

Package ‘GofCens’

July 29, 2024

Type Package

Title Goodness-of-Fit Methods for Complete and Right-Censored Data

Version 1.1

Date 2024-07-29

Description Graphical tools and goodness-of-fit tests for complete and right-censored data:

1. Kolmogorov-Smirnov, Cramér-von Mises, and Anderson-Darling tests, which use the empirical distribution function for complete data and are extended for right-censored data.
2. Generalized chi-squared-type test, which is based on the squared differences between observed and expected counts using random cells with right-censored data.
3. A series of graphical tools such as probability or cumulative hazard plots to guide the decision about the most suitable parametric model for the data.

License GPL (>= 2)

Encoding UTF-8

Depends R (>= 3.5.0), survival, actuar

Imports fitdistrplus, grid, ggplot2, gridExtra, survminer, boot

NeedsCompilation no

Author Klaus Langohr [aut, cre],
Mireia Besalú [aut],
Matilde Francisco [aut],
Guadalupe Gómez [aut]

Maintainer Klaus Langohr <klaus.langohr@upc.edu>

Repository CRAN

Date/Publication 2024-07-29 08:40:02 UTC

Contents

GofCens-package	2
ADcens	4

chisqcens	6
cumhazPlot	8
CvMcens	10
gofcens	12
kmPlot	14
KScens	16
nba	18
print.ADcens	19
print.chisqcens	20
print.CvMcens	21
print.gofcens	22
print.KScens	24
probPlot	25
Index	28

GofCens-package

Goodness-of-Fit Methods for Complete and Right-Censored Data.

Description

This package implements both graphical tools and goodness-of-fit tests for complete and right-censored data. It has implemented:

1. Kolmogorov-Smirnov, Cramér-von Mises, and Anderson-Darling tests, which use the empirical distribution function for complete data and are extended for right-censored data.
2. Generalized chi-squared-type test, which is based on the squared differences between observed and expected counts using random cells with right-censored data.
3. A series of graphical tools such as probability or cumulative hazard plots to guide the decision about the most suitable parametric model for the data.

Details

The GofCens package can be used to check the goodness of fit of the following 8 distributions. The list shows the parametrizations of the survival functions.

1. Exponential Distribution [Exp(β)]

$$S(t) = e^{-\frac{t}{\beta}}$$

2. Weibull Distribution [Wei(α, β)]

$$S(t) = e^{-\left(\frac{t}{\beta}\right)^\alpha}$$

3. Gumbel Distribution [Gum(μ, β)]

$$S(t) = 1 - e^{-e^{-\frac{t-\mu}{\beta}}}$$

4. Log-Logistic Distribution [LLogis(α, β)]

$$S(t) = \frac{1}{1 + \left(\frac{t}{\beta}\right)^\alpha}$$

5. Logistic Distribution [Logis(μ, β)]

$$S(t) = \frac{e^{-\frac{t-\mu}{\beta}}}{1 + e^{-\frac{t-\mu}{\beta}}}$$

6. Log-Normal Distribution [LN(μ, β)]

$$S(t) = \int_{\frac{\log t - \mu}{\beta}}^{\infty} \frac{1}{\sqrt{2\pi}}$$

7. Normal Distribution [N(μ, β)]

$$S(t) = \int_t^{\infty} \frac{1}{\beta\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\beta^2}} dx$$

8. 4-Param. Beta Distribution [Beta(α, γ, a, b)]

$$S(t) = 1 - \frac{B_{(\alpha, \gamma, a, b)}(t)}{B(\alpha, \gamma)}$$

The list of the parameters of the theoretical distribution can be set manually using the argument `params` of each function. In that case, the correspondence is: α is the shape value, γ is the shape2 value, μ is the location value and β is the scale value.

Package: GofCens
 Type: Package
 Version: 1.1
 Date: 2024-07-29
 License: GPL (>= 2)

Author(s)

Klaus Langohr, Mireia Besalú, Matilde Francisco, Guadalupe Gómez

Maintainer: Klaus Langohr <klaus.langohr@upc.edu>

ADcens

*Anderson-Darling test for complete and right-censored data***Description**

Function ADcens computes the Anderson-Darling test statistic and p-value for complete and right-censored data against eight possible distributions using bootstrapping.

Usage

```
## Default S3 method:
ADcens(times, cens = rep(1, length(times)),
       distr = c("exponential", "gumbel", "weibull", "normal",
                "lognormal", "logistic", "loglogistic", "beta"),
       betaLimits = c(0, 1), igumb = c(10, 10), degs = 3, BS = 999,
       params0 = list(shape = NULL, shape2 = NULL,
                     location = NULL, scale = NULL), tol = 1e-04, ...)
## S3 method for class 'formula'
ADcens(formula, data, ...)
```

Arguments

times	Numeric vector of times until the event of interest.
cens	Status indicator (1, exact time; 0, right-censored time). If not provided, all times are assumed to be exact.
distr	A string specifying the name of the distribution to be studied. The possible distributions are the exponential ("exponential"), the Weibull ("weibull"), the Gumbel ("gumbel"), the normal ("normal"), the lognormal ("lognormal"), the logistic ("logistic"), the loglogistic ("loglogistic"), and the beta ("beta") distribution.
betaLimits	Two-components vector with the lower and upper bounds of the Beta distribution. This argument is only required, if the beta distribution is considered.
igumb	Two-components vector with the initial values for the estimation of the Gumbel distribution parameters.
degs	Integer indicating the number of decimal places of the numeric results of the output.
BS	Number of bootstrap samples.
params0	List specifying the parameters of the theoretical distribution. By default, parameters are set to NULL and estimated with the maximum likelihood method. This argument is only considered, if all parameters of the studied distribution are specified.
tol	Precision of survival times.
formula	A formula with a numeric vector as response (which assumes no censoring) or Surv object.
data	Data frame for variables in formula.
...	Additional arguments.

