#Assignment 5

Importing the given dataset

```
dat <-
read.table("C:/Users/Aravind/Desktop/CGS698C/recognition_f785f9bc-de94-4583-bb8f-25ae4cc8be87.csv",sep=",",header
= T)[,-1]
head(dat)
```

```
##          Tw      Tnw
## 1 285.0780 296.8060
## 2 267.5184 280.1157
## 3 289.9203 310.4417
## 4 399.0674 324.8276
## 5 359.9884 373.8152
## 6 403.3993 269.8220
```

Next we define the grid for the mu as well as delta values

```
mu_values <- seq(200, 400, length=40)
delta_grid <- seq(-100, 100, length=10000)
```

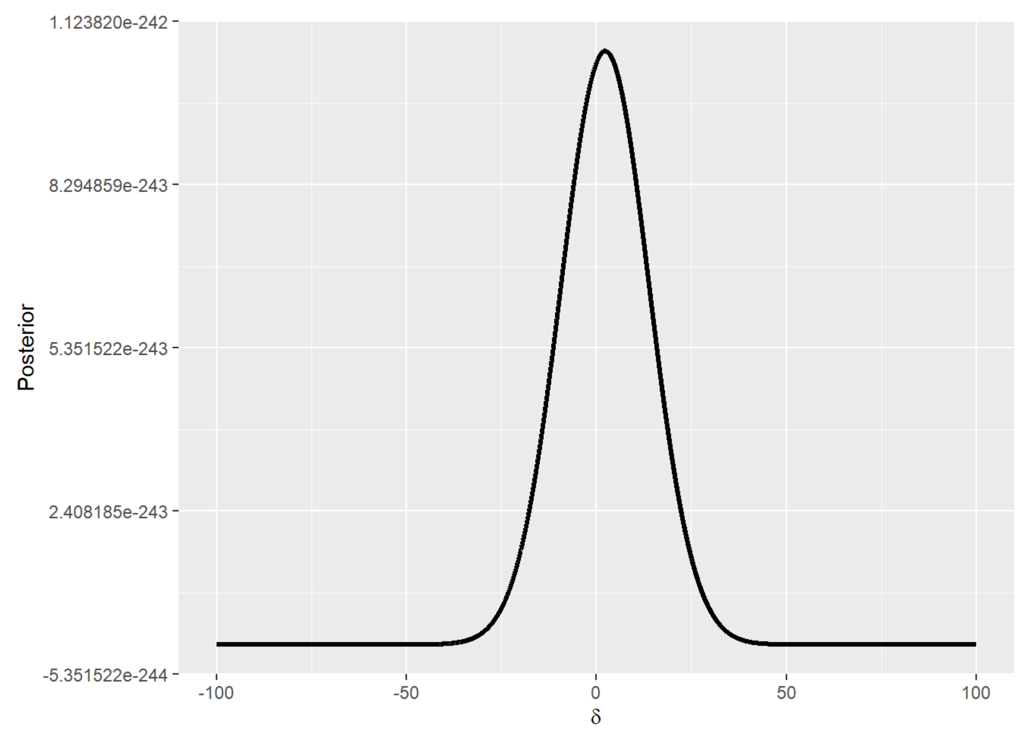
Now we need to integrate over all the mu values for getting the posterior distribution with delta

```
likelihood<-rep(0,each=length(delta_grid))
prior<-dnorm(delta_grid,0,50)
Tw=dat$Tw
Tnw=dat$Tnw
for(i in 1:length(delta_grid)){
 for(j in 1:length(mu_values)){
  likelihood[i]=likelihood[i]+prod(dnorm(Tw,mu_values[j],60))*

prod(dnorm(Tnw,mu_values[j]+delta_grid[i],60))*dnorm(mu_values[j],300,50)
 }
}
```

Lets plot this posterior distribution with the delta values

```
delta_posterior_grid<-likelihood*prior
grid_delta<-data.frame(delta_grid,delta_posterior_grid)
library(ggplot2)
ggplot(grid_delta,aes(x=delta_grid,y=delta_posterior_grid))+geom_line(size=1.2)+ xlab(expression(delta))+ylab("Po
sterior")
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## ℹ Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



2-For the next question we need to calculate the marginal likelihood using Monte Carlo Integration Let us initialize the mu and delta values

