Package 'BoundEdgeworth'

September 9, 2022

Type Package

Title Bound on the Error of the First-Order Edgeworth Expansion

Version 0.1.1

Description Computes uniform bounds on the distance between the cumulative distribution function of a standardized sum of random variables and its first-order Edgeworth expansion, following the article Derumigny, Girard, Guyonvarch (2021) <arXiv:2101.05780>.

License GPL-3

Encoding UTF-8

Imports expint, mathjaxr

RdMacros mathjaxr

RoxygenNote 7.2.1

BugReports https://github.com/AlexisDerumigny/BoundEdgeworth/issues

R topics documented:

Bound_BE *Compute a Berry-Esseen-type bound*

Description

This function returns a valid value δ_n for the bound

$$\sup_{x \in R} |\text{Prob}(S_n \leq x) - \Phi(x)| \leq \delta_n,$$

Usage

```
Bound_BE(
  setup = list(continuity = FALSE, iid = FALSE, no_skewness = FALSE),
  n,
  K4 = 9,
  K3 = NULL,
  lambda3 = NULL,
  K3tilde = NULL,
  regularity = list(C0 = 1, p = 2),
  eps = 0.1
)
```

Arguments

`setup` logical vector of size 3 made up of the following components:

- `continuity`: if `TRUE`, assume that the distribution is continuous.
- `iid`: if `TRUE`, assume that the random variables are i.i.d.
- `no_skewness`: if `TRUE`, assume that the distribution is unskewed.

`n` sample size (= number of random variables that appear in the sum).

`K4` bound on the 4th normalized moment of the random variables. We advise to use K4 = 9 as a general case which covers most "usual" distributions.

`K3` bound on the 3rd normalized moment. If not given, an upper bound on `K3` will be derived from the value of `K4`.

`lambda3` (average) skewness of the variables. If not given, an upper bound on $abs(lambda3)$ will be derived from the value of `K4`.

`K3tilde` value of

$$K_{3,n} + \frac{1}{n} \sum_{i=1}^{n} E|X_i| \sigma_{X_i}^2 / \overline{B}_n^3$$

where $\overline{B}_n := \sqrt{(1/n) \sum_{i=1}^n E[X_i^2]}$. If not given, an upper bound on `K3tilde` will be derived from the value of `K4`.

`regularity` list of length up to 3 (only used in the `continuity=TRUE` framework) with the following components:

- `C0` and `p`: only used in the `iid=FALSE` case. It corresponds to the assumption of a polynomial bound on f_{S_n}: $|f_{S_n}(u)| \leq C_0 \times u^{-p}$ for every $u > a_n$, where $a_n := 2t_1^* \pi \sqrt{n} / K3tilde$.
- `kappa`: only used in the `iid=TRUE` case. Corresponds to a bound on the modulus of the characteristic function of the standardized X_n. More precisely, `kappa` is an upper bound on $kappa :=$ sup of modulus of $f_{X_n/\sigma_n}(t)$ over all t such that $|t| \geq 2t_1^* \pi / K3tilde$.

`eps` a value between 0 and 1/3 on which several terms depends. Any value of `eps` will give a valid upper bound but some may give tighter results than others.

Details

where $X_1, \dots, X_n$ be n independent centered variables, and S_n be their normalized sum, in the sense that $S_n := \sum_{i=1}^n X_i / \text{sd}(\sum_{i=1}^n X_i)$. This bounds follows from the triangular inequality and the bound on the difference between a cdf and its 1st-order Edgeworth Expansion.

Note that the variables $X_1, \dots, X_n$ must be independent but may have different distributions (if `setup$iid = FALSE`).

Value

A vector of the same size as n with values δ_n such that

$$\sup_{x \in R} |\text{Prob}(S_n \le x) - \Phi(x)| \le \delta_n.$$

References

Derumigny A., Girard L., and Guyonvarch Y. (2021). Explicit non-asymptotic bounds for the distance to the first-order Edgeworth expansion, ArXiv preprint arxiv:2101.05780.

See Also

Bound_EE1() for a bound on the distance to the first-order Edgeworth expansion.

Examples

```
setup = list(continuity = FALSE, iid = FALSE, no_skewness = FALSE)
regularity = list(C0 = 1, p = 2, kappa = 0.99)

computedBound_EE1 <- Bound_EE1(
  setup = setup, n = 150, K4 = 9,
  regularity = regularity, eps = 0.1 )

computedBound_BE <- Bound_BE(
  setup = setup, n = 150, K4 = 9,
  regularity = regularity, eps = 0.1 )

print(c(computedBound_EE1, computedBound_BE))
```

Bound_EE1 *Uniform bound on Edgeworth expansion*

Description

This function computes a non-aymptotically uniform bound on the difference between the cdf of a normalized sum of random varialbles and its 1st order Edgeworth expansion. It returns a valid value δ_n such that

$$\sup_{x \in R} \left| \text{Prob}(S_n \le x) - \Phi(x) - \frac{\lambda_{3,n}}{6\sqrt{n}}(1 - x^2)\varphi(x) \right| \le \delta_n,$$

where $X_1, \ldots, X_n$ be n independent centered variables, and S_n be their normalized sum, in the sense that $S_n := \sum_{i=1}^n X_i / \text{sd}(\sum_{i=1}^n X_i)$. Here $\lambda_{3,n}$ denotes the average skewness of the variables $X_1, \ldots, X_n$.

