Package ‘wast’

May 4, 2023

Type Package

Title Change-plane testing in the generalized estimating equations

Version 1.0.1

Author Xu Liu [aut,cre]

Maintainer Xu Liu <liu.xu@sufe.edu.cn>

Description Provide a method to calculate p-value of the test statistic for subgroup detecting in in the framework of general estimating equation (EE). In the paper Liu (2022), we propose a novel U-like statistic by taking the weighted average over the nuisance parametric space. The proposed test statistics not only improve power, but also save dramatically computational time. Many common and useful models are considered, including models with change point or change plane. We propose a novel U-like test statistic to detect multiple change planes in the framework of EE.

License GPL (>= 2)

Depends R (>= 3.2.0)

LazyData true

NeedsCompilation yes

Repository github

URL https://github.com/xliusufe/wast

Encoding UTF-8

Archs i386, x64

R topics documented:

`wast-package` *Change-plane testing in the generalized estimating equations*

Description

Provide a method to calculate p-value of the test statistic for subgroup detecting in in the framework of general estimating equation (EE). In the paper Liu (2022), we propose a novel U-like statistic by taking the weighted average over the nuisance parametric space. The proposed test statistics not only improve power, but also save dramatically computational time. Many common and useful models are considered, including models with change point or change plane. We propose a novel U-like test statistic to detect multiple change planes in the framework of EE.

Details

Package:	wast
Type:	Package
Version:	1.0.1
Date:	2022-05-5
License:	GPL (>= 2)

References

Andrews, D. W. K. and Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative. Econometrica, 62(6):1383-1414.

Fan, A., Rui, S., and Lu, W. (2017). Change-plane analysis for subgroup detection and sample size calculation. Journal of the American Statistical Association, 112(518):769-778.

Huang, Y., Cho, J., and Fong, Y. (2021). Threshold-based subgroup testing in logistic regression models in two phase sampling designs. Journal of the Royal Statistical Society: Series C. 291-311.

Liu, X. (2023). Subgroup testing in change-plane regression with high-dimensional grouping variables. Manuscript.

`exams` *Examples for Subgroup Test in Generalized Linear Models*

Description

Examples for Family 'Gaussian', 'binomial', and 'Poisson'.

Usage

```
exams(family = "gaussian", method = "wast", tau = 0.5, B = 1000, K = 1000)
```

Arguments

`family`	Family for generalized linear models, including 'gaussian', 'binomial', 'poisson', 'probit', 'quantile', and 'semiparam'.
`method`	There are there methods, including the proposed 'wast', 'sst', and 'slrt'.
`tau`	The given quantile τ when family is 'quantile', a scale in the unit inteval. Default is `tau = 0.5`.
`B`	An integer, the number of bootstrap samples.
`K`	An integer, the number of threshold values for 'sst' and 'slrt'.

Value

`pvals`	P-value of the corresponding test statistic.

References

Andrews, D. W. K. and Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative. Econometrica, 62(6):1383-1414.

Fan, A., Rui, S., and Lu, W. (2017). Change-plane analysis for subgroup detection and sample size calculation. Journal of the American Statistical Association, 112(518):769-778.

Huang, Y., Cho, J., and Fong, Y. (2021). Threshold-based subgroup testing in logistic regression models in two phase sampling designs. Journal of the Royal Statistical Society: Series C. 291-311.

Liu, X. (2023). Subgroup testing in change-plane regression with high-dimensional grouping variables. Manuscript.

Examples

```
pvals <- exams(family = "gaussian", method = "wast")
pvals

pvals <- exams(family = "binomial", method = "wast")
pvals

pvals <- exams(family = "poisson", method = "wast")
pvals

pvals <- exams(family = "quantile", method = "wast", tau = 0.5)
pvals
```

`pvalglm` *P-value for Subgroup Test in Generalized Linear Models*

Description

Provide p-value for subgroup test in generalized linear models, including three methods 'wast', 'sst', and 'slrt'.

