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

Symbol	Description	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:

$$\begin{aligned}
P(B|WB) &= \frac{P(B,WB)}{P(WB)} \qquad (1.1)\\
&= \frac{P(WB|B)\times P(B)}{P(WB|B)\times P(B)+(1-P(WB|B))\times(1-P(B))}\\
&= \frac{0.8\times 0.15}{0.8\times 0.15+0.2\times 0.85}\\
&= 0.41
\end{aligned}$$

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:

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

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

```
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,

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

$$
\begin{aligned}
&= \left(\frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)}\lambda^{\alpha_0-1}e^{-\lambda\beta_0}\right)\left(\frac{\lambda^{\sum_{i=1}^N y_i}e^{-N\lambda}}{\prod_{i=1}^N y_i!}\right) \\
&= \left(\frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)}\frac{1}{\prod_{i=1}^N y_i!}\right)\lambda^{\sum_{i=1}^N y_i+\alpha_0-1}e^{-\lambda(\beta_0+N)} \\
&\propto \lambda^{\sum_{i=1}^N y_i+\alpha_0-1}e^{-\lambda(\beta_0+N)}. \qquad (1.5)
\end{aligned}
$$

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

$$
\begin{aligned}
E[\lambda|\mathbf{y}] &= \frac{\alpha_n}{\beta_n} \qquad (1.6) \\
&= \frac{\sum_{i=1}^N y_i + \alpha_0}{\beta_0 + N} \\
&= \left(\frac{N}{\beta_0+N}\right)\bar{y} + \left(\frac{\beta_0}{\beta_0+N}\right)\frac{\alpha_0}{\beta_0} \\
&= w\bar{y} + (1-w)E[\lambda],
\end{aligned}
$$

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

$$
\begin{aligned}
\pi(Y_0|\mathbf{y}) &= \int_0^\infty \frac{\lambda^{y_0}e^{-\lambda}}{y_0!}\pi(\lambda|\mathbf{y})\,d\lambda \qquad (1.7) \\
&= \int_0^\infty \left(\frac{\lambda^{y_0}e^{-\lambda}}{y_0!}\right)\left(\frac{\beta_n^{\alpha_n}}{\Gamma(\alpha_n)}\lambda^{\alpha_n-1}e^{-\lambda\beta_n}\right)d\lambda \\
&= \frac{\beta_n^{\alpha_n}}{\Gamma(\alpha_n)\,y_0!}\int_0^\infty \lambda^{y_0+\alpha_n-1}e^{-\lambda(1+\beta_n)}d\lambda \\
&= \frac{\beta_n^{\alpha_n}}{\Gamma(\alpha_n)\,y_0!}\frac{\Gamma(y_0+\alpha_n)}{(1+\beta_n)^{y_0+\alpha_n}} \\
&= \frac{\Gamma(y_0+\alpha_n)}{\Gamma(\alpha_n)\,y_0!}\left(\frac{1}{1+\beta_n}\right)^{y_0}\left(\frac{\beta_n}{1+\beta_n}\right)^{\alpha_n}
\end{aligned}
$$

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

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

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

R code. Health insurance, predictive distribution using vague hyperparameters

```r
set.seed(010101)
y <- c(0, 3, 2, 1, 0) # Data
N <- length(y)

# Predictive distribution
ProbBo <- function(y, a0, b0){
  N <- length(y)
  #sample size
  aN <- a0 + sum(y)
  # Posterior shape parameter
  bN <- b0 + N
  # Posterior scale parameter
  p <- 1 / (bN + 1)
  # Probability negative binomial density
  Pr <- 1 - pnbinom(0, size = aN, prob = (1 - p))
  # Probability of visiting the Doctor
  # Observe that in R there is a slightly
  # different parametrization.
  return(Pr)
}

# Using a vague prior:
a0 <- 0.001 # Prior shape parameter
b0 <- 0.001 # Prior scale parameter
PriMeanV <- a0 / b0 # Prior mean
PriVarV <- a0 / b0^2 # Prior variance
Pp <- ProbBo(y, a0 = 0.001, b0 = 0.001)
# This setting is vague prior information.
Pp
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

$$\begin{aligned}
\pi(p|\mathbf{y}) &= \left[\frac{\Gamma(\alpha_0+\beta_0)}{\Gamma(\alpha_0)\Gamma(\beta_0)} p^{\alpha_0-1}(1-p)^{\beta_0-1}\right]\left[\prod_{i=1}^{N} p^{y_i}(1-p)^{1-y_i}\right] \\
&= \frac{\Gamma(\alpha_0+\beta_0)}{\Gamma(\alpha_0)\Gamma(\beta_0)} p^{\sum_{i=1}^N y_i+\alpha_0-1}(1-p)^{\beta_0+N-\sum_{i=1}^N y_i-1} \\
&\propto p^{\sum_{i=1}^N y_i+\alpha_0-1}(1-p)^{\beta_0+N-\sum_{i=1}^N y_i-1}.
\end{aligned} \tag{1.9}$$

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

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

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

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

$$\begin{aligned}
\pi(Y_0 = 1|\mathbf{y}) &= \int_0^1 P(Y_0 = 1|\mathbf{y}, p)\pi\left(p|\mathbf{y}\right) dp \\
&= \int_0^1 p \times \pi\left(p|\mathbf{y}\right) dp \\
&= \mathbb{E}\left[p|\mathbf{y}\right] \\
&= \frac{\alpha_n}{\alpha_n + \beta_n}.
\end{aligned} \tag{1.11}$$

To use empirical Bayes, we have the following setting

$$\begin{bmatrix} \hat{\alpha_0} & \hat{\beta_0} \end{bmatrix} = \arg\max_{\alpha_0, \beta_0} \ln(p(\mathbf{y})),$$

where

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

R code. Car claim, predictive distribution using vague hyperparameters

```
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

```
# 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(a0+b0) + lgamma(aN) + lgamma(b0+N-sum(y)) -lgamma(a0) - lgamma(b0) - lgamma(a0+b0+N)}
  # log marginal likelihood
  return(-lnp)
}
theta0 <- c(0.1, 0.1)
# 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$par[1]
# Prior shape using empirical Bayes
b0EB <- EmpBay$par[2]
# Prior scale using empirical Bayes
PriMeanEB <- a0EB /(a0EB + b0EB)
# Prior mean
PriVarEB <- (a0EB*b0EB)/(((a0EB+b0EB)^2)*(a0EB+b0EB+1))
# Prior variance
PpEB <- ProbBo(y, a0 = a0EB, b0 = b0EB)
# This setting is using empirical Bayes.
PpEB
0.47
```

R code. Car claim, density plots

```r
# Density figures
lambda <- seq(0.001, 1, 0.001)
# Values of lambda
VaguePrior <- dbeta(lambda, shape1 = a0, shape2 = b0)
EBPrior <- dbeta(lambda, shape1 = a0EB, shape2 = b0EB)
PosteriorV <- dbeta(lambda, shape1 = a0 + sum(y), shape2 = b0 + N - sum(y))
PosteriorEB <- dbeta(lambda, shape1 = a0EB + sum(y), shape2 = b0EB + N - sum(y))
# Likelihood function
Likelihood <- function(theta, y){
  LogL <- dbinom(y, 1, theta, log = TRUE)
  #  LogL <- dbern(y, theta)
  Lik <- prod(exp(LogL))
  return(Lik)
}
Liks <- sapply(lambda, function(par) {Likelihood(par, y = y)})
Sc <- max(PosteriorEB)/max(Liks)
#Scale for displaying in figure
LiksScale <- Liks * Sc
data <- data.frame(cbind(lambda, VaguePrior, EBPrior, PosteriorV, PosteriorEB, LiksScale))
#Data frame
require(ggplot2)
# Cool figures
require(latex2exp)
# LaTeX equations in figures
require(ggpubr)
# Multiple figures in one page
fig1 <- ggplot(data = data, aes(lambda, VaguePrior)) +
  geom_line() +
  xlab(TeX("$p$")) + ylab("Density") + ggtitle("Prior: Vague Beta")
fig2 <- ggplot(data = data, aes(lambda, EBPrior)) +
  geom_line() +
  xlab(TeX("$p$")) + ylab("Density") +
  ggtitle("Prior: Empirical Bayes Beta")
fig3 <- ggplot(data = data, aes(lambda, PosteriorV)) +
  geom_line() +
  xlab(TeX("$p$")) + ylab("Density") +
  ggtitle("Posterior: Vague Beta")
fig4 <- ggplot(data = data, aes(lambda, PosteriorEB)) +
  geom_line() +
  xlab(TeX("$p$")) + ylab("Density") +
  ggtitle("Posterior: Empirical Bayes Beta")
FIG <- ggarrange(fig1, fig2, fig3, fig4,
ncol = 2, nrow = 2)
annotate_figure(FIG,
top = text_grob("Vague versus Empirical Bayes: Beta-Bernoulli model", color = "black", face = "bold", size = 14))
```

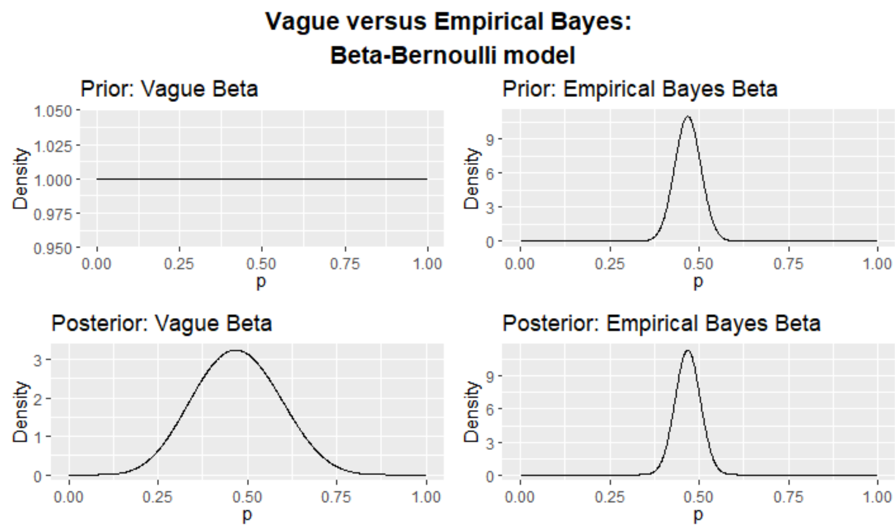

FIGURE 1.1
Vague versus Empirical Bayes: Bernoulli-Beta model.

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

