Package 'KSgeneral'

March 15, 2024

Type Package

Version 1.1.3

Title Computing P-Values of the K-S Test for (Dis)Continuous Null Distribution

Author Dimitrina S. Dimitrova <D.Dimitrova@city.ac.uk>, Yun Jia <yunjia2019@gmail.com>, Vladimir K. Kaishev <Vladimir.Kaishev.1@city.ac.uk> and Senren Tan <raymondtsrtsr@outlook.com>

Maintainer Dimitrina S. Dimitrova <D.Dimitrova@city.ac.uk>

Depends R (>= 3.3.0)

SystemRequirements fftw3 (>=3.3.4)

Description Computes a p-value of the one-sample two-sided (or one-sided, as a special case) Kolmogorov-Smirnov (KS) statistic, for any fixed critical level, and an arbitrary, possibly large sample size for a pre-specified purely discrete, mixed or continuous cumulative distribution function (cdf) under the null hypothesis. If a data sample is supplied, 'KSgeneral' computes the p-value corresponding to the value of the KS test statistic computed based on the user provided data sample. The package 'KSgeneral' implements a novel, accurate and efficient method named Exact-KS-FFT, expressing the p-value as a double-boundary non-crossing probability for a homogeneous Poisson process, which is then efficiently computed using Fast Fourier Transform (FFT). The package can also be used to compute and plot the complementary cdf of the KS statistic which is known to depend on the hypothesized distribution when the latter is discontinuous (i.e. purely discrete or mixed). To cite this package in publication use: Dimitrina S. Dimitrova, Vladimir K. Kaishev, and Senren Tan. Computing the Kolmogorov-Smirnov Distribution When the Underlying CDF is Purely Discrete, Mixed, or Continuous. Journal of Statistical Software. 2020; 95(10): 1--42. <doi:10.18637/jss.v095.i10>.

License GPL (>= 2.0)

URL https://github.com/raymondtsr/KSgeneral

Encoding UTF-8

LazyData true

Imports Rcpp (>= 0.12.12), MASS, dgof

LinkingTo Rcpp

NeedsCompilation yes

Repository CRAN

Date/Publication 2024-03-15 22:30:18 UTC

R topics documented:

`KSgeneral-package` *Computing P-Values of the K-S Test for (Dis)Continuous Null Distribution*

Description

The one-sample two-sided Kolmogorov-Smirnov (KS) statistic is one of the most popular goodness-of-fit test statistics that is used to measure how well the distribution of a random sample agrees with a prespecified theoretical distribution. Given a random sample $\{X_1, ..., X_n\}$ of size n with an empirical cdf $F_n(x)$, the two-sided KS statistic is defined as $D_n = \sup |F_n(x) - F(x)|$, where $F(x)$ is the cdf of the prespecified theoretical distribution under the null hypothesis H_0, that $\{X_1, ..., X_n\}$ comes from $F(x)$.

The package **KSgeneral** implements a novel, accurate and efficient Fast Fourier Transform (FFT)-based method, referred as Exact-KS-FFT method to compute the complementary cdf, $P(D_n \geq q)$, at a fixed $q \in [0, 1]$ for a given (hypothezied) purely discrete, mixed or continuous underlying cdf $F(x)$, and arbitrary, possibly large sample size n. A plot of the complementary cdf $P(D_n \geq q)$, $0 \leq q \leq 1$, can also be produced.

In other words, the package computes the p-value, $P(D_n \geq q)$ for any fixed critical level $q \in [0, 1]$. If a data sample, $\{x_1, ..., x_n\}$ is supplied, **KSgeneral** computes the p-value $P(D_n \geq d_n)$, where d_n is the value of the KS test statistic computed based on $\{x_1, ..., x_n\}$.

Remark: The description of the package and its functions are primarily tailored to computing the (complementary) cdf of the two-sided KS statistic, D_n. It should be noted however that one can compute the (complementary) cdf for the one-sided KS statistics D_n^- or D_n^+ (cf., Dimitrova, Kaishev, Tan (2020)) by appropriately specifying correspondingly $A_i = 0$ for all i or $B_i = 1$ for all i, in the function `ks_c_cdf_Rcpp`.

Details

The Exact-KS-FFT method underlying **KSgeneral** is based on expressing the p-value $P(D_n \geq q)$ in terms of an appropriate rectangle probability with respect to the uniform order statistics, as noted by Gleser (1985) for $P(D_n > q)$. The latter representation is used to express $P(D_n \geq q)$ via a double-boundary non-crossing probability for a homogeneous Poisson process, with intensity n, which is then efficiently computed using FFT, ensuring total run-time of order $O(n^2 log(n))$ (see Dimitrova, Kaishev, Tan (2020) and also Moscovich and Nadler (2017) for the special case when $F(x)$ is continuous).

KSgeneral represents an R wrapper of the original C++ code due to Dimitrova, Kaishev, Tan (2020) and based on the C++ code developed by Moscovich and Nadler (2017). The package includes the functions `disc_ks_c_cdf`, `mixed_ks_c_cdf` and `cont_ks_c_cdf` that compute the complementary cdf $P(D_n \geq q)$, for a fixed q, $0 \leq q \leq 1$, when $F(x)$ is purely discrete, mixed or continuous, respectively. **KSgeneral** includes also the functions `disc_ks_test`, `mixed_ks_test` and `cont_ks_test` that compute the p-value $P(D_n \geq d_n)$, where d_n is the value of the KS test statistic computed based on a user provided data sample $\{x_1, ..., x_n\}$, when $F(x)$ is purely discrete, mixed or continuous, respectively.

The functions `disc_ks_test` and `cont_ks_test` represent accurate and fast (run time $O(n^2 log(n))$) alternatives to the functions `ks.test` from the package **dgof** and the function `ks.test` from the package **stat**, which compute p-values of $P(D_n \geq d_n)$, assuming $F(x)$ is purely discrete or continuous, respectively.

The package also includes the function `ks_c_cdf_Rcpp` which gives the flexibility to compute the complementary cdf (p-value) for the one-sided KS test statistics D_n^- or D_n^+. It also allows for faster computation time and possibly higher accuracy in computing $P(D_n \geq q)$.

