Package ‘NAVAECI’

January 21, 2026

Type Package

Title Non-Asymptotically Valid and Asymptotically Exact (NAVAE) Confidence Intervals

Version 0.1.1

Description Implements the non-asymptotically valid and asymptotically exact confidence intervals in two cases:
estimation of the mean, and estimation of (a linear combination of) the coefficients in a linear regression model,
following (Derumigny, Girard and Guyonvarch, 2025) <doi:10.48550/arXiv.2507.16776>.

License GPL-3

Encoding UTF-8

Imports BoundEdgeworth, expm

RoxygenNote 7.3.3

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

NeedsCompilation no

Author Alexis Derumigny [aut, cre] (ORCID: <https://orcid.org/0000-0002-6163-8097>),
Lucas Girard [aut],
Yannick Guyonvarch [aut]

Maintainer Alexis Derumigny <a.f.f.derumigny@tudelft.nl>

Repository CRAN

Date/Publication 2026-01-21 20:00:13 UTC

Contents

`.computeTuningParameters_OLS`
Compute tuning parameters for the NAVAE confidence interval in the linear regression case

---

Description

Compute tuning parameters for the NAVAE confidence interval in the linear regression case

Usage

```
.computeTuningParameters_OLS(n, a = NULL, omega = NULL)

## S3 method for class 'NAVAE_CI_OLS_TuningParameters'
print(x, ...)
```

Arguments

`n`	sample size
`a`	parameter a in the function `Navae_ci_ols`
`omega`	parameter omega in the function `Navae_ci_ols`
`x`	object to be printed
`...`	other arguments to passed to `print`, currently unused.

Value

`.computeTuningParameters_OLS` returns an object of class `NAVAE_CI_OLS_TuningParameters` with the values of the tuning parameters and some information on how they were determined.

`print` displays information about the tuning parameters and returns `x` invisibly.

Examples

```
.computeTuningParameters_OLS(n = 1000)
.computeTuningParameters_OLS(n = 1000, a = 2)
.computeTuningParameters_OLS(n = 1000, a = list(power_of_n_for_b = -1/3))
.computeTuningParameters_OLS(n = 1000, omega = 0.2)
.computeTuningParameters_OLS(n = 1000, omega = list(power_of_n_for_omega = -0.2))
```

`Navae_ci_mean` *Compute NAVAE CI for the expectation based on empirical mean estimator and Berry-Esseen (BE) or Edgeworth Expansions (EE) bounds*

Description

Compute NAVAE CI for the expectation based on empirical mean estimator and Berry-Esseen (BE) or Edgeworth Expansions (EE) bounds

Usage

```
Navae_ci_mean(
  data,
  alpha = 0.05,
  a = "best",
  bound_K = NULL,
  known_variance = NULL,
 param_BE_EE = list(choice = "best", setup = list(continuity = FALSE, iid = TRUE,
   no_skewness = FALSE), regularity = list(C0 = 1, p = 2), eps = 0.1),
  na.rm = FALSE
)
```

Arguments

`data` vector of univariate observations.

`alpha` this is 1 minus the confidence level of the CI; in other words, the nominal level is `1 - alpha`. By default, `alpha` is set to `0.05`, yielding a 95% CI.

`a` the free parameter a (or a_n) of the interval. It must be either

- a numeric value larger than 1, taken as the value of a,
- the character value `"best"` which is the default. It selects the `a` such that the confidence interval has the smallest length.
- a list such as `list(power_of_n_for_b = -2/5)` giving a way to compute `a` as `a = 1 + n^power_of_n_for_b`. Note that `-2/5` is the optimal (theoretical) rate.

`bound_K` bound on the kurtosis K_4(theta) of the distribution of the observations that are assumed to be i.i.d. The choice of `9` covers most "usual" distributions. If the argument is not provided (default argument `NULL`), the value used is the plug-in counterpart $\widehat{K}$, that is, the empirical kurtosis of the observations.

`known_variance` by default NULL, in this case, the function computes the CI in the general case with an unknown variance (which is estimated). Otherwise, a scalar numeric vector equal to the (assumed/known) variance. (NB: if the option is used, one must provide the variance and not the standard deviation.)