Details

The parameter estimation is accomplished with the `fitdistcens` function of the **fitdistrplus** package.

To avoid long computation times due to bootstrapping, an alternative with complete data is the function `ad.test` of the **gofest** package.

The precision of the survival times is important mainly in the data generation step of the bootstrap samples.

Value

ADcens returns an object of class "ADcens".

An object of class "ADcens" is a list containing the following components:

Distribution	Null distribution.
Null hypothesis	Parameters under the null hypothesis (if <code>params0</code> is provided).
AD	Value of the Anderson-Darling statistic.
p-value	Estimated p-value.
Parameters	List with the maximum likelihood estimates of the parameters of the distribution under study.

Warning

If the amount of data is large, the execution time of the function can be elevated. The parameter `BS` can limit the number of random censored samples generated and reduce the execution time.

Author(s)

K. Langohr, M. Besalú, M. Francisco, G. Gómez.

References

G. Marsaglia and J. Marsaglia. *Evaluating the Anderson-Darling Distribution*. In: Journal of Statistical Software, Articles, 9 (2) (2004), 1-5.

See Also

Function `ad.test` (Package **gofest**) for complete data and function `gofcens` for statistics and p-value of the Kolmogorov-Smirnov, Cramér von-Mises and Anderson-Darling together for right-censored data.

Examples

```
# Complete data
set.seed(123)
ADcens(times = rweibull(100, 12, scale = 4), distr = "weibull",
        BS = 199)
```

```
## Not run:
# Censored data
data(colon)
set.seed(123)
colonsamp <- colon[sample(nrow(colon), 300), ]
ADcens(Surv(time, status) ~ 1, colonsamp, distr = "normal")

## End(Not run)
```

chisqcens

General chi-squared statistics for right-censored data.

Description

Function `chisqcens` computes the general chi-squared test statistic for right-censored data introduced by Kim (1993) and the respective p-value using bootstrapping.

Usage

```
## Default S3 method:
chisqcens(times, cens = rep(1, length(times)), M,
          distr = c("exponential", "gumbel", "weibull", "normal",
                   "lognormal", "logistic", "loglogistic", "beta"),
          betaLimits=c(0, 1), igumb = c(10, 10), degs = 3, BS = 999,
          params0 = list(shape = NULL, shape2 = NULL,
                         location = NULL, scale = NULL), tol = 1e-04, ...)

## S3 method for class 'formula'
chisqcens(formula, data,...)
```

Arguments

<code>times</code>	Numeric vector of times until the event of interest.
<code>cens</code>	Status indicator (1, exact time; 0, right-censored time). If not provided, all times are assumed to be exact.
<code>M</code>	Number indicating the number of cells that will be considered.
<code>distr</code>	A string specifying the name of the distribution to be studied. The possible distributions are the exponential ("exponential"), the Weibull ("weibull"), the Gumbel ("gumbel"), the normal ("normal"), the lognormal ("lognormal"), the logistic ("logistic"), the loglogistic ("loglogistic"), and the beta ("beta") distribution.
<code>betaLimits</code>	Two-components vector with the lower and upper bounds of the Beta distribution. This argument is only required, if the beta distribution is considered.
<code>igumb</code>	Two-components vector with the initial values for the estimation of the Gumbel distribution parameters.
<code>degs</code>	Integer indicating the number of decimal places of the numeric results of the output.

BS	Number of bootstrap samples.
params0	List specifying the parameters of the theoretical distribution. By default, parameters are set to NULL and estimated with the maximum likelihood method. This argument is only considered, if all parameters of the studied distribution are specified.
tol	Precision of survival times.
formula	A formula with a numeric vector as response (which assumes no censoring) or Surv object.
data	Data frame for variables in formula.
...	Additional arguments.

Details

The function implements the test introduced by Kim (1993) and returns the value of the test statistic. The cell boundaries of the test are obtained via the quantiles, which are based on the Kaplan-Meier estimate of the distribution function. In the presence of right-censored data, it is possible that not all quantiles are estimated, and in this case, the value of M provided by the user is reduced.

The parameter estimation is accomplished with the `fitdistcens` function of the **fitdistrplus** package.

The precision of the survival times is important mainly in the data generation step of the bootstrap samples.

Value

`chisqcens` returns an object of class "chisqcens".

An object of class "chisqcens" is a list containing the following components:

Distribution	Null distribution.
Null hypothesis	Parameters under the null hypothesis (if <code>params0</code> is provided).
Statistic	Value of the test statistic.
p-value	Estimated p-value.
Parameters	The values of the parameters of the null distribution. If the user has set the parameters manually, these will be the returned parameters, otherwise the maximum likelihood estimates are returned.
CellNumber	Vector with two values: the original cell number introduced by the user and the final cell number used.

Author(s)

K. Langohr, M. Besalú, M. Francisco, G. Gómez.

References

J. H. Kim. *Chi-Square Goodness-of-Fit Tests for Randomly Censored Data*. In: *The Annals of Statistics*, 21 (3) (1993), 1621-1639.

