

Package ‘MixtureInf’

April 7, 2016

Type Package

Version 1.1

Date 2016-03-21

Title Inference for Finite Mixture Models

Author Shaoting Li, Jiahua Chen and Pengfei Li.

Maintainer Davor Cubranic <cubranic@stat.ubc.ca>

Description

Functions for computing the penalized maximum likelihood estimate (PMLE) or maximum likelihood estimate (MLE), testing the order of a finite mixture model using EM-test, drawing histogram of observations and the fitted density or probability mass function of the mixture model.

Imports quadprog

License AGPL (>= 3)

URL <http://cran.r-project.org/package=MixtureInf>

NeedsCompilation no

Repository CRAN

Date/Publication 2016-04-07 01:10:26

R topics documented:

agesch	2
earthquake	3
emtest.binom	4
emtest.exp	5
emtest.norm	7
emtest.norm0	8
emtest.pois	10
grains	12
lengthofstay	12
pearson	13
plotmix.binom	14
plotmix.exp	15

plotmix.norm	16
plotmix.norm0	17
plotmix.pois	18
pmle.binom	20
pmle.exp	21
pmle.norm	22
pmle.norm0	23
pmle.pois	25
residual1	26
residual2	27
rmix.binom	27
rmix.exp	28
rmix.norm	29
rmix.pois	30
SLC	31
timesoffailure	31
zscores	32

Index	34
--------------	-----------

agesch	<i>age of onset of schizophrenia</i>
--------	--------------------------------------

Description

This data set is from a schizophrenia study by Levine (1981), who collected the results of seven studies on the age of onset of schizophrenia of 99 females and 152 males. This data set contains observed values of only male data. Everitt et al., (2001) fitted the male data using a two-component normal mixture model. The data frame has 152 rows and 1 column.

Usage

```
#Load this data set for subsequent analysis:
data(agesch)
```

Format

This data frame contains one column:
age: age of onset of schizophrenia of 152 males.

References

- Levine, R. (1981). Sex differences in schizophrenia: Timing or subtypes? *Psychological Bulletin*, 90, 432–444.
- Everitt, B. S., Landau, S., and Leese, M. (2001). *Cluster Analysis* (4th ed.). New York, NY: Oxford University Press, Inc.

Examples

```
#fit a finite normal mixture model of order 2 to this data set,  
#plot the histogram of the observations and the fitted density function.  
data(agesch)  
out <- pmle.norm(agesch,2,1)  
plotmix.norm(agesch,out)
```

earthquake	<i>number of earthquakes</i>
------------	------------------------------

Description

This data set contains the number of major earthquakes (magnitude 7 or greater) in the world from 1900 to 2006. The data are available in Table 1.1 of Zucchini & MacDonald (2009). Zucchini & MacDonald (2009) suggested using a Poisson mixture model to fit the data. The data frame has 107 rows and 1 column.

Usage

```
#Load this data set for subsequent analysis:  
data(earthquake)
```

Format

This data frame contains one column:
number: number of major earthquakes in a year.

References

Zucchini W, MacDonald IL (2009). Hidden Markov models for time series: an introduction using R. CRC Press.

Examples

```
#fit a finite Poisson mixture model of order 2 to this data set,  
#plot the histogram of the observations and the fitted probability mass function.  
data(earthquake)  
out <- pmle.pois(earthquake,2,1)  
plotmix.pois(earthquake,out)
```

emtest.binom	<i>test the order of a mixture of binomials</i>
--------------	---

Description

EM-test for the order of a finite mixture of binomials.

Usage

```
emtest.binom(x, size, m0 = 1, C = NULL,
             inival = NULL, len = 10, niter = 50, tol = 1e-06, k = 3, rformat = FALSE)
```

Arguments

x	data, can be either a vector or a matrix with the 1st column being the observed values and the 2nd column being the corresponding frequencies.
size	number of trials.
m0	order of the finite mixture model under the null hypothesis, default value: m0 = 1.
C	optional tuning parameter for EM-test procedure, default value: C = NULL. (if not provided, it will be determined by the formulas described in Chen and Li, 2011).
inival	initial values for the EM-algorithm to compute the MLE under the null model, a 2m0-dimension vector including m0 mixing proportions and m0 component parameters, or a matrix with 2m0 columns, default value: inival = NULL. (if not provided, random initial values are used.)
len	number of random initial values for the EM-algorithm, default value: len = 10.
niter	number of iterations for all initial values in the EM-algorithm. The algorithm runs EM-iteration niter times from each initial value. The iteration will restart from the parameter value with the highest likelihood value at the point and run until convergence. default value: niter = 50.
tol	tolerance level for the convergence of the EM-algorithm, default value: tol = 1e-6.
k	number of EM iterations, default value: k = 3.
rformat	form of the digital output: default of R package is used when rformat = T; If rformat = T, the digital output is rounded to the 3rd decimal place if it is larger than 0.001, keeps 3 significant digits otherwise. The default value of rformat is F.

