Reproduciblity: Real Data Example: VCG Data

# Loading the R Package

library(RBVNF)

##   
## Attaching package: 'RBVNF'

## The following object is masked from 'package:base':  
##   
## norm

load\_packages()

## Loading required package: numDeriv

## Loading required package: MASS

## Loading required package: Rcpp

## Loading required package: RcppZiggurat

## Loading required package: RcppParallel

##   
## Attaching package: 'RcppParallel'

## The following object is masked from 'package:Rcpp':  
##   
## LdFlags

##   
## Rfast: 2.1.0

## \_\_\_ \_\_ \_\_ \_\_ \_\_ \_\_ \_\_ \_\_ \_\_ \_\_ \_ \_ \_\_ \_\_ \_\_ \_\_ \_\_ \_\_ \_\_ \_\_ \_\_ \_\_ \_\_   
## | \_\_ \_\_ \_\_ \_\_ | | \_\_ \_\_ \_\_ \_\_ \_/ / \ | \_\_ \_\_ \_\_ \_\_ / /\_\_ \_\_ \_ \_ \_\_ \_\_\   
## | | | | | | / \_ \ | | / /   
## | | | | | | / / \ \ | | / /   
## | | | | | | / / \ \ | | / /   
## | |\_\_ \_\_ \_\_ \_\_| | | |\_\_ \_\_ \_\_ \_\_ / / \ \ | |\_\_ \_\_ \_\_ \_\_ \_ / /\_\_/\   
## | \_\_ \_\_ \_\_ \_\_| | \_\_ \_\_ \_\_ \_\_| / /\_\_ \_ \_\_\ \ |\_ \_\_ \_\_ \_\_ \_ | / \_\_\_ /   
## | \ | | / \_ \_ \_ \_ \_ \_ \ | | \/ / /   
## | |\ \ | | / / \ \ | | / /   
## | | \ \ | | / / \ \ | | / /   
## | | \ \ | | / / \ \ | | / /   
## | | \ \\_\_ \_\_ \_ | | / / \ \ \_ \_\_ \_\_ \_\_ \_| | / /   
## |\_| \\_\_ \_\_ \_\_\ |\_| /\_/ \\_\ /\_ \_\_ \_\_ \_\_ \_\_\_| \/ team

## Loading required package: cowplot

##   
## Attaching package: 'mvtnorm'

## The following objects are masked from 'package:Rfast':  
##   
## Crossprod, dmvnorm, dmvt, rmvnorm, rmvt, Tcrossprod

## Loading required package: Matrix

## Loaded glmnet 4.1-8

load\_additional\_packages()

##   
## Attaching package: 'plotly'

## The following object is masked from 'package:ggplot2':  
##   
## last\_plot

## The following object is masked from 'package:MASS':  
##   
## select

## The following object is masked from 'package:stats':  
##   
## filter

## The following object is masked from 'package:graphics':  
##   
## layout

## Loading required package: dplyr

##   
## Attaching package: 'dplyr'

## The following object is masked from 'package:gridExtra':  
##   
## combine

## The following object is masked from 'package:Rfast':  
##   
## nth

## The following object is masked from 'package:MASS':  
##   
## select

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ forcats 1.0.0 ✔ stringr 1.5.0  
## ✔ lubridate 1.9.3 ✔ tibble 3.2.1  
## ✔ purrr 1.0.2 ✔ tidyr 1.3.0  
## ✔ readr 2.1.4   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::combine() masks gridExtra::combine()  
## ✖ tidyr::expand() masks Matrix::expand()  
## ✖ dplyr::filter() masks plotly::filter(), stats::filter()  
## ✖ purrr::is\_integer() masks Rfast::is\_integer()  
## ✖ dplyr::lag() masks stats::lag()  
## ✖ dplyr::nth() masks Rfast::nth()  
## ✖ tidyr::pack() masks Matrix::pack()  
## ✖ dplyr::select() masks plotly::select(), MASS::select()  
## ✖ lubridate::stamp() masks cowplot::stamp()  
## ✖ purrr::transpose() masks Rfast::transpose()  
## ✖ tidyr::unpack() masks Matrix::unpack()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

