Effects of Latitude and Elevation on Temperature

Jina Park - 30129680

4/10/2022

Abstract

Bayes' Theorem uses the prior probability distribution and the likelihood of the data to generate a posterior probability distribution. Models, hypotheses, and parameter estimates can be stated with a function of posterior probability distributions.

This paper is collecting the data from 16 counties in Texas and will mainly be addressing the effects of latitude and elevation (altitude) on temperature using Bayesian inference in order to find the relationship between them and to make temperature forecasts depending on latitude and elevation. In addition, metropolis algorithm and non-informative prior distribution are expected to be used and compared to draw a conclusion.

Introduction

Bayes' Theorem is widely used to predict posterior distributions using the prior distribution and the likelihood of the data, treating model parameters as random variables. With Bayesian inference, it is not only can obtain a quantitative measure of the probability of a hypothesis being true in light of the available data (P(H|Y)), but also figure out probability as an individual's degree of belief in the likelihood of an event. The probability of the hypothesis being conditional on the sample data can be found by the following formula:

$P(H \mid Y) = \frac{f(Y|H)\pi(H)}{P(Y)}$

It is well known that change in latitude and altitude affects on temperature. The lower latitude and elevation, the warmer the temperature (Key et al. 1997). When we are able to figure out the relationship of temperature with latitude and altitude, we can predict temperature with corresponding latitude and altitude with certain percentage of confidence. Therefore, this paper will be using Bayesian inference to observe the effects of latitude and elevation on temperature by taking the data that measures average high temperature in January of 16 counties in Texas with latitude and altitude level as an example, and forecast the temperature in January in 16 counties of Texas.

We are expecting to obtain the best model with Aikaike's information criterion (AIC; Burnham & Anderson 2002), maximum likelihood estimates of the model and posterior distribution with non-informative prior distribution. We will assume non-informative prior distribution as $\beta \sim N$ (0,1000) to approximate the posterior distribution.

I will be using RStudio throughout the paper to conduct parameter estimation and model selection. The best model will be chosen from the set of models by Aikaike's information criterion (Burnham & Anderson 2002). Metropolis algorithm and non-informative prior distribution will be used to estimate the posterior probability distribution.

Data and Analysis

The original data set (University of Florida 2013) not only includes variables latitude and elevation but also longitude, however, longitude has no relationship (no effect) with temperature change((National Geographic Society 2019) so longitude is excluded from the data to be worked with. Hence, the working data will be the following:

```
texas <- read_excel("texas.dat.xlsx",
                        range = "A1:D17")
texas
```

```
## # A tibble: 16 x 4
##     Name        'Temp (F)' Latitude Elevation
##     <chr>           <dbl>    <dbl>     <dbl>
##  1 Harris             56     29.8        41
##  2 Dallas             48     32.8       440
##  3 Kennedy            60     26.9        25
##  4 Midland            46     32.0      2851
##  5 Deaf Smith         38     34.8      3840
##  6 Knox               46     33.4      1461
##  7 Maverick           53     28.7       815
##  8 Nolan              46     32.4      2380
##  9 El Paso            44     31.8      3918
## 10 Collington         41     34.8      2040
## 11 Pecos              47     30.9      3000
## 12 Sherman            36     36.4      3693
## 13 Travis             52     30.3       597
## 14 Zapata             60     26.9       315
## 15 Lasalle            56     28.4       459
## 16 Cameron            62     25.9        19
```

Out of this data set, I examined simple additive models (temperature as a function of latitude and elevation) and models that included all possible interaction terms. The 'best' model was chosen from the set of candidate models by minimizing Aikaike's information criterion (AIC; Burnham & Anderson 2002):

AIC = -2 $\log(x)$ $(L(\hat{\beta} \mid data))$ +2k

$(L(\hat{\beta} \mid data))$ stands for the likelihood of the model with parameters β_i conditional on the data, and k stands for the number of parameters in the model.

The response variable is the average Jan high temp in 16 Texas counties. Here, It uses a linear model to relate a discrete (Poission) random variable (temperature) to two continuous predictor variables (latitude and elevation).

```
lat.mod <- lm(texas$'Temp (F)'~texas$Latitude, data = texas)
ele.mod <-  lm(texas$'Temp (F)'~texas$Elevation, data =texas)
lat.ele.mod <- lm(texas$'Temp (F)'~texas$Latitude+texas$Elevation, data = texas)
interaction.mod <- lm(texas$'Temp (F)'~texas$Latitude*texas$Elevation, data = texas)
all.mod <- lm(texas$'Temp (F)'~texas$Latitude+texas$Elevation+texas$Latitude*texas$Elevation, data = te

models <- list(lat.mod,ele.mod,lat.ele.mod,interaction.mod,all.mod)
model.names <- c('L','E','L+E','LxE','L+E+LxE')

aictab(cand.set=models, modnames = model.names)
```

```
## Warning in aictab.AIClm(cand.set = models, modnames = model.names):
## Check model structure carefully as some models may be redundant
```

```
##
## Model selection based on AICc:
##
##          K  AICc Delta_AICc AICcWt Cum.Wt     LL
## L+E      4 48.54       0.00   0.81   0.81 -18.45
## LxE      5 52.84       4.30   0.09   0.91 -18.42
## L+E+LxE  5 52.84       4.30   0.09   1.00 -18.42
## L        3 74.80      26.26   0.00   1.00 -33.40
## E        3 96.39      47.84   0.00   1.00 -44.19
```