Author(s)

Dimitrina S. Dimitrova <D.Dimitrova@city.ac.uk>, Yun Jia <yunjia2019@gmail.com>, Vladimir K. Kaishev <Vladimir.Kaishev.1@city.ac.uk> and Senren Tan <raymondtsrtsr@outlook.com>

Maintainer: Dimitrina S. Dimitrova <D.Dimitrova@city.ac.uk>

References

Dimitrina S. Dimitrova, Vladimir K. Kaishev, Senren Tan. (2020) "Computing the Kolmogorov-Smirnov Distribution When the Underlying CDF is Purely Discrete, Mixed or Continuous". Journal of Statistical Software, **95**(10): 1-42. doi:10.18637/jss.v095.i10.

Gleser L.J. (1985). "Exact Power of Goodness-of-Fit Tests of Kolmogorov Type for Discontinuous Distributions". Journal of the American Statistical Association, **80**(392), 954-958.

Moscovich A., Nadler B. (2017). "Fast Calculation of Boundary Crossing Probabilities for Poisson Processes". Statistics and Probability Letters, **123**, 177-182.

`cont_ks_cdf` *Computes the cumulative distribution function of the two-sided Kolmogorov-Smirnov statistic when the cdf under the null hypothesis is continuous*

Description

Computes the cdf $P(D_n \leq q) \equiv P(D_n < q)$ at a fixed q, $q \in [0, 1]$, for the one-sample two-sided Kolmogorov-Smirnov statistic, D_n, for a given sample size n, when the cdf $F(x)$ under the null hypothesis is continuous.

Usage

```
cont_ks_cdf(q, n)
```

Arguments

`q`	numeric value between 0 and 1, at which the cdf $P(D_n \leq q)$ is computed
`n`	the sample size

Details

Given a random sample $\{X_1, ..., X_n\}$ of size `n` with an empirical cdf $F_n(x)$, the Kolmogorov-Smirnov goodness-of-fit statistic is defined as $D_n = \sup |F_n(x) - F(x)|$, where $F(x)$ is the cdf of a prespecified theoretical distribution under the null hypothesis H_0, that $\{X_1, ..., X_n\}$ comes from $F(x)$.

The function `cont_ks_cdf` implements the FFT-based algorithm proposed by Moscovich and Nadler (2017) to compute the cdf $P(D_n \leq q)$ at a value q, when $F(x)$ is continuous. This algorithm ensures a total worst-case run-time of order $O(n^2 log(n))$ which makes it more efficient and numerically stable than the algorithm proposed by Marsaglia et al. (2003). The latter is used by many existing packages computing the cdf of D_n, e.g., the function `ks.test` in the package **stats** and the function `ks.test` in the package **dgof**. More precisely, in these packages, the exact p-value, $P(D_n \geq q)$ is computed only in the case when $q = d_n$, where d_n is the value of the KS statistic computed based on a user provided sample $\{x_1, ..., x_n\}$. Another limitation of the functions `ks.test` is that the sample size should be less than 100, and the computation time is $O(n^3)$. In contrast, the function `cont_ks_cdf` provides results with at least 10 correct digits after the decimal point for sample sizes n up to 100000 and computation time of 16 seconds on a machine with an 2.5GHz Intel Core i5 processor with 4GB RAM, running MacOS X Yosemite. For `n` > 100000, accurate results can still be computed with similar accuracy, but at a higher computation time. See Dimitrova, Kaishev, Tan (2020), Appendix B for further details and examples.

Value

Numeric value corresponding to $P(D_n \leq q)$.

Source

Based on the C++ code available at https://github.com/mosco/crossing-probability developed by Moscovich and Nadler (2017). See also Dimitrova, Kaishev, Tan (2020) for more details.

References

Dimitrina S. Dimitrova, Vladimir K. Kaishev, Senren Tan. (2020) "Computing the Kolmogorov-Smirnov Distribution When the Underlying CDF is Purely Discrete, Mixed or Continuous". Journal of Statistical Software, **95**(10): 1-42. doi:10.18637/jss.v095.i10.

Marsaglia G., Tsang WW., Wang J. (2003). "Evaluating Kolmogorov's Distribution". Journal of Statistical Software, **8**(18), 1-4.

Moscovich A., Nadler B. (2017). "Fast Calculation of Boundary Crossing Probabilities for Poisson Processes". Statistics and Probability Letters, **123**, 177-182.

Examples

```
## Compute the value for P(D_{100} <= 0.05)

KSgeneral::cont_ks_cdf(0.05, 100)


## Compute P(D_{n} <= q)
## for n = 100, q = 1/500, 2/500, ..., 500/500
## and then plot the corresponding values against q

n<-100
q<-1:500/500
plot(q, sapply(q, function(x) KSgeneral::cont_ks_cdf(x, n)), type='l')

## Compute P(D_{n} <= q) for n = 40, nq^{2} = 0.76 as shown
## in Table 9 of Dimitrova, Kaishev, Tan (2020)

KSgeneral::cont_ks_cdf(sqrt(0.76/40), 40)
```

`cont_ks_c_cdf` *Computes the complementary cumulative distribution function of the two-sided Kolmogorov-Smirnov statistic when the cdf under the null hypothesis is continuous*

Description

Computes the complementary cdf $P(D_n \geq q) \equiv P(D_n > q)$ at a fixed q, $q \in [0, 1]$, for the one-sample two-sided Kolmogorov-Smirnov statistic, D_n, for a given sample size n, when the cdf $F(x)$ under the null hypothesis is continuous.

Usage

```
cont_ks_c_cdf(q, n)
```

Arguments

q	numeric value between 0 and 1, at which the complementary cdf $P(D_n \geq q)$ is computed
n	the sample size

Details

Given a random sample $\{X_1, ..., X_n\}$ of size n with an empirical cdf $F_n(x)$, the two-sided Kolmogorov-Smirnov goodness-of-fit statistic is defined as $D_n = \sup |F_n(x) - F(x)|$, where $F(x)$ is the cdf of a prespecified theoretical distribution under the null hypothesis H_0, that $\{X_1, ..., X_n\}$ comes from $F(x)$.

