Package 'BAGEL'

June 30, 2020

Type Package

Title BAGEL: A Bayesian Graphical Model for Inferring Drug Effect Longitudinally on Depression in People with HIV

Version 0.1.0

Description This package implements BAGEL, a Bayesian graphical model to investigate the longitudinal effects of ART drugs on a range of depression symptoms while adjusting for participants' demographic, behavior, and clinical characteristics, and taking into account the heterogeneous population through a Bayesian nonparametric prior.

License GPL-3

Encoding UTF-8

LazyData true

LinkingTo Rcpp, RcppArmadillo, RcppDist

Imports Rcpp

RoxygenNote 7.0.2

R topics documented:

---

depression_prob_predict

The prediction function

Description

This function returns the predicted probability of reporting depression for each depression item for a new visit, provided covariates, ART usage information, and previous visit time

Usage

```
depression_prob_predict(mcmc, z_new, x_new, seed = 1)
```

Arguments

`mcmc`	the MCMC output given by main_mcmc
`z_new`	the binary indicator matrix for drug usage information for new visit to predict
`x_new`	the covariates matrix for new visit to predict, it also includes the patient ID (1st column) and visit time (2nd column)
`seed`	the starting number used to generate random numbers

Value

the predicted probability of reporting depression for each depression item

See Also

See main_mcmc for an example.

---

generate_simulate_data

The function generates simulated data

Description

This function generates simulated data, provided the scale of the dataset

Usage

```
generate_simulate_data(
  D,
  Q,
  S,
  N,
  H,
  B = 10,
  rho = 0.1,
  prob = NULL,
  f = NULL,
  Comega = NULL,
  a = c(-100, 0, 1, 2, 100),
  sigma_square = 1,
  seed = 1
)
```

Arguments

`D`	the number of drugs
`Q`	the number of depression scores
`S`	the number of covariates
`N`	the number of patients in the dataset
`H`	the number of clusters

B	the number of bases for P-spline
rho	the prior probability of the binary indicator cube entries being 1
prob	the probabilities of patients being assigned to clusters. If not given, it will be equal probabilities.
f	the function designed to capture the longitudinally on depression. If not given, it will be constant 0
Comega	the correlation matrix which captures the dependencies among depression items. If not given, it will be an identity matrix
a	the thresholds that connects the latent and observed depression scores
sigma_square	the variance of noise for the latent continuous depression scores
seed	the starting number used to generate random numbers

Value

a simulated dataset with observed data and unobserved parameters

Examples

```
D = 5; Q=3; S=5; N=100;H=10
simulated_data <- generate_simulate_data(D, Q, S, N, H)
```

---

main_mcmc	*The function returns the posted-brun-in MCMC samples thinned by the thinning factor*

Description

This function returns the posted-brun-in MCMC samples thinned by the thinning factor, provided observed data and initial values for some parameters.

Usage

```
main_mcmc(Data, Burnin, Niter, thin = 1, hyper_parameter = list(), seed = 1)
```

Arguments

Data	a list giving observed data. It includes: U giving the depression scores, each row corresponds to one visit; Z giving the binary indicator matrix for drug usage information, each row corresponds to one visit; X giving the covariates matrix, it also includes the patient ID (1st column) and visit time (2nd column); D giving the number of drugs; Q giving the number of depression scores; S giving the number of covariates; N giving the number of patients in the dataset; B giving the number of bases for P-spline.
Burnin	number of burn-in iterations in MCMC
Niter	number of iterations after burn-in in MCMC
thin	the MCMC thinning factor
hyper_parameter	a list giving hyperparameters used in MCMC. See details below for more information.
seed	the starting number used to generate random numbers