Examples

```
# Complete data
set.seed(123)
chisqcens(times = rgumbel(100, 12, scale = 4), M = 8, distr = "gumbel",
          BS = 99)

## Not run:
# Censored data
data(colon)
set.seed(123)
colonsamp <- colon[sample(nrow(colon), 300), ]
chisqcens(Surv(time, status) ~ 1, colonsamp, M = 6, distr = "normal")

## End(Not run)
```

cumhazPlot	<i>Cumulative hazard plots to check the goodness of fit of parametric models</i>
------------	--

Description

Function cumhazPlot uses the cumulative hazard plot to check if a certain distribution is an appropriate choice for the data.

Usage

```
## Default S3 method:
cumhazPlot(times, cens = rep(1, length(times)), distr = "all6", colour = 1,
          betaLimits = c(0, 1), igumb = c(10, 10), ggp = FALSE, m = NULL,
          prnt = TRUE, degs = 3, ...)
## S3 method for class 'formula'
cumhazPlot(formula, data, ...)
```

Arguments

times	Numeric vector of times until the event of interest.
cens	Status indicator (1, exact time; 0, right-censored time). If not provided, all times are assumed to be exact.
distr	A string specifying the names of the distributions to be studied. The possible distributions are the exponential ("exponential"), the Weibull ("weibull"), the Gumbel ("gumbel"), the normal ("normal"), the lognormal ("lognormal"), the logistic ("logistic"), the loglogistic ("loglogistic"), and the beta ("beta") distribution. By default, distr is set to "all6", which means that the cumulative hazard plots are drawn for the Weibull, loglogistic, lognormal, Gumbel, logistic, and normal distributions.
colour	Colour of the points. Default colour: black.

betaLimits	Two-components vector with the lower and upper bounds of the Beta distribution. This argument is only required, if the beta distribution is considered.
igumb	Two-components vector with the initial values for the estimation of the Gumbel distribution parameters.
ggp	Logical to use or not the ggplot2 package to draw the plots. Default is FALSE.
m	Optional layout for the plots to be displayed.
prnt	Logical to indicate if the maximum likelihood estimates of the parameters of all distributions considered should be printed. Default is TRUE.
degs	Integer indicating the number of decimal places of the numeric results of the output.
formula	A formula with a numeric vector as response (which assumes no censoring) or Surv object.
data	Data frame for variables in formula.
...	Optional arguments for function par, if ggplot = FALSE.

Details

The cumulative hazard plot is based on transforming the cumulative hazard function Λ in such a way that it becomes linear in t or $\log(t)$. This transformation is specific for each distribution. The function uses the data to compute the Nelson-Aalen estimator of the cumulative hazard function, $\hat{\Lambda}$, and the maximum likelihood estimators of the parameters of the theoretical distribution under study. If the distribution fits the data, the plot is expected to be a straight line.

The parameter estimation is accomplished with the `fitdistcens` function of the **fitdistrplus** package.

Value

If `prnt = TRUE`, the following output is returned:

Parameter estimates

A list with the maximum likelihood estimates of the parameters of all distributions considered.

In addition, a list with the same contents is returned invisibly.

Author(s)

K. Langohr, M. Besalú, M. Francisco, G. Gómez.

Examples

```
# Complete data and default distributions
set.seed(123)
x <- rlogis(1000, 50, 5)
cumhazPlot(x, lwd = 2)

# Censored data comparing three distributions
data(nba)
cumhazPlot(Surv(survtime, cens) ~ 1, nba, distr = c("expo", "normal", "gumbel"))
```

CvMcens

*Cramér-von Mises test for complete and right-censored data***Description**

Function CvMcens computes the Cramér-von Mises statistic and p-value for complete and right-censored data against eight possible distributions.

Usage

```
## Default S3 method:
CvMcens(times, cens = rep(1, length(times)),
        distr = c("exponential", "gumbel", "weibull", "normal",
                  "lognormal", "logistic", "loglogistic", "beta"),
        betaLimits = c(0, 1), igumb = c(10, 10), degs = 3, BS = 999,
        params0 = list(shape = NULL, shape2 = NULL,
                       location = NULL, scale = NULL), tol = 1e-04, ...)

## S3 method for class 'formula'
CvMcens(formula, data, ...)
```

Arguments

times	Numeric vector of times until the event of interest.
cens	Status indicator (1, exact time; 0, right-censored time). If not provided, all times are assumed to be exact.
distr	A string specifying the name of the distribution to be studied. The possible distributions are the exponential ("exponential"), the Weibull ("weibull"), the Gumbel ("gumbel"), the normal ("normal"), the lognormal ("lognormal"), the logistic ("logistic"), the loglogistic ("loglogistic"), and the beta ("beta") distribution.
betaLimits	Two-components vector with the lower and upper bounds of the Beta distribution. This argument is only required, if the beta distribution is considered.
igumb	Two-components vector with the initial values for the estimation of the Gumbel distribution parameters.
degs	Integer indicating the number of decimal places of the numeric results of the output.
BS	Number of bootstrap samples.
params0	List specifying the parameters of the theoretical distribution. By default, parameters are set to NULL and estimated with the maximum likelihood method. This argument is only considered, if all parameters of the studied distribution are specified.
tol	Precision of survival times.
formula	A formula with a numeric vector as response (which assumes no censoring) or Surv object.
data	Data frame for variables in formula.
...	Additional arguments.

Details

Koziol and Green (1976) proposed a Cramér-von Mises statistic for randomly censored data. This function reproduces this test for a given survival data and a theoretical distribution. In presence of ties, different authors provide slightly different definitions of the product-limit estimator, what might provide different values of the test statistic.

The parameter estimation is accomplished with the `fitdistcens` function of the **fitdistrplus** package.

