Bayesian Methods Lab 02

Rakhshanda Jabeen

11/04/2020

Contents

1. Linear and Polynomial Regression

The considered “TempLinkoping” dataset contains daily temperatures (in Celcius degrees) at Malmslätt, Linköping over the course of the year 2016 (366 days since 2016 was a leap year). The response variable is temp and the covariate is,

$$time = \frac{\text{the number of days since beginning of year}}{366}$$

The task is to perform a Bayesian analysis of a quadratic regression

$$temp = \beta_0 + \beta_1\ .time + \beta_2\ .time^2 + \epsilon, \qquad \epsilon \sim \mathcal{N}(0, \sigma^2)$$

1.1 Prior Distribution of the Model Parameters

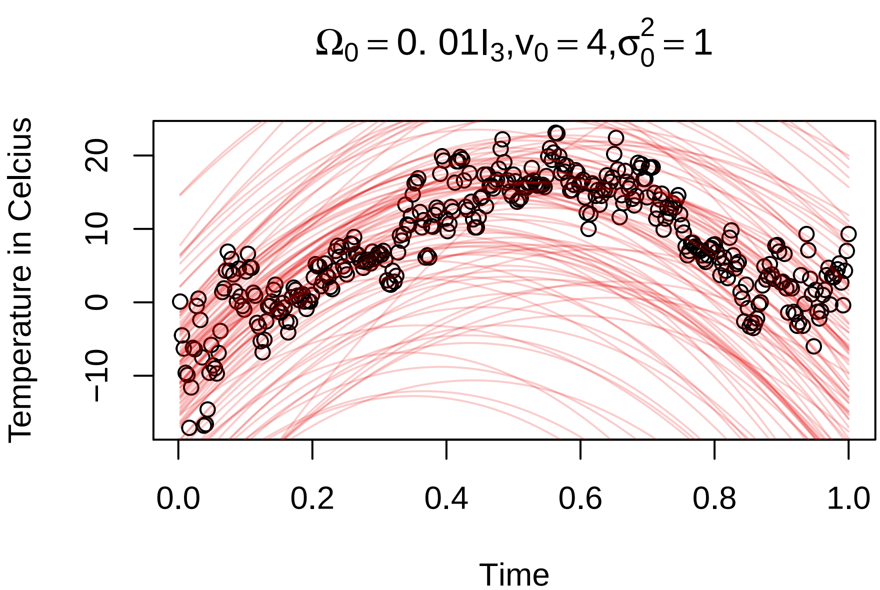
In this task we are using tyhe conjugate prior for the linear regression model. We are supposed to set the most reasonable values of the hyperparameters $\mu_0,\ v_0,\ \sigma_0^2,$ and Ω_0.