Details

hyper_parameter is a list giving hyperparameters used in MCMC. It includes: Comega_init giving the intial value of Comega, which is the correlation matrix which captures the dependencies among depression items; rho_init giving the intial value of rho, which is the prior probability of the binary indicator cube entries being 1; H_init giving the intial value of H, which is the number of clusters; a_init giving the intial value of a, which is the thresholds that connects the latent and observed depression scores; e_init giving the intial value of e, which is the clustering membership vector; alpha_init giving the intial value of alpha, which is the linear coefficients dependent on drug effect (it should be a cube, each slice corresponds to one cluster, each slice should be a 'S' by 'DQ' matrix); gamma_init giving the intial value of gamma, which is the linear coefficients for the P-spline (it should be a cube, each slice corresponds to one cluster, each slice should be a 'B' by 'DQ' matrix); beta_init giving the intial value of beta, which is the linear coefficients not dependent on drug effect (it should be a cube, each slice corresponds to one cluster, each slice should be a 'Q' by '1+S' matrix with the first column being the intercept); R_init giving the intial value of R, which is the binary indicator cube (it should be a cube, each slice corresponds to one cluster, each slice should be a 'Q' by 'D' matrix); sigma_square_alpha_init giving the intial value of sigma_square_alpha, which is the variance of alpha; sigma_square_beta_init giving the intial value of sigma_square_beta, which is the variance of beta; sigma_square_gamma_init giving the intial value of sigma_square_gamma, which is the variance of gamma; sigma_square giving the variance of noise for the latent continuous depression scores; m0 giving the concentration parameter for Dirichlet Process; H_max giving the maximum number of clusters allowed; a_rho giving the first hyper-parameter of rho with default value being 1, where we assume rho~Beta(alpha, beta); b_rho giving the second hyper-parameter of rho; step_a giving the step size for Metropolis–Hastings algorithm updating a; step_Comega giving the step size for Metropolis–Hastings algorithm updating Comega; a_beta giving the first hyper-parameter of sigma_square_beta, where we assume variance of beta is sampled from inv-Gamma(a,b); b_beta giving the second hyper-parameter of sigma_square_beta; a_alpha giving the first hyper-parameter of sigma_square_alpha, where we assume variance of alpha is sampled from inv-Gamma(a,b); b_alpha giving the second hyper-parameter of sigma_square_alpha; a_gamma giving the first hyper-parameter of sigma_square_gamma, where we assume variance of gamma is sampled from inv-Gamma(a,b); b_gamma giving the second hyper-parameter of sigma_square_gamma.

Value

the posted-brun-in MCMC samples thinned by the thinning factor, along with the following: D the number of drugs; Q the number of depression scores; S the number of covariates; knots the knots for P-spline; B the number of bases for P-spline; sigma_square the variance of noise for the latent continuous depression scores; m0 the concentration parameter for Dirichlet Process.

Examples

```
## Not run:
library(BAGEL)
data("simulated_data")
mcmc <- main_mcmc(Data=simulated_data, 2,10,5)
x_new <- simulated_data$X[10,]
z_new <- simulated_data$Z[10,]
depression_prob_predict(mcmc, z_new, x_new)

## End(Not run)
```

---

simulated_data   *An example simulated dataset.*

Description

It is a list generated by function generate_simulate_data (the example code). We assumed 5 drugs, 3 depression items, 5 covariates, 10 bases, and 100 patients.

Usage

```
data(simulated_data)
```

Format

A list giving the following variables.

U   the depression scores, each row corresponds to one visit

Z   the binary indicator matrix for drug usage information, each row corresponds to one visit

X   the covariates matrix, it also includes the patient ID (1st column) and visit time (2nd column). Each row corresponds to one visit

D   the number of drugs

Q   the number of depression scores

S   the number of covariates

N   the number of patients in the dataset

B   the number of bases for P-spline

H   the number of clusters

beta   the linear coefficients not dependent on drug effect

alpha   the linear coefficients dependent on drug effect

gamma   the linear coefficients for the P-spline

R   the binary indicator cube that introduce sparsity

e   the clustering membership vector

a   the thresholds that connects the latent and observed depression scores

Comega   the correlation matrix which captures the dependencies among depression items

Examples

```
# output true clustering membership
data("simulated_data")
simulated_data$e

# get the details of the list
str(simulated_data)
```

Index