To avoid long computation times due to bootstrapping, an alternative with complete data is the function `cvm.test` of the **gofest** package.

The precision of the survival times is important mainly in the data generation step of the bootstrap samples.

Value

CvMcens returns an object of class "CvMcens".

An object of class "CvMcens" is a list containing the following components:

Distribution	Null distribution.
Null hypothesis	Parameters under the null hypothesis (if <code>params0</code> is provided).
CvM	Value of the Cramér-von Mises statistic.
p-value	Estimated p-value.
Parameters	List with the maximum likelihood estimates of the parameters of the distribution under study.

Warning

If the amount of data is large, the execution time of the function can be elevated. The parameter `BS` can limit the number of random censored samples generated and reduce the execution time.

Author(s)

K. Langohr, M. Besalú, M. Francisco, G. Gómez.

References

- J. A. Koziol and S. B. Green. *A Cramér-von Mises statistic for randomly censored data*. In: *Biometrika*, 63 (3) (1976), 465-474.
- A. N. Pettitt and M. A. Stephens. *Modified Cramér-von Mises statistics for censored data*. In: *Biometrika*, 63 (2) (1976), 291-298.

See Also

Function `cvm.test` (Package **gofest**) for complete data and `gofcens` for statistics and p-value of Kolmogorov-Smirnov, Cramér von-Mises and Anderson-Darling together for right-censored data.

Examples

```
# Complete data
set.seed(123)
CvMcens(times = rweibull(100, 12, scale = 4), distr = "weibull",
        BS = 199)

## Not run:
# Censored data
data(colon)
set.seed(123)
colonsamp <- colon[sample(nrow(colon), 300), ]
CvMcens(Surv(time, status) ~ 1, colonsamp, distr = "normal")

## End(Not run)
```

gofcens	<i>Kolmogorov-Smirnov, Cramér-von Mises, and Anderson-Darling statistics for complete and right-censored data</i>
---------	---

Description

Function `gofcens` computes the Kolmogorov-Smirnov, Cramér-von Mises, and Anderson-Darling statistics and p-values for complete and right-censored data against eight possible distributions.

Usage

```
## Default S3 method:
gofcens(times, cens = rep(1, length(times)),
        distr = c("exponential", "gumbel", "weibull", "normal",
                  "lognormal", "logistic", "loglogistic", "beta"),
        betaLimits = c(0, 1), igumb = c(10, 10), degs = 3, BS = 999,
        params0 = list(shape = NULL, shape2 = NULL, location = NULL,
                       scale = NULL), ...)

## S3 method for class 'formula'
gofcens(formula, data, ...)
```

Arguments

<code>times</code>	Numeric vector of times until the event of interest.
<code>cens</code>	Status indicator (1, exact time; 0, right-censored time). If not provided, all times are assumed to be exact.
<code>distr</code>	A string specifying the name of the distribution to be studied. The possible distributions are the exponential ("exponential"), the Weibull ("weibull"), the Gumbel ("gumbel"), the normal ("normal"), the lognormal ("lognormal"), the logistic ("logistic"), the loglogistic ("loglogistic"), and the beta ("beta") distribution.

betaLimits	Two-components vector with the lower and upper bounds of the Beta distribution. This argument is only required, if the beta distribution is considered.
igumb	Two-components vector with the initial values for the estimation of the Gumbel distribution parameters.
degs	Integer indicating the number of decimal places of the numeric results of the output.
BS	Number of bootstrap samples.
params0	List specifying the parameters of the theoretical distribution. By default, parameters are set to NULL and estimated with the maximum likelihood method. This argument is only considered, if all parameters of the studied distribution are specified.
formula	A formula with a numeric vector as response (which assumes no censoring) or Surv object.
data	Data frame for variables in formula.
...	Additional arguments.

Details

Fleming et al. (1980) proposed a modified Kolmogorov-Smirnov test to be used with right-censored data. Koziol and Green (1976) proposed a Cramér-von Mises statistic for randomly censored data. This function reproduces this test for a given survival data and a theoretical distribution. In presence of ties, different authors provide slightly different definitions of the product-limit estimator, what might provide different values of the test statistic.

When dealing with complete data, we recommend the use of functions `ks.test` of the **stats** package and `cvm.test` and `ad.test` of the **gofest** package.

Value

`gofcens` returns an object of class `"gofcens"`.

An object of class `"gofcens"` is a list containing the following components:

Distribution	Null distribution.
Null hypothesis	Parameters under the null hypothesis (if <code>params0</code> is provided).
Tests statistics	Values of the Kolmogorov-Smirnov, Cramér-von Mises, and Anderson-Darling test statistics.
p-value	P-values associated with each test.
Parameters	List with the maximum likelihood estimates of the parameters of the distribution under study.

Warning

If the amount of data is large, the execution time of the function can be elevated. The parameter `BS` can limit the number of random censored samples generated and reduce the execution time.

Author(s)

K. Langohr, M. Besalú, M. Francisco, G. Gómez.

References

T. R. Fleming et al. *Modified Kolmogorov-Smirnov test procedure with application to arbitrarily right-censored data*. In: *Biometrics* 36 (1980), 607-625.

J. A. Koziol and S. B. Green. *A Cramér-von Mises statistic for randomly censored data*. In: *Biometrika*, 63 (3) (1976), 465-474.

A. N. Pettitt and M. A. Stephens. *Modified Cramér-von Mises statistics for censored data*. In: *Biometrika*, 63 (2) (1976), 291-298.