The joint prior distribution has the following form,

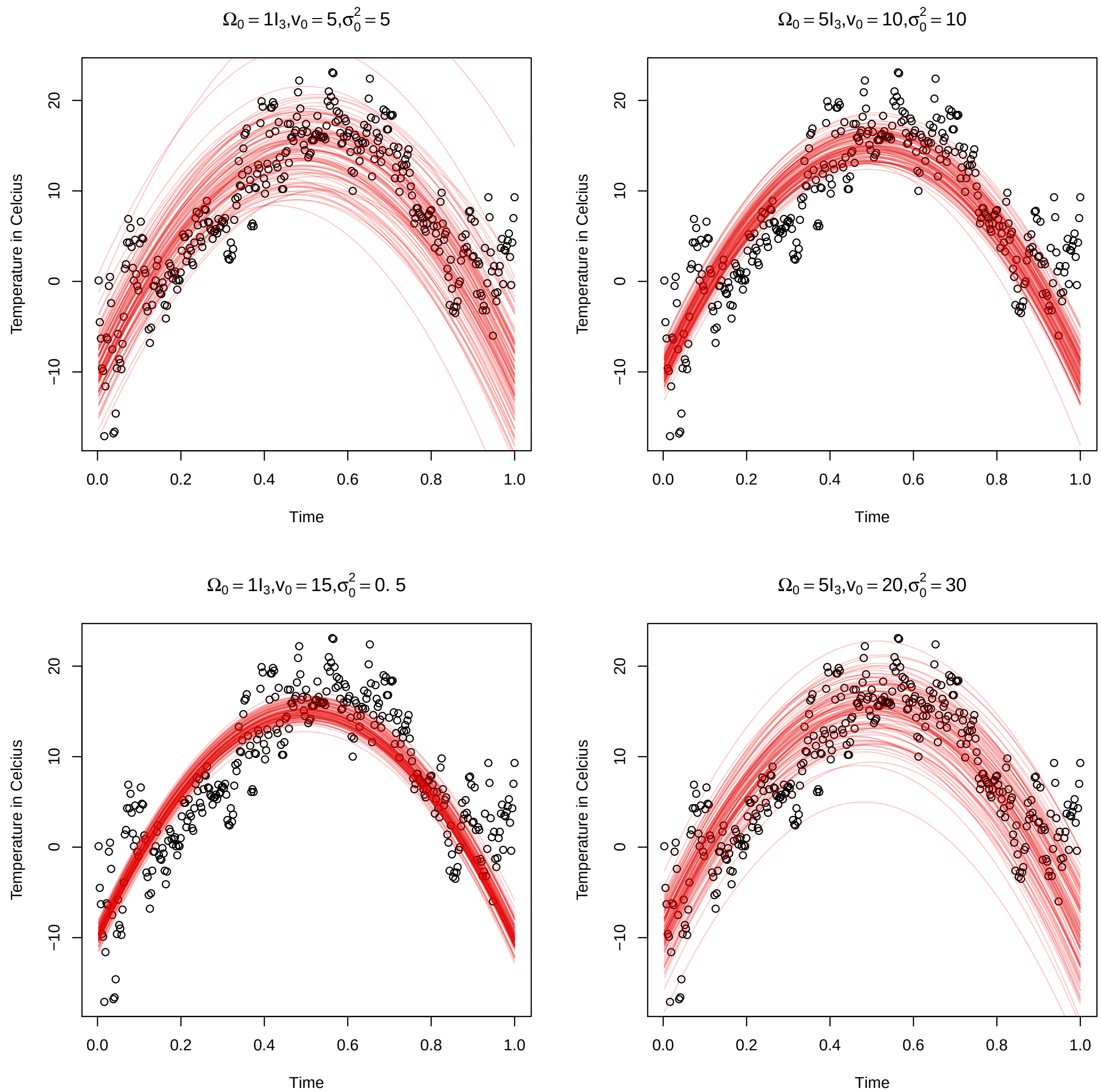
$$p(\beta, \sigma^2) = p(\beta|\sigma^2)\ p(\sigma^2)$$
$$\beta|\sigma^2 \sim \mathcal{N}(\mu_0, \sigma^2\Omega_0^{-1})$$
$$\sigma^2 \sim Inv - \chi^2(v_0, \sigma_0)$$

Simmulating from joint prior with the follwing hyperparameters:

- $\mu_0 = (-10, 100, -100)^T$
- $\Omega_0 = 0.01 * I_3$
- $v_0 = 4\ \&\ \sigma_0^2 = 1$



We can clearly see in the plot that all of the regression curves are way neyond the actual data.



By Looking at the plots we select the following values of the hyperparameters:

- $\mu_0 = (-10, 100, -100)^T$
- $\Omega_0 = 1 * I_3$
- $v_0 = 10$
- $\sigma_0^2 = 5$

1.2. Joint Posterior Distribution of $\beta_0, \beta_1, \beta_2$ and σ^2

In this task we are simmulationg random numbers from joint and marginals posterior distributions of $\beta_0, \beta_1, \beta_2$ and σ^2.