The function `cont_ks_c_cdf` implements the FFT-based algorithm proposed by Moscovich and Nadler (2017) to compute the complementary cdf, $P(D_n \geq q)$ at a value q, when $F(x)$ is continuous. This algorithm ensures a total worst-case run-time of order $O(n^2 log(n))$ which makes it more efficient and numerically stable than the algorithm proposed by Marsaglia et al. (2003). The latter is used by many existing packages computing the cdf of D_n, e.g., the function `ks.test` in the package **stats** and the function `ks.test` in the package **dgof**. More precisely, in these packages, the exact p-value, $P(D_n \geq q)$ is computed only in the case when $q = d_n$, where d_n is the value of the KS test statistic computed based on a user provided sample $\{x_1, ..., x_n\}$. Another limitation of the functions `ks.test` is that the sample size should be less than 100, and the computation time is $O(n^3)$. In contrast, the function `cont_ks_c_cdf` provides results with at least 10 correct digits after the decimal point for sample sizes n up to 100000 and computation time of 16 seconds on a machine with an 2.5GHz Intel Core i5 processor with 4GB RAM, running MacOS X Yosemite. For n > 100000, accurate results can still be computed with similar accuracy, but at a higher computation time. See Dimitrova, Kaishev, Tan (2020), Appendix C for further details and examples.

Value

Numeric value corresponding to $P(D_n \geq q)$.

Source

Based on the C++ code available at `https://github.com/mosco/crossing-probability` developed by Moscovich and Nadler (2017). See also Dimitrova, Kaishev, Tan (2020) for more details.

References

Dimitrina S. Dimitrova, Vladimir K. Kaishev, Senren Tan. (2020) "Computing the Kolmogorov-Smirnov Distribution When the Underlying CDF is Purely Discrete, Mixed or Continuous". Journal of Statistical Software, **95**(10): 1-42. doi:10.18637/jss.v095.i10.

Marsaglia G., Tsang WW., Wang J. (2003). "Evaluating Kolmogorov's Distribution". Journal of Statistical Software, **8**(18), 1-4.

Moscovich A., Nadler B. (2017). "Fast Calculation of Boundary Crossing Probabilities for Poisson Processes". Statistics and Probability Letters, **123**, 177-182.

Examples

```
## Compute the value for P(D_{100} >= 0.05)

KSgeneral::cont_ks_c_cdf(0.05, 100)


## Compute P(D_{n} >= q)
## for n = 100, q = 1/500, 2/500, ..., 500/500
## and then plot the corresponding values against q

n <- 100
q <- 1:500/500
plot(q, sapply(q, function(x) KSgeneral::cont_ks_c_cdf(x, n)), type='l')

## Compute P(D_{n} >= q) for n = 141, nq^{2} = 2.1 as shown
## in Table 18 of Dimitrova, Kaishev, Tan (2020)

KSgeneral::cont_ks_c_cdf(sqrt(2.1/141), 141)
```

`cont_ks_test` *Computes the p-value for a one-sample two-sided Kolmogorov-Smirnov test when the cdf under the null hypothesis is continuous*

Description

Computes the p-value $P(D_n \geq d_n) \equiv P(D_n > d_n)$, where d_n is the value of the KS test statistic computed based on a data sample $\{x_1, ..., x_n\}$, when $F(x)$ is continuous.

Usage

```
cont_ks_test(x, y, ...)
```

Arguments

`x`	a numeric vector of data sample values $\{x_1, ..., x_n\}$.
`y`	a pre-specified continuous cdf, $F(x)$ under the null hypothesis. Note that `y` should be a character string naming a continuous cumulative distribution function such as `pexp`, `pnorm`, etc. Only continuous cdfs are valid!
`...`	values of the parameters of the cdf, $F(x)$ specified (as a character string) by `y`.

Details

Given a random sample $\{X_1, ..., X_n\}$ of size `n` with an empirical cdf $F_n(x)$, the two-sided Kolmogorov-Smirnov goodness-of-fit statistic is defined as $D_n = \sup |F_n(x) - F(x)|$, where $F(x)$ is the cdf of a prespecified theoretical distribution under the null hypothesis H_0, that $\{X_1, ..., X_n\}$ comes from $F(x)$.

The function cont_ks_test implements the FFT-based algorithm proposed by Moscovich and Nadler (2017) to compute the p-value $P(D_n \geq d_n)$, where d_n is the value of the KS test statistic computed based on a user provided data sample $\{x_1, ..., x_n\}$, assuming $F(x)$ is continuous. This algorithm ensures a total worst-case run-time of order $O(n^2 log(n))$ which makes it more efficient and numerically stable than the algorithm proposed by Marsaglia et al. (2003). The latter is used by many existing packages computing the cdf of D_n, e.g., the function ks.test in the package **stats** and the function ks.test in the package **dgof**. A limitation of the functions ks.test is that the sample size should be less than 100, and the computation time is $O(n^3)$. In contrast, the function cont_ks_test provides results with at least 10 correct digits after the decimal point for sample sizes n up to 100000 and computation time of 16 seconds on a machine with an 2.5GHz Intel Core i5 processor with 4GB RAM, running MacOS X Yosemite. For n > 100000, accurate results can still be computed with similar accuracy, but at a higher computation time. See Dimitrova, Kaishev, Tan (2020), Appendix C for further details and examples.

Value

A list with class "htest" containing the following components:

statistic	the value of the statistic.
p.value	the p-value of the test.
alternative	"two-sided".
data.name	a character string giving the name of the data.

Source

Based on the C++ code available at https://github.com/mosco/crossing-probability developed by Moscovich and Nadler (2017). See also Dimitrova, Kaishev, Tan (2020) for more details.

References

Dimitrina S. Dimitrova, Vladimir K. Kaishev, Senren Tan. (2020) "Computing the Kolmogorov-Smirnov Distribution When the Underlying CDF is Purely Discrete, Mixed or Continuous". Journal of Statistical Software, **95**(10): 1-42. doi:10.18637/jss.v095.i10.

Moscovich A., Nadler B. (2017). "Fast Calculation of Boundary Crossing Probabilities for Poisson Processes". Statistics and Probability Letters, **123**, 177-182.

Examples

```
## Comparing the p-values obtained by stat::ks.test
## and KSgeneral::cont_ks_test

x<-abs(rnorm(100))
p.kt <- ks.test(x, "pexp", exact = TRUE)$p
p.kt_fft <- KSgeneral::cont_ks_test(x, "pexp")$p
abs(p.kt-p.kt_fft)
```

`disc_ks_c_cdf` *Computes the complementary cumulative distribution function of the two-sided Komogorov-Smirnov statistic when the cdf under the null hypothesis is purely discrete*

Description

Computes the complementary cdf, $P(D_n \geq q)$ at a fixed q, $q \in [0, 1]$, of the one-sample two-sided Kolmogorov-Smirnov (KS) statistic, when the cdf $F(x)$ under the null hypothesis is purely discrete, using the Exact-KS-FFT method expressing the p-value as a double-boundary non-crossing probability for a homogeneous Poisson process, which is then efficiently computed using FFT (see Dimitrova, Kaishev, Tan (2020)). Moreover, for comparison purposes, `disc_ks_c_cdf` gives, as an option, the possibility to compute (an approximate value for) the asymptotic $P(D_n \geq q)$ using the simulation-based algorithm of Wood and Altavela (1978).