norm<-function(x){sqrt(sum(x^2))}  
######  
Effect\_of\_var<-function(beta\_EM, sel\_var ){  
  
 baseLine\_dir=beta\_EM[1,]/ norm(beta\_EM[1,])  
 beta\_sel\_dir<-beta\_EM[sel\_var, ]/norm(beta\_EM[sel\_var, ])  
  
  
 val\_return<-as.numeric(acos(t(baseLine\_dir)%\*%beta\_sel\_dir))  
 return(val\_return)  
}

########################################################  
 ######## Example 2######################################  
 #########################################################  
 VCG\_file="data/RealData/vcg.csv"  
 vcg\_data = read.csv(file = VCG\_file)

X1 = vcg\_data[,3:5]  
 X2 = vcg\_data[,6:8]  
  
 X1\_M = vcg\_data[,12:14]  
 X2\_M = vcg\_data[,15:17]  
  
 ##### Entire data  
 #output\_file = "./vcg\_data/vcg\_output.RData"  
  
 N = 98  
 YY = array(c(0,0,0,0,0,0), c(3, 2, N))  
 YY\_M = YY  
  
 for (i in 1:N){  
 YY[,1,i] = t(X1[i,])  
 YY[,2,i] = t(X2[i,])  
  
 YY\_M[,1,i] = t(X1\_M[i,])  
 YY\_M[,2,i] = t(X2\_M[i,])  
 }  
  
 data = YY;data\_M = YY\_M  
  
  
 XX<-as.data.frame(model.matrix( ~ as.factor(AgeSex)-1, data=vcg\_data ))  
  
 Gender= XX[, 3]+XX[,4] # Female =1  
 Age= XX[, 1]+XX[,3] # Age between 2 to 10 =1  
  
 Y=t(YY[,1,] )  
 X=as.matrix(cbind(Intercept = replicate(length(Age), 1), Gender=Gender, Age=Age, Gender\_Age\_Int=Gender\*Age ))  
 #X= cbind(replicate(dim(Y)[1], 1),as.integer(Maps$Maze)-1, as.integer(Maps$Trial.type)-1, Maps$L.c)

# Preparation of the data for Regression:

SetMcLength=2000;  
SetMcLength=max(SetMcLength, 2000)  
  
 n=dim(Y)[1] # NUmber of the samples  
 p=dim(X)[2] # NUmber of the regression covariates  
 d=dim(Y)[2] # Number of direcions in the direcional data  
 #### bbeta is a matrix of dimension p\times d  
 #bbeta=matrix( rnorm(p\*d), nrow=p, ncol=d)  
 sigma\_square=1  
 tau\_square=1000

# Running the Optimization Algorithm to obtain Posterior Mode:

Start\_time= Sys.time()  
 beta\_EM=EM\_Dir\_regression\_optimizer\_V1(Y=Y, X=X, prior=NULL, beta\_init = NULL, EM\_tolerence = .00001)