1.2.1 Marginal Posterior Distrbutions

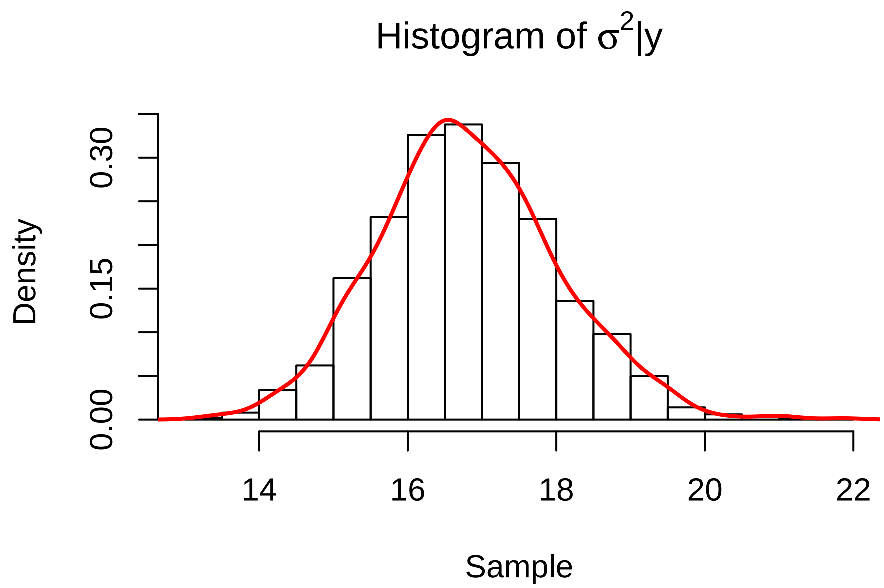
Marginal posterior distribution of σ^2

The marginal posterior distribution of σ^2 can be written as:

$$
\begin{aligned}
p(\sigma^2|y) &= \frac{p(\beta, \sigma^2|y)}{p(\beta|\sigma^2, y)} \\
&\sim Inv - \chi^2(n-k, s^2)
\end{aligned}
$$

Where,

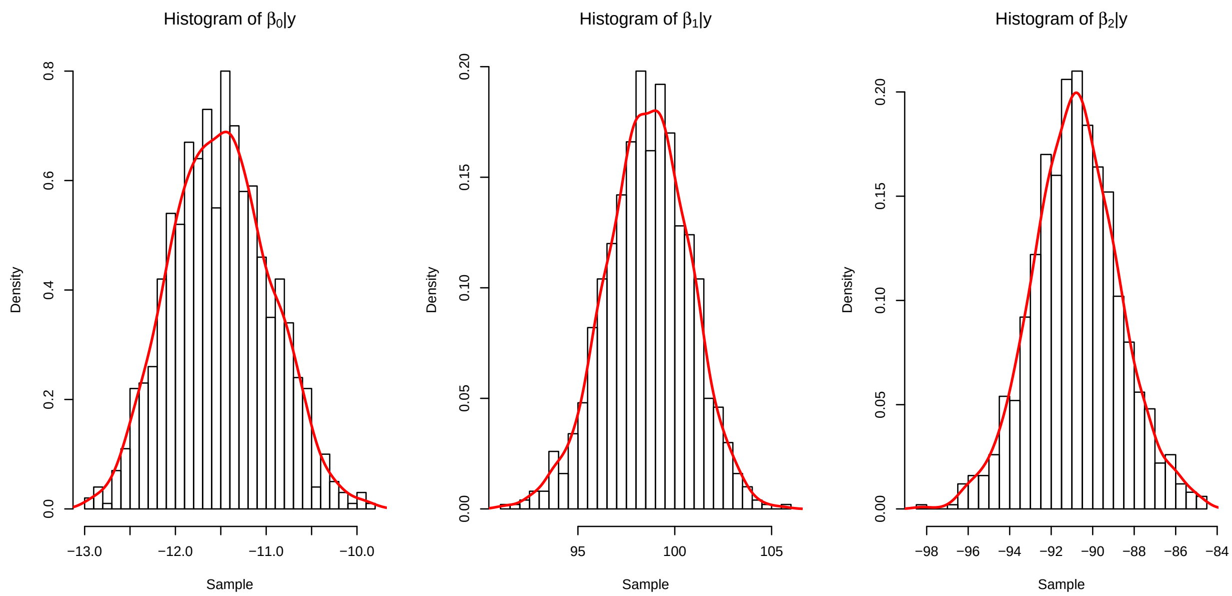
$$
\begin{aligned}
s^2 &= \frac{1}{n-k}\left(y - X\hat{\beta}\right)^T\left(y - X\hat{\beta}\right) \\
\hat{\beta} &= (X^T X)^{-1} X^T y
\end{aligned}
$$