Usage

```
disc_ks_c_cdf(q, n, y, ..., exact = NULL, tol = 1e-08, sim.size = 1e+06, num.sim = 10)
```

Arguments

`q`	numeric value between 0 and 1, at which the complementary cdf $P(D_n \geq q)$ is computed
`n`	the sample size
`y`	a pre-specified discrete cdf, $F(x)$ under the null hypothesis. Note that `y` should be a step function within the class: `stepfun`, of which `ecdf` is a subclass!
`...`	values of the parameters of the cdf, $F(x)$, specified (as a character string) by `y`.
`exact`	logical variable specifying whether one wants to compute exact p-value $P(D_n \geq q)$ using the Exact-KS-FFT method, in which case `exact = TRUE` or wants to compute an approximate p-value $P(D_n \geq q)$ using the simulation-based algorithm of Wood and Altavela (1978), in which case `exact = FALSE`. When `exact = NULL` and `n <= 100000`, the exact $P(D_n \geq q)$ will be computed using the Exact-KS-FFT method. Otherwise, the asymptotic complementary cdf is computed based on Wood and Altavela (1978). By default, `exact = NULL`.
`tol`	the value of ϵ that is used to compute the values of A_i and B_i, $i = 1, ..., n$, as detailed in Step 1 of Section 2.1 in Dimitrova, Kaishev and Tan (2020) (see also (ii) in the Procedure Exact-KS-FFT therein). By default, `tol = 1e-08`. Note that a value of `NA` or `0` will lead to an error!
`sim.size`	the required number of simulated trajectories in order to produce one Monte Carlo estimate (one MC run) of the asymptotic complementary cdf using the algorithm of Wood and Altavela (1978). By default, `sim.size = 1e+06`.
`num.sim`	the number of MC runs, each producing one estimate (based on `sim.size` number of trajectories), which are then averaged in order to produce the final estimate for the asymptotic complementary cdf. This is done in order to reduce the variance of the final estimate. By default, `num.sim = 10`.

Details

Given a random sample $\{X_1, ..., X_n\}$ of size n with an empirical cdf $F_n(x)$, the two-sided Kolmogorov-Smirnov goodness-of-fit statistic is defined as $D_n = \sup |F_n(x) - F(x)|$, where $F(x)$ is the cdf of a prespecified theoretical distribution under the null hypothesis H_0, that $\{X_1, ..., X_n\}$ comes from $F(x)$.

The function `disc_ks_c_cdf` implements the Exact-KS-FFT method, proposed by Dimitrova, Kaishev, Tan (2020) to compute the complementary cdf $P(D_n \geq q)$ at a value q, when $F(x)$ is purely discrete. This algorithm ensures a total worst-case run-time of order $O(n^2 log(n))$ which makes it more efficient and numerically stable than the only alternative algorithm developed by Arnold and Emerson (2011) and implemented as the function `ks.test` in the package **dgof**. The latter only computes a p-value $P(D_n \geq d_n)$, corresponding to the value of the KS test statistic d_n computed based on a user provided sample $\{x_1, ..., x_n\}$. More precisely, in the package **dgof** (function `ks.test`), the p-value for a one-sample two-sided KS test is calculated by combining the approaches of Gleser (1985) and Niederhausen (1981). However, the function `ks.test` only provides exact p-values for $n \leq 30$, since as noted by the authors (see Arnold and Emerson (2011)), when n is large, numerical instabilities may occur. In the latter case, `ks.test` uses simulation to approximate p-values, which may be rather slow and inaccurate (see Table 6 of Dimitrova, Kaishev, Tan (2020)).

Thus, making use of the Exact-KS-FFT method, the function `disc_ks_c_cdf` provides an exact and highly computationally efficient (alternative) way of computing $P(D_n \geq q)$ at a value q, when $F(x)$ is purely discrete.

Lastly, incorporated into the function `disc_ks_c_cdf` is the MC simulation-based method of Wood and Altavela (1978) for estimating the asymptotic complementary cdf of D_n. The latter method is the default method behind `disc_ks_c_cdf` when the sample size n is $n \geq 100000$.

Value

Numeric value corresponding to $P(D_n \geq q)$.

References

Arnold T.A., Emerson J.W. (2011). "Nonparametric Goodness-of-Fit Tests for Discrete Null Distributions". The R Journal, **3**(2), 34-39.

Dimitrina S. Dimitrova, Vladimir K. Kaishev, Senren Tan. (2020) "Computing the Kolmogorov-Smirnov Distribution When the Underlying CDF is Purely Discrete, Mixed or Continuous". Journal of Statistical Software, **95**(10): 1-42. doi:10.18637/jss.v095.i10.

Gleser L.J. (1985). "Exact Power of Goodness-of-Fit Tests of Kolmogorov Type for Discontinuous Distributions". Journal of the American Statistical Association, **80**(392), 954-958.

Niederhausen H. (1981). "Sheffer Polynomials for Computing Exact Kolmogorov-Smirnov and Renyi Type Distributions". The Annals of Statistics, 58-64.

Wood C.L., Altavela M.M. (1978). "Large-Sample Results for Kolmogorov-Smirnov Statistics for Discrete Distributions". Biometrika, **65**(1), 235-239.