```
# Prior, likelihood and posterior:
#Empirical Bayes Binonial-Beta model
dataNew <- data.frame(cbind(rep(lambda, 3),
c(EBPrior, PosteriorEB, LiksScale),
rep(1:3, each = 1000)))
#Data frame

colnames(dataNew) <- c("Lambda", "Density", "Factor")
dataNew$Factor <- factor(dataNew$Factor, levels=c("1", "3",
"2"), labels=c("Prior", "Likelihood", "Posterior"))

ggplot(data = dataNew, aes_string(x = "Lambda",
y = "Density", group = "Factor")) +
geom_line(aes(color = Factor)) +
xlab(TeX("$\\lambda$")) + ylab("Density") +
ggtitle("Prior, likelihood and posterior: Empirical Bayes
 Poisson-Gamma model") +
guides(color=guide_legend(title="Information")) +
scale_color_manual(values = c("red", "yellow", "blue"))
```

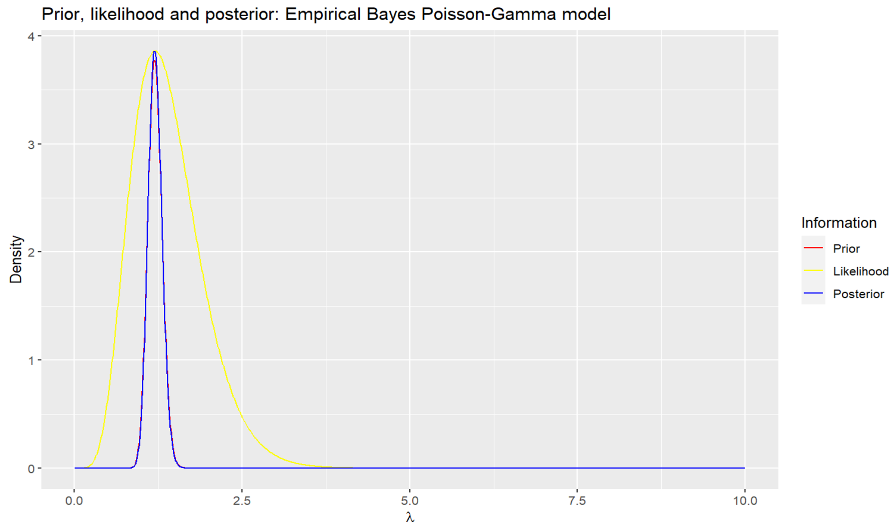

FIGURE 1.2
Prior, likelihood and posterior: Bernoulli-Beta model.

R code. Car claim, predictive probabilities plots

```
# 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))
```

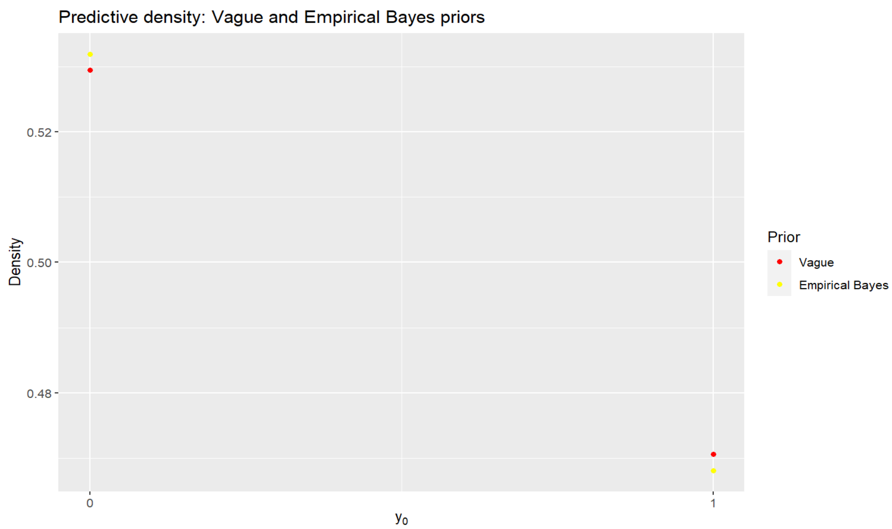

FIGURE 1.3
Predictive probabilities: Bernoulli-Beta model.

R code. Car claim, Bayesian model average

```
# Posterior odds: Vague vs Empirical Bayes
PO12 <- exp(-LogMgLik(c(a0EB, b0EB), y = y))/exp(-LogMgLik(c(a0, b0), y = y))
PostProMEM <- PO12/(1 + PO12)
# Posterior model probability Empirical Bayes
PostProMEM
0.757
PostProbMV <- 1 - PostProMEM
# Posterior model probability vague prior
PostProbMV
0.242
# Bayesian model average (BMA)
PostMeanEB <- (a0EB + sum(y)) / (a0EB + b0EB + N)
# Posterior mean Empirical Bayes
PostMeanV <- (a0 + sum(y)) / (a0 + b0 + N)
# Posterior mean vague priors
BMAmean <- PostProMEM * PostMeanEB + PostProbMV * PostMeanV
# BMA posterior mean
PostVarEB <- (a0EB + sum(y))*(b0EB + N - sum(y)) / ((a0EB + b0EB + N)^2)*(a0EB + b0EB + N -1)
# Posterior variance Empirical Bayes
PostVarV <- (a0 + sum(y))*(b0 + N - sum(y)) / ((a0 + b0 + N)^2)*(a0 + b0 + N -1)
# Posterior variance vague prior
BMAVar <- PostProMEM * PostVarEB + PostProbMV * PostVarV + PostProMEM * (PostMeanEB - BMAmean)^2 + PostProbMV * (PostMeanV - BMAmean)^2
# BMA posterior variance
# BMA: Predictive
BMAPred <- PostProMEM * PredEB + PostProbMV * PredVague
dataPredBMA <- as.data.frame(cbind(y0, BMAPred))
colnames(dataPredBMA) <- c("y0", "PredictiveBMA")
ggplot(data = dataPredBMA) +
  geom_point(aes(y0, PredictiveBMA, color = "red")) +
  xlab(TeX("$y_0$")) + ylab("Density") +
  ggtitle("Predictive density: BMA") +
  guides(color = guide_legend(title="BMA")) +
  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

```
# 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,

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

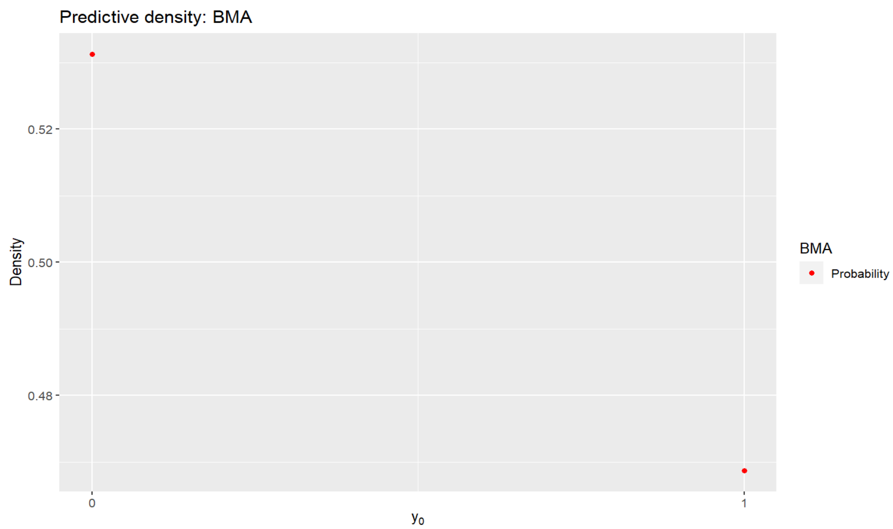

FIGURE 1.4
Predictive probabilities: Bernoulli-Beta Bayesian model average.

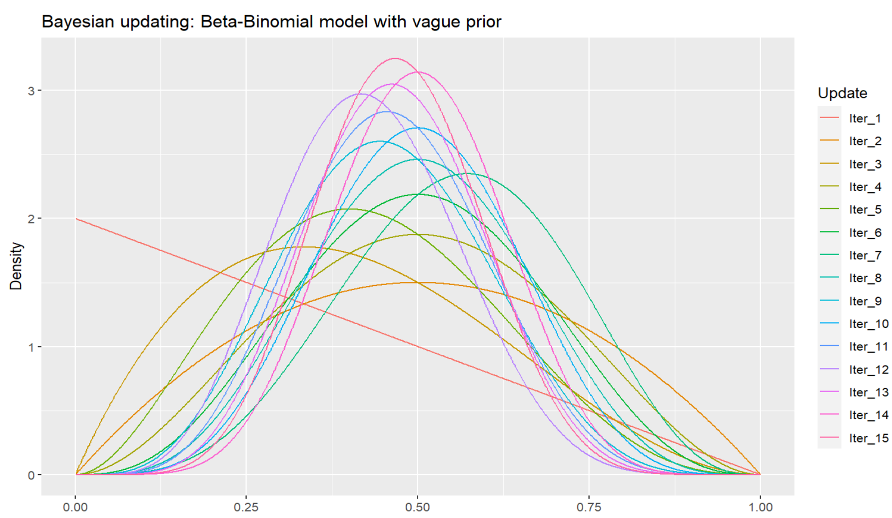

FIGURE 1.5
Predictive probabilities: Bernoulli-Beta Bayesian model updating.

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** [4, 6] 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

$$
\begin{aligned}
p(y|H_1) &= \int_0^1 \binom{n}{y}\theta^y(1-\theta)^{n-y}d\theta \\
&= \binom{n}{y}B(y+1, n-k+1) \\
&= \frac{\Gamma(N+1)}{\Gamma(y+1)\Gamma(n-y+1)}\frac{\Gamma(y+1)\Gamma(N-y+1)}{\Gamma(N+2)} \\
&= \frac{N!}{(N+1)!} \\
&= \frac{1}{N+1} \\
&\approx 1.016 \times 10^{-5}.
\end{aligned}
$$

Then, $PO_{01} = \frac{1.95\times10^{-4}}{1.016\times10^{-5}} = 19.19$, 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,

$$
\begin{aligned}
p &= 2\int_{49{,}581}^{\infty}(2\pi\sigma^2)^{-1/2}\exp\left\{-\frac{1}{2\sigma^2}(y-\mu)^2\right\}dy \\
&= 0.0235,
\end{aligned}
$$

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 ([5] 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

   $p(\mathbf{y}|\mathcal{M}_1) = \frac{\pi^{-N/2}}{2}\Gamma(N/2)2^{N/2}\left(\frac{1}{\alpha_n\hat{\sigma}^2}\right)^{N/2}\left(\frac{N}{\alpha_n\hat{\sigma}^2}\right)^{-1/2}\frac{\Gamma(1/2)\Gamma(\alpha_n/2)}{\Gamma((\alpha_N+1)/2)}$ and $p(\mathbf{y}|\mathcal{M}_0)=(2\pi)^{-N/2}\left[\frac{2}{\Gamma(N/2)}\left(\frac{N}{2}\frac{\sum_{i=1}^N(y_i-\mu_0)^2}{N}\right)^{N/2}\right]^{-1}$. Then,

$$\begin{aligned}PO_{01}&=\frac{p(\mathbf{y}|\mathcal{M}_0)}{p(\mathbf{y}|\mathcal{M}_1)}\\&=\frac{\Gamma((\alpha_n+1)/2)}{\Gamma(1/2)\Gamma(\alpha_N/2)}(\alpha_n\hat{\sigma}^2/N)^{-1/2}\left[1+\frac{(\mu_0-\bar{y})^2}{\alpha_n\hat{\sigma}^2/N}\right]^{-\left(\frac{\alpha_n+1}{2}\right)},\end{aligned}$$

   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

$$\begin{aligned}p(\mathbf{y}|\mathcal{M}_1)&=\int_{-\infty}^{\infty}\int_0^{\infty}(2\pi)^{-N/2}\sigma^{-N}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^N(y_i-\mu)^2\right\}\frac{1}{\sigma}d\sigma d\mu\\&=(2\pi)^{-N/2}\int_{-\infty}^{\infty}\int_0^{\infty}\sigma^{-(N+1)}\exp\left\{-\frac{N}{2\sigma^2}\frac{\sum_{i=1}^N(y_i-\mu)^2}{N}\right\}d\sigma d\mu\\&=(2\pi)^{-N/2}\frac{\Gamma(N/2)}{2}2^{N/2}\int_{-\infty}^{\infty}\left[\sum_{i=1}^N(y_i-\mu)^2\right]^{-N/2}d\mu\\&=(2\pi)^{-N/2}\frac{\Gamma(N/2)}{2}2^{N/2}\int_{-\infty}^{\infty}\left[\sum_{i=1}^N[(y_i-\bar{y})-(\mu-\bar{y})]^2\right]^{-N/2}d\mu\\&=(2\pi)^{-N/2}\frac{\Gamma(N/2)}{2}2^{N/2}\int_{-\infty}^{\infty}\left[\alpha_n\hat{\sigma}^2+N(\mu-\bar{y})^2\right]^{-N/2}d\mu\\&=(2\pi)^{-N/2}\frac{\Gamma(N/2)}{2}2^{N/2}\left(\frac{\alpha_n\hat{\sigma}^2}{\alpha_n\hat{\sigma}^2}\right)^{-N/2}\int_{-\infty}^{\infty}\left[\alpha_n\hat{\sigma}^2+N(\mu-\bar{y})^2\right]^{-N/2}d\mu\\&=(2\pi)^{-N/2}\frac{\Gamma(N/2)}{2}2^{N/2}\left(\alpha_n\hat{\sigma}^2\right)^{-N/2}\int_{-\infty}^{\infty}\left[1+\frac{N(\mu-\bar{y})^2}{\alpha_n\hat{\sigma}^2}\right]^{-N/2}d\mu\\&=\frac{\pi^{-N/2}}{2}\Gamma(N/2)2^{N/2}\left(\frac{1}{\alpha_n\hat{\sigma}^2}\right)^{N/2}\left(\frac{N}{\alpha_n\hat{\sigma}^2}\right)^{-1/2}\frac{\Gamma(1/2)\Gamma(\alpha_n/2)}{\Gamma((\alpha_N+1)/2)}.\end{aligned}$$

   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 [10].

$$\begin{aligned}
p(\mathbf{y}|\mathcal{M}_0) &= \int_0^{\infty} (2\pi)^{-N/2}\sigma^{-N} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i-\mu_0)^2\right\}\frac{1}{\sigma}d\sigma \\
&= (2\pi)^{-N/2}\int_0^{\infty} \sigma^{-(N+1)} \exp\left\{-\frac{N}{2\sigma^2}\frac{\sum_{i=1}^{N}(y_i-\mu_0)^2}{N}\right\}d\sigma \\
&= (2\pi)^{-N/2}\left[\frac{2}{\Gamma(N/2)}\left(\frac{N}{2}\frac{\sum_{i=1}^{N}(y_i-\mu_0)^2}{N}\right)^{N/2}\right]^{-1}.
\end{aligned}$$

The third line takes into account that the integral in the second line is the kernel of an inverted-gamma distribution [10].

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

In addition,

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

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

