Package ‘mhde’

October 23, 2015

Type Package
Title Minimum Hellinger Distance Test for Normality
Version 1.0-1
Date 2015-10-21
Author Paul W. Eslinger [aut, cre],
       Heather Orr [aut, ctb]
Maintainer Paul W. Eslinger <paul.w.eslinger@pnnl.gov>
Description Implementation of a goodness-of-fit test for normality using the Minimum Hellinger Distance.
License GPL-2 | GPL-3
LazyData TRUE
URL `https://cran.r-project.org/package=mhde`
Imports graphics, stats
NeedsCompilation no
Repository CRAN
Date/Publication 2015-10-23 08:50:31

R topics documented:

mhde.cn ................................................................. 2
mhde.crit ............................................................. 2
mhde.plot ............................................................. 3
mhde.test ............................................................. 3

Index 6
\textbf{mhde.cn} \quad \textit{Utility function for retrieving c\_n value for any sample size}

\textbf{Description}

This function evaluates the $c_n$ bandwidth value for the Epanechnikov kernel based on the sample size.

\textbf{Usage}

\texttt{mhde.cn(Nsize)}

\textbf{Arguments}

\begin{itemize}
\item \texttt{Nsize} \quad \text{Number of values in the sample data set}
\end{itemize}

\textbf{Value}

The Epanechnikov kernel bandwidth parameter for the sample size.

\textbf{mhde.crit} \quad \textit{Utility function for retrieving the critical value for any sample size and desired $\alpha$ level}

\textbf{Description}

This function calculates the critical value for the minimized Hellinger distance for a given sample size and probability level for a goodness of fit test for normality.

P values below 0.0001 return the value for p=0.0001 P values above 0.9999 return the value for p=0.9999

This utility function is not used by the other functions. It is provided as a "curiosity" for users.

\textbf{Usage}

\texttt{mhde.crit(Nsize, Plevel = 0.95)}

\textbf{Arguments}

\begin{itemize}
\item \texttt{Nsize} \quad \text{The sample size is defined using Nsize. No default value is set.}
\item \texttt{Plevel} \quad \text{The probability level associated with the critical value is set using Plevel. The default value is 0.95. The $\alpha$ level of the test is 1-Plevel.}
\end{itemize}

\textbf{Value}

\texttt{pval} \quad \text{Critical value for the minimized Hellinger distance}
Examples

```r
## example using default alpha level of .05
mhde.crit(45)

## example using alpha level of .01
mhde.crit(45, 0.99)
```

---

**mhde.plot**  
*Plot the non-parametric and normal densities*

---

**Description**

Simple plot function to display the data-based kernel density and the normal density for the final location and scale estimates.

**Usage**

```r
mhde.plot(ListOut)
```

**Arguments**

- `ListOut`: The output list from `mhde.test`. The argument `Small` for `mhde.test` must have the value `FALSE`.

---

**mhde.test**  
*Minimum Hellinger Distance Test for Normality*

---

**Description**

This function fits a normal distribution to a data set using a minimum Hellinger distance approach and then performs a test of hypothesis that the data are from a normal distribution.

**Usage**

```r
mhde.test(DataVec, NGauss = 100, MaxIter = 25, InitLocation, InitScale, EpsLoc = 1e-04, EpsSca = 1e-04, Silent = FALSE, Small = FALSE)
```

**Arguments**

- `DataVec`: The data are supplied by the user in a numeric vector. The length of the vector determines the number of data values.
- `NGauss`: The number of subintervals for the Gauss-Legendre integration techniques is controlled by `NGauss`. A default value of 100 is used. A minimum of 25 is enforced.
**MaxIter**  The maximum number of iterations that can occur in evaluating the minimum Hellinger distance is controlled by `MaxIter`. A default of 25 is used. A minimum of 1 is enforced.

**InitLocation**  An optional initial location estimate can be defined using `InitLocation`. The data median is the default value.

**InitScale**  An optional initial scale estimate can be defined using `InitScale`. The data median absolute deviation is the default value.

**EpsLoc**  The epsilon (in data units) below which the iterative minimization approach declares convergence in the location estimate is controlled by `EpsLoc`. `EpsLoc` should be set to give approximately 5 digits of accuracy in the location estimate. A default value of 0.0001 is used.

**EpsSca**  The epsilon (in data units) below which the iterative minimization approach declares convergence in the SCALE estimate is controlled by `EpsSca`. `EpsSca` should be set to give approximately 5 digits of accuracy in the scale estimate. A default value of 0.0001 is used.

**Silent**  A value of `FALSE` for `Silent` writes several results to the R console. Use `Silent=TRUE` to eliminate the output.

**Small**  A value of `FALSE` for `Small` returns a list of 11 objects. Use `Small=TRUE` to return a shorter list containing only the Hellinger distance and the p-value.

**Details**

Let $f(x)$ and $g(x)$ be absolutely continuous probability density functions. The square of the Hellinger distance can be written as $H^2 = 1 - \int \sqrt{f(x)g(x)} \, dx$. For this package, $f(x)$ denotes the family of normal densities and $g$ is a data-based density obtained by using the Ephanechnikov kernel. The kernel has the form $w(z) = 0.75(1 - z^2)$ for $-1 < z < 1$ and 0 elsewhere. Let the $n$ sample data be denoted by $x_1, ..., x_n$. The data-based kernel density at any point $y$ is calculated from

$$g_n(y) = \frac{1}{n s_n c_n} \sum_{i=1}^{n} w(y - x_i)$$

A Newton-Rhapson method with analytical derivatives is to determine the minimum Hellinger distance. Numerical integration is done using a 6-point composite Gauss-Legendre technique.

**Value**

Values returned in a list include the following items:

- Minimized Hellinger distance
- p-value for the minimized distance
- Initial location used in the iterative solution
- Initial scale used in the iterative solution
- Final location estimate
- Final scale estimate
- Sample size
mhde.test

- Kernel density bandwidth parameter
- Vector of x values used in the integration for the Hellinger distance
- Vector of nonparametric density values at the x values used in the integration
- Vector of normal density values for the estimated location and scale at the x values used in integration

Author(s)

Paul W. Eslinger and Heather M. Orr

References


Examples

```r
## example with a normal data set
mhde.test(rnorm(20),0.0,1.0),Small=TRUE)

## example with a uniform data set including example plot
MyList <- mhde.test(runif(25),min=2,max=4)
mhde.plot(MyList)
```
Index

mhde.cn, 2
mhde.crit, 2
mhde.plot, 3
mhde.test, 3