Package ‘wastee’

May 11, 2022

Type Package

Title Testing under loss of identifiability

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 mixture models and 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/wastee

Encoding UTF-8

Archs i386, x64

R topics documented:

`wastee-package` *Testing under loss of identifiability*

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 mixture models and 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:	wastee
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.

Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika, 64(2):247-254.

Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika, 74(1):33-43.

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. (2022). Testing under loss of identifiability. Manuscript.

`exams` *Examples for subgroup test in the framework of general estimating equation*

Description

Examples of p-value of test statitics for 'Gaussian mixture', 'Exponential mixture', 'Poission mixture', 'Quantile regression', 'Probit regression', and 'Semiparamtric models'.

Usage

```
exams(family = "mixexp", method = "wast", M = 1000, K = 1000, tau = 0.5)
```

Arguments

`family`	Family for the general estimating equation, including Gaussian mixture ('mixnorm'), Exponential mixture ('mixexp'), Poission mixture ('mixpoiss'), Quantile regression ('quantile'), Probit regression ('probit'), and Semiparamtric models ('semiparam').
`method`	There are two methods, including the proposed 'wast' and 'sst'.
`M`	An integer, the number of bootstrap samples.
`K`	An integer, the number of threshold values for 'sst'.
`tau`	The given quantile τ, a scale in the unit inteval. It is only available if `family="quantile"`.

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.

Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika, 64(2):247-254.

Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika, 74(1):33-43.

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. (2022). Testing under loss of identifiability. Manuscript.

Examples

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

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

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

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

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

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

pval_mixexp *P-value for subgroup test in Exponential mixture models*

Description

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

Usage

```
pval_mixexp(y, method = "wast", M = 1000, K = 1000, lambda = 1, alpha = NULL)
```

Arguments

`y`	The data.
`method`	There are three methods, including the proposed 'wast', 'sst', and 'davies'. Refer method 'davies' to Davies (1977, 1987) which is only avaliable for known true parameter $\alpha = \alpha_0$.
`M`	An integer, the number of bootstrap samples.
`K`	An integer, the number of threshold values for 'sst' and 'davies'.
`lambda`	The rate parameter of weight that is a Exponential distribution.
`alpha`	Rate parameter of Exponential distribution under the null. Default is `NULL`, in which the rate parameter α is estimated by the reciprocal of sample mean, that is $\hat{\alpha} = \frac{1}{\bar{y}}$.

Details

Exponential mixture model

$$f(y) = (1-\beta)\alpha \exp(-\alpha y) + \beta\theta \exp(-\theta y),$$

where $\beta \in [0, 1]$, $\alpha \in \boldsymbol{R}$ and $\theta \in \boldsymbol{R}$ are unknown parameters.

The hypothesis test problem is

$$H_0 : \beta = 0 \quad versus \quad H_1 : \beta \neq 0.$$

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.

Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika, 64(2):247-254.

Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika, 74(1):33-43.

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. (2022). Testing under loss of identifiability. Manuscript.

Examples

```
data(simulatedData_mixexp)
pvals <- pval_mixexp(y = data_mixexp, method = "wast")
pvals

pvals <- pval_mixexp(y = data_mixexp, method = "sst")
pvals

pvals <- pval_mixexp(y = data_mixexp, method = "davies", alpha = 1)
pvals
```

`pval_mixnorm` *P-value for subgroup test in Gaussian mixture models*

Description

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

Usage

```
pval_mixnorm(y, method = "wast", M = 1000, K = 1000,
                      mu = 0, sigma2 = 1, alpha = NULL)
```

Arguments

`y`	The data.
`method`	There are three methods, including the proposed 'wast', 'sst', and 'davies'. Refer method 'davies' to Davies (1977, 1987), which is only avaliable for known true parameter $\alpha = \alpha_0$.
`M`	An integer, the number of bootstrap samples.
`K`	An integer, the number of threshold values for 'sst' and 'davies'.
`mu`	The mean μ of weight that is a normal distribution.
`sigma2`	The variance σ^2 of weight that is a normal distribution.
`alpha`	Mean under the null. Default is `NULL`, in which the mean α is estimated by the sample mean.