```
N <- 50 # Sample size
y_bar <- 102 # Sample mean
s2 <- 10 # Sample variance
alpha <- N - 1
serror <- (s2/N)^0.5
y.H0 <- c(100, 100.5, 101, 101.5, 102)
test <- (y.H0 - y_bar)/serror
pval <- 2*pt(test, alpha)
pval
0.0000459 0.0015431 0.0299338 0.2690040 1
# p-values
P001 <- (gamma(N/2)*((N-1)*serror^2)^(-0.5)*(1+test^2/alpha)^(-N/2))/(gamma(1/2)*gamma((N-1)/2))
P001/(1+P001)
0.0001705 0.0050345 0.0725330 0.3210223 0.4702050
# Posterior model probability of the null hypothesis.
```

3

Objective and subjective Bayesian approaches

3.1 Solutions of Exercises

1. Elicitation ...

R code. Example: Math test

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

4

Cornerstone models: Conjugate families

4.1 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 $\mathbf{y} = [y_1, y_2, \dots, y_N]^\top$ 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

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

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

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

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

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

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

Finally, the predictive distribution is

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

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 $\mathbf{y} = [y_1, y_2, \dots, y_N]^\top$ 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

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

where tr and vec are the trace and vectorization operators, respectively.
Then, $h(\mathbf{y}) = (2\pi)^{-N/2}$, $\eta(\boldsymbol{\mu},\boldsymbol{\Sigma}) = \left[\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\quad vec\left(\boldsymbol{\Sigma}^{-1}\right)\right]$, $T(\mathbf{y}) = \left[\mathbf{y}\quad -\frac{1}{2}vec(\mathbf{y}\mathbf{y}^{\top})\right]$ 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

$$\begin{aligned}
p(\mathbf{Y}) &= \int_{\mathcal{R}^p}\int_{\mathcal{S}}(2\pi)^{-pN/2}|\boldsymbol{\Sigma}|^{-N/2}\exp\left\{-\frac{1}{2}tr[(\mathbf{S}+N(\boldsymbol{\mu}-\hat{\boldsymbol{\mu}})(\boldsymbol{\mu}-\hat{\boldsymbol{\mu}})^{\top})\boldsymbol{\Sigma}^{-1}]\right\}\\
&\times(2\pi)^{-p/2}\beta_0^{p/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\left\{-\frac{\beta_0}{2}tr[(\boldsymbol{\mu}-\boldsymbol{\mu}_0)(\boldsymbol{\mu}-\boldsymbol{\mu}_0)^{\top}\boldsymbol{\Sigma}^{-1}]\right\}\\
&\times|\boldsymbol{\Sigma}|^{-(\alpha_0+p+1)/2}\frac{2^{-\alpha_0 p/2}|\boldsymbol{\Psi}_0|^{\alpha_0/2}}{\Gamma_p(\alpha_0/2)}\exp\left\{-\frac{1}{2}tr(\boldsymbol{\Psi}_0\boldsymbol{\Sigma}^{-1})\right\}d\boldsymbol{\Sigma}d\boldsymbol{\mu}\\
&= \frac{(2\pi)^{-frac12(pN+p)}|\boldsymbol{\Psi}_0|^{\alpha_0/2}\beta_0^{p/2}2^{-\alpha_0 p/2}}{\Gamma_p(\alpha_0/2)}\int_{\mathcal{R}^p}\int_{\mathcal{S}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}(N+1+\alpha_0+p+1)}\\
&\times\exp\left\{-\frac{1}{2}tr[(\mathbf{S}+N(\boldsymbol{\mu}-\hat{\boldsymbol{\mu}})(\boldsymbol{\mu}-\hat{\boldsymbol{\mu}})^{\top}+\beta_0(\boldsymbol{\mu}-\boldsymbol{\mu}_0)(\boldsymbol{\mu}-\boldsymbol{\mu}_0)^{\top}+\boldsymbol{\Psi}_0)\boldsymbol{\Sigma}^{-1}]\right\}d\boldsymbol{\Sigma}d\boldsymbol{\mu}.
\end{aligned}$$

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

$$\begin{aligned}
p(\mathbf{Y}) &= \frac{\Gamma_p\left(\frac{N+1+\alpha_0}{2}\right)|\boldsymbol{\Psi}_0|^{\alpha_0/2}\beta_0^{p/2}}{\Gamma_p(\alpha_0/2)\pi^{p(N+1)/2}}\\
&\times \int_{\mathcal{R}^p} |\mathbf{S}+\boldsymbol{\Psi}_0+(N+\beta_0)(\boldsymbol{\mu}-\boldsymbol{\mu}_n)(\boldsymbol{\mu}-\boldsymbol{\mu}_n)^\top\\
&+ N\beta_0/(N+\beta_0)(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu}_0)(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu}_0)^\top|d\boldsymbol{\mu}\\
&= \frac{\Gamma_p\left(\frac{N+1+\alpha_0}{2}\right)|\boldsymbol{\Psi}_0|^{\alpha_0/2}\beta_0^{p/2}}{\Gamma_p(\alpha_0/2)\pi^{p(N+1)/2}}\\
&\times \int_{\mathcal{R}^p} |\boldsymbol{\Psi}_n||1+\beta_n(\boldsymbol{\mu}-\boldsymbol{\mu}_n)\boldsymbol{\Psi}_n^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_n)^\top|^{-\frac{1}{2}(\alpha_n+1)}d\boldsymbol{\mu}\\
&= \frac{\Gamma_p\left(\frac{\alpha_n+1}{2}\right)|\boldsymbol{\Psi}_0|^{\alpha_0/2}\beta_0^{p/2}}{\Gamma_p(\alpha_0/2)\pi^{p(N+1)/2}}|\boldsymbol{\Psi}_n|^{-\frac{1}{2}(\alpha_n+1)}\\
&\times \int_{\mathcal{R}^p} [1+\beta_n(\boldsymbol{\mu}-\boldsymbol{\mu}_n)^\top\boldsymbol{\Psi}_n^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_n)]^{-\frac{1}{2}(\alpha_n+1)}d\boldsymbol{\mu}.
\end{aligned}$$

The last equality uses the definition of $\boldsymbol{\Psi}_n$, β_n and α_n, and the Sylvester's determinant theorem. Observe that we have the kernel of a multivariate t distribution [7]. Then,

$$\begin{aligned}
p(\mathbf{Y}) &= \frac{\Gamma_p\left(\frac{\alpha_n+1}{2}\right)|\boldsymbol{\Psi}_0|^{\alpha_0/2}\beta_0^{p/2}}{\Gamma_p(\alpha_0/2)\pi^{p(N+1)/2}}|\boldsymbol{\Psi}_n|^{-\frac{1}{2}(\alpha_n+1)}\\
&\times \int_{\mathcal{R}^p}\left[1+\frac{1}{\alpha_n+1-p}(\boldsymbol{\mu}-\boldsymbol{\mu}_n)^\top\left(\frac{\boldsymbol{\Psi}_n}{\beta_n(\alpha_n+1-p)}\right)^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_n)\right]^{-\frac{1}{2}(\alpha_n+1-p+p)} d\boldsymbol{\mu}\\
&= \frac{\Gamma_p\left(\frac{\alpha_n+1}{2}\right)\Gamma_p\left(\frac{\alpha_n+1-p}{2}\right)|\boldsymbol{\Psi}_0|^{\alpha_0/2}\beta_0^{p/2}(\alpha_n+1-p)^{p/2}\pi^{p/2}|\boldsymbol{\Psi}_n|^{-\frac{1}{2}(\alpha_n+1)}}{\Gamma_p(\alpha_0/2)\pi^{p(N+1)/2}\Gamma_p\left(\frac{\alpha_n+1-p+p}{2}\right)\left(\frac{\boldsymbol{\Psi}_n}{\alpha_n+1-p}\right)^{-1/2}}\\
&= \frac{\Gamma_p\left(\frac{v_n}{2}\right)}{\Gamma_p\left(\frac{\alpha_0}{2}\right)}\frac{|\boldsymbol{\Psi}_0|^{\alpha_0/2}}{|\boldsymbol{\Psi}_n|^{\alpha_n/2}}\left(\frac{\beta_0}{\beta_n}\right)^{p/2}(2\pi)^{-Np/2},
\end{aligned}$$

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

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

   Answer

$$\begin{aligned}
p(\mathbf{Y}_0|\mathbf{Y}) &\propto \int_{\mathcal{R}^p}\int_{\mathcal{S}} |\mathbf{\Sigma}|^{-1/2}\exp\left\{-\frac{1}{2}tr[(\mathbf{y}_0-\boldsymbol{\mu})(\mathbf{y}_0-\boldsymbol{\mu})^{\top}\mathbf{\Sigma}^{-1}]\right\}\\
&\times |\mathbf{\Sigma}|^{-1/2}\exp\left\{-\frac{\beta_n}{2}tr[(\boldsymbol{\mu}-\boldsymbol{\mu}_n)(\boldsymbol{\mu}-\boldsymbol{\mu}_n)^{\top}\mathbf{\Sigma}^{-1}]\right\}\\
&\times |\mathbf{\Sigma}|^{-(\alpha_n+p+1)/2}\exp\left\{-\frac{1}{2}tr(\mathbf{\Psi}_n\mathbf{\Sigma}^{-1})\right\}d\mathbf{\Sigma}d\boldsymbol{\mu}\\
&\propto \int_{\mathcal{R}^p} |(\mathbf{y}_0-\boldsymbol{\mu})(\mathbf{y}_0-\boldsymbol{\mu})^{\top}+(\boldsymbol{\mu}-\boldsymbol{\mu}_n)(\boldsymbol{\mu}-\boldsymbol{\mu}_n)^{\top}+\mathbf{\Psi}_n|^{-(\alpha_n+2)/2}d\boldsymbol{\mu}.
\end{aligned}$$

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

Taking into account that

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

Then,

$$\begin{aligned}
p(\mathbf{Y}_0|\mathbf{Y}) &\propto \int_{\mathcal{R}^p} |(\mathbf{y}_0-\boldsymbol{\mu})(\mathbf{y}_0-\boldsymbol{\mu})^{\top}+(\boldsymbol{\mu}-\boldsymbol{\mu}_n)(\boldsymbol{\mu}-\boldsymbol{\mu}_n)^{\top}+\mathbf{\Psi}_n|^{-(\alpha_n+2)/2}d\boldsymbol{\mu}\\
&= \int_{\mathcal{R}^p}\left|(1+\beta_n)\left(\boldsymbol{\mu}-\frac{(\mathbf{y}_0+\beta_n\boldsymbol{\mu}_n)}{1+\beta_n}\right)\left(\boldsymbol{\mu}-\frac{(\mathbf{y}_0+\beta_n\boldsymbol{\mu}_n)}{1+\beta_n}\right)^{\top}\right.\\
&\left.+\frac{\beta_n}{1+\beta_n}(\mathbf{y}_0-\boldsymbol{\mu}_n)(\mathbf{y}_0-\boldsymbol{\mu}_n)^{\top}+\mathbf{\Psi}_n\right|^{-(\alpha_n+2)/2}d\boldsymbol{\mu}\\
&= \int_{\mathcal{R}^p}\left[\left|\underbrace{\mathbf{\Psi}_n+\frac{\beta_n}{1+\beta_n}(\mathbf{y}_0-\boldsymbol{\mu}_n)(\mathbf{y}_0-\boldsymbol{\mu}_n)^{\top}}_{\mathbf{\Lambda}_n}\right|\right.\\
&\left.\left|1+(1+\beta_n)\left(\boldsymbol{\mu}-\frac{(\mathbf{y}_0+\beta_n\boldsymbol{\mu}_n)}{1+\beta_n}\right)^{\top}\frac{1}{\alpha_n+2-p}\left(\frac{\mathbf{\Lambda}_n}{\alpha_n+2-p}\right)^{-1}\left(\boldsymbol{\mu}-\frac{(\mathbf{y}_0+\beta_n\boldsymbol{\mu}_n)}{1+\beta_n}\right)\right|\right]^{-(\alpha_n+2-p+p)/2}d\boldsymbol{\mu}\\
&\propto \left|\mathbf{\Psi}_n+\frac{\beta_n}{1+\beta_n}(\mathbf{y}_0-\boldsymbol{\mu}_n)(\mathbf{y}_0-\boldsymbol{\mu}_n)^{\top}\right|^{-(\alpha_n+2)/2}\\
&\times \left|\mathbf{\Psi}_n+\frac{\beta_n}{1+\beta_n}(\mathbf{y}_0-\boldsymbol{\mu}_n)(\mathbf{y}_0-\boldsymbol{\mu}_n)^{\top}\right|^{1/2}\\
&= \left|\mathbf{\Psi}_n+\frac{\beta_n}{1+\beta_n}(\mathbf{y}_0-\boldsymbol{\mu}_n)(\mathbf{y}_0-\boldsymbol{\mu}_n)^{\top}\right|^{-(\alpha_n+1)/2}\\
&\propto \left[1+(\mathbf{y}_0-\boldsymbol{\mu}_n)^{\top}\frac{1}{\alpha_n+1-p}\left(\frac{\mathbf{\Psi}_n(1+\beta_n)}{(\alpha_n+1-p)\beta_n}\right)^{-1}(\mathbf{y}_0-\boldsymbol{\mu}_n)\right]^{-(\alpha_n+1-p+p)}.
\end{aligned}$$

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 $\mathbf{Y}_0|\mathbf{Y} \sim T_{N_0,M}(\alpha_n - M + 1, \mathbf{X}_0\mathbf{B}_n, \mathbf{I}_{N_0} + \mathbf{X}_0\mathbf{V}_n\mathbf{X}_0^\top, \boldsymbol{\Psi}_n)$.

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

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,

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

The second proportionality follows from the Sylvester's theorem. Observe that this is the kernel of a matrix t distribution with α_n+1-M degrees of freedom, location $\mathbf{X}_0\mathbf{B}_n$ and scale matrices $\mathbf{\Psi}_n$ 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 $\mathbf{\Psi}_n=\mathbf{\Psi}_0+\mathbf{S}+(\hat{\mathbf{B}}-\mathbf{B}_0)^{\top}\mathbf{V}_n(\hat{\mathbf{B}}-\mathbf{B}_0)$ in the linear multivariate regression model.

Answer

Taking into account that

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

Observe that

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

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 [9]:

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

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,

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

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 $\widehat{\mathbf{B}}=(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}$.

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 $(\mathbf{A}+\mathbf{B})^{-1}=\mathbf{A}^{-1}-\mathbf{A}^{-1}(\mathbf{A}^{-1}+\mathbf{B}^{-1})^{-1}\mathbf{A}^{-1}$ [9], 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 $\boldsymbol{\beta}_{**}=\mathbf{X}_0\boldsymbol{\beta}_n$,

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

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 $(\mathbf{Y}-\mathbf{X}\mathbf{B})^\top(\mathbf{Y}-\mathbf{X}\mathbf{B}) = \mathbf{S} + (\mathbf{B}-\widehat{\mathbf{B}})^\top\mathbf{X}^\top\mathbf{X}(\mathbf{B}-\widehat{\mathbf{B}})$ 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

$$
\begin{aligned}
(\mathbf{Y}-\mathbf{X}\mathbf{B})^\top(\mathbf{Y}-\mathbf{X}\mathbf{B}) &= (\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}}+\mathbf{X}\hat{\mathbf{B}}-\mathbf{X}\mathbf{B})^\top(\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}}+\mathbf{X}\hat{\mathbf{B}}-\mathbf{X}\mathbf{B})\\
&= (\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}})^\top(\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}}) + 2(\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}})^\top(\mathbf{X}\hat{\mathbf{B}}-\mathbf{X}\mathbf{B})\\
&+ (\mathbf{X}\mathbf{B}-\mathbf{X}\hat{\mathbf{B}})^\top(\mathbf{X}\mathbf{B}-\mathbf{X}\hat{\mathbf{B}})\\
&= \mathbf{S} + (\mathbf{B}-\hat{\mathbf{B}})^\top\mathbf{X}^\top\mathbf{X}(\mathbf{B}-\hat{\mathbf{B}}),
\end{aligned}
$$

   given that $(\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}})^\top(\mathbf{X}\hat{\mathbf{B}}-\mathbf{X}\mathbf{B}) = \hat{\mathbf{U}}^\top\mathbf{X}(\hat{\mathbf{B}}-\mathbf{B})$, 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 18^{th} 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