Value

Return an object of class EM-test with the following elements:

MLE of the parameters under the null hypothesis (order = m0)

Parameter estimates under the specific alternative whose order is 2m0

EM-test statistic
P-value
Level of penalty

Author(s)

Shaoting Li, Jiahua Chen and Pengfei Li

References

Chen, J. and Li, P. (2011). Tuning the EM-test for the order of finite mixture models. *The Canadian Journal of Statistics*. 39, 389–404.

Li, P. and Chen, J. (2010). Testing the order of a finite mixture model. *JASA*. 105, 1084–1092.

Li, P., Chen, J. and Marriott, P. (2009). Non-finite Fisher information and homogeneity: The EM approach. *Biometrika*. 96, 411–426.

See Also

plotmix.binom, pmle.binom, rmix.binom

Examples

```
#generate a random sample from a 2 component binomial mixture,  
#test for order = 2 by the EM-test,  
#plot the histogram of the observations and the fitted probability mass function  
#under the null model.  
x <- rmix.binom(200,c(0.3,0.7),c(0.4,0.6),50)  
out <- emtest.binom(x,50)  
plotmix.binom(x,50,out)
```

emtest.exp

test the order of a mixture of exponentials

Description

EM-test for the order of a finite mixture of exponentials.

Usage

```
emtest.exp(x, m0 = 1, C = NULL,  
          inival = NULL, len = 10, niter = 50, tol = 1e-6, k = 3, rformat = FALSE)
```

Arguments

x	data, can be either a vector or a matrix with the 1st column being the observed values and the 2nd column being the corresponding frequencies.
m0	order of the finite mixture model under the null hypothesis, default value: m0 = 1.
C	optional tuning parameter for EM-test procedure, default value: C = NULL. (if not provided, it will be determined by the formulas described in Chen and Li, 2011).
inival	initial values for the EM-algorithm to compute the MLE under the null model, a 2m0-dimension vector including m0 mixing proportions and m0 component parameters, or a matrix with 2m0 columns, default value: inival = NULL. (if not provided, random initial values are used.)
len	number of random initial values for the EM-algorithm, default value: len = 10.
niter	number of iterations for all initial values in the EM-algorithm. The algorithm runs EM-iteration niter times from each initial value. The iteration will restart from the parameter value with the highest likelihood value at the point and run until convergence. default value: niter = 50.
tol	tolerance level for the convergence of the EM-algorithm, default value: tol = 1e-6.
k	number of EM iterations: default value: k = 3.
rformat	form of the digital output: default of R package is used when rformat = T; If rformat = T, the digital output is rounded to the 3rd decimal place if it is larger than 0.001, keeps 3 significant digits otherwise. The default value of rformat is F.

Value

Return an object of class EM-test with the following elements:
 MLE of the parameters under the null hypothesis (order = m0)
 Parameter estimates under the specific alternative whose order is 2m0
 EM-test statistic
 P-value
 Level of penalty

Author(s)

Shaoting Li, Jiahua Chen and Pengfei Li

References

- Chen, J. and Li, P. (2011). Tuning the EM-test for the order of finite mixture models. *The Canadian Journal of Statistics*. 39, 389–404.
- Li, P. and Chen, J. (2010). Testing the order of a finite mixture model. *JASA*. 105, 1084–1092.
- Li, P., Chen, J. and Marriott, P. (2009). Non-finite Fisher information and homogeneity: The EM approach. *Biometrika*. 96, 411–426.

See Also

plotmix.exp, pmle.exp, rmix.exp

Examples

```
#generate a random sample from a 2 component exponential mixture,
#test for order = 2 by the EM-test,
#plot the histogram of the observations and the fitted density under the null model.
x <- rmix.exp(100,c(0.3,0.7),c(2,10))
out <- emtest.exp(x,2)
plotmix.exp(x,out)
```

emtest.norm

test the order of a mixture of normals with unequal variance

Description

EM-test for the order of a finite mixture of normals with unequal variance.