See Also

[ks.test](#) (Package stats), [cvm.test](#) (Package goftest), and [ad.test](#) (Package goftest) for complete data, and [KScens](#) for the Kolmogorov-Smirnov test for right-censored data, which returns the p-value.

Examples

```
## Not run:
# Complete data
set.seed(123)
gofcens(times = rweibull(100, 12, scale = 4), distr = "weibull", BS = 499)

# Censored data
data(colon)
set.seed(123)
colonsamp <- colon[sample(nrow(colon), 300), ]
gofcens(Surv(time, status) ~ 1, colonsamp, distr = "normal")

## End(Not run)
```

kmPlot

Plot of the Kaplan-Meier and parametric estimations

Description

Function `kmPlot` is a function that generates a plot that combines a Kaplan-Meier survival curve and a parametric survival curve in the same graph. It is useful for comparing non-parametric survival estimates with the fitted parametric survival model.

Usage

```
## Default S3 method:
kmPlot(times, cens = rep(1, length(times)), distr = "all6", colour = 1,
        betaLimits = c(0, 1), igumb = c(10, 10), ggp = FALSE, m = NULL,
        prnt = TRUE, degs = 3, ...)
## S3 method for class 'formula'
kmPlot(formula, data, ...)
```

Arguments

times	Numeric vector of times until the event of interest.
cens	Status indicator (1, exact time; 0, right-censored time). If not provided, all times are assumed to be exact.
distr	A string specifying the name of the distribution to be studied. The possible distributions are the Weibull ("weibull"), the Gumbel ("gumbel"), the normal ("normal"), the lognormal ("lognormal"), the logistic ("logistic"), the loglogistic ("loglogistic"), the exponential ("exponential") and the beta ("beta") distribution. Default is "all6" and includes the first 6 listed which are the most used distributions.
colour	Vector indicating the colours of the displayed plots.
betaLimits	Two-components vector with the lower and upper bounds of the Beta distribution. This argument is only required, if the beta distribution is considered.
igumb	Two-components vector with the initial values for the estimation of the Gumbel distribution parameters.
ggp	Logical to use or not the ggplot2 package to draw the plots. Default is FALSE.
m	Optional layout for the plots to be displayed.
prnt	Logical to indicate if the maximum likelihood estimates of the parameters should be printed. Default is TRUE.
degs	Integer indicating the number of decimal places of the numeric results of the output.
formula	A formula with a numeric vector as response (which assumes no censoring) or Surv object.
data	Data frame for variables in formula.
...	Optional arguments for function par, if ggp = FALSE.

Details

The parameter estimation is accomplished with the `fitdistcens` function of the **fitdistrplus** package.

Value

If `prnt = TRUE`, the following output is returned:

Distribution	Distribution under study.
--------------	---------------------------

Parameters List with the maximum likelihood estimates of the parameters of the distribution under study.

In addition, a list with the same contents is returned invisibly.

Author(s)

K. Langohr, M. Besalú, M. Francisco, G. Gómez.

References

Peterson Jr, Arthur V. *Expressing the Kaplan-Meier estimator as a function of empirical subsurvival functions*. In: Journal of the American Statistical Association 72.360a (1977): 854-858.

Examples

```
# Plots for complete data and default distributions
set.seed(123)
x <- rexp(1000, 0.1)
kmPlot(x)

# Plots for censored data using ggplot2
data(colon)
kmPlot(Surv(time, status) ~ 1, colon, distr = "lognormal", ggp = TRUE)

# Plots for censored data from three distributions
data(nba)
kmPlot(Surv(survtime, cens) ~ 1, nba, distr = c("normal", "weibull", "lognormal"),
       prnt = FALSE)
```

KScens

Kolmogorov-Smirnov test for complete and right-censored data

Description

Function KScens computes the Kolmogorov-Smirnov statistic and p-value for complete and right-censored data against eight possible distributions.

Usage

```
## Default S3 method:
KScens(times, cens = rep(1, length(times)),
       distr = c("exponential", "gumbel", "weibull", "normal",
                "lognormal", "logistic", "loglogistic", "beta"),
       betaLimits = c(0, 1), igumb = c(10, 10), degs = 3,
       params0 = list(shape = NULL, shape2 = NULL, location = NULL,
                      scale = NULL), ...)

## S3 method for class 'formula'
KScens(formula, data, ...)
```


Arguments

<code>times</code>	Numeric vector of times until the event of interest.
<code>cens</code>	Status indicator (1, exact time; 0, right-censored time). If not provided, all times are assumed to be exact.
<code>distr</code>	A string specifying the name of the distribution to be studied. The possible distributions are the exponential ("exponential"), the Weibull ("weibull"), the Gumbel ("gumbel"), the normal ("normal"), the lognormal ("lognormal"), the logistic ("logistic"), the loglogistic ("loglogistic"), and the beta ("beta") distribution.
<code>betaLimits</code>	Two-components vector with the lower and upper bounds of the Beta distribution. This argument is only required, if the beta distribution is considered.
<code>igumb</code>	Two-components vector with the initial values for the estimation of the Gumbel distribution parameters.
<code>degs</code>	Integer indicating the number of decimal places of the numeric results of the output.
<code>params0</code>	List specifying the parameters of the theoretical distribution. By default, parameters are set to NULL and estimated with the maximum likelihood method. This argument is only considered, if all parameters of the studied distribution are specified.
<code>formula</code>	A formula with a numeric vector as response (which assumes no censoring) or Surv object.
<code>data</code>	Data frame for variables in formula.
<code>...</code>	Additional arguments.

