta}_n \\
&= \boldsymbol{\beta}_n^{\top}\mathbf{X}_0^{\top}\mathbf{C}\mathbf{X}_0\boldsymbol{\beta}_n \\
&= \boldsymbol{\beta}_{**}^{\top}\mathbf{C}\boldsymbol{\beta}_{**}.
\end{aligned}
$$

Let's show that $\boldsymbol{\beta}_{**} = \mathbf{X}_0\boldsymbol{\beta}_n$,

$$\boldsymbol{\beta}_{**} = \mathbf{C}^{-1}\mathbf{X}_0\mathbf{M}^{-1}\mathbf{B}_n^{-1}\boldsymbol{\beta}_n$$
$$= (\mathbf{I}_{N_0} + \mathbf{X}_0\mathbf{B}_n\mathbf{X}_0^\top)\mathbf{X}_0\mathbf{M}^{-1}\mathbf{B}_n^{-1}\boldsymbol{\beta}_n$$
$$= (\mathbf{I}_{N_0} + \mathbf{X}_0\mathbf{B}_n\mathbf{X}_0^\top)\mathbf{X}_0(\mathbf{B}_n - \mathbf{B}_n((\mathbf{X}_0^\top\mathbf{X}_0)^{-1} + \mathbf{B}_n)^{-1}\mathbf{B}_n)\mathbf{B}_n^{-1}\boldsymbol{\beta}_n$$
$$= (\mathbf{I}_{N_0} + \mathbf{X}_0\mathbf{B}_n\mathbf{X}_0^\top)(\mathbf{X}_0\boldsymbol{\beta}_n - \mathbf{X}_0\mathbf{B}_n((\mathbf{X}_0^\top\mathbf{X}_0)^{-1} + \mathbf{B}_n)^{-1}\boldsymbol{\beta}_n)$$
$$= \mathbf{X}_0\boldsymbol{\beta}_n - \mathbf{X}_0\mathbf{B}_n((\mathbf{X}_0^\top\mathbf{X}_0)^{-1} + \mathbf{B}_n)^{-1}\boldsymbol{\beta}_n + \mathbf{X}_0\mathbf{B}_n\mathbf{X}_0^\top\mathbf{X}_0\boldsymbol{\beta}_n$$
$$- \mathbf{X}_0\mathbf{B}_n\mathbf{X}_0^\top\mathbf{X}_0\mathbf{B}_n((\mathbf{X}_0^\top\mathbf{X}_0)^{-1} + \mathbf{B}_n)^{-1}\boldsymbol{\beta}_n$$
$$= \mathbf{X}_0\boldsymbol{\beta}_n - \mathbf{X}_0\mathbf{B}_n[((\mathbf{X}_0^\top\mathbf{X}_0)^{-1} + \mathbf{B}_n)^{-1} - \mathbf{X}_0^\top\mathbf{X}_0 + \mathbf{X}_0^\top\mathbf{X}_0\mathbf{B}_n((\mathbf{X}_0^\top\mathbf{X}_0)^{-1} + \mathbf{B}_n)^{-1}]\boldsymbol{\beta}_n.$$

Using that $(\mathbf{A} + \mathbf{B})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{A} + \mathbf{B})^{-1}$, we observe that the expression in brackets is equal to $\mathbf{0}$, then we have the result.

9. Show that $(\mathbf{Y} - \mathbf{XB})^\top(\mathbf{Y} - \mathbf{XB}) = \mathbf{S} + (\mathbf{B} - \widehat{\mathbf{B}})^\top\mathbf{X}^\top\mathbf{X}(\mathbf{B} - \widehat{\mathbf{B}})$ where $\mathbf{S} = (\mathbf{Y} - \mathbf{X}\widehat{\mathbf{B}})^\top(\mathbf{Y} - \mathbf{X}\widehat{\mathbf{B}})$, $\widehat{\mathbf{B}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}$ in the multivariate regression model.

Answer

$$(\mathbf{Y} - \mathbf{XB})^\top(\mathbf{Y} - \mathbf{XB}) = (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}} + \mathbf{X}\hat{\mathbf{B}} - \mathbf{XB})^\top(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}} + \mathbf{X}\hat{\mathbf{B}} - \mathbf{XB})$$
$$= (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})^\top(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}) + 2(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})^\top(\mathbf{X}\hat{\mathbf{B}} - \mathbf{XB})$$
$$+ (\mathbf{XB} - \mathbf{X}\hat{\mathbf{B}})^\top(\mathbf{XB} - \mathbf{X}\hat{\mathbf{B}})$$
$$= \mathbf{S} + (\mathbf{B} - \hat{\mathbf{B}})^\top\mathbf{X}^\top\mathbf{X}(\mathbf{B} - \hat{\mathbf{B}}),$$

given that $(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})^\top(\mathbf{X}\hat{\mathbf{B}} - \mathbf{XB}) = \hat{\mathbf{U}}^\top\mathbf{X}(\hat{\mathbf{B}} - \mathbf{B})$, using that $\hat{\mathbf{B}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}$ which implies $\mathbf{X}^\top\mathbf{X}\hat{\mathbf{B}} = \mathbf{X}^\top\mathbf{Y} = \mathbf{X}^\top\mathbf{X}\hat{\mathbf{B}} + \mathbf{X}^\top\hat{\mathbf{U}}$, then $\mathbf{X}^\top\hat{\mathbf{U}} = \mathbf{0}$.

10. **What is the probability that the Sun will rise tomorrow?**

This is the most famous Richard Price's example developed in the Appendix of the Bayes' theorem paper [1]. Here, we implicitly use *Laplace's Rule of Succession* to solve this question. In particular, if we were a priori uncertain about the probability the Sun will rise on a specified day, we can assume a prior uniform distribution over (0,1), that is, a beta (1,1) distribution. Then, what is the probability that the Sun will rise tomorrow?

Answer

This exercise is an application of the Bernoulli-beta model. Thus, the likelihood is given by a binomial distribution where the probability of success is θ, $p(\mathbf{y}|\theta) \propto \theta^{\sum_{i=1}^N y_i}(1-\theta)^{N-\sum_{i=1}^N y_i}$. In addition, the prior distribution is beta, that is, $\pi(\theta) \propto \theta^{\alpha_0-1}(1-\theta)^{\beta_0-1}$, where $\alpha_0 = \beta_0 = 1$. Then, the predictive distribution that the sun will rise tomorrow is $p(Y_0 = 1|\mathbf{y}) = \frac{1+S}{2+N}$, where $S = \sum_{i=1}^N y_i$ is the number of successes (the Sun rise). $\frac{1+S}{2+N}$ is known

as the *Laplace's Rule of Succession* that was introduced by Laplace in the 18th century in the course of treating the sunrise problem.

11. Using information from Public Policy Polling in September 27th-28th for the 2016 presidential five-way race in USA, there are 411, 373 and 149 sampled people supporting Hillary Clinton, Donald Trump and other, respectively.

   - Find the posterior probability of the percentage difference of people supporting Hillary versus Trump according to this data using a non-informative prior, that is, $\alpha_0 = [1\ 1\ 1]$ in the multinomial-Dirichlet model. What is the probability of having more supporters of Hillary vs Trump?
   - What is the probability that sampling one hundred independent individuals 44, 40 and 16 support Hillary, Trump and other, respectively?

Answer

R code. Multinomial-Dirichlet model: Polling 2016 USA presidential race

```r
set.seed(010101)
# Multinomial-Dirichlet example:
# Polling 2016 USA presidential race
y <- c(411, 373, 149)
# Clinton, Trump, Other
# Public Policy Polling September 27-28,
# 2016 five-way race
alpha0 <- rep(1, 3)
# Hyperparameters: non-informative distribution
alphan <- alpha0 + y
S <- 100000
# Sample draws of posterior
thetas <- MCMCpack::rdirichlet(S, alphan)
colnames(thetas) <- c("Clinton", "Trump", "Other")
head(thetas)
        Clinton      Trump      Other
[1,]  0.4211346  0.4188607  0.1600046
[2,]  0.4244207  0.4224523  0.1531270
[3,]  0.4349268  0.3843953  0.1806779
[4,]  0.4533499  0.4005530  0.1460972
[5,]  0.4381799  0.3968502  0.1649699
[6,]  0.4436852  0.3971321  0.1591827
dif <- thetas[,1] - thetas[,2]
# Difference of shares Hillary vs Trump
data <- data.frame(dif)
names(data) <- c("Difference")
library(ggplot2)
p <- ggplot(data) +
  geom_histogram(aes(x = Difference), binwidth = 0.01) +
  geom_vline(xintercept=0.0, lwd=1, colour="red") +
  ggtitle("Percentage difference Clinton vs Trump 2016
    presidential race") + xlab("Percentage Difference") +
    ylab("")
difmcmc <- coda::mcmc(dif)
# Declaring a MCMC object
summary(difmcmc)

Iterations = 1:1e+05
Thinning interval = 1
Number of chains = 1
Sample size per chain = 1e+05

1. Empirical mean and standard deviation for each
variable, plus standard error of the mean:

Mean              SD        Naive SE Time-series SE
4.062e-02      2.996e-02      9.474e-05      9.474e-05
2. Quantiles for each variable:

2.5%      25%       50%       75%     97.5%
-0.01817  0.02033  0.04058  0.06089  0.09923
CW <- mean(difmcmc>0)
CW
0.91339
```

FIGURE 3.1
Percentage difference: Hillary Clinton vs Donald Trump, five-way race.

There is a 95% probability that the percentage difference between Hillary and Trump according to this poll is (-1.8%, 9.9%). The probability of Hillary having more supporters is 91.3%

R code. Multinomial-Dirichlet model: Polling 2016 USA presidential race

```
1  # Predictive distribution by simulation
2  y0 <- c(44, 40, 16)
3  Pred <- apply(thetas, 1, function(p) {rmultinom(1, size =
       sum(y0), prob = p)})
4  sum(sapply(1:S, function(s) {sum(Pred[,s] == y0) == 3}))/S
5  0.00825
6  # Predictive distribution by analytical expression
7  PredY0 <- function(y0){
8    n <- sum(y0)
9    Res1 <- sum(sapply(1:length(y), function(l){lgamma(alphan[
       l]+y0[l]) - lgamma(alphan[l])-lfactorial(y0[l])}))
10   Res <- lfactorial(n)+lgamma(sum(alphan))-lgamma(sum(alphan
       )+n) + Res1
11   return(exp(Res))
12 }
13 PredY0(y0)
14 0.00850
```

The probability that from one hundred random selected people 44 support Hillary, 40 support Trump and 16 support other candidate is 0.85%.

12. **Math test example continues**

You have a random sample of math scores of size $N = 50$ from a normal distribution, $Y_i \sim \mathcal{N}(\mu, \sigma)$. The sample mean and variance are equal to 102 and 10, respectively. Using the normal-normal/inverse-gamma model where $\mu_0 = 100$, $\beta_0 = 1$, $\alpha_0 = \delta_0 = 0.001$

- Get 95% confidence and credible intervals for μ.
- What is the posterior probability that $\mu > 103$?

Answer

R code. Math test example continues

```r
1  set.seed(010101)
2  N <- 50
3  # Sample size
4  muhat <- 102
5  # Sample mean
6  sig2hat <- 10
7  # Sample variance
8  # Hyperparameters
9  mu0 <- 100
10 beta0 <- 1
11 delta0 <- 0.001
12 alpha0 <- 0.001
13 S <- 100000
14 # Posterior draws
15 alphan <- alpha0 + N
16 deltan <- sig2hat*(N - 1) + delta0 + beta0*N/(beta0 + N)*(
       muhat - mu0)^2
17 sig2Post <- invgamma::rinvgamma(S, shape = alphan, rate =
       deltan)
18 summary(sig2Post)
19 betan <- beta0 + N
20 mun <- (beta0*mu0 + N*muhat)/betan
21 muPost <- sapply(sig2Post, function(s2){rnorm(1, mun, sd = (
       s2/betan)^0.5)})
22 muPostq <- quantile(muPost, c(0.025, 0.5, 0.975))
23 muPostq
24      2.5%        50%      97.5%
25 101.0929 101.9625 102.8311
26 cutoff <- 103
27 PmuPostcutoff <- mean(muPost > cutoff)
28 PmuPostcutoff
29 0.00994
30 # Using Student's t
31 muPost_t <- ((deltan/(alphan*betan))^0.5)*rt(S, alphan) +
       mun
32 c1 <- rgb(173,216,230,max = 255, alpha = 50, names = "lt.
       blue")
33 c2 <- rgb(255,192,203, max = 255, alpha = 50, names = "lt.
       pink")
34 hist(muPost, main = "Histogram: Posterior mean", xlab = "
       Posterior mean", col = c2)
35 hist(muPost_t, main = "Histogram: Posterior mean", xlab = "
       Posterior mean", add = T, col = c1)
36 muPost_tq <- quantile(muPost_t, c(0.025, 0.5, 0.975))
37 muPost_tq
38 2.5%        50%      97.5%
39 101.0837 101.9608 102.8435
40 PmuPost_tcutoff <- mean(muPost_t > cutoff)
41 PmuPost_tcutoff
42 0.01087
```

We perform our calculations using the posterior conditional distribution, and the posterior marginal distribution. Both procedures give similar results as we can observe from Figure 3.2.

Histogram: Posterior mean

FIGURE 3.2
Histogram using the posterior conditional distribution and the posterior marginal distribution

We have that the 95% credible interval is (101.08, 102.84), and the probability of having a value greater than 103 is 1.09%.

13. **Demand of electricity example continues**

   Set c_0 such that maximizes the marginal likelihood in the specifications with and without electricity price in the example of demand of electricity (empirical Bayes). Then, calculate the Bayes factor, and conclude if there is evidence supporting the inclusion of the price of electricity in the demand equation.

   Answer

R code. Demand of electricity

```r
1  rm(list = ls())
2  set.seed(010101)
3  # Electricity demand
4  DataUt <- read.csv("https://raw.githubusercontent.com/
       besmarter/BSTApp/refs/heads/master/DataApp/Utilities.csv
       ", sep = ",", header = TRUE, quote = "")
5  DataUtEst <- DataUt %>%
6  filter(Electricity != 0)
7  attach(DataUtEst)
8  # Dependent variable: Monthly consumption (kWh) in log
9  Y <- log(Electricity)
10 N <- length(Y)
11 # Regressors quantity including intercept
12 X <- cbind(LnPriceElect, IndSocio1, IndSocio2, Altitude,
       Nrooms, HouseholdMem, Children, Lnincome, 1)
13 # Regressor without price
14 Xnew <- cbind(IndSocio1, IndSocio2, Altitude, Nrooms,
       HouseholdMem, Children, Lnincome, 1)
15 # Log marginal function (multiply by -1 due to minimization)
16 LogMarLikLM <- function(X, c0){
17   k <- dim(X)[2]
18   N <- dim(X)[1]
19   # Hyperparameters
20   B0 <- c0*diag(k)
21   b0 <- rep(0, k)
22   # Posterior parameters
23   bhat <- solve(t(X)%*%X)%*%t(X)%*%Y
24   # Force this matrix to be symmetric
25   Bn <- as.matrix(Matrix::forceSymmetric(solve(solve(B0) + t
       (X)%*%X)))
26   bn <- Bn%*%(solve(B0)%*%b0 + t(X)%*%X%*%bhat)
27   dn <- as.numeric(d0 + t(Y)%*%Y+t(b0)%*%solve(B0)%*%b0-t(bn
       )%*%solve(Bn)%*%bn)
28   an <- a0 + N
29   # Log marginal likelihood
30   logpy <- (N/2)*log(1/pi)+(a0/2)*log(d0)-(an/2)*log(dn) +
       0.5*log(det(Bn)/det(B0)) + lgamma(an/2)-lgamma(a0/2)
31   return(-logpy)
32 }
33 # Hyperparameters
34 d0 <- 0.001/2
35 a0 <- 0.001/2
36 # Empirical Bayes: Obtain c0 maximizing the log marginal
       likelihood
37 c0 <- 1000
38 EB <- optim(c0, fn = LogMarLikLM, method = "Brent", lower =
       0.0001, upper = 10^6, X = X)
39 EBnew <- optim(c0, fn = LogMarLikLM, method = "Brent", lower
        = 0.0001, upper = 10^6, X = Xnew)
40 # Change of order to take into account the -1 in the
       LogMarLikLM function
41 BFEM <- exp(EBnew$value - EB$value)
42 BFEM
43 71897938
```

The Bayes factor based on the empirical Bayes of the model with electricity price versus the model without electricity price is equal to 71897938, this gives very strong evidence to include the price in the specification.

14. **Utility demand**

Use the file *Utilities.csv* to estimate a multivariate linear regression model where $\mathbf{Y}_i = [\log(\text{electricity}_i) \ \log(\text{water}_i) \ \log(\text{gas}_i)]$ as function of $\log(\text{electricity price}_i)$, $\log(\text{water price}_i)$, $\log(\text{gas price}_i)$, IndSocio1_i, IndSocio2_i, Altitude_i, Nrooms_i, HouseholdMem_i, Children_i, and $\log(\text{Income}_i)$, where electricity, water and gas are monthly consumption of electricity (kWh), water (m^3) and gas (m^3), and other definitions are given in the Example of Section 4.3. Omit households that do not consume any of the utilities in this exercise.

Set a non-informative prior framework, $\mathbf{B}_0 = [0]_{11 \times 3}$, $\mathbf{V}_0 = 1000\mathbf{I}_{11}$, $\mathbf{\Psi}_0 = 1000\mathbf{I}_3$ and $\alpha_0 = 3$, where we have $K = 11$ (regressors plus intercept) and $M = 3$ (equations) in this exercise.

(a) Find the posterior mean estimates and the highest posterior density intervals at 95% of $\mathbf{B}$ and $\mathbf{\Sigma}$. Use the marginal distribution and the conditional distribution to obtain the posterior estimates of $\mathbf{B}$, and compare the results.

(b) Find the Bayes factor comparing the baseline model in this exercise with the same specification but using the income in dollars. Now, calculate the Bayes factor using the income in thousand dollars. Is there any difference?

(c) Find the predictive distribution for the monthly demand of electricity, water and gas in the baseline specification of a household located in the lowest socioeconomic condition in a municipality located below 1000 meters above the sea level, 2 rooms, 3 members with children, a monthly income equal to USD 500, an electricity price equal to USD/kWh 0.15, a water price equal to USD/M^3 0.70, and a gas price equal to USD/M^3 0.75.

Answer

We see that the posterior estimates of the location parameters based on the marginal distribution and the conditional distribution are very similar (conditional on $\mathbf{\Sigma}$). This is important as many times there is no analytical solutions in well-known forms of marginal posterior distributions, and consequently, we should get draws of the posterior distributions based on conditional distributions of block of parameters (See Chapter 4).

We find that the Bayes factor of the baseline model (log(Income)) versus the two alternative models using income in dollars and thousand dollars are 108925764 and 0.1089261. The former gives strong evidence in favor of the baseline model, whereas the latter gives positive evidence for the

model using the income in thousand dollars. This result despite that the location coefficients are the same in the two alternative specifications, except for the change in scale of the coefficients associated with income. This example shows that Bayes factors are sensitive to units of measure, and consequently, it is relevant to think carefully about the priors when performing hypothesis testing using a Bayesian framework. Observe that a nice feature in Bayesian inference is that we followed the same conceptual framework (Bayes factor) in the previous exercise and this exercise. In one hand, the previous exercise is an example of nested models, that is, one model is a restricted version of a more general model. On the other hand, this exercise is an example of non-nested models. This is not the case in the Frequentist approach. The statistical framework is not the same when testing nested and non-nested models.