Usage

```
emtest.norm(x, m0 = 1, pens = NULL,
            inival = NULL, len = 10, niter = 50, tol = 1e-6, k = 3, rformat = FALSE)
```

Arguments

x	data, can be either a vector or a matrix with the 1st column being the the observed values and the 2nd column being the corresponding frequencies.
m0	order of the finite mixture model under the null hypothesis. default value: m0 = 1.
pens	a 2-dimensions vector with the level of penalty functions for mixing proportion and variance, default value: pens = NULL. (if not provided, it will be determined by the formulas described in Chen, Li and Fu, 2012)
inival	initial values for the EM-algorithm to compute the MLE under the null model, a 3m0-dimensions vector including m0 mixing proportions, m0 component means and m0 component variances, or a matrix with 3m0 columns, default value: inival = NULL. (if not provided, random initial values are used.)
len	number of random initial values for the EM-algorithm, default value: len = 10.
niter	number of iterations for all initial values in the EM-algorithm. The algorithm runs EM-iteration niter times from each initial value. The iteration will restart from the parameter value with the highest likelihood value at the point and run until convergence. default value: niter = 50.
tol	tolerance level for the convergence of the EM-algorithm, default value: tol = 1e-6.
k	number of EM iterations: default value: k = 3.

`rformat` form of the digital output: default of R package is used when `rformat = T`; If `rformat = T`, the digital output is rounded to the 3rd decimal place if it is larger than 0.001, keeps 3 significant digits otherwise. The default value of `rformat` is `F`.

Value

Return an object of class EM-test with the following elements:

MLE of the parameters under the null hypothesis (`order = m0`)

Parameter estimates under the specific alternative whose order is `2m0`

EM-test statistic

P-value

Level of penalty

Author(s)

Shaoting Li, Jiahua Chen and Pengfei Li

References

Chen, J. and Li, P. (2009). Hypothesis test for normal mixture models: The EM approach. *The Annals of Statistics*. 37, 2523–2542.

Chen, J., Li, P. and Fu, Y. (2012). Inference on the order of a normal mixture. *JASA*. 107, 1096–1105.

See Also

`plotmix.norm`, `pmle.norm`, `rmix.norm`

Examples

```
#load the grains data set,
#conduct homogeneity test by the EM-test.
data(grains)
emtest.norm(grains)
```

`emtest.norm0` *test the order of a mixture of normals with equal and known variance*

Description

EM-test for the order of a finite mixture of normals with equal and known variance.

Usage

```
emtest.norm0(x, var, m0 = 1, C = NULL,
             inival = NULL, len = 10, niter = 50, tol = 1e-6, k = 3, rformat = FALSE)
```


Arguments

x	data, can be either a vector or a matrix with the 1st column being the observed values and the 2nd column being the corresponding frequencies.
var	known component variance.
m0	order of the finite mixture model under the null hypothesis, default value: m0 = 1.
C	optional tuning parameter for EM-test procedure, default value: C = NULL. (if not provided, it will be determined by the formulas described in Chen and Li, 2011).
inival	initial values for the EM-algorithm to compute the MLE under the null model, a 2m0-dimension vector including m0 mixing proportions and m0 component parameters, or a matrix with 2m0 columns, default value: inival = NULL. (if not provided, random initial values are used.)
len	number of random initial values for the EM-algorithm, default value: len = 10.
niter	number of iterations for all initial values in the EM-algorithm. The algorithm runs EM-iteration niter times from each initial value. The iteration will restart from the parameter value with the highest likelihood value at the point and run until convergence. default value: niter = 50.
tol	tolerance level for the convergence of the EM-algorithm, default value: tol = 1e-6.
k	number of EM iterations: default value: k = 3.
rformat	form of the digital output: default of R package is used when rformat = T; If rformat = T, the digital output is rounded to the 3rd decimal place if it is larger than 0.001, keeps 3 significant digits otherwise. The default value of rformat is F.

Value

Return an object of class EM-test with the following elements:

MLE of the parameters under the null hypothesis (order = m0)

Parameter estimates under the specific alternative whose order is 2m0

EM-test statistic

P-value

Level of penalty

Author(s)

Shaoting Li, Jiahua Chen and Pengfei Li

References

Chen, J. and Li, P. (2011). Tuning the EM-test for the order of finite mixture models. *The Canadian Journal of Statistics*. 39, 389–404.

Li, P. and Chen, J. (2010). Testing the order of a finite mixture model. *JASA*. 105, 1084–1092.

Li, P., Chen, J. and Marriott, P. (2009). Non-finite Fisher information and homogeneity: The EM approach. *Biometrika*. 96, 411–426.

See Also

plotmix.norm0, pmle.norm0, rmix.norm

Examples

```
#generate a random sample from a 2 component normal mixture with equal variance,
#conduct homogeneity test by the EM-test.
x <- rmix.norm(200,c(0.3,0.7),c(-1,2),c(2,2))
emtest.norm0(x,var=4)
```

<code>emtest.pois</code>	<i>test the order of a mixture of Poissons</i>
--------------------------	--

Description

EM-test for the order of a finite mixture of Poissons.