```
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
```

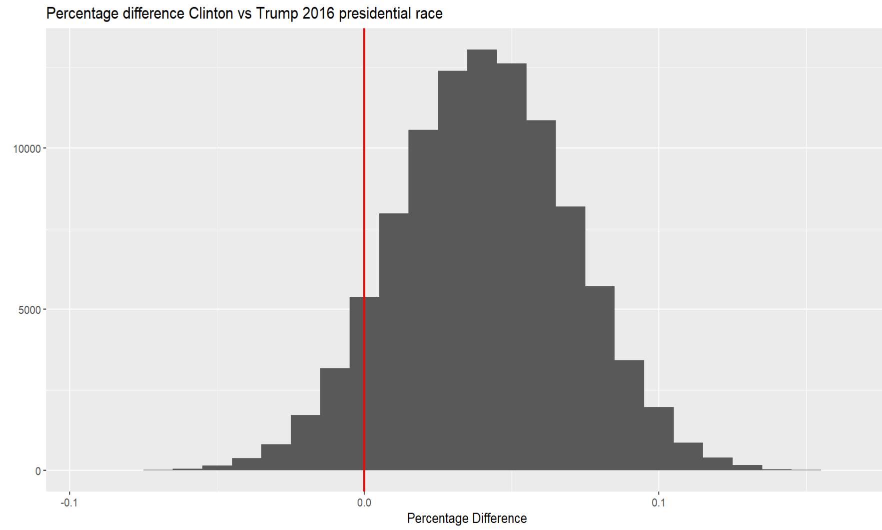

FIGURE 4.1
Percentage difference: Hillary Clinton vs Donald Trump, five-way race.

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

