### download data

### metabolite data preprocessing

library(readxl)

library(mvtnorm)

library(NORTARA)

# download real metabolic data as pilot data.#

truedata <- read.csv(file.choose())

truemat <- as.matrix(t(truedata))

#data preprocessing:#

#1.An offset is added to avoid negative values.#

truemat <- truemat + 1

#To efficiently simulate the long-tailed distributions and correlations. #

# present in biochemical measurements, pilot data are log-transformed. #

truemat <- log(truemat)

#To estimate mean and covariance from the log-transformed pilot data. #

mu <- apply(truemat, 2, mean)

Sigma <- cov(truemat)

#To generate new data using 'rmvnorm' function. #

N <- 100

simdata <- rmvnorm(N, mean = mu, sigma = Sigma)

#New data are then exponentiated to the original scale.#

simdata <- exp(simdata)

#The offset is subtracted to generate the final simulated data set corresponding to a multivariate log-normal distribution.#

simdata <- simdata - 1

#A small number of remaining negative values are set to zero.#

simdata[simdata < 0] <- 0

# To estimate mean and covariance from the simulated data.#

newmu <- apply(simdata, 2, mean)

sigma <- cov(simdata)

##################################################################

#m means variable number#

#n means sample number#

# set up correlation coefficient matrix#

#normalize or not#

gen\_Normal\_Negbin <- function(m,n,cor\_mat)

{

library(mvtnorm)

library(NORTARA)

u<-c(19,66)

p<-m/2

#set up new null matrix#

nor.matrix<-matrix(nrow=m,ncol=n)

neg.matrix<-matrix(nrow=m,ncol=n)

#specified the name of the dataset#

invcdfnames <- c(rep(c("qnorm","qnbinom"),each=p))

#Specify the parameters, where the normal distribution is approximate to the simulation data from the pilot data and the negative binomial distribution NB (20, 0.5) #

paramslists <- list()

#Set parameters including mean and variance for matrix A#

m1 = list()

for(i in 1:p) {

m1[[i]] <- list(mean = newmu[i], sd = sqrt(diag(sigma)[i]))

}

#Set parameters for the negative binomial distribution#

m3 = list(size = 20, prob = 0.5)

m3\_p <- rep(list(m3), p)

#Parameters are merged together#

paramslists <- c(m1\_p, m3\_p)

nth <- paste0(c("m"),1:m)

names(paramslists)<-nth

#specified the correlation coefficient matrix#  
cor\_matrix <- matrix(cornum,m,m)  
diag(cor\_matrix ) <- 1

#Generate the dataset. #

res <- genNORTARA(n, cor\_mat, invcdfnames, paramslists)

#Extract the multivariate normal distribution matrix#

norm <- res[,1:p]

#Extract the negative binomial distribution matrix#

negbin <- res[,(p+1):m]

#Output#

resu<-list()

resu[["norm"]]<-norm

resu[["negbin"]]<-negbin

return(resu)

}

#-0.4#

results\_0.4<-gen\_Normal\_Negbin(4,100,-0.4)

matrix\_norm\_0.4<-results\_0.4$norm

matrix\_dirichlet\_0.4<-results\_0.4$dirichlet