Package 'BayTetra'

August 31, 2023

Title BayTetra: A Bayesian Semiparametric Approach for Testing Trajectory Differences

Version 0.1.0

Author Wei Jin, Qiuxin Gao, Yanxun Xu <yanxun.xu@jhu.edu>

Maintainer Qiuxin Gao <qgao13@jh.edu>

Description BayTetra addresses an important task in the field of biomedical applications, testing differences in longitudinal trajectories among distinct groups of populations. The package offers a Bayesian semiparametric approach for modeling multivariate longitudinal data. It accounts for correlations among different responses and employs B-splines, along with spike-and-slab priors on the spline coefficients, for flexible and parsimonious trajectory estimation. The package is inspired by the research paper ``BayTetra — A Bayesian Semiparametric Approach for Testing Trajectory Differences" by Wei, J & Gao, Q & Xu, Y (2023).

License GPL-3

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.2.3

LinkingTo Rcpp(>= 1.0.5),RcppArmadillo(>= 0.12.2.0.0)

Imports MCMCpack(>= 1.6-3),MASS(>= 7.3-58.4),Rcpp(>= 1.0.10),splines(>= 4.3.0), dplyr(>= 1.1.2),tmvtnorm(>= 1.5),Matrix(>= 1.5-4),rstan(>= 2.20.0), mvtnorm(>= 1.1-1),truncnorm(>= 1.0-8),pracma(>= 2.3.3),loo(>= 2.6.0)

Depends R (>= 4.2)

SystemRequirements GNU make

NeedsCompilation yes

LazyData true

R topics documented:

ex_data *Example Longitudinal Data*

Description

Example longitudinal data

Usage

```
ex_data
```

Format

A dataframe with following variables:

- **ID**: Identity for individuals
- **VISIT**: Individuals' visit index
- **time**: Time variable
- **cov1**: Covariate
- **Group**: Group memberships for individuals
- **R1**: Response variable 1
- **R2**: Response variable 2

Examples

```
data(ex_data)
```

Generate_simulated_data *Generate Simulated Data*

Description

This function generates a simulated dataset used in Scenario #1 of the paper "BayTetra: A Bayesian Semiparametric Approach for Testing Trajectory Differences"

Usage

```
Generate_simulated_data()
```

Value

Longitudinal data with following variables:

- **ID**: Identity for individuals
- **VISIT**: Individuals' visit index
- **time**: Time variable
- **cov1**: Covariate
- **Group**: Group memberships for individuals
- **R1**: Response variable 1
- **R2**: Response variable 2

Examples

```
## Not run:
ex_data = Generate_simulated_data()
head(ex_data)

## End(Not run)
```

`mcmc_BayTetra` *Posterior inference for BayTetra*

Description

Draw posterior samples of the parameters of interest from BayTetra

Usage

```
mcmc_BayTetra(
  data,
  v_rsp,
  v_covs,
  v_grp,
  v_time,
  df,
  prior = list(),
  mcmc = list(),
  display_process = TRUE
)
```

Arguments

`data`	longitudinal data with ID, VISIT, Group, and Covariates, Responses, Time.
`v_rsp`	Column names corresponding to responses.
`v_covs`	Column names corresponding to covariates.
`v_grp`	Column name corresponding to group memberships.
`v_time`	Column name corresponding to time.
`df`	This parameter specifies the degree of freedom of B-spline and is used to select the number of interior knots. Default value is 4 and minimum value is 3. • df = 3: Function uses a degree 2 B-spline with 0 interior knots. • df = 4: Function uses a degree 3 B-spline with 0 interior knots. • df >= 5: Function uses a degree 3 B-spline with (df - 4) interior knots.
`prior`	A list giving the prior information. • `mu_alpha`: The mean in normal prior for α_q. Default value is a zero vector. • `V_alpha`: The covariance matrix in normal prior for α_q. Default value is 100 * I where I is the identity matrix. • `nu_0`: The hyperparameter ν_0 in prior for γ_{kq} and γ_{kq0}. Default value is 2.5e-5. • `a_nu`: The hyperparameter a_ν in prior for ν_{kq} and ν_{kq0}. Default value is 5.