In the table above, L indicates latitude and E indicates elevation. According to this data, the smallest value of AICc is 48.54, which has a simple addictive model T = L+E. Therefore, we will be working on the model that has an intercept (β_1) and main effects (β_2,β_3) without interaction term:
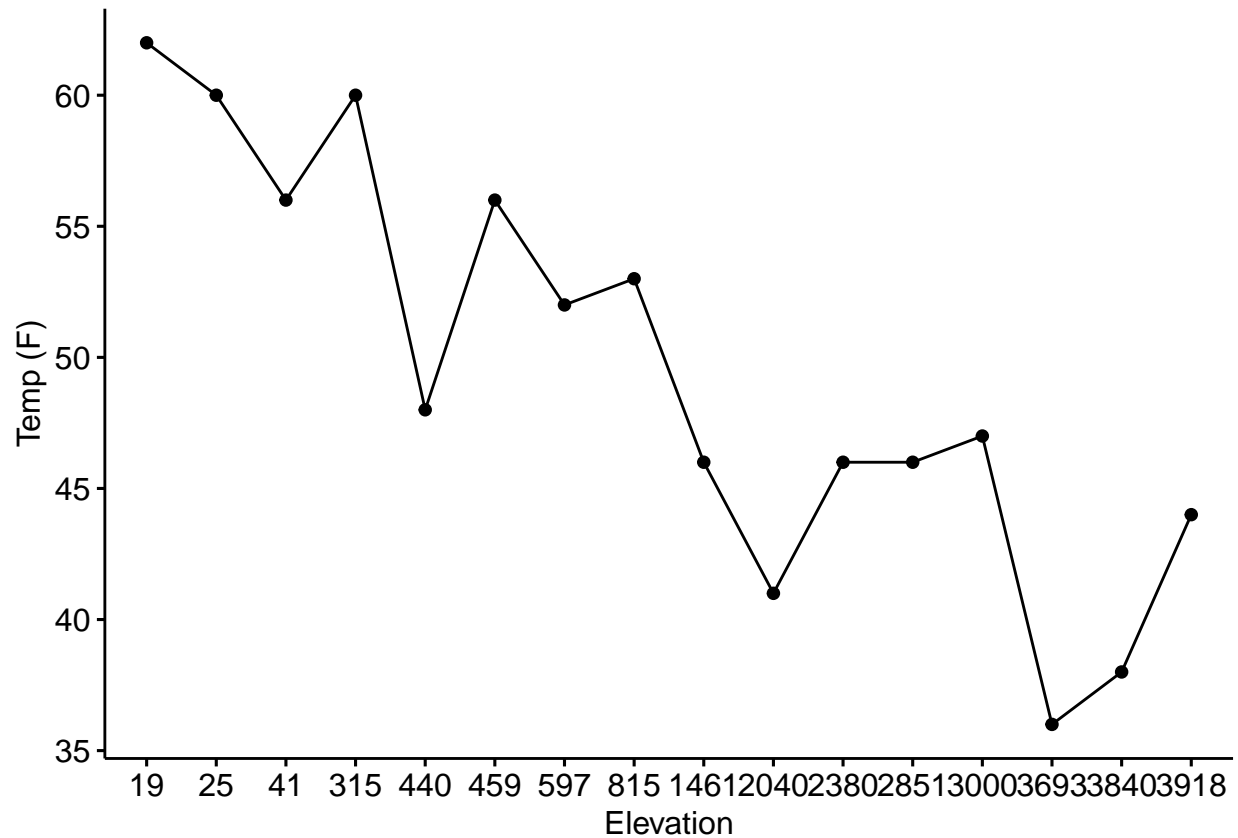
$$(log(\hat{T}_i)) = \hat{\beta}_1 + \hat{\beta}_2 latitude_i + \hat{\beta}_3 elevation_i$$

The fit of this model enables us to understand the relationship of temperature with latitude and elevation, and it can be illustrated with graphs.

```
ggline(data=texas, x="Latitude", y="Temp (F)")
```
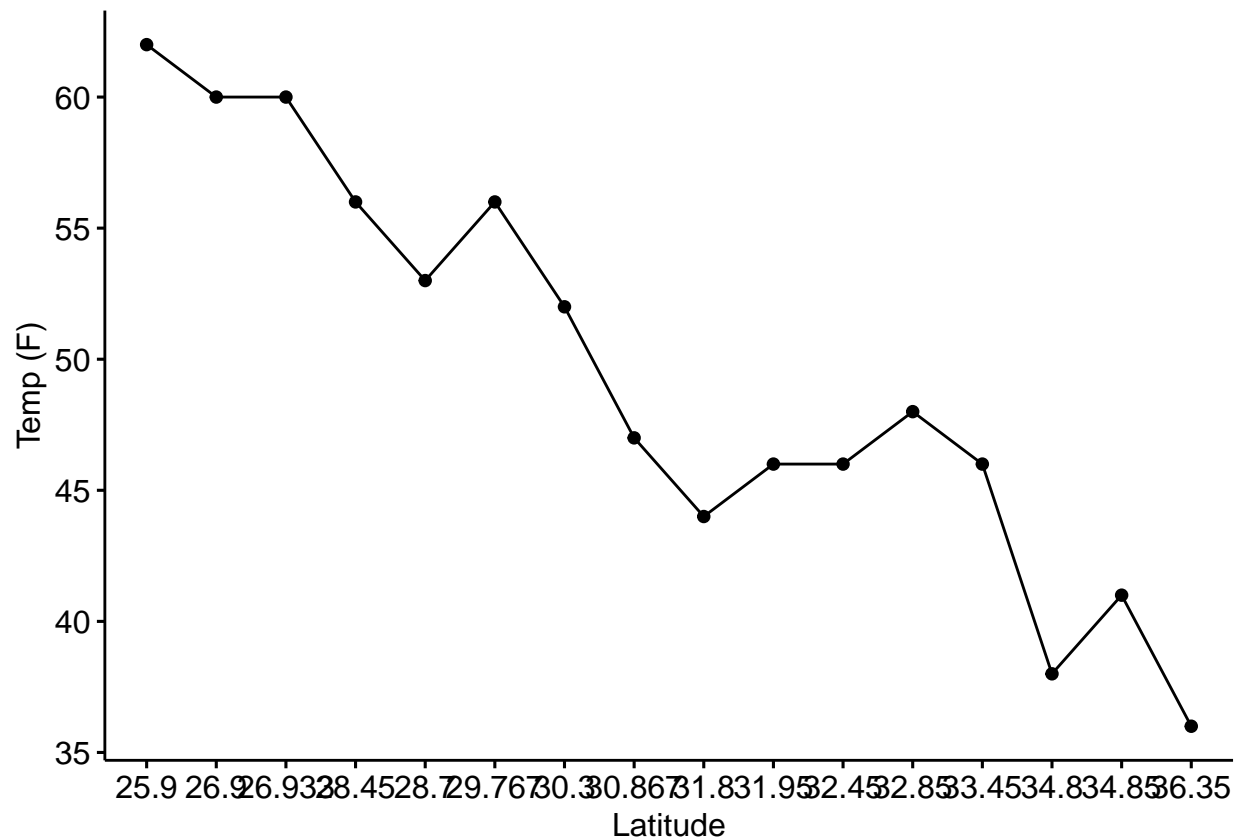
```
ggline(data=texas, x="Elevation", y="Temp (F)")
```