Usage

```
emtest.pois(x, m0 = 1, C = NULL,
            inival = NULL, len = 10, niter = 50, tol = 1e-6, k = 3, rformat = FALSE)
```

Arguments

<code>x</code>	data, can be either a vector or a matrix with the 1st column being the observed values and the 2nd column being the corresponding frequencies.
<code>m0</code>	order of the finite mixture model under the null hypothesis, default value: <code>m0 = 1</code> .
<code>C</code>	optional tuning parameter for EM-test procedure, default value: <code>C = NULL</code> . (if not provided, it will be determined by the formulas described in Chen and Li, 2011).
<code>inival</code>	initial values for the EM-algorithm to compute the MLE under the null model, a $2m_0$ -dimension vector including <code>m0</code> mixing proportions and <code>m0</code> component parameters, or a matrix with $2m_0$ columns, default value: <code>inival = NULL</code> . (if not provided, random initial values are used.)
<code>len</code>	number of random initial values for the EM-algorithm, default value: <code>len = 10</code> .
<code>niter</code>	number of iterations for all initial values in the EM-algorithm. The algorithm runs EM-iteration <code>niter</code> times from each initial value. The iteration will restart from the parameter value with the highest likelihood value at the point and run until convergence. default value: <code>niter = 50</code> .
<code>tol</code>	tolerance level for the convergence of the EM-algorithm, default value: <code>tol = 1e-6</code> .

k	number of EM iterations: default value: $k = 3$.
rformat	form of the digital output: default of R package is used when $rformat = T$; If $rformat = T$, the digital output is rounded to the 3rd decimal place if it is larger than 0.001, keeps 3 significant digits otherwise. The default value of rformat is F.

Value

Return an object of class EM-test with the following elements:

MLE of the parameters under the null hypothesis (order = m_0)

Parameter estimates under the specific alternative whose order is $2m_0$.

EM-test statistic

P-value

Level of penalty

Author(s)

Shaoting Li, Jiahua Chen and Pengfei Li

References

Chen, J. and Li, P. (2011). Tuning the EM-test for the order of finite mixture models. *The Canadian Journal of Statistics*. 39, 389–404.

Li, P. and Chen, J. (2010). Testing the order of a finite mixture model. *JASA*. 105, 1084–1092.

Li, P., Chen, J. and Marriott, P. (2009). Non-finite Fisher information and homogeneity: The EM approach. *Biometrika*. 96, 411–426.

See Also

plotmix.pois, pmle.pois, rmix.pois

Examples

```
#load the earthquake data set,  
#conduct homogeneity test by the EM-test.  
data(earthquake)  
emtest.pois(earthquake,2)
```

grains

grains data

Description

The data set contains the square root of the total number of grains for each planty from Loisel et al., (1994). Loisel et al., (1994) suggested using a finite normal mixture model. The grains data frame has 150 rows and 1 column.

Usage

```
#Load this data set for subsequent analysis:  
data(grains)
```

Format

This data frame contains one column:

x: square root of the total number of grains for each planty.

References

Loisel, P., Goffinet, B., Monod, H., and Montes De Oca, G. (1994). Detecting a major gene in an F2 population. *Biometrics*, 50, 512–516.

Examples

```
#fit a finite noraml mixture model of order 2 to this data set.  
data(grains)  
pmle.norm(grains,2,1)
```

lengthofstay*length of stay*

Description

This data set contains the length of stay (in days) for 469 geriatric patients in a psychiatric hospital in northeast London in 1991. Harrison and Millard (1991) used a mixture of two exponential distributions in their analysis. The data frame has 469 rows and 1 column.

Usage

```
#Load this data set for subsequent analysis:  
data(lengthofstay)
```

Format

This data frame contains one column,
length: length of stay for 469 geriatric patients in a psychiatric hospital.

References

Harrison G, Millard P (1991). Balancing acute and long-term care: the mathematics of throughput in departments of geriatric medicine. *Methods of information in medicine*, 30(3), 221–228.

Examples

```
#fit a 2 component exponential model to this data set,  
#plot the histogram of the observations and the fitted density.  
data(lengthofstay)  
out <- pmle.exp(lengthofstay,2,1)  
plotmix.exp(lengthofstay,out)
```

pearson

Pearson's crab data

Description

This data set contains the ratio of "forehead" breadth to body length for 1000 crabs sampled at Naples by Professor W.F.R. Weldon. Pearson (1894) used a two component normal mixture model to fit this data set. The data frame has 29 rows and 2 columns.