- b_nu: The hyperparameter b_ν in prior for ν_{kq} and ν_{kq0}. Default value is 25.
- a_rho: The hyperparameter a_ρ in prior for ρ_0. Default value is 1.
- b_rho: The hyperparameter b_ρ in prior for ρ_0. Default value is 1.
- h_1: The hyperparameter a_σ in prior for σ_q^2. Default value is 1.
- h_2: The hyperparameter b_σ in prior for σ_q^2. Default value is 1.

mcmc A list giving the MCMC parameters.

- Nit: The number of iterations for the MCMC chain. Default is 4000.
- burn_in: The number of burn-in samples in the MCMC chain. Default is 2000.
- thin_factor: The thinning factor for the chain. Default is 10.

display_process
A bool value; if TRUE, progress will be displayed every 1000 iteration by default.

Details

The model of the BayTetra is:

$$y_{iqj} = \boldsymbol{Z}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_q + \sum_{l=1}^{L-1}\widetilde{\beta}_{Kql}\widetilde{\boldsymbol{B}}_l\left(t_{iqj}\right) + \sum_{k=1}^{K-1}\mathbb{I}\left(g_i = k\right)\left(\widetilde{\beta}_{kq0} + \sum_{l=1}^{L-1}\widetilde{\beta}_{kql}\widetilde{\boldsymbol{B}}_l\left(t_{iqj}\right)\right) + \omega_{iq} + \epsilon_{iqj},$$

$$\boldsymbol{\omega}_i = \left(\omega_{i1}, \ldots, \omega_{iQ}\right) \sim \mathcal{N}\left(\mathbf{0}, \Sigma_\omega\right), \epsilon_{iqj} \sim \mathcal{N}\left(0, \sigma_q^2\right),$$

where $\widetilde{\boldsymbol{B}}_l\left(t_{iqj}\right)$ denote the l-th basis function for the $L-1$ dimensional cubic B-spline expansion at time t_{iqj}.

We set $\widetilde{\beta}_{Kq0}$ to 0 for identifiability and denote $\widetilde{\boldsymbol{\beta}}_{kq} = \left(\widetilde{\beta}_{kq0}, \widetilde{\beta}_{kq1}, \ldots, \widetilde{\beta}_{kq,L-1}\right)^{\mathrm{T}} = \left(\widetilde{\beta}_{kq0}, \left(\widetilde{\boldsymbol{\beta}}_{kq}^{-}\right)^{\mathrm{T}}\right)^{\mathrm{T}}$.

We assign priors:

$$\widetilde{\boldsymbol{\beta}}_{kq}^{-} = \eta_{kq}\boldsymbol{\xi}_{kq},$$
$$\eta_{kq} \sim \mathcal{N}\left(0, \gamma_{kq}\nu_{kq}\right),$$
$$\gamma_{kq} \sim \rho\delta_1\left(\gamma_{kq}\right) + (1-\rho)\delta_{\nu_0}\left(\gamma_{kq}\right),$$
$$\nu_{kq} \sim \text{Inverse-Gamma}\left(a_\nu, b_\nu\right),$$
$$\rho \sim \text{Beta}(a_\rho, b_\rho), \xi_{kql} \sim \mathcal{N}(m_{kql}, 1),$$
$$m_{kql} \sim \frac{1}{2}\delta_1(m_{kql}) + \frac{1}{2}\delta_{-1}(m_{kql}), l = 1, 2, \ldots, L-1.$$

For the intercept $\widetilde{\beta}_{kq0}$, we assume its prior:

$$\widetilde{\beta}_{kq0} \sim \mathcal{N}\left(0, \gamma_{kq0}\nu_{kq0}\right),$$

$$\gamma_{kq0} \sim \rho_0\delta_1\left(\gamma_{kq0}\right) + \left(1-\rho_0\right)\delta_{\nu_0}\left(\gamma_{kq0}\right).$$

The prior of other parameters are:

$$\boldsymbol{\alpha}_q \sim \mathcal{N}\left(0, \Sigma_\alpha\right), p\left(\Sigma_\omega\right) \propto 1,$$

$$\sigma_q^2 \sim \text{ Inverse-Gamma } \left(a_\sigma, b_\sigma\right).$$

Value

An object of class 'Post_BayTetra' containing posterior samples:

- pos.alpha: Posterior samples for α_q.
- pos.beta: Posterior samples for $\widetilde{\boldsymbol{\beta}}_{kq}$.
- pos.gamma_kq: Posterior samples for γ_{kq}.
- pos.gamma_kq0: Posterior samples for γ_{kq0}.
- pos.Sigma_omega: Posterior samples for Σ_ω.
- pos.sigma2: Posterior samples for σ_q^2.