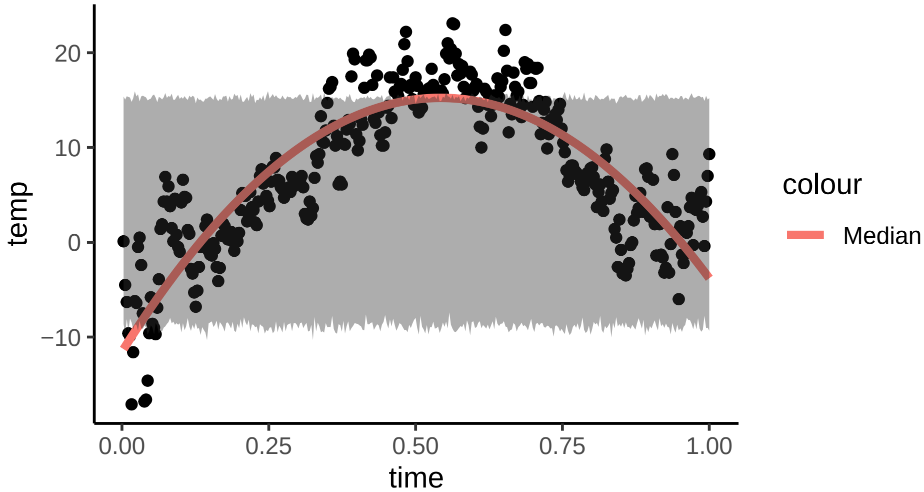
Marginal Posterior Distribution of $\beta_0, \beta_1, \beta_2$

The marginal posterior distribution of β is defnided as:

$$
\beta|y \sim t_{n-k}[\hat{\beta}, s^2(X^T X)^{-1}]
$$



1.2.2 Posterior Median and 95% Credible Interval



We can see in the plot that the most of the data points lies in the 95% equal tail confidence interval band.

1.3. Highest Expected Temperature

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon, \qquad \epsilon \sim \mathcal{N}(0, \sigma^2)$$

Differentiation with respect to x.

$$f'(x) = \beta_1 + 2\beta_2 \; x$$

$$\implies \tilde{x} = \frac{-\beta_1}{\beta_2} \qquad \qquad \big(\text{Inflexion Point}\big)$$

Second Order Dervative Test,

$$f''(x)|_{x=\tilde{x}} = 2\beta_2 < 0$$

This show that the function has a maximum value at $\frac{-\beta_1}{\beta_2}$. Thus the highest expected time in Malmslätt (2016) is at time 0.543311 i.e. 16th July 2016.

1.4. Prior for a Polynomial Model of Order 7

As we know that polynomial are too global thus in order to avoid overfitting we can do spline regression instead of polynomial regression. As we know that $\Omega_0 = \lambda I$ determines the smoothness/shrinkage factor. Thus by increasing the value of λ we can get a smoother fit which can immitate a higher degree polynomial.We also know that as

$$\lambda \to \infty, \qquad \tilde{\beta} \to 0$$

Thus in order to eliminate higher order parameters we can change the value of Ω_0 by increasing λ and give μ_0 values closer to zero.

2. Posterior Approximation for Classification with Logistic Regression

2.1. Logistic Regression Model

In this task we are performing logistic regression on the considered data set using glm() R function.

```
glmModel = glm(Work ~ 0 + ., data = data, family = binomial)
summary(glmModel)
```

```
Call:
glm(formula = Work ~ 0 + ., family = binomial, data = data)

Deviance Residuals:
   Min       1Q   Median       3Q      Max
-2.166   -0.930    0.439    0.949    2.058

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
Constant      0.6443     1.5231    0.42  0.67227
HusbandInc   -0.0198     0.0159   -1.24  0.21375
EducYears     0.1799     0.0791    2.27  0.02302 *
ExpYears      0.1675     0.0660    2.54  0.01114 *
ExpYears2    -0.1444     0.2359   -0.61  0.54049
Age          -0.0823     0.0270   -3.05  0.00228 **
NSmallChild  -1.3625     0.3900   -3.49  0.00048 ***
NBigChild    -0.0254     0.1417   -0.18  0.85759
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 277.26  on 200  degrees of freedom
Residual deviance: 222.73  on 192  degrees of freedom
AIC: 238.7

Number of Fisher Scoring iterations: 4
```

Table 1: Confusion Matrix

	0	1
0	66	31
1	27	76

The table above represents the confusion matrix of the fitted glm() model. The error rate of the regression model is 29%.

2.2. Posterior Distribution of vector β

The 8-dim parameter vectoe β has the following distribution:

$$\beta|y, X \sim \mathcal{N}\Big(\tilde{\beta}, J_y^{-1}(\tilde{\beta})\Big)$$

where $\tilde{\beta}$ is the posterior mode and $J(\tilde{\beta}) = -\frac{\partial^2 \ln\ p(\beta|y)}{\partial\beta\ \partial\beta^T}|_{\beta=\tilde{\beta}}$ is the negative of obsrved Hessian evaluated at the posterior mode.