Usage

```
#Load this data set for subsequent analysis:  
data(pearson)
```

Format

This data frame contains the following columns:
ratio: the boundaries of grouping intervals.
freq: the frequencies of observation falling into each interval.

References

Pearson K (1894). Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London. A*, 185, 71–110.

Examples

```
#fit a 2 component normal mixture model to this data set.  
data(pearson)  
pmle.norm(pearson,2,1)
```

plotmix.binom	<i>Histogram of the observations and the fitted probability mass function of a mixture of binomials</i>
---------------	---

Description

Plot the histogram of the observations and the fitted probability mass function of a mixture of binomials.

Usage

```
plotmix.binom(x, size, theta, hist = 1, comp = TRUE, h = 1, main = "",
             xlab = "Observations", ylab = "")
```

Arguments

x	data, can be either a vector or a matrix with the 1st column being the observed values and the 2nd column being the corresponding frequencies.
size	number of trials.
theta	parameter values, output of emtest.binom or pmle.binom, or a vector of parameter values, include mixing proportions and component parameters.
hist	style of histogram, hist = 0,1,2 are permitted. hist = 0 means no histogram, hist = 1,2 give two styles, default value: hist = 1.
comp	a parameter specifies the form of the component probability functions in the graph. The component probability functions will be plotted only if the order is 3 or lower. comp = T switches on the component probability functions in the graph. If comp=F, no component probability functions will be included. default value: comp = T.
h	a number between 0 and 1, specifying the height of the probability function, default value: h = 1 (complete probability function).
main	title of graph, default value: main = "".
xlab	label of x-axis, default value: xlab = "Observations".
ylab	label of y-axis, default value: ylab = "".

Value

Return the histogram of the observations and the plot of the fitted probability mass function.

Author(s)

Shaoting Li, Jiahua Chen and Pengfei Li

See Also

emtest.binom, pmle.binom, rmix.binom

Examples

```
#generate a random sample from a 2 component binomial mixture,
#compute the PMLE of parameters based on the 2 component binomial mixture model,
#plot the histogram of the observations and the fitted probability mass function.
x <- rmix.binom(200,c(0.3,0.7),c(0.4,0.6),50)
out <- pmle.binom(x,50,2)
plotmix.binom(x,50,out)
```

plotmix.exp	<i>Histogram of the observations and the fitted density of a mixture of exponentials</i>
-------------	--

Description

Plot the histogram of the observations and the fitted density of a mixture of exponentials.

Usage

```
plotmix.exp(x, theta, hist = 1, comp = TRUE, k = 20, h = 1, main = "",
           xlab = "Observations", ylab = "")
```

Arguments

x	data, can be either a vector or a matrix with the 1st column being the observed values and the 2nd column being the corresponding frequencies.
theta	parameter values, output of emtest.exp or pmle.exp, or a vector of parameter values, include mixing proportions and component parameters.
hist	style of histogram. hist = 0,1,2 are permitted. hist = 0 means no histogram, hist = 1,2 give two styles, default value: hist = 1.
comp	a parameter specifies the form of the component densities in the graph. comp = T switches on the component densities in the graph. If comp = F, no component densities will be included. default value: comp = T.
k	number of bins for the histogram, default value: k = 20.
h	a number between 0 and 1, specifying the height of the density function, default value: h = 1 (complete density function).
main	title of graph, default value: main = "".
xlab	label of x-axis, default value: xlab = "Observations".
ylab	label of y-axis, default value: ylab = "".

Value

Return the histogram of observations and the plot of the fitted density.

Author(s)

Shaoting Li, Jiahua Chen and Pengfei Li

See Also

emtest.exp, pmle.exp, rmix.exp

Examples

```
#load the lengthofstay data set,
#fit a 2 compoent exponential mixture model,
#plot the histogram of the observations and the fitted density.
data(lengthofstay)
out <- pmle.exp(lengthofstay,2)
plotmix.exp(lengthofstay,out)
```

plotmix.norm	<i>Histogram of the observations and the fitted density of a mixture of normals with unequal variance</i>
--------------	---

Description

Plot the histogram of the observations and the fitted density of a mixture of normals with unequal variance.

