Week 4 lab solutions – MAST90125: Bayesian Statistical learning

Question One

We have seen residual plots in lectures. One example is a case where observations y_i are simulated using

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2) \quad i = 1, \cdots, n, \tag{1}$$

but the model fitted was

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \cdots, n. \tag{2}$$

Simulate 2000 datapoints according to (1) and fit the model (2) to this data and answer the following.

```r
#Step 1: Simulate an explanatory variable
set.seed(123456)
n<-2000; x<-rnorm(n)
beta<-c(1.5,1.5,-0.5) #Specify co-efficient values
X<-cbind(rep(1,n),x,x^2) #Create predictor matrix
y<-X%*%beta + rnorm(n)*1.4 #Create response vector with i.i.d. errors
y<-as.numeric(y)
#We have shown a t-test can be interpreted in a Bayesian way, so can fixed-effect
#regression. Therefore just estimate co-efficients using the lm function.
#Then extract co-efficients and estimate of residual variance.
#Note the question says to fit a model with only a linear term for x.
Xm <-X[,1:2] #Component of X matrix used in model fit.
mod<-lm(y~0+Xm) #Estimate was coded in X matrix, suppress this with 0 in lm.
  #Note as intercept was included, remove second term.
betaest<-mod$coef; betaest
```

```
##       Xm      Xmx
## 1.032173 1.549181
```

```r
s2 <-summary(mod)$sigma^2; s2
```

```
## [1] 2.405377
```

a) For your simulated data, generate replicate data from the posterior predictive distribution, and construct two test statistics such that one will suggest poor model fit and the other will suggest good model fit.

   Hint: In week 2 lab, we found a Bayesian interpretation of a t-test in the context of estimating a mean. Generalise this to the regression case.

b) Perform a marginal check, that is calculate

$$p_i = \Pr(y_i^{\text{rep}} \leq y_i | y_1, \cdots, y_n)$$

   Comment on the distribution of p_i, including a discussion of whether this marginal check was appropriate for checking model plausibility in this example. Graphical summaries may prove useful.

To answer **part a)**, draw $\boldsymbol{\beta}$ from multivariate normal with mean $\hat{\boldsymbol{\beta}}_{MLE}$ and variance $\sigma^2(X^\top X)^{-1}$, or from multivariate t with $df = n - p$, location $\hat{\boldsymbol{\beta}}_{MLE}$, and scale matrix $s^2(X^\top X)^{-1}$. Also draw $\tau = (\sigma^2)^{-1}$ from gamma distribution with parameters $a = (n-p)/2$, $b = (n-p)s2/2$. In this case $p = 2$ as we estimate an intercept and slope.

```
library(mvtnorm)
```

```
## Warning: package 'mvtnorm' was built under R version 4.3.1
```

```
XTXinv <-solve(crossprod(Xm)) #Note drop the last column as it was not fitted in the lm.
n<-length(y);p<-dim(Xm)[2]
iter = 3000
vyrep1<-vyrep2<-rangeyrep1<-rangeyrep2<-0 #Storing test statistics.
for(i in 1:iter){
  tau <-rgamma(1,0.5*(n-p),0.5*(n-p)*s2) #draw a precision
  sigma2 <-1/tau #convert to variance
  betarep1<- rmvnorm(1,mean=betaest,sigma=sigma2*XTXinv ) #draw co-efficient
  betarep2<- rmvt(1, sigma=s2*XTXinv, df=n-2,delta=betaest)
  Xbetarep1 <- Xm%*%as.vector(betarep1)
  Xbetarep2 <- Xm%*%as.vector(betarep2)
  yrep1 <-Xbetarep1+rnorm(n)*sqrt(sigma2) #create replicate dataset 1.
  yrep2 <-Xbetarep2+rnorm(n)*sqrt(sigma2) #create replicate dataset 2.
  yrep1 <-as.numeric(yrep1); yrep2 <-as.numeric(yrep2)
  vyrep1[i] <-var(yrep1) #Example of poor choice of diagnostic.
  vyrep2[i] <-var(yrep2) #Example of poor choice of diagnostic.
  rangeyrep1[i] <-diff(range(yrep1)) # A better choice of test diagnostic.
  rangeyrep2[i] <-diff(range(yrep2)) # A better choice of test diagnostic.
}
table(vyrep1>var(y))
```

```
##
## FALSE  TRUE
##  1466  1534
```

```
table(vyrep2>var(y))
```

```
##
## FALSE  TRUE
##  1496  1504
```

```
table(rangeyrep1>diff(range(y)))
```