In this task we are using the prior $\beta \sim \mathcal{N}(0, \tau^2 I)$, with $\tau = 10$.

2.2.1. Numerical Values for $\tilde{\beta}$:

Constant	HusbandInc	EducYears	ExpYears	ExpYears2	Age	NSmallChild	NBigChild
0.626729	-0.019791	0.180219	0.167567	-0.144597	-0.082066	-1.35913	-0.024684

2.2.2. The Matrix $J_y^{-1}(\tilde{\beta})$:

Table 2: Inverse of Hessian Matrix

2.266023	0.003339	-0.065451	-0.011791	0.045781	-0.030293	-0.188748	-0.098024
0.003339	0.000253	-0.000561	-0.000031	0.000141	-0.000036	0.000507	-0.000144
-0.065451	-0.000561	0.006218	-0.000356	0.001896	-0.000003	-0.006135	0.001753
-0.011791	-0.000031	-0.000356	0.004352	-0.014249	-0.000134	-0.001469	0.000544
0.045781	0.000141	0.001896	-0.014249	0.055579	-0.000330	0.003208	0.000512
-0.030293	-0.000036	-0.000003	-0.000134	-0.000330	0.000718	0.005184	0.001095
-0.188748	0.000507	-0.006135	-0.001469	0.003208	0.005184	0.151262	0.006769
-0.098024	-0.000144	0.001753	0.000544	0.000512	0.001095	0.006769	0.019972

2.2.3. Approximate 95% Credible Interval for the Variable NSmallChild

The 95% credible interval of the variable NSmallChild variable is as follows:

```
       2.5%       97.5%
-2.13072129 -0.59850961
```

As the expected value of regression coefficient of NsmallChild is -1.35913 whose absolute value is greater than 1 and also it is greater than all other regression coefficients thus we can infer that this feature is an important determinant of the probability that a woman works or not. The negative sign indicates that greater the value of this feature means that there is a higher probability that a woman does not work.

2.3. Prediction of the Response Variable

In this task we are predicting response variable “work” for the following data:

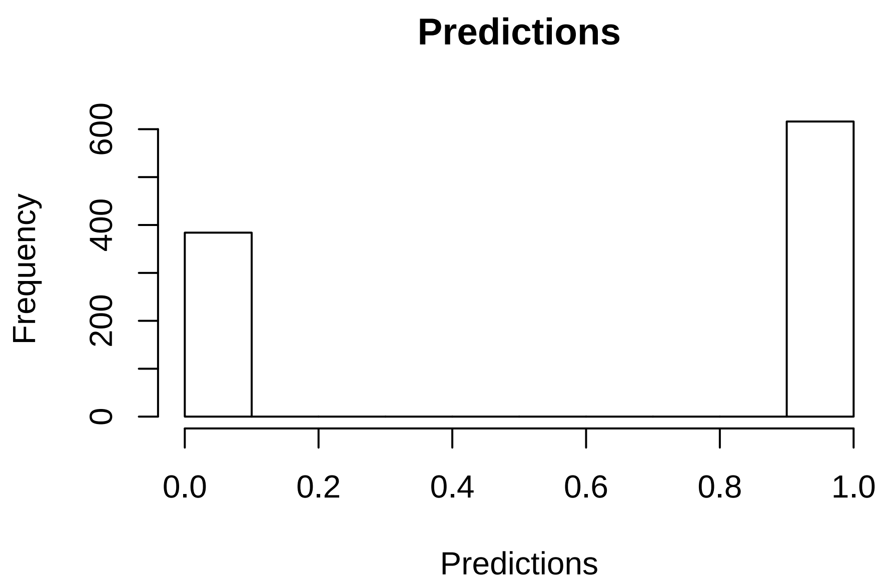
Table 3: Data

Constant	HusbandInc	EducYears	ExpYears	ExpYears2	Age	NSmallChild	NBigChild
1	10	8	10	1	40	1	1

We simmulated 1000 random draws from the posterior using 2.2.The predicted values of these random numbers is shown in the following table:

Table 4: Predictions

Var1	Freq
0	355
1	645



According to these predictions we can say that there is higher probability that a woman with these attributes is a working woman.