As can be seen from the graphs above, they provide us with the information that the higher latitude and elevation, the lower the temperature. Also, the following graph and the ordinary least square(OLS) estimating table illustrate that the model fits nicely.

```
ggplot(texas, aes(Latitude+Elevation, 'Temp (F)')) + geom_point()+geom_smooth(method = lm)
```

```
## 'geom_smooth()' using formula 'y ~ x'
```
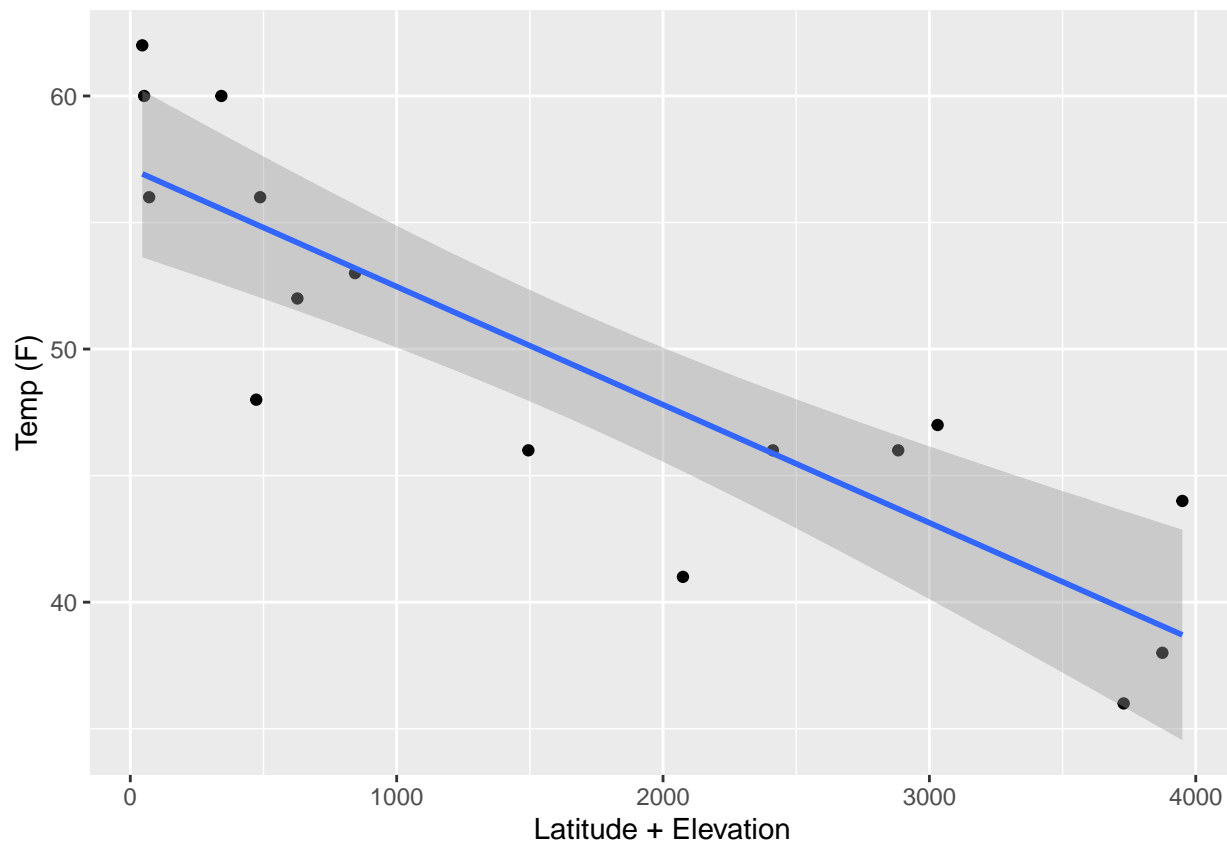
```r
#The ordinary least square (OLS) estimator
lmfit<-lm(texas$'Temp (F)'~texas$Latitude+texas$Elevation)
summary(lmfit)
```

```
##
## Call:
## lm(formula = texas$'Temp (F)' ~ texas$Latitude + texas$Elevation)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.1711 -0.3172  0.1458  0.5542  1.3548
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.093e+02  2.979e+00  36.682 1.64e-14 ***
## texas$Latitude  -1.832e+00  1.038e-01 -17.651 1.83e-10 ***
## texas$Elevation -1.846e-03  2.188e-04  -8.439 1.24e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8506 on 13 degrees of freedom
## Multiple R-squared:   0.99,  Adjusted R-squared:  0.9885
## F-statistic: 644.4 on 2 and 13 DF,  p-value: 9.906e-14
```