R code. Utilities demand: Multivariate regression, posterior inference

```r
1  rm(list = ls())
2  set.seed(010101)
3  library(dplyr)
4  # Electricity demand
5  DataUt <- read.csv("https://raw.githubusercontent.com/
      besmarter/BSTApp/refs/heads/master/DataApp/Utilities.csv
      ", sep = ",", header = TRUE, quote = "")
6  DataUtEst <- DataUt %>%
7  filter(Electricity != 0 & Water !=0 & Gas != 0)
8  attach(DataUtEst)
9  Y <- cbind(log(Electricity), log(Water), log(Gas))
10 X <- cbind(LnPriceElect, LnPriceWater, LnPriceGas, IndSocio1
      , IndSocio2, Altitude, Nrooms, HouseholdMem, Children,
      Lnincome, 1)
11 M <- dim(Y)[2]
12 K <- dim(X)[2]
13 N <- dim(Y)[1]
14 # Hyperparameters
15 B0 <- matrix(0, K, M)
16 c0 <- 1000
17 V0 <- c0*diag(K)
18 Psi0 <- c0*diag(M)
19 a0 <- M
20 # Posterior parameters
21 Bhat <- solve(t(X)%*%X)%*%t(X)%*%Y
22 S <- t(Y - X%*%Bhat)%*%(Y - X%*%Bhat)
23 Vn <- solve(solve(V0) + t(X)%*%X)
24 Bn <- Vn%*%(solve(V0)%*%B0 + t(X)%*%X%*%Bhat)
25 Psin <- Psi0 + S + t(B0)%*%solve(V0)%*%B0 + t(Bhat)%*%t(X)%*
      %X%*%Bhat - t(Bn)%*%solve(Vn)%*%Bn
26 an <- a0 + N
27 #Posterior draws
28 s <- 10000 #Number of posterior draws
29 SIGs <- replicate(s, LaplacesDemon::rinvwishart(an, Psin))
30 BsCond <- sapply(1:s, function(s) {MixMatrix::rmatrixnorm(n
      = 1, mean=Bn, U = Vn,V = SIGs[,,s])})
31 summary(coda::mcmc(t(BsCond)))
32 Bs <- sapply(1:s, function(s) {MixMatrix::rmatrixt(n = 1,
      mean=Bn, U = Vn,V = Psin, df = an + 1 - M)})
33 summary(coda::mcmc(t(Bs)))
34 SIGMs <- t(sapply(1:s, function(l) {gdata::lowerTriangle(
      SIGs[,,l], diag=TRUE, byrow=FALSE)}))
35 summary(coda::mcmc(SIGMs))
36 hdiBs <- HDInterval::hdi(t(BsCond), credMass = 0.95) #
      Highest posterior density credible interval
37 hdiBs
38 hdiSIG <- HDInterval::hdi(SIGMs, credMass = 0.95) # Highest
      posterior density credible interval
39 hdiSIG
40
```

R code. Utilities demand: Multivariate regression, Bayes factors

```r
1  # Log marginal function (multiply by -1 due to minimization)
2  LogMarLikLM <- function(X, c0){
3    c10 <- c0[1]; c20 <- c0[2]
4    k <- dim(X)[2]
5    N <- dim(X)[1]
6    # Hyperparameters
7    V0 <- c10*diag(K)
8    Psi0 <- c20*diag(M)
9    # Posterior parameters
10   Bhat <- solve(t(X)%*%X)%*%t(X)%*%Y
11   S <- t(Y - X%*%Bhat)%*%(Y - X%*%Bhat)
12   Vn <- solve(solve(V0) + t(X)%*%X)
13   Bn <- Vn%*%(solve(V0)%*%B0 + t(X)%*%X%*%Bhat)
14   Psin <- Psi0 + S + t(B0)%*%solve(V0)%*%B0 + t(Bhat)%*%t(X)
       %*%X%*%Bhat - t(Bn)%*%solve(Vn)%*%Bn
15   # Log marginal likelihood
16   logpy <- (N*M/2)*log(1/pi)+(a0/2)*log(det(Psi0)) - (an/2)*
       log(det(Psin)) + (M/2)*(log(det(Vn)) - log(det(V0))) +
       lgamma(an/2)-lgamma(a0/2)
17   return(-logpy)
18 }
19 c0 <- rep(1000, 2)
20 LogML <- LogMarLikLM(X=X, c0 = c0)
21 # Using income in dollars as regressor
22 Xnew <- cbind(LnPriceElect, LnPriceWater, LnPriceGas,
       IndSocio1, IndSocio2, Altitude, Nrooms, HouseholdMem,
       Children, exp(Lnincome), 1)
23 LogMLnew <- LogMarLikLM(X=Xnew, c0 = c0)
24 # Bayes factor
25 BF12 <- exp(LogMLnew - LogML)
26 BF12
27 # Using income in thousand dollars as regressor
28 XnewT <- cbind(LnPriceElect, LnPriceWater, LnPriceGas,
       IndSocio1, IndSocio2, Altitude, Nrooms, HouseholdMem,
       Children, exp(Lnincome)/1000, 1)
29 LogMLnewT <- LogMarLikLM(X=XnewT, c0 = c0)
30 # Bayes factor
31 BF13 <- exp(LogMLnewT - LogML)
32 BF13
33
```

R code. Utilities demand: Multivariate regression, predictive distribution

```r
# Predictive distribution
Xpred <- c(log(0.15), log(0.70), log(0.75), 1, 0, 0, 2, 3,
    1, log(500), 1)
Mean <- Xpred%*%Bn
Hn <- 1+t(Xpred)%*%Vn%*%Xpred
UtilDemand <- exp(replicate(s, MixMatrix::rmatrixt(n = 1,
    mean=Mean, U = Hn, V = Psin, df = an + 1 - M)))
ElePred <- UtilDemand[1,1,]
WatPred <- UtilDemand[1,2,]
GasPred <- UtilDemand[1,3,]
data <- data.frame(cbind(ElePred, WatPred, GasPred)) #Data
    frame
annotations1 <- data.frame(
x = round(quantile(data$ElePred, c(0.025, 0.5, 0.975)),1),
y = c(600, 1000, 600),
label = c("2.5%:", "50%:", "97.5%:")
)
annotations2 <- data.frame(
x = round(quantile(data$WatPred, c(0.025, 0.5, 0.975)),1),
y = c(600, 1000, 600),
label = c("2.5%:", "50%:", "97.5%:")
)
annotations3 <- data.frame(
x = round(quantile(data$GasPred, c(0.025, 0.5, 0.975)),1),
y = c(600, 1000, 600),
label = c("2.5%:", "50%:", "97.5%:")
)
require(ggplot2) # Cool figures
require(ggpubr) # Multiple figures in one page
require(latex2exp) # LaTeX equations in figures
fig1 <- ggplot(data = data, aes(ElePred)) + geom_histogram(
    bins = 40, color = "#000000", fill = "#0099F8") +  xlab(
    "kWh") + ylab("Frequency") + ggtitle("Electricity") +
    xlim(0, 1050) + geom_text(data = annotations1, aes(x = x
    , y = y, label = paste(label, x)), size = 3, fontface =
    "bold")
fig2 <- ggplot(data = data, aes(WatPred)) + geom_histogram(
    bins = 40, color = "#000000", fill = "#0099F8") +  xlab(
    TeX("$M^3$")) + ylab("Frequency") +  ggtitle("Water") +
    xlim(0, 100) + geom_text(data = annotations2, aes(x = x,
     y = y, label = paste(label, x)), size = 3, fontface = "
    bold")
fig3 <- ggplot(data = data, aes(GasPred)) + geom_histogram(
    bins = 40, color = "#000000", fill = "#0099F8") +  xlab(
    TeX("$M^3$")) + ylab("Frequency") +  ggtitle("Gas") +
    xlim(0, 80) + geom_text(data = annotations3, aes(x = x,
    y = y, label = paste(label, x)), size = 3, fontface = "
    bold")

```

Figures 3.3, 3.4 and 3.5 show the marginal predictive distributions of electricity, water and gas for the reference household. The median predictive values are kWh 168.8, M^3 12.3 and M^3 10.1, respectively. In addition, the 95% credible intervals are (27.7, 1028.9), (1.5, 98.7) and (1.5, 67.5) for electricity, water and gas.

FIGURE 3.3
Histogram using the posterior predictive distribution of electricity demand

FIGURE 3.4
Histogram using the posterior predictive distribution of water demand

FIGURE 3.5
Histogram using the posterior predictive distribution of gas demand

4

Simulation methods

4.1 Solutions of Exercises

1. The inverse transform method ...

R code. Example: Math test

```
1  N <- 50 # Sample size
2  y_bar <- 102 # Sample mean
3  s2 <- 10 # Sample variance
4  alpha <- N - 1
5  serror <- (s2/N)^0.5
6  y.H0 <- c(100, 100.5, 101, 101.5, 102)
7  test <- (y.H0 - y_bar)/serror
8  pval <- 2*pt(test, alpha)
9  pval
10 0.0000459 0.0015431 0.0299338 0.2690040 1
11 # p-values
12 PO01 <- (gamma(N/2)*((N-1)*serror^2)^(-0.5)*(1+test^2/alpha)
      ^(-N/2))/(gamma(1/2)*gamma((N-1)/2))
13 PO01/(1+PO01)
14 0.0001705 0.0050345 0.0725330 0.3210223 0.4702050
15 # Posterior model probability of the null hypothesis.
```

Part II

Regression models: A GUIded tour

5

Univariate models

5.1 Solutions of Exercises

1. Get the posterior conditional distributions of the Gaussian linear model assuming independent priors $\pi(\boldsymbol{\beta}, \sigma^2) = \pi(\boldsymbol{\beta}) \times \pi(\sigma^2)$, where $\boldsymbol{\beta} \sim N(\boldsymbol{\beta}_0, \mathbf{B}_0)$ and $\sigma^2 \sim IG(\alpha_0/2, \delta_0/2)$.

Answer

The joint posterior distribution of the parameters is

$$\pi(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}, \mathbf{X}) \propto p(\mathbf{y} | \boldsymbol{\beta}, \sigma^2, \mathbf{X}) \pi(\boldsymbol{\beta}) \pi(\sigma^2)$$

$$\propto (\sigma^2)^{-\frac{N}{2}} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right\}$$

$$\times \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^\top \mathbf{B}_0^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\right\}$$

$$\times \frac{(\delta_0/2)^{(\alpha_0/2)}}{\Gamma(\alpha_0/2)} \frac{1}{(\sigma^2)^{(\alpha_0/2+1)}} \exp\left\{-\frac{\delta_0}{2\sigma^2}\right\}$$

$$\propto \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^\top \mathbf{B}_0^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\right\}$$

$$\times \frac{1}{(\sigma^2)^{(\alpha_0+N)/2+1}} \exp\left\{-\frac{\delta_0 + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{2\sigma^2}\right\}$$

$$= \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^\top \mathbf{B}_0^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\right\} \times \underbrace{\frac{1}{(\sigma^2)^{(\frac{\alpha_n}{2}+1)}} \exp\left\{-\frac{\delta_n}{2\sigma^2}\right\}}_{1}.$$

Observe that (1) is the kernel of an inverse-gamma density function. Thus, $\sigma^2 | \boldsymbol{\beta}, \mathbf{y}, \mathbf{X} \sim IG(\alpha_n/2, \delta_n/2)$, where $\alpha_n = \alpha_0 + N$ and $\delta_n = \delta_0 + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$.

Let's see the posterior distribution of $\boldsymbol{\beta}$,

$$\pi(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}, \mathbf{X}) \propto p(\mathbf{y}|\boldsymbol{\beta}, \sigma^2, \mathbf{X})\pi(\boldsymbol{\beta})\pi(\sigma^2)$$

$$\propto (\sigma^2)^{-\frac{N}{2}} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right\}$$

$$\times \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^\top \mathbf{B}_0^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\right\}$$

$$\times \frac{(\delta_0/2)^{(\alpha_0/2)}}{\Gamma(\alpha_0/2)} \frac{1}{(\sigma^2)^{(\alpha_0/2+1)}} \exp\left\{-\frac{\delta_0}{2\sigma^2}\right\}$$

$$= (\sigma^2)^{-\frac{N}{2}} \exp\left\{-\frac{\sigma^{-2}}{2}[\mathbf{y}^\top\mathbf{y} - \mathbf{y}^\top X\boldsymbol{\beta} - \boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{y} + \boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta}]\right\}$$

$$\times \exp\left\{-\frac{1}{2}[\boldsymbol{\beta}^\top\mathbf{B}_0^{-1}\boldsymbol{\beta} - \boldsymbol{\beta}^\top\mathbf{B}_0^{-1}\boldsymbol{\beta}_0 - \boldsymbol{\beta}_0^\top\mathbf{B}_0^{-1}\boldsymbol{\beta} + \boldsymbol{\beta}_0^\top\mathbf{B}_0^{-1}\boldsymbol{\beta}_0]\right\}$$

$$\times \frac{(\delta_0/2)^{(\alpha_0/2)}}{\Gamma(\alpha_0/2)} \frac{1}{(\sigma^2)^{(\alpha_0/2+1)}} \exp\left\{-\frac{\delta_0}{2\sigma^2}\right\}$$

$$\propto \exp\left\{-\frac{1}{2}[\boldsymbol{\beta}^\top(\mathbf{B}_0^{-1} + \sigma^{-2}\mathbf{X}^\top\mathbf{X})\boldsymbol{\beta} - 2\boldsymbol{\beta}^\top(\mathbf{B}_0^{-1}\boldsymbol{\beta}_0 + \sigma^{-2}\mathbf{X}^\top\mathbf{X}\hat{\boldsymbol{\beta}})]\right\}$$

$$\times \frac{1}{(\sigma^2)^{(\alpha_0+N)/2+1}} \exp\left\{-\frac{\delta_0 + \mathbf{y}^\top\mathbf{y}}{2\sigma^2}\right\},$$

where $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}$.

Adding and subtracting $\boldsymbol{\beta}_n^\top\mathbf{B}_n^{-1}\boldsymbol{\beta}_n$ where

$$\mathbf{B}_n = (\mathbf{B}_0^{-1} + \sigma^{-2}\mathbf{X}^\top\mathbf{X})^{-1}$$

$$\boldsymbol{\beta}_n = \mathbf{B}_n(\mathbf{B}_0^{-1}\boldsymbol{\beta}_0 + \sigma^{-2}\mathbf{X}^\top\mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{B}_n(\mathbf{B}_0^{-1}\boldsymbol{\beta}_0 + \sigma^{-2}\mathbf{X}^\top\mathbf{y}),$$

and completing the square

$$\pi(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}, \mathbf{X}) \propto \exp\left\{-\frac{1}{2}[\boldsymbol{\beta}^\top(\mathbf{B}_0^{-1} + \sigma^{-2}\mathbf{X}^\top\mathbf{X})\boldsymbol{\beta} - 2\boldsymbol{\beta}^\top\mathbf{B}_n^{-1}\mathbf{B}_n(\mathbf{B}_0^{-1}\boldsymbol{\beta}_0 + \sigma^{-2}\mathbf{X}^\top\mathbf{X}\hat{\boldsymbol{\beta}})\right.$$

$$\left. +\boldsymbol{\beta}_n^\top\mathbf{B}_n^{-1}\boldsymbol{\beta}_n - \boldsymbol{\beta}_n^\top\mathbf{B}_n^{-1}\boldsymbol{\beta}_n]\right\} \times \frac{1}{(\sigma^2)^{(\alpha_0+N)/2+1}} \exp\left\{-\frac{\delta_0 + \mathbf{y}^\top\mathbf{y}}{2\sigma^2}\right\}$$

$$= \exp\left\{-\frac{1}{2}[\boldsymbol{\beta}^\top\mathbf{B}_n^{-1}\boldsymbol{\beta} - 2\boldsymbol{\beta}^\top\mathbf{B}_n^{-1}\boldsymbol{\beta}_n + \boldsymbol{\beta}_n^\top\mathbf{B}_n^{-1}\boldsymbol{\beta}_n]\right\}$$

$$\times \frac{1}{(\sigma^2)^{(\alpha_0+N)/2+1}} \exp\left\{-\frac{\delta_0 + \mathbf{y}^\top\mathbf{y} - \sigma^2\boldsymbol{\beta}_n^\top\mathbf{B}_n^{-1}\boldsymbol{\beta}_n}{2\sigma^2}\right\}$$

$$= \underbrace{\exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_n)^\top\mathbf{B}_n^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}_n)\right\}}_{1}$$

$$\times (\sigma^2)^{-(\frac{\alpha_n}{2}+1)} \exp\left\{-\frac{\delta^*}{2\sigma^2}\right\},$$

where $\delta^* = \delta_0 + \mathbf{y}^\top \mathbf{y} + \sigma^2 \boldsymbol{\beta}_n^\top \mathbf{B}_n^{-1} \boldsymbol{\beta}_n$ does not dependent on $\boldsymbol{\beta}$.

We can see that (1) is the kernel of a multivariate normal distribution with mean equal to $\boldsymbol{\beta}_n$ and covariance matrix $\mathbf{B}_n$, that is, $\boldsymbol{\beta}|\sigma^2, \mathbf{y}, \mathbf{X} \sim N(\boldsymbol{\beta}_n, \mathbf{B}_n)$.

We see that the posterior distributions are from the same family as the prior distributions.

2. Show that the posterior conditional distributions of the Gaussian linear model with heteroskedasticity assuming independent priors $\pi(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\tau}) = \pi(\boldsymbol{\beta}) \times \pi(\sigma^2) \times \prod_{i=1}^{N} \pi(\tau_i)$, where $\boldsymbol{\beta} \sim N(\boldsymbol{\beta}_0, \mathbf{B}_0)$, $\sigma^2 \sim IG(\alpha_0/2, \delta_0/2)$ and $\tau_i \sim G(v/2, v/2)$ are $\boldsymbol{\beta}|\sigma^2, \boldsymbol{\tau}, \mathbf{y}, \mathbf{X} \sim N(\boldsymbol{\beta}_n, \mathbf{B}_n)$, $\sigma^2|\boldsymbol{\beta}, \boldsymbol{\tau}, \mathbf{y}, \mathbf{X} \sim IG(\alpha_n/2, \delta_n/2)$ and $\tau_i|\boldsymbol{\beta}, \sigma^2, \mathbf{y}, \mathbf{X} \sim G(v_{1n}/2, v_{2in}/2)$, where $\boldsymbol{\tau} = [\tau_1, \ldots, \tau_n]^\top$, $\mathbf{B}_n = (\mathbf{B}_0^{-1} + \sigma^{-2} \mathbf{X}^\top \boldsymbol{\Psi} \mathbf{X})^{-1}$, $\boldsymbol{\beta}_n = \mathbf{B}_n(\mathbf{B}_0^{-1} \boldsymbol{\beta}_0 + \sigma^{-2} \mathbf{X}^\top \boldsymbol{\Psi} \mathbf{y})$, $\alpha_n = \alpha_0 + N$, $\delta_n = \delta_0 + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \boldsymbol{\Psi}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$, $v_{1n} = v + 1$, $v_{2in} = v + \sigma^{-2}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2$, and $\boldsymbol{\Psi} = \text{diagonal}\{\tau_i\}$.

Answer

The joint posterior distribution of the parameters is

$$\pi(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\tau}|\mathbf{y}, \mathbf{X}) \propto p(\mathbf{y}|\boldsymbol{\beta}, \sigma^2, \boldsymbol{\tau}, \mathbf{X})\pi(\boldsymbol{\beta})\pi(\sigma^2) \prod_{i=1}^{N} \pi(\tau_i)$$

$$\propto \left(\prod_{i=1}^{N} \tau_i^{1/2} \right) (\sigma^2)^{-\frac{N}{2}} \exp \left\{ -\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \boldsymbol{\Psi}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\}$$

$$\times \exp \left\{ -\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^\top \mathbf{B}_0^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) \right\}$$

$$\times \frac{1}{(\sigma^2)^{(\alpha_0/2+1)}} \exp \left\{ -\frac{\delta_0}{2\sigma^2} \right\} \prod_{i=1}^{N} \tau_i^{v/2-1} \exp \left\{ -v\tau_i/2 \right\}.$$

Thus, the conditional posterior distribution of $\sigma^2|\boldsymbol{\beta}, \boldsymbol{\tau}, \mathbf{y}, \mathbf{X}$ is given by

$$\pi(\sigma^2|\boldsymbol{\beta}, \boldsymbol{\tau}, \mathbf{y}, \mathbf{X}) \propto (\sigma^2)^{-\frac{N+\alpha_0}{2}} \exp \left\{ -\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \boldsymbol{\Psi}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \delta_0 \right\}.$$

This is the kernel of an inverse-gamma distribution with shape parameter $\alpha_n/2$ and rate parameter $\delta_n/2$.

The conditional posterior distribution of $\tau_i|\boldsymbol{\beta}, \sigma^2, \mathbf{y}, \mathbf{X}$ is

$$\pi(\tau_i|\boldsymbol{\beta}, \sigma^2, \mathbf{y}, \mathbf{X}) \propto \tau_i^{(v+1)/2-1} \exp \left\{ -\frac{\tau_i}{2}[\sigma^{-2}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2 + v] \right\}.$$

This is the kernel of a gamma distribution with parameter $(v+1)/2$ and $(\sigma^{-2}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2 + v)/2$.