Details

Fleming et al. (1980) proposed a modified Kolmogorov-Smirnov test to use with right-censored data. This function reproduces this test for a given survival data and a theoretical distribution. The p-value is computed following the results of Koziol and Byar (1975) and the output of the function follows the notation of Fleming et al. (1980).

In presence of ties, different authors provide slightly different definitions of $\widehat{F}_n(t)$, with which other values of the test statistic might be obtained.

An alternative with complete data is the function `ks.test` of the **stats** package.

The parameter estimation is accomplished with the `fitdistcens` function of the **fitdistrplus** package.

Value

KScens returns an object of class "KScens".

An object of class "KScens" is a list containing the following components:

<code>Distribution</code>	Null distribution.
<code>Null hypothesis</code>	Parameters under the null hypothesis (if <code>params0</code> is provided).

A	Value of the modified Kolmogorov-Smirnov statistic.
p-value	P-value.
F(y_m)	Estimation of the image of the last recorded time.
y_m	Last recorded time.
Parameters	List with the maximum likelihood estimates of the parameters of the distribution under study.

Author(s)

K. Langohr, M. Besalú, M. Francisco, G. Gómez.

References

T. R. Fleming et al. *Modified Kolmogorov-Smirnov test procedure with application to arbitrarily right-censored data*. In: *Biometrics* 36 (1980), 607-625.

J.A. Koziol and P. Byar. *Percentage Points of the Asymptotic Distributions of One and Two Sample K-S statistics for Truncated or Censored Data*. In: *Technometrics* 17 (4) (1975), 507-510.

See Also

Function [ks.test](#) (Package [stats](#)) for complete data and [gofcens](#) for statistics and p-value of Kolmogorov-Smirnov, Cramér von-Mises and Anderson-Darling together for right-censored data.

Examples

```
# Complete data
set.seed(123)
KScens(times = rweibull(1000, 12, scale = 4), distr = "weibull")

# Censored data
KScens(Surv(time, status) ~ 1, colon, distr = "norm")

data(nba)
KScens(Surv(survtime, cens) ~ 1, nba, "logis", degs = 2)
KScens(Surv(survtime, cens) ~ 1, nba, "beta", betaLimits = c(0, 80))
```

nba

Survival times of former NBA players.

Description

Survival times of former NBA players after their NBA career.

Usage

```
data("nba")
```

Format

A data frame with 3962 observations on the following 3 variables.

id Player ID

survtime Time (in years) from end of NBA career until either death or July 31, 2019.

cens Death indicator (1, exact survival time; 0, right-censored survival time).

Details

The survival times of former NBA players were analyzed by Martínez et al. (2022).

Source

J. A. Martínez, K. Langohr, J. Felipo, L. Consuegra and M. Casals. *Data set on mortality of national basketball association (NBA) players*. In: *Data in Brief*, 45 (2022).

Examples

```
data(nba)
cumhazPlot(Surv(survtime, cens) ~ 1, nba)
```

<code>print.ADcens</code>	<i>Printing method for ADcens object.</i>
---------------------------	---

Description

Printing method for ADcens object.

Usage

```
## S3 method for class 'ADcens'
print(x, prnt = TRUE, outp = c("list", "table"), ...)
```

Arguments

<code>x</code>	An object of class ADcens.
<code>prnt</code>	Logical to indicate if the estimations of the Anderson-Darling statistic and p-value should be printed. Default is TRUE.
<code>outp</code>	Indicator of how the output will be displayed. The possible formats are <code>list</code> and <code>table</code> .
<code>...</code>	Additional arguments.

Value

If `prnt = TRUE`, a list or table (if `outp = "table"`) containing the following components:

Distribution	Null distribution.
Null hypothesis	Parameters under the null hypothesis (if <code>params0</code> is provided).
AD	Value of the Anderson-Darling statistic.
p-value	Estimated p-value.
Parameters	List with the maximum likelihood estimates of the parameters of the distribution under study.

The list is also returned invisibly.

Author(s)

K. Langohr, M. Besalú, M. Francisco, G. Gómez.

Examples

```
# List output
set.seed(123)
ADcens(times = rweibull(100, 12, scale = 4), distr = "weibull",
        BS = 149)

# Table output
set.seed(123)
print(ADcens(times = rweibull(100, 12, scale = 4), distr = "weibull",
             BS = 99), outp = "table")
```

`print.chisqcens` *Printing method for chisqcens object.*

Description

Printing method for `chisqcens` object.

Usage

```
## S3 method for class 'chisqcens'
print(x, prnt = TRUE, outp = c("list", "table"), ...)
```

Arguments

<code>x</code>	An object of class <code>chisqcens</code> .
<code>prnt</code>	Logical to indicate if the estimations of the chi-squared statistic and p-value should be printed. Default is <code>TRUE</code> .
<code>outp</code>	Indicator of how the output will be displayed. The possible formats are <code>list</code> and <code>table</code> .
<code>...</code>	Additional arguments.

Value

If `prnt = TRUE`, a list or table (if `outp = "table"`) containing the following components:

Distribution	Null distribution.
Null hypothesis	Parameters under the null hypothesis (if <code>params0</code> is provided).
Statistic	Value of the test statistic.
p-value	Estimated p-value.
Parameters	The values of the parameters of the null distribution. If the user has set the parameters manually, these will be the returned parameters, otherwise the maximum likelihood estimates are returned.
CellNumber	Vector with two values: the original cell number introduced by the user and the final cell number used.

The list is also returned invisibly.

