Package 'EL'

January 20, 2025

Version 1.3

Date 2024-10-28

Title Two-Sample Empirical Likelihood

License GPL (>= 2)

Description Empirical likelihood (EL) inference for two-sample problems. The following statistics are included: the difference of two-sample means, smooth Huber estimators, quantile (qdiff) and cumulative distribution functions (ddiff), probability-probability (P-P) and quantile-quantile (Q-Q) plots as well as receiver operating characteristic (ROC) curves. EL calculations are based on J. Valeinis, E. Cers (2011) <http://home.lu.lv/~valeinis/lv/petnieciba/EL_TwoSample_2011.pdf>.

Author Janis Valeinis [aut] (<https://orcid.org/0000-0003-0989-0444>),
Edmunds Cers [aut],
Janis Gredzens [cre],
Reinis Alksnis [ctb]

Maintainer Janis Gredzens <janis.gredzens@lu.lv>

Repository CRAN

NeedsCompilation yes

Date/Publication 2024-10-28 10:20:02 UTC

RoxygenNote 7.3.2

Encoding UTF-8

Imports stats, nleqslv, ggplot2

Contents

BEL.means — *The two-sample blockwise empirical likelihood statistic for differences in means*

Description

Calculates blockwise empirical likelihood test for the difference of two sample means.

Usage

```
BEL.means(X, Y, M_1, M_2, Delta = 0)
```

Arguments

X, Y	vectors of data values.
M_1, M_2	positive integers specifying block length for X and Y, respectively.
Delta	hypothesized difference of two populations.

Value

A list of class "htest" containing following components: method - the character string of the test. data.name - a character string with the names of the input data. Delta0 - the specified hypothesized value of mean differences under the null hypothesis statistic - the value of the test statistic. p.value - the p-value for the test.

Examples

```
# Basic example
Delta0 <- 1.5
X <- arima.sim(n = 400, model = list(ar = .3))
Y <- arima.sim(n = 400, model = list(ar = .5)) + Delta0
BEL.means(X, Y, 10, 20, Delta = Delta0)
```

EL.Huber — *Empirical likelihood test for the difference of smoothed Huber estimators*

Description

Empirical likelihood inference for the difference of smoothed Huber estimators. This includes a test for the null hypothesis for a constant difference of smoothed Huber estimators, confidence interval and EL estimator.

Usage

```
EL.Huber(X, Y, mu = 0, conf.level = 0.95,
         scaleX=1, scaleY=1, VX = 2.046, VY = 2.046, k = 1.35)
```

Arguments

X	a vector of data values.
Y	a vector of data values.
mu	a number specifying the null hypothesis.
conf.level	confidence level of the interval.
scaleX	the scale estimate of sample 'X'.
scaleY	the scale estimate of sample 'Y'.
VX	the asymptotic variance of initial (nonsmooth) Huber estimator for the sample 'X'.
VY	the asymptotic variance of initial (nonsmooth) Huber estimator for the sample 'Y'.
k	tuning parameter for the Huber estimator.

Details

A common choice for a robust scale estimate (parameters scaleX and scaleY) is the mean absolute deviation (MAD).

Value

A list of class 'htest' containing the following components:

estimate	the empirical likelihood estimate for the difference of two smoothed Huber estimators.
conf.int	a confidence interval for the difference of two smoothed Huber estimators.
p.value	the p-value for the test.
statistic	the value of the test statistic.
method	the character string 'Empirical likelihood smoothed Huber estimator difference test'.
null.value	the specified hypothesized value of the mean difference 'mu' under the null hypothesis.
data.name	a character string giving the names of the data.

Author(s)

E. Cers, J. Valeinis

References

J. Valeinis, E. Cers. Extending the two-sample empirical likelihood. To be published. Preprint available at http://home.lanet.lv/~valeinis/lv/petnieciba/EL_TwoSample_2011.pdf.

F. Hampel, C. Hennig and E. A. Ronchetti (2011). A smoothing principle for the Huber and other location M-estimators, Computational Statistics & Data Analysis, 55(1), 324-337.