```
# Predictive distribution by simulation
y0 <- c(44, 40, 16)
Pred <- apply(thetas, 1, function(p) {rmultinom(1, size = sum(y0), prob = p)})
sum(sapply(1:S, function(s) {sum(Pred[,s] == y0) == 3}))/S
0.00825
# Predictive distribution by analytical expression
PredY0 <- function(y0){
  n <- sum(y0)
  Res1 <- sum(sapply(1:length(y), function(l){lgamma(alphan[l]+y0[l]) - lgamma(alphan[l])-lfactorial(y0[l])}))
  Res <- lfactorial(n)+lgamma(sum(alphan))-lgamma(sum(alphan)+n) + Res1
  return(exp(Res))
}
PredY0(y0)
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
set.seed(010101)
N <- 50
# Sample size
muhat <- 102
# Sample mean
sig2hat <- 10
# Sample variance
# Hyperparameters
mu0 <- 100
beta0 <- 1
delta0 <- 0.001
alpha0 <- 0.001
S <- 100000
# Posterior draws
alphan <- alpha0 + N
deltan <- sig2hat*(N - 1) + delta0 + beta0*N/(beta0 + N)*(muhat - mu0)^2
sig2Post <- invgamma::rinvgamma(S, shape = alphan, rate = deltan)
summary(sig2Post)
betan <- beta0 + N
mun <- (beta0*mu0 + N*muhat)/betan
muPost <- sapply(sig2Post, function(s2){rnorm(1, mun, sd = (s2/betan)^0.5)})
muPostq <- quantile(muPost, c(0.025, 0.5, 0.975))
muPostq
   2.5%      50%    97.5%
101.0929 101.9625 102.8311
cutoff <- 103
PmuPostcutoff <- mean(muPost > cutoff)
PmuPostcutoff
0.00994
# Using Student's t
muPost_t <- ((deltan/(alphan*betan))^0.5)*rt(S, alphan) + mun
c1 <- rgb(173,216,230,max = 255, alpha = 50, names = "lt.blue")
c2 <- rgb(255,192,203, max = 255, alpha = 50, names = "lt.pink")
hist(muPost, main = "Histogram: Posterior mean", xlab = "Posterior mean", col = c2)
hist(muPost_t, main = "Histogram: Posterior mean", xlab = "Posterior mean", add = T, col = c1)
muPost_tq <- quantile(muPost_t, c(0.025, 0.5, 0.975))
muPost_tq
2.5%      50%    97.5%
101.0837 101.9608 102.8435
PmuPost_tcutoff <- mean(muPost_t > cutoff)
PmuPost_tcutoff
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 4.2.

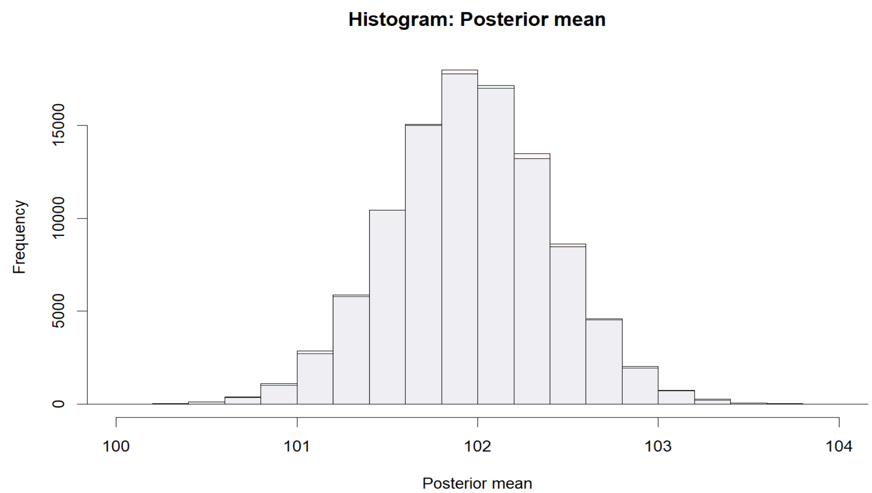

FIGURE 4.2
Histogram using the posterior conditional distribution and the posterior marginal distribution

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
rm(list = ls())
set.seed(010101)
# Electricity demand
DataUt <- read.csv("DataApplications/Utilities.csv", sep = ",", header = TRUE, fileEncoding = "latin1")
DataUtEst <- DataUt %>%
filter(Electricity != 0)
attach(DataUtEst)
# Dependent variable: Monthly consumption (kWh) in log
Y <- log(Electricity)
N <- length(Y)
# Regressors quantity including intercept
X <- cbind(LnPriceElect, IndSocio1, IndSocio2, Altitude, Nrooms, HouseholdMem, Children, Lnincome, 1)
# Regressor without price
Xnew <- cbind(IndSocio1, IndSocio2, Altitude, Nrooms, HouseholdMem, Children, Lnincome, 1)
# Log marginal function (multiply by -1 due to minimization)
LogMarLikLM <- function(X, c0){
  k <- dim(X)[2]
  N <- dim(X)[1]
  # Hyperparameters
  B0 <- c0*diag(k)
  b0 <- rep(0, k)
  # Posterior parameters
  bhat <- solve(t(X)%*%X)%*%t(X)%*%Y
  # Force this matrix to be symmetric
  Bn <- as.matrix(Matrix::forceSymmetric(solve(solve(B0) + t(X)%*%X)))
  bn <- Bn%*%(solve(B0)%*%b0 + t(X)%*%X%*%bhat)
  dn <- as.numeric(d0 + t(Y)%*%Y+t(b0)%*%solve(B0)%*%b0-t(bn)%*%solve(Bn)%*%bn)
  an <- a0 + N
  # Log marginal likelihood
  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)
  return(-logpy)
}
# Hyperparameters
d0 <- 0.001/2
a0 <- 0.001/2
# Empirical Bayes: Obtain c0 maximizing the log marginal likelihood
c0 <- 1000
EB <- optim(c0, fn = LogMarLikLM, method = "Brent", lower = 0.0001, upper = 10^6, X = X)
EBnew <- optim(c0, fn = LogMarLikLM, method = "Brent", lower = 0.0001, upper = 10^6, X = Xnew)
# Change of order to take into account the -1 in the LogMarLikLM function
BFEM <- exp(EBnew$value - EB$value)
BFEM
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(electricity price$_i$), log(water price$_i$), log(gas price$_i$), IndSocio1$_i$, IndSocio2$_i$, Altitude$_i$, Nrooms$_i$, HouseholdMem$_i$, Children$_i$, and log(Income$_i$). 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 5).

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
rm(list = ls())
set.seed(010101)
library(dplyr)
# Electricity demand
DataUt <- read.csv("DataApplications/Utilities.csv", sep = ",", header = TRUE, fileEncoding = "latin1")
DataUtEst <- DataUt %>%
filter(Electricity != 0 & Water !=0 & Gas != 0)
attach(DataUtEst)
Y <- cbind(log(Electricity), log(Water), log(Gas))
X <- cbind(LnPriceElect, LnPriceWater, LnPriceGas, IndSocio1, IndSocio2, Altitude, Nrooms, HouseholdMem, Children, Lnincome, 1)
M <- dim(Y)[2]
K <- dim(X)[2]
N <- dim(Y)[1]
# Hyperparameters
B0 <- matrix(0, K, M)
c0 <- 1000
V0 <- c0*diag(K)
Psi0 <- c0*diag(M)
a0 <- M
# Posterior parameters
Bhat <- solve(t(X)%*%X)%*%t(X)%*%Y
S <- t(Y - X%*%Bhat)%*%(Y - X%*%Bhat)
Vn <- solve(solve(V0) + t(X)%*%X)
Bn <- Vn%*%(solve(V0)%*%B0 + t(X)%*%X%*%Bhat)
Psin <- Psi0 + S + t(B0)%*%solve(V0)%*%B0 + t(Bhat)%*%t(X)%*%X%*%Bhat - t(Bn)%*%solve(Vn)%*%Bn
an <- a0 + N
#Posterior draws
s <- 10000 #Number of posterior draws
SIGs <- replicate(s, LaplacesDemon::rinvwishart(an, Psin))
BsCond <- sapply(1:s, function(s) {MixMatrix::rmatrixnorm(n = 1, mean=Bn, U = Vn,V = SIGs[,,s])})
summary(coda::mcmc(t(BsCond)))
Bs <- sapply(1:s, function(s) {MixMatrix::rmatrixt(n = 1, mean=Bn, U = Vn,V = Psin, df = an + 1 - M)})
summary(coda::mcmc(t(Bs)))
SIGMs <- t(sapply(1:s, function(l) {gdata::lowerTriangle(SIGs[,,l], diag=TRUE, byrow=FALSE)}))
summary(coda::mcmc(SIGMs))
hdiBs <- HDInterval::hdi(t(BsCond), credMass = 0.95) # Highest posterior density credible interval
hdiBs
hdiSIG <- HDInterval::hdi(SIGMs, credMass = 0.95) # Highest posterior density credible interval
hdiSIG

```

R code. Utilities demand: Multivariate regression, Bayes factors

```
# Log marginal function (multiply by -1 due to minimization)
LogMarLikLM <- function(X, c0){
  c10 <- c0[1]; c20 <- c0[2]
  k <- dim(X)[2]
  N <- dim(X)[1]
  # Hyperparameters
  V0 <- c10*diag(K)
  Psi0 <- c20*diag(M)
  # Posterior parameters
  Bhat <- solve(t(X)%*%X)%*%t(X)%*%Y
  S <- t(Y - X%*%Bhat)%*%(Y - X%*%Bhat)
  Vn <- solve(solve(V0) + t(X)%*%X)
  Bn <- Vn%*%(solve(V0)%*%B0 + t(X)%*%X%*%Bhat)
  Psin <- Psi0 + S + t(B0)%*%solve(V0)%*%B0 + t(Bhat)%*%t(X)
    %*%X%*%Bhat - t(Bn)%*%solve(Vn)%*%Bn
  # Log marginal likelihood
  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)
  return(-logpy)
}
c0 <- rep(1000, 2)
LogML <- LogMarLikLM(X=X, c0 = c0)
# Using income in dollars as regressor
Xnew <- cbind(LnPriceElect, LnPriceWater, LnPriceGas,
    IndSocio1, IndSocio2, Altitude, Nrooms, HouseholdMem,
    Children, exp(Lnincome), 1)
LogMLnew <- LogMarLikLM(X=Xnew, c0 = c0)
# Bayes factor
BF12 <- exp(LogMLnew - LogML)
BF12
# Using income in thousand dollars as regressor
XnewT <- cbind(LnPriceElect, LnPriceWater, LnPriceGas,
    IndSocio1, IndSocio2, Altitude, Nrooms, HouseholdMem,
    Children, exp(Lnincome)/1000, 1)
LogMLnewT <- LogMarLikLM(X=XnewT, c0 = c0)
# Bayes factor
BF13 <- exp(LogMLnewT - LogML)
BF13
```

R code. Utilities demand: Multivariate regression, predictive distribution

```
# 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 4.3, 4.4 and 4.5 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.