See Also

`ks.test`

Examples

```
## Example to compute the exact complementary cdf for D_{n}
## when the underlying cdf F(x) is a binomial(3, 0.5) distribution,
## as shown in Example 3.4 of Dimitrova, Kaishev, Tan (2020)

binom_3 <- stepfun(c(0:3), c(0,pbinom(0:3,3,0.5)))
KSgeneral::disc_ks_c_cdf(0.05, 400, binom_3)

## Not run:
## Compute P(D_{n} >= q) for n = 100,
## q = 1/5000, 2/5000, ..., 5000/5000, when
## the underlying cdf F(x) is a binomial(3, 0.5) distribution,
## as shown in Example 3.4 of Dimitrova, Kaishev, Tan (2020),
## and then plot the corresponding values against q,
## i.e. plot the resulting complementary cdf of D_{n}

n <- 100
q <- 1:5000/5000
binom_3 <- stepfun(c(0:3), c(0,pbinom(0:3,3,0.5)))
plot(q, sapply(q, function(x) KSgeneral::disc_ks_c_cdf(x, n, binom_3)), type='l')

## End(Not run)

## Not run:
## Example to compute the asymptotic complementary cdf for D_{n}
## based on Wood and Altavela (1978),
## when the underlying cdf F(x) is a binomial(3, 0.5) distribution,
## as shown in Example 3.4 of Dimitrova, Kaishev, Tan (2020)

binom_3 <- stepfun(c(0: 3), c(0, pbinom(0 : 3, 3, 0.5)))
KSgeneral::disc_ks_c_cdf(0.05, 400, binom_3, exact = FALSE, tol = 1e-08,
sim.size = 1e+06, num.sim = 10)

## End(Not run)
```

`disc_ks_test` *Computes the p-value for a one-sample two-sided Kolmogorov-Smirnov test when the cdf under the null hypothesis is purely discrete*

Description

Computes the p-value $P(D_n \geq d_n)$, where d_n is the value of the KS test statistic computed based on a data sample $\{x_1, ..., x_n\}$, when $F(x)$ is purely discrete, using the Exact-KS-FFT method expressing the p-value as a double-boundary non-crossing probability for a homogeneous Poisson process, which is then efficiently computed using FFT (see Dimitrova, Kaishev, Tan (2020)).

Usage

```
disc_ks_test(x, y, ..., exact = NULL, tol = 1e-08, sim.size = 1e+06, num.sim = 10)
```

Arguments

x	a numeric vector of data sample values $\{x_1, ..., x_n\}$.
y	a pre-specified discrete cdf, $F(x)$, under the null hypothesis. Note that y should be a step function within the class: stepfun, of which ecdf is a subclass!
...	values of the parameters of the cdf, $F(x)$, specified (as a character string) by y.
exact	logical variable specifying whether one wants to compute exact p-value $P(D_n \geq d_n)$ using the Exact-KS-FFT method, in which case exact = TRUE or wants to compute an approximate p-value $P(D_n \geq d_n)$ using the simulation-based algorithm of Wood and Altavela (1978), in which case exact = FALSE. When exact = NULL and n <= 100000, the exact $P(D_n \geq d_n)$ will be computed using the Exact-KS-FFT method. Otherwise, the asymptotic complementary cdf is computed based on Wood and Altavela (1978). By default, exact = NULL.
tol	the value of ϵ that is used to compute the values of A_i and B_i, $i = 1, ..., n$, as detailed in Step 1 of Section 2.1 in Dimitrova, Kaishev and Tan (2020) (see also (ii) in the Procedure Exact-KS-FFT therein). By default, tol = 1e-08. Note that a value of NA or 0 will lead to an error!
sim.size	the required number of simulated trajectories in order to produce one Monte Carlo estimate (one MC run) of the asymptotic p-value using the algorithm of Wood and Altavela (1978). By default, sim.size = 1e+06.
num.sim	the number of MC runs, each producing one estimate (based on sim.size number of trajectories), which are then averaged in order to produce the final estimate for the asymptotic p-value. This is done in order to reduce the variance of the final estimate. By default, num.sim = 10.

Details

Given a random sample $\{X_1, ..., X_n\}$ of size n with an empirical cdf $F_n(x)$, the two-sided Kolmogorov-Smirnov goodness-of-fit statistic is defined as $D_n = \sup |F_n(x) - F(x)|$, where $F(x)$ is the cdf of a prespecified theoretical distribution under the null hypothesis H_0, that $\{X_1, ..., X_n\}$ comes from $F(x)$.

The function disc_ks_test implements the Exact-KS-FFT method expressing the p-value as a double-boundary non-crossing probability for a homogeneous Poisson process, which is then efficiently computed using FFT (see Dimitrova, Kaishev, Tan (2020)). It represents an accurate and fast (run time $O(n^2 log(n))$) alternative to the function ks.test from the package **dgof**, which computes a p-value $P(D_n \geq d_n)$, where d_n is the value of the KS test statistic computed based on a user provided data sample $\{x_1, ..., x_n\}$, assuming $F(x)$ is purely discrete.

In the function ks.test, the p-value for a one-sample two-sided KS test is calculated by combining the approaches of Gleser (1985) and Niederhausen (1981). However, the function ks.test due to Arnold and Emerson (2011) only provides exact p-values for n ≤ 30, since as noted by the authors, when n is large, numerical instabilities may occur. In the latter case, ks.test uses simulation to approximate p-values, which may be rather slow and inaccurate (see Table 6 of Dimitrova, Kaishev, Tan (2020)).

Thus, making use of the Exact-KS-FFT method, the function disc_ks_test provides an exact and highly computationally efficient (alternative) way of computing the p-value $P(D_n \geq d_n)$, when $F(x)$ is purely discrete.

Lastly, incorporated into the function disc_ks_test is the MC simulation-based method of Wood and Altavela (1978) for estimating the asymptotic p-value of D_n. The latter method is the default method behind disc_ks_test when the sample size n is n $\geq$ 100000.

Value

A list with class "htest" containing the following components:

statistic	the value of the statistic.
p.value	the p-value of the test.
alternative	"two-sided".
data.name	a character string giving the name of the data.

References

Arnold T.A., Emerson J.W. (2011). "Nonparametric Goodness-of-Fit Tests for Discrete Null Distributions". The R Journal, **3**(2), 34-39.

Dimitrina S. Dimitrova, Vladimir K. Kaishev, Senren Tan. (2020) "Computing the Kolmogorov-Smirnov Distribution When the Underlying CDF is Purely Discrete, Mixed or Continuous". Journal of Statistical Software, **95**(10): 1-42. doi:10.18637/jss.v095.i10.

Gleser L.J. (1985). "Exact Power of Goodness-of-Fit Tests of Kolmogorov Type for Discontinuous Distributions". Journal of the American Statistical Association, **80**(392), 954-958.

Niederhausen H. (1981). "Sheffer Polynomials for Computing Exact Kolmogorov-Smirnov and Renyi Type Distributions". The Annals of Statistics, 58-64.

Wood C.L., Altavela M.M. (1978). "Large-Sample Results for Kolmogorov-Smirnov Statistics for Discrete Distributions". Biometrika, **65**(1), 235-239.

