

# Package ‘MKLE’

August 21, 2023

**Type** Package

**Title** Maximum Kernel Likelihood Estimation

**Version** 1.0.1

**Author** Thomas Jaki

**Maintainer** Thomas Jaki <thomas.jaki@protonmail.com>

**Description** Package for fast computation of the maximum kernel likelihood estimator (mkle).

**License** GPL

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2023-08-21 08:02:38 UTC

## R topics documented:

|                        |   |
|------------------------|---|
| MKLE-package . . . . . | 1 |
| klik . . . . .         | 3 |
| mkle . . . . .         | 4 |
| mkle.ci . . . . .      | 5 |
| opt.bw . . . . .       | 6 |
| state . . . . .        | 8 |

|              |          |
|--------------|----------|
| <b>Index</b> | <b>9</b> |
|--------------|----------|

---

|              |   |
|--------------|---|
| MKLE-package | <i>Maximum kernel likelihood estimation</i> |
|--------------|---|

---

## Description

Computes the maximum kernel likelihood estimator using fast fourier transforms.

**Details**

Package: MKLE  
 Type: Package  
 Version: 1.01  
 Date: 2023-08-21  
 License: GPL

The maximum kernel likelihood estimator is defined to be the value  $\hat{\theta}$  that maximizes the estimated kernel likelihood based on the general location model,

$$f(x|\theta) = f_0(x - \theta).$$

This model assumes that the mean associated with  $f_0$  is zero which of course implies that the mean of  $X_i$  is  $\theta$ . The kernel likelihood is the estimated likelihood based on the above model using a kernel density estimate,  $\hat{f}(\cdot|h, X_1, \dots, X_n)$ , and is defined as

$$\hat{L}(\theta|X_1, \dots, X_n) = \prod_{i=1}^n \hat{f}(X_i - (\bar{X} - \theta)|h, X_1, \dots, X_n).$$

The resulting estimator therefore is an estimator of the mean of  $X_i$ .

**Author(s)**

Thomas Jaki

Maintainer: Thomas Jaki <jaki.thomas@gmail.com>

**References**

Jaki T., West R. W. (2008) Maximum kernel likelihood estimation. *Journal of Computational and Graphical Statistics* Vol. 17(No 4), 976-993.

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, 2nd ed.

**Examples**

```
data(state)
mkle(state$CRIME)
```

---

klik *Kernel log likelihood*

---

### Description

The function computes the kernel log likelihood for a given  $\hat{\theta}$ .

### Usage

```
klik(delta , data, kde, grid, min)
```

### Arguments

|       |   |
|-------|---|
| delta | the difference of the parameter theta for which the kernel log likelihood will be computed and the sample mean. |
| data  | the data for which the kernel log likelihood will be computed.  |
| kde   | an object of the class "density".   |
| grid  | the stepsize between the x-values in kde.   |
| min   | the smallest x-value in kde.  |

### Details

This function is intended to be called through the function [mkle](#) and is optimized for fast computation.

### Value

The log likelihood based on the shifted kernel density estimator.

### Author(s)

Thomas Jaki

### References

Jaki T., West R. W. (2008) Maximum kernel likelihood estimation. *Journal of Computational and Graphical Statistics* Vol. 17(No 4), 976-993.

### See Also

[mkle](#)

**Examples**

```

data(state)
attach(state)
bw<-2*sd(CRIME)
kdensity<-density(CRIME,bw=bw,kernel="biweight",
                  from=min(CRIME)-2*bw,to=max(CRIME)+2*bw,n=2^12)
min<-kdensity$x[1]
grid<-kdensity$x[2]-min

# finds the kernel log likelihood at the sample mean
klik(0,CRIME, kdensity, grid, min)

```

---

mkle

*Maximum kernel likelihood estimation*


---

**Description**

Computes the maximum kernel likelihood estimator for a given dataset and bandwidth.

**Usage**

```

mkle(data,bw=2*sd(data),kernel=c("gaussian", "epanechnikov", "rectangular", "triangular",
                                "biweight", "cosine", "optcosine"),gridsize=2^14)

```

**Arguments**

|          |  |
|----------|--|
| data     | the data for which the estimator should be found.  |
| bw       | the smoothing bandwidth to be used.  |
| kernel   | a character string giving the smoothing kernel to be used. This must be one of "gaussian", "rectangular", "triangular", "epanechnikov", "biweight", "cosine" or "optcosine", with default "gaussian". May be abbreviated to a unique prefix (single letter). |
| gridsize | the number of points at which the kernel density estimator is to be evaluated with $2^{14}$ as the default.  |

**Details**

The default for the bandwidth is  $2s$ , which is the near-optimal value if a Gaussian kernel is used. If the bandwidth is zero, the sample mean will be returned.

Larger gridsize results in more accurate estimates but also longer computation times. The use of gridsizes between  $2^{11}$  and  $2^{20}$  is recommended.

**Value**

The maximum kernel likelihood estimator.

**Note**

[optimize](#) is used for the optimization and [density](#) is used to estimate the kernel density.

**Author(s)**

Thomas Jaki

**References**

Jaki T., West R. W. (2008) Maximum kernel likelihood estimation. *Journal of Computational and Graphical Statistics* Vol. 17(No 4), 976-993.

