within and between trial variation addition

Michael Seo

5 Oktober 2019

setwd("~/GitHub/phd/varselect")  
#setwd("C:/Users/mike/Desktop/Github/phd/varselect")  
  
library(MASS) # Used for data simulation  
  
library(lme4) #for fitting glmm

## Warning: package 'lme4' was built under R version 3.5.3

## Loading required package: Matrix

library(glmnet) #for lasso

## Warning: package 'glmnet' was built under R version 3.5.3

## Loading required package: foreach

## Warning: package 'foreach' was built under R version 3.5.3

## Loaded glmnet 2.0-18

source("run.simulation.functions.modified2.R")  
source("helpful.functions.R")  
  
### parameters to change ###  
col\_labels <- c(paste0("X", 1:10, ":treat"), "treat")  
col\_labels\_glmmLasso <- c(paste0("X", 1:10, "\_treat"), "treat")  
  
correct\_em <- c(0, 0, 0, 1, 1, 0, 0, 0, 0, 0)  
correct\_em\_values <- c(0, 0, 0, 0.3, 0.5, 0, 0, 0, 0, 0)  
  
Nstudies <- 5  
Ncovariate <- 10  
continuous.cov <- c(1, 2, 4, 6, 7, 8)  
pf <- c(1,2,3,4,5)  
em <- c(4,5)  
b1 <- c(0.1, 0.1, 0.5, 0.3, 0.5)  
b2 <- c(0.3, 0.5)  
  
glmm\_oracle\_formula <- as.formula("y ~ studyid + X1 + X2 + X3 + X4 + X5 + X4 \* treat + X5 \* treat + treat + (-1 + treat|studyid)")  
glmm\_full\_formula <- as.formula("y ~ studyid + (X1 + X2 + X3 + X4 + X5 + X6 + X7 + X8 + X9 + X10)\*treat + (-1 + treat|studyid)")  
step\_full\_formula <- as.formula("y ~ studyid + (X1 + X2 + X3 + X4 + X5 + X6 + X7 + X8 + X9 + X10)\*treat")  
glmmLasso\_formula <- as.formula("y ~ as.factor(studyid) + X1 + X2 + X3 + X4 + X5 + X6 + X7 + X8 + X9 + X10 + X1\_treat + X2\_treat + X3\_treat + X4\_treat + X5\_treat + X6\_treat + X7\_treat + X8\_treat + X9\_treat + X10\_treat + treat")  
  
model.type = "gaussian"  
  
  
data <-generate.simulation(Nstudies = Nstudies, Ncovariate = Ncovariate, continuous.cov = continuous.cov, pf = pf, em = em, b1 = b1, b2 = b2, model.type = model.type)  
  
library(dplyr)

## Warning: package 'dplyr' was built under R version 3.5.3

##   
## Attaching package: 'dplyr'

## 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

library(tibble)

## Warning: package 'tibble' was built under R version 3.5.3

newdata1 <-  
 as.data.frame(data[,c(2:11, 23)] %>%  
 group\_by(studyid) %>%  
 mutate\_all(funs(Z = .- mean(.), Zbar = mean(.))))

## `mutate\_all()` ignored the following grouping variables:  
## Column `studyid`  
## Use `mutate\_at(df, vars(-group\_cols()), myoperation)` to silence the message.

## Warning: funs() is soft deprecated as of dplyr 0.8.0  
## please use list() instead  
##   
## # Before:  
## funs(name = f(.))  
##   
## # After:   
## list(name = ~ f(.))  
## This warning is displayed once per session.

newdata1 <- cbind(newdata1, treat = data$treat, y = data$y)  
newdata2 <- model.matrix(y~ studyid + X1 + X2 + X3 + X4 + X5 + X6 + X7 + X8 + X9 + X10 +   
 (X1\_Zbar + X2\_Zbar + X3\_Zbar + X4\_Zbar + X5\_Zbar + X6\_Zbar + X7\_Zbar + X8\_Zbar + X9\_Zbar + X10\_Zbar):treat +   
 (X1\_Z + X2\_Z + X3\_Z + X4\_Z + X5\_Z + X6\_Z + X7\_Z + X8\_Z + X9\_Z + X10\_Z):treat + treat, data = newdata1)  
newdata2 <- newdata2[,-1]   
newdata2 <- cbind(y = data$y, newdata2= newdata2)   
  