Author(s)

K. Langohr, M. Besalú, M. Francisco, G. Gómez.

Examples

```
# List output
set.seed(123)
chisqcens(times = rweibull(100, 12, scale = 4), M = 8, distr = "weibull",
           BS = 149)

# Table output
set.seed(123)
print(chisqcens(times = rweibull(100, 12, scale = 4), M = 8, distr = "weibull",
               BS = 99), outp = "table")
```

print.CvMcens *Printing method for CvMcens object.*

Description

Printing method for CvMcens object.

Usage

```
## S3 method for class 'CvMcens'
print(x, prnt = TRUE, outp = c("list", "table"), ...)
```

Arguments

x	An object of class CvMcens.
prnt	Logical to indicate if the estimations of the Cramér-von Mises statistic and p-value should be printed. Default is TRUE.
outp	Indicator of how the output will be displayed. The possible formats are list and table.
...	Additional arguments.

Value

If prnt = TRUE, a list or table (if outp = "table") containing the following components:

Distribution	Null distribution.
Null hypothesis	Parameters under the null hypothesis (if params0 is provided).
CvM	Value of the Cramér-von Mises statistic.
p-value	Estimated p-value.
Parameters	List with the maximum likelihood estimates of the parameters of the distribution under study.

The list is also returned invisibly.

Author(s)

K. Langohr, M. Besalú, M. Francisco, G. Gómez.

Examples

```
# List output
set.seed(123)
CvMcens(times = rweibull(100, 12, scale = 4), distr = "weibull",
        BS = 149)

# Table output
set.seed(123)
print(CvMcens(times = rweibull(100, 12, scale = 4), distr = "weibull",
             BS = 99), outp = "table")
```

print.gofcens

Printing method for gofcens object.

Description

Printing method for gofcens object.

Usage

```
## S3 method for class 'gofcens'
print(x, prnt = TRUE, outp = c("list", "table"), ...)
```

Arguments

x	An object of class gofcens.
prnt	Logical to indicate if the values of the Kolmogorov-Smirnov, Cramér-von Mises, and Anderson-Darling test statistics along with the p-values should be printed. Default is TRUE.
outp	Indicator of how the output will be displayed. The possible formats are list and table.
...	Additional arguments.

Value

If prnt = TRUE, a list or table (if outp = "table") containing the following components:

Distribution	Null distribution.
Null hypothesis	Parameters under the null hypothesis (if params0 is provided).
Tests statistics	Values of the Kolmogorov-Smirnov, Cramér-von Mises, and Anderson-Darling test statistics.
p-value	P-values associated with each test.
Parameters	List with the maximum likelihood estimates of the parameters of the distribution under study.

The list is also returned invisibly.

Author(s)

K. Langohr, M. Besalú, M. Francisco, G. Gómez.

Examples

```
## Not run:
# List output
set.seed(123)
gofcens(times = rweibull(100, 12, scale = 4), distr = "weibull",
        BS = 149)

# Table output
set.seed(123)
print(gofcens(times = rweibull(100, 12, scale = 4), distr = "weibull",
             BS = 149), outp = "table")

## End(Not run)
```

print.KScens *Printing method for KScens object.*

Description

Printing method for KScens object.

Usage

```
## S3 method for class 'KScens'
print(x, prnt = TRUE, outp = c("list", "table"), ...)
```

Arguments

x	An object of class KScens.
prnt	Logical to indicate if the estimations of the Kolmogorov-Smirnov statistic and p-value should be printed. Default is TRUE.
outp	Indicator of how the output will be displayed. The possible formats are list and table.
...	Additional arguments.

Value

If prnt = TRUE, a list or table (if outp = "table") containing the following components:

Distribution	Null distribution.
Null hypothesis	Parameters under the null hypothesis (if params0 is provided).
A	Value of the modified Kolmogorov-Smirnov statistic.
p-value	P-value.
F(y_m)	Estimation of the image of the last recorded time.
y_m	Last recorded time.
Parameters	List with the maximum likelihood estimates of the parameters of the distribution under study.

The list is also returned invisibly.

Author(s)

K. Langohr, M. Besalú, M. Francisco, G. Gómez.

Examples

```
# List output
set.seed(123)
KScens(times = rweibull(1000, 12, scale = 4), distr = "weibull")

# Table output
set.seed(123)
print(KScens(times = rweibull(1000, 12, scale = 4), distr = "weibull"),
      outp = "table")
```

 probPlot

Probability plots to check the goodness of fit of parametric models

Description

probPlot provides four types of probability plots: P-P plot, Q-Q plot, Stabilised probability plot, and Empirically Rescaled plot to check if a certain distribution is an appropriate choice for the data.

Usage

```
## Default S3 method:
probPlot(times, cens = rep(1, length(times)),
         distr = c("exponential", "gumbel", "weibull", "normal",
                  "lognormal", "logistic", "loglogistic", "beta"),
         plots = c("PP", "QQ", "SP", "ER"),
         colour = c("green4", "deepskyblue4", "yellow3",
                   "mediumvioletred"), mtitle = TRUE, ggp = FALSE,
         m = NULL, betaLimits = c(0, 1), igumb = c(10, 10),
         prnt = TRUE, degs = 3,
         params0 = list(shape = NULL, shape2 = NULL,
                       location = NULL, scale = NULL), ...)

## S3 method for class 'formula'
probPlot(formula, data, ...)
```