See Also

ks.test

Examples

```
# Comparison of results obtained from dgof::ks.test
# and KSgeneral::disc_ks_test, when F(x) follows the discrete
# Uniform[1, 10] distribution as in Example 3.5 of
# Dimitrova, Kaishev, Tan (2020)

# When the sample size is larger than 100, the
# function dgof::ks.test will be numerically
# unstable

x3 <- sample(1:10, 25, replace = TRUE)
KSgeneral::disc_ks_test(x3, ecdf(1:10), exact = TRUE)
dgof::ks.test(x3, ecdf(1:10), exact = TRUE)
KSgeneral::disc_ks_test(x3, ecdf(1:10), exact = TRUE)$p -
          dgof::ks.test(x3, ecdf(1:10), exact = TRUE)$p

x4 <- sample(1:10, 500, replace = TRUE)
```

```
KSgeneral::disc_ks_test(x4, ecdf(1:10), exact = TRUE)
dgof::ks.test(x4, ecdf(1:10), exact = TRUE)
KSgeneral::disc_ks_test(x4, ecdf(1:10), exact = TRUE)$p -
          dgof::ks.test(x4, ecdf(1:10), exact = TRUE)$p

# Using stepfun() to specify the same discrete distribution as defined by ecdf():

steps <- stepfun(1:10, cumsum(c(0, rep(0.1, 10))))
KSgeneral::disc_ks_test(x3, steps, exact = TRUE)
```

`ks_c_cdf_Rcpp` *R function calling directly the C++ routines that compute the complementary cumulative distribution function of the two-sided (or one-sided, as a special case) Kolmogorov-Smirnov statistic, when the cdf under the null hypothesis is arbitrary (i.e., purely discrete, mixed or continuous)*

Description

Function calling directly the C++ routines that compute the complementary cdf for the one-sample two-sided Kolmogorov-Smirnov statistic, given the sample size `n` and the file "Boundary_Crossing_Time.txt" in the working directory. The latter file contains A_i and B_i, $i = 1, ..., n$, specified in Steps 1 and 2 of the Exact-KS-FFT method (see Equation (5) in Section 2 of Dimitrova, Kaishev, Tan (2020)). The latter values form the n-dimensional rectangular region for the uniform order statistics (see Equations (3), (5) and (6) in Dimitrova, Kaishev, Tan (2020)), namely $P(D_n \geq q) = 1 - P(A_i \leq U_{(i)} \leq B_i, 1 \leq i \leq n) = 1 - P(g(t) \leq nU_n(t) \leq h(t), 0 \leq t \leq 1)$, where the upper and lower boundary functions $h(t)$, $g(t)$ are defined as $h(t) = \sum_{i=1}^{n} 1_{(A_i < t)}$, $g(t) = \sum_{i=1}^{n} 1_{(B_i \leq t)}$, or equivalently, noting that $h(t)$ and $g(t)$ are correspondingly left and right continuous functions, we have $\sup\{t \in [0,1] : h(t) < i\} = A_i$ and $\inf\{t \in [0,1] : g(t) > i - 1\} = B_i$.

Note that on can also compute the (complementary) cdf for the one-sided KS statistics D_n^- or D_n^+ (cf., Dimitrova, Kaishev, Tan (2020)) by appropriately specifying correspondingly $A_i = 0$ for all i or $B_i = 1$ for all i, in the function `ks_c_cdf_Rcpp`.

Usage

```
ks_c_cdf_Rcpp(n)
```

Arguments

`n`	the sample size

Details

Note that all calculations here are done directly in C++ and output in R. That leads to faster computation time, as well as in some cases, possibly higher accuracy (depending on the accuracy of the pre-computed values A_i and B_i, $i = 1, ..., n$, provided in the file "Boundary_Crossing_Time.txt") compared to the functions `cont_ks_c_cdf`, `disc_ks_c_cdf`, `mixed_ks_c_cdf`.

Given a random sample $\{X_1, ..., X_n\}$ of size n with an empirical cdf $F_n(x)$, the two-sided Kolmogorov-Smirnov goodness-of-fit statistic is defined as $D_n = \sup |F_n(x) - F(x)|$, where $F(x)$ is the cdf of a prespecified theoretical distribution under the null hypothesis H_0, that $\{X_1, ..., X_n\}$ comes from $F(x)$. The one-sided KS test statistics are correspondingly defined as $D_n^- = \sup_x(F(x) - F_n(x))$ and $D_n^+ = \sup_x(F_n(x) - F(x))$.

The function ks_c_cdf_Rcpp implements the Exact-KS-FFT method, proposed by Dimitrova, Kaishev, Tan (2020), to compute the complementary cdf, $P(D_n \geq q)$ at a value q, when $F(x)$ is arbitrary (i.e. purely discrete, mixed or continuous). It is based on expressing the complementary cdf as

$P(D_n \geq q) = 1 - P(A_i \leq U_{(i)} \leq B_i, 1 \leq i \leq n)$, where A_i and B_i are defined as in Step 1 of Dimitrova, Kaishev, Tan (2020).

The complementary cdf is then re-expressed in terms of the conditional probability that a homogeneous Poisson process, $\xi_n(t)$ with intensity n will not cross an upper boundary $h(t)$ and a lower boundary $g(t)$, given that $\xi_n(1) = n$ (see Steps 2 and 3 in Section 2.1 of Dimitrova, Kaishev, Tan (2020)). This conditional probability is evaluated using FFT in Step 4 of the method in order to obtain the value of the complementary cdf $P(D_n \geq q)$. This algorithm ensures a total worst-case run-time of order $O(n^2 log(n))$ which makes it highly computationally efficient compared to other known algorithms developed for the special cases of continuous or purely discrete $F(x)$.

The values A_i and B_i, $i = 1, ..., n$, specified in Steps 1 and 2 of the Exact-KS-FFT method (see Dimitrova, Kaishev, Tan (2020), Section 2) must be pre-computed (in R or, if needed, using alternative softwares offering high accuracy, e.g. Mathematica) and saved in a file with the name "Boundary_Crossing_Time.txt" (in the current working directory).

