Package ‘ebci’

September 6, 2021

Title Robust Empirical Bayes Confidence Intervals

Version 1.0.0

Description Computes empirical Bayes confidence estimators and confidence intervals in a normal means model. The intervals are robust in the sense that they achieve correct coverage regardless of the distribution of the means. If the means are treated as fixed, the intervals have an average coverage guarantee. The implementation is based on Armstrong, Kolesár and Plagborg-Møller (2020) <arXiv:2004.03448>.

Depends R (>= 4.1.0)

License MIT + file LICENSE

Encoding UTF-8

LazyData true

Suggests spelling, testthat (>= 2.1.0), lpSolve, knitr, rmarkdown

Language en-US

URL `https://github.com/kolesarm/ebci`

BugReports `https://github.com/kolesarm/ebci/issues`

RoxygenNote 7.1.1

VignetteBuilder knitr

NeedsCompilation no

Author Michal Kolesár [aut, cre] (<`https://orcid.org/0000-0002-2482-7796`>),
Tim Armstrong [ctb],
Mikkel Plagborg-Møller [ctb]

Maintainer Michal Kolesár <`kolesarmi@googlemail.com`>

Repository CRAN

Date/Publication 2021-09-06 08:40:05 UTC

R topics documented:

cva *Compute average coverage critical value under moment constraints.*

Description

Computes the critical value $cva_\alpha(m_2, \kappa)$ from Armstrong, Kolesár, and Plagborg-Møller (2020).

Usage

```
cva(m2, kappa = Inf, alpha = 0.05, check = TRUE)
```

Arguments

m2	Bound on second moment of the normalized bias, m_2
kappa	Bound on the kurtosis of the normalized bias, κ
alpha	Determines confidence level, $1 - \alpha$.
check	If TRUE, verify accuracy of the solution by checking that the implied least favorable distribution satisfies the m2 and kappa constraints and yields the same non-coverage rate. If this fails (perhaps due to numerical accuracy issues), solve a finite-grid approximation (by discretizing the support of the normalized bias) to the primal linear programming problem, and check that it agrees with the dual solution.

Value

Returns a list with 4 components:

cv Critical value for constructing two-sided confidence intervals.

alpha The argument alpha.

x Support points for the least favorable distribution for the squared normalized bias, b^2.

p Probabilities associated with the support points.

References

Armstrong, Timothy B., Kolesár, Michal, and Plagborg-Møller, Mikkel (2020): Robust Empirical Bayes Confidence Intervals, https://arxiv.org/abs/2004.03448

Examples

```
# Usual critical value
cva(m2=0, kappa=Inf, alpha=0.05)
# Larger critical value that takes bias into account. Only uses second moment
# constraint on normalized bias.
cva(m2=4, kappa=Inf, alpha=0.05)
# Add a constraint on kurtosis. This tightens the critical value.
cva(m2=4, kappa=3, alpha=0.05)
```

`cz` *Neighborhood effects data from Chetty and Hendren (2018)*

Description

This dataset contains a subset of the publicly available data from Chetty and Hendren (2018). It contains raw estimates and standard errors of neighborhood effects at the commuting zone level

Usage

```
cz
```

Format

A data frame with 741 rows corresponding to commuting zones (CZ) and 10 columns corresponding to the variables:

cz Commuting zone ID

czname Name of CZ

state 2-digit state code

pop Population according to the year 2000 Census

theta25 Fixed-effect estimate of the causal effect of living in the CZ for one year on children's percentile rank in the national distribution of household earnings at age 26 relative to others in the same birth cohort for children growing up with parents at the 25th percentile of national income distribution

theta75 Fixed-effect estimate of the causal effect of living in the CZ for one year on children's percentile rank in the national distribution of household earnings at age 26 relative to others in the same birth cohort for children growing up with parents at the 75th percentile of national income distribution

se25 Standard error of `theta25`

se75 Standard error of `theta75`

stayer25 Average percentile rank in the national distribution of household earnings at age 26 relative to others in the same birth cohort for stayers (children who grew up in the CZ and did not move) with parents at the 25th percentile of national income distribution.

stayer75 Average percentile rank in the national distribution of household earnings at age 26 relative to others in the same birth cohort for stayers (children who grew up in the CZ and did not move) with parents at the 75th percentile of national income distribution.