The conditional posterior distribution of $\boldsymbol{\beta}|\sigma^2, \boldsymbol{\tau}, \mathbf{y}, \mathbf{X}$ is given by

$$\pi(\boldsymbol{\beta}|\sigma^2, \boldsymbol{\tau}, \mathbf{y}, \mathbf{X}) \propto \exp\left\{-\frac{1}{2}[\sigma^{-2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^\top \Psi(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) + (\boldsymbol{\beta}-\boldsymbol{\beta}_0)^\top \mathbf{B_0}^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta_0})]\right\}.$$

Following same steps as in the previous exercise we get $\boldsymbol{\beta}|\sigma^2, \boldsymbol{\tau}, \mathbf{y}, \mathbf{X} \sim N(\boldsymbol{\beta}_n, \mathbf{B}_n)$.

3. **The market value of soccer players in Europe continues**

   Use the setting of the previous exercise to perform inference using a Gibbs sampling algorithm of the the market value of soccer players in Europe setting $v = 5$ and same other hyperparameters as the homoscedastic case. Is there any meaningful difference for the coefficient associated with the national team compared to the application in the homoscedastic case?

   Answer

R. code. The value of soccer players, programming our Gibbs sampler (heteroskedastic case)

```r
rm(list = ls())
set.seed(010101)
######################### Linear regression: Value of
    soccer players ##########################
Data <- read.csv("https://raw.githubusercontent.com/
    besmarter/BSTApp/refs/heads/master/DataApp/1
    ValueFootballPlayers.csv", sep = ",", header = TRUE,
    quote = "")
attach(Data)
y <- log(Value)
# Value: Market value in Euros (2017) of soccer players
# Regressors quantity including intercept
X <- cbind(1, Perf, Age, Age2, NatTeam, Goals, Exp, Exp2)
# Perf: Performance. Perf2: Performance squared. Age: Age;
    Age: Age squared.
# NatTeam: Indicator of national team. Goals: Scored goals.
    Goals2: Scored goals squared
# Exp: Years of experience. Exp2: Years of experience
    squared. Assists: Number of assists
k <- dim(X)[2]
N <- dim(X)[1]
# Hyperparameters
d0 <- 0.001
a0 <- 0.001
b0 <- rep(0, k)
c0 <- 1000
B0 <- c0*diag(k)
B0i <- solve(B0)
v <- 5
# MCMC parameters
mcmc <- 5000
burnin <- 5000
tot <- mcmc + burnin
thin <- 1
# Posterior distributions programming the Gibbs sampling
# Auxiliary parameters
an <- a0 + N
v1n <- v + 1
# Gibbs sampling functions
PostSig2 <- function(Beta, tau){
  dn <- d0 + t(y - X%*%Beta)%*%diag(tau)%*%(y - X%*%Beta)
  sig2 <- invgamma::rinvgamma(1, shape = an/2, rate = dn/2)
  return(sig2)
}
PostBeta <- function(sig2, tau){
  Bn <- solve(B0i + sig2^(-1)*t(X)%*%diag(tau)%*%X)
  bn <- Bn%*%(B0i%*%b0 + sig2^(-1)*t(X)%*%diag(tau)%*%y)
  Beta <- MASS::mvrnorm(1, bn, Bn)
  return(Beta)
}
PostTau <- function(sig2, Beta, i){
  v2n <- v + sig2^(-1)*(y[i]-X[i,]%*%Beta)^2
  taui <- rgamma(1, v1n/2, v2n/2)
  return(taui)
}
```

R. code. The value of soccer players, programming our Gibbs sampler (heteroskedastic case)

```
1  PostBetas <- matrix(0, mcmc+burnin, k)
2  PostSigma2 <- rep(0, mcmc+burnin)
3  Beta <- rep(0, k)
4  tau <- rep(1, N)
5  # create progress bar in case that you want to see
       iterations progress
6  pb <- winProgressBar(title = "progress bar", min = 0, max =
       tot, width = 300)
7  for(s in 1:tot){
8    sig2 <- PostSig2(Beta = Beta, tau = tau)
9    PostSigma2[s] <- sig2
10   Beta <- PostBeta(sig2 = sig2, tau = tau)
11   PostBetas[s,] <- Beta
12   tau <- sapply(1:N, function(i){PostTau(sig2 = sig2, Beta =
       Beta, i)})
13   setWinProgressBar(pb, s, title=paste( round(s/tot*100, 0),
       "% done"))
14 }
15 close(pb)
16 keep <- seq((burnin+1), tot, thin)
17 PosteriorBetas <- PostBetas[keep,]
18 colnames(PosteriorBetas) <- c("Intercept", "Perf", "Age", "
       Age2", "NatTeam", "Goals", "Exp", "Exp2")
19 summary(coda::mcmc(PosteriorBetas))
20 PosteriorSigma2 <- PostSigma2[keep]
21 summary(coda::mcmc(PosteriorSigma2))
22 summary(coda::mcmc(exp(PosteriorBetas[,5])-1))
23 Iterations = 1:5000
24 Thinning interval = 1
25 Number of chains = 1
26 Sample size per chain = 5000
27 1. Empirical mean and standard deviation for each variable,
28 plus standard error of the mean:
29 Mean              SD          Naive SE Time-series SE
30 1.256091       0.254974        0.003606        0.004579
31 2. Quantiles for each variable:
32 2.5%    25%     50%     75%  97.5%
33 0.8078 1.0746 1.2367 1.4189 1.7959
```

We see in this application that the value of a top soccer player in Europe increases 124% $(\exp(0.80) - 1))$ on average when he has played in the national team, the credible interval at 95% is (81%, 180%). These values are not very different from the application assuming homoscedasticity in the book.

4. **Example: Determinants of hospitalization continues**

Program a Gibbs sampling algorithm in the application of determinants of hospitalization.

Answer

R. code. Determinants of hospitalization, programming our Gibbs sampler

```r
1  set.seed(010101)
2  Data <- read.csv("https://raw.githubusercontent.com/
       besmarter/BSTApp/refs/heads/master/DataApp/2HealthMed.
       csv", sep = ",", header = TRUE, quote = "")
3  attach(Data)
4  str(Data)
5  y <- Hosp # Dependent variables
6  X <- cbind(1, SHI, Female, Age, Age2, Est2, Est3, Fair, Good
       , Excellent) # Regressors
7  K <- dim(X)[2]
8  N <- dim(X)[1]
9  # Hyperparameters
10 b0 <- rep(0, K) # Prio mean
11 B0 <- diag(K) # Prior covariance
12 B0i <- solve(B0)
13 mcmc <- 1000; burnin <- 500; thin <- 2; tot <- mcmc + burnin
       ; keep <- seq(burnin, tot, thin)
14 # Posterior distributions programming the Gibbs sampling
15 # Auxiliary parameters
16 XtX <- t(X)%*%X
17 # Gibbs sampling functions
18 PostBeta <- function(Yl){
19   Bn <- solve(B0i + XtX)
20   bn <- Bn%*%(B0i%*%b0 + t(X)%*%Yl)
21   Beta <- MASS::mvrnorm(1, bn, Bn)
22   return(Beta)
23 }
24 PostYl <- function(Beta, i){
25   Ylmean <- X[i,]%*%Beta
26   if(y[i] == 1){
27     Yli <- truncnorm::rtruncnorm(1, a = 0, b = Inf, mean =
       Ylmean, sd = 1)
28   }else{
29     Yli <- truncnorm::rtruncnorm(1, a = -Inf, b = 0, mean =
       Ylmean, sd = 1)
30   }
31   return(Yli)
32 }
33 PostBetas <- matrix(0, mcmc+burnin, K)
34 Beta <- rep(0, K)
35 # create progress bar in case that you want to see
       iterations progress
36 pb <- winProgressBar(title = "progress bar", min = 0, max =
       tot, width = 300)
37 for(s in 1:tot){
38   Yl <- sapply(1:N, function(i){PostYl(Beta = Beta, i)})
39   Beta <- PostBeta(Yl = Yl)
40   PostBetas[s,] <- Beta
41   setWinProgressBar(pb, s, title=paste( round(s/tot*100, 0),
       "% done"))
42 }
43 close(pb)
44 keep <- seq((burnin+1), tot, thin)
45 PosteriorBetas <- PostBetas[keep,]
46 colnames(PosteriorBetas) <- c("Intercept", "SHI", "Female",
       "Age", "Age2", "Est2", "Est3", "Fair", "Good", "
       Excellent")
47 summary(coda::mcmc(PosteriorBetas))
```

5. **Choice of the fishing mode continues**

- Run the Algorithm 3 of the book to show the results of the Geweke [3], Raftery [9] and Heidelberger [4] tests using our GUI.

- Use the command *rmnpGibbs* to do the example of the choice of the fishing mode.

Answer

R. code. Fishing choice mode, results our GUI

```
1  GewekeTestLocationCoef
2  Fraction in 1st window = 0.1
3  Fraction in 2nd window = 0.5
4  cte_1  cte_2  cte_3 NAS_1_1 NAS_1_2 NAS_1_3    AS_1    AS_2
5  -1.821 -0.714  0.792  2.275 -3.944 -2.071  1.627 -2.729
6
7  RafteryTestLocationCoef
8  Quantile (q) = 0.5
9  Accuracy (r) = +/- 0.025
10 Probability (s) = 0.95
11 Burn-in  Total   Lower bound  Dependence
12 (M)      (N)     (Nmin)       factor (I)
13 cte_1    780     365690  1537         238.0
14 cte_2    360     193950  1537         126.0
15 cte_3    660     340120  1537         221.0
16 NAS_1_1 120      70320   1537          45.8
17 NAS_1_2 475      243960  1537         159.0
18 NAS_1_3 440      248930  1537         162.0
19 AS_1     3010    1438135 1537         936.0
20 AS_2     550     297770  1537         194.0
21
22 HeidelTestLocationCoef
23 Stationarity start      p-value
24 test          iteration
25 cte_1   passed        6001      6.54e-01
26 cte_2   failed         NA       3.72e-02
27 cte_3   failed         NA       4.99e-02
28 NAS_1_1 failed         NA       4.77e-07
29 NAS_1_2 failed         NA       1.82e-05
30 NAS_1_3 failed         NA       1.19e-04
31 AS_1    passed       2001       3.71e-01
32 AS_2    passed       8001       4.48e-01
33 Halfwidth Mean     Halfwidth
34 test
35 cte_1   passed     -0.34236 0.017025
36 cte_2   <NA>            NA        NA
37 cte_3   <NA>            NA        NA
38 NAS_1_1 <NA>            NA        NA
39 NAS_1_2 <NA>            NA        NA
40 NAS_1_3 <NA>            NA        NA
41 AS_1    passed     -0.00708 0.000306
42 AS_2    passed      0.27982 0.009994
```

R. code. Fishing choice mode, library bayesm

```r
remove(list = ls()); set.seed(12345)
Data <- read.csv("https://raw.githubusercontent.com/
    besmarter/BSTApp/refs/heads/master/DataApp/3Fishing.csv"
    , sep = ",", header = TRUE, quote = "")
attach(Data); str(Data)
p <- 4; na <- 2; nd <- 1; N <- dim(Data)[1]
Xa <- Data[,2:9]
Xd <- matrix(income, N, 1)
X <- bayesm::createX(p = p, na = na, nd = nd, Xa = Xa, Xd =
    Xd, INT = TRUE, base = p, DIFF = TRUE)
df <- list(y = mode, X = X, p = 4)
# Hyperparameters
k <- dim(X)[2]; b0 <- rep(0, k); c0 <- 1000
B0 <- c0*diag(k); B0i <- solve(B0)
a0 <- p - 1 + 3; Psi0 <- a0*diag(p-1)
Prior <- list(betabar = b0, A = B0i, nu = a0, V = Psi0)
# MCMC parameters
mcmc <- 100000; thin <- 5
Mcmc <- list(R = mcmc, keep = thin)
Results <- bayesm::rmnpGibbs(Data = df, Prior = Prior, Mcmc
    = Mcmc)
betatilde <- Results$betadraw / sqrt(Results$sigmadraw[,1])
attributes(betatilde)$class <- "bayesm.mat"
summary(coda::mcmc(betatilde))
Quantiles for each variable:
            2.5%        25%        50%        75%      97.5%
var1 -6.371e-01 -4.670e-01 -3.834e-01 -3.028e-01 -1.422e-01
var2 -2.149e-01 -6.681e-02 -7.870e-03  4.414e-02  1.493e-01
var3 -8.873e-01 -6.071e-01 -4.982e-01 -4.040e-01 -2.446e-01
var4  1.859e-05  4.577e-05  5.952e-05  7.326e-05  9.861e-05
var5 -2.589e-05 -3.050e-06  8.212e-06  2.089e-05  5.330e-05
var6  4.549e-05  7.451e-05  9.144e-05  1.098e-04  1.555e-04
var7 -9.425e-03 -7.612e-03 -6.586e-03 -5.573e-03 -3.136e-03
var8  1.214e-01  2.068e-01  2.592e-01  3.154e-01  4.477e-01
sigmadraw <-  Results$sigmadraw / Results$sigmadraw[,1]
attributes(sigmadraw)$class = "bayesm.var"
summary(coda::mcmc(sigmadraw))
```

6. **Simulation exercise of the multinomial logit model continues**

   Perform inference in the simulation of the multinomial logit model using the command *rmnlIndepMetrop* from the *bayesm* package of **R** and using our GUI.

   Answer

R. code. Simulation of the multinomial logit model

```r
1  remove(list = ls())
2  set.seed(12345)
3  # Simulation of data
4  N<-1000  # Sample Size
5  B<-c(0.5,0.8,-3)
6  B1<-c(-2.5,-3.5,0)
7  B2<-c(1,1,0)
8  # Alternative specific attributes of choice 1, for instance,
          price, quality and duration of choice 1
9  X1<-matrix(cbind(rnorm(N,0,1),rnorm(N,0,1),rnorm(N,0,1)),N,
          length(B))
10 # Alternative specific attributes of choice 2, for instance,
          price, quality and duration of choice 2
11 X2<-matrix(cbind(rnorm(N,0,1),rnorm(N,0,1),rnorm(N,0,1)),N,
          length(B))
12 # Alternative specific attributes of choice 3, for instance,
          price, quality and duration of choice 3
13 X3<-matrix(cbind(rnorm(N,0,1),rnorm(N,0,1),rnorm(N,0,1)),N,
          length(B))
14 X4<-matrix(rnorm(N,1,1),N,1)
15 V1<-B2[1]+X1%*%B+B1[1]*X4
16 V2<-B2[2]+X2%*%B+B1[2]*X4
17 V3<-B2[3]+X3%*%B+B1[3]*X4
18 suma<-exp(V1)+exp(V2)+exp(V3)
19 p1<-exp(V1)/suma
20 p2<-exp(V2)/suma
21 p3<-exp(V3)/suma
22 p<-cbind(p1,p2,p3)
23 y<- apply(p,1, function(x)sample(1:3, 1, prob = x, replace =
          TRUE))
24 table(y)
25 L <- length(table(y))
26 dat <-data.frame(mode,X1[,1],X2[,1],X3[,1],X1[,2],X2[,2],X3
          [,2],X1[,3],X2[,3],X3[,3],X4)
27 colnames(dat) <- c("mode","V1.1","V1.2","V1.3","V2.1","V2.2"
          ,"V2.3","V3.1","V3.2","V3.3","V4")
28 attach(dat)
29 LongData <- mlogit::mlogit.data(dat, shape = "wide", varying
          =2:10, choice = "mode")
30 Xa <- cbind(LongData$V1, LongData$V2, LongData$V3)
31 Xa <- cbind(X1[,1],X2[,1],X3[,1],X1[,2],X2[,2],X3[,2],X1
          [,3],X2[,3],X3[,3])
32 na <- 3
33 Xd <- X4
34 X <- bayesm::createX(p = L, na = na, nd = 1, Xa = Xa, Xd =
          Xd, base = L)
35 DataMlogit <- list(y=y, X = X, p = L)
36 # MCMC parameters
37 mcmc <- 11000+1
38 thin <- 5
39 df <- 6
40 mcmcpar <- list(R = mcmc, keep = 5, nu = df)
41 PostBeta <- bayesm::rmnlIndepMetrop(Data = DataMlogit, Mcmc
          = mcmcpar)
42 summary(PostBeta[["betadraw"]])
```

7. **Simulation of the ordered probit model**

   Simulate an ordered probit model where the first regressor distributes $N(6,5)$ and the second distributes $G(1,1)$, the location parameter is $\boldsymbol{\beta} = [0.5 \ -0.25 \ 0.5]^\top$, and the cutoffs is the vector $\boldsymbol{\alpha} = [0 \ 1 \ 2.5]^\top$. Program from scratch a Metropolis-within-Gibbs sampling algorithm to perform inference in this simulation.

   Answer

R. code. Simulation of the ordered probit model

```r
1  rm(list = ls()); set.seed(010101); N <- 1000
2  x1 <- rnorm(N, 6, 5); x2 <- rgamma(N, shape = 1, scale = 1)
3  X <- cbind(1, x1, x2)
4  beta <- c(0.5, -0.25, 0.5); cutoffs <- c(0, 1, 2.5)
5  e <- rnorm(N,0,1)
6  y_latent <- X%*%beta + e; y <- rep(0,N)
7  for (i in 1:N) {
8    if (y_latent[i] < cutoffs[1]){
9      y[i] <- 0}else{
10     if (y_latent[i] >= cutoffs[1] & y_latent[i] < cutoffs
       [2]) {
11       y[i] <- 1
12     }else{
13       if (y_latent[i] >= cutoffs[2] & y_latent[i] < cutoffs
       [3]) {
14         y[i] <- 2
15       }else{y[i] <- 3
16       }
17     }
18   }
19 }
20 # Likelihood function
21 LogLikOP <- function(param){
22   beta_g <- param[1:ncol(X)]
23   delta <- param[(ncol(X)+1):(ncol(X) + dplyr::n_distinct(y)
       - 1)]
24   Xbeta <- X%*%beta_g
25   logLik <- 0
26   for (i in 1:length(y)){
27     if (y[i]==0){logLiki <- log(pnorm(-Xbeta[i]))
28     }else if (y[i]==1){
29       logLiki <- log(pnorm(exp(delta[1]) - Xbeta[i]) - pnorm
       (-Xbeta[i]))
30     }else if (y[i]==2){
31       logLiki <- log(pnorm(exp(delta[2]) + exp(delta[1]) -
       Xbeta[i]) - pnorm(exp(delta[1]) - Xbeta[i]))
32     }else {logLiki <- log(1 - pnorm(exp(delta[2]) + exp(
       delta[1]) - Xbeta[i]))
33     }
34     logLik <- logLik + logLiki
35   }
36   return(-logLik)
37 }
38 # ML Estimation
39 param0 <- rep(0, ncol(X) + n_distinct(y)-2)
40 mle <- optim(param0, LogLikOP, hessian = T, method = "BFGS")
41 mle$par
42 exp(mle$par[length(beta)+1])
43 exp(mle$par[length(beta)+1])+exp(mle$par[length(beta)+2])
44 CovarML <- solve(mle$hessian)
```