```
N<-10000
mu_values<-rep(0,each=N)
delta<-rep(0,each=N)
likelihood<-rep(0,each=N)
```

Now we need to independently sample from the priors and calculate the likelihood

```
for(i in 1:N){
 mu_values[i]<-rnorm(1,300,50)
 delta[i]<-rnorm(1,0,50)
 likelihood[i]<-
prod(dnorm(Tw,mu_values[i],60))*prod(dnorm(Tnw,mu_values[i]+delta[i],60))
}
```

Now we need to take the average of all the likelihoods to get the estimated Marginal Likelihood

```
ML<-mean(likelihood)
ML
```

```
## [1] 1.565406e-240
```

3- For Importance Sampling we take the proposal distribution as: μ~Normal(300,60) and δ~Normal(0,60)

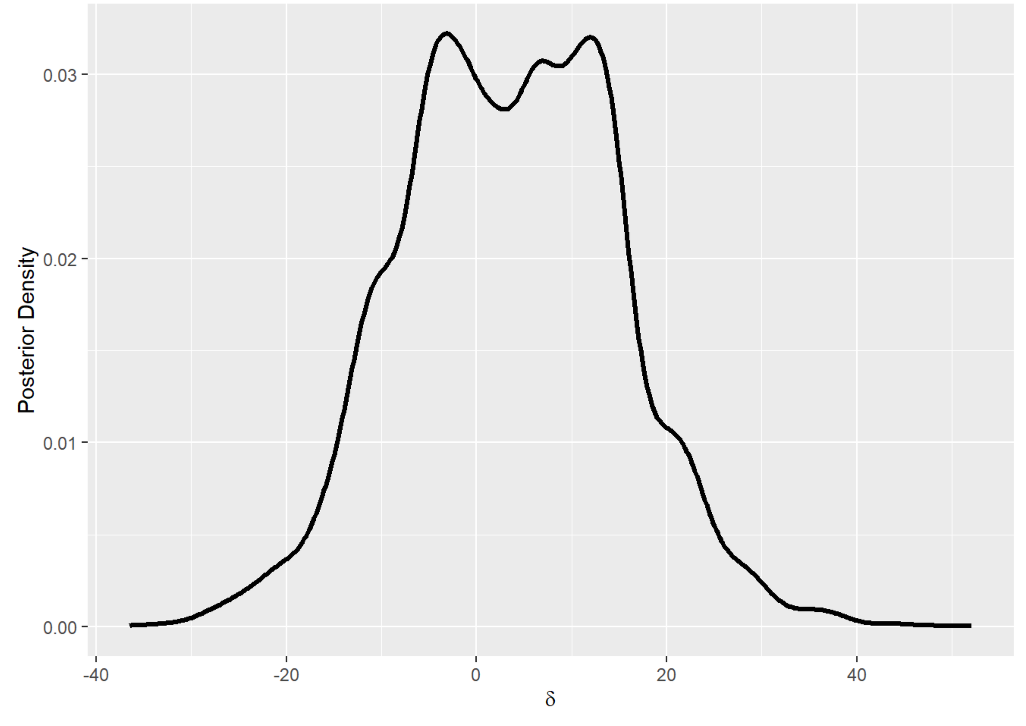
```
N<-10000
proposal_mu<-rnorm(N,300,60)
proposal_delta<-rnorm(N,0,60)
proposal_density_mu<-dnorm(proposal_mu,300,60)
proposal_density_delta<-dnorm(proposal_delta,0,60)
weight<-rep(0,each=N)
```

Lets compute the likelihoods,prior and weights

```
for(i in 1:N){
 likelihood<-prod(dnorm(Tw,proposal_mu[i],60))*
  prod(dnorm(Tnw,proposal_delta[i]+proposal_mu[i],60))
 prior<-dnorm(proposal_mu[i],300,50)*dnorm(proposal_delta[i],0,50)
 weight[i]<-
(likelihood*prior)/proposal_density_delta[i]*proposal_density_mu[i]
}
```

For drawing N/2 samples based on importance sampling