By using the first equation stated earlier, we can derive the following equation to estimate the posterior:

5

$P(\beta \mid data) \propto P(data \mid \beta) \text{ X } \pi(H)$

We can figure out the maximum likelihood estimates (MLE) using the following R codes.

```r
texas.mle=glm(texas$'Temp (F)'~texas$Latitude+texas$Elevation, family="poisson")
summary(texas.mle)
```

```
##
## Call:
## glm(formula = texas$'Temp (F)' ~ texas$Latitude + texas$Elevation,
##     family = "poisson")
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -0.25339  -0.09661  -0.00216   0.11189   0.21280
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)      5.081e+00  5.005e-01  10.152   <2e-16 ***
## texas$Latitude  -3.636e-02  1.759e-02  -2.067   0.0387 *
## texas$Elevation -3.964e-05  3.792e-05  -1.045   0.2958
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 19.19011  on 15  degrees of freedom
## Residual deviance:  0.27631  on 13  degrees of freedom
## AIC: 97.952
##
## Number of Fisher Scoring iterations: 3
```

Now, we are going to run Metropolis algorithm.

Metropolis algorithm gives an acceptance rate approximately 41%. Effective sample size are 1313.3506, 552.0609 and 1513.4282 for β_1, β_2 and β_3, respectively.