Appendix

```r
knitr::opts_chunk$set(echo = TRUE)
setwd("E:/Bayesian Learning/LABS/lab_02")
library(mvtnorm)
library(readr)
library(ggplot2)
library(knitr)
library(shape)
library(latex2exp)
library(kableExtra)
options("scipen"=100, "digits"=6)
data = read.table(file = "TempLinkoping.txt", header = T)
Y = as.matrix(data$temp)
X = matrix(c(data$time^0, data$time, data$time^2), byrow = FALSE, ncol = 3)
n = nrow(data)
k = ncol(X)
#________(1.1)_________#
rInvchisq <- function(n, df, scale) (df*scale)/rchisq(n,df=df)
n = nrow(data)
mu = c(-10,100,-100)
prior = function(v0 = 4, mu0, sigma20 = 1, omega0= 0.01, size=100)
{
  n = nrow(data)
  reg_curve = matrix(nrow = n, ncol = size)
  for (i in 1:size) {
    omega0 = diag(3)*omega0
    sigma2 = rInvchisq(n = 1,df = v0,scale = sigma20)
    beta = rmvnorm(1, mean = mu0, sigma = sigma2*solve(omega0))
    reg_curve[,i] = X %*% t(beta)
  }
  plot(data$time, Y, xlab="Time", ylab="Temperature in Celcius",
       main = TeX(paste0("$\\Omega_0 \ = \ $",omega0, "$\\I_3, \\v_0 \ = \ $",v0,
                         "$, \\sigma_0^2 \ = \ $",sigma20)))
  for (i in 1:size) {
    lines(data$time,reg_curve[,i],type = "l", col=rgb(0.9,0,0,0.2))
  }
}
mu = c(-10,100,-100)
prior(mu0 = mu)
par(mfrow = c(2,2))
prior(mu0 = mu, omega0 = 1, v0 = 5,sigma20 = 5)
prior(mu0 = mu, omega0 = 5, v0 = 10,sigma20 = 10)
prior(mu0 = mu, omega0 = 1, v0 = 15,sigma20 = 0.5)
prior(mu0 = mu, omega0 = 5, v0 = 20,sigma20 = 30)
#_____Joint posterior distribution______#
mu0 = matrix(c(-10,100,-100), nrow = 3, ncol = 1)
omega0 = diag(3)*1
v0 = 10
sigma20 = 5
#_____Joint posterior distribution______#
BayesLinReg = function(y, X, mu0, Omega0, v0, sigma20, nIter){
  n = nrow(X)
```

```r
  k = ncol(X)
  omega0 = diag(k)*omega0
  beta_hat = solve(t(X)%*%X) %*% t(X)%*%Y
  mu_n = solve(t(X)%*%X + omega0) %*% (t(X)%*%X %*% beta_hat+omega0%*%mu0)
  vn = v0+n
  omega_n = t(X)%*%X + omega0
  sigma2_n = (v0*sigma20)+ t(Y)%*%Y + t(mu0)%*%omega0%*%mu0 - t(mu_n)%*%omega_n%*%mu_n

  variance = c()
  betas = matrix(nrow = nIter, ncol = ncol(X))
  for (i in 1:nIter) {
    sigma2 = rInvchisq(n = 1,df = vn,scale = sigma2_n/vn)
    variance[i] = sigma2
    beta = rmvnorm(1, mean = mu_n, sigma = c(sigma2)*solve(omega_n))
    betas[i,] = beta
  }
  return(cbind(betas, variance))
}
posterior = BayesLinReg(Y,X, mu0, Omega0,v0, sigma20, nIter = 1000)
var_marginal = posterior[,4]
beta_marginal = posterior[,1:3]
set.seed(12345)
#___Marginal Distribution of sigma2__#
hist(var_marginal, breaks = 30,probability = T,xlab = 'Sample',
     main = TeX(paste0("Histogram of $\\sigma^2|y $")))
lines(density(var_marginal), col = 'red', lwd = 2)
#___Marginal Distribution of betas__#
par(mfrow = c(1,3))
hist(beta_marginal[,1], breaks = 30,probability = T,xlab = 'Sample',
     main = TeX(paste0("Histogram of $\\beta_0|y $")))
lines(density(beta_marginal[,1]), col = 'red', lwd = 2)

hist(beta_marginal[,2], breaks = 30,probability = T,xlab = 'Sample',
     main = TeX(paste0("Histogram of $\\beta_1|y $")))
lines(density(beta_marginal[,2]), col = 'red', lwd = 2)

hist(beta_marginal[,3], breaks = 30,probability = T,xlab = 'Sample',
     main = TeX(paste0("Histogram of $\\beta_2|y $")))