See Also

EL.means

Examples

```
X <- rnorm(100)
Y <- rnorm(100)
t.test(X, Y)
EL.means(X, Y)
EL.Huber(X, Y)
```

`EL.means` *Empirical likelihood test for the difference of two sample means*

Description

Empirical likelihood inference for the difference of two sample means. This includes a test for the null hypothesis for a constant difference of mean difference, confidence interval and EL estimator.

Usage

```
EL.means(X, Y, mu = 0, conf.level = 0.95)
```

Arguments

`X`	a vector of data values.
`Y`	a vector of data values.
`mu`	a number specifying the null hypothesis.
`conf.level`	confidence level of the interval.

Value

A list of class 'htest' containing the following components:

`estimate`	the empirical likelihood estimate of the mean difference.
`conf.int`	a confidence interval for the mean difference.
`p.value`	the p-value for the test.
`statistic`	the value of the test statistic.
`method`	the character string 'Empirical likelihood mean difference test'.

null.value	the specified hypothesized value of mean differences 'mu' under the null hypothesis.
data.name	a character string giving the names of the data.

Author(s)

E. Cers, J. Valeinis

References

J. Valeinis, E. Cers and J. Cielens (2010). Two-sample problems in statistical data modelling. Mathematical modelling and analysis, 15(1), 137-151.

J. Valeinis, E. Cers. Extending the two-sample empirical likelihood. To be published. Preprint available at http://home.lanet.lv/~valeinis/lv/petnieciba/EL_TwoSample_2011.pdf.

See Also

EL.Huber

Examples

```
X <- rnorm(100)
Y <- rnorm(100)
t.test(X, Y)
EL.means(X, Y)
EL.Huber(X, Y)
```

EL.plot *Draws plots using the smoothed two-sample empirical likelihood method*

Description

Draws P-P and Q-Q plots, ROC curves, quantile differences (qdiff) and CDF differences (ddiff) and their respective confidence bands (pointwise or simultaneous) using the empirical likelihood method.

Usage

```
EL.plot(method, X, Y, bw = bw.nrd0, conf.level = NULL,
        simultaneous = FALSE, bootstrap.samples = 300,
        more.warnings = FALSE, ...)
```

Arguments

method	"pp", "qq", "roc", "qdiff" or "fdiff".
X	a vector of data values.
Y	a vector of data values.
bw	a function taking a vector of values and returning the corresponding bandwidth or a vector of two values corresponding to the respective bandwidths of X and Y.
conf.level	confidence level for the intervals. A number between 0 and 1 or NULL when no confidence bands should be calculated. Depending on the value of 'simultaneous' either pointwise intervals or simultaneous confidence bands will be drawn.
simultaneous	if this is TRUE, simultaneous confidence bands will be constructed, using a nonparametric bootstrap procedure to select the level of confidence bands. The default is FALSE, in which case simple pointwise confidence bands are calculated.
bootstrap.samples	the number of samples used to bootstrap the simultaneous confidence bands when 'simultaneous = TRUE'.
more.warnings	if this is FALSE (the default) a single warning will be produced if there is any problem calculating the estimate or the confidence bands. If this is set to TRUE a warning will be produced for every point at which there was a problem.
...	further arguments passed to plot.

Details

The plotting interval for P-P plots, ROC curves and differences of quantile functions is [0, 1] (where these functions are defined). The Q-Q plot is drawn from the minimum to the maximum of 'Y'. Finally, for the plot of distribution function differences the interval from max(min(X), min(Y)) to min(max(X), max(Y)) is used.

Confidence bands are drawn only if 'conf.level' is not 'NULL'.

When constructing simultaneous confidence bands, the plot is drawn on an interval that is narrowed by 5% on both sides, since the procedure is usually sensitive at the end-points, which can result in large bands. The confidence level for the simultaneous confidence bands is bootstrapped using 50 evenly spaced points in this interval. If the default interval produces too large confidence bands, use the function 'EL.smooth' where the intervals are specified manually. Note that calculation of simultaneous confidence bands can take a long time.

Value

none.

Author(s)

E. Cers, J. Valeinis

References

J. Valeinis, E. Cers. Extending the two-sample empirical likelihood. To be published. Preprint available at http://home.lanet.lv/~valeinis/lv/petnieciba/EL_TwoSample_2011.pdf.

P. Hall and A. Owen (1993). Empirical likelihood bands in density estimation. Journal of Computational and Graphical statistics, 2(3), 273-289.