R. code. Simulation of the ordered probit model

```r
1  # M-H within Gibbs
2  mhop <- function(param0, G){
3    betasamples <- matrix(c(0), nrow = G, ncol = ncol(X))
4    betasamples[1,] <- param0[1:ncol(X)]
5    tau <- matrix(c(0), nrow = G, ncol = dplyr::n_distinct(y)
       - 2)
6    tau[1,] <- param0[(ncol(X)+1):(ncol(X) + dplyr::n_distinct
       (y) - 2)]
7    yl <- rep(0,length(y)); ar <- rep(0,G); B1 <- solve(t(X)%*
       %X+solve(B0))
8    pb <- winProgressBar(title = "progress bar", min = 0, max
       = G, width = 300)
9    for(g in 2:G){
10     bg <- betasamples[g-1,]; tg <- tau[g-1,]
11     #Random walk M-H for delta
12     delta_prime <- tg + mvtnorm::rmvnorm(1, mean = rep(0,2),
        sigma = VarProp)
13     alpha <- min(1,(mvtnorm::dmvnorm(delta_prime, mean = d0,
        sigma = D0)*exp(-LogLikOP(c(bg, delta_prime)) +
       LogLikOP(c(bg, tg))))/mvtnorm::dmvnorm(tg, mean = d0,
       sigma = D0))
14     if(is.nan(alpha) | is.na(alpha)) {
15       alpha <- 0
16     }
17     #Acceptance step
18     u <- runif(1, min = 0, max = 1)
19     if(u<=alpha){tau[g,] <- delta_prime; ar[g] <- 1
20     }else{tau[g,] <- tg
21     }
22     #Generation of latent variables
23     for (i in 1:length(y)){
24       if (y[i]==0) {
25         yl[i] <- EnvStats::rnormTrunc(1, mean = X[i,]%*%bg,
       sd = 1, max = 0)
26       }else if(y[i]==1){
27         yl[i] <- EnvStats::rnormTrunc(1, mean = X[i,]%*%bg,
       sd = 1, min = 0, max = exp(tau[g,1]))
28       }else if(y[i]==2){
29         yl[i] <- EnvStats::rnormTrunc(1, mean = X[i,]%*%bg,
       sd = 1, min = exp(tau[g,1]), max = exp(tau[g,2])+exp(tau
       [g,1]))
30       }else{
31         yl[i] <- EnvStats::rnormTrunc(1, mean = X[i,]%*%bg,
       sd = 1, min = exp(tau[g,2])+exp(tau[g,1]))
32       }
33     }
34     #Gibbs sampling for beta
35     if(sum(is.nan(yl))>0 | sum(is.na(yl))>0 | sum(yl)==Inf){
36       betasamples[g,] <- betasamples[g-1,]
37     }else{
38       b1 <- B1%*%(t(X)%*%yl + solve(B0)%*%b0)
39       betasamples[g,] <- mvrnorm(1, mu = b1, Sigma = B1)
40     }
41     setWinProgressBar(pb, g, title=paste( round(g/G*100, 0),
       "% done"))
42   }
43   close(pb)
44   return(cbind(betasamples, tau, ar))
45 }
```

R. code. Simulation of the ordered probit model

```r
#Hyperparameters
d0 <- rep(0,2)
D0 <- diag(2)*10000
b0 <- rep(0,ncol(X))
B0 <- diag(ncol(X))*10000
#Estimation
param0 <- rep(0, ncol(X) + dplyr::n_distinct(y)-1)
G <- 1000
tun <- 1
VarProp <- tun*solve(solve(CovarML[4:5, 4:5]) + solve(D0))
param_sample <- mhop(param0, G)
#Burn in
B <- round(0.2*G)
param_sample <- param_sample[(B+1):G,]
mcmc0 <- coda::mcmc(param_sample[, 1:(ncol(X) + dplyr::n_
    distinct(y) - 2)])
summary(mcmc0)
Iterations = 1:800
Thinning interval = 1
Number of chains = 1
Sample size per chain = 800
1. Empirical mean and standard deviation for each variable,
plus standard error of the mean:
Mean       SD  Naive SE Time-series SE
0.49120 0.08465 0.0029929       0.007140
-0.24919 0.01222 0.0004319       0.001269
0.49440 0.03942 0.0013937       0.002739
0.06716 0.06419 0.0022695       0.008479
0.41926 0.07414 0.0026212       0.009479
2. Quantiles for each variable:
2.5%       25%       50%      75%    97.5%
0.31947   0.43558   0.49710   0.5479   0.6496
-0.27180  -0.25740  -0.24948  -0.2408  -0.2238
0.42229   0.46706   0.49253   0.5189   0.5762
-0.06338   0.02328   0.06819   0.1104   0.1932
0.25857   0.37195   0.41742   0.4672   0.5558
summary(coda::mcmc(cbind(exp(param_sample[, 4]),exp(param_
    sample[, 4])+exp(param_sample[, 5]))))
Iterations = 1:800
Thinning interval = 1
Number of chains = 1
Sample size per chain = 800
1. Empirical mean and standard deviation for each variable,
plus standard error of the mean:
Mean       SD Naive SE Time-series SE
[1,]  1.072 0.06854 0.002423       0.008999
[2,]  2.597 0.13593 0.004806       0.019700
2. Quantiles for each variable:
2.5%    25%    50%    75% 97.5%
var1 0.9386 1.024 1.071 1.117 1.213
var2 2.3224 2.500 2.603 2.686 2.877
```

All posterior mean estimates are close to the population parameters, and the 95% credible intervals encompass the population parameters. We use the definition of γ in the last line of the code.

8. **Simulation of the negative binomial model continues**

Perform inference in the simulation of the negative binomial model using the *bayesm* package in **R** software.

Answer

R. code. Simulation of the negative binomial model

```r
rm(list = ls())
set.seed(010101)
N <- 2000 # Sample size
x1 <- runif(N); x2 <- rnorm(N)
X <- cbind(1, x1, x2)
k <- dim(X)[2]
B <- rep(1, k)
alpha <- 1.2
gamma <- exp(alpha)
lambda <- exp(X%*%B)
y <- rnbinom(N, mu = lambda, size = gamma)
table(y)
# MCMC parameters
mcmc <- 10000
burnin <- 1000
thin <- 5
iter <- mcmc + burnin
keep <- seq(burnin, iter, thin)
sbeta <- 2.93/sqrt(k); salpha <- 2.93
# Hyperparameters: Priors
B0 <- 1000*diag(k); b0 <- rep(0, k)
alpha0 <- 0.5; delta0 <- 0.1
DataNB <- list(y = y, X = X)
mcmcNB <- list(R = mcmc, keep = thin, s_beta = sbeta, s_
    alpha = salpha)
PriorNB <- list(betabar = b0, A = solve(B0), a = alpha0, b =
     delta0)
ResultBayesm <- bayesm::rnegbinRw(Data = DataNB, Mcmc =
    mcmcNB, Prior = PriorNB)
summary(ResultBayesm$alphadraw)
summary(ResultBayesm$betadraw)
```

9. **The market value of soccer players in Europe continues**

Perform the application of the value of soccer players with left censuring

at one million Euros in our GUI using the Algorithm A7, and the hyper-
parameters of the example.

Answer

These are the results of running the application in our GUI. These are very
similar compared with the results in the book.

R. code. The value of soccer players with left censoring in our GUI

```
1  Summary
2  Iterations = 10001:60000
3  Thinning interval = 1
4  Number of chains = 1
5  Sample size per chain = 50000
6  1. Empirical mean and standard deviation for each variable,
7  plus standard error of the mean:
8               Mean        SD   Naive SE Time-series SE
9  (Intercept)  1.014765 2.626395 1.175e-02      1.659e-02
10 Perf         0.033935 0.004529 2.025e-05      2.245e-05
11 Age          1.027285 0.212413 9.499e-04      1.326e-03
12 Age2        -0.021914 0.003989 1.784e-05      2.528e-05
13 NatTeam      0.847412 0.124884 5.585e-04      6.423e-04
14 Goals        0.010092 0.001644 7.351e-06      7.712e-06
15 Exp          0.175324 0.069653 3.115e-04      3.727e-04
16 Exp2        -0.005696 0.002950 1.319e-05      1.527e-05
17 sigma2       0.983337 0.096194 4.302e-04      6.724e-04
18 2. Quantiles for each variable:
19               2.5%       25%        50%        75%       97.5%
20 (Intercept) -4.212563 -0.732856   1.038281   2.796603   6.051e
       +00
21 Perf         0.025121  0.030873   0.033896   0.036988   4.286e
       -02
22 Age          0.618743  0.883011   1.026760   1.168860   1.450e
       +00
23 Age2        -0.029854 -0.024580  -0.021886  -0.019203  -1.424e
       -02
24 NatTeam      0.605030  0.763154   0.846671   0.930593   1.095e
       +00
25 Goals        0.006893  0.008987   0.010097   0.011201   1.330e
       -02
26 Exp          0.038522  0.128032   0.175261   0.222110   3.118e
       -01
27 Exp2        -0.011484 -0.007671  -0.005685  -0.003706   5.502e
       -05
28 sigma2       0.813118  0.915664   0.977298   1.043661   1.189e
       +00
```

---

10. **The market value of soccer players in Europe continues**

    Program from scratch the Gibbs sampling algorithm in the example of the market value of soccer players at the 0.75 quantile.

    Answer

R. code. The value of soccer players, quantile regression

```r
1  rm(list = ls()); set.seed(010101)
2  Data <- read.csv("https://raw.githubusercontent.com/
       besmarter/BSTApp/refs/heads/master/DataApp/1
       ValueFootballPlayers.csv", sep = ",", header = TRUE,
       quote = "")
3  attach(Data)
4  y <- log(Value); X <- cbind(1, Perf, Age, Age2, NatTeam,
       Goals, Exp, Exp2)
5  RegLS <- lm(y ~ X -1)
6  k <- dim(X)[2]; N <- dim(X)[1]
7  # Hyperparameters
8  b0 <- rep(0, k); c0 <- 1000
9  B0 <- c0*diag(k); B0i <- solve(B0)
10 # MCMC parameters
11 mcmc <- 5000; burnin <- 1000
12 tot <- mcmc + burnin; thin <- 1
13 # Quantile
14 tau <- 0.5; theta <- (1-2*tau)/(tau*(1-tau))
15 psi2 <- 2/(tau*(1-tau)); an2 <- 2+theta^2/psi2
16 # Gibbs sampler
17 PostBeta <- function(e){
18   Bn <- solve(B0i + psi2^(-1)*t(X)%*%diag(1/e)%*%X)
19   bn <- Bn%*%(B0i%*%b0 + psi2^(-1)*t(X)%*%diag(1/e)%*%(y-
       theta*e))
20   Beta <- MASS::mvrnorm(1, bn, Bn)
21   return(Beta)
22 }
23 PostE <- function(Beta, i){
24   dn2 <-(y[i]-X[i,]%*%Beta)^2/psi2
25   ei <- GIGrvg::rgig(1, chi = dn2, psi = an2, lambda = 1/2)
26   return(ei)
27 }
28 PostBetas <- matrix(0, mcmc+burnin, k)
29 Beta <- RegLS$coefficients
30 pb <- winProgressBar(title = "progress bar", min = 0, max =
       tot, width = 300)
31 for(s in 1:tot){
32   e <- sapply(1:N, function(i){PostE(Beta = Beta, i)})
33   Beta <- PostBeta(e = e)
34   PostBetas[s,] <- Beta
35   setWinProgressBar(pb, s, title=paste( round(s/tot*100, 0),
       "% done"))
36 }
37 close(pb)
38 keep <- seq((burnin+1), tot, thin)
39 PosteriorBetas <- PostBetas[keep,]
40 colnames(PosteriorBetas) <- c("Intercept", "Perf", "Age", "
       Age2", "NatTeam", "Goals", "Exp", "Exp2")
41 summary(coda::mcmc(PosteriorBetas))
```

11. Use the *bayesboot* package to perform inference in the simulation exercise of Section 7.10, and compared the results with the ones that we get using our GUI setting $S = 10000$.

Answer

R. code. The value of soccer players, quantile regression

```r
########################### Bayesian bootstrap: Simulation
    ###########################
rm(list = ls())
set.seed(010101)
N <- 1000 # Sample size
x1 <- runif(N); x2 <- rnorm(N)
X <- cbind(x1, x2)
k <- dim(X)[2]
B <- rep(1, k+1)
sig2 <- 1
u <- rnorm(N, 0, sig2)
y <- cbind(1, X)%*%B + u
data <- as.data.frame(cbind(y, X))
names(data) <- c("y", "x1", "x2")
Reg <- function(d){
  Reg <- lm(y ~ x1 + x2, data = d)
  Bhat <- Reg$coef
  return(Bhat)
}
Reg(data)
S <- 10000
BB <- bayesboot::bayesboot(data = data, statistic = Reg, R =
    S)
plot(BB)
```

6

Multivariate models

6.1 Solutions of Exercises

1. Show that $\mathbb{E}[u_1\text{PAER}] = \frac{\alpha_1}{1-\beta_1\alpha_1}\sigma_1^2$ assuming that $\mathbb{E}[u_1 u_2] = 0$ where $Var(u_1) = \sigma_1^2$ in the effect of institutions on per capita GDP.

Answer

The point of departure is the following *simultaneous structural* economic model:

$$\log(\text{pcGDP95}_i) = \beta_1 + \beta_2\text{PAER}_i + \beta_3\text{Africa} + \beta_4\text{Asia} + \beta_5\text{Other} + u_{1i}, \tag{6.1}$$

$$\text{PAER}_i = \alpha_1 + \alpha_2\log(\text{pcGDP95}_i) + \alpha_3\log(\text{Mort}_i) + u_{2i}, \tag{6.2}$$

where *pcGDP95*, *PAER* and *Mort* are the per capita gross domestic product (GDP) in 1995, the average index of protection against expropriation between 1985 and 1995, and the settler mortality rate during the time of colonization. *Africa*, *Asia* and *Other* are dummies for continents, with *America* as the baseline group.

Replacing Equation 6.2 into Equation 6.1, and solving for $\log(pcGDP95)$,

$$\log(\text{pcGDP95}_i) = \pi_1 + \pi_2\log(\text{Mort}_i) + \pi_3\text{Africa} + \pi_4\text{Asia} + \pi_5\text{Other} + e_{1i}, \tag{6.3}$$

where $\pi_1 = \frac{\beta_1+\beta_2\alpha_1}{1-\beta_2\alpha_2}$, $\pi_2 = \frac{\beta_2\alpha_3}{1-\beta_2\alpha_2}$, $\pi_3 = \frac{\beta_3}{1-\beta_2\alpha_2}$, $\pi_4 = \frac{\beta_4}{1-\beta_2\alpha_2}$, $\pi_5 = \frac{\beta_5}{1-\beta_2\alpha_2}$, and $e_1 = \frac{\beta_2 u_2 + u_1}{1-\beta_2\alpha_2}$.

Then, replacing Equation 6.3 into Equation 6.2, and solving for *PAER*,

$$\text{PAER}_i = \gamma_1 + \gamma_2\log(\text{Mort}_i) + \gamma_3\text{Africa} + \gamma_4\text{Asia} + \gamma_5\text{Other} + e_{2i}, \tag{6.4}$$

where $\gamma_1 = \frac{\alpha_1+\alpha_2\beta_1}{1-\beta_2\alpha_2}$, $\gamma_2 = \frac{\alpha_3}{1-\beta_2\alpha_2}$, $\gamma_3 = \frac{\alpha_2\beta_3}{1-\beta_2\alpha_2}$, $\gamma_4 = \frac{\alpha_2\beta_4}{1-\beta_2\alpha_2}$, $\gamma_5 = \frac{\alpha_2\beta_5}{1-\beta_2\alpha_2}$, and $e_2 = \frac{\alpha_2 u_1 + u_2}{1-\beta_2\alpha_2}$.

Then, $\mathbb{E}[u_1\text{PAER}] = \mathbb{E}\left[u_1(\gamma_1 + \gamma_2\log(\text{Mort}_i) + \gamma_3\text{Africa} + \gamma_4\text{Asia} + \gamma_5\text{Other} + e_{2i})\right] = \mathbb{E}\left[u_1\left(\frac{\alpha_2 u_1 + u_2}{1-\beta_2\alpha_2}\right)\right] = \frac{\alpha_2}{1-\beta_2\alpha_2}\sigma_1^2$ given the assumptions.

2. Show that $\beta_2 = \pi_2/\gamma_2$ in the effect of institutions on per capita GDP.

   Answer

   Given that $\pi_2 = \frac{\beta_2\alpha_3}{1-\beta_2\alpha_2}$ and $\gamma_2 = \frac{\alpha_3}{1-\beta_2\alpha_2}$ from the previous exercise, consequently, $\beta_2 = \pi_2/\gamma_2$.

3. **The effect of institutions on per capita gross domestic product continues**

   Use the *rmultireg* command from the *bayesm* package to perform inference in the example of the effect of institutions on per capita GDP.

   Answer

   ### R code. The effect of institutions on per capita GDP

   ```r
   1  rm(list = ls())
   2  set.seed(010101)
   3  DataInst <- read.csv("https://raw.githubusercontent.com/
        besmarter/BSTApp/refs/heads/master/DataApp/4Institutions
        .csv", sep = ",", header = TRUE, quote = "")
   4  attach(DataInst)
   5  Y <- cbind(logpcGDP95, PAER)
   6  X <- cbind(1, logMort, Africa, Asia, Other)
   7  M <- dim(Y)[2]
   8  K <- dim(X)[2]
   9  # Hyperparameters
   10 B0 <- matrix(0, K, M)
   11 c0 <- 100
   12 V0 <- c0*diag(K)
   13 Psi0 <- 5*diag(M)
   14 a0 <- 5
   15 S <- 10000 #Number of posterior draws
   16 betadraw = matrix(double(S*K*M), ncol=K*M)
   17 Sigmadraw = matrix(double(S*M*M), ncol=M*M)
   18 pb <- winProgressBar(title = "progress bar", min = 0, max =
        S, width = 300)
   19 for (s in 1:S) {
   20   Results <- bayesm::rmultireg(Y, X, Bbar = B0, A = solve(V0
        ), nu = a0, V = Psi0)
   21   betadraw[s,] <- Results$B
   22   Sigmadraw[s,] <- Results$Sigma
   23   setWinProgressBar(pb, s, title=paste( round(s/S*100, 0), "
        % done"))
   24 }
   25 close(pb)
   26 summary(coda::mcmc(betadraw))
   27 summary(coda::mcmc(Sigmadraw))
   ```

4. **Demand and supply simulation**

Given the structural demand-supply model:

$$q_i^d = \beta_1 + \beta_2 p_i + \beta_3 y_i + \beta_4 pc_i + \beta_5 ps_i + u_{i1}$$
$$q_i^s = \alpha_1 + \alpha_2 p_i + \alpha_3 er_i + u_{i2},$$

where q^d is demand, q^s is supply, p, y, pc, ps and er are price, income, complementary price, substitute price, and exchange rate. Complementary and substitute prices are prices of a complementary and substitute goods of q. Assume that $\boldsymbol{\beta} = [5 \ -0.5 \ 0.8 \ -0.4 \ 0.7]^\top$, $\boldsymbol{\alpha} = [-2 \ 0.5 \ -0.4]^\top$, $u_1 \sim N(0, 0.5^2)$ and $u_2 \sim N(0, 0.5^2)$. In addition, assume that $y \sim N(10, 1)$, $pc \sim N(5, 1)$, $ps \sim N(5, 1)$ and $tc \sim N(15, 1)$.

- Find the *reduce-form* model using that in equilibrium demand and supply are equal, that is, $q^d = q^s$. This condition defines the observable quantity (q).
- Simulate p and q from the *reduce-form* equations.
- Preform inference of the *reduce-form* model using the *rmultireg* command from the *bayesm* package.
- Use the posterior draws of the *reduce-form* parameters to perform inference of the *structural* parameters. Any issue? Hint: Are all *structural* parameters exactly identified?

Answer

We should equal demand and supply, and solve for price,

$$p = \pi_1 + \pi_2 er + \pi_3 y + \pi_4 pc + \pi_5 ps + v_1,$$

where $\pi_1 = \frac{\alpha_1 - \beta_1}{\beta_2 - \alpha_2}$, $\pi_2 = \frac{\alpha_3}{\beta_2 - \alpha_2}$, $\pi_3 = \frac{-\beta_3}{\beta_2 - \alpha_2}$, $\pi_4 = \frac{-\beta_4}{\beta_2 - \alpha_2}$, $\pi_5 = \frac{-\beta_5}{\beta_2 - \alpha_2}$, and $v_1 = \frac{u_2 - u_1}{\beta_2 - \alpha_2}$ given $\beta_2 \neq \alpha_2$, that is, the equations should be independent. This condition is given by economic theory due to $\beta_2 < 0$ and $\alpha_2 > 0$, the effect of price on demand and supply should be negative and positive, respectively.

The equation of price into the demand equation gives

$$q = \tau_1 + \tau_2 er + \tau_3 y + \tau_4 pc + \tau_5 ps + v_2,$$

where $\tau_1 = \beta_1 + \beta_2 \pi_1$, $\tau_2 = \beta_2 \pi_2$, $\tau_3 = \beta_2 \pi_3 + \beta_3$, $\tau_4 = \beta_2 \pi_4 + \beta_4$, $\tau_5 = \beta_2 \pi_5 + \beta_5$, and $v_2 = \beta_2 v_1 + u_1$. We can use the equations para π_k and τ_k, $k = \{1, 2, 3, 4, 5\}$ to simulate the *reduced-form* equations.