Usage

```
pvalglm(data, family = "gaussian", method = 'wast', B = 1000, K = 2000,
       isBeta = FALSE, shape1 = 1, shape2 = 1, N0 = 5000, MU = NULL, ZK = NULL)
```

Arguments

data	A list, including Y (response), X (baseline variable), Z (grouping difference variable), and U (grouping variable).
family	Family for generalized linear models, including 'gaussian', 'binomial', and 'poisson'.
method	There are there methods, including the proposed 'wast', 'wastapprox', 'sst', and 'slrt'.
B	An integer, the number of bootstrap samples.
K	An integer, the number of threshold values for 'sst' and 'slrt'.
isBeta	A bool value. The weight $w(\gamma)$ is chosen to be Beta distribution if `isBeta=TRUE`, which can be used if the grouping difference variable is bounded in $[0, 1]$. Default is `FALSE`.
shape1	The first parameter of Best distribution if `isBeta = TRUE`.
shape2	The second parameter of Best distribution if `isBeta = TRUE`.
N0	An integer, the number of samples to approximate ω_{ij} for 'wastapprox'. Default is `N0 = 5000`.
MU	A vector with same length as $\boldsymbol{U}$, which is the mean of weight to approximate ω_{ij} for 'wastapprox'. Default is `MU = NULL`, in which `MU = runif(p3) -0.5`.
ZK	A vector with length `N0`, which is normal sample to approximate ω_{ij} for 'wastapprox'. Default is `ZK = NULL`, in which `ZK = rnorm(N0)`.

Details

Generalized linear models

$$f(\boldsymbol{V}_i; \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}) = \exp\left\{\frac{y_i\mu_i - c(\mu_i)}{a(\phi)}\right\} h(y_i),$$

where

$$\mu_i = \boldsymbol{X}_i^T\boldsymbol{\alpha} + \boldsymbol{Z}_i^T\boldsymbol{\beta}\mathbf{1}(\boldsymbol{U}_i^T\boldsymbol{\gamma} \geq 0).$$

The hypothesis test problem is

$$H_0 : \boldsymbol{\beta} = \mathbf{0} \quad versus \quad H_1 : \boldsymbol{\beta} \neq \mathbf{0}.$$

Value

pvals	P-value of the corresponding test statistic.

References

Andrews, D. W. K. and Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative. Econometrica, 62(6):1383-1414.

Fan, A., Rui, S., and Lu, W. (2017). Change-plane analysis for subgroup detection and sample size calculation. Journal of the American Statistical Association, 112(518):769-778.

Huang, Y., Cho, J., and Fong, Y. (2021). Threshold-based subgroup testing in logistic regression models in two phase sampling designs. Journal of the Royal Statistical Society: Series C. 291-311.

Liu, X. (2023). Subgroup testing in change-plane regression with high-dimensional grouping variables. Manuscript.

Examples

```
data(simulatedData_gaussian)
pvals <- pvalglm(data = data_gaussian, family = "gaussian")
pvals

data(simulatedData_binomial)
pvals <- pvalglm(data = data_binomial, family = "binomial")
pvals

data(simulatedData_poisson)
pvals <- pvalglm(data = data_poisson, family = "poisson")
pvals
```

---

pval_probit *P-value for subgroup test in probit regression models*

Description

Provide p-value for subgroup test in probit regression models, including two methods 'wast' and 'sst'.