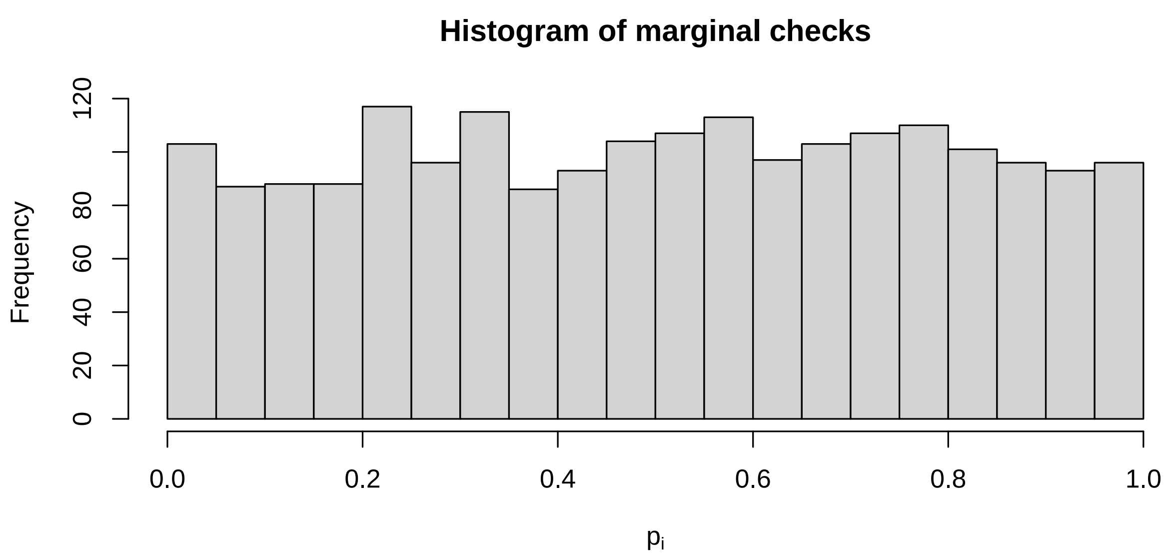
```
##
## FALSE  TRUE
##  2994     6
```

```
table(rangeyrep2>diff(range(y)))
```

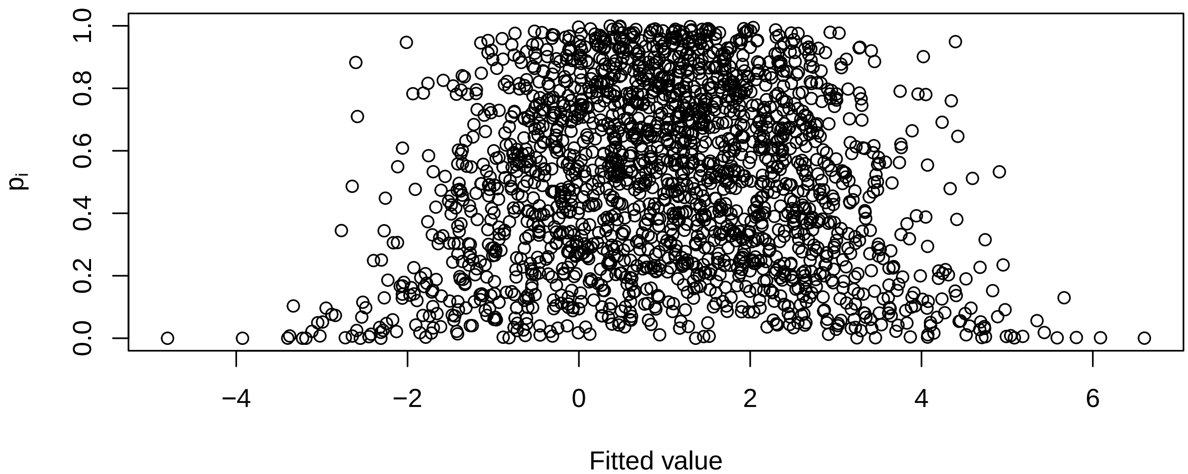
```
##
## FALSE  TRUE
##  2994     6
```