`param_BE_EE` parameters to compute the BE or EE bound δ_n used to construct the confidence interval. If `param_BE_EE` is exactly equal to `"BE"`, then the bound used is the best up-to-date BE bound from Shevtsova (2013) combined with a convexity inequality. Otherwise, `param_BE_EE` is a list of four objects:

- choice: If equal to "EE", the bound used is Derumigny et al. (2023)'s bound computed using the parameters specified by the rest of param_BE_EE, namely
- setup: itself a logical vector of size 3,
- regularity: itself a list of length up to 3,
- eps: value between 0 and 1/3,

as described in the arguments of the function BoundEdgeworth::Bound_EE1. Together, they specify the bounds and assumptions used to compute the bound δ_n from Derumigny et al. (2023). Finally, if choice is equal to "best", the bound used is the minimum between the previous one (with choice = "EE") and the bound "BE".

By default, following Remark 3.3 of the article, "best" is used and Derumigny et al. (2025)'s bounds is computed assuming i.i.d data and no other regularity assumptions (continuous or unskewed distribution) and the bound on kurtosis used is the one specified in the previous the argument bound_K.

na.rm logical, should missing values in data be removed?

Value

Navae_ci_mean returns an object of class NAVAE_CI_Mean, containing:

- ci_navae: the NAVAE confidence interval
- ci_asymp: the classical "asymptotic" CI based on CLT (as a comparison)
- indicator_R_regime: 1 if we are in the $\mathbb{R}$ regime and 0 else.
- delta_n, delta_n_from: respectively the numerical value of the bound $delta_n$ used, and a character string BE or EE indicating which type of inequality was used.
- minimal_alpha_to_exit_R_regime: the minimal alpha to exit the $\mathbb{R}$ regime.
- bound_K_value, bound_K_method: the value K used and the method to compute it.

References

For the confidence interval:

Derumigny, A., Girard, L., & Guyonvarch, Y. (2025). Can we have it all? Non-asymptotically valid and asymptotically exact confidence intervals for expectations and linear regressions. ArXiv preprint, doi:10.48550/arXiv.2507.16776.

For the underlying Edgeworth expansion bounds:

Derumigny A., Girard L., and Guyonvarch Y. (2023). Explicit non-asymptotic bounds for the distance to the first-order Edgeworth expansion, Sankhya A. doi:10.1007/s13171023003320y ArXiv preprint: doi:10.48550/arxiv.2101.05780.

See Also

Navae_ci_ols the corresponding function for the linear regression case.

Some methods for the returned object: print.NAVAE_CI_Mean and as.data.frame.NAVAE_CI_Mean.

Examples

```
n = 10000
x = rexp(n, 1)
Navae_ci_mean(x, bound_K = 9, alpha = 0.2)

Navae_ci_mean(x, bound_K = 9, alpha = 0.2, a = 1 + n^(-2/5))
# Same as:
Navae_ci_mean(x, bound_K = 9, alpha = 0.2, a = list(power_of_n_for_b = -2/5))

# plug-in for K ( = data-driven choice of K)
Navae_ci_mean(x, alpha = 0.2)

listParams1 = list(
  choice = "best",
  setup = list(continuity = FALSE, iid = TRUE, no_skewness = FALSE),
  regularity = list(C0 = 1, p = 2),
  eps = 0.1)

listParams2 = list(
  choice = "best",
  setup = list(continuity = TRUE, iid = TRUE, no_skewness = FALSE),
  regularity = list(kappa = 0.99), eps = 0.1)

Navae_ci_mean(x, alpha = 0.1, param_BE_EE = listParams1)
Navae_ci_mean(x, alpha = 0.1, param_BE_EE = listParams2)
Navae_ci_mean(x, alpha = 0.05, param_BE_EE = listParams1)
Navae_ci_mean(x, alpha = 0.05, param_BE_EE = listParams2)
```

Navae_ci_ols

Compute NAVAE CI for coefficients of a linear regression based on the OLS estimator and Berry-Esseen (BE) or Edgeworth Expansions (EE) bounds

---

Description

Compute NAVAE CI for coefficients of a linear regression based on the OLS estimator and Berry-Esseen (BE) or Edgeworth Expansions (EE) bounds

Usage

```
Navae_ci_ols(
  Y,
  X,
  alpha = 0.05,
  a = NULL,
  omega = NULL,
  bounds = list(lambda_reg = NULL, K_reg = NULL, K_eps = NULL, K_xi = NULL, C = NULL, B =
    NULL),
```

```
  K_xi = NULL,
  param_BE_EE = list(choice = "best", setup = list(continuity = FALSE, iid = TRUE,
    no_skewness = FALSE), regularity = list(C0 = 1, p = 2), eps = 0.1),
  intercept = TRUE,
  options = list(center = FALSE, bounded_case = FALSE, with_Exp_regime = FALSE),
  matrix_u = NULL,
  verbose = 0
)
```

Arguments

`Y` — vector of observations of the explained variables

`X`, `intercept` — `X` is the matrix of explanatory variables. If `intercept = TRUE`, a constant column of 1 (intercept) is added too. Note that the number of rows of `X` must be the same as the length of `Y`.