Observe that estimating the *reduced-form* equations, we can get the *structural* parameters for the demand equation, $\beta_2 = \tau_2 / \pi_2$ $(\pi_2 \neq 0)$, $\beta_3 = \tau_3 - \beta_2 \pi_3$, $\beta_4 = \tau_4 - \beta_2 \pi_4$ and $\beta_5 = \tau_5 - \beta_2 \pi_5$, whereas the *structural* parameters of the supply equation cannot be recovered just in one way,

$\alpha_2 = \beta_2 + \beta_3/\pi_3 = \beta_2 + \beta_4/\pi_4 = \beta_2 + \beta_5/\pi_5$. This in turn implies different values for $\alpha_3 = \pi_2(\beta_2 - \alpha_2)$. This is because the demand equation is *exactly identified*, whereas the supply equation is *over identified*.

In this exercise, $K = 5$, $M = 2$, $K_1 = 4$, $K_2 = 2$, $M_1 = 2$ and $M_2 = 2$. This means that $K - K_1 = 1 = M - 1$ and $K - K_2 = 3 > M - 1 = 1$, that is, the order condition says that both equations (demand and supply) satisfy the necessary condition of identification, the demand would be *exactly identified*, and the supply equation would be *over identified*.

Regarding the rank condition (necessary and sufficient), let's see the identification matrix:

TABLE 6.1
Identification matrix.

q	p	constant	er	y	pc	ps
1	$-\beta_2$	$-\beta_1$	0	$-\beta_3$	$-\beta_4$	$-\beta_5$
1	$-\alpha_2$	$-\alpha_1$	$-\alpha_3$	0	0	0

The demand equation excludes the exchange rate (*exclusion restriction*), and given $\alpha_3 \neq 0$, that is, the exchange rate is relevant in the supply equation, then the rank condition is satisfied in the demand equation. The supply equation excludes the income, the complementary price and substitute price (*exclusion restrictions*), then as far as $\beta_k \neq 0$ for any $k = \{3, 4, 5\}$, the rank condition is satisfied in the supply equation.

The following code shows how to do the simulation, and perform inference in this exercise. We can see that all 95% credible intervals encompass the population parameters, and the posterior means are very close to them.

R code. Demand and supply simulation

```r
rm(list = ls()); set.seed(12345)
B0 <- 5; B1 <- -0.5; B2 <- 0.8; B3 <- -0.4; B4 <- 0.7; SD <-
    0.5
A0 <- -2; A1 <- 0.5; A2 <- -0.4; SS <- 0.5
P0 <- (A0-B0)/(B1-A1); P2 <- -B2/(B1-A1); P3 <- -B3/(B1-A1);
    P1 <- A2/(B1-A1); P4 <- -B4/(B1-A1)
T0 <- B0+B1*P0; T2 <- B2+B1*P2; T3 <- B3+B1*P3; T1 <- B1*P1;
    T4 <- B4+B1*P4;
n <- 5000
ED <- rnorm(n, 0, SD); ES <- rnorm(n, 0, SS)
VP <- (ES-ED)/(B1-A1); UQ <- B1*VP+ED
y <- rnorm(n, 10, 1); pc <- rnorm(n, 5, 1); er <- rnorm(n,
    15, 1); ps <- rnorm(n, 5, 1);
p <- P0+P1*er+P2*y+P3*pc+P4*ps+VP
q <- T0+T1*er+T2*y+T3*pc+T4*ps+UQ
#Inference
Y <- cbind(p, q); X <- cbind(1, er, y, pc, ps)
M <- dim(Y)[2]; K <- dim(X)[2]
# Hyperparameters
B0 <- matrix(0, K, M); c0 <- 100; V0 <- c0*diag(K)
Psi0 <- 5*diag(M); a0 <- 5; S <- 10000 #Posterior draws
betadraw = matrix(double(S*K*M), ncol=K*M)
Sigmadraw = matrix(double(S*M*M), ncol=M*M)
pb <- winProgressBar(title = "progress bar", min = 0, max =
    S, width = 300)
for (s in 1:S) {
  Results <- bayesm::rmultireg(Y, X, Bbar = B0, A = solve(V0
    ), nu = a0, V = Psi0)
  betadraw[s,] <- Results$B
  Sigmadraw[s,] <- Results$Sigma
  setWinProgressBar(pb, s, title=paste( round(s/S*100, 0), "
    % done"))
}
close(pb)
summary(coda::mcmc(betadraw))
summary(coda::mcmc(Sigmadraw))
beta2 <- betadraw[,7]/betadraw[,2] # Effect of price on
    demand
summary(coda::mcmc(beta2))
beta3 <- betadraw[,8] - beta2*betadraw[,3] # Effect of
    income on demand
summary(coda::mcmc(beta3))
beta4 <- betadraw[,9] - beta2*betadraw[,4] # Effect of
    complementary price on demand
summary(coda::mcmc(beta4))
beta5 <- betadraw[,10] - beta2*betadraw[,5] # Effect of
    substitute price  on demand
summary(coda::mcmc(beta5))
```

5. **Utility demand continues**

- Run the **Utility demand** application using our GUI and the information in the dataset *Utilities.csv*. Hint: This file should be modified to agree the structure that requires our GUI (see the dataset *5Institutions.csv* in the folder *DataApp* of our GitHub repository - **https://github.com/besmarter/BSTApp**- for a template).

- Program from scratch the Gibbs sampler algorithm in this application.

Answer

R code. Utility demand in Colombia, Gibss sampler

```r
1  rm(list = ls()); set.seed(010101); library(dplyr)
2  DataUt <- read.csv("https://raw.githubusercontent.com/
       besmarter/BSTApp/refs/heads/master/DataApp/Utilities.csv
       ", sep = ",", header = TRUE, quote = "")
3  DataUtEst <- DataUt %>% filter(Electricity != 0 & Water !=0
       & Gas != 0)
4  attach(DataUtEst)
5  y1 <- log(Electricity); y2 <- log(Water); y3 <- log(Gas)
6  X1 <- cbind(1, LnPriceElect, LnPriceWater, LnPriceGas,
       IndSocio1, IndSocio2, Altitude, Nrooms, HouseholdMem,
       Lnincome)
7  X2 <- cbind(1, LnPriceElect, LnPriceWater, LnPriceGas,
       IndSocio1, IndSocio2, Nrooms, HouseholdMem)
8  X3 <- cbind(1, LnPriceElect, LnPriceWater, LnPriceGas,
       IndSocio1, IndSocio2, Altitude, Nrooms, HouseholdMem)
9  y <- c(y1, y2, y3)
10 X <- as.matrix(Matrix::bdiag(X1, X2, X3))
11 M <- 3; K1 <- dim(X1)[2]; K2 <- dim(X2)[2]; K3 <- dim(X3)[2]
12 K <- K1 + K2 + K3; N <- length(y1)
13 # Hyperparameters
14 b0 <- rep(0, K); c0 <- 100; B0 <- c0*diag(K)
15 B0i <- solve(B0); Psi0 <- 5*diag(M); Psi0i <- solve(Psi0)
16 a0 <- M; IN <- diag(N); an <- a0 + N
17 #Posterior draws
18 S <- 6000; burnin <- 1000; thin <- 1
19 tot <- S+burnin
20 # Gibbs functions
21 PostBeta <- function(Sigma){
22   Aux <- solve(Sigma)%x%IN
23   Bn <- solve(B0i + t(X)%*%Aux%*%X)
24   bn <- Bn%*%(B0i%*%b0 + t(X)%*%Aux%*%y)
25   Beta <- MASS::mvrnorm(1, bn, Bn)
26   return(Beta)
27 }
28 PostSigma <- function(Beta){
29   B1 <- Beta[1:K1]; B2 <- Beta[(K1+1):(K1+K2)]; B3 <- Beta[(
       K1+K2+1):(K1+K2+K3)]
30   U1 <- y1 - X1%*%B1; U2 <- y2 - X2%*%B2; U3 <- y3 - X3%*%B3
31   U <- cbind(U1, U2, U3)
32   Psin <- solve(Psi0i + t(U)%*%U)
33   Sigmai <- rWishart::rWishart(1, df = an, Sigma = Psin)
34   Sigma <- solve(Sigmai[,,1])
35   return(Sigma)
36 }
37 PostBetas <- matrix(0, tot, K)
38 PostSigmas <- matrix(0, tot, M*(M+1)/2); Beta <- rep(1, K)
39 pb <- winProgressBar(title = "progress bar", min = 0, max =
       tot, width = 300)
40 for(s in 1:tot){
41   Sigma <- PostSigma(Beta = Beta); PostSigmas[s,] <-
       matrixcalc::vech(Sigma)
42   Beta <- PostBeta(Sigma = Sigma);  PostBetas[s,] <- Beta
43   setWinProgressBar(pb, s, title=paste( round(s/tot*100, 0),
       "% done"))
44 }
45 close(pb)
46 keep <- seq((burnin+1), tot, thin)
47 Bs <- PostBetas[keep,]; summary(coda::mcmc(Bs))
48 Sigmas <- PostSigmas[keep,]; summary(coda::mcmc(Sigmas))
```

R code. Utility demand in Colombia, results

```
1  keep <- seq((burnin+1), tot, thin)
2  Bs <- PostBetas[keep,]; summary(coda::mcmc(Bs))
3  Sigmas <- PostSigmas[keep,]; summary(coda::mcmc(Sigmas))
4  Quantiles for each variable:
5                 2.5%      25%       50%       75%     97.5%
6  Const          0.43304   1.04508   1.337751  1.64728   2.25228
7  LnPriceElect  -2.40343  -2.05533  -1.882159 -1.70463  -1.35979
8  LnPriceWater  -0.44027  -0.38524  -0.356049 -0.32676  -0.27021
9  LnPriceGas    -0.21665  -0.13807  -0.098877 -0.05712   0.01673
10 IndSocio1     -0.87408  -0.78454  -0.736561 -0.68898  -0.59137
11 IndSocio2     -0.24504  -0.18344  -0.150827 -0.11845  -0.06029
12 Altitude      -0.27153  -0.23866  -0.221539 -0.20468  -0.17325
13 Nrooms         0.04598   0.06191   0.069971  0.07837   0.09379
14 HouseholdMem   0.06610   0.07999   0.087320  0.09406   0.10708
15 Lnincome       0.03897   0.05516   0.063130  0.07152   0.08751
16 Const          0.88247   1.73546   2.162290  2.60793   3.48450
17 LnPriceElect  -0.83128  -0.31258  -0.051838  0.20446   0.72087
18 LnPriceWater  -0.49760  -0.40946  -0.363138 -0.31943  -0.23498
19 LnPriceGas     0.05880   0.16932   0.228734  0.28677   0.39226
20 IndSocio1     -0.63928  -0.50417  -0.427978 -0.35529  -0.21024
21 IndSocio2     -0.50363  -0.40976  -0.359576 -0.30929  -0.21097
22 Nrooms         0.05733   0.08126   0.093291  0.10550   0.12935
23 HouseholdMem   0.09984   0.12098   0.132126  0.14265   0.16290
24 Const         -2.26403  -1.57842  -1.201124 -0.84494  -0.15985
25 LnPriceElect  -2.41128  -2.00330  -1.786150 -1.57292  -1.17185
26 LnPriceWater  -0.10791  -0.03978  -0.003399  0.03329   0.10082
27 LnPriceGas    -0.76450  -0.67575  -0.627062 -0.57949  -0.48807
28 IndSocio1     -0.91346  -0.80094  -0.743421 -0.68334  -0.57228
29 IndSocio2     -0.31613  -0.24086  -0.202828 -0.16367  -0.08824
30 Altitude       0.24986   0.28987   0.310784  0.33183   0.37126
31 Nrooms         0.06071   0.07899   0.089606  0.09949   0.11842
32 HouseholdMem   0.14485   0.16148   0.169509  0.17797   0.19467
33 Sigmas <- PostSigmas[keep,]
34 summary(coda::mcmc(Sigmas))
35          2.5%      25%      50%      75%    97.5%
36 var1 0.19664 0.20567 0.21072 0.21593 0.22636
37 var2 0.08201 0.09289 0.09865 0.10447 0.11623
38 var3 0.05222 0.06015 0.06452 0.06903 0.07799
39 var4 0.47390 0.49575 0.50789 0.52046 0.54446
40 var5 0.07355 0.08672 0.09381 0.10061 0.11427
41 var6 0.29259 0.30618 0.31384 0.32140 0.33700
```

6. **Simulation exercise of instrumental variables continues**

   (a) Use the setting of the simulation exercise of instrumental variables
       to analyze what happens when the instrument is weak, for instance,

setting $\gamma_2 = 0.2$, and compare the performance of the posterior means of the ordinary and instrumental models.

(b) Perform a simulation that helps to analyze how the degree of exogeneity of the instrument affects the performance of the posterior mean of the instrumental variable model.
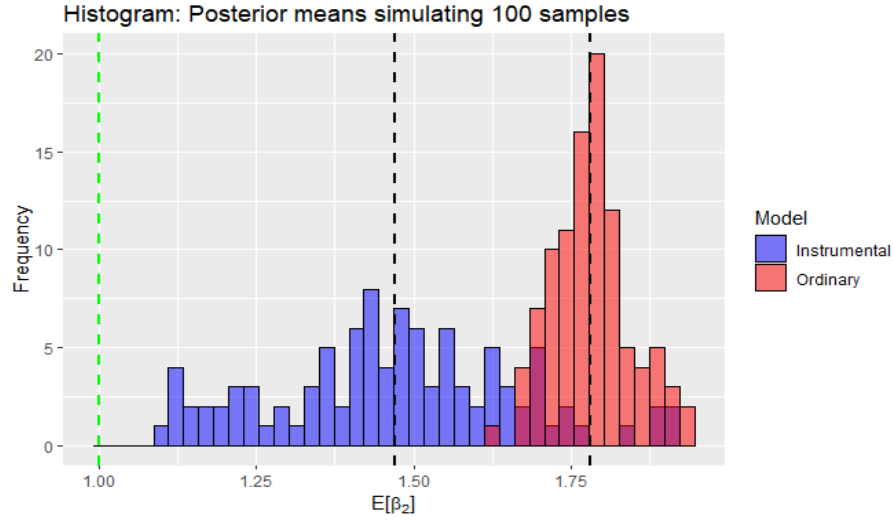
Answer

R code. Simulation exercise, sampling properties of instrumental variable

```r
1  rm(list = ls())
2  set.seed(010101)
3  N <- 100
4  k <- 2
5  B <- rep(1, k)
6  G <- c(1, 0.2)
7  s12 <- 0.8
8  SIGMA <- matrix(c(1, s12, s12, 1), 2, 2)
9  z <- rnorm(N); # w <- rnorm(N)
10 Z <- cbind(1, z); w <- matrix(1,N,1)
11 S <- 100
12 U <- replicate(S, MASS::mvrnorm(n = N, mu = rep(0, 2), SIGMA
     ))
13 x <- G[1] + G[2]*z + U[,2,]
14 y <- B[1] + B[2]*x + U[,1,]
15 VarX <- G[2]^2+1 # Population variance of x
16 EU1U2 <- s12 # Covariance U1
17 BiasPopB2 <- EU1U2/VarX
18 # Hyperparameters
19 d0 <- 0.001/2
20 a0 <- 0.001/2
21 b0 <- rep(0, k)
22 c0 <- 1000
23 B0 <- c0*diag(k)
24 B0i <- solve(B0)
25 g0 <- rep(0, 2)
26 G0 <- 1000*diag(2)
27 G0i <- solve(G0)
28 nu <- 3
29 Psi0 <- nu*diag(2)
30 # MCMC parameters
31 mcmc <- 5000
32 burnin <- 1000
33 tot <- mcmc + burnin
34 thin <- 1
35 # Gibbs sampling
36 Gibbs <- function(x, y){
37   Data <- list(y = y, x = x, w = w, z = Z)
38   Mcmc <- list(R = mcmc, keep = thin, nprint = 0)
39   Prior <- list(md = g0, Ad = G0i, mbg = b0, Abg = B0i, nu =
        nu, V = Psi0)
40   RestIV <- bayesm::rivGibbs(Data = Data, Mcmc = Mcmc, Prior
        = Prior)
41   PostBIV <- mean(RestIV[["betadraw"]])
42   ResLM <- MCMCpack::MCMCregress(y ~ x + w - 1, b0 = b0, B0
        = B0i, c0 = a0, d0 = d0)
43   PostB <- mean(ResLM[,1])
44   Res <- c(PostB,PostBIV)
45   return(Res)
46 }
47 PosteriorMeans <- sapply(1:S, function(s) {Gibbs(x = x[,s],
     y = y[,s])})
48 rowMeans(PosteriorMeans)
49 1.781513 1.470648
```

FIGURE 6.1
Histogram of posterior means to assess the effects of *weakness* of the instrument: Ordinary and instrumental models.

Figure 6.1 shows that having a *weak instrument* generates poor performance of the posterior means of the variable instrumental model. Observe that the mean of the posterior means is less precise and biased. This is due to an identification problem in the likelihood function of y and x_s given z (see [10, Chap. 7]).

R code. Simulation exercise, sampling properties of instrumental variable

```r
1  # Endogenous instruments
2  rm(list = ls())
3  set.seed(010101)
4  N <- 100
5  k <- 2
6  B <- rep(1, k)
7  G <- c(1, 1)
8  delta <- 0.2 # Effect of instrument on y
9  s12 <- 0.8
10 SIGMA <- matrix(c(1, s12, s12, 1), 2, 2)
11 z <- rnorm(N); # w <- rnorm(N)
12 Z <- cbind(1, z); w <- matrix(1,N,1)
13 S <- 100
14 U <- replicate(S, MASS::mvrnorm(n = N, mu = rep(0, 2), SIGMA
       ))
15 x <- G[1] + G[2]*z + U[,2,]
16 y <- B[1] + B[2]*x + delta*z + U[,1,]
17 VarX <- G[2]^2+1 # Population variance of x
18 EU1U2 <- s12 # Covariance U1
19 BiasPopB2 <- EU1U2/VarX
20 # Hyperparameters
21 d0 <- 0.001/2
22 a0 <- 0.001/2
23 b0 <- rep(0, k)
24 c0 <- 1000
25 B0 <- c0*diag(k)
26 B0i <- solve(B0)
27 g0 <- rep(0, 2)
28 G0 <- 1000*diag(2)
29 G0i <- solve(G0)
30 nu <- 3
31 Psi0 <- nu*diag(2)
32 # MCMC parameters
33 mcmc <- 5000
34 burnin <- 1000
35 tot <- mcmc + burnin
36 thin <- 1
37 # Gibbs sampling
38 Gibbs <- function(x, y){
39   Data <- list(y = y, x = x, w = w, z = Z)
40   Mcmc <- list(R = mcmc, keep = thin, nprint = 0)
41   Prior <- list(md = g0, Ad = G0i, mbg = b0, Abg = B0i, nu =
        nu, V = Psi0)
42   RestIV <- bayesm::rivGibbs(Data = Data, Mcmc = Mcmc, Prior
        = Prior)
43   PostBIV <- mean(RestIV[["betadraw"]])
44   ResLM <- MCMCpack::MCMCregress(y ~ x + w - 1, b0 = b0, B0
        = B0i, c0 = a0, d0 = d0)
45   PostB <- mean(ResLM[,1])
46   Res <- c(PostB,PostBIV)
47   return(Res)
48 }
49 PosteriorMeans <- sapply(1:S, function(s) {Gibbs(x = x[,s],
        y = y[,s])})
50 rowMeans(PosteriorMeans)
51 1.508009 1.217419
```
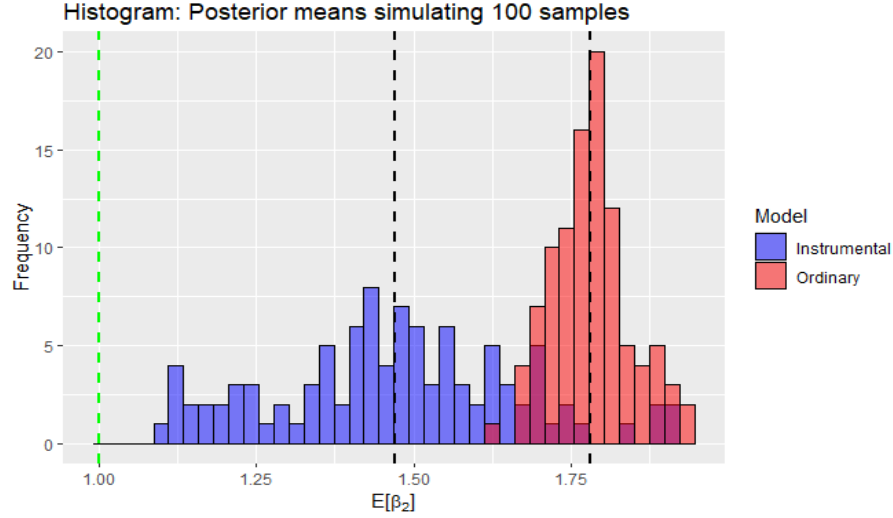
FIGURE 6.2

Histogram of posterior means to assess the effects of *endogeneity* of the instrument: Ordinary and instrumental models.