To answer **part b)**:

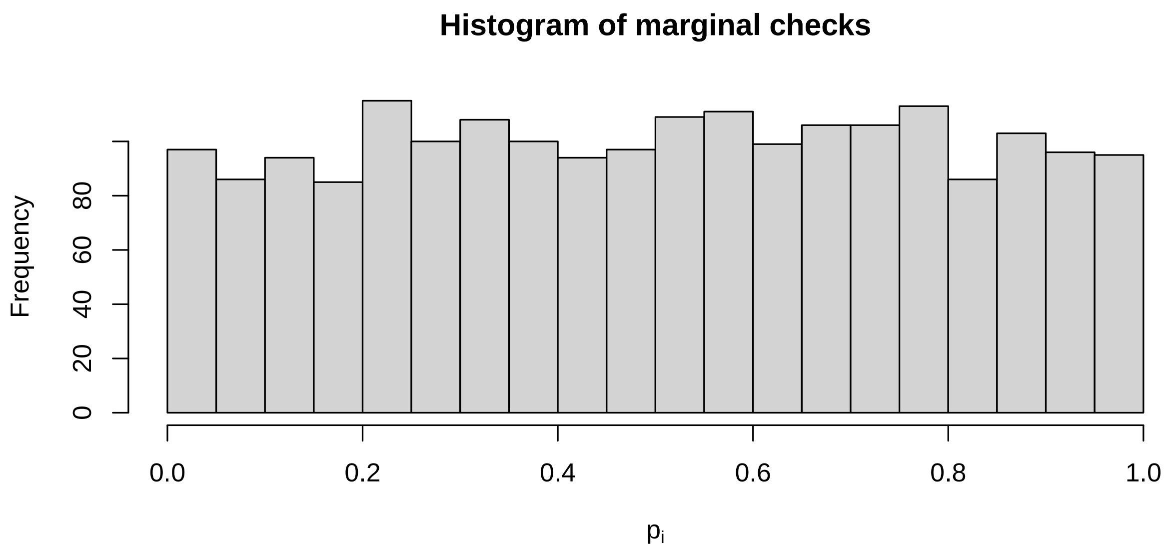
```
pvec1<- pvec2 <-rep(0,n)
for(i in 1:iter){
  tau <-rgamma(1,0.5*(n-p),0.5*(n-p)*s2) #draw a precision
  sigma2 <-1/tau #convert to variance
  betarep1<- rmvnorm(1,mean=betaest,sigma=sigma2*XTXinv ) #draw co-efficient
  betarep2<- rmvt(1, sigma=s2*XTXinv, df=n-2,delta=betaest)
  Xbetarep1 <- Xm%*%as.vector(betarep1)
  Xbetarep2 <- Xm%*%as.vector(betarep2)
  yrep1 <-Xbetarep1+rnorm(n)*sqrt(sigma2) #create replicate dataset 1.
  yrep2 <-Xbetarep2+rnorm(n)*sqrt(sigma2) #create replicate dataset 2.
  yrep1 <-as.numeric(yrep1); yrep2 <-as.numeric(yrep2)
  pvec1[yrep1 < y] <-pvec1[yrep1 <y]+1
  pvec2[yrep2 < y] <-pvec2[yrep2 <y]+1
}
#Histogram of marginal p-values.
hist(pvec1/iter,xlab=expression(p[i]),main='Histogram of marginal checks',breaks=20)
```