#penalize only within study variation   
p.fac <- c(rep(0, Nstudies - 1), rep(0, Ncovariate), 0, rep(0, Ncovariate), rep(1, Ncovariate))  
  
family <- ifelse(model.type == "gaussian", "gaussian", "binomial")  
cvfit <- cv.glmnet(as.matrix(newdata2[,-1]), as.matrix(newdata2[,1]), penalty.factor = p.fac, family = family, standardize = FALSE, type.measure = "deviance")   
aa <- coef(cvfit, s = "lambda.min")  
aa

## 36 x 1 sparse Matrix of class "dgCMatrix"  
## 1  
## (Intercept) -0.09496795  
## studyid2 0.90188730  
## studyid3 -0.51841724  
## studyid4 -0.05270306  
## studyid5 0.60955760  
## X1 0.14685369  
## X2 0.12943860  
## X3 0.52074812  
## X4 0.30320249  
## X5 0.58709280  
## X6 0.02469916  
## X7 -0.03923289  
## X8 0.04195872  
## X9 0.01296768  
## X10 -0.03155997  
## treat 1.02833114  
## X1\_Zbar:treat 0.46093567  
## X2\_Zbar:treat -0.27794603  
## X3\_Zbar:treat 1.94568407  
## X4\_Zbar:treat 0.04658924  
## X5\_Zbar:treat -0.38635198  
## X6\_Zbar:treat -0.42048725  
## X7\_Zbar:treat 0.56011288  
## X8\_Zbar:treat -1.02508246  
## X9\_Zbar:treat 0.65587838  
## X10\_Zbar:treat -0.14915627  
## treat:X1\_Z .   
## treat:X2\_Z .   
## treat:X3\_Z .   
## treat:X4\_Z 0.29432492  
## treat:X5\_Z 0.35626969  
## treat:X6\_Z .   
## treat:X7\_Z .   
## treat:X8\_Z -0.07867843  
## treat:X9\_Z .   
## treat:X10\_Z .

#penalize both within study variation and between study variation  
p.fac <- c(rep(0, Nstudies - 1), rep(0, Ncovariate), 0, rep(1, Ncovariate), rep(1, Ncovariate))  
  
cvfit <- cv.glmnet(as.matrix(newdata2[,-1]), as.matrix(newdata2[,1]), penalty.factor = p.fac, family = family, standardize = FALSE, type.measure = "deviance")   
aa <- coef(cvfit, s = "lambda.min")  
aa

## 36 x 1 sparse Matrix of class "dgCMatrix"  
## 1  
## (Intercept) -0.237472419  
## studyid2 1.026165833  
## studyid3 -0.331527995  
## studyid4 0.219979452  
## studyid5 0.673011175  
## X1 0.143831617  
## X2 0.126625412  
## X3 0.522331669  
## X4 0.280058878  
## X5 0.559915499  
## X6 0.031087174  
## X7 -0.034906085  
## X8 0.071107024  
## X9 0.007796418  
## X10 -0.043999006  
## treat 1.033341546  
## X1\_Zbar:treat .   
## X2\_Zbar:treat .   
## X3\_Zbar:treat 1.087889003  
## X4\_Zbar:treat .   
## X5\_Zbar:treat .   
## X6\_Zbar:treat .   
## X7\_Zbar:treat .   
## X8\_Zbar:treat .   
## X9\_Zbar:treat .   
## X10\_Zbar:treat .   
## treat:X1\_Z 0.009667298  
## treat:X2\_Z .   
## treat:X3\_Z -0.009222194  
## treat:X4\_Z 0.334874500  
## treat:X5\_Z 0.402023987  
## treat:X6\_Z 0.002492273  
## treat:X7\_Z .   
## treat:X8\_Z -0.130533499  
## treat:X9\_Z .   
## treat:X10\_Z 0.036316316