Examples

```
## Not run:
mcmc_result = mcmc_BayTetra(data = ex_data,
                            v_rsp = paste("R", 1:2,sep = ""),
                            v_covs = "cov1",
                            v_grp = "Group",
                            v_time = "time",
                            df = 5)

## End(Not run)
```

Model_selection_BayTetra

Model Selection for BayTetra

Description

Function implements the model selection of BayTetra for the degree of freedom in B-splines.

Usage

```
Model_selection_BayTetra(
  df_min,
  df_max,
  data,
  v_rsp,
  v_covs,
  v_grp,
  v_time,
  mcmc = list(),
  prior = list(),
  display_process = TRUE
)
```

Arguments

df_min	Minimum value for degrees of freedom (df). Must be 3 or greater. See mcmc_BayTetra documentation.
df_max	Maximum value for degrees of freedom (df).
data	See mcmc_BayTetra documentation.
v_rsp	See mcmc_BayTetra documentation.
v_covs	See mcmc_BayTetra documentation.
v_grp	See mcmc_BayTetra documentation.
v_time	See mcmc_BayTetra documentation.
mcmc	See mcmc_BayTetra documentation.
prior	See mcmc_BayTetra documentation.
display_process	See mcmc_BayTetra documentation.

Value

A list containing two elements:

- selection_elpd: A list containing a named vector of elpd values, which elements' names are the degrees of freedom. It also contains the optimal degree of freedom.
- best_mcmc_result: The best model's MCMC posterior samples (model with the highest elpd).

See Also

mcmc_BayTetra()

Examples

```
## Not run:
selection_result = Model_selection_BayTetra(
                         df_min = 4, df_max = 6,
                         data = df_new,
                         v_rsp = paste("R", 1:2, sep = ""),
                         v_covs = "cov1",
                         v_grp = "Group",
                         v_time = "time"
                         )

## End(Not run)
```

summary.Test_BayTetra *Summarize Results of BayTetra Hypothesis Test*

Description

Summarize result generated by the Test_BayTetra function in a tidy way

Usage

```
## S3 method for class 'Test_BayTetra'
summary(object, ...)
```

Arguments

object	An object of class "Test_BayTetra", typically the result of calling "Test_BayTetra()".
...	Additional arguments affecting the summary produced (currently not used).

Value

The function print the object.

See Also

Test_BayTetra The function that generates the "Test_BayTetra" object.

Examples

```
## Not run:
test_result <- Test_BayTetra(mcmc_result,
                             v_Rsp = paste("R", 1:2,sep = ""))
summary(test_result)

## End(Not run)
```

Test_BayTetra *Hypothesis Testing for BayTetra*

Description

This function implements hypothesis test based on BayTetra posterior samples.

Usage

```
Test_BayTetra(object, v_rsp)
```

Arguments

object	An object of class "Post_BayTetra" containing MCMC posterior samples from "mcmc_BayTetra".
v_rsp	A character vector of response variables.

Value

Return an object of class "Test_BayTetra" containing three elements:

- BayTetra_summary: A dataframe that summarizes the testing information for $\widetilde{\beta}_{kq0}$ and $\widetilde{\boldsymbol{\beta}}_{kq}^{-}$. The 'Pr(Signal)' column represents the confidence level $P(\gamma_{kq} = 1)$.
- pairwise_significance: A matrix indicating pairwise hypothesis testing results. For each pair of groups, their longitudinal trajectories are significantly different if the corresponding responses present in the table.
- diff_among_all_grps: A named logical vector indicating if there exists difference among all groups for each response.

Examples

```
## Not run:
mcmc_result = mcmc_BayTetra(data = ex_data,
                            v_rsp = paste("R", 1:2,sep = ""),
                            v_covs = "cov1",
                            v_grp = "Group",
                            v_time = "time",
                            df = 5)
test_result <- hypo_BayTetra(mcmc_result,
                           v_rsp = paste("R", 1:2,sep = ""))

## End(Not run)
```

Index