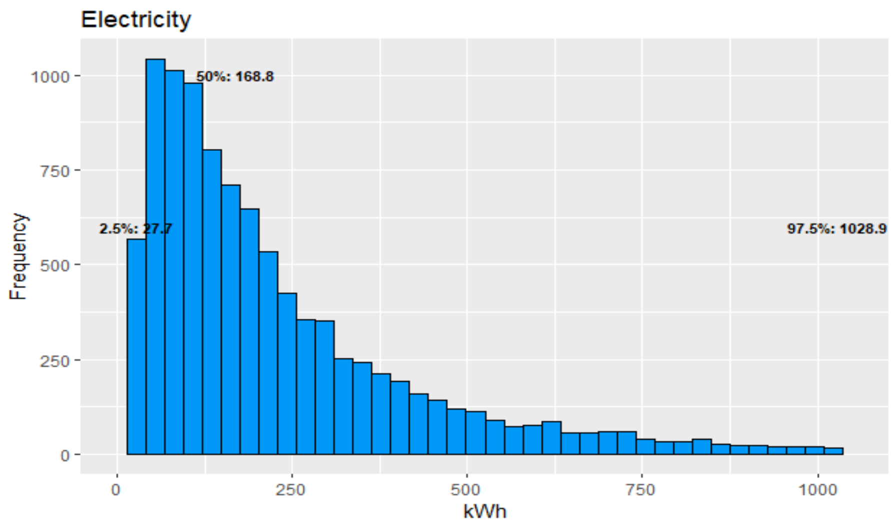

FIGURE 4.3
Histogram using the posterior predictive distribution of electricity demand

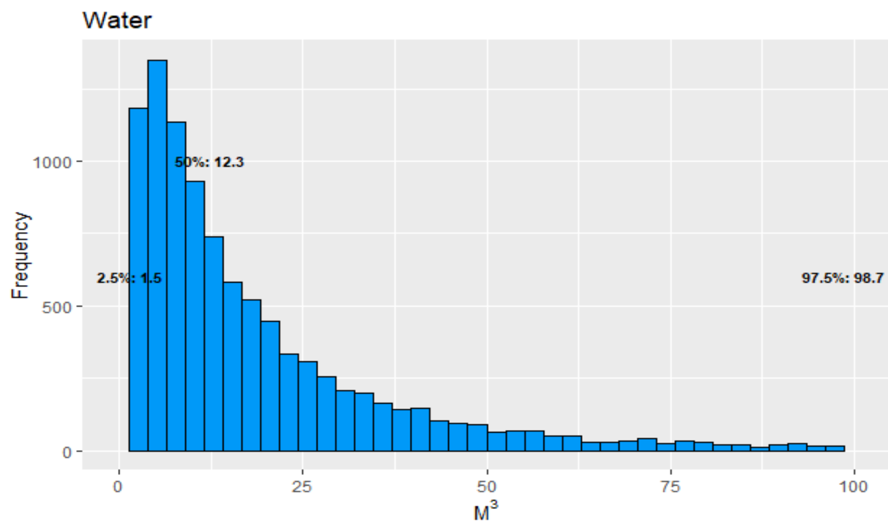

FIGURE 4.4
Histogram using the posterior predictive distribution of water demand

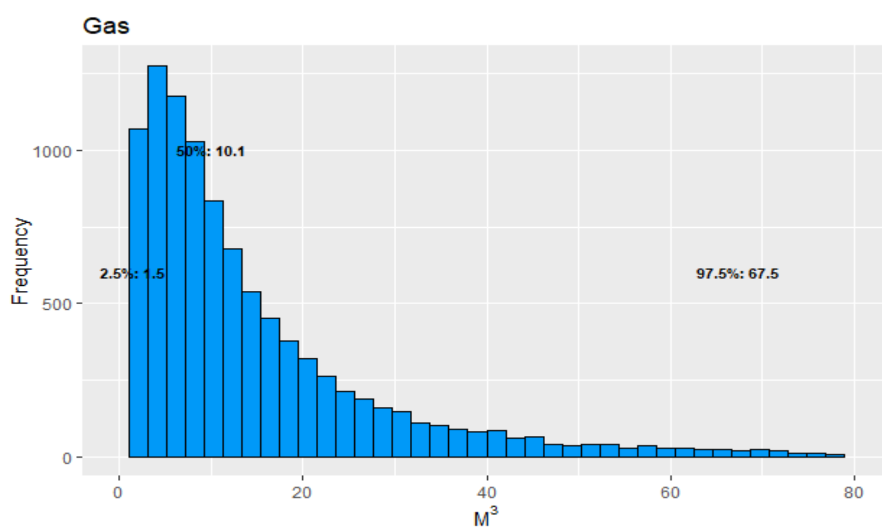

FIGURE 4.5
Histogram using the posterior predictive distribution of gas demand

5

Simulation methods

5.1 Solutions of Exercises

1. The inverse transform method ...

R code. Example: Math test

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

Part II

Regression models: A GUIded tour

6

Univariate models

6.1 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

$$
\begin{aligned}
\pi(\boldsymbol{\beta},\sigma^2|\mathbf{y},\mathbf{X}) &\propto p(\mathbf{y}|\boldsymbol{\beta},\sigma^2,\mathbf{X})\pi(\boldsymbol{\beta})\pi(\boldsymbol{\sigma^2})\\
&\propto (\sigma^2)^{-\frac{N}{2}}\exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^\top(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right\}\\
&\times\exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^\top\mathbf{B_0^{-1}}(\boldsymbol{\beta}-\boldsymbol{\beta_0})\right\}\\
&\times\frac{(\delta_0/2)^{(\alpha_0/2)}}{\Gamma(\alpha_0/2)}\frac{1}{(\sigma^2)^{(\alpha_0/2+1)}}\exp\left\{-\frac{\delta_0}{2\sigma^2}\right\}\\
&\propto\exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^\top\mathbf{B_0^{-1}}(\boldsymbol{\beta}-\boldsymbol{\beta_0})\right\}\\
&\times\frac{1}{(\sigma^2)^{(\alpha_0+N)/2+1}}\exp\left\{-\frac{\delta_0+(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^\top(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}{2\sigma^2}\right\}\\
&=\exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^\top\mathbf{B_0^{-1}}(\boldsymbol{\beta}-\boldsymbol{\beta_0})\right\}\times\underbrace{\frac{1}{(\sigma^2)^{(\frac{\alpha_n}{2}+1)}}\exp\left\{-\frac{\delta_n}{2\sigma^2}\right\}}_{1}.
\end{aligned}
$$

   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 $\boldsymbol{\beta}$,

$$\begin{aligned}
\pi(\boldsymbol{\beta},\sigma^2|\mathbf{y},\mathbf{X}) &\propto p(\mathbf{y}|\boldsymbol{\beta},\sigma^2,\mathbf{X})\pi(\boldsymbol{\beta})\pi(\sigma^2)\\
&\propto (\sigma^2)^{-\frac{N}{2}}\exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^{\top}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right\}\\
&\times\exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^{\top}\mathbf{B}_0^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\right\}\\
&\times\frac{(\delta_0/2)^{(\alpha_0/2)}}{\Gamma(\alpha_0/2)}\frac{1}{(\sigma^2)^{(\alpha_0/2+1)}}\exp\left\{-\frac{\delta_0}{2\sigma^2}\right\}\\
&=(\sigma^2)^{-\frac{N}{2}}\exp\left\{-\frac{\sigma^{-2}}{2}[\mathbf{y}^{\top}\mathbf{y}-\mathbf{y}^{\top}X\boldsymbol{\beta}-\boldsymbol{\beta}^{\top}\mathbf{X}^{\top}\mathbf{y}+\boldsymbol{\beta}^{\top}\mathbf{X}^{\top}\mathbf{X}\boldsymbol{\beta}]\right\}\\
&\times\exp\left\{-\frac{1}{2}[\boldsymbol{\beta}^{\top}\mathbf{B}_0^{-1}\boldsymbol{\beta}-\boldsymbol{\beta}^{\top}\mathbf{B}_0^{-1}\boldsymbol{\beta}_0-\boldsymbol{\beta}_0^{\top}\mathbf{B}_0^{-1}\boldsymbol{\beta}+\boldsymbol{\beta}_0^{\top}\mathbf{B}_0^{-1}\boldsymbol{\beta}_0]\right\}\\
&\times\frac{(\delta_0/2)^{(\alpha_0/2)}}{\Gamma(\alpha_0/2)}\frac{1}{(\sigma^2)^{(\alpha_0/2+1)}}\exp\left\{-\frac{\delta_0}{2\sigma^2}\right\}\\
&\propto\exp\left\{-\frac{1}{2}[\boldsymbol{\beta}^{\top}(\mathbf{B}_0^{-1}+\sigma^{-2}\mathbf{X}^{\top}\mathbf{X})\boldsymbol{\beta}-2\boldsymbol{\beta}^{\top}(\mathbf{B}_0^{-1}\boldsymbol{\beta}_0+\sigma^{-2}\mathbf{X}^{\top}\mathbf{X}\hat{\boldsymbol{\beta}})]\right\}\\
&\times\frac{1}{(\sigma^2)^{(\alpha_0+N)/2+1}}\exp\left\{-\frac{\delta_0+\mathbf{y}^{\top}\mathbf{y}}{2\sigma^2}\right\},
\end{aligned}$$

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

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

and completing the square

$$\begin{aligned}
\pi(\boldsymbol{\beta},\sigma^2|\mathbf{y},\mathbf{X}) &\propto\exp\left\{-\frac{1}{2}[\boldsymbol{\beta}^{\top}(\mathbf{B}_0^{-1}+\sigma^{-2}\mathbf{X}^{\top}\mathbf{X})\boldsymbol{\beta}-2\boldsymbol{\beta}^{\top}\mathbf{B}_n^{-1}\mathbf{B}_n(\mathbf{B}_0^{-1}\boldsymbol{\beta}_0+\sigma^{-2}\mathbf{X}^{\top}\mathbf{X}\hat{\boldsymbol{\beta}})\right.\\
&\left.+\boldsymbol{\beta}_n^{\top}\mathbf{B}_n^{-1}\boldsymbol{\beta}_n-\boldsymbol{\beta}_n^{\top}\mathbf{B}_n^{-1}\boldsymbol{\beta}_n]\right\}\times\frac{1}{(\sigma^2)^{(\alpha_0+N)/2+1}}\exp\left\{-\frac{\delta_0+\mathbf{y}^{\top}\mathbf{y}}{2\sigma^2}\right\}\\
&=\exp\left\{-\frac{1}{2}[\boldsymbol{\beta}^{\top}\mathbf{B}_n^{-1}\boldsymbol{\beta}-2\boldsymbol{\beta}^{\top}\mathbf{B}_n^{-1}\boldsymbol{\beta}_n+\boldsymbol{\beta}_n^{\top}\mathbf{B}_n^{-1}\boldsymbol{\beta}_n]\right\}\\
&\times\frac{1}{(\sigma^2)^{(\alpha_0+N)/2+1}}\exp\left\{-\frac{\delta_0+\mathbf{y}^{\top}\mathbf{y}-\sigma^2\boldsymbol{\beta}_n^{\top}\mathbf{B}_n^{-1}\boldsymbol{\beta}_n}{2\sigma^2}\right\}\\
&=\underbrace{\exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}_n)^{\top}\mathbf{B}_n^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}_n)\right\}}_{1}\\
&\times(\sigma^2)^{-(\frac{\alpha_n}{2}+1)}\exp\left\{-\frac{\delta^*}{2\sigma^2}\right\},
\end{aligned}$$