A simple way to assess the effects of *endogeneity* of the instrument in the instrumental variable model is to introduce the instrument into the *structural equation*, but assume that $\mathbb{E}[yz|x_s] = 0$. In particular, we generate $y_i = \beta_1 + \beta_2 x_{si} + \beta_3 z_i + \mu_i$, but estimate $y_i = \beta_1 + \beta_2 x_{si} + \mu_i$. Figure 6.2 shows that assuming that an instrument is exogenous, when actually it is not, generates bias in the posterior mean of the instrumental variable model. Observe that *exogeneity* of the instrument is equivalent to the dogmatic prior belief that the instrument does not affect the structural equation. Thus, [2] propose to assess the plausibility of the *exclusion restrictions* ($\beta_3 = 0$ and $\gamma_2 \neq 0$) defining *plausible exogeneity* as having prior information that the effect of the instrument in the *structural equation* is near zero, but perhaps not exactly zero.

7. **Simulation exercise of instrumental variables continues II**

   Program from scratch the Gibbs sampling algorithm of the instrumental model for the simulation exercise of the instrumental variables.

   Answer

   The following code shows how to implement the Gibbs sampling algorithm from scratch. We observe that all posterior means are close to the population values, and 95% credible intervals encompass them.

R code. Simulation exercise, Gibss sampling instrumental model

```r
1  rm(list = ls()); set.seed(010101)
2  N <- 100; k <- 2; kz <- 2
3  B <- rep(1, k); G <- rep(1, kz)
4  s12 <- 0.8; SIGMA <- matrix(c(1, s12, s12, 1), 2, 2)
5  z <- rnorm(N); Z <- cbind(1, z); w <- matrix(1,N,1)
6  S <- 100; U <- MASS::mvrnorm(n = N, mu = rep(0, 2), SIGMA)
7  x <- G[1] + G[2]*z + U[,2]; y <- B[1] + B[2]*x + U[,1]
8  X <- cbind(w, x)
9  # Hyperparameters
10 b0 <- rep(0, k); c0 <- 1000; B0 <- c0*diag(k)
11 B0i <- solve(B0); g0 <- rep(0, kz)
12 G0 <- 1000*diag(kz); G0i <- solve(G0)
13 nu <- 3; Psi0 <- nu*diag(2); Psi0i <- solve(Psi0)
14 # MCMC parameters
15 mcmc <- 5000; burnin <- 1000; tot <- mcmc + burnin
16 thin <- 1
17 # Auxiliary elements
18 XtX <- t(X)%*%X; ZtZ <- t(Z)%*%Z; nun <- nu + N
19 # Gibbs sampling
20 PostBeta <- function(Sigma, Gamma){
21   w1 <- Sigma[1,1] - Sigma[1,2]^2/Sigma[2,2]
22   Bn <- solve(w1^(-1)*XtX + B0i)
23   yaux <- y - (Sigma[1,2]/Sigma[2,2])*(x - Z%*%Gamma)
24   bn <- Bn%*%(B0i%*%b0 + w1^(-1)*t(X)%*%yaux)
25   Beta <- MASS::mvrnorm(1, bn, Bn)
26   return(Beta)
27 }
28 PostGamma <- function(Sigma, Beta){
29   w2 <- Sigma[2,2] - Sigma[1,2]^2/Sigma[1,1]
30   Gn <- solve(w2^(-1)*ZtZ + G0i)
31   xaux <- x - (Sigma[1,2]/Sigma[1,1])*(y - X%*%Beta)
32   gn <- Gn%*%(G0i%*%g0 + w2^(-1)*t(Z)%*%xaux)
33   Gamma <- MASS::mvrnorm(1, gn, Gn)
34   return(Gamma)
35 }
36 PostSigma <- function(Beta, Gamma){
37   Uy <- y - X%*%Beta; Ux <- x - Z%*%Gamma
38   U <- cbind(Uy, Ux)
39   Psin <- solve(Psi0i + t(U)%*%U)
40   Sigmai <- rWishart::rWishart(1, df = nun, Sigma = Psin)
41   Sigma <- solve(Sigmai[,,1])
42   return(Sigma)
43 }
```

R code. Simulation exercise, Gibss sampling instrumental model

```r
1  PostBetas <- matrix(0, tot, k); PostGammas <- matrix(0, tot,
       kz)
2  PostSigmas <- matrix(0, tot, 2*(2+1)/2); Beta <- rep(0, k);
       Gamma <- rep(0, kz)
3  pb <- winProgressBar(title = "progress bar", min = 0, max =
       tot, width = 300)
4  for(s in 1:tot){
5    Sigma <- PostSigma(Beta = Beta, Gamma = Gamma)
6    Beta <- PostBeta(Sigma = Sigma, Gamma = Gamma)
7    Gamma <- PostGamma(Sigma = Sigma, Beta = Beta)
8    PostBetas[s,] <- Beta;  PostGammas[s,] <- Gamma;
       PostSigmas[s,] <- matrixcalc::vech(Sigma)
9    setWinProgressBar(pb, s, title=paste( round(s/tot*100, 0),
       "% done"))
10 }
11 close(pb)
12 keep <- seq((burnin+1), tot, thin)
13 Bs <- PostBetas[keep,]
14 Gs <- PostGammas[keep,]
15 Sigmas <- PostSigmas[keep,]
16 summary(coda::mcmc(Bs))
17 1. Empirical mean and standard deviation for each variable,
18 plus standard error of the mean:
19 Mean      SD Naive SE Time-series SE
20 [1,] 1.1519 0.1683  0.00238       0.006321
21 [2,] 0.9896 0.1146  0.00162       0.005177
22 summary(coda::mcmc(Gs))
23 1. Empirical mean and standard deviation for each variable,
24 plus standard error of the mean:
25 Mean      SD Naive SE Time-series SE
26 [1,] 1.133 0.1068 0.001511      0.003096
27 [2,] 1.030 0.1066 0.001507      0.004193
28 summary(coda::mcmc(Sigmas))
29 1. Empirical mean and standard deviation for each variable,
30 plus standard error of the mean:
31 Mean      SD Naive SE Time-series SE
32 [1,] 1.3883 0.3149 0.004454      0.012707
33 [2,] 0.9992 0.2133 0.003017      0.007166
34 [3,] 1.1637 0.1685 0.002382      0.002481
```

8. **The effect of institutions on per capita gross domestic product continues II**

Estimate the structural Equation 6.1 using the instrumental variable model where the instrument of PAER is log(*Mort*). Compare the effect of prop-

erty rights on per capita GDP of this model with the effect estimated in the example of the effect of institutions on per capita gross domestic product. Use the file *6Institutions.csv* to do this exercise in our GUI, and set $\boldsymbol{B}_0 = 100\boldsymbol{I}_5$, $\boldsymbol{\beta}_0 = \boldsymbol{0}_5$, $\boldsymbol{\gamma}_0 = \boldsymbol{0}_2$, $\boldsymbol{G}_0 = 100\boldsymbol{I}_2$ $\alpha_0 = 3$ and $\boldsymbol{\Psi}_0 = 3\boldsymbol{I}_2$. The MCMC iterations, burn-in and thinning parameters are 50000, 1000 and 5, respectively.

Answer

The following code shows how to do this using the command *rivGibss* from the *bayesm* package. We see that the posterior mean of the effect of property rights on per capita GDP is 0.72, and the 95% credible interval is (0.47, 1.00). The posterior mean estimate of this effect was 0.98 in the example, and the 95% credible interval was (0.56, 2.87) (see Section 8.1 in the book). We can observe that the indirect approach, where the *reduced-form* parameters are estimated, and then, plug into the equations of the *structural parameters* produce less efficient posterior estimations.

R code. The effects of institutions on GDP, variable instrumental model

```r
rm(list = ls()); set.seed(010101)
DataInst <- read.csv("https://raw.githubusercontent.com/
    besmarter/BSTApp/refs/heads/master/DataApp/6Institutions
    .csv", sep = ",", header = TRUE, quote = "")
attach(DataInst)
y <- logpcGDP95; x <- PAER
w <- cbind(1, Africa, Asia, Other); Z <- cbind(1, logMort)
# Hyperparameters
k <- 5; kz <- 2; b0 <- rep(0, k); c0 <- 100
B0 <- c0*diag(k); B0i <- solve(B0); g0 <- rep(0, kz)
G0 <- 100*diag(kz); G0i <- solve(G0); nu <- 5
Psi0 <- nu*diag(2)
# MCMC parameters
mcmc <- 50000; burnin <- 10000; tot <- mcmc + burnin; thin
    <- 5
# Gibbs sampling
Data <- list(y = y, x = x, w = w, z = Z)
Mcmc <- list(R = mcmc, keep = thin, nprint = 100)
Prior <- list(md = g0, Ad = G0i, mbg = b0, Abg = B0i, nu =
    nu, V = Psi0)
RestIV <- bayesm::rivGibbs(Data = Data, Mcmc = Mcmc, Prior =
     Prior)
summary(RestIV[["deltadraw"]])
summary(coda::mcmc(RestIV[["betadraw"]]))
Iterations = 1:10000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 10000
1. Empirical mean and standard deviation for each variable,
plus standard error of the mean:
Mean            SD        Naive SE Time-series SE
0.759497        0.140874       0.001409        0.002477
2. Quantiles for each variable:
2.5%    25%    50%    75%  97.5%
0.5110 0.6654 0.7502 0.8424 1.0707
summary(RestIV[["gammadraw"]])
summary(RestIV[["Sigmadraw"]])
```

9. **Multivariate probit with different regressors**

   Let's do a simulation exercise where $y_{i1}^* = 0.5 - 1.2x_{i11} + 0.7x_{i12} + 0.8x_{i3} + \mu_{i1}$ and $y_{i2}^* = 1.5 - 0.8x_{i21} + 0.5x_{i22} + \mu_{i2}$, $\boldsymbol{\Sigma} = \begin{Bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{Bmatrix}$, where all regressors distribute standard normal, and $N = 5000$. Use $\boldsymbol{\beta}_0 = \mathbf{0}$, $\boldsymbol{B}_0 = 1000\boldsymbol{B}$,

$\alpha_0 = 4$ and $\boldsymbol{\Psi}_0 = 4\boldsymbol{I}_2$. Set number of iterations 2000 and a thinning parameter equal to 5.

- Perform inference using the setting of Section **??**, that is, assuming that x_{i3} could have an effect on y_{i2}.

- Program a Gibss sampling algorithm taking into account that there are different regressors in each binary decision, that is, x_{i3} does not have an effect on y_{i2}.

Answer

The following code shows how to perform inference assuming same number of regressors in both binary results. We can check that all 95% credible intervals encompass the population values, and the posterior mean estimates are very close to the population parameters.

R code. Multivariate probit model: Same number of regressors in each binary response

```r
remove(list = ls())
n <-5000   #Individuals
p <-2  #Number dependent variables (For instance to buy or
     not to buy different products)
xo <- 2  #Number of choice dependent variables (For instance
      price of products)
xi <- 1  #Number of regressors that dependt on individuals (
     For instance income)
B1 <- c(0.5,-1.2,0.7,0.8)
B2 <- c(1.5,-0.8,0.5,0) # The last regressor is not relevant
      for the second product
Cor <- matrix(c(1,0.5,0.5,1),p,p)
yX <- NULL; yl <- NULL
for (i in 1:n){
  ei <- MASS::mvrnorm(1, c(0,0), Cor);  xs <- rnorm(xi)
  x1i <- c(1, rnorm(xo), xs);   yl1i <- sum(x1i*B1)
  y1i <- yl1i + ei[1] > 0;  x2i <- c(1, rnorm(xo), xs)
  yl2i <- sum(x2i*B2);  y2i <- yl2i + ei[2] > 0
  yXi <- rbind(c(y1i, x1i), c(y2i, x2i))
  yXi <- cbind(i, yXi)
  colnames(yXi) <- c("id", "y", "cte", "x1o", "x2o", "xi")
  yX <- rbind(yX, yXi)
  yl <- c(yl, c(yl1i + ei[1], yl2i + ei[2]))
}

nd <- 4 # Number of regressors
Xd <- as.matrix(yX[,3:6])
XcreateMP<-function(p,nxs,nind,Data){
  pandterm = function(message) {
    stop(message, call. = FALSE)
  }
  if (missing(nxs))
  pandterm("requires number of regressors: include intercept
       if required")
  if (missing(nind))
  pandterm("requires number of units (individuals)")
  if (missing(Data))
  pandterm("requires dataset")
  if (nrow(Data)!=nind*2)
  pandterm("check dataset! number of units times number
     alternatives should be equal to dataset rows")
  XXDat<-array(0,c(p,1+nxs,nind))
  XX<-array(0,c(p,nxs*p,nind))
  YY<-array(0,c(p,1,nind))
  is<- seq(p,nind*p,p)
  cis<- seq(nxs,nxs*p+1,nxs)
  for(i in is){
    j<-which(i==is)
    XXDat[,,j]<-as.matrix(Data[c((i-(p-1)):i),-1])
    YY[,,j]<-XXDat[,1,j]
    for(l in 1:p){
      XX[l,((cis[l]-(nxs-1)):cis[l]),j]<-XXDat[l,-1,j]
    }
  }
  return(list(y=YY,X=XX))
}
Dat <- XcreateMP(p = p, nxs = nd, nind = n, Data = yX)
```

R code. Multivariate probit model: Same number of regressors in each binary response

```
 1  y<-NULL
 2  X<-NULL
 3  for(i in 1:dim(Dat$y)[3]){
 4    y<-c(y,Dat$y[,,i])
 5    X<-rbind(X,Dat$X[,,i])
 6  }
 7  DataMP = list(p=p, y=y, X=X)
 8  # Hyperparameters
 9  k  <- dim(X)[2]
10  b0 <- rep(0, k)
11  c0 <- 1000
12  B0 <- c0*diag(k)
13  B0i <- solve(B0)
14  a0 <- p - 1 + 3
15  Psi0 <- a0*diag(p)
16  Prior <- list(betabar = b0, A = B0i, nu = a0, V = Psi0)
17  # MCMC parameters
18  mcmc <- 2000
19  burnin <- 1000
20  thin <- 5
21  Mcmc <- list(R = mcmc, keep = thin)
22  Results <- bayesm::rmvpGibbs(Data = DataMP, Mcmc = Mcmc,
        Prior = Prior)
23  betatildeEq1 <- Results$betadraw[,1:4] / sqrt(Results$
        sigmadraw[,1])
24  summary(coda::mcmc(betatildeEq1))
25  betatildeEq2 <- Results$betadraw[,5:8] / sqrt(Results$
        sigmadraw[,4])
26  summary(coda::mcmc(betatildeEq2))
27  sigmadraw12 <-  Results$sigmadraw[,2] / (Results$sigmadraw
        [,1]*Results$sigmadraw[,4])^0.5
28  summary(coda::mcmc(sigmadraw12))
```

The following code shows the implementation of a Gibbs sampling algorithm using the *bayesm* package, particularly, the *rmvpGibbs* command.

R code. Multivariate probit model: Different number of regressors in each binary response, the *rmvpGibbs* command

```r
1  # Omitting last regressor for y2
2  Xnew <- X[,-8]
3  DataMP = list(p=p, y=y, X=Xnew)
4  # Hyperparameters
5  k <- dim(Xnew)[2]
6  b0 <- rep(0, k)
7  c0 <- 1000
8  B0 <- c0*diag(k)
9  B0i <- solve(B0)
10 Prior <- list(betabar = b0, A = B0i, nu = a0, V = Psi0)
11 Results <- bayesm::rmvpGibbs(Data = DataMP, Mcmc = Mcmc,
       Prior = Prior)
12 betatildeEq1 <- Results$betadraw[,1:4] / sqrt(Results$
       sigmadraw[,1])
13 summary(coda::mcmc(betatildeEq1))
14 betatildeEq2 <- Results$betadraw[,5:7] / sqrt(Results$
       sigmadraw[,4])
15 summary(coda::mcmc(betatildeEq2))
16 sigmadraw12 <-  Results$sigmadraw[,2] / (Results$sigmadraw
       [,1]*Results$sigmadraw[,4])^0.5
17 summary(coda::mcmc(sigmadraw12))
```

The following code shows how to implement a Gibbs sampling algorithm from scratch.

R code. Multivariate probit model: Different number of regressors in each binary response, Gibbs from scratch

```r
PostBetaNew <- function(Sigma, Yl){
  Sigmai <- solve(Sigma); C <- chol(Sigmai)
  Xstar <- matrix(0, n*p, k); Ystar <- matrix(0, n*p, 1)
  for(i in seq(1, n*p, 2)){
    Xstari <- C%*%Xnew[i:(i+1),]
    Xstar[i:(i+1),] <- Xstari
    Ystari <- C%*%Yl[i:(i+1)]
    Ystar[i:(i+1)] <- Ystari
  }
  Bn <- solve(B0i + t(Xstar)%*%Xstar)
  bn <- Bn%*%(B0i%*%b0 + t(Xstar)%*%Ystar)
  Beta <- MASS::mvrnorm(1, bn, Bn)
  return(Beta)
}
PostSigmaNew <- function(Beta, Yl){
  U <- matrix(0, p, p)
  for(i in seq(1, n*p, 2)){
    Ui <- (Yl[i:(i+1)] - Xnew[i:(i+1),]%*%Beta)%*%t(Yl[i:(i
    +1)] - Xnew[i:(i+1),]%*%Beta)
    U <- U + Ui
  }
  Psin <- solve(Psi0i + U)
  Sigmai <- rWishart::rWishart(1, df = an, Sigma = Psin)
  Sigma <- solve(Sigmai[,,1])
}
PostYlNew <- function(Beta, Sigma, Yl){
  w1 <- Sigma[1, ]
  w2 <- Sigma[2, ]
  f1 <- w1[-1]*Sigma[2, 2]^(-1)
  f2 <- w2[-2]*Sigma[1, 1]^(-1)
  t11 <- Sigma[1, 1] - w1[-1]*Sigma[2, 2]^(-1)*w1[-1]
  t22 <- Sigma[2, 2] - w2[-2]*Sigma[1, 1]^(-1)*w2[-2]
  Ylnew <- Yl
  for(i in seq(1, n*p, 2)){
    Ylmean1 <- Xnew[i,]%*%Beta + f1*(Yl[i+1] - Xnew[i+1,]%*%
    Beta)
    if(y[i] == 1){
      Ylnew[i] <- truncnorm::rtruncnorm(1, a = 0, b = Inf,
    mean = Ylmean1, sd = t11^0.5)
    }else{
      Ylnew[i] <- truncnorm::rtruncnorm(1, a = -Inf, b = 0,
    mean = Ylmean1, sd = t11^0.5)
    }
    Ylmean2 <- Xnew[i+1,]%*%Beta + f2*(Ylnew[i] - Xnew[i,]%*
    %Beta)
    if(y[i+1] == 1){
      Ylnew[i+1] <- truncnorm::rtruncnorm(1, a = 0, b = Inf,
     mean = Ylmean2, sd = t22^0.5)
    }else{
      Ylnew[i+1] <- truncnorm::rtruncnorm(1, a = -Inf, b =
    0, mean = Ylmean2, sd = t22^0.5)
    }
  }
  return(Ylnew)
}
```

R code. Multivariate probit model: Different number of regressors in each binary response, Gibbs from scratch

```r
an <- a0 + n
Psi0i <- solve(Psi0)
K1 <- 4; K2 <- 3; K <- K1 + K2
tot <- mcmc + burnin
PostBetas <- matrix(0, tot, K)
PostSigmas <- matrix(0, tot, p*(p+1)/2)
Beta <- rep(1, K)
Sigma <- diag(p)
pb <- winProgressBar(title = "progress bar", min = 0, max =
    tot, width = 300)
for(s in 1:tot){
  Yl <- PostYlNew(Beta = Beta, Sigma = Sigma, Yl = yl)
  Sigma <- PostSigmaNew(Beta = Beta, Yl = Yl)
  Beta <- PostBetaNew(Sigma = Sigma, Yl = Yl)
  PostBetas[s,] <- Beta
  PostSigmas[s,] <- matrixcalc::vech(Sigma)
  setWinProgressBar(pb, s, title=paste( round(s/tot*100, 0),
    "% done"))
}
close(pb)
keep <- seq((burnin+1), tot, thin)
Bs <- PostBetas[keep,]
betatilde1 <- Bs[,1:K1] / sqrt(PostSigmas[keep,1])
summary(coda::mcmc(betatilde1))
betatilde2 <- Bs[,(K1+1):(K1+K2)] / sqrt(PostSigmas[keep,3])
summary(coda::mcmc(betatilde2))
sigmadraw12 <-  PostSigmas[keep,2] / (PostSigmas[keep,1]*
    PostSigmas[keep,3])^0.5
summary(coda::mcmc(sigmadraw12))
```

7

Longitudinal/Panel data models

7.1 Solutions of Exercises

1. Show that the posterior distribution of $\boldsymbol{\beta}|\sigma^2, \boldsymbol{D}$ is $N(\boldsymbol{\beta}_n, \boldsymbol{B}_n)$, where $\boldsymbol{B}_n = (\boldsymbol{B}_0^{-1} + \sum_{i=1}^{N} \boldsymbol{X}_i^\top \boldsymbol{V}_i^{-1} \boldsymbol{X}_i)^{-1}$ and $\boldsymbol{\beta}_n = \boldsymbol{B}_n(\boldsymbol{B}_0^{-1}\boldsymbol{\beta}_0 + \sum_{i=1}^{N} \boldsymbol{X}_i^\top \boldsymbol{V}_i^{-1} \boldsymbol{y}_i)$.