Usage

```
pval_probit(data, method = "wast", B = 1000, K = 2000,
      isBeta = FALSE, shape1 = 1, shape2 = 1, N0 = 5000, MU = NULL, ZK = NULL)
```

Arguments

data	A list, including Y (response), X (baseline variable), Z (grouping difference variable), and U (grouping variable).
method	There are two methods, including the proposed 'wast', 'wastapprox' and 'sst'.
B	An integer, the number of bootstrap samples.
K	An integer, the number of threshold values for 'sst'.
isBeta	A bool value. The weight $w(\gamma)$ is chosen to be Beta distribution if isBeta=TRUE, which can be used if the grouping difference variable is bounded in $[0, 1]$. Default is FALSE.
shape1	The first parameter of Best distribution if isBeta = TRUE.
shape2	The second parameter of Best distribution if isBeta = TRUE.
N0	An integer, the number of samples to approximate ω_{ij} for 'wastapprox'. Default is N0 = 5000.
MU	A vector with same length as $\boldsymbol{U}$, which is the mean of weight to approximate ω_{ij} for 'wastapprox'. Default is MU = NULL, in which MU = runif(p3) -0.5.
ZK	A vector with length N0, which is normal sample to approximate ω_{ij} for 'wastapprox'. Default is ZK = NULL, in which ZK = rnorm(N0).

Details

Probit regression models

$$f(\boldsymbol{V}_i) = \Phi(h(\boldsymbol{V}_i, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}))^{Y_i} + \Phi(-h(\boldsymbol{V}_i, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}))^{1-Y_i},$$

where $\Phi(\cdot)$ is the cumulative distribution function of standard normal distribution, and

$$h(\boldsymbol{V}_i, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}) = \boldsymbol{X}_i\boldsymbol{\alpha} + \boldsymbol{Z}_i^T\boldsymbol{\beta}\mathbf{1}(\boldsymbol{U}_i^T\boldsymbol{\theta} \geq 0).$$

The hypothesis test problem is

$$H_0 : \boldsymbol{\beta} = \mathbf{0} \quad versus \quad H_1 : \boldsymbol{\beta} \neq \mathbf{0}.$$

Value

`pvals` P-value of the corresponding test statistic.

References

Andrews, D. W. K. and Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative. Econometrica, 62(6):1383-1414.

Fan, A., Rui, S., and Lu, W. (2017). Change-plane analysis for subgroup detection and sample size calculation. Journal of the American Statistical Association, 112(518):769-778.

Huang, Y., Cho, J., and Fong, Y. (2021). Threshold-based subgroup testing in logistic regression models in two phase sampling designs. Journal of the Royal Statistical Society: Series C. 291-311.

LEE, S., SEO, M. H. and SHIN, Y. (2011). Testing for Threshold Effects in Regression Models. Journal of the American Statistical Association 106, 220-231.

Liu, X. (2023). Subgroup testing in change-plane regression with high-dimensional grouping variables. Manuscript.

Examples

```
data(simulatedData_probit)
pvals <- pval_probit(data = data_probit, method = "wast")
pvals

pvals <- pval_probit(data = data_probit, method = "sst")
pvals
```

`pval_quantile` *P-value for subgroup test in quantile regression models*

Description

Provide p-value for subgroup test in quantile regression.

Usage

```
pval_quantile(data, method = "wast", tau = 0.5, B = 1000, K = 2000,
      isBeta = FALSE, shape1 = 1, shape2 = 1, N0 = 5000, MU = NULL, ZK = NULL)
```

Arguments

data	A list, including Y (response), X (baseline variable), Z (grouping difference variable), and U (grouping variable).
method	There are two methods, including the proposed 'wast', 'wastapprox' and 'sst'.
tau	The given quantile τ, a scale in the unit inteval. Default is `tau = 0.5`.
B	An integer, the number of bootstrap samples.
K	An integer, the number of threshold values.
isBeta	A bool value. The weight $w(\gamma)$ is chosen to be Beta distribution if `isBeta=TRUE`, which can be used if the grouping difference variable is bounded in $[0, 1]$. Default is `FALSE`.
shape1	The first parameter of Best distribution if `isBeta = TRUE`.
shape2	The second parameter of Best distribution if `isBeta = TRUE`.
N0	An integer, the number of samples to approximate ω_{ij} for 'wastapprox'. Default is `N0 = 5000`.
MU	A vector with same length as $\boldsymbol{U}$, which is the mean of weight to approximate ω_{ij} for 'wastapprox'. Default is `MU = NULL`, in which `MU = runif(p3) -0.5`.
ZK	A vector with length `N0`, which is normal sample to approximate ω_{ij} for 'wastapprox'. Default is `ZK = NULL`, in which `ZK = rnorm(N0)`.