See Also

EL.smooth EL.statistic

Examples

```
## The examples showcase all available graphs

X1 <- rchisq(100, 2.5)
X2 <- rnorm(100, 0, 1)


X1 <- rchisq(100, 2.5)
X2 <- rnorm(100, 0, 1)

# Intro
xlim <- c(min(X1, X2) - 0.5, max(X1, X2) + 0.5)
D1 <- density(X1)
D2 <- density(X2)
ylim <- c(min(D1$y, D2$y), max(D1$y, D2$y))
df <- data.frame(x1 = D1$x, y1 = D1$y, x2 = D2$x, y2 = D2$y)
p1 <- ggplot2::ggplot(data = df) +
   ggplot2::geom_line(ggplot2::aes(x=x2, y=y2, color=paste0('X2 (bw=', round(D2$bw, 2), ')'))) +
   ggplot2::geom_line(ggplot2::aes(x=x1, y=y1, color=paste0('X1 (bw=', round(D1$bw, 2), ')'))) +
    ggplot2::guides(color = ggplot2::guide_legend(title = NULL)) +
    ggplot2::theme_minimal() +
    ggplot2::theme(legend.position = "top") +
    ggplot2::labs(x="X", y="Density")
p1

# CDF differences
p2 <- EL.plot("fdiff", X1, X2, main="F difference", conf.level=0.95)
tt <- seq(max(c(min(X1), min(X2))), min(c(max(X1), max(X2))), length=30)
ee <- ecdf(X2)(tt) - ecdf(X1)(tt)
p2 <- p2 + ggplot2::geom_point(data=data.frame(tt = tt, ee = ee), ggplot2::aes(x=tt, y=ee))
p2

# Quantile differences
p3 <- EL.plot("qdiff", X1, X2, main="Quantile difference", conf.level = 0.95)
tt <- seq(0.01, 0.99, length=30)
ee <- quantile(X2, tt) - quantile(X1, tt)
p3 <- p3 + ggplot2::geom_point(data=data.frame(tt = tt, ee = ee), ggplot2::aes(x=tt, y=ee))
p3

# Q-Q plot
```

```
p4 <- EL.plot("qq", X1, X2, main="Q-Q plot", conf.level=0.95)
tt <- seq(min(X2), max(X2), length=30)
ee <- quantile(X1, ecdf(X2)(tt))
p4 <- p4 + ggplot2::geom_point(data=data.frame(tt = tt, ee = ee), ggplot2::aes(x=tt, y=ee))
p4

# P-P plot
p5 <- EL.plot("pp", X1, X2, main="P-P plot", conf.level=0.95, ylim=c(0,1))
tt <- seq(0.01, 0.99, length=30)
ee <- ecdf(X1)(quantile(X2, tt))
p5 <- p5 + ggplot2::geom_point(data=data.frame(tt = tt, ee = ee), ggplot2::aes(x=tt, y=ee))
p5

# ROC curve
p6 <- EL.plot("roc", X1, X2, main="ROC curve", conf.level=0.95, ylim=c(0,1))
tt <- seq(0.01, 0.99, length=30)
ee <- 1- ecdf(X1)(quantile(X2, 1-tt))
p6 <- p6 + ggplot2::geom_point(data=data.frame(tt = tt, ee = ee), ggplot2::aes(x=tt, y=ee))
p6

# Showing all plots at once is outside of scope from
# these examples but to do so run the following:
# require(cowplot)
# cowplot::plot_grid(p1, p2, p3, p4, p5, p6, ncol = 2)
```

`EL.smooth` *Smooth estimates and confidence intervals (or simultaneous bands) using the smoothed two-sample EL method*

Description

Calculates estimates and pointwise confidence intervals (or simultaneous bands) for P-P and Q-Q plots, ROC curves, quantile differences (qdiff) and CDF differences (ddiff) using the smoothed empirical likelihood method.

Usage

```
EL.smooth(method, X, Y, t, bw = bw.nrd0,
          conf.level = NULL, simultaneous = FALSE,
          bootstrap.samples = 300, more.warnings = FALSE)
```

Arguments

`method`	"pp", "qq", "roc", "qdiff" or "ddiff".
`X`	a vector of data values.
`Y`	a vector of data values.
`t`	a vector of points for which to calculate the estimates and confidence intervals.

