Package: glmpermu (via r-universe)

October 9, 2024

Title Permutation-Based Inference for Generalized Linear Models

Version 0.0.1

Description In practical applications, the assumptions underlying generalized linear models frequently face violations, including incorrect specifications of the outcome variable's distribution or omitted predictors. These deviations can render the results of standard generalized linear models unreliable. As the sample size increases, what might initially appear as minor issues can escalate to critical concerns. To address these challenges, we adopt a permutation-based inference method tailored for generalized linear models. This approach offers robust estimations that effectively counteract the mentioned problems, and its effectiveness remains consistent regardless of the sample size.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.3.1

Maintainer Xuekui Zhang <xuekui@uvic.ca>

NeedsCompilation no

Author Xuekui Zhang [aut, cre], Li Xing [aut], Jing Zhang [aut], Soojeong Kim [aut]

Date/Publication 2024-03-12 20:00:06 UTC

Repository https://ubcxzhang.r-universe.dev

RemoteUrl https://github.com/cran/glmpermu

RemoteRef HEAD

RemoteSha a6490814ae2ffc14b012dc28bb412b9619402335

Contents

---

glm.permu *Permutation-Based Inference for Generalized Linear Models*

Description

In practical applications, the assumptions underlying generalized linear models frequently face violations, including incorrect specifications of the outcome variable's distribution or omitted predictors. These deviations can render the results of standard generalized linear models unreliable. As the sample size increases, what might initially appear as minor issues can escalate to critical concerns. To address these challenges, we adopt a permutation-based inference method tailored for generalized linear models. This approach offers robust estimations that effectively counteract the mentioned problems, and its effectiveness remains consistent regardless of the sample size.

Usage

```
glm.permu(outcome, predictors, family, npermu = 1000, CI.percent = 0.95)
```

Arguments

outcome	a vector of the response variable.
predictors	a data frame of all the predictors.
family	a description of the error distribution and link function to be used in the model. We can handle all families supported by glm function.
npermu	the number of permutation times. The default value is 1000.
CI.percent	the confidence level. The default value is 0.95.

Details

In the big data era, the need to reevaluate traditional statistical methods is paramount due to the challenges posed by vast datasets. While larger samples theoretically enhance accuracy and hypothesis testing power without increasing false positives, practical concerns about inflated Type-I errors persist. The prevalent belief is that larger samples can uncover subtle effects, necessitating dual consideration of p-value and effect size. Yet, the reliability of p-values from large samples remains debated.

The fact is that larger samples can exacerbate minor issues into significant errors, leading to false conclusions when assumption violations exist. In response, a permutation-based test is introduced to counterbalance the effects of sample size and assumption discrepancies by neutralizing them between actual and permuted data. This approach effectively stabilizes nominal Type I error rates across various sample sizes, thereby ensuring robust statistical inferences even amidst breached conventional assumptions in big data.

There are many situations can lead to the assumption violations in generalized linear models such as a scenario of distribution misspecification and a scenario involving unobserved predictors.

For example, consider the problem of fitting a Poisson regression to analyze a dataset comprising one outcome variable y and one predictor x_1. The objective is to determine the statistical significance of the predictor's association with the outcome variable, primarily through the p-value of the

regression coefficient for the predictor. In the first scenario, the actual distribution of the outcomes diverges from the Poisson distribution that the model presumes. In the second scenario, outcomes are influenced by an unobserved predictor x_2. Under both situations, the Type I error rates cannot be accurately estimated, and their biases increase as the sample size grows.

To utilize an interaction term, a more complex model is required, which cannot be directly applied using this function.

Value

a data frame of estimates of regression coefficients with their permutation p-values and permutation confidence intervals.

Examples

```
set.seed(0)
x1 = rnorm(10, 0, 1)
x2 = rnorm(10, 0, 2)
x3 = rnorm(10, 0, 3)
lambda = exp(x3)
y = rpois(10, lambda)
X = as.data.frame(cbind(x1, x2))
glm.fit = glm(y~., "poisson", data = cbind(y, X))
summary(glm.fit)$coef
glm.permu(y, X, "poisson")
```

---

permu   *Permutation Function*

Description

In practical applications, the assumptions underlying generalized linear models frequently face violations, including incorrect specifications of the outcome variable's distribution or omitted predictors. These deviations can render the results of standard generalized linear models unreliable. As the sample size increases, what might initially appear as minor issues can escalate to critical concerns. To address these challenges, we adopt a permutation-based inference method tailored for generalized linear models. This approach offers robust estimations that effectively counteract the mentioned problems, and its effectiveness remains consistent regardless of the sample size.

Usage

```
permu(variables, family, npermu = 1000, CI.percent = 0.95)
```

Arguments

variables	a data frame of all the variables.
family	a description of the error distribution and link function to be used in the model.
npermu	the number of permutation times. The default value is 1000.
CI.percent	the confidence level. The default value is 0.95.

Value

a vector.

Index