`alpha` — this is 1 minus the confidence level of the CI; in other words, the nominal level is 1 - alpha. By default, `alpha` is set to 0.05, yielding a 95% CI.

`a` — the free parameter a (or a_n) of the interval. It must be either

- a numeric value larger than 1, taken as the value of a,
- the character value `"best"` which is the default. It selects the `a` such that the confidence interval has the smallest length.
- a list such as `list(power_of_n_for_b = -2/5)` giving a way to compute `a` as `a = 1 + n^power_of_n_for_b`. Note that `-2/5` is the optimal (theoretical) rate.
- `NULL`, interpreted as the default value `a = 1 + 100 * n^(-2/5)`.

`omega` — the free parameter $omega$ (or $omega_n$) of the interval. It must be either

- a numeric value larger than 1, taken as the value of $omega$,
- the character value `"best"` which is the default. It selects the `omega` such that the confidence interval has the smallest length.
- a list such as `list(power_of_n_for_omega = -1/5)` giving a way to compute `omega` as `omega = n^power_of_n_for_omega`. Note that `-1/5` is the optimal (theoretical) rate.
- `NULL`, interpreted as the default value `omega = n^(-1/5)`.

`bounds`, `K_xi` — list of bounds for the DGP. Note that `K_xi` can also be provided as a separate argument, for convenience. It can contain the following items:

- `lambda_reg`
- `K_eps`
- `K_xi`
- `K3_xi`
- `lambda3_xi`
- `K3tilde_xi`
- `B`, `C` Bounds for the concentration of || Xi tilde
- `K_reg` Bound on $E[||vec(\widetilde{X}\widetilde{X}' - \mathbb{I}_p)||^2]$ Defined in Assumption 3.2 (ii).

The bounds that are not given are replaced by plug-ins. For K3_xi, lambda3_xi and K3tilde_xi, the bounds are obtained from K_xi (= K4_xi).

param_BE_EE — parameters to compute the BE or EE bound δ_n used to construct the confidence interval. Otherwise, param_BE_EE is a list of four objects:

- choice:
  - If equal to "EE", the bound used is Derumigny et al. (2023)'s bound computed using the parameters specified by the rest of param_BE_EE, as described in the arguments of the function BoundEdgeworth::Bound_EE1. Together, these last three items of the list specify the bounds and assumptions used to compute the bound δ_n from Derumigny et al. (2023).
  - If equal to "BE", then the bound used is the best up-to-date BE bound from Shevtsova (2013) combined with a convexity inequality.
  - If equal to "best", both bounds are computed and the smallest of both is used.
    By default, following Remark 3.3 of the article, "best" is used and Derumigny et al. (2023)'s bound is computed assuming i.i.d data and no other regularity assumptions (continuous or unskewed distribution). The bound on kurtosis that is used is the one specified in the previous argument K_xi.
- setup: itself a logical vector of size 3,
- regularity: itself a list of length up to 3,
- eps: value between 0 and 1/3,

options — a list of other options (experimental).

matrix_u — each row of this matrix is understood as a new vector u for which a confidence interval should be computed. By default matrix_u is the identity matrix, corresponding to the canonical basis of R^p.

verbose — If verbose = 0, this function is silent and does not print anything. Increasing values of verbose print more details about the progress of the computations and, in particular, the different terms that are computed.

Value

Navae_ci_ols returns an object of class NAVAE_CI_OLS, containing

- ci_navae: the NAVAE confidence interval
- ci_asymp: the classical "asymptotic" CI based on CLT (as a comparison)
- allTuningParameters, allBounds: information concerning the tuning parameters and the bounds used (numerical value and origin)
- about_delta_n, delta_n_from: respectively the numerical value of the bound $delta_n$ used, and a character string BE or EE indicating which type of inequality was used.
- minimal_alpha_to_exit_R_regime: the minimal alpha to exit the $\mathbb{R}$ regime.
- bound_K_value, bound_K_method: the value K used and the method to compute it.

References

For the confidence interval:

Derumigny, A., Girard, L., & Guyonvarch, Y. (2025). Can we have it all? Non-asymptotically valid and asymptotically exact confidence intervals for expectations and linear regressions. ArXiv preprint, doi:10.48550/arXiv.2507.16776.