   Answer

$$
\pi(\boldsymbol{\beta}|\sigma^2, \boldsymbol{D}, \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{W}) \propto \exp\left\{ -\frac{1}{2} \sum_{i=1}^{N} (\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})^\top \boldsymbol{V}_i^{-1} (\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}) \right\}
$$

$$
\times \exp\left\{ -\frac{1}{2} (\boldsymbol{\beta} - \boldsymbol{\beta}_0)^\top \boldsymbol{B}_0^{-1} (\boldsymbol{\beta} - \boldsymbol{\beta}_0) \right\}
$$

$$
\propto \exp\left\{ -\frac{1}{2} \left(-2\boldsymbol{\beta}^\top \left(\sum_{i=1}^{N} \boldsymbol{X}_i^\top \boldsymbol{V}_i^{-1} \boldsymbol{y}_i + \boldsymbol{B}_0^{-1}\boldsymbol{\beta}_0 \right) + \boldsymbol{\beta}^\top \left(\sum_{i=1}^{N} \boldsymbol{X}_i^\top \boldsymbol{V}_i^{-1} \boldsymbol{X}_i + \boldsymbol{B}_0^{-1} \right) \boldsymbol{\beta} \right) \right\}
$$

$$
= \exp\left\{ -\frac{1}{2}(-2\boldsymbol{\beta}^\top \boldsymbol{B}_n^{-1} \boldsymbol{B}_n \left(\sum_{i=1}^{N} \boldsymbol{X}_i^\top \boldsymbol{V}_i^{-1} \boldsymbol{y}_i + \boldsymbol{B}_0^{-1}\boldsymbol{\beta}_0 \right) + \boldsymbol{\beta}^\top \boldsymbol{B}_n^{-1}\boldsymbol{\beta}) \right\}
$$

$$
= \exp\left\{ -\frac{1}{2}(-2\boldsymbol{\beta}^\top \boldsymbol{B}_n^{-1}\boldsymbol{\beta}_n + \boldsymbol{\beta}^\top \boldsymbol{B}_n^{-1}\boldsymbol{\beta}) \right\}.
$$

   We can complete the square in this expression by adding and subtracting $\boldsymbol{\beta}_n^\top \boldsymbol{B}_n^{-1}\boldsymbol{\beta}_n$. Thus,

$$
\pi(\boldsymbol{\beta}|\sigma^2, \boldsymbol{D}, \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{W}) \propto \exp\left\{ -\frac{1}{2}(-2\boldsymbol{\beta}^\top \boldsymbol{B}_n^{-1}\boldsymbol{\beta}_n + \boldsymbol{\beta}^\top \boldsymbol{B}_n^{-1}\boldsymbol{\beta} + \boldsymbol{\beta}_n^\top \boldsymbol{B}_n^{-1}\boldsymbol{\beta}_n - \boldsymbol{\beta}_n^\top \boldsymbol{B}_n^{-1}\boldsymbol{\beta}_n) \right\}
$$

$$
\propto \exp\left\{ -\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_n)^\top \boldsymbol{B}_n^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}_n) \right\}.
$$

   This is the kernel of a multivariate random variable with mean $\boldsymbol{\beta}_n$ and variance matrix $\boldsymbol{B}_n$.

2. **The relation between productivity and public investment example continues**

   - Perform inference of this example using our GUI.

- Program from scratch a Gibbs sampling algorithm to perform this application.

- Perform inference in this example assuming that there is heteroskedasticity, $\mu_{it}|\tau_{it} \sim N(0, \sigma^2/\tau_{it})$ and $\tau_{it} \sim G(v/2, v/2)$ setting $v = 5$.

Answer

R code. The relationship between productivity and public investment, programming from scratch the Gibbs sampler

```r
rm(list = ls())
set.seed(12345)
DataGSP <- read.csv("https://raw.githubusercontent.com/
    besmarter/BSTApp/refs/heads/master/DataApp/8PublicCap.
    csv", sep = ",", header = TRUE, quote = "")
attach(DataGSP)
N <- length(unique(id))
y <- log(gsp)
NT <- length(y)
X <- cbind(1, log(pcap), log(pc), log(emp), unemp)
K1 <- dim(X)[2]
W <- matrix(rep(1, NT), NT, 1)
K2 <- dim(W)[2]
mcmc <- 10000; burnin <- 5000; thin <- 1; tot <- mcmc +
    burnin
b0 <- rep(0, K1); B0 <- diag(K1); B0i <- solve(B0)
r0 <- 5; R0 <- diag(K2); a0 <- 0.001; d0 <- 0.001
PostBeta <- function(sig2, D){
  XVX <- matrix(0, K1, K1)
  XVy <- matrix(0, K1, 1)
  for(i in 1:N){
    ids <- which(id == i)
    Ti <- length(ids)
    Wi <- W[ids, ]
    Vi <- sig2*diag(Ti) + Wi%*%D%*%t(Wi)
    ViInv <- solve(Vi)
    Xi <- X[ids, ]
    XVXi <- t(Xi)%*%ViInv%*%Xi
    XVX <- XVX + XVXi
    yi <- y[ids]
    XVyi <- t(Xi)%*%ViInv%*%yi
    XVy <- XVy + XVyi
  }
  Bn <- solve(B0i + XVX)
  bn <- Bn%*%(B0i%*%b0 + XVy)
  Beta <- MASS::mvrnorm(1, bn, Bn)
  return(Beta)
}
```

R code. The relationship between productivity and public investment, programming from scratch the Gibbs sampler

```r
1  Postb <- function(Beta, sig2, D){
2    Di <- solve(D)
3    bis <- matrix(0, N, K2)
4    for(i in 1:N){
5      ids <- which(id == i)
6      Wi <- W[ids, ]
7      Xi <- X[ids, ]
8      yi <- y[ids]
9      Wtei <- sig2^(-1)*t(Wi)%*%(yi - Xi%*%Beta)
10     Bni <- solve(sig2^(-1)*t(Wi)%*%Wi + Di)
11     bni <- Bni%*%Wtei
12     bi <- MASS::mvrnorm(1, bni, Bni)
13     bis[i, ] <- bi
14   }
15   return(as.matrix(bis))
16 }
17 PostSig2 <- function(Beta, bs){
18   an <- a0 + 0.5*NT
19   ete <- 0
20   for(i in 1:N){
21     ids <- which(id == i)
22     Xi <- X[ids, ]
23     yi <- y[ids]
24     Wi <- W[ids, ]
25     ei <- yi - Xi%*%Beta - Wi*bs[i, ]
26     etei <- t(ei)%*%ei
27     ete <- ete + etei
28   }
29   dn <- d0 + 0.5*ete
30   sig2 <- MCMCpack::rinvgamma(1, shape = an, scale = dn)
31   return(sig2)
32 }
33 PostD <- function(bs){
34   rn <- r0 + N
35   btb <- matrix(0, K2, K2)
36   for(i in 1:N){
37     bsi <- bs[i, ]
38     btbi <- bsi%*%t(bsi)
39     btb <- btb + btbi
40   }
41   Rn <- d0*R0 + btb
42   Sigma <- MCMCpack::riwish(v = rn, S = Rn)
43   return(Sigma)
44 }
```

R code. The relationship between productivity and public investment, programming from scratch the Gibbs sampler

```r
1  PostBetas <- matrix(0, tot, K1); PostDs <- matrix(0, tot, K2
        *(K2+1)/2)
2  PostSig2s <- rep(0, tot); Postbs <- array(0, c(N, K2, tot))
3  RegLS <- lm(y ~ X - 1); SumLS <- summary(RegLS)
4  Beta <- SumLS[["coefficients"]][,1]
5  sig2 <- SumLS[["sigma"]]^2; D <- diag(K2)
6  pb <- winProgressBar(title = "progress bar", min = 0, max =
        tot, width = 300)
7  for(s in 1:tot){
8    bs <- Postb(Beta = Beta, sig2 = sig2, D = D)
9    D <- PostD(bs = bs)
10   Beta <- PostBeta(sig2 = sig2, D = D)
11   # Beta <- PostBetaNew(sig2 = sig2, D = D)
12   sig2 <- PostSig2(Beta = Beta, bs = bs)
13   PostBetas[s,] <- Beta
14   PostDs[s,] <- matrixcalc::vech(D)
15   PostSig2s[s] <- sig2
16   Postbs[, , s] <- bs
17   setWinProgressBar(pb, s, title=paste( round(s/tot*100, 0),
        "% done"))
18 }
19 close(pb)
20 keep <- seq((burnin+1), tot, thin)
21 Bs <- PostBetas[keep,]; Ds <- PostDs[keep,]
22 bs <- Postbs[, , keep]; sig2s <- PostSig2s[keep]
23 summary(coda::mcmc(Bs))
24 Quantiles for each variable:
25        2.5%      25%       50%       75%       97.5%
26 var1  1.727227  1.934457  2.037411  2.138829  2.325961
27 var2 -0.033664  0.001015  0.018974  0.037977  0.076626
28 var3  0.274521  0.303619  0.318302  0.332881  0.363323
29 var4  0.652937  0.692479  0.712605  0.730885  0.768415
30 var5 -0.008724 -0.007265 -0.006548 -0.005844 -0.004508
31 summary(coda::mcmc(Ds))
32 summary(coda::mcmc(sig2s))
33 # Convergence diagnostics
34 coda::geweke.diag(Bs)
35 coda::raftery.diag(Bs,q=0.5,r=0.05,s = 0.95)
36 coda::heidel.diag(Bs)
```

The likelihood function in the case with heteroskedasticity is proportional

to

$$p(\boldsymbol{\beta}, \boldsymbol{b}, \sigma^2, \boldsymbol{\tau} | \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{W}) \propto \prod_{i=1}^{N} |\sigma^2 \boldsymbol{\Psi}_i|^{-1/2}$$
$$\times \exp \left\{ -\frac{1}{2\sigma^2} (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta} - \boldsymbol{W}_i \boldsymbol{b}_i)^\top \boldsymbol{\Psi}_i^{-1} (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta} - \boldsymbol{W}_i \boldsymbol{b}_i) \right\},$$

where $\boldsymbol{b} = [\boldsymbol{b}_1^\top, \boldsymbol{b}_2^\top, \dots, \boldsymbol{b}_N^\top]^\top$, $\boldsymbol{\tau} = [\tau_{it}]^\top$ and $\boldsymbol{\Psi}_i = diag\left\{ \tau_{it}^{-1} \right\}$.

Following the same procedure as in Section 9.1 of the book, and Exercise 1 of this chapter, we have that

$$\boldsymbol{\beta} | \sigma^2, \boldsymbol{\tau}, \boldsymbol{D}, \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{W} \sim N(\boldsymbol{\beta}_n, \boldsymbol{B}_n),$$

where $\boldsymbol{B}_n = (\boldsymbol{B}_0^{-1} + \sum_{i=1}^{N} \boldsymbol{X}_i^\top \boldsymbol{V}_i^{-1} \boldsymbol{X}_i)^{-1}$, $\boldsymbol{\beta}_n = \boldsymbol{B}_n (\boldsymbol{B}_0^{-1} \boldsymbol{\beta}_0 + \sum_{i=1}^{N} \boldsymbol{X}_i^\top \boldsymbol{V}_i^{-1} \boldsymbol{y}_i)$ and $\boldsymbol{V}_i = \sigma^2 \boldsymbol{\Psi}_i + \sigma_b^2 \boldsymbol{i}_{T_i} \boldsymbol{i}_{T_i}^\top$.

$$\boldsymbol{b}_i | \boldsymbol{\beta}, \sigma^2, \boldsymbol{\tau}, \boldsymbol{D}, \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{W} \sim N(\boldsymbol{b}_{ni}, \boldsymbol{B}_{ni}),$$

where $\boldsymbol{B}_{ni} = (\sigma^{-2} \boldsymbol{W}_i^\top \boldsymbol{\Psi}_i^{-1} \boldsymbol{W}_i + \boldsymbol{D}^{-1})^{-1}$ and $\boldsymbol{b}_{ni} = \boldsymbol{B}_{ni} (\sigma^{-2} \boldsymbol{W}_i^\top \boldsymbol{\Psi}_i^{-1} (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta}))$.

$$\sigma^2 | \boldsymbol{\beta}, \boldsymbol{b}, \boldsymbol{\tau}, \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{W} \sim IG(\alpha_n, \delta_n),$$

where $\alpha_n = \alpha_0 + \frac{1}{2} \sum_{i=1}^{N} T_i$ and $\delta_n = \delta_0 + \frac{1}{2} \sum_{i=1}^{N} (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta} - \boldsymbol{W}_i \boldsymbol{b}_i)^\top \boldsymbol{\Psi}_i^{-1} (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta} - \boldsymbol{W}_i \boldsymbol{b}_i)$.

$$\boldsymbol{D} | \boldsymbol{b} \sim IW(d_n, \boldsymbol{D}_n),$$

where $d_n = d_0 + N$ and $\boldsymbol{D}_n = d_0 \boldsymbol{D}_0 + \sum_{i=1}^{N} \boldsymbol{b}_i \boldsymbol{b}_i^\top$. And

$$\tau_{it} | \sigma^2, \boldsymbol{\beta}, \boldsymbol{b}, \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{W} \sim G(v_{1n}/2, v_{2ni}/2),$$

where $v_{1n} = v + 1$ and $v_{2ni} = v + \sigma^{-2} (y_{it} - \boldsymbol{x}_{it}^\top \boldsymbol{\beta} - \boldsymbol{w}_{it}^\top \boldsymbol{b}_i)^2$.

The following code shows to implement this for the productivity application. The results of the *fixed effects* are very similar compared to the results without taken into account heteroskedasticity.

R code. The relationship between productivity and public investment, programming from scratch the Gibbs sampler with heteroskedasticity

```r
1  rm(list = ls())
2  set.seed(12345)
3  DataGSP <- read.csv("https://raw.githubusercontent.com/
       besmarter/BSTApp/refs/heads/master/DataApp/8PublicCap.
       csv", sep = ",", header = TRUE, quote = "")
4  attach(DataGSP)
5  N <- length(unique(id))
6  y <- log(gsp)
7  NT <- length(y)
8  X <- cbind(1, log(pcap), log(pc), log(emp), unemp)
9  K1 <- dim(X)[2]
10 W <- matrix(rep(1, NT), NT, 1)
11 K2 <- dim(W)[2]
12 mcmc <- 10000; burnin <- 5000; thin <- 1; tot <- mcmc +
       burnin
13 b0 <- rep(0, K1); B0 <- diag(K1); B0i <- solve(B0)
14 r0 <- 5; R0 <- diag(K2); a0 <- 0.001; d0 <- 0.001; v <- 5
15 # Gibbs by hand
16 PostBeta <- function(sig2, D, tau){
17   XVX <- matrix(0, K1, K1)
18   XVy <- matrix(0, K1, 1)
19   for(i in 1:N){
20     ids <- which(id == i)
21     Ti <- length(ids)
22     Wi <- W[ids, ]
23     taui <- tau[ids]
24     Vi <- sig2*solve(diag(1/taui)) + Wi%*%D%*%t(Wi)
25     ViInv <- solve(Vi)
26     Xi <- X[ids, ]
27     XVXi <- t(Xi)%*%ViInv%*%Xi
28     XVX <- XVX + XVXi
29     yi <- y[ids]
30     XVyi <- t(Xi)%*%ViInv%*%yi
31     XVy <- XVy + XVyi
32   }
33   Bn <- solve(B0i + XVX)
34   bn <- Bn%*%(B0i%*%b0 + XVy)
35   Beta <- MASS::mvrnorm(1, bn, Bn)
36   return(Beta)
37 }
```

R code. The relationship between productivity and public investment, programming from scratch the Gibbs sampler with heteroskedasticity

```r
Postb <- function(Beta, sig2, D, tau){
  Di <- solve(D);   bis <- matrix(0, N, K2)
  for(i in 1:N){
    ids <- which(id == i)
    Wi <- W[ids, ]; Xi <- X[ids, ]
    yi <- y[ids]; taui <- tau[ids]
    Taui <- solve(diag(1/taui))
    Wtei <- sig2^(-1)*t(Wi)%*%Taui%*%(yi - Xi%*%Beta)
    Bni <- solve(sig2^(-1)*t(Wi)%*%Taui%*%Wi + Di)
    bni <- Bni%*%Wtei
    bi <- MASS::mvrnorm(1, bni, Bni)
    bis[i, ] <- bi
  }
  return(bis)
}
PostSig2 <- function(Beta, bs, tau){
  an <- a0 + 0.5*NT; ete <- 0
  for(i in 1:N){
    ids <- which(id == i)
    Xi <- X[ids, ]; yi <- y[ids]
    Wi <- W[ids, ]; taui <- tau[ids]
    Taui <- solve(diag(1/taui))
    ei <- yi - Xi%*%Beta - Wi*bs[i]
    etei <- t(ei)%*%Taui%*%ei
    ete <- ete + etei
  }
  dn <- d0 + 0.5*ete
  sig2 <- MCMCpack::rinvgamma(1, shape = an, scale = dn)
  return(sig2)
}
PostD <- function(bs){
  rn <- r0 + N
  btb <- matrix(0, K2, K2)
  for(i in 1:N){
    bsi <- bs[i, ]
    btbi <- bsi%*%t(bsi)
    btb <- btb + btbi
  }
  Rn <- d0*R0 + btb
  Sigma <- MCMCpack::riwish(v = rn, S = Rn)
  return(Sigma)
}
PostTau <- function(sig2, Beta, bs){
  v1n <- v + 1
  v2n <- NULL
  for(i in 1:NT){
    Xi <- X[i, ]; yi <- y[i]
    Wi <- W[i, ]; bi <- bs[id[i],]
    v2ni <- v + sig2^(-1)*(yi - Xi%*%Beta - Wi%*%bi)^2
    v2n <- c(v2n, v2ni)
  }
  tau <- rgamma(NT, shape = rep(v1n/2, NT), rate = v2n/2)
  return(tau)
}
```