Details

Quantile regression models

$$Q_{Y_i}(\tau|\boldsymbol{X}_i, \boldsymbol{Z}_i, \boldsymbol{U}_i) = \boldsymbol{X}_i^T\boldsymbol{\alpha}(\tau) + \boldsymbol{Z}_i^T\boldsymbol{\beta}(\tau)\mathbf{1}(\boldsymbol{U}_i^T\boldsymbol{\theta}(\tau) \geq 0).$$

The hypothesis test problem is

$$H_0 : \boldsymbol{\beta} = \mathbf{0} \quad versus \quad H_1 : \boldsymbol{\beta} \neq \mathbf{0}.$$

Value

pvals	P-value of the corresponding test statistic.

References

Andrews, D. W. K. and Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative. Econometrica, 62(6):1383-1414.

Fan, A., Rui, S., and Lu, W. (2017). Change-plane analysis for subgroup detection and sample size calculation. Journal of the American Statistical Association, 112(518):769-778.

Huang, Y., Cho, J., and Fong, Y. (2021). Threshold-based subgroup testing in logistic regression models in two phase sampling designs. Journal of the Royal Statistical Society: Series C. 291-311.

LEE, S., SEO, M. H. and SHIN, Y. (2011). Testing for Threshold Effects in Regression Models. Journal of the American Statistical Association 106, 220-231.

Liu, X. (2023). Subgroup testing in change-plane regression with high-dimensional grouping variables. Manuscript.

Examples

```
data(simulatedData_quantile)
pvals <- pval_quantile(data = data_quantile, tau = 0.5, method = "wast")
pvals

pvals <- pval_quantile(data = data_quantile, tau = 0.3, method = "wast")
pvals
```

`pval_semiparam` *P-value for subgroup test in semiparamtric models*

Description

Provide p-value for subgroup test in semiparamtric models, including two methods 'wast' and 'sst'.

Usage

```
pval_semiparam(data, method = "wast", B = 1000, K = 2000,
      isBeta = FALSE, shape1 = 1, shape2 = 1, N0 = 5000, MU = NULL, ZK = NULL)
```

Arguments

`data`	A list, including Y (response), A (treatment indicator), $X1$ and $X2$ (baseline variables), Z (grouping difference variable), and U (grouping variable).
`method`	There are two methods, including the proposed 'wast', 'wastapprox' and 'sst'.
`B`	An integer, the number of bootstrap samples.
`K`	An integer, the number of threshold values for 'sst'.
`isBeta`	A bool value. The weight $w(\gamma)$ is chosen to be Beta distribution if `isBeta=TRUE`, which can be used if the grouping difference variable is bounded in $[0, 1]$. Default is `FALSE`.
`shape1`	The first parameter of Best distribution if `isBeta = TRUE`.
`shape2`	The second parameter of Best distribution if `isBeta = TRUE`.
`N0`	An integer, the number of samples to approximate ω_{ij} for 'wastapprox'. Default is `N0 = 5000`.
`MU`	A vector with same length as $\boldsymbol{U}$, which is the mean of weight to approximate ω_{ij} for 'wastapprox'. Default is `MU = NULL`, in which `MU = runif(p3) -0.5`.
`ZK`	A vector with length `N0`, which is normal sample to approximate ω_{ij} for 'wastapprox'. Default is `ZK = NULL`, in which `ZK = rnorm(N0)`.