Usage

```
plotmix.norm(x, theta, hist = 1, comp = TRUE, k = 20, h = 1, main = "",
             xlab = "Observations", ylab = "")
```

Arguments

x	data, can be either a vector or a matrix with the 1st column being the observed values and the 2nd column being the corresponding frequencies.
theta	parameter values, output of emtest.exp or pmle.exp, or a vector of parameter values, include mixing proportions, mixing means and mixing variances.
hist	style of histogram. hist = 0,1,2 are permitted. hist = 0 means no histogram, hist = 1,2 give two styles, default value: hist = 1.
comp	a parameter specifies the form of the component densities in the graph. comp = T switches on the component densities in the graph. If comp = F, no component densities will be included. default value: comp = T.
k	number of bins for the histogram, default value: k = 20.
h	a number between 0 and 1, specifying the height of the density function, default value: h = 1 (complete density function).
main	title of graph, default value: main = "".
xlab	label of x-axis, default value: xlab = "Observations".
ylab	label of y-axis, default value: ylab = "".

Value

Return the histogram of the observations and the plot of the fitted density.

Author(s)

Shaoting Li, Jiahua Chen and Pengfei Li

See Also

emtest.norm, pmle.norm, rmix.norm

Examples

```
#generate a random sample from a 2 component normal mixture,
#plot the histogram of the observations and the density function with a set of
#given parameters.
x <- rmix.norm(200,c(0.3,0.7),c(2,5),c(1,2))
theta <- c(0.25,0.75,1.8,5.2,1.1,2.1)
plotmix.norm(x,theta)
```

plotmix.norm0	<i>Histogram of the observations and the fitted density of a mixture of normals with equal and known variance</i>
---------------	---

Description

Plot the histogram of the observations and the fitted density of a mixture of normals with equal and known variance.

Usage

```
plotmix.norm0(x, var, theta, hist = 1, comp = TRUE, k = 20, h = 1, main = "",
             xlab = "Observations", ylab = "")
```

Arguments

x	data, can be either a vector or a matrix with the 1st column being the observed values and the 2nd column being the corresponding frequencies.
var	known component variance.
theta	parameter values, output of emtest.exp or pmle.exp, or a vector of parameter values, include mixing proportions and component parameters.
hist	style of histogram. hist = 0,1,2 are permitted. hist = 0 means no histogram, hist = 1,2 give two styles, default value: hist = 1.
comp	a parameter specifies the form of the component densities in the graph. comp = T switches on the component densities in the graph. If comp = F, no component densities will be included. default value: comp = T.

k	number of bins for the histogram, default value: k = 20.
h	a number between 0 and 1, specifying the height of the density function, default value: h = 1 (complete density function).
main	title of graph, default value: main = "".
xlab	label of x-axis, default value: xlab = "Observations".
ylab	label of y-axis, default value: ylab = "".

Value

Return the histogram of the observations and the plot of the fitted density.

Author(s)

Shaoting Li, Jiahua Chen and Pengfei Li

See Also

emtest.norm0, pmle.norm0, rmix.norm

Examples

```
#generate a random sample from a 2 component normal mixture,
#compute the PMLE of parameters under a 2 component normal mixture model,
#plot the histogram of the observations and the fitted density function.
x <- rmix.norm(200,c(0.3,0.7),c(-2,3))
out <- pmle.norm0(x,var=1,2)
plotmix.norm0(x,var=1,out)
```

plotmix.pois	<i>Histogram of the observations and the fitted probability mass function of a mixture of Poissons</i>
--------------	--

Description

Plot the histogram of the observations and the fitted probability mass function of a mixture of Poissons.

Usage

```
plotmix.pois(x, theta, hist = 1, comp = TRUE, h = 1, main = "",
            xlab = "Observations", ylab = "")
```

Arguments

x	data, can be either a vector or a matrix with the 1st column being the observed values and the 2nd column being the corresponding frequencies.
theta	parameter values, output of emtest.pois or pmle.pois, or a vector of parameter values, include mixing proportions and component parameters.
hist	style of histogram. hist = 0,1,2 are permitted. hist = 0 means no histogram, hist = 1,2 give two styles, default value: hist = 1.
comp	a parameter specifies the form of the component probability functions in the graph. The component probability functions will be plotted only if the order is 3 or lower. comp = T switches on the component probability functions in the graph. If comp = F, no component probability functions will be included. default value: comp = T.
h	a number between 0 and 1, specifying the height of the probability function, default value: h=1 (complete probability function).
main	title of graph, default value: main = "".
xlab	label of x-axis, default value: xlab = "Observations".
ylab	label of y-axis, default value: ylab = "".

Value

Return the histogram of the observations and the plot of the fitted probability mass function.

Author(s)

Shaoting Li, Jiahua Chen and Pengfei Li

See Also

emtest.pois, pmle.pois, rmix.pois

Examples

```
#generate a random sample from a 2 component Poisson mixture,  
#plot the histogram of the observations and the probability mass function with a set of  
#given parameters.  
x <- rmix.pois(200,c(0.3,0.7),c(2,5))  
theta <- c(0.25,0.75,1.8,5.2)  
plotmix.pois(x,theta)
```

pmlе.binom	<i>compute the PMLE or MLE of the parameters under a mixture of binomials</i>
------------	---

Description

Compute the PMLE or MLE of the parameters under a mixture of binomials. When the level of penalty is 0, PMLE reduces to MLE.