Plots of the Markov chain in β_2 and β_3 along with autocorrelation functions are illustrated below:

```r
#Beta2
par(mar=c(3,3,1,1),mgp=c(1.75,.75,0))
par(mfrow=c(1,3))

blabs <-c(expression(beta[1]),expression(beta[2]),expression(beta[3]))
j<-2 #beta2
thin <- c(1,seq(10,S,by=25))
plot(thin,BETA[thin,j],type="l",xlab="iteration",ylab=blabs[j])
abline(h=mean(BETA[,j]))
acf(BETA[,j],ci.col="gray",xlab="lag")
acf(BETA[thin,j],xlab="lag/25",ci.col="red")
```
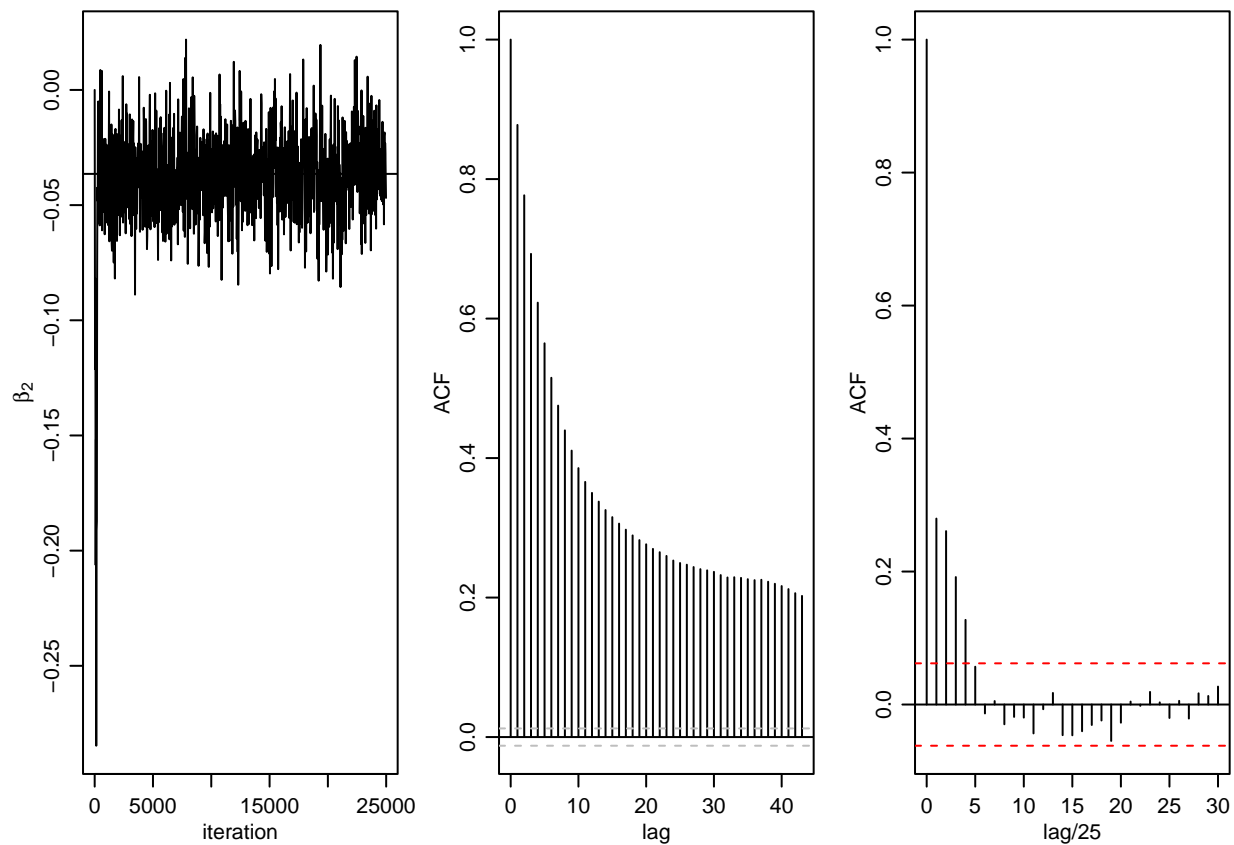
```
#Beta3
par(mar=c(3,3,1,1),mgp=c(1.75,.75,0))
par(mfrow=c(1,3))

blabs <-c(expression(beta[1]),expression(beta[2]),expression(beta[3]))
j<-3 #beta3
thin <- c(1,seq(10,S,by=25))
plot(thin,BETA[thin,j],type="l",xlab="iteration",ylab=blabs[j])
abline(h=mean(BETA[,j]))
acf(BETA[,j],ci.col="gray",xlab="lag")
acf(BETA[thin,j],xlab="lag/25",ci.col="red")
```
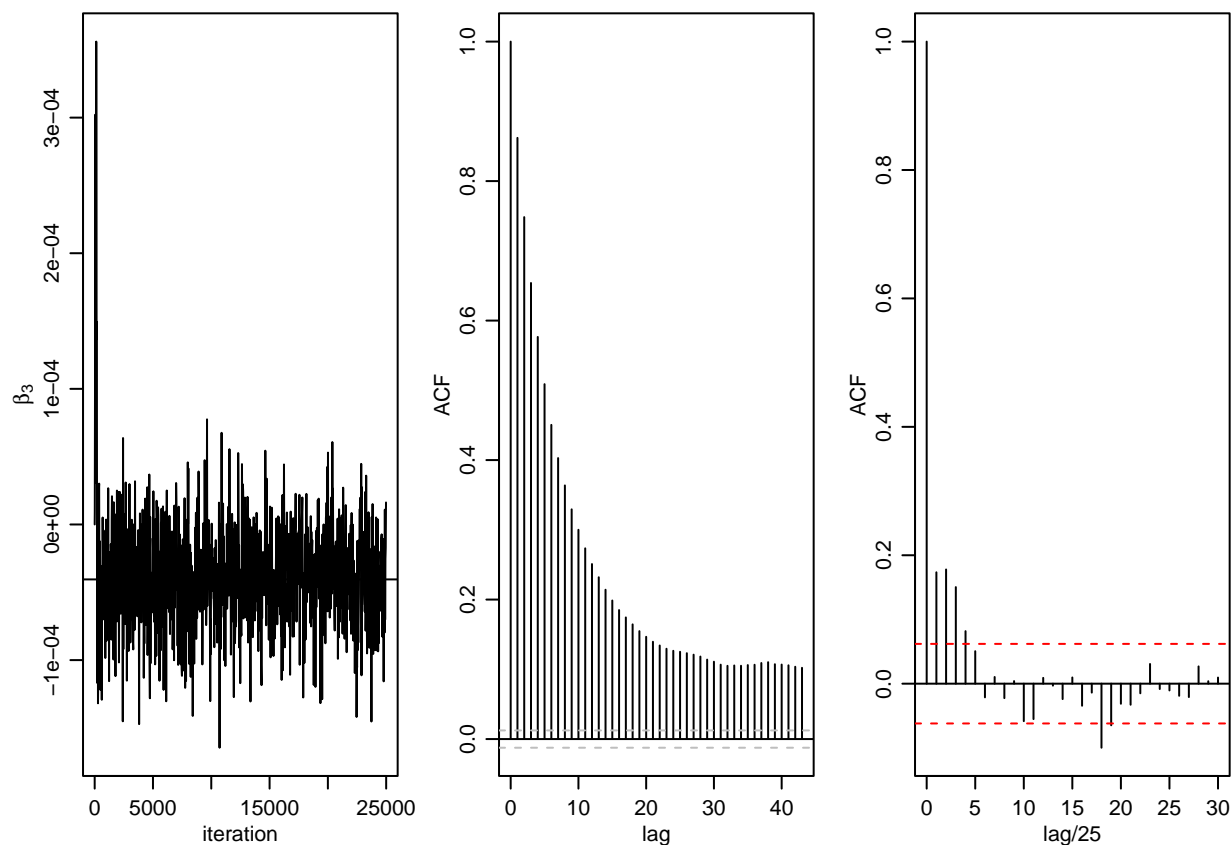
The graphs with lag/25 in x-axis gives us the autocorrelation function of every 25th value of parameters from Markov chain. For β_2, Markov chain is highly correlated. However, we find there is a correlated errors for β_1. Therefore, we are going to generalize this regression model.

```r
n<-dim(texas)[1]
y<-texas$'Temp (F)'
X<-cbind(rep(1,length(y)),texas$Elevation)
DY <- abs(outer((1:n),(1:n),"-"))
lmfit<-lm(y~-1+X)
beta<-lmfit$coef
s2<-summary(lmfit)$sigma^2
phi<-acf(lmfit$residuals,plot=FALSE)$acf[2]

nu0<-1
s20<-1
T0<-diag(1/1000,nrow=2)

set.seed(1)
S<-25000
odens<-S/1000
OUT<-NULL
ac<-0
par(mfrow=c(1,2))
for (s in 1:S)
{
  Cor<-phi^DY
```

```r
    iCor<-solve(Cor)
    V.beta<-solve(t(X)%*%iCor%*%X/s2+T0)