```
weight<-weight/sum(weight)
N_half<-N/2
post_samples_importance<-
sample(proposal_delta,size=N_half,replace=TRUE,prob=weight)
post_df<-data.frame(delta=post_samples_importance)
ggplot(post_df,aes(x=delta))+geom_density(size=1.2)+
 xlab(expression(delta))+
 ylab("Posterior Density")
```

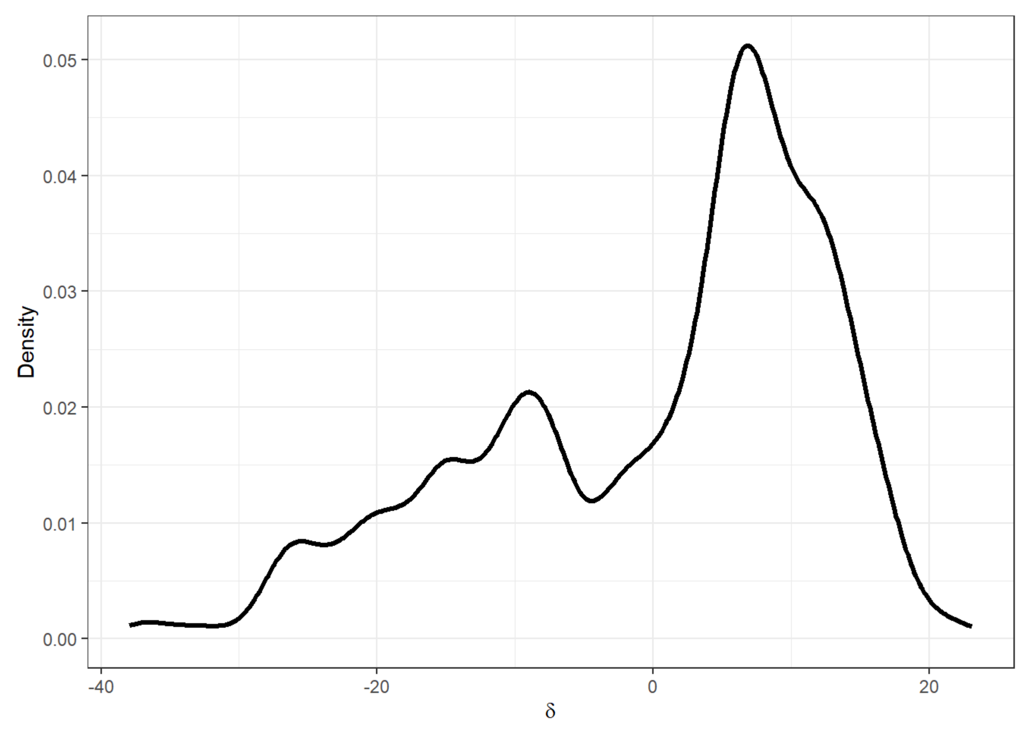


4-Lets use Markov Chain Monte Carlo

```
nsamp<-10000
mu_chain<-rep(NA,nsamp)
delta_chain<-rep(NA,nsamp)

mu_chain[1]<-rnorm(1,300,50)
delta_chain[1]<-rnorm(1,0,50)

i<-1
reject<-0
step<-0.5
while(i<nsamp){
 proposal_mu<-rnorm(1,mu_chain[i],step)
 proposal_delta<-rnorm(1,delta_chain[i],step)
 post_new<-prod(dnorm(Tw,proposal_mu,60))*
  prod(dnorm(Tnw,proposal_mu+proposal_delta,60))*
  dnorm(proposal_mu,300,50)*dnorm(proposal_delta,0,50)
 post_prev<-prod(dnorm(Tw,mu_chain[i],60))*
  prod(dnorm(Tnw,mu_chain[i]+delta_chain[i],60))*
  dnorm(mu_chain[i],300,50)*dnorm(delta_chain[i],0,50)
 Hastings_ratio<-(post_new*dnorm(mu_chain[i],proposal_mu,step)*dnorm(delta_chain[i],proposal_delta,step))/