Usage

```
pmlе.binom(x, size, m0 = 1, lambda = 0,
           inival=NULL, len = 10, niter = 50, tol = 1e-06, rformat = FALSE)
```

Arguments

x	data, can be either a vector or a matrix with the 1st column being the observed values and the 2nd column being the corresponding frequencies.
size	number of trials.
m0	order of the finite mixture model, default value: m0 = 1.
lambda	level of penalty, default value: lambda = 0.
inival	initial values for the EM-algorithm, a 2m0-dimension vector including m0 mixing proportions and m0 component parameters, or a matrix with 2m0 columns, default value: inival = NULL. (if not provided, random initial values are used.)
len	number of random initial values for the EM-algorithm, default value: len = 10.
niter	number of iterations for all initial values in the EM-algorithm. The algorithm runs EM-iteration niter times from each initial value. The iteration will restart from the parameter value with the highest likelihood value at the point and run until convergence. default value: niter = 50.
tol	tolerance level for the convergence of the EM-algorithm, default value: tol = 1e-6.
rformat	form of the digital output: default of R package is used when rformat = T; If rformat = T, the digital output is rounded to the 3rd decimal place if it is larger than 0.001, keeps 3 significant digits otherwise. The default value of rformat is F.

Value

Return the PMLE or MLE of the parameters with order = m0 (mixing proportions and component parameters), log-likelihood value at the PMLE or MLE and the penalized log-likelihood value at the PMLE.

Author(s)

Shaoting Li, Jiahua Chen and Pengfei Li

See Also

emtest.binom, plotmix.binom, rmix.binom

Examples

```
#load the residual2 data set,
#fit a 2 component binomial mixture model.
data(residual2)
pmle.binom(residual2,12,2,1)
```

pmle.exp	<i>compute the PMLE or MLE of the parameters under a mixture of exponentials</i>
----------	--

Description

Compute the PMLE or MLE of the parameters under a mixture of exponentials. When the level of penalty is 0, PMLE reduces to MLE.

Usage

```
pmle.exp(x, m0 = 1, lambda = 0,
         inival = NULL, len = 10, niter = 50, tol = 1e-06, rformat = FALSE)
```

Arguments

x	data, can be either a vector or a matrix with the 1st column being the observed values and the 2nd column being the corresponding frequencies.
m0	order of the finite mixture model, default value: m0 = 1.
lambda	level of penalty, default value: lambda = 0.
inival	initial values for the EM-algorithm, a 2m0-dimension vector including m0 mixing proportions and m0 component parameters, or a matrix with 2m0 columns, default value: inival = NULL. (if not provided, random initial values are used.)
len	number of random initial values for the EM-algorithm, default value: len = 10.
niter	number of iterations for all initial values in the EM-algorithm. The algorithm runs EM-iteration niter times from each initial value. The iteration will restart from the parameter value with the highest likelihood value at the point and run until convergence. default value: niter = 50.
tol	tolerance level for the convergence of the EM-algorithm, default value: tol = 1e-6.
rformat	form of the digital output: default of R package is used when rformat = T; If rformat = T, the digital output is rounded to the 3rd decimal place if it is larger than 0.001, keeps 3 significant digits otherwise. The default value of rformat is F.

Value

Return the PMLE or MLE of the parameters with order = m0 (mixing proportions and component parameters), log-likelihood value at the PMLE or MLE and the penalized log-likelihood value at the PMLE.

Author(s)

Shaoting Li, Jiahua Chen and Pengfei Li

See Also

emtest.exp, plotmix.exp, rmix.exp

Examples

```
#generate a random sample from a 2 component exponential mixture,
#compute the PMLE of the parameters under 2 component exponential mixture model,
#plot the histogram of the observations and the fitted density.
x <- rmix.exp(200,c(0.3,0.7),c(2,8))
out <- pmle.exp(x,2,1)
plotmix.exp(x,out)
```

pmle.norm	<i>compute the PMLE or MLE of the parameters under a mixture of normals with unequal variance</i>
-----------	---

Description

Compute the PMLE or MLE of the parameters under a mixture of normals with unequal variance.