    E.beta<-V.beta%*%(t(X)%*%iCor%*%y/s2)
    beta<-t(rmvnorm(1,E.beta,V.beta))

    s2<-1/rgamma(1,(nu0+n)/2,(nu0*s20+t(y-X%*%beta)%*%iCor%*%(y-X%*%beta))/2)

    phi.p<-abs(runif(1,phi-.1,phi+.1))
    phi.p<- min(phi.p,2-phi.p)
    lr<- -.5*(determinant(phi.p^DY,log=TRUE)$mod-
              determinant(phi^DY,log=TRUE)$mod+
            sum(diag((y-X%*%beta)%*%t(y-X%*%beta)%*%(solve(phi.p^DY)-solve(phi^DY))))/s2)

    if(log(runif(1))<lr) {phi<-phi.p ; ac<-ac+1}
    if(s%%odens==0)
    {#cat(s,ac/s,beta,s2,phi,"\n")
      OUT<-rbind(OUT,c(beta,s2,phi))
    }
}

OUT.25000<-OUT
#Effective sample size
#apply(OUT.25000,2,effectiveSize) = 1000.0000 1000.0000 463.9249  380.8224

#figure
par(mar=c(3,3,1,1),mgp=c(1.75,.75,0))
par(mfrow=c(1,2))
plot(OUT.25000[,4],xlab="scan/25",ylab=expression(rho),type="l")
acf(OUT.25000[,4],ci.col="red",xlab="lag/25")
```
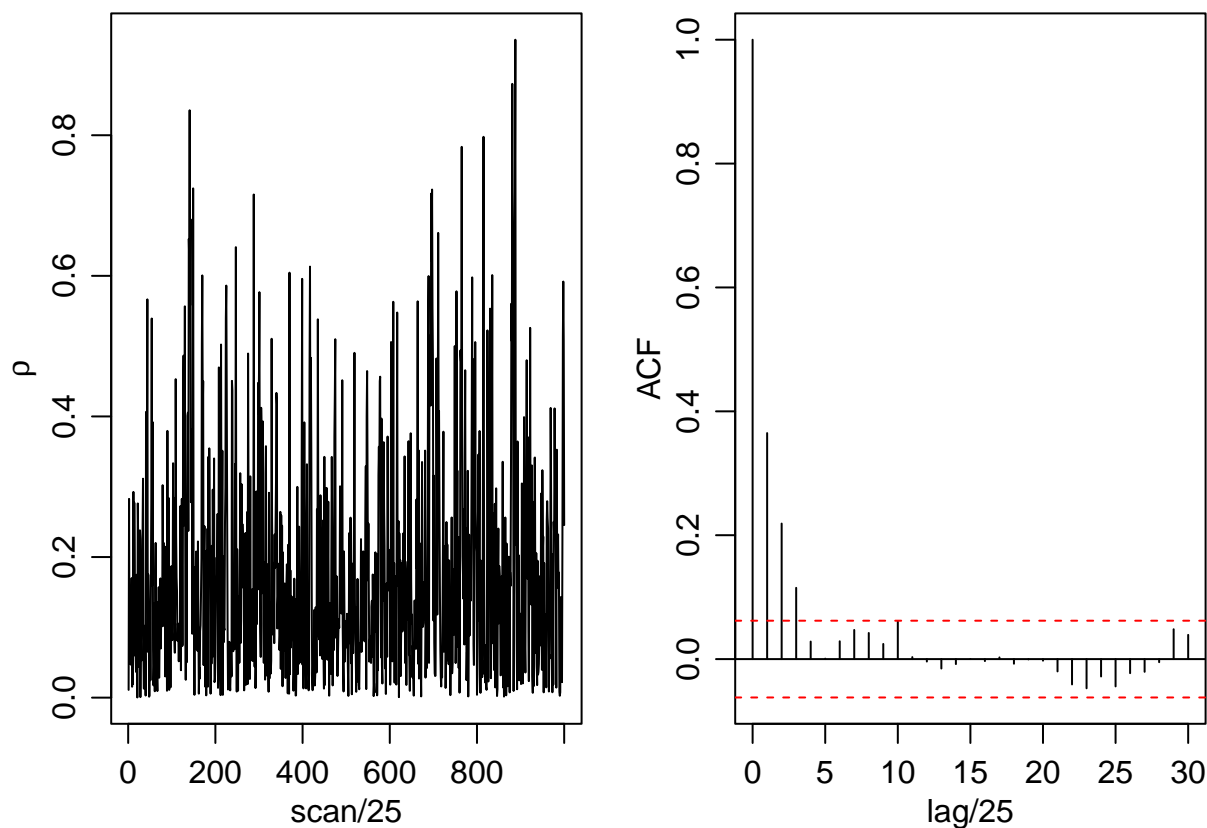
Now, we have a highly correlated Markov chain. This suggests that we have nearly the equivalent of 1000 independent samples of β_2 and β_3 with which to approximate the posterior distribution. We can visualize the MCMC approximations. MCMC approximations to the posterior marginal distributions of β_2 and β_3 are illustrated in black.

```
par(mar=c(3,3,1,1),mgp=c(1.75,.75,0))
par(mfrow=c(1,2))

plot(density(BETA[,2],adj=2),lwd=2,type="l",xlab=expression(beta[2]),ylab=expression(paste(italic("p(")
abline(h=0,col="gray")
abline(v=0,col="gray")
plot(density(BETA[,3],adj=2),lwd=2,type="l",xlab=expression(beta[3]),ylab=expression(paste(italic("p(")
abline(h=0,col="gray")
abline(v=0,col="gray")
```