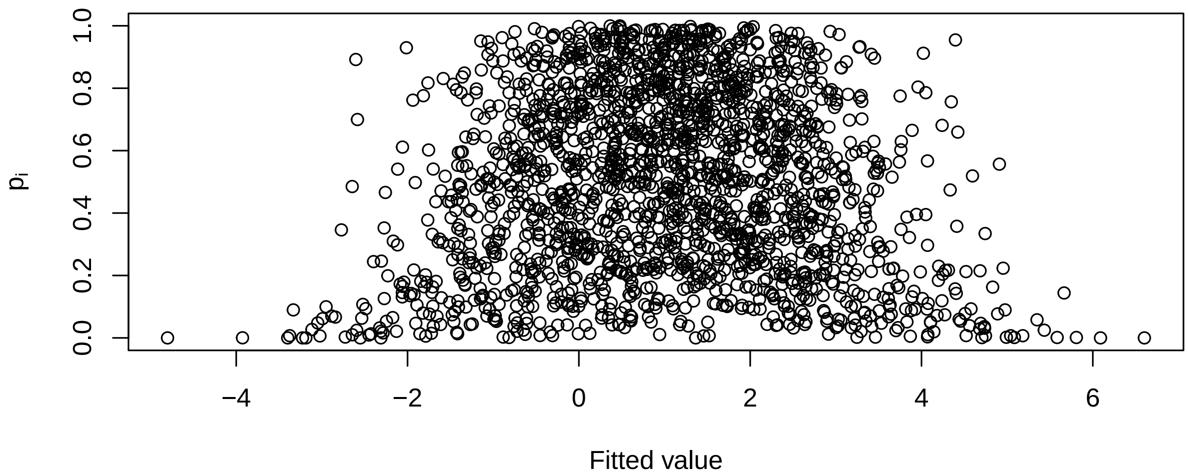

```
#Compare to prediction of fitted values.
plot(predict(mod),pvec1/iter,ylab=expression(p[i]),xlab='Fitted value')
```




```
hist(pvec2/iter,xlab=expression(p[i]),main='Histogram of marginal checks',breaks=20)
```




```
#Compare to prediction of fitted values.
plot(predict(mod),pvec2/iter,ylab=expression(p[i]),xlab='Fitted value')
```



If model (2) is a good approximation of model (1), each p_i value should be around 0.5. Both histograms of all p_i values, however, look like a uniform distribution over $[0, 1]$. This suggests model (2) is not a good approximation of model (1). Same conclusion can be obtained from the two scatter plots of p_i versus y_i. It can be verified that the parameter β_2 in model (1) is significantly different from 0, confirming that model (2) is not a good estimation of the true model (1).

Question Two

In week 2 Lab, we looked at the posterior distribution for the parameters of a normal distribution assuming Jeffreys' priors. For this example, determine

$$\text{i) } \mathrm{Var}(\mu|\mathbf{y}), \quad \text{ii) } E(\mathrm{Var}(\mu|\mathbf{y})), \quad \text{iii) } \mathrm{Var}(E(\mu|\mathbf{y})), \quad \text{with } \mathbf{y} = (y_1, \cdots, y_n)^\top$$

and using the law of total variance, deduce what this implies about $\mathrm{Var}(\mu)$.

Hint: If z is drawn from a student-t distribution with ν degrees of freedom, then $E(z) = 0$ and $\mathrm{Var}(z) = \frac{\nu}{\nu-2}$.

Answer:

In week 2 lab, we showed that the posterior of μ is t with $n-1$ degrees of freedom, location parameter $\bar{y}$ and scale parameter s^2/n. This can be converted to a standard student t by writing $z = \dfrac{\mu - \bar{y}}{\sqrt{s^2/n}}$.

Hence $\mathrm{Var}(\mu|\mathbf{y}) = \dfrac{s^2}{n} \times \dfrac{n-1}{n-3} = \dfrac{(n-1)s^2}{n(n-3)}$ and $E(\mathrm{Var}(\mu|\mathbf{y})) = \dfrac{(n-1)E(s^2)}{n(n-3)} = \dfrac{(n-1)\sigma^2}{n(n-3)}$.

Also $\mathrm{Var}(E(\mu|\mathbf{y})) = \mathrm{Var}(\bar{y}) = \dfrac{\sigma^2}{n}$. This implies that

$$\mathrm{Var}(\mu) = \frac{(n-1)\sigma^2}{n(n-3)} + \frac{\sigma^2}{n} = \frac{\sigma^2}{n}\left(\frac{n-1}{n-3} + 1\right) = \frac{\sigma^2}{n}\left(2 + \frac{2}{n-3}\right)$$

is a function of n. This highlights that the prior is improper and does not have a defined variance.