

Package: `indexevent` (via `r-universe`)

September 12, 2024

Type Package

Title Adjustment for index event bias

Version 0.2.0

Author Frank Dudbridge

Maintainer Frank Dudbridge <frank.dudbridge@leicester.ac.uk>

Description Adjusts association statistics for index event bias in the context of a genome-wide association study for a subsequent event.

License GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 7.1.1

Repository <https://mrcieu.r-universe.dev>

RemoteUrl <https://github.com/DudbridgeLab/indexevent>

RemoteRef HEAD

RemoteSha e660ca3faca4e89becd130750933d8c11b165428

Contents

<code>indexevent</code>	2
<code>simexllhd</code>	4
<code>simexprofileCI</code>	4
<code>simexprofilellhd</code>	5
<code>testData</code>	6

Index	8
--------------	----------

 indexevent

Adjust association statistics for index event bias

Description

Given effect sizes and standard errors for predictors of an index trait and a subsequent trait, this function adjusts the statistics for the subsequent trait for selection bias through the index trait.

Usage

```
indexevent(
  xbeta,
  xse,
  ybeta,
  yse,
  weighted = T,
  prune = NULL,
  method = c("CWLS", "Hedges-Olkin", "Simex"),
  tol = 1e-06,
  B = 10,
  lambda = seq(0.25, 5, 0.25),
  seed = 2018
)
```

Arguments

xbeta	Vector of effects on the index trait
xse	Vector of standard errors of xbeta
ybeta	Vector of effects on the subsequent trait
yse	Vector of standard errors of ybeta
weighted	If true (default), regression of ybeta on xbeta is weighted by the inverse of yse^2 .
prune	Vector containing the indices of an approximately independent subset of the predictors in xbeta and ybeta. If unspecified, all predictors will be used.
method	Method to adjust for regression dilution (weak instruments) in the regression of $ybeta[prune]$ on $xbeta[prune]$. "CWLS" (default) applies Corrected Weighted Least Squares from Cai et al (2022). "Hedges-Olkin" applies the correction from Dudbridge et al (2019), equivalent to CWLS for unweighted regression. "Simex" applies a more time-consuming correction which may be more accurate than CWLS.
B	Number of simulations performed in each stage of the Simex adjustment.
lambda	Vector of lambdas for which the Simex simulations are performed.
seed	Random number seed for the Simex adjustment

Details

Effect sizes are on a linear scale, so could be the coefficients from linear regression, or log odds ratios, or log hazard ratios. Effects on the subsequent trait are regressed on the effects on the index trait. By default, the regression is weighted by the inverse variances of the subsequent trait effects. The regression is adjusted for sampling variation in the index trait effects, and the residuals then used to obtain adjusted effect sizes and standard errors for the subsequent trait.

The regression should be performed on a subset of predictors that are independent. In the context of a genome-wide association study, these would be LD-pruned SNPs. In terms of the input parameters, the regression command is `lm(ybeta[prune]~xbeta[prune], weights=1/yse[prune]^2)`.

The effects in `xbeta` and `ybeta` should be aligned for the same variables and the same direction prior to running `indexevent`.

The default value of `B` is 10 to get a quick result, but higher values are recommended, eg 1000.

Value

An object of class "indexevent" which contains:

- `ybeta.adj` Adjusted effects on the subsequent trait
- `yse.adj` Adjusted standard errors of `ybeta.adj`
- `ychisq.adj` Chi-square statistics for $(ybeta.adj/yse.adj)^2$
- `yp.adj` P-values for `ychisq.adj` on 1df
- `b` Coefficient of the regression of `ybeta[prune]` on `xbeta[prune]`, after correction for regression dilution
- `b.se` Standard error of `b`
- `b.ci` Lower and upper confidence limits for `b`
- `b.raw` Regression coefficient without correction for regression dilution
- `simex.estimates` Regression coefficients under simulated measurement error

Author(s)

Frank Dudbridge

References

Cai S, Hartley A, Mahmoud O, Tilling K, Dudbridge F (2022) Adjusting for collider bias in genetic association studies using instrumental variable methods. *Genetic Epidemiol* 46:303-316

Dudbridge F, Allen RJ, Sheehan NA, Schmidt AF, Lee JC, Jenkins RG, Wain LV, Hingorani AD, Patel RS (2019) Adjustment for index event bias in genome-wide association studies of subsequent events. *Nat Commun* 10:1561

simexllhd	<i>Log-likelihood for SIMEX data</i>
-----------	--------------------------------------

Description

Calculates the log-likelihood in a simple linear regression model with measurement error in the predictor, using the SIMEX method.

Usage

```
simexllhd(pvar, pmean, simex.estimateds)
```

Arguments

pvar	Ratio of the sampling variance to the variance of the true predictors, on the log scale.
pmean	Slope of the simple linear regression.
simex.estimateds	Matrix containing data simulated by SIMEX.