Arguments

times	Numeric vector of times until the event of interest.
cens	Status indicator (1, exact time; 0, right-censored time). If not provided, all times are assumed to be exact.
distr	A string specifying the name of the distribution to be studied. The possible distributions are the exponential ("exponential"), the Weibull ("weibull"), the Gumbel ("gumbel"), the normal ("normal"), the lognormal ("lognormal"), the logistic ("logistic"), the loglogistic ("loglogistic"), and the beta ("beta") distribution.
plots	Vector stating the plots to be displayed. Possible choices are the P-P plot ("PP"), the Q-Q plot ("QQ"), the Stabilised Probability plot ("SP"), and the Empirically Rescaled plot ("ER"). By default, all four plots are displayed.

colour	Vector indicating the colours of the displayed plots. The vector will be recycled if its length is smaller than the number of plots to be displayed.
mtitle	Logical to add or not the title "Probability plots for a distr distribution" to the plot. Default is TRUE.
ggp	Logical to use or not the ggplot2 package to draw the plots. Default is FALSE.
m	Optional layout for the plots to be displayed.
betaLimits	Two-components vector with the lower and upper bounds of the Beta distribution. This argument is only required, if the beta distribution is considered.
igumb	Two-components vector with the initial values for the estimation of the Gumbel distribution parameters.
prnt	Logical to indicate if the maximum likelihood estimates of the parameters should be printed. Default is TRUE.
degs	Integer indicating the number of decimal places of the numeric results of the output.
params0	List specifying the parameters of the theoretical distribution. By default, parameters are set to NULL and estimated with the maximum likelihood method. This argument is only considered, if all parameters of the studied distribution are specified.
formula	A formula with a numeric vector as response (which assumes no censoring) or Surv object.
data	Data frame for variables in formula.
...	Optional arguments for function par, if ggp = FALSE.

Details

By default, function `probPlot` draws four plots: P-P plot, SP plot, Q-Q plot, and EP plot. Following, a description is given for each plot.

The **Probability-Probability plot** (P-P plot) depicts the empirical distribution, $\widehat{F}(t)$, which is obtained with the Kaplan-Meier estimator if data are right-censored, versus the theoretical cumulative distribution function (cdf), $\widehat{F}_0(t)$. If the data come from the chosen distribution, the points of the resulting graph are expected to lie on the identity line.

The **Stabilised Probability plot** (SP plot), proposed by Michael (1983), is a transformation of the P-P plot. It stabilises the variance of the plotted points. If $F_0 = F$ and the parameters of F_0 are known, $\widehat{F}_0(t)$ corresponds to the cdf of a uniform order statistic, and the arcsin transformation stabilises its variance. If the data come from distribution F_0 , the SP plot will resemble the identity line.

The **Quartile-Quartile plot** (Q-Q plot) is similar to the P-P plot, but it represents the sample quantiles versus the theoretical ones, that is, it plots t versus $\widehat{F}_0^{-1}(\widehat{F}(t))$. Hence, if F_0 fits the data well, the resulting plot will resemble the identity line.

A drawback of the Q-Q plot is that the plotted points are not evenly spread. Waller and Turnbull (1992) proposed the **Empirically Rescaled plot** (EP plot), which plots $\widehat{F}_u(t)$ against $\widehat{F}_u(\widehat{F}_0^{-1}(\widehat{F}(t)))$, where $\widehat{F}_u(t)$ is the empirical cdf of the points corresponding to the uncensored observations. Again, if \widehat{F}_0 fits the data well, the ER plot will resemble the identity line.

By default, all four probability plots are drawn and the maximum likelihood estimates of the parameters of the chosen parametric model are returned. The parameter estimation is accomplished with the `fitdistrib` function of the **fitdistrplus** package.

Value

If `prnt = TRUE`, the following output is returned:

Distribution	Distribution under study.
Parameters	Parameters used to draw the plots (if <code>params0</code> is provided).
Estimates	Maximum likelihood estimates of the parameters.

In addition, a list with the same contents is returned invisibly.

Author(s)

K. Langohr, M. Besalú, M. Francisco, G. Gómez.

References

J. R. Michael. *The Stabilized Probability Plot*. In: *Biometrika* 70 (1) (1983), 11-17.
 L.A. Waller and B.W. Turnbull. *Probability Plotting with Censored Data*. In: *American Statistician* 46 (1) (1992), 5-12.

Examples

```
# P-P, Q-Q, SP, and EP plots for complete data
set.seed(123)
x <- rlnorm(1000, 3, 2)
probPlot(x)
probPlot(x, distr = "lognormal")

# P-P, Q-Q, SP, and EP plots for censored data using ggplot2
data(colon)
probPlot(Surv(time, status) ~ 1, colon, "weibull", ggp = TRUE)

# P-P, Q-Q and SP plots for censored data and lognormal distribution
data(nba)
probPlot(Surv(survtime, cens) ~ 1, nba, "lognorm", plots = c("PP", "QQ", "SP"),
         ggp = TRUE, m = matrix(1:3, nr = 1))
```

Index

* datasets

nba, [18](#)

ADcens, [4](#)

chisqcens, [6](#)

class, [5](#), [7](#), [11](#), [13](#), [17](#)

cumhazPlot, [8](#)

CvMcens, [10](#)

GofCens (GofCens-package), [2](#)

gofcens, [5](#), [11](#), [12](#), [18](#)

GofCens-package, [2](#)

kmPlot, [14](#)

ks.test, [13](#), [14](#), [17](#), [18](#)

KScens, [14](#), [16](#)

nba, [18](#)

plot.probPlot (probPlot), [25](#)

print.ADcens, [19](#)

print.chisqcens, [20](#)

print.CvMcens, [21](#)

print.gofcens, [22](#)

print.KScens, [24](#)

print.probPlot (probPlot), [25](#)

probPlot, [25](#)