## [1] 2  
## [1] 3  
## [1] 4  
## [1] 5  
## [1] 6  
## [1] 7  
## [1] 8  
## [1] 9  
## [1] 10  
## [1] 11  
## [1] 12  
## [1] 13  
## [1] 14  
## [1] 15  
## [1] 16  
## [1] 17  
## [1] 18  
## [1] 19  
## [1] 20  
## [1] 21  
## [1] 22  
## [1] 23  
## [1] 24  
## [1] 25  
## [1] 26  
## [1] 27  
## [1] 28  
## [1] 29  
## [1] 30  
## [1] 31  
## [1] 32  
## [1] 33  
## [1] 34  
## [1] 35  
## [1] 36  
## [1] 37  
## [1] 38  
## [1] 39  
## [1] 40  
## [1] 41  
## [1] 42  
## [1] 43  
## [1] 44  
## [1] 45  
## [1] 46  
## [1] 47  
## [1] 48  
## [1] 49  
## [1] 50  
## [1] 51  
## [1] 52  
## [1] 53  
## [1] 54  
## [1] 55  
## [1] 56  
## [1] 57  
## [1] 58  
## [1] 59  
## [1] 60  
## [1] 61  
## [1] 62  
## [1] 63  
## [1] 64  
## [1] 65  
## [1] 66  
## [1] 67  
## [1] 68  
## [1] 69  
## [1] 70  
## [1] 71  
## [1] 72  
## [1] 73  
## [1] 74  
## [1] 75  
## [1] 76  
## [1] 77  
## [1] 78  
## [1] 79  
## [1] 80  
## [1] 81  
## [1] 82  
## [1] 83  
## [1] 84  
## [1] 85  
## [1] 86  
## [1] 87  
## [1] 88  
## [1] 89  
## [1] 90  
## [1] 91  
## [1] 92  
## [1] 93  
## [1] 94  
## [1] 95  
## [1] 96  
## [1] 97  
## [1] 98  
## [1] 99  
## [1] 100  
## [1] 101  
## [1] 102  
## [1] 103  
## [1] 104  
## [1] 105  
## [1] 106  
## [1] 107  
## [1] 108  
## [1] 109  
## [1] 110  
## [1] 111  
## [1] 112  
## [1] 113  
## [1] 114  
## [1] 115  
## [1] 116  
## [1] 117  
## [1] 118  
## [1] 119  
## [1] 120  
## [1] 121  
## [1] 122  
## [1] 123  
## [1] 124  
## [1] 125  
## [1] 126  
## [1] 127  
## [1] 128  
## [1] 129  
## [1] 130  
## [1] 131  
## [1] 132  
## [1] 133  
## [1] 134  
## [1] 135  
## [1] 136  
## [1] 137  
## [1] 138  
## [1] 139  
## [1] 140  
## [1] 141  
## [1] 142  
## [1] 143  
## [1] 144  
## [1] 145  
## [1] 146  
## [1] 147  
## [1] 148  
## [1] 149  
## [1] 150  
## [1] 151  
## [1] 152  
## [1] 153  
## [1] 154  
## [1] 155  
## [1] 156  
## [1] 157  
## [1] 158  
## [1] 159  
## [1] 160  
## [1] 161  
## [1] 162  
## [1] 163  
## [1] 164  
## [1] 165  
## [1] 166  
## [1] 167

print("Estimated Posterior Mode:=")

## [1] "Estimated Posterior Mode:="

colnames(beta\_EM)=gsub("Y\_", "Beta\_Y\_", colnames(beta\_EM))  
print(beta\_EM)

## Beta\_Y\_1 Beta\_Y\_2 Beta\_Y\_3  
## Intercept 5.0500146 6.5213712 6.008248  
## Gender 2.0992814 5.0777478 3.365241  
## Age 0.3731888 -0.6799576 -1.690946  
## Gender\_Age\_Int -1.8891929 -5.0347649 -4.048214

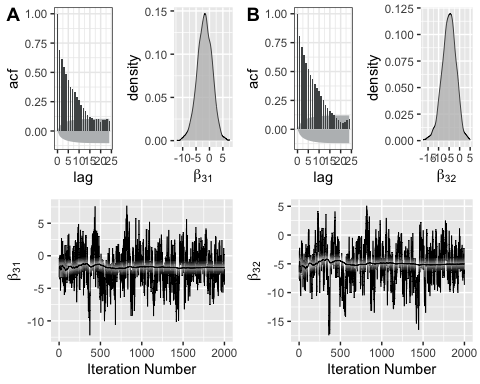
# Running the Data Augmentation- MCMC Algorithm to obtain Posterior Mode:

lst=MCMC\_Dir\_regression\_sampler\_V1(Y=Y, X=X, prior=NULL, beta\_init = NULL, MCSamplerSize =SetMcLength)