For the underlying Edgeworth expansion bounds:

Derumigny A., Girard L., and Guyonvarch Y. (2023). Explicit non-asymptotic bounds for the distance to the first-order Edgeworth expansion, Sankhya A. doi:10.1007/s13171023003200y ArXiv preprint: doi:10.48550/arxiv.2101.05780.

See Also

The methods to display and process the output of this function: `print.NAVAE_CI_OLS` and `as.data.frame.NAVAE_CI_OLS`.

`Navae_ci_mean` which is the corresponding function for the estimation of the mean.

Examples

```
n = 4000
X1 = rnorm(n, sd = 1)
true_eps = rnorm(n)
Y = 2 + 8 * X1 + true_eps

myCI <- Navae_ci_ols(Y, X1, K_xi = 3, a = 1.1)

print(myCI)
```

`print.NAVAE_CI_Mean` *Print and coerce a NAVAE_CI_Mean object*

Description

Print and coerce a NAVAE_CI_Mean object

Usage

```
## S3 method for class 'NAVAE_CI_Mean'
print(x, verbose = 0, ...)

## S3 method for class 'NAVAE_CI_Mean'
as.data.frame(x, ...)
```

Arguments

`x`	the object
`verbose`	if zero, only basic printing is done. Higher values corresponds to more detailed output.
`...`	other arguments, currently ignored.

Value

print.Navae_ci_ols prints information about x and returns it invisibly.

as.data.frame returns a data.frame with 2 rows.

References

Derumigny, A., Girard, L., & Guyonvarch, Y. (2025). Can we have it all? Non-asymptotically valid and asymptotically exact confidence intervals for expectations and linear regressions. ArXiv preprint, doi:10.48550/arXiv.2507.16776

See Also

The function to generate such objects Navae_ci_mean.

The corresponding methods for the regression (OLS): print.NAVAE_CI_OLS and as.data.frame.NAVAE_CI_OLS.

Examples

```
n = 10000
x = rexp(n, 1)
myCI = Navae_ci_mean(x, bound_K = 9, alpha = 0.2)

print(myCI)
as.data.frame(myCI)
```

---

print.NAVAE_CI_OLS *Print and coerce a NAVAE_CI_OLS object*

Description

This also displays CLT-based confidence intervals. The results are different from the confidence intervals that can be obtained via confint(lm( )) since they are robust to heteroscedasticity.

Usage

```
## S3 method for class 'NAVAE_CI_OLS'
print(x, verbose = 0, ...)

## S3 method for class 'NAVAE_CI_OLS'
as.data.frame(x, ...)
```

Arguments

x	the object
verbose	if zero, only basic printing is done. Higher values corresponds to more detailed output.
...	additional arguments, currently ignored.

Value

`print.Navae_ci_ols` prints information about `x` and returns it invisibly.

`as.data.frame.NAVAE_CI_OLS` returns a data frame consisting of two observations for each vector u given as a line of `matrix_u`, with the following columns:

- `name`: name of the estimateed coefficient in the linear model
- `lower`: lower bound of the confidence interval
- `upper`: upper bound of the confidence interval
- `estimate`: the estimated value of the coefficient
- `length`: the length of the interval
- `method`: the method used for the computation of the confidence intervals. This is either "Asymptotic (CLT-based), or "NAVAE (BE-based)", or "NAVAE (EE-based)".
- `regime`: the regime used for the computation of the CI (only applicable for NAVAE confidence intervals). Four regimes are possible:
  - the degenerate regimes `R1` and `R2` in which the confidence interval is `(-Inf, Inf)`.
  - the exponential regime `Exp`
  - the Edgeworth regime `Edg`.

References

Derumigny, A., Girard, L., & Guyonvarch, Y. (2025). Can we have it all? Non-asymptotically valid and asymptotically exact confidence intervals for expectations and linear regressions. ArXiv preprint, doi:10.48550/arXiv.2507.16776

See Also

The function to generate such objects `Navae_ci_ols`.

The corresponding methods for the mean: `print.NAVAE_CI_Mean` and `as.data.frame.NAVAE_CI_Mean`.

Examples

```
n = 4000
X1 = rnorm(n, sd = 1)
true_eps = rnorm(n)
Y = 8 * X1 + true_eps
X = cbind(X1)

myCI <- Navae_ci_ols(Y, X, K_xi = 3, intercept = TRUE, a = 1.1)

print(myCI)
as.data.frame(myCI)
```

Index