Details

Gaussian mixture model

$$f(y) = (1-\beta)\phi(y-\alpha) + \beta\phi(y-\theta),$$

where $\phi(\cdot)$ is the density of standard normal distribution, and $\beta \in [0,1]$, $\alpha \in \boldsymbol{R}$ and $\theta \in \boldsymbol{R}$ are unknown parameters.

The hypothesis test problem is

$$H_0 : \beta = 0 \quad versus \quad H_1 : \beta \neq 0.$$

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.

Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika, 64(2):247-254.

Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika, 74(1):33-43.

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. (2022). Testing under loss of identifiability. Manuscript.

Examples

```
data(simulatedData_mixnorm)
pvals <- pval_mixnorm(y = data_mixnorm, method = "wast")
pvals

pvals <- pval_mixnorm(y = data_mixnorm, method = "sst")
pvals

pvals <- pval_mixnorm(y = data_mixnorm, method = "davies", alpha = 0)
pvals
```

pval_mixpoiss *P-value for subgroup test in Poission mixture models*

Description

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

Usage

```
pval_mixpoiss(y, method = "wast", M = 1000, K = 1000,
                    lambda = 1, tau = 2, alpha = NULL)
```

Arguments

y	The data.
method	There are three methods, including the proposed 'wast', 'sst', and 'davies'. Refer method 'davies' to Davies (1977, 1987) which is only avaliable for known true parameter $\alpha = \alpha_0$.
M	An integer, the number of bootstrap samples.
K	An integer, the number of threshold values for 'sst' and 'davies'.
lambda	The rate parameter λ of weight that is a Gamma distribution.
tau	The shape parameter τ of weight that is a Gamma distribution.
alpha	Mean under the null. Default is NULL, in which the mean α is estimated by the sample mean.

Details

Poisson mixture model

$$f(y) = (1-\beta)\frac{\alpha^y \exp(-\alpha)}{y!} + \beta\frac{\theta^y \exp(-\theta)}{y!},$$

where $\beta \in [0,1]$, $\alpha \in \boldsymbol{R}$ and $\theta \in \boldsymbol{R}$ are unknown parameters.

The hypothesis test problem is

$$H_0 : \beta = 0 \quad versus \quad H_1 : \beta \neq 0.$$

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.

Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika, 64(2):247-254.

Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika, 74(1):33-43.

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. (2022). Testing under loss of identifiability. Manuscript.

Examples

```
data(simulatedData_mixpoiss)
pvals <- pval_mixpoiss(y = data_mixpoiss, method = "wast")
pvals

pvals <- pval_mixpoiss(y = data_mixpoiss, method = "sst")
pvals

pvals <- pval_mixpoiss(y = data_mixpoiss, method = "davies", alpha = 1)
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", M = 1000, K = 2000,
                 isBeta = FALSE, shape1 = 1, shape2 = 1)
```

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' and 'sst'.
`M`	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`.

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

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

The hypothesis test problem is

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

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. (2022). Testing under loss of identifiability. 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 models, including two methods 'wast' and 'sst'.

Usage

```
pval_quantile(data, method = "wast", tau = 0.5, M = 1000, K = 2000,
                              isBeta = FALSE, shape1 = 1, shape2 = 1)
```

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' and 'sst'.
tau	The given quantile τ, a scale in the unit inteval.
M	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`.

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. (2022). Testing under loss of identifiability. 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", M = 1000, K = 2000,
                isBeta = FALSE, shape1 = 1, shape2 = 1)
```

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' and 'sst'.
`M`	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`.

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. (2022). Testing under loss of identifiability. 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 the framework of general estimating equations*

Description

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

- 'Gaussian mixture' (`simulatedData_mixnorm`),
- 'Exponential mixture' (`simulatedData_mixexp`),
- 'Poission mixture' (`simulatedData_mixpoiss`),
- 'Quantile regression' (`simulatedData_quantile`),
- 'Probit regression' (`simulatedData_probit`), and
- 'Semiparamtric models' (`simulatedData_semiparam`).

Usage

```
data(simulatedData_probit)
```

Details

We simulated data generated from general estimating equations, for instance, 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

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

- 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. (2022). Testing under loss of identifiability. Manuscript.

Examples

```
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