Package ‘wast’

July 10, 2022

Type Package

Title Subgroup testing in generalized linear models

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 generalized linear models. In the paper Liu (2022), we consider hypothesis test of coefficients in the generalized linear models (GLM) to detect the existence of the subgroups, which can serve as the optimal individualized treatment recommendation in practice. Test we consider in this paper is one of the class of test problems when a part of parameters is not identifiable under the null. We propose a novel U-like statistic by taking the weighted average over the grouping parameter's space. The proposed test statistic not only improves significantly the power but also is computationally efficient.

License GPL (>= 2)

Depends R (>= 3.2.0)

LazyData true

NeedsCompilation yes

Repository github

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

Encoding UTF-8

R topics documented:

`wast-package` *Subgroup testing in generalized linear models*

Description

Provide a method to calculate p-value of the test statistic for subgroup detecting in generalized linear models. In the paper Liu (2022), we consider hypothesis test of coefficients in the generalized linear models (GLM) to detect the existence of the subgroups, which can serve as the optimal individualized treatment recommendation in practice. Test we consider in this paper is one of the class of test problems when a part of parameters is not identifiable under the null. We propose a novel U-like statistic by taking the weighted average over the grouping parameter's space. The proposed test statistic not only improves significantly the power but also is computationally efficient.

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. (2022). Subgroup testing in generalized linear models. Manuscript.

`estglm` *Estimation in Generalized Linear Models with subgroups*

Description

Provide estimators of coefficients in generalized linear models with subgroups.

Usage

```
estglm(data, family = "gaussian", h = NULL, smooth = "sigmoid", maxIter = 100, tol = 0.0001)
```

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'.
`h`	A numeric number, which is the bandwidth in the smooth function. Default is `h = log(n)/sqrt(n)`
`smooth`	The smooth function. Either "sigmoid" (the default), "pnorm", or "mixnorm", see details below.
`maxIter`	An integer, the maximum number of iterations. Default is `maxIter = 100`.
`tol`	Convergence threshhold. Default is `tol = 0.0001`.

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 smooth functioms: (a) sigmoid function ("sigmoid")

$$S(u) = 1/(1 + e^u);$$

(b) norm CDF ("pnorm")

$$S(u) = \Phi(u);$$

(c) mixture of norm CDF and density ("mixnorm")

$$S(u) = \Phi(u) + u\phi(u),$$

where $\Phi(u)$ and $\phi(u)$ are the CDF and density of starndard norm distribution, that is,

$$\Phi(u) = \int_{-\infty}^{u} \frac{1}{2\pi} \exp\left(-\frac{s^2}{2}\right) ds,$$

and

$$\phi(u) = \frac{1}{2\pi} \exp\left(-\frac{u^2}{2}\right).$$

Value

`alpha`	Estimator of the baseline parameter $\boldsymbol{\alpha}$.
`beta`	Estimator of the grouping difference parameter $\boldsymbol{\beta}$.
`gamma`	Estimator of the grouping parameter $\boldsymbol{\gamma}$.
`delta`	A vector with length n. Estimator of the indicator function $I(\boldsymbol{U}^T\boldsymbol{\gamma} \geq 0)$.

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. (2022). Subgroup detecting in generalized linear models. Manuscript.

Examples

```
data(simulatedData_gaussian)
fit <- estglm(data = data_gaussian, family = "gaussian")
fit$alpha

data(simulatedData_binomial)
fit <- estglm(data = data_binomial, family = "binomial")
fit$beta

data(simulatedData_poisson)
fit <- estglm(data = data_poisson, family = "poisson")
fit$alpha
fit$beta
```

---

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

Description

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

Usage

```
exams(family = "gaussian", method = "wast", M = 1000, K = 1000)
```

Arguments

`family`	Family for generalized linear models, including 'Gaussian', 'binomial', and 'Poisson'.
`method`	There are there methods, including the proposed 'wast', 'sst', and 'slrt'.
`M`	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. (2022). Subgroup detecting in generalized linear models. Manuscript.

Examples

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

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

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

`pval` *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

```
pval(data, family = "gaussian", method = 'wast', M=1000, K = 2000)
```

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', 'sst', and 'slrt'.
`M`	An integer, the number of bootstrap samples.
`K`	An integer, the number of threshold values for 'sst' and 'slrt'.

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. (2022). Subgroup detecting in generalized linear models. Manuscript.

Examples

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

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

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

`simulatedData` *Simulated data from generalized linear models*

Description

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. (2022). Subgroup detecting in generalized linear models. 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)
```

Index