(post_prev*dnorm(proposal_mu,mu_chain[i],step))*dnorm(proposal_delta,delta_chain[i],step)
 p_str<-min(Hastings_ratio,1)
 if(p_str>runif(1,0,1)){
  mu_chain[i+1]<-proposal_mu
  delta_chain[i+1]<-proposal_delta
  i<-i+1
 }else{
  reject<-reject+1
 }
}
```

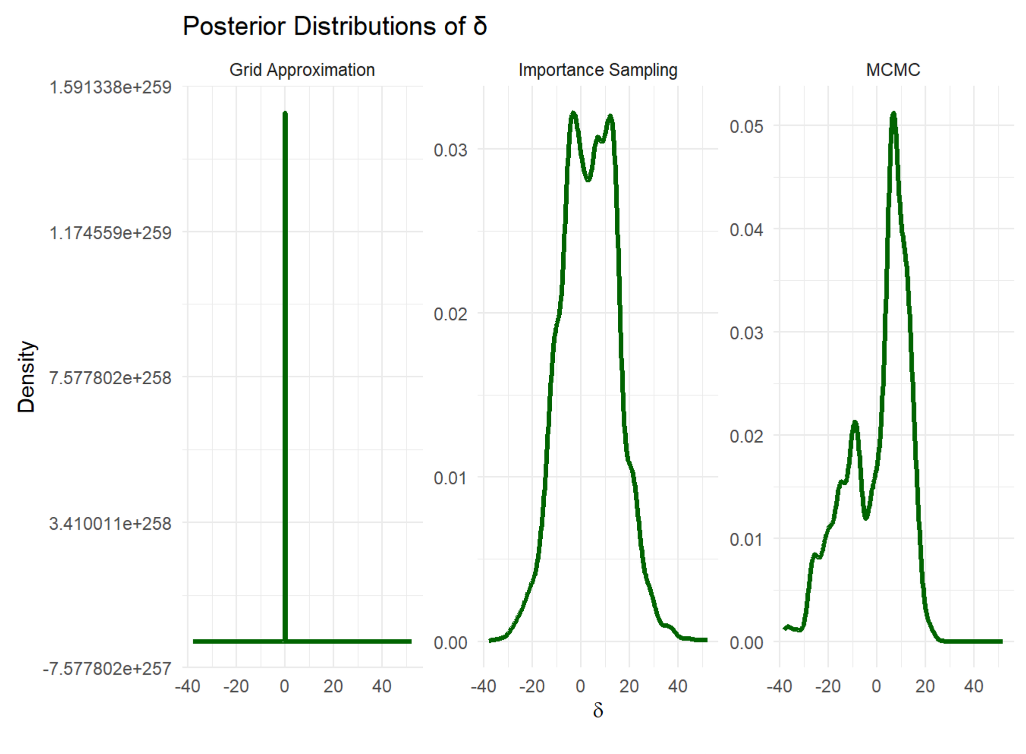
Now Lets plot the posterior density of delta

```
post<-data.frame(delta_chain)
ggplot(post,aes(x=delta_chain))+
 geom_density(size=1.2)+
 theme_bw()+xlab(expression(delta))+
 ylab("Density")
```



5- Comparing the different plots of Delta

```
delta_posteriors <- data.frame(
 delta = c(delta_posterior_grid, post_samples_importance, delta_chain),
 method = c(rep("Grid Approximation", length(delta_posterior_grid)),rep("Importance Sampling", length(post_sample
s_importance)),
 rep("MCMC", length(delta_chain)))
)
ggplot(delta_posteriors,aes(x=delta))+
 geom_density(size = 1.2, color = "darkgreen") + facet_wrap(~method, scales = "free_y") + xlab(expression(delta))
+ylab("Density") +
 theme_minimal() +
 ggtitle("Posterior Distributions of δ")
```



6-Yes,The MCMC posterior distribution provides stronger evidence for the lexical-access hypothesis. The majority of its density is concentrated in the negative region, indicating that recognition times for words are longer than for non-words.