density.default(x = BETA[, 3], adjust
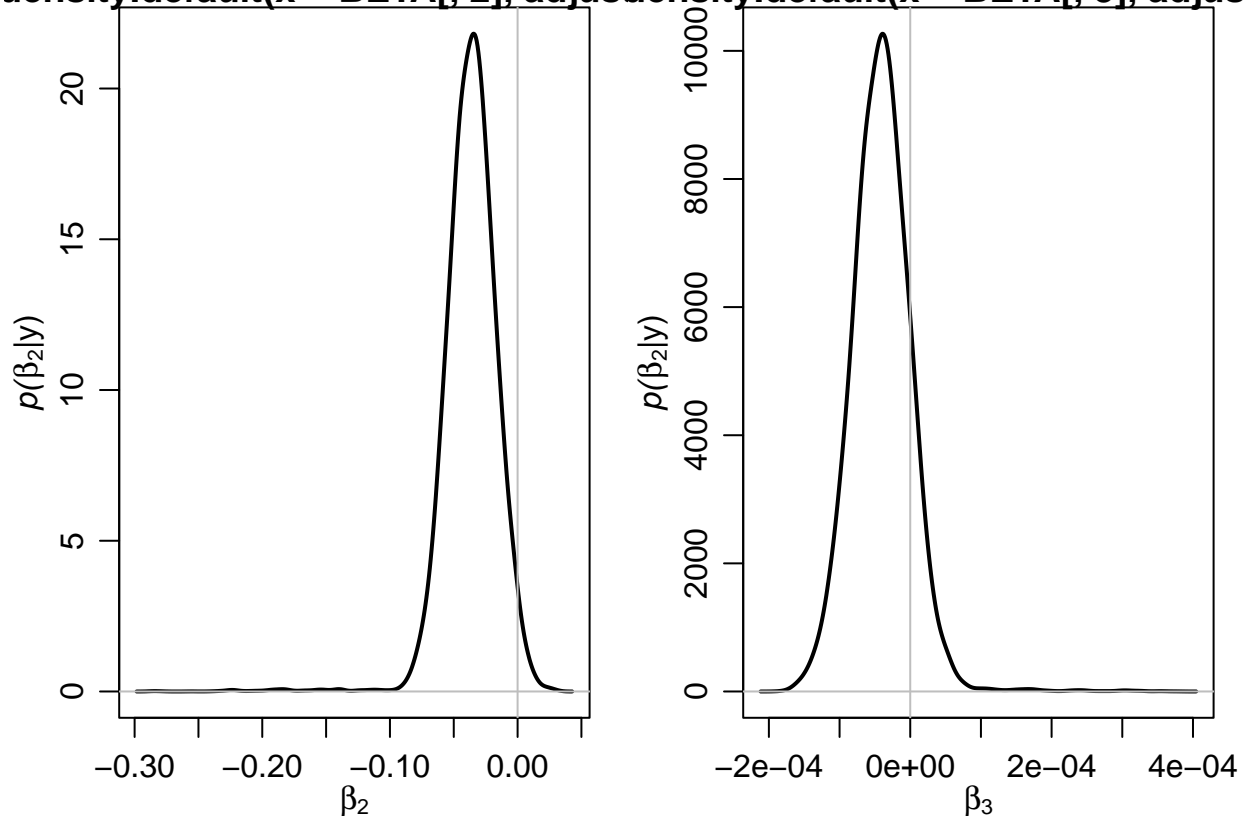
Apart from MCMC approximation, we can also work on non-informative prior to approximate the posterior distribution. I will be using a generalized linear regression model, assuming non-informative prior distribution as $\beta \sim N\ (0,1000)$.

```
n=16
nu0=1
s20=summary(lm(y~-1+X))$sigma^2

lm.gprior<-function(y,X,g=dim(X)[1],nu0=1,s20=try(summary(lm(y~-1+X))$sigma^2,silent=TRUE),S=1000)
{
  n<-dim(X)[1]
  p<-3
  Hg<-(g/(g+1))*X%*%solve(t(X)%*%X)%*%t(X)
  SSRg<-t(y)%*%(diag(1,nrow=n)-Hg)%*%y
  s2<-1/rgamma(S,(nu0+n)/2, (nu0*s20+SSRg)/2)

  Vb<-g*solve(t(X)%*%X)/(g+1)
  Eb<-Vb%*%t(X)%*%y
  E<-matrix(rnorm(S*p,0,sqrt(s2)),S,p)
  beta<-t(t(E%*%chol(Vb))+c(Eb))
  list(beta=beta,s2=s2)
}
x1<-texas$Latitude
x2<-texas$Elevation
y<-texas$'Temp (F)'
n<-length(y)
```

```
X<-cbind(rep(1,n),x1,x2)
p<-dim(X)[2]
tmp<-lm.gprior(y,X)
beta.post<-tmp$beta
beta.ols <-lm(y~-1+X)$coef
beta.ols #OLS beta estimates
```

```
##            X           Xx1            Xx2
## 109.258872211  -1.832160909  -0.001846349
```

```
g<-n
nu0<-1
s20<-summary(lm(y~-1+X))$sigma^2
beta.ols*g/(g+1) #Posterior mean
```

```
##            X           Xx1           Xx2
## 102.83187973  -1.72438674  -0.00173774
```

```
iXX<-solve(t(X)%*%X)

#Figure
#Marginal prior distribution
mdt<-function(t,mu,sig,nu){
  gamma(.5*(nu+1))*(1+((t-mu)/sig)^2/nu)^(-.5+(nu+1))/(sqrt(nu*pi)*sig*gamma(nu/2))
}

par(mfrow=c(1,2),mar=c(2.75,2.75,.5,.5),mgp=c(1.7,.7,0))
x<-seq(-85,130,length=200)
plot(density(beta.post[,2],adj=2),xlab=expression(beta[2]),main="",ylab="",lwd=2)
abline(v=0,col="grey")
lines(x,mdt(x,0,sqrt(n*s20*iXX[2,2]),nu0),col="red")
x<-seq(-5,5,length=100)
plot(density(beta.post[,3],adj=2),xlab=expression(beta[3]),main = "",ylab="",lwd=2)
abline(v=0,col="grey")
lines(x,mdt(x,0,sqrt(n*s20*iXX[3,3]),nu0),col="red")
```
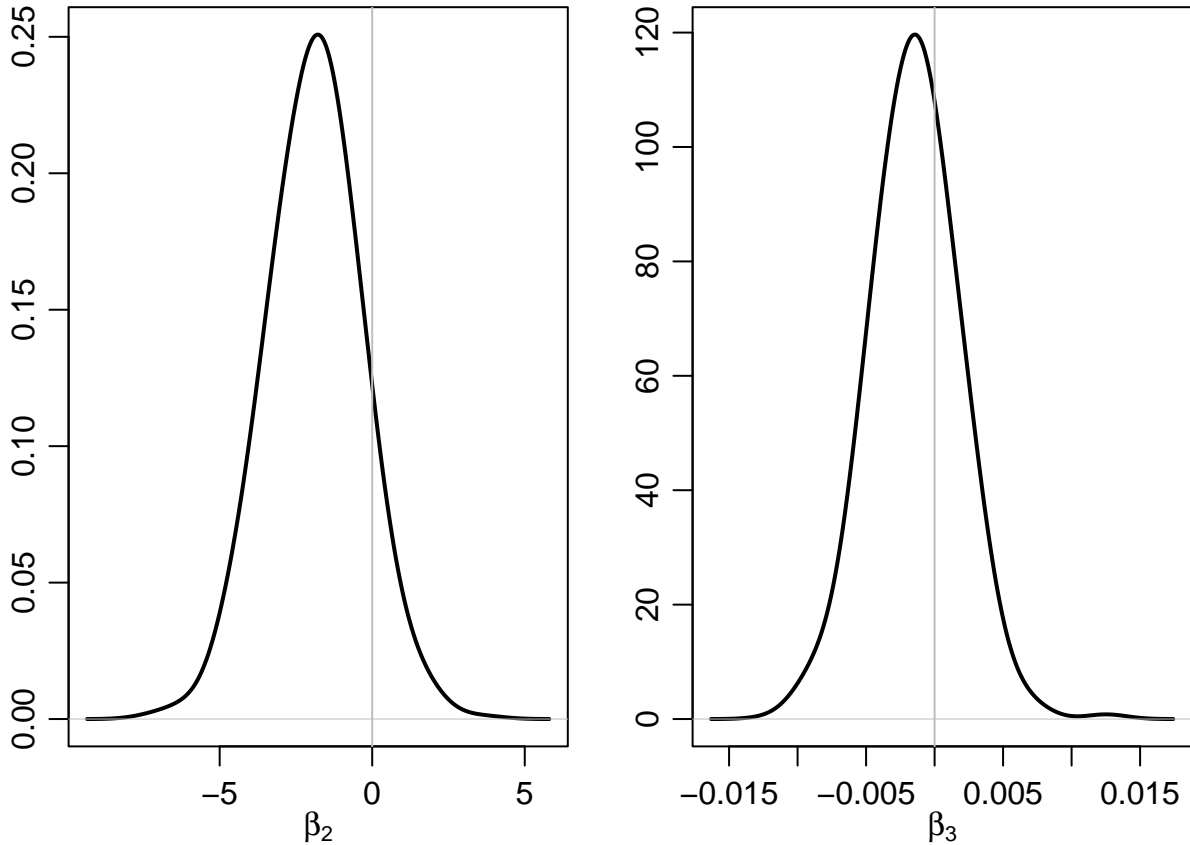
Conclusion