`conf.level` confidence level for the intervals. A number between 0 and 1 or NULL when no confidence bands should be calculated. Depending on the value of 'simultaneous' either pointwise intervals or simultaneous confidence bands will be calculated.

`simultaneous` if this is TRUE, simultaneous confidence bands will be constructed, using a nonparametric bootstrap procedure to select the level of confidence bands. The default is FALSE, in which case simple pointwise confidence bands are calculated.

`bootstrap.samples` the number of samples used to bootstrap the simultaneous confidence bands when 'simultaneous = TRUE'.

`bw` a function taking a vector of values and returning the corresponding bandwidth or a vector of two values corresponding to the respective bandwidths of X and Y.

`more.warnings` if this is FALSE (the default) a single warning will be produced if there is any problem calculating the estimate or the confidence bands. If this is set to TRUE a warning will be produced for every point at which there was a problem.

Details

Confidence bands are drawn only if 'conf.level' is not 'NULL'.

When constructing simultaneous confidence bands, it is advisable to check whether the chosen range of 't' values does not produce too large bands (for example, for the P-P plot in the example below the interval [0.05, 0.95] was a sensible choice). This has to be checked for each data sample separately by hand. Note that the calculation of simultaneous confidence bands can take a long time.

Value

`estimate` the estimated values at points 't'.

`conf.int` a two column matrix where each row represents the lower and upper bounds of the confidence bands corresponding to the values at points 't'.

`simultaneous.conf.int` will be a true value if simultaneous confidence bands are constructed.

`bootstrap.crit` the critical value from the bootstrapped -2 * log-likelihood statistic for simultaneous confidence bands using the confidence level 'conf.level'. Only calculated when 'conf.level' is not NULL and 'simultaneous' is TRUE.

Author(s)

E. Cers, J. Valeinis

References

J. Valeinis and E. Cers. Extending the two-sample empirical likelihood. To be published. Preprint available at http://home.lanet.lv/~valeinis/lv/petnieciba/EL_TwoSample_2011.pdf

P. Hall and A. Owen (1993). Empirical likelihood bands in density estimation. Journal of Computational and Graphical statistics, 2(3), 273-289.

See Also

EL.plot EL.statistic

Examples

```
#### Simultaneous confidence bands for a P-P plot
X1 <- rnorm(200)
X2 <- rnorm(200, 1)

x <- seq(0.05, 0.95, length=19)
y <- EL.smooth("pp", X1, X2, x, conf.level=0.95,
               simultaneous=TRUE, bw=c(0.3, 0.3))
conf.int <- data.frame(x = x, ci.l = y$conf.int[1,], ci.u = y$conf.int[2,])

## Plot the graph with both pointwise and simultaneous confidence bands
EL.plot("pp", X1, X2, conf.level=0.95, bw=c(0.3, 0.3)) +
    ggplot2::geom_line(data= conf.int, ggplot2::aes(x=x, y=ci.u), lty="dotted") +
    ggplot2::geom_line(data= conf.int, ggplot2::aes(x=x, y=ci.l), lty="dotted")
```

EL.statistic *The two-sample empirical likelihood statistic*

Description

Calculates -2 times the log-likelihood ratio statistic when the function of interest (either of P-P or Q-Q plot, ROC curve, difference of quantile or distribution functions) at some point 't' is equal to 'd'.

Usage

```
EL.statistic(method, X, Y, d, t, bw = bw.nrd0)
```

Arguments

method	"pp", "qq", "roc", "qdiff" or "fdiff".
X	a vector of data values.
Y	a vector of data values.
d	a number
t	a number.
bw	a function taking a vector of values and returning the corresponding bandwidth or a vector of two values corresponding to the respective bandwidths of X and Y.

Value

-2 times the logarithm of the two-sample empirical likelihood ratio.

Author(s)

E. Cers, J. Valeinis

References

J.Valeinis, E.Cers. Extending the two-sample empirical likelihood. To be published. Preprint available at http://home.lanet.lv/

See Also

EL.smooth

Examples

```
EL.statistic("pp", rnorm(100), rnorm(100), 0.5, 0.5)
```

Index