Details

simex.estimateds is a matrix with three columns. Column 1 contains the values of lambda under which measurement error is simulated. Column 2 contains the estimated slopes for each value of lambda. Column 3 contains the sampling variances of the estimated slopes.

The likelihood is a function of two parameters, the true slope of the simple linear regression and a parameter representing the ratio of the sampling variance to the variance of the true predictors. As this parameter must be positive, it is estimated on the log scale.

Value

Log-likelihood evaluated at pvar and pmean for the data in simex.estimateds.

simexprofileCI	<i>Profile likelihood confidence interval for SIMEX</i>
----------------	---

Description

Obtains a maximum likelihood estimate of the slope in a simple linear regression model with measurement error in the predictor, using the SIMEX method.

Usage

```
simexprofileCI(simex.estimateds, variance.ratio)
```

Arguments

- `simex.estimate` Matrix containing data simulated by SIMEX.
- `variance.ratio` Ratio of the variance of the predictor to the variance of the outcome.

Details

`simex.estimate` is a matrix with three columns. Column 1 contains the values of lambda under which measurement error is simulated. Column 2 contains the estimated slopes for each value of lambda. Column 3 contains the sampling variances of the estimated slopes.

The likelihood is a profile likelihood for the true regression slope, with the profile taken over a nuisance parameter representing the ratio of the sampling variance to the variance of the true predictors. The profiling step requires a value for the `variance.ratio`.

Value

A vector with three elements, the estimated slope and its lower and upper 95% confidence limits.

<code>simexprofilellhd</code>	<i>Profile log-likelihood for SIMEX data</i>
-------------------------------	--

Description

Calculates the profile log-likelihood of the slope in a simple linear regression model with measurement error in the predictor, using the SIMEX method.

Usage

```
simexprofilellhd(p, simex.estimate, variance.ratio)
```

Arguments

- `p` Slope of the simple linear regression.
- `simex.estimate` Matrix containing data simulated by SIMEX.
- `variance.ratio` Ratio of the variance of the predictor to the variance of the outcome.

Details

`simex.estimate` is a matrix with three columns. Column 1 contains the values of lambda under which measurement error is simulated. Column 2 contains the estimated slopes for each value of lambda. Column 3 contains the sampling variances of the estimated slopes.

The likelihood is a profile likelihood for the true regression slope, with the profile taken over a nuisance parameter representing the ratio of the sampling variance to the variance of the true predictors. The lower bound for this nuisance parameter depends on `p` and `variance.ratio`.

Value

Profile log-likelihood evaluated at p for the data in `simex.estimate`s.

testData	<i>Simulated effects on incidence and prognosis</i>
----------	---

Description

A simulated dataset consisting of regression coefficients on incidence and prognosis, with their standard errors, for 10,000 variables (eg SNPs). 500 variables have effects on incidence only, 500 on prognosis only, and 500 on both. The effects on incidence and prognosis are independent. The estimates are obtained from linear regression in a simulated dataset of 20,000 individuals.

Usage

```
testData
```

Format

A data frame with 10,000 rows and 4 variables:

xbeta Regression coefficient on incidence

xse Standard error of xbeta

ybeta Regression coefficient on prognosis

yse Standard error of ybeta

Examples

```
Default analysis with CWLS
indexevent(testData$xbeta, testData$xse, testData$ybeta, testData$yse)
# [1] "Coefficient -0.416773273239147"
# [1] "Standard error 0.0196993218284169"
# [1] "95% CI -0.455383234542707 -0.378163311935586"

# Hedges-Olkin adjustment for regression dilution
# Equivalent to an unweighted regression with CWLS
indexevent(testData$xbeta, testData$xse, testData$ybeta, testData$yse, method="Hedges-Olkin")
# [1] "Coefficient -0.441061156526639"
# [1] "Standard error 0.0211910391231297"
# [1] "95% CI -0.482594830002953 -0.399527483050326"

# SIMEX adjustment with 100 simulations for each step
indexevent(testData$xbeta, testData$xse, testData$ybeta, testData$yse, method="SIMEX", B=100)
# [1] "Coefficient -0.446543628582032"
# [1] "Standard error 0.011576233488927"
# [1] "95% CI -0.470301533547 -0.424923532117153"

# First few unadjusted effects on prognosis
```

```
testData$ybeta[1:5]
# [1] 0.032240 0.057070 -0.006959 0.080460 0.032820
# Adjusted effects
indexevent(testData$xbeta, testData$yse, testData$ybeta, testData$yse)$ybeta.adj[1:5]
# [1] 0.05109482 0.06088181 -0.01446092 0.08931226 0.01435694
```

Index

* **datasets**

testData, 6

indexevent, 2

simexllhd, 4

simexprofileCI, 4

simexprofilellhd, 5

testData, 6