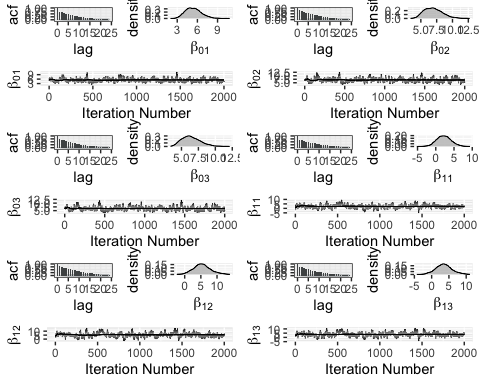
## [1] "Default Procedure using EM is being used to obtain initial value of the regression coefficients that will be used to start the MCMC Data Augmentation Algorithm. Iteration number of EM algorithm is being printed untill convergence."  
## [1] 2  
## [1] 3  
## [1] 4  
## [1] 5  
## [1] 6  
## [1] 7  
## [1] 8  
## [1] 9  
## [1] 10  
## [1] 11  
## [1] 12  
## [1] 13  
## [1] 14  
## [1] 15  
## [1] 16  
## [1] 17  
## [1] 18  
## [1] 19  
## [1] 20  
## [1] 21  
## [1] 22  
## [1] 23  
## [1] 24  
## [1] 25  
## [1] 26  
## [1] 27  
## [1] 28  
## [1] 29  
## [1] 30  
## [1] 31  
## [1] 32  
## [1] 33  
## [1] 34  
## [1] 35  
## [1] 36  
## [1] 37  
## [1] 38  
## [1] 39  
## [1] 40  
## [1] 41  
## [1] 42  
## [1] 43  
## [1] 44  
## [1] 45  
## [1] 46  
## [1] 47  
## [1] 48  
## [1] 49  
## [1] 50  
## [1] 51  
## [1] 52  
## [1] 53  
## [1] 54  
## [1] 55  
## [1] 56  
## [1] 57  
## [1] 58  
## [1] 59  
## [1] 60  
## [1] 61  
## [1] 62  
## [1] 63  
## [1] 64  
## [1] 65  
## [1] 66  
## [1] 67  
## [1] 68  
## [1] 69  
## [1] 70  
## [1] 71  
## [1] 72  
## [1] 73  
## [1] 74  
## [1] 75  
## [1] 76  
## [1] 77  
## [1] 78  
## [1] 79  
## [1] 80  
## [1] 81  
## [1] 82  
## [1] 83  
## [1] 84  
## [1] 85  
## [1] 86  
## [1] 87  
## [1] 88  
## [1] 89  
## [1] 90  
## [1] 91  
## [1] 92  
## [1] 93  
## [1] 94  
## [1] 95  
## [1] 96  
## [1] 97  
## [1] 98  
## [1] 99  
## [1] 100  
## [1] 101  
## [1] 102  
## [1] 103  
## [1] 104  
## [1] 105  
## [1] 106  
## [1] 107  
## [1] 108  
## [1] 109  
## [1] 110  
## [1] 111  
## [1] 112  
## [1] 113  
## [1] 114  
## [1] 115  
## [1] 116  
## [1] 117  
## [1] 118  
## [1] 119  
## [1] 120  
## [1] 121  
## [1] 122  
## [1] 123  
## [1] 124  
## [1] 125  
## [1] 126  
## [1] 127  
## [1] 128  
## [1] 129  
## [1] 130  
## [1] 131  
## [1] 132  
## [1] 133  
## [1] 134  
## [1] 135  
## [1] 136  
## [1] 137  
## [1] 138  
## [1] 139  
## [1] 140  
## [1] 141  
## [1] 142  
## [1] 143  
## [1] 144  
## [1] 145  
## [1] 146  
## [1] 147  
## [1] 148  
## [1] 149  
## [1] 150  
## [1] 151  
## [1] 152  
## [1] 153  
## [1] 154  
## [1] 155  
## [1] 156  
## [1] 157  
## [1] 158  
## [1] 159  
## [1] 160  
## [1] 161  
## [1] 162  
## [1] 163  
## [1] 164  
## [1] 165  
## [1] 166  
## [1] 167  
## [1] " Initial value and prior information obtained successfully. The MCMC samples are being generated. This step may take significnt amount of time depending on the MCMC sample size to be Generated. "  
## [1] "MC\_Iter=100completed"  
## [1] "MC\_Iter=200completed"  
## [1] "MC\_Iter=300completed"  
## [1] "MC\_Iter=400completed"  
## [1] "MC\_Iter=500completed"  
## [1] "MC\_Iter=600completed"  
## [1] "MC\_Iter=700completed"  
## [1] "MC\_Iter=800completed"  
## [1] "MC\_Iter=900completed"  
## [1] "MC\_Iter=1000completed"  
## [1] "MC\_Iter=1100completed"  
## [1] "MC\_Iter=1200completed"  
## [1] "MC\_Iter=1300completed"  
## [1] "MC\_Iter=1400completed"  
## [1] "MC\_Iter=1500completed"  
## [1] "MC\_Iter=1600completed"  
## [1] "MC\_Iter=1700completed"  
## [1] "MC\_Iter=1800completed"  
## [1] "MC\_Iter=1900completed"  
## [1] "MC\_Iter=2000completed"