Source

https://opportunityinsights.org/data/?paper_id=599

References

Chetty, R., & Hendren, N. (2018). The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates. The Quarterly Journal of Economics, 133(3), 1163–1228. doi: 10.1093/qje/qjy006

`ebci` *Compute empirical Bayes confidence intervals by shrinking toward regression*

Description

Computes empirical Bayes estimators based on shrinking towards a regression, and associated robust empirical Bayes confidence intervals (EBCIs), as well as length-optimal robust EBCIs.

Usage

```
ebci(
  formula,
  data,
  se,
  weights = NULL,
  alpha = 0.1,
  kappa = NULL,
  wopt = FALSE,
  fs_correction = "PMT"
)
```

Arguments

`formula`	object of class `"formula"` (or one that can be coerced to that class) of the form `Y ~ predictors`, where `Y` is a preliminary unbiased estimator, and `predictors` are predictors X that guide the direction of shrinkage. For shrinking toward the grand mean, use `Y ~ 1`, and for shrinking toward `0` use `Y ~ 0`
`data`	optional data frame, list or environment (or object coercible by `as.data.frame` to a data frame) containing the preliminary estimator `Y` and the predictors. If not found in `data`, these variables are taken from `environment(formula)`, typically the environment from which the function is called.
`se`	Standard errors σ associated with the preliminary estimates `Y`
`weights`	An optional vector of weights to be used in the fitting process in computing δ, μ_2 and κ. Should be `NULL` or a numeric vector.
`alpha`	Determines confidence level, $1-\alpha$.
`kappa`	If non-`NULL`, use pre-specified value for the kurtosis κ of $\theta - X'\delta$ (such as `Inf`), instead of computing it.

`wopt` If `TRUE`, also compute length-optimal robust EBCIs. These are robust EBCIs centered at estimates with the shrinkage factor w_i chosen to minimize the length of the resulting EBCI.

`fs_correction` Finite-sample correction method used to compute μ_2 and κ. These corrections ensure that we do not shrink the preliminary estimates `Y` all the way to zero. If `"PMT"`, use posterior mean truncation, if `"FPLIB"` use limited information Bayesian approach with a flat prior, and if `"none"`, truncate the estimates at `0` for μ_2 and `1` for κ.

Value

Returns a list with the following components:

`mu2` Estimated second moment of $\theta - X'\delta$, μ_2. Vector of length 2, the first element corresponds to the estimate after the finite-sample correction as specified by `fs_correction`, the second element is the uncorrected estimate.

`kappa` Estimated kurtosis κ of $\theta - X'\delta$. Vector of length 2 with the same structure as `mu2`.

`delta` Estimated regression coefficients δ

`X` Matrix of regressors

`alpha` Determines confidence level $1 - \alpha$ used.

`df` Data frame with components described below.

`df` has the following components:

`w_eb` EB shrinkage factors, $\mu_2/(\mu_2 + \sigma_i^2)$

`w_opt` Length-optimal shrinkage factors

`ncov_pa` Maximal non-coverage of parametric EBCIs

`len_eb` Half-length of robust EBCIs based on EB shrinkage, so that the intervals take the form `cbind(th_eb-len_eb,th_eb+len_eb)`

`len_op` Half-length of robust EBCIs based on length-optimal shrinkage, so that the intervals take the form `cbind(th_op-len_op,th_op+len_op)`

`len_pa` Half-length of parametric EBCIs, which take the form `cbind(th_eb-len_pa,th_eb+len_a)`

`len_us` Half-length of unshrunk CIs, which take the form `cbind(th_us-len_us,th_us+len_us)`

`th_us` Unshrunk estimate Y

`th_eb` EB estimate.

`th_op` Estimate based on length-optimal shrinkage.

`se` Standard error σ, as supplied by the argument `se`

`weights` Weights used

`residuals` The residuals $Y_i - X_i\delta$

References

Armstrong, Timothy B., Kolesár, Michal, and Plagborg-Møller, Mikkel (2020): Robust Empirical Bayes Confidence Intervals, https://arxiv.org/abs/2004.03448

Examples

```
## Same specification as in empirical example in Armstrong, Kolesár
## and Plagborg-Møller (2020), but only use data on NY commuting zones
r <- ebci(theta25 ~ stayer25, data=cz[cz$state=="NY", ],
          se=se25, weights=1/se25^2)
```

Index