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 $\boldsymbol{\beta}|\sigma^2,\boldsymbol{\tau},\mathbf{y},\mathbf{X}\sim N(\boldsymbol{\beta}_n,\mathbf{B}_n)$, $\sigma^2|\boldsymbol{\beta},\boldsymbol{\tau},\mathbf{y},\mathbf{X}\sim IG(\alpha_n,\delta_n)$ and $\tau_i|\boldsymbol{\beta},\sigma^2,\mathbf{y},\mathbf{X}\sim G(v_{1n},v_{2in})$, where $\boldsymbol{\tau}=[\tau_1,\dots,\tau_n]^\top$, $\mathbf{B}_n=(\mathbf{B}_0^{-1}+\sigma^{-2}\mathbf{X}^\top\Psi\mathbf{X})^{-1}$, $\boldsymbol{\beta}_n=\mathbf{B}_n(\mathbf{B}_0^{-1}\boldsymbol{\beta}_0+\sigma^{-2}\mathbf{X}^\top\Psi\mathbf{y})$, $\alpha_n=\alpha_0+N$, $\delta_n=\delta_0+(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^\top\Psi(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})$, $v_{1n}=v+1$, $v_{2in}=v+\sigma^{-2}(y_i-\mathbf{x}_i^\top\boldsymbol{\beta})^2$, and $\Psi=\text{diagonal}\{\tau_i\}$.

Answer

The joint posterior distribution of the parameters is

$$\begin{aligned}
\pi(\boldsymbol{\beta},\sigma^2,\boldsymbol{\tau}|\mathbf{y},\mathbf{X}) &\propto p(\mathbf{y}|\boldsymbol{\beta},\sigma^2,\boldsymbol{\tau},\mathbf{X})\pi(\boldsymbol{\beta})\pi(\sigma^2)\prod_{i=1}^N\pi(\tau_i)\\
&\propto\left(\prod_{i=1}^N\tau_i^{1/2}\right)(\sigma^2)^{-\frac{N}{2}}\exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^\top\Psi(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right\}\\
&\times\exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^\top\mathbf{B_0^{-1}}(\boldsymbol{\beta}-\boldsymbol{\beta_0})\right\}\\
&\times\frac{1}{(\sigma^2)^{(\alpha_0/2+1)}}\exp\left\{-\frac{\delta_0}{2\sigma^2}\right\}\prod_{i=1}^N\tau_i^{v/2-1}\exp\left\{-v\tau_i/2\right\}.
\end{aligned}$$

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

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

This is the kernel of an inverse-gamma distribution with shape parameter α_n and rate parameter δ_n.

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

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

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)

```
rm(list = ls())
set.seed(010101)
########################## Linear regression: Value of soccer players ##########################
Data <- read.csv("DataApplications/1ValueFootballPlayers.csv", sep = ",", header = TRUE, fileEncoding = "latin1")
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 <- 0.001/2
a0 <- 0.001/2
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, v2n)
  return(taui)
}
```

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

```
PostBetas <- matrix(0, mcmc+burnin, k)
PostSigma2 <- rep(0, mcmc+burnin)
Beta <- rep(0, k)
tau <- rep(1, N)
for(s in 1:tot){
  sig2 <- PostSig2(Beta = Beta, tau = tau)
  PostSigma2[s] <- sig2
  Beta <- PostBeta(sig2 = sig2, tau = tau)
  PostBetas[s,] <- Beta
  tau <- sapply(1:N, function(i){PostTau(sig2 = sig2, Beta =
     Beta, i)})
}
keep <- seq((burnin+1), tot, thin)
PosteriorBetas <- PostBetas[keep,]
colnames(PosteriorBetas) <- c("Intercept", "Perf", "Age", "
    Age2", "NatTeam", "Goals", "Exp", "Exp2")
summary(coda::mcmc(PosteriorBetas))
PosteriorSigma2 <- PostSigma2[keep]
summary(coda::mcmc(PosteriorSigma2))
summary(coda::mcmc(exp(PosteriorBetas[,5])-1))
Iterations = 1:5000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 5000
1. Empirical mean and standard deviation for each variable,
plus standard error of the mean:
Mean             SD       Naive SE Time-series SE
1.246823       0.235881       0.003336       0.004010
2. Quantiles for each variable:
2.5%   25%   50%   75% 97.5%
0.815 1.086 1.238 1.395 1.744
```

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%, 174%). 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
set.seed(010101)
Data <- read.csv("DataApplications/2HealthMed.csv", sep = ",", header = TRUE, fileEncoding = "latin1")
attach(Data)
str(Data)
y <- Hosp # Dependent variables
X <- cbind(1, SHI, Female, Age, Age2, Est2, Est3, Fair, Good, Excellent) # Regressors
K <- dim(X)[2]
N <- dim(X)[1]
# Hyperparameters
b0 <- rep(0, K) # Prio mean
B0 <- diag(K) # Prior covariance
B0i <- solve(B0)
mcmc <- 1000; burnin <- 500; thin <- 2; tot <- mcmc + burnin; keep <- seq(burnin, tot, thin)
# Posterior distributions programming the Gibbs sampling
# Auxiliary parameters
XtX <- t(X)%*%X
# Gibbs sampling functions
PostBeta <- function(Yl){
  Bn <- solve(B0i + XtX)
  bn <- Bn%*%(B0i%*%b0 + t(X)%*%Yl)
  Beta <- MASS::mvrnorm(1, bn, Bn)
  return(Beta)
}
PostYl <- function(Beta, i){
  Ylmean <- X[i,]%*%Beta
  if(y[i] == 1){
    Yli <- truncnorm::rtruncnorm(1, a = 0, b = Inf, mean = Ylmean, sd = 1)
  }else{
    Yli <- truncnorm::rtruncnorm(1, a = -Inf, b = 0, mean = Ylmean, sd = 1)
  }
  return(Yli)
}
PostBetas <- matrix(0, mcmc+burnin, K)
Beta <- rep(0, K)
# create progress bar in case that you want to see iterations progress
pb <- winProgressBar(title = "progress bar", min = 0, max = tot, width = 300)
for(s in 1:tot){
  Yl <- sapply(1:N, function(i){PostYl(Beta = Beta, i)})
  Beta <- PostBeta(Yl = Yl)
  PostBetas[s,] <- Beta
  setWinProgressBar(pb, s, title=paste( round(s/tot*100, 0), "% done"))
}
close(pb)
keep <- seq((burnin+1), tot, thin)
PosteriorBetas <- PostBetas[keep,]
colnames(PosteriorBetas) <- c("Intercept", "SHI", "Female", "Age", "Age2", "Est2", "Est3", "Fair", "Good", "Excellent")
summary(coda::mcmc(PosteriorBetas))
```

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

   Run the Algorithm A3 of the book to show the results of the Geweke [2], Raftery [8] and Heidelberger [3] tests using our GUI.

   Answer

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

```
GewekeTestLocationCoef
Fraction in 1st window = 0.1
Fraction in 2nd window = 0.5
cte_1  cte_2  cte_3 NAS_1_1 NAS_1_2 NAS_1_3    AS_1   AS_2
-1.821 -0.714  0.792  2.275 -3.944 -2.071  1.627 -2.729

RafteryTestLocationCoef
Quantile (q) = 0.5
Accuracy (r) = +/- 0.025
Probability (s) = 0.95
Burn-in  Total   Lower bound  Dependence
(M)      (N)     (Nmin)       factor (I)
cte_1   780      365690  1537         238.0
cte_2   360      193950  1537         126.0
cte_3   660      340120  1537         221.0
NAS_1_1 120      70320   1537          45.8
NAS_1_2 475      243960  1537         159.0
NAS_1_3 440      248930  1537         162.0
AS_1    3010     1438135 1537         936.0
AS_2    550      297770  1537         194.0

HeidelTestLocationCoef
Stationarity start     p-value
test         iteration
cte_1   passed       6001      6.54e-01
cte_2   failed         NA      3.72e-02
cte_3   failed         NA      4.99e-02
NAS_1_1 failed         NA      4.77e-07
NAS_1_2 failed         NA      1.82e-05
NAS_1_3 failed         NA      1.19e-04
AS_1    passed       2001      3.71e-01
AS_2    passed       8001      4.48e-01
Halfwidth Mean     Halfwidth
test
cte_1   passed    -0.34236 0.017025
cte_2   <NA>            NA       NA
cte_3   <NA>            NA       NA
NAS_1_1 <NA>            NA       NA
NAS_1_2 <NA>            NA       NA
NAS_1_3 <NA>            NA       NA
AS_1    passed    -0.00708 0.000306
AS_2    passed     0.27982 0.009994
```

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