# Plotting and Related Output:

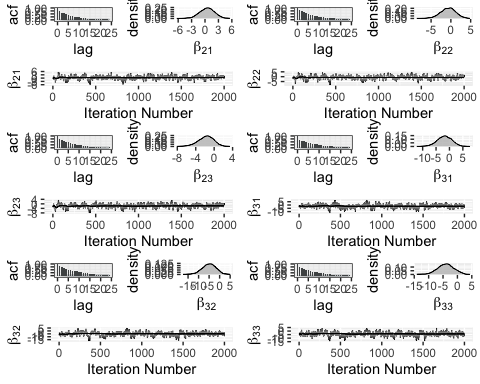
i=4;j= 1  
 p1=Plot\_MCMC\_Diag\_Triplet(lst$MC$Mc\_Beta[,i,j],y\_lab\_text = bquote(beta[.(i-1)][.(j)]))  
 i=4;j= 2  
 p2=Plot\_MCMC\_Diag\_Triplet(lst$MC$Mc\_Beta[,i,j],y\_lab\_text = bquote(beta[.(i-1)][.(j)]))  
  
 library(cowplot)  
  
  
  
  
#pdf(file="/Users/subhadippal/Dropbox/projects/Regression of Directional data/DirReg\_WriteUpShared/fig/Plot\_VCG\_beta31\_beta32\_TripletPlot1.pdf", width = 10, height= 5)  
 plot\_grid(p1, p2, labels = c('A', 'B'))



#dev.off()  
  
  
  
 ### All plots for appendix  
 #Set1  
  
 i=1;j= 1  
 p1=Plot\_MCMC\_Diag\_Triplet(lst$MC$Mc\_Beta[,i,j],y\_lab\_text = bquote(beta[.(i-1)][.(j)]))  
 j=2  
 p2=Plot\_MCMC\_Diag\_Triplet(lst$MC$Mc\_Beta[,i,j],y\_lab\_text = bquote(beta[.(i-1)][.(j)]))  
 j= 3  
 p3=Plot\_MCMC\_Diag\_Triplet(lst$MC$Mc\_Beta[,i,j],y\_lab\_text = bquote(beta[.(i-1)][.(j)]))  
 i=2;j=1  
 p4=Plot\_MCMC\_Diag\_Triplet(lst$MC$Mc\_Beta[,i,j],y\_lab\_text = bquote(beta[.(i-1)][.(j)]))  
 j=2  
 p5=Plot\_MCMC\_Diag\_Triplet(lst$MC$Mc\_Beta[,i,j],y\_lab\_text = bquote(beta[.(i-1)][.(j)]))  
 j= 3  
 p6=Plot\_MCMC\_Diag\_Triplet(lst$MC$Mc\_Beta[,i,j],y\_lab\_text = bquote(beta[.(i-1)][.(j)]))  
  