The graphs with two different algorithm give us the similar posterior distribution. From the Bayesian analysis, we could conclude that the additive model is a believable description of how latitude and elevation can be used to predict temperature. We have approximately 41% of acceptance rate from Metropolis method. However, there was a limitation to the model selection as there can be other factors that influences the temperature changes. Since Bayes' Theorem allows us to estimate the posterior distribution with a less prior information or data, other models, even with less information, can also be used to estimate the effects of latitude and elevation on temperature. In addition, other factors also can be considered to estimate temperature.

References

- Burnham, K.P. & Anderson, D.R. (2002). Model Selection and Inference: A Practical Information-theoretic Approach, 2nd edn. SpringerVerlag, New York.

- Key, J. R., Collins, J. B., Fowler, C., & Stone, R. S. (1997). High-latitude surface temperature estimates from thermal satellite data. Remote Sensing of Environment, 61(2), 302-309.

- National Geographic Society. (2019, September 19). Latitude, longitude, and temperature. National Geographic Society. Retrieved April 11, 2022, from https://www.nationalgeographic.org/activity/ latitude-longitude-temperature/#:~:text=Latitude%20and%20longitude%20make%20up,and%20cooler%20approaching

- University of Florida. (2013, October 29). Index of /~winner/data. (n.d.). Retrieved April 11, 2022, from https://users.stat.ufl.edu/~winner/data/texas1.dat

- Aaron M. Ellison. (2004, May 10). Bayesian inference in ecology - ellison - 2004 - ecology . . . Retrieved April 12, 2022, from https://onlinelibrary.wiley.com/doi/full/10.1111/j.1461-0248.2004.00603.x