Details

Semiparamtric models (see details in my paper Liu (2022))

$$Y_i = h(\boldsymbol{X}_{1i}) + A_i \boldsymbol{Z}_i^T \boldsymbol{\beta} \mathbf{1}(\boldsymbol{U}_i^T \boldsymbol{\theta} \geq 0) + \epsilon_i,$$

where $h(\boldsymbol{X}_1)$ is an unknown baseline mean function for patients in treatment $A = 0$ which can be set a linear function $h(\boldsymbol{X}_1) = X_1^T \boldsymbol{\alpha}_1$, and $P(A = 1|\boldsymbol{X}_2)$ can be modelled by a logistic regression model

$$P(A = 1|\boldsymbol{X}_2) = \pi(\boldsymbol{X}_2, \boldsymbol{\alpha}_2) = \frac{\exp\{\boldsymbol{X}_2^T \boldsymbol{\alpha}_2\}}{1 + \exp\{\boldsymbol{X}_2^T \boldsymbol{\alpha}_2\}}.$$

The hypothesis test problem is

$$H_0 : \boldsymbol{\beta} = \mathbf{0} \quad versus \quad H_1 : \boldsymbol{\beta} \neq \mathbf{0}.$$

Value

`pvals` P-value of the corresponding test statistic.

References

Andrews, D. W. K. and Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative. Econometrica, 62(6):1383-1414.

Fan, A., Rui, S., and Lu, W. (2017). Change-plane analysis for subgroup detection and sample size calculation. Journal of the American Statistical Association, 112(518):769-778.

Huang, Y., Cho, J., and Fong, Y. (2021). Threshold-based subgroup testing in logistic regression models in two phase sampling designs. Journal of the Royal Statistical Society: Series C. 291-311.

LEE, S., SEO, M. H. and SHIN, Y. (2011). Testing for Threshold Effects in Regression Models. Journal of the American Statistical Association 106, 220-231.

Liu, X. (2023). Subgroup testing in change-plane regression with high-dimensional grouping variables. Manuscript.

Examples

```
data(simulatedData_semiparam)
pvals <- pval_semiparam(data = data_semiparam, method = "wast")
pvals

pvals <- pval_semiparam(data = data_semiparam, method = "sst")
pvals
```

`simulatedData` *Simulated data from generalized linear models*

Description

Simulated data from the framework of general estimating equations, including model

- 'Quantile regression' (`simulatedData_quantile`),
- 'Probit regression' (`simulatedData_probit`),
- 'Semiparamtric models' (`simulatedData_semiparam`),
- Simulated data from generalized linear models, including family 'gaussian' (`simulatedData_gaussian`), 'binomial' (`simulatedData_binomial`), and 'poisson' (`simulatedData_poisson`).

Usage

```
data(simulatedData_gaussian)
```

Details

We simulated data generated from generalized linear models

$$f(\boldsymbol{V}_i; \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}) = \exp\left\{\frac{y_i\mu_i - c(\mu_i)}{a(\phi)}\right\} h(y_i),$$

where

$$\mu_i = \boldsymbol{X}_i^T\boldsymbol{\alpha} + \boldsymbol{Z}_i^T\boldsymbol{\beta}\mathbf{1}(\boldsymbol{U}_i^T\boldsymbol{\gamma} \geq 0).$$

- Y: the response, an n-vector
- X: the baseline variable with dimension $n \times p$
- Z: the grouping difference variable with dimension $n \times q$
- U: the grouping variable with dimension $n \times r$

References

Liu, X. (2023). Subgroup testing in change-plane regression with high-dimensional grouping variables. Manuscript.

Examples

```
data(simulatedData_gaussian)

y <- data_gaussian$Y[1:5]
x <- dim(data_gaussian$X)
z <- dim(data_gaussian$Z)
u <- dim(data_gaussian$U)

data(simulatedData_probit)
y <- data_probit$Y[1:5]
x <- dim(data_probit$X)
z <- dim(data_probit$Z)
u <- dim(data_probit$U)

data(simulatedData_semiparam)
y <- data_semiparam$Y[1:5]
x1 <- dim(data_semiparam$X1)
x2 <- dim(data_semiparam$X2)
z <- dim(data_semiparam$Z)
u <- dim(data_semiparam$U)
```

Index