lines(density(beta_marginal[,3]), col = 'red', lwd = 2)
#___ Posterior Median and 95% Credible Interval___#
posterior = BayesLinReg(y = Y, X, mu0,Omega0, v0, sigma20, nIter = nrow(data))
Median = X%*%c(apply(posterior[,1:3],2,median))
lower = c()
upper = c()
for (i in 1:nrow(posterior)) {
  reg_cuve = X%*%posterior[i,1:3]
  lower[i] = quantile(reg_cuve,0.025)
  upper[i] = quantile(reg_cuve,0.975)
}
df = cbind(data, Median,lower,upper)

ggplot(df)+ geom_point(mapping = aes(x = time, y = temp))+theme_classic()+
  geom_line(aes(x = time, y = Median, col = 'Median'), size = 1.5)+
```

```r
  geom_ribbon(mapping = aes(x= time, ymin= lower, ymax =upper),
              alpha = 0.4)
#____(c. Highest Temperature)_____#
maxima = mean(-posterior[,2]/(2*posterior[,3]))
data = read.table(file = "WomenWork.dat", header = T)
#__________(a)________#
glmModel = glm(Work ~ 0 + ., data = data, family = binomial)
summary(glmModel)
fit = predict(glmModel,newdata = data, type = 'response')
y_hat = ifelse(fit > 0.5, 1L, 0L)
conf_mat = table(data$Work, y_hat)
n = nrow(data)
error_rate = (1-sum(diag(conf_mat))/n)*100
kable(conf_mat,"latex", caption = "Confusion Matrix", booktabs = T) %>%
  kable_styling(latex_options = "HOLD_position")
#________(b)__________#
y = as.vector(data$Work)
X = as.matrix(data[,2:9])
nPara = ncol(X)
#___Setting up Log-prior___#
mu = as.vector(rep(0,nPara))
tau = 10
Sigma = tau^2*diag(nPara)
mu0 = rep(0,ncol(X))
#___Log Posterior Logistic____#
logistic_postrior = function(beta,y,X,mu,sigma){
  nPara = length(beta)
  linPred = X%*%beta
  log_lik = sum( linPred*y -log(1 + exp(linPred)))
  if (abs(log_lik) == Inf) log_lik = -20000
  log_prior = dmvnorm(beta, matrix(0,nPara,1), Sigma, log=TRUE)
  return(log_lik + log_prior)
}
Optimal = optim(rep(0,8),logistic_postrior,gr=NULL,y,X,mu,sigma,method=c("BFGS"), control=list(fnscale=
Betas = matrix( Optimal$par,ncol = 8, nrow = 1)
kable(Betas, "latex", booktabs = T,col.names = colnames(X)) %>%
kable_styling(latex_options = "HOLD_position",bootstrap_options = "striped")
options("scipen"=100, "digits"=8)
J_inverse = round(solve(-1*Optimal$hessian),6)
kable(J_inverse, "latex",caption ='Inverse of Hessian Matrix', booktabs = T) %>%
kable_styling(latex_options = "HOLD_position")
#____Aprroximation of 95% CI for NSmallChild____#
mu = Optimal$par
Sigma = solve(-1*Optimal$hessian)
set.seed(12345678)
beta_7 = rnorm(n = 10000,mean = mu[7],sd = sqrt(Sigma[7,7]))
ci = quantile(beta_7,probs = c(0.025,0.975))
ci
#___Prediction of the Response Variable___#
options("scipen"=100, "digits"=5)
work_data = matrix(c(1,10,8,10,(10/10)^2, 40, 1,1), nrow = 1)
colnames(work_data) = colnames(X)
kable(work_data, "latex",caption ='Data', booktabs = T) %>%
```

```
kable_styling(latex_options = "HOLD_position")
set.seed(123456)
predictions = function(size=1000){
  draws = c()
  for (i in 1:size) {
    beta = rmvnorm(n = 1, mean = mu,sigma = Sigma)
    pred = work_data*beta
    probs = exp(pred)/(1+exp(pred))
    draws[i] = rbinom(n = 1,size = 1,probs)
  }
    draws
}
kable(table(predictions()), "latex",caption ='Predictions', booktabs = T) %>%
kable_styling(latex_options = "HOLD_position")
r = predictions()
hist(r, main='Predictions', xlab = 'Predictions')
``
```