Package 'BoundEdgeworth'

June 16, 2023

Type Package

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

Version 0.1.2.1

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 7.2.3

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

Suggests spelling,
testthat (>= 3.0.0)

Config/testthat/edition 3

Language en-US

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 $X_1, \ldots, X_n$ be n independent centered variables, and S_n be their normalized sum, in the sense that $S_n := \sum_{i=1}^n X_i / \mathrm{sd}(\sum_{i=1}^n X_i)$. 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 $X_1, \ldots, X_n$ 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

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

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_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-asymptotically uniform bound on the difference between the cdf of a normalized sum of random variables and its 1st order Edgeworth expansion. It returns a valid value δ_n such that

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

where $X_1, \ldots, X_n$ be n independent centered variables, and S_n be their normalized sum, in the sense that $S_n := \sum_{i=1}^n X_i / \mathrm{sd}(\sum_{i=1}^n X_i)$. 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,
  verbose = 0
)
```

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.

verbose   if it is `0` the function is silent (no printing). Higher values of verbose give more precise information about the computation. verbose = 1 prints the values of the intermediary terms that are summed to produce the final bound. This can be useful to understand which term has the largest contribution to the bound.

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.

Gauss_test_powerAnalysis() for a power analysis of the classical Gauss test that is uniformly valid based on this bound on the Edgeworth expansion.

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)
```

Gauss_test_powerAnalysis

Computation of power and sufficient sample size for the one-sided Gauss test

---

Description

Let $X_1, \ldots, X_n$ be n i.i.d. variables with mean μ, variance σ^2. Assume that we want to test the hypothesis $H_0 : \mu \leq \mu_0$ against the alternative $H_1 : \mu \leq \mu_0$. Let $\eta := (\mu - \mu_0)/\sigma$ be the effect size, i.e. the distance between the null and the alternative hypotheses, measured in terms of standard deviations.

Usage

```
Gauss_test_powerAnalysis(
  eta = NULL,
  n = NULL,
  beta = NULL,
  alpha = 0.05,
  K4 = 9,
  kappa = 0.99
)
```

Arguments

eta	the effect size η that characterizes the alternative hypothesis
n	sample size
beta	the power of detecting the effect eta using the sample size n
alpha	the level of the test
K4	the kurtosis of the X_i
kappa	Regularity parameter of the distribution of the X_i It 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 $|t| \geq 2t_1^* \pi / K3tilde$.

Details

There is a relation between the sample size n, the effect size `eta` and the power `beta`. This function takes as an input two of these quantities and return the third one.

Value

The computed value of either the sufficient sample size n, or the minimum effect size `eta`, or the power `beta`.

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.

Examples

```
# Sufficient sample size to detect an effect of 0.5 standard deviation with probability 80%
Gauss_test_powerAnalysis(eta = 0.5, beta = 0.8)
# We can detect an effect of 0.5 standard deviations with probability 80% for n >= 548

# Power of an experiment to detect an effect of 0.5 with a sample size of n = 800
Gauss_test_powerAnalysis(eta = 0.5, n = 800)
# We can detect an effect of 0.5 standard deviations with probability 85.1% for n = 800

# Smallest effect size that can be detected with a probability of 80% for a sample size of n = 800
Gauss_test_powerAnalysis(n = 800, beta = 0.8)
# We can detect an effect of 0.114 standard deviations with probability 80% for n = 800
```

Index