  
  
 # pdf(file="/Users/subhadippal/Dropbox/projects/Regression of Directional data/DirReg\_WriteUpShared/fig/Plot\_VCG\_Set1\_TripletPlot1.pdf", width = 10, height= 15)  
 plot\_grid(p1, p2,p3,p4,p5,p6, ncol = 2)



# dev.off()  
  
  
  
 #Set2  
 i=3;j= 1  
 p1=Plot\_MCMC\_Diag\_Triplet(lst$MC$Mc\_Beta[,i,j],y\_lab\_text = bquote(beta[.(i-1)][.(j)]))  
 j=2  
 p2=Plot\_MCMC\_Diag\_Triplet(lst$MC$Mc\_Beta[,i,j],y\_lab\_text = bquote(beta[.(i-1)][.(j)]))  
 j= 3  
 p3=Plot\_MCMC\_Diag\_Triplet(lst$MC$Mc\_Beta[,i,j],y\_lab\_text = bquote(beta[.(i-1)][.(j)]))  
 i=4;j=1  
 p4=Plot\_MCMC\_Diag\_Triplet(lst$MC$Mc\_Beta[,i,j],y\_lab\_text = bquote(beta[.(i-1)][.(j)]))  
 j=2  
 p5=Plot\_MCMC\_Diag\_Triplet(lst$MC$Mc\_Beta[,i,j],y\_lab\_text = bquote(beta[.(i-1)][.(j)]))  
 j= 3  
 p6=Plot\_MCMC\_Diag\_Triplet(lst$MC$Mc\_Beta[,i,j],y\_lab\_text = bquote(beta[.(i-1)][.(j)]))  
  
  
  
 #pdf(file="/Users/subhadippal/Dropbox/projects/Regression of Directional data/DirReg\_WriteUpShared/fig/Plot\_VCG\_Set2\_TripletPlot1.pdf", width = 10, height= 15)  
 plot\_grid(p1, p2,p3,p4,p5,p6, ncol = 2)



# dev.off()

# Estimated Parameters

Beta\_est<-apply(lst$MC$Mc\_Beta, MARGIN = c(2,3), FUN = mean)  
  
 Beta\_sd<-apply(lst$MC$Mc\_Beta, MARGIN = c(2,3), FUN = sd)  
 Beta\_est<- matrix(paste0(round(c(Beta\_est),2),"(", round(c(Beta\_sd),2),")"), nrow=4)  
 #write.csv(Beta\_est, file="/Users/subhadippal/Dropbox/projects/Regression of Directional data/VCG\_Beta\_est.csv")  
 print(Beta\_est)

## [,1] [,2] [,3]   
## [1,] "5.4(1.17)" "6.98(1.44)" "6.44(1.36)"   
## [2,] "2.23(1.96)" "5.36(2.64)" "3.53(2.33)"   
## [3,] "0.46(1.65)" "-0.72(1.92)" "-1.79(1.71)"  
## [4,] "-1.81(2.73)" "-5.04(3.31)" "-4.07(2.83)"

i=4; j=1  
 xx<-(lst$MC$Mc\_Beta[1001:SetMcLength, i, j])  
 round(c(quantile(xx, .025), quantile(xx, .975)),2)

## 2.5% 97.5%   
## -7.12 2.84

End\_time= Sys.time()  
 print(paste0("Total Run Time=",End\_time- Start\_time, "minutes" ))

## [1] "Total Run Time=1.74406028588613minutes"