R code. The relationship between productivity and public investment, programming from scratch the Gibbs sampler with heteroskedasticity

```r
PostBetas <- matrix(0, tot, K1); PostDs <- matrix(0, tot, K2
    *(K2+1)/2)
PostSig2s <- rep(0, tot); Postbs <- array(0, c(N, K2, tot))
PostTaus <- matrix(0, tot, NT); RegLS <- lm(y ~ X - 1)
SumLS <- summary(RegLS); Beta <- SumLS[["coefficients"]][,1]
sig2 <- SumLS[["sigma"]]^2; D <- diag(K2)
tau <- rgamma(NT, shape = v/2, rate = v/2)
pb <- winProgressBar(title = "progress bar", min = 0, max =
    tot, width = 300)
for(s in 1:tot){
  bs <- Postb(Beta = Beta, sig2 = sig2, D = D, tau = tau)
  D <- PostD(bs = bs)
  Beta <- PostBeta(sig2 = sig2, D = D, tau = tau)
  sig2 <- PostSig2(Beta = Beta, bs = bs, tau = tau)
  tau <- PostTau(sig2 = sig2, Beta = Beta, bs = bs)
  PostBetas[s,] <- Beta
  PostDs[s,] <- matrixcalc::vech(D)
  PostSig2s[s] <- sig2
  Postbs[, , s] <- bs
  PostTaus[s,] <- tau
  setWinProgressBar(pb, s, title=paste( round(s/tot*100, 0),
    "% done"))
}
close(pb)
keep <- seq((burnin+1), tot, thin)
Bs <- PostBetas[keep,]
Ds <- PostDs[keep,]
bs <- Postbs[, , keep]
sig2s <- PostSig2s[keep]
taus <- PostTaus[keep,]
summary(coda::mcmc(Bs))
Quantiles for each variable:
        2.5%        25%        50%        75%       97.5%
var1   1.61764   1.891242   2.029985   2.165677   2.427349
var2  -0.07088  -0.016921   0.013080   0.044681   0.107360
var3   0.27259   0.312197   0.333106   0.353678   0.395123
var4   0.59507   0.664490   0.699526   0.734612   0.796977
var5  -0.01046  -0.008533  -0.007586  -0.006666  -0.004918
summary(coda::mcmc(Ds))
summary(coda::mcmc(sig2s))
```

3. **Simulation exercise of the longitudinal normal model continues**

Assume that $y_{it} = \beta_0 + \beta_1 x_{it1} + \beta_2 x_{it2} + \beta_3 x_{it3} + \beta_4 z_{i1} + b_i + w_{it1} b_{i1} + \mu_{it}$ where $x_{itk} \sim N(0,1)$, $k = 1,2,3$, $z_{i1} \sim B(0.5)$, $w_{it1} \sim N(0,1)$, $b_i \sim N(0, 0.7^{1/2})$,

$b_{i1} \sim N(0, 0.6^{1/2})$, $\mu_{it} \sim N(0, 0.1^{1/2})$ $\boldsymbol{\beta} = [0.5\ 0.4\ 0.6\ -0.6\ 0.7]^{\top}$, $i = 1, 2, \ldots, 50$, and the sample size is 2000 in an *unbalanced panel structure.* In addition, we assume that $\boldsymbol{b}_i$ dependents on $\boldsymbol{z}_i = [1\ z_{i1}]^{\top}$ such that $\boldsymbol{b}_i \sim N(\boldsymbol{Z}_i \boldsymbol{\gamma}, \boldsymbol{D})$ where $\boldsymbol{Z}_i = \boldsymbol{I}_{K_2} \otimes \boldsymbol{z}_i^{\top}$, and $\boldsymbol{\gamma} = [1\ 1\ 1\ 1]$. The prior for $\boldsymbol{\gamma}$ is $N(\boldsymbol{\gamma}_0, \boldsymbol{\Gamma}_0)$ where we set $\boldsymbol{\gamma}_0 = \boldsymbol{0}_4$ and $\boldsymbol{\Gamma}_0 = \boldsymbol{I}_4$.

- Perform inference in this model without taking into account the dependence between $\boldsymbol{b}_i$ and z_{i1}, and compare the posterior estimates with the population parameters.

- Perform inference in this model taking into account the dependence between $\boldsymbol{b}_i$ and z_{i1}, and compare the posterior estimates with the population parameters.

Answer

Following the same procedure as in Section 9.1 of the book, and Exercise 1 of this chapter, we have that

$$\boldsymbol{\beta} | \sigma^2, \boldsymbol{\gamma}, \boldsymbol{D}, \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{W}, \boldsymbol{Z} \sim N(\boldsymbol{\beta}_n, \boldsymbol{B}_n),$$

where $\boldsymbol{B}_n = (\boldsymbol{B}_0^{-1} + \sum_{i=1}^{N} \boldsymbol{X}_i^{\top} \boldsymbol{V}_i^{-1} \boldsymbol{X}_i)^{-1}$, $\boldsymbol{\beta}_n = \boldsymbol{B}_n(\boldsymbol{B}_0^{-1} \boldsymbol{\beta}_0 + \sum_{i=1}^{N} \boldsymbol{X}_i^{\top} \boldsymbol{V}_i^{-1} (\boldsymbol{y}_i - \boldsymbol{W}_i \boldsymbol{Z}_i \boldsymbol{\gamma}))$ and $\boldsymbol{V}_i = \sigma^2 \boldsymbol{I}_{T_i} + \sigma_b^2 \boldsymbol{i}_{T_i} \boldsymbol{i}_{T_i}^{\top}$.

$$\boldsymbol{b}_i | \boldsymbol{\beta}, \sigma^2, \boldsymbol{\gamma}, \boldsymbol{D}, \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{W}, \boldsymbol{Z} \sim N(\boldsymbol{b}_{ni}, \boldsymbol{B}_{ni}),$$

where $\boldsymbol{B}_{ni} = (\sigma^{-2} \boldsymbol{W}_i^{\top} \boldsymbol{W}_i + \boldsymbol{D}^{-1})^{-1}$ and $\boldsymbol{b}_{ni} = \boldsymbol{B}_{ni}(\sigma^{-2} \boldsymbol{W}_i^{\top} (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta}) + \boldsymbol{D}^{-1} \boldsymbol{Z}_i \boldsymbol{\gamma})$.

$$\sigma^2 | \boldsymbol{\beta}, \boldsymbol{b}, \boldsymbol{\tau}, \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{W} \sim IG(\alpha_n, \delta_n),$$

where $\alpha_n = \alpha_0 + \frac{1}{2} \sum_{i=1}^{N} T_i$ and $\delta_n = \delta_0 + \frac{1}{2} \sum_{i=1}^{N} (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta} - \boldsymbol{W}_i \boldsymbol{b}_i)^{\top} (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta} - \boldsymbol{W}_i \boldsymbol{b}_i)$.

$$\boldsymbol{D} | \boldsymbol{b}, \boldsymbol{\gamma}, \boldsymbol{Z} \sim IW(d_n, \boldsymbol{D}_n),$$

where $d_n = d_0 + N$ and $\boldsymbol{D}_n = d_0 \boldsymbol{D}_0 + \sum_{i=1}^{N} (\boldsymbol{b}_i - \boldsymbol{Z}_i \boldsymbol{\gamma})(\boldsymbol{b}_i - \boldsymbol{Z}_i \boldsymbol{\gamma})^{\top}$. And

$$\boldsymbol{\gamma} | \boldsymbol{b}, \boldsymbol{D}, \boldsymbol{Z} \sim N(\boldsymbol{\gamma}_n, \boldsymbol{\Gamma}_n),$$

where $\boldsymbol{\Gamma}_n = (\sum_{i=1}^{N} \boldsymbol{Z}_i^{\top} \boldsymbol{D}^{-1} \boldsymbol{Z}_i + \boldsymbol{\Gamma}_0^{-1})^{-1}$ and $\boldsymbol{\gamma}_n = \boldsymbol{\Gamma}_n(\sum_{i=1}^{N} \boldsymbol{Z}_i^{\top} \boldsymbol{D}^{-1} \boldsymbol{b}_i + \boldsymbol{\Gamma}_0^{-1} \boldsymbol{\gamma}_0)$.

The following code shwos how to implement this longitudinal normal model where *random effects* are correlated with regressors. We set MCMC iterations, burn-in and thinning parameters equal to 15000, 5000 and 5, respectively. In addition, $\boldsymbol{\beta}_0 = \boldsymbol{0}_5$, $\boldsymbol{B}_0 = \boldsymbol{I}_5$, $\alpha_0 = \delta_0 = 0.001$, $d_0 = 2$, $\boldsymbol{D}_0 = \boldsymbol{I}_2$, $\boldsymbol{\gamma}_0 = \boldsymbol{0}_4$ and $\boldsymbol{\Gamma}_0 = \boldsymbol{I}_4$.

R code. Simulation exercise: Longitudinal normal model with fixed effects correlated to regressors

```r
rm(list = ls())
set.seed(010101)
NT <- 2000
N <- 50
id <- c(1:N, sample(1:N, NT - N,replace=TRUE))
table(id)
x1 <- rnorm(NT); x2 <- rnorm(NT)
x3 <- rnorm(NT); z1 <- rbinom(N, 1, 0.5)
zdata <- NULL
for(i in 1:NT){
  zdatai <- z1[id[i]]
  zdata <- c(zdata, zdatai)
}
X <- cbind(1, x1, x2, x3, zdata)
K1 <- dim(X)[2]
w1 <- rnorm(NT)
W <- cbind(1, w1)
K2 <- dim(W)[2]
B <- c(0.5, 0.4, 0.6, -0.6, 0.7)
D <- c(0.7, 0.6)
sig2 <- 0.1
u <- rnorm(NT, 0, sd = sig2^0.5)
Z <- cbind(1, z1)
K3 <- dim(Z)[2]
G <- rep(1, K3*K2)
b <- matrix(0, N, K2)
for(i in 1:N){
  ZGi <- t(kronecker(diag(K2), Z[i, ]))%*%G
  b[i, ] <- MASS::mvrnorm(1, ZGi, diag(D))
}
y <- NULL
for(i in 1:NT){
  yi <- X[i,]%*%B + W[i,]%*%b[id[i],] + u[i]
  y <- c(y, yi)
}
Data <- as.data.frame(cbind(y, x1, x2, x3, zdata, w1, id))
mcmc <- 15000; burnin <- 5000; thin <- 10; tot <- mcmc +
    burnin
b0 <- rep(0, K1); B0 <- diag(K1); B0i <- solve(B0)
r0 <- K2; R0 <- diag(K2); a0 <- 0.001; d0 <- 0.001
g0 <- rep(0, K2*K3); G0 <- diag(K2*K3); G0i <- solve(G0)
Resultshreg <- MCMCpack::MCMChregress(fixed = y~x1 + x2 + x3
     + zdata, random = ~w1, group="id",
data = Data, burnin = burnin, mcmc = mcmc, thin = thin,
mubeta = b0, Vbeta = B0,
r = r0, R = R0, nu = a0, delta = d0)
Betas <- Resultshreg[["mcmc"]][,1:K1]
Sigma2RanEff <- Resultshreg[["mcmc"]][,c(K2*N+K1+1, 2*N+K1+
    K2^2)]
Sigma2 <- Resultshreg[["mcmc"]][,K2*N+K1+K2^2+1]
summary(Betas)
summary(Sigma2RanEff)
summary(Sigma2)
```

R code. Simulation exercise: Longitudinal normal model with fixed effects correlated to regressors

```r
1  PostBeta <- function(sig2, Gamma, D){
2    XVX <- matrix(0, K1, K1)
3    XVy <- matrix(0, K1, 1)
4    for(i in 1:N){
5      ids <- which(id == i)
6      Ti <- length(ids)
7      Wi <- W[ids, ]
8      Vi <- sig2*diag(Ti) + Wi%*%D%*%t(Wi)
9      ViInv <- solve(Vi)
10     Xi <- X[ids, ]
11     XVXi <- t(Xi)%*%ViInv%*%Xi
12     XVX <- XVX + XVXi
13     Zi <- Z[i, ]
14     yi <- y[ids]
15     ZGi <- t(kronecker(diag(K2), Zi))%*%Gamma
16     XVyi <- t(Xi)%*%ViInv%*%(yi - Wi%*%ZGi)
17     XVy <- XVy + XVyi
18   }
19   Bn <- solve(B0i + XVX)
20   bn <- Bn%*%(B0i%*%b0 + XVy)
21   Beta <- MASS::mvrnorm(1, bn, Bn)
22   return(Beta)
23 }
24 Postb <- function(Beta, sig2, Gamma, D){
25   Di <- solve(D)
26   bis <- matrix(0, N, K2)
27   for(i in 1:N){
28     ids <- which(id == i)
29     Wi <- W[ids, ]
30     Xi <- X[ids, ]
31     yi <- y[ids]
32     Wtei <- sig2^(-1)*t(Wi)%*%(yi - Xi%*%Beta)
33     Bni <- solve(sig2^(-1)*t(Wi)%*%Wi + Di)
34     Zi <- Z[i, ]
35     ZGi <- t(kronecker(diag(K2), Zi))%*%Gamma
36     bni <- Bni%*%(Wtei + Di%*%ZGi)
37     bi <- MASS::mvrnorm(1, bni, Bni)
38     bis[i, ] <- bi
39   }
40   return(bis)
41 }
42 PostSig2 <- function(Beta, bs){
43 an <- a0 + 0.5*NT
44 ete <- 0
45 for(i in 1:N){
46   ids <- which(id == i)
47   Xi <- X[ids, ]
48   yi <- y[ids]
49   Wi <- W[ids, ]
50   ei <- yi - Xi%*%Beta - Wi%*%bs[i, ]
51   etei <- t(ei)%*%ei
52   ete <- ete + etei
53 }
54 dn <- d0 + 0.5*ete
55 sig2 <- MCMCpack::rinvgamma(1, shape = an, scale = dn)
56 return(sig2)
57 }
```

R code. Simulation exercise: Longitudinal normal model with fixed effects correlated to regressors

```r
PostGamma <- function(bs, D){
  Di <- solve(D)
  ZDZ <- matrix(0, K2*K3, K2*K3)
  ZDb <- matrix(0, K2*K3, 1)
  for(i in 1:N){
    Zi <- Z[i, ]
    ZZi <- t(kronecker(diag(K2), Zi))
    ZDZi <- t(ZZi)%*%Di%*%ZZi
    ZDZ <- ZDZ + ZDZi
    bi <- bs[i,]
    ZDbi <- t(ZZi)%*%Di%*%bi
    ZDb <- ZDb + ZDbi
  }
  Gn <- solve(G0i + ZDZ)
  gn <- Gn%*%(G0i%*%g0 + ZDb)
  Gamma <- MASS::mvrnorm(1, gn, Gn)
  return(Gamma)
}
PostD <- function(bs, Gamma){
  rn <- r0 + N
  btb <- matrix(0, K2, K2)
  for(i in 1:N){
    bsi <- bs[i, ]
    Zi <- Z[i, ]
    ZGi <- t(kronecker(diag(K2), Zi))%*%Gamma
    btbi <- (bsi-ZGi)%*%t(bsi-ZGi)
    btb <- btb + btbi
  }
  Rn <- d0*R0 + btb
  Sigma <- MCMCpack::riwish(v = rn, S = Rn)
  return(Sigma)
}
PostBetas <- matrix(0, tot, K1); PostGammas <- matrix(0, tot
    , K2*K3)
PostDs <- matrix(0, tot, K2*(K2+1)/2); PostSig2s <- rep(0,
    tot)
Postbs <- array(0, c(N, K2, tot)); RegLS <- lm(y ~ X - 1)
SumLS <- summary(RegLS); Beta <- SumLS[["coefficients"]][,1]
sig2 <- SumLS[["sigma"]]^2; Gamma <- rep(0, K2*K3)
D <- diag(K2)
pb <- winProgressBar(title = "progress bar", min = 0, max =
    tot, width = 300)
for(s in 1:tot){
  bs <- Postb(Beta = Beta, sig2 = sig2, Gamma = Gamma, D = D
      )
  D <- PostD(bs = bs, Gamma = Gamma)
  Beta <- PostBeta(sig2 = sig2, Gamma = Gamma, D = D)
  sig2 <- PostSig2(Beta = Beta, bs = bs)
  Gamma <- PostGamma(bs = bs, D = D)
  PostBetas[s,] <- Beta
  PostDs[s,] <- matrixcalc::vech(D)
  PostSig2s[s] <- sig2
  Postbs[, , s] <- bs
  PostGammas[s,] <- Gamma
  setWinProgressBar(pb, s, title=paste( round(s/tot*100, 0),
      "% done"))
}
close(pb)
```

R code. Simulation exercise: Longitudinal normal model with fixed effects correlated to regressors

```
1  keep <- seq((burnin+1), tot, thin)
2  Bs <- PostBetas[keep,]
3  Ds <- PostDs[keep,]
4  bs <- Postbs[, , keep]
5  sig2s <- PostSig2s[keep]
6  Gs <- PostGammas[keep, ]
7  summary(coda::mcmc(Bs))
8  Quantiles for each variable:
9           2.5%      25%      50%      75%    97.5%
10 var1 -0.3437   0.4087   0.7846   1.2108   2.1137
11 var2  0.3520   0.3800   0.3891   0.3977   0.4243
12 var3  0.5757   0.6008   0.6102   0.6193   0.6444
13 var4 -0.6247  -0.5989  -0.5887  -0.5780  -0.5530
14 var5 -0.3534   0.4210   0.8600   1.3199   2.1592
15 summary(coda::mcmc(Ds))
16 Quantiles for each variable:
17 2.5%      25%      50%      75%   97.5%
18 var1  0.6405   0.8765   1.0900   1.45774 3.3695
19 var2 -0.4688  -0.2151  -0.1182  -0.02945 0.1423
20 var3  0.4133   0.5314   0.6151   0.70782 0.9530
21 summary(coda::mcmc(sig2s))
22 Quantiles for each variable:
23 2.5%     25%      50%      75%   97.5%
24 0.1075 0.2088 0.3826 0.7386 2.5551
25 summary(coda::mcmc(Gs))
26 Quantiles for each variable:
27           2.5%      25%      50%     75% 97.5%
28 var1 -0.4446 0.3318 0.7535 1.168 2.002
29 var2 -0.4786 0.4345 0.8332 1.225 1.996
30 var3  0.8643 1.0627 1.1615 1.272 1.482
31 var4  0.5410 0.8468 0.9992 1.138 1.436
```

In one hand, the 95% credible intervals of the model without taking into account the dependence between the *random effects* and z_{i1} overestimates the posterior mean of β_0, β_4 and the variance of b_{i1}. In addition, the credible intervals are very narrow. However, the variance of μ_i is well estimated. On the other hand, all the 95% credible intervals of the model taking into account the dependence between the *random effects* and z_{i1} encompass the population parameters. However, there is a lot of variability in β_0 and β_4 posterior estimates. The posterior mean estimate of the variance of b_{i1} is very close to the population value, and the location parameters associated with z_{i1} in the mean of the *random effects* is also well estimated by the posterior mean. However, the variance of the model is overestimated.

Part III

Advanced methods: Theory, applications and programming

Bibliography

[1] Thomas Bayes. LII. an essay towards solving a problem in the doctrine of chances. by the late rev. mr. bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFR S. *Philosophical transactions of the Royal Society of London*, 53:370–418, 1763.

[2] T. Conley, C. Hansen, and P. Rossi. Plausibly exogenous. *The Review of Economics and Statistics*, 94(1):260–272, 2012.

[3] J. Geweke. *Bayesian Statistics*, chapter Evaluating the accuracy of sampling-based approaches to calculating posterior moments. Clarendon Press, Oxford, UK., 1992.

[4] P. Heidelberger and P. D. Welch. Simulation run length control in the presence of an initial transient. *Operations Research*, 31(6):1109–1144, 1983.

[5] H. Jeffreys. *Theory of Probability*. Oxford University Press, London, 1961.

[6] Valen E Johnson and David Rossell. On the use of non-local prior densities in bayesian hypothesis tests. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 72(2):143–170, 2010.

[7] Dennis V Lindley. A statistical paradox. *Biometrika*, 44(1/2):187–192, 1957.

[8] Kevin P Murphy. Conjugate bayesian analysis of the Gaussian distribution. *def*, $1(2\sigma2)$:16, 2007.

[9] A.E. Raftery and S.M. Lewis. One long run with diagnostics: Implementation strategies for Markov chain Monte Carlo. *Statistical Science*, 7:493–497, 1992.

[10] P. Rossi, G. M. Allenby, and R. McCulloch. *Bayesian Statistics and Marketing*. John Wiley & Sons, 2005.

[11] A. F. M. Smith. A General Bayesian Linear Model. *Journal of the Royal Statistical Society. Series B (Methodological).*, 35(1):67–75, 1973.

[12] A. Zellner. *Introduction to Bayesian inference in econometrics*. John Wiley & Sons Inc., 1996.