The function ks_c_cdf_Rcpp is called in R and it first reads the file "Boundary_Crossing_Time.txt" and then computes the value for the complementaty cdf $P(D_n \geq q) = 1 - P(A_i \leq U_{(i)} \leq B_i, 1 \leq i \leq n) = 1 - P(g(t) \leq nU_n(t) \leq h(t), 0 \leq t \leq 1)$ in C++ and output in R (or as noted above, as a special case, computes the value of the complementary cdf $P(D_n^+ \geq q) = 1 - P(A_i \leq U_{(i)} \leq 1, 1 \leq i \leq n)$ or $P(D_n^- \geq q) = 1 - P(0 \leq U_{(i)} \leq B_i, 1 \leq i \leq n))$.

Value

Numeric value corresponding to $P(D_n \geq q) = 1 - P(A_i \leq U_{(i)} \leq B_i, 1 \leq i \leq n) = 1 - P(g(t) \leq \eta_n(t) \leq h(t), 0 \leq t \leq 1)$ (or, as a special case, to $P(D_n^+ \geq q)$ or $P(D_n^- \geq q)$), given a sample size n and the file "Boundary_Crossing_Time.txt" containing A_i and B_i, $i = 1, ..., n$, specified in Steps 1 and 2 of the Exact-KS-FFT method (see Dimitrova, Kaishev, Tan (2020), Section 2).

References

Dimitrina S. Dimitrova, Vladimir K. Kaishev, Senren Tan. (2020) "Computing the Kolmogorov-Smirnov Distribution When the Underlying CDF is Purely Discrete, Mixed or Continuous". Journal of Statistical Software, **95**(10): 1-42. doi:10.18637/jss.v095.i10.

Moscovich A., Nadler B. (2017). "Fast Calculation of Boundary Crossing Probabilities for Poisson Processes". Statistics and Probability Letters, **123**, 177-182.

Examples

```
## Computing the complementary cdf P(D_{n} >= q)
```

```
## for n = 10 and q = 0.1, when F(x) is continuous,
## In this case,
## B_i = (i-1)/n + q
## A_i =  i/n - q


n <- 10
q <- 0.1
up_rec <- ((1:n)-1)/n + q
low_rec <- (1:n)/n - q
df <- data.frame(rbind(up_rec, low_rec))
write.table(df,"Boundary_Crossing_Time.txt", sep = ", ",
                  row.names = FALSE, col.names = FALSE)
ks_c_cdf_Rcpp(n)
```

`mixed_ks_c_cdf` *Computes the complementary cumulative distribution function of the two-sided Kolmogorov-Smirnov statistic when the cdf under the null hypothesis is mixed*

Description

Computes the complementary cdf, $P(D_n \geq q)$ at a fixed q, $q \in [0, 1]$, of the one-sample two-sided Kolmogorov-Smirnov statistic, when the cdf $F(x)$ under the null hypothesis is mixed, using the Exact-KS-FFT method expressing the p-value as a double-boundary non-crossing probability for a homogeneous Poisson process, which is then efficiently computed using FFT (see Dimitrova, Kaishev, Tan (2020)).

Usage

```
mixed_ks_c_cdf(q, n, jump_points, Mixed_dist, ..., tol = 1e-10)
```

Arguments

`q`	numeric value between 0 and 1, at which the complementary cdf $P(D_n \geq q)$ is computed
`n`	the sample size
`jump_points`	a numeric vector containing the points of (jump) discontinuity, i.e. where the underlying cdf $F(x)$ has jump(s)
`Mixed_dist`	a pre-specified (user-defined) mixed cdf, $F(x)$, under the null hypothesis.
`...`	values of the parameters of the cdf, $F(x)$ specified (as a character string) by `Mixed_dist`.
`tol`	the value of ϵ that is used to compute the values of A_i and B_i, $i = 1, ..., n$, as detailed in Step 1 of Section 2.1 in Dimitrova, Kaishev and Tan (2020) (see also (ii) in the Procedure Exact-KS-FFT therein). By default, `tol = 1e-10`. Note that a value of `NA` or `0` will lead to an error!

Details

Given a random sample $\{X_1, ..., X_n\}$ of size n with an empirical cdf $F_n(x)$, the Kolmogorov-Smirnov goodness-of-fit statistic is defined as $D_n = \sup |F_n(x) - F(x)|$, where $F(x)$ is the cdf of a prespecified theoretical distribution under the null hypothesis H_0, that $\{X_1, ..., X_n\}$ comes from $F(x)$.

The function mixed_ks_c_cdf implements the Exact-KS-FFT method, proposed by Dimitrova, Kaishev, Tan (2020) to compute the complementary cdf $P(D_n \geq q)$ at a value q, when $F(x)$ is mixed. This algorithm ensures a total worst-case run-time of order $O(n^2 log(n))$.

We have not been able to identify alternative, fast and accurate, method (software) that has been developed/implemented when the hypothesized $F(x)$ is mixed.

Value

Numeric value corresponding to $P(D_n \geq q)$.

References

Dimitrina S. Dimitrova, Vladimir K. Kaishev, Senren Tan. (2020) "Computing the Kolmogorov-Smirnov Distribution When the Underlying CDF is Purely Discrete, Mixed or Continuous". Journal of Statistical Software, **95**(10): 1-42. doi:10.18637/jss.v095.i10.

Examples

```
# Compute the complementary cdf of D_{n}
# when the underlying distribution is a mixed distribution
# with two jumps at 0 and log(2.5),
# as in Example 3.1 of Dimitrova, Kaishev, Tan (2020)

## Defining the mixed distribution

Mixed_cdf_example <- function(x)
{
     result <- 0
     if (x < 0){
         result <- 0
     }
     else if (x == 0){
         result <- 0.5
     }
     else if (x < log(2.5)){
         result <- 1 - 0.5 * exp(-x)
     }
     else{
         result <- 1
     }

     return (result)
 }

KSgeneral::mixed_ks_c_cdf(0.1, 25, c(0, log(2.5)), Mixed_cdf_example)
```

```
## Not run:
## Compute P(D_{n} >= q) for n = 5,
## q = 1/5000, 2/5000, ..., 5000/5000
## when the underlying distribution is a mixed distribution
## with four jumps at 0, 0.2, 0.8, 1.0,
## as in Example 2.8 of Dimitrova, Kaishev, Tan (2020)

n <- 5
q <- 1:5000/5000

Mixed_cdf_example <- function(x)
{
  result <- 0
  if (x < 0){
    result <- 0
  }
  else if (x == 0){
    result <- 0.2
  }
  else if (x < 0.2){
    result <- 0.2 + x
  }
  else if (x < 0.8){
    result <- 0.5
  }
  else if (x < 1){
    result <- x - 0.1
  }
  else{
    result <- 1
  }

  return (result)
}

plot(q, sapply(q, function(x) KSgeneral::mixed_ks_c_cdf(x, n,
     c(0, 0.2, 0.8, 1.0), Mixed_cdf_example)), type='l')


## End(Not run)
```

`mixed_ks_test` *Computes the p-value for a one-sample two-sided Kolmogorov-Smirnov test when the cdf under the null hypothesis is mixed*

Description

Computes the p-value $P(D_n \geq d_n)$, where d_n is the value of the KS test statistic computed based on a data sample $\{x_1, ..., x_n\}$, when $F(x)$ is mixed, using the Exact-KS-FFT method expressing the p-value as a double-boundary non-crossing probability for a homogeneous Poisson process, which is then efficiently computed using FFT (see Dimitrova, Kaishev, Tan (2020)).