Usage

```
Bound_EE1(
  setup = list(continuity = FALSE, iid = FALSE, no_skewness = FALSE),
  n,
  K4 = 9,
  K3 = NULL,
```

```
    lambda3 = NULL,
    K3tilde = NULL,
    regularity = list(C0 = 1, p = 2),
    eps = 0.1
)
```

Arguments

`setup` logical vector of size 3 made up of the following components:

- `continuity`: if `TRUE`, assume that the distribution is continuous.
- `iid`: if `TRUE`, assume that the random variables are i.i.d.
- `no_skewness`: if `TRUE`, assume that the distribution is unskewed.

`n` sample size (= number of random variables that appear in the sum).

`K4` bound on the 4th normalized moment of the random variables. We advise to use K4 = 9 as a general case which covers most "usual" distributions.

`K3` bound on the 3rd normalized moment. If not given, an upper bound on `K3` will be derived from the value of `K4`.

`lambda3` (average) skewness of the variables. If not given, an upper bound on $abs(lambda3)$ will be derived from the value of `K4`.

`K3tilde` value of

$$K_{3,n} + \frac{1}{n}\sum_{i=1}^{n} E|X_i|\sigma_{X_i}^2/\overline{B}_n^3$$

where $\overline{B}_n := \sqrt{(1/n)\sum_{i=1}^{n} E[X_i^2]}$. If not given, an upper bound on `K3tilde` will be derived from the value of `K4`.

`regularity` list of length up to 3 (only used in the `continuity=TRUE` framework) with the following components:

- `C0` and `p`: only used in the `iid=FALSE` case. It corresponds to the assumption of a polynomial bound on f_{S_n}: $|f_{S_n}(u)| \leq C_0 \times u^{-p}$ for every $u > a_n$, where $a_n := 2t_1^* \pi \sqrt{n}/K3tilde$.
- `kappa`: only used in the `iid=TRUE` case. Corresponds to a bound on the modulus of the characteristic function of the standardized X_n. More precisely, `kappa` is an upper bound on $kappa :=$ sup of modulus of $f_{X_n/\sigma_n}(t)$ over all t such that $|t| \geq 2t_1^* \pi/K3tilde$.

`eps` a value between 0 and 1/3 on which several terms depends. Any value of `eps` will give a valid upper bound but some may give tighter results than others.

Details

Note that the variables $X_1, \ldots, X_n$ must be independent but may have different distributions (if `setup$iid = FALSE`).

Value

A vector of the same size as `n` with values δ_n such that

$$\sup_{x \in R} \left| \text{Prob}(S_n \leq x) - \Phi(x) - \frac{\lambda_{3,n}}{6\sqrt{n}}(1 - x^2)\varphi(x) \right| \leq \delta_n.$$

References

Derumigny A., Girard L., and Guyonvarch Y. (2021). Explicit non-asymptotic bounds for the distance to the first-order Edgeworth expansion, ArXiv preprint arxiv:2101.05780.

See Also

Bound_BE() for a Berry-Esseen bound.

Examples

```
setup = list(continuity = TRUE, iid = FALSE, no_skewness = TRUE)
regularity = list(C0 = 1, p = 2)

computedBound <- Bound_EE1(
  setup = setup, n = c(150, 2000), K4 = 9,
  regularity = regularity, eps = 0.1 )

setup = list(continuity = TRUE, iid = TRUE, no_skewness = TRUE)
regularity = list(kappa = 0.99)

computedBound2 <- Bound_EE1(
  setup = setup, n = c(150, 2000), K4 = 9,
  regularity = regularity, eps = 0.1 )

setup = list(continuity = FALSE, iid = FALSE, no_skewness = TRUE)

computedBound3 <- Bound_EE1(
  setup = setup, n = c(150, 2000), K4 = 9, eps = 0.1 )

setup = list(continuity = FALSE, iid = TRUE, no_skewness = TRUE)

computedBound4 <- Bound_EE1(
  setup = setup, n = c(150, 2000), K4 = 9, eps = 0.1 )

print(computedBound)
print(computedBound2)
print(computedBound3)
print(computedBound4)
```

Index