**See Also**

[klik](#)

**Examples**

```
data(state)
plot(density(state$CRIME))
abline(v=mean(state$CRIME),col='red')
abline(v=mkle(state$CRIME),col='blue')
```

---

mkle.ci

---

*Confidence intervals for the maximum kernel likelihood estimator*


---

**Description**

Computes different confidence intervals for the maximum kernel likelihood estimator for a given dataset and bandwidth.

**Usage**

```
mkle.ci(data, bw=2*sd(data), alpha=0.1, kernel=c("gaussian", "epanechnikov",
"rectangular", "triangular", "biweight", "cosine", "optcosine"),
method=c("percentile", "wald", "boott"), B=1000, gridsize=2^14)
```

**Arguments**

|        |  |
|--------|--|
| data   | the data for which the confidence interval should be found.  |
| bw     | the smoothing bandwidth to be used.  |
| alpha  | the significance level.  |
| kernel | a character string giving the smoothing kernel to be used. This must be one of "gaussian", "rectangular", "triangular", "epanechnikov", "biweight", "cosine" or "optcosine", with default "gaussian", and may be abbreviated to a unique prefix (single letter). |

|          |   |
|----------|---|
| method   | a character string giving the type of interval to be used. This must be one of "percentile", "wald" or "boott". |
| B        | number of resamples used to estimate the mean squared error with 1000 as the default.                           |
| gridsize | the number of points at which the kernel density estimator is to be evaluated with $2^{14}$ as the default.     |

### Details

The method can be a vector of strings containing the possible choices.

The bootstrap-t-interval can be very slow for large datasets and a large number of resamples as a two layered resampling is necessary.

### Value

A dataframe with the requested intervals.

### Author(s)

Thomas Jaki

### References

Jaki T., West R. W. (2008) Maximum kernel likelihood estimation. *Journal of Computational and Graphical Statistics* Vol. 17(No 4), 976-993.

Davison, A. C. and Hinkley, D. V. (1997), *Bootstrap Methods and their Applications*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.

### See Also

[mkle](#)

### Examples

```
data(state)
mkle.ci(state$CRIME,method=c('wald','percentile'),B=100,gridsize=2^11)
```

---

opt.bw

*Optimal bandwidth for the maximum kernel likelihood estimator*

---

### Description

Estimates the optimal bandwidth for the maximum kernel likelihood estimator using a Gaussian kernel for a given dataset using the bootstrap.

### Usage

```
opt.bw(data, bws=c(sd(data),4*sd(data)), B=1000, gridsize=2^14)
```

**Arguments**

|          |   |
|----------|---|
| data     | the data for which the optimal bandwidth should be found.   |
| bws      | a vector with the upper and lower bound for the bandwidth.  |
| B        | number of resamples used to estimate the mean squared error with 1000 as the default.                       |
| gridsize | the number of points at which the kernel density estimator is to be evaluated with $2^{14}$ as the default. |

**Details**

The bandwidth considered fall between one and 4 standard deviations. In addition the mse of the [mkle](#) for a bandwidth of zero will also be included.

The estimation of the optimal bandwidth might take several minutes depending on the number of bootstrap resamples and the gridsize used.

**Value**

The estimated optimal bandwidth.

**Note**

The [optimize](#) is used for the optimization.

**Author(s)**

Thomas Jaki

**References**

Jaki T., West R. W. (2008) Maximum kernel likelihood estimation. Submitted to *Journal of Computational and Graphical Statistics* Vol. 17(No 4), 976-993.

Davison, A. C. and Hinkley, D. V. (1997), *Bootstrap Methods and their Applications*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.

**See Also**

[mkle](#)

**Examples**

```
data(state)
opt.bw(state$CRIME, B=10)
```

---

state

*Violent death in the USA*

---

**Description**

The dataset gives the number of violent death per 100,000 population per state

**Usage**

```
data(state)
```

**Format**

A data frame with 50 observations on the following 2 variables.

STATE a factor with levels AK AL AR AZ CA CO CT DE FL GA HI IA ID IL IN KS KY LA MA MD ME MI MN  
MO MS MT NC ND NE NH NJ NM NV NY OH OK OR PA RI SC SD TN TX UT VA VT WA WI WV WY

CRIME a numeric vector

**Source**

Shapiro, Robert~J. 1998. Statistical Abstract of the United States. 118 edn. U.S. Bureau of the Census.

**Examples**

```
data(state)
hist(state$CRIME)
mkle(state$CRIME)
```

# Index

- \* **datasets**
  - state, 8
- \* **distribution**
  - klik, 3
- \* **htest**
  - mkle.ci, 5
- \* **nonparametric**
  - klik, 3
  - mkle, 4
  - mkle.ci, 5
  - opt.bw, 6
- \* **package**
  - MKLE-package, 1
- \* **univar**
  - mkle, 4

density, 5

klik, 3, 5

MKLE (MKLE-package), 1

mkle, 3, 4, 6, 7

MKLE-package, 1

mkle.ci, 5

opt.bw, 6

optimize, 5, 7

state, 8