Usage

```
mixed_ks_test(x, jump_points, Mixed_dist, ..., tol = 1e-10)
```

Arguments

`x`	a numeric vector of data sample values $\{x_1, ..., x_n\}$.
`jump_points`	a numeric vector containing the points of (jump) discontinuity, i.e. where the underlying cdf $F(x)$ has jump(s)
`Mixed_dist`	a pre-specified (user-defined) mixed cdf, $F(x)$, under the null hypothesis.
`...`	values of the parameters of the cdf, $F(x)$ specified (as a character string) by `Mixed_dist`.
`tol`	the value of ϵ that is used to compute the values of A_i and B_i, $i = 1, ..., n$, as detailed in Step 1 of Section 2.1 in Dimitrova, Kaishev and Tan (2020) (see also (ii) in the Procedure Exact-KS-FFT therein). By default, `tol = 1e-10`. Note that a value of `NA` or `0` will lead to an error!

Details

Given a random sample $\{X_1, ..., X_n\}$ of size `n` with an empirical cdf $F_n(x)$, the Kolmogorov-Smirnov goodness-of-fit statistic is defined as $D_n = \sup |F_n(x) - F(x)|$, where $F(x)$ is the cdf of a prespecified theoretical distribution under the null hypothesis H_0, that $\{X_1, ..., X_n\}$ comes from $F(x)$.

The function `mixed_ks_test` implements the Exact-KS-FFT method expressing the p-value as a double-boundary non-crossing probability for a homogeneous Poisson process, which is then efficiently computed using FFT (see Dimitrova, Kaishev, Tan (2020)). This algorithm ensures a total worst-case run-time of order $O(n^2 log(n))$.

The function `mixed_ks_test` computes the p-value $P(D_n \geq d_n)$, where d_n is the value of the KS test statistic computed based on a user-provided data sample $\{x_1, ..., x_n\}$, when $F(x)$ is mixed,

We have not been able to identify alternative, fast and accurate, method (software) that has been developed/implemented when the hypothesized $F(x)$ is mixed.

Value

A list with class "htest" containing the following components:

`statistic`	the value of the statistic.
`p.value`	the p-value of the test.
`alternative`	"two-sided".
`data.name`	a character string giving the name of the data.

References

Dimitrina S. Dimitrova, Vladimir K. Kaishev, Senren Tan. (2020) "Computing the Kolmogorov-Smirnov Distribution When the Underlying CDF is Purely Discrete, Mixed or Continuous". Journal of Statistical Software, **95**(10): 1-42. doi:10.18637/jss.v095.i10.

Examples

```
# Example to compute the p-value of the one-sample two-sided KS test,
# when the underlying distribution is a mixed distribution
# with two jumps at 0 and log(2.5),
# as in Example 3.1 of Dimitrova, Kaishev, Tan (2020)

# Defining the mixed distribution

Mixed_cdf_example <- function(x)
{
     result <- 0
     if (x < 0){
         result <- 0
     }
     else if (x == 0){
         result <- 0.5
     }
     else if (x < log(2.5)){
         result <- 1 - 0.5 * exp(-x)
     }
     else{
         result <- 1
     }

     return (result)
}
test_data <- c(0,0,0,0,0,0,0.1,0.2,0.3,0.4,
            0.5,0.6,0.7,0.8,log(2.5),log(2.5),
            log(2.5),log(2.5),log(2.5),log(2.5))
KSgeneral::mixed_ks_test(test_data, c(0, log(2.5)),
                         Mixed_cdf_example)


## Compute the p-value of a two-sided K-S test
## when F(x) follows a zero-and-one-inflated
## beta distribution, as in Example 3.3
## of Dimitrova, Kaishev, Tan (2020)

## The data set is the proportion of inhabitants
## living within a 200 kilometer wide costal strip
## in 232 countries in the year 2010

data("Population_Data")
mu <- 0.6189
phi <- 0.6615
a <- mu * phi
```

```
b <- (1 - mu) * phi

Mixed_cdf_example <- function(x)
{
     result <- 0
     if (x < 0){
         result <- 0
     }
     else if (x == 0){
         result <- 0.1141
     }
     else if (x < 1){
         result <- 0.1141 + 0.4795 * pbeta(x, a, b)
     }
     else{
         result <- 1
     }

     return (result)
}
KSgeneral::mixed_ks_test(Population_Data, c(0, 1), Mixed_cdf_example)
```

`Population_Data` *The proportion of inhabitants living within a 200 kilometer wide costal strip in 232 countries in the year 2010*

Description

This data set contains the proportion of inhabitants living within a 200 kilometer wide costal strip in 232 countries in the year 2010. In Example 3.3 of Dimitrova, Kaishev, Tan (2020), the data set is modelled using a zero-and-one-inflated beta distribution in the null hypothesis and a one-sample two-sided Kolmogorov-Smirnov test is performed to test whether the proposed distribution fits the data well enough.

Usage

```
data("Population_Data")
```

Format

A data frame with 232 observations on the proportion of inhabitants living within a 200 kilometer wide costal strip in 2010.

Source

https://sedac.ciesin.columbia.edu/data/set/nagdc-population-landscape-climate-estimates-v3

References

Dimitrina S. Dimitrova, Vladimir K. Kaishev, Senren Tan. (2020) "Computing the Kolmogorov-Smirnov Distribution When the Underlying CDF is Purely Discrete, Mixed or Continuous". Journal of Statistical Software, **95**(10): 1-42. doi:10.18637/jss.v095.i10.

Index