```
remove(list = ls())
set.seed(12345)
# Simulation of data
N<-1000  # Sample Size
B<-c(0.5,0.8,-3)
B1<-c(-2.5,-3.5,0)
B2<-c(1,1,0)
# Alternative specific attributes of choice 1, for instance, price, quality and duration of choice 1
X1<-matrix(cbind(rnorm(N,0,1),rnorm(N,0,1),rnorm(N,0,1)),N,length(B))
# Alternative specific attributes of choice 2, for instance, price, quality and duration of choice 2
X2<-matrix(cbind(rnorm(N,0,1),rnorm(N,0,1),rnorm(N,0,1)),N,length(B))
# Alternative specific attributes of choice 3, for instance, price, quality and duration of choice 3
X3<-matrix(cbind(rnorm(N,0,1),rnorm(N,0,1),rnorm(N,0,1)),N,length(B))
X4<-matrix(rnorm(N,1,1),N,1)
V1<-B2[1]+X1%*%B+B1[1]*X4
V2<-B2[2]+X2%*%B+B1[2]*X4
V3<-B2[3]+X3%*%B+B1[3]*X4
suma<-exp(V1)+exp(V2)+exp(V3)
p1<-exp(V1)/suma
p2<-exp(V2)/suma
p3<-exp(V3)/suma
p<-cbind(p1,p2,p3)
y<- apply(p,1, function(x)sample(1:3, 1, prob = x, replace = TRUE))
table(y)
L <- length(table(y))
dat <-data.frame(mode,X1[,1],X2[,1],X3[,1],X1[,2],X2[,2],X3[,2],X1[,3],X2[,3],X3[,3],X4)
colnames(dat) <- c("mode","V1.1","V1.2","V1.3","V2.1","V2.2","V2.3","V3.1","V3.2","V3.3","V4")
attach(dat)
LongData <- mlogit::mlogit.data(dat, shape = "wide", varying=2:10, choice = "mode")
Xa <- cbind(LongData$V1, LongData$V2, LongData$V3)
Xa <- cbind(X1[,1],X2[,1],X3[,1],X1[,2],X2[,2],X3[,2],X1[,3],X2[,3],X3[,3])
na <- 3
Xd <- X4
X <- bayesm::createX(p = L, na = na, nd = 1, Xa = Xa, Xd = Xd, base = L)
DataMlogit <- list(y=y, X = X, p = L)
# MCMC parameters
mcmc <- 11000+1
thin <- 5
df <- 6
mcmcpar <- list(R = mcmc, keep = 5, nu = df)
PostBeta <- bayesm::rmnlIndepMetrop(Data = DataMlogit, Mcmc = mcmcpar)
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
rm(list = ls()); set.seed(010101); N <- 1000
x1 <- rnorm(N, 6, 5); x2 <- rgamma(N, shape = 1, scale = 1)
X <- cbind(1, x1, x2)
beta <- c(0.5, -0.25, 0.5); cutoffs <- c(0, 1, 2.5)
e <- rnorm(N,0,1)
y_latent <- X%*%beta + e; y <- rep(0,N)
for (i in 1:N) {
  if (y_latent[i] < cutoffs[1]){
    y[i] <- 0}else{
    if (y_latent[i] >= cutoffs[1] & y_latent[i] < cutoffs[2]) {
      y[i] <- 1
    }else{
      if (y_latent[i] >= cutoffs[2] & y_latent[i] < cutoffs[3]) {
        y[i] <- 2
      }else{y[i] <- 3
      }
    }
  }
}
# Likelihood function
LogLikOP <- function(param){
  beta_g <- param[1:ncol(X)]
  delta <- param[(ncol(X)+1):(ncol(X) + dplyr::n_distinct(y) - 1)]
  Xbeta <- X%*%beta_g
  logLik <- 0
  for (i in 1:length(y)){
    if (y[i]==0){logLiki <- log(pnorm(-Xbeta[i]))
    }else if (y[i]==1){
      logLiki <- log(pnorm(exp(delta[1]) - Xbeta[i]) - pnorm(-Xbeta[i]))
    }else if (y[i]==2){
      logLiki <- log(pnorm(exp(delta[2]) + exp(delta[1]) - Xbeta[i]) - pnorm(exp(delta[1]) - Xbeta[i]))
    }else {logLiki <- log(1 - pnorm(exp(delta[2]) + exp(delta[1]) - Xbeta[i]))
    }
    logLik <- logLik + logLiki
  }
  return(-logLik)
}
# ML Estimation
param0 <- rep(0, ncol(X) + n_distinct(y)-2)
mle <- optim(param0, LogLikOP, hessian = T, method = "BFGS")
mle$par
exp(mle$par[length(beta)+1])
exp(mle$par[length(beta)+1])+exp(mle$par[length(beta)+2])
CovarML <- solve(mle$hessian)
```

R. code. Simulation of the ordered probit model

```
# M-H within Gibbs
mhop <- function(param0, G){
  betasamples <- matrix(c(0), nrow = G, ncol = ncol(X))
  betasamples[1,] <- param0[1:ncol(X)]
  tau <- matrix(c(0), nrow = G, ncol = dplyr::n_distinct(y) - 2)
  tau[1,] <- param0[(ncol(X)+1):(ncol(X) + dplyr::n_distinct(y) - 2)]
  yl <- rep(0,length(y)); ar <- rep(0,G); B1 <- solve(t(X)%*%X+solve(B0))
  pb <- winProgressBar(title = "progress bar", min = 0, max = G, width = 300)
  for(g in 2:G){
    bg <- betasamples[g-1,]; tg <- tau[g-1,]
    #Random walk M-H for delta
    delta_prime <- tg + mvtnorm::rmvnorm(1, mean = rep(0,2), sigma = VarProp)
    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))
    if(is.nan(alpha) | is.na(alpha)) {
      alpha <- 0
    }
    #Acceptance step
    u <- runif(1, min = 0, max = 1)
    if(u<=alpha){tau[g,] <- delta_prime; ar[g] <- 1
    }else{tau[g,] <- tg
    }
    #Generation of latent variables
    for (i in 1:length(y)){
      if (y[i]==0) {
        yl[i] <- EnvStats::rnormTrunc(1, mean = X[i,]%*%bg, sd = 1, max = 0)
      }else if(y[i]==1){
        yl[i] <- EnvStats::rnormTrunc(1, mean = X[i,]%*%bg, sd = 1, min = 0, max = exp(tau[g,1]))
      }else if(y[i]==2){
        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]))
      }else{
        yl[i] <- EnvStats::rnormTrunc(1, mean = X[i,]%*%bg, sd = 1, min = exp(tau[g,2])+exp(tau[g,1]))
      }
    }
    #Gibbs sampling for beta
    if(sum(is.nan(yl))>0 | sum(is.na(yl))>0 | sum(yl)==Inf){
      betasamples[g,] <- betasamples[g-1,]
    }else{
      b1 <- B1%*%(t(X)%*%yl + solve(B0)%*%b0)
      betasamples[g,] <- mvrnorm(1, mu = b1, Sigma = B1)
    }
    setWinProgressBar(pb, g, title=paste( round(g/G*100, 0), "% done"))
  }
  close(pb)
  return(cbind(betasamples, tau, ar))
}
```

R. code. Simulation of the ordered probit model

```
#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

```
Summary
Iterations = 10001:60000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 50000
1. Empirical mean and standard deviation for each variable,
plus standard error of the mean:
               Mean        SD  Naive SE Time-series SE
(Intercept)  1.014765 2.626395 1.175e-02      1.659e-02
Perf         0.033935 0.004529 2.025e-05      2.245e-05
Age          1.027285 0.212413 9.499e-04      1.326e-03
Age2        -0.021914 0.003989 1.784e-05      2.528e-05
NatTeam      0.847412 0.124884 5.585e-04      6.423e-04
Goals        0.010092 0.001644 7.351e-06      7.712e-06
Exp          0.175324 0.069653 3.115e-04      3.727e-04
Exp2        -0.005696 0.002950 1.319e-05      1.527e-05
sigma2       0.983337 0.096194 4.302e-04      6.724e-04
2. Quantiles for each variable:
             2.5%       25%       50%       75%      97.5%
(Intercept) -4.212563 -0.732856  1.038281  2.796603  6.051e
    +00
Perf         0.025121  0.030873  0.033896  0.036988  4.286e
    -02
Age          0.618743  0.883011  1.026760  1.168860  1.450e
    +00
Age2        -0.029854 -0.024580 -0.021886 -0.019203 -1.424e
    -02
NatTeam      0.605030  0.763154  0.846671  0.930593  1.095e
    +00
Goals        0.006893  0.008987  0.010097  0.011201  1.330e
    -02
Exp          0.038522  0.128032  0.175261  0.222110  3.118e
    -01
Exp2        -0.011484 -0.007671 -0.005685 -0.003706  5.502e
    -05
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
rm(list = ls()); set.seed(010101)
Data <- read.csv("DataApplications/1ValueFootballPlayers.csv", sep = ",", header = TRUE, fileEncoding = "latin1")
attach(Data)
y <- log(Value); X <- cbind(1, Perf, Age, Age2, NatTeam, Goals, Exp, Exp2)
RegLS <- lm(y ~ X -1)
k <- dim(X)[2]; N <- dim(X)[1]
# Hyperparameters
b0 <- rep(0, k); c0 <- 1000
B0 <- c0*diag(k); B0i <- solve(B0)
# MCMC parameters
mcmc <- 5000; burnin <- 1000
tot <- mcmc + burnin; thin <- 1
# Quantile
tau <- 0.5; theta <- (1-2*tau)/(tau*(1-tau))
psi2 <- 2/(tau*(1-tau)); an2 <- 2+theta^2/psi2
# Gibbs sampler
PostBeta <- function(e){
  Bn <- solve(B0i + psi2^(-1)*t(X)%*%diag(1/e)%*%X)
  bn <- Bn%*%(B0i%*%b0 + psi2^(-1)*t(X)%*%diag(1/e)%*%(y-theta*e))
  Beta <- MASS::mvrnorm(1, bn, Bn)
  return(Beta)
}
PostE <- function(Beta, i){
  dn2 <-(y[i]-X[i,]%*%Beta)^2/psi2
  ei <- GIGrvg::rgig(1, chi = dn2, psi = an2, lambda = 1/2)
  return(ei)
}
PostBetas <- matrix(0, mcmc+burnin, k)
Beta <- RegLS$coefficients
pb <- winProgressBar(title = "progress bar", min = 0, max = tot, width = 300)
for(s in 1:tot){
  e <- sapply(1:N, function(i){PostE(Beta = Beta, i)})
  Beta <- PostBeta(e = e)
  PostBetas[s,] <- Beta
  setWinProgressBar(pb, s, title=paste( round(s/tot*100, 0), "% done"))
}
close(pb)
keep <- seq((burnin+1), tot, thin)
PosteriorBetas <- PostBetas[keep,]
colnames(PosteriorBetas) <- c("Intercept", "Perf", "Age", "Age2", "NatTeam", "Goals", "Exp", "Exp2")
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

```
########################## 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)
```

7

Multivariate models

7.1 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:

$$\log(\text{pcGDP95}_i) = \beta_1 + \beta_2\text{PAER}_i + \beta_3\text{Africa} + \beta_4\text{Asia} + \beta_5\text{Other} + u_{1i}, \tag{7.1}$$

$$\text{PAER}_i = \alpha_1 + \alpha_2 \log(\text{pcGDP95}_i) + \alpha_3 \log(\text{Mort}_i) + u_{2i}, \tag{7.2}$$

   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 7.2 into Equation 7.1, and solving for log(*pcGDP95*),

$$\log(\text{pcGDP95}_i) = \pi_1 + \pi_2 \log(\text{Mort}_i) + \pi_3\text{Africa} + \pi_4\text{Asia} + \pi_5\text{Other} + e_{1i}, \tag{7.3}$$

   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 7.3 into Equation 7.2, and solving for *PAER*,

$$\text{PAER}_i = \gamma_1 + \gamma_2 \log(\text{Mort}_i) + \gamma_3\text{Africa} + \gamma_4\text{Asia} + \gamma_5\text{Other} + e_{2i}, \tag{7.4}$$

   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

```
rm(list = ls())
set.seed(010101)
DataInst <- read.csv("DataApplications/4Institutions.csv", sep = ",", header = TRUE, fileEncoding = "latin1")
attach(DataInst)
Y <- cbind(logpcGDP95, PAER)
X <- cbind(1, logMort, Africa, Asia, Other)
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 #Number of 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))
```

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 7.1
Identification matrix.

q	p	constant	er	y	pc	ps
1	-β_2	-β_1	0	-β_3	-β_4	-β_5
1	-α_2	$-\alpha_1$	-α_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))
```

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

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

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

[5] 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.

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

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

[8] 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.

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

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