Usage

```
pmle.norm(x, m0 = 1, lambda = 0,
  inival=NULL, len = 10, niter = 50, tol = 1e-06, rformat = FALSE)
```

Arguments

x	data, can be either a vector or a matrix with the 1st column being the observed data values and the 2nd column being the corresponding frequencies.
m0	order of the finite mixture model, default value: m0 = 1.
lambda	level of penalty for mixing proportions, default value: lambda = 0.
inival	initial values for the EM-algorithm, a 3m0-dimension vector including m0 mixing proportions, m0 component means and m0 component variances, or a matrix with 3m0 columns, default value: inival = NULL. (if not provided, random initial values are used.)
len	number of random initial values for the EM-algorithm, default value: len = 10.

niter	number of iterations for all initial values in the EM-algorithm. The algorithm runs EM-iteration niter times from each initial value. The iteration will restart from the parameter value with the highest likelihood value at the point and run until convergence. default value: niter = 50.
tol	tolerance level for the convergence of the EM-algorithm, default value: tol = 1e-6.
rformat	form of the digital output: default of R package is used when rformat = T; If rformat = T, the digital output is rounded to the 3rd dicimal place if it is larger than 0.001, keeps 3 significant digits otherwise. The default value of rformat is F.

Value

Return the PMLE or MLE of the parameters with order = m0 (mixing proportions, mixing means and mixing variances), log-likelihood value at the PMLE or MLE and the penalized log-likelihood value at the PMLE.

Author(s)

Shaoting Li, Jiahua Chen and Pengfei Li

See Also

emtest.norm, plotmix.norm, rmix.norm

Examples

```
#load the pearson's crab data,
#fit the 2 and 3 component normal mixture models,
#plot the histgorams of the observations and the fitted densities.
data(pearson)
out1 <- pmle.norm(pearson,2,1)
plotmix.norm(pearson,out1)

## Not run:
out2 <- pmle.norm(pearson,3,1)
plotmix.norm(pearson,out2)
par(mfrow=c(1,1))

## End(Not run)
```

pmle.norm0	<i>compute the PMLE or MLE of the parameters under a mixture of normals with equal and known variance</i>
------------	---

Description

Compute the PMLE or MLE of the parameters under a mixture of normals with equal and known variance. When the level of penalty is 0, PMLE reduces to MLE.

Usage

```
pmle.norm0(x, var, m0 = 1, lambda = 0,
           inival = NULL, len = 10, niter = 50, tol = 1e-06, rformat = FALSE)
```

Arguments

x	data, can be either a vector or a matrix with the 1st column being the observed values and the 2nd column being the corresponding frequencies.
var	known component variance.
m0	order of the finite mixture model, default value: m0 = 1.
lambda	level of penalty, default value: lambda = 0.
inival	initial values for the EM-algorithm, a 2m0-dimension vector including m0 mixing proportions and m0 component parameters, or a matrix with 2m0 columns, default value: inival = NULL. (if not provided, random initial values are used.)
len	number of random initial values for the EM-algorithm, default value: len = 10.
niter	number of iterations for all initial values in the EM-algorithm. The algorithm runs EM-iteration niter times from each initial value. The iteration will restart from the parameter value with the highest likelihood value at the point and run until convergence. default value: niter = 50.
tol	tolerance level for the convergence of the EM-algorithm, default value: tol = 1e-6.
rformat	form of the digital output: default of R package is used when rformat = T; If rformat = T, the digital output is rounded to the 3rd decimal place if it is larger than 0.001, keeps 3 significant digits otherwise. The default value of rformat is F.

Value

Return the PMLE or MLE of the parameters with order = m0 (mixing proportions and component parameters), log-likelihood value at the PMLE or MLE and the penalized log-likelihood value at the PMLE.

Author(s)

Shaoting Li, Jiahua Chen and Pengfei Li

See Also

emtest.norm0, plotmix.norm0, rmix.norm

Examples

```
#generate a random sample from a 2 component normal mixture,
#compute the PMLE of the parameters under the 2 component normal mixture model with
#known variance 1.
x <- rmix.norm(200,c(0.3,0.7),c(-1,2))
pmle.norm0(x,var=1,2)
```

pmle.pois	<i>compute the PMLE or MLE of the parameters under a mixture of Possions</i>
-----------	--

Description

Compute the PMLE or MLE of the parameters under a mixture of Possions. When the level of penalty is 0, PMLE reduces to MLE.

Usage

```
pmle.pois(x, m0 = 1, lambda = 0,
          inival = NULL, len = 10, niter = 50, tol = 1e-06, rformat = FALSE)
```

Arguments

x	data, can be either a vector or a matrix with the 1st column being the observed values and the 2nd column being the corresponding frequencies.
m0	order of the finite mixture model, default value: m0 = 1.
lambda	level of penalty, default value: lambda = 0.
inival	initial values for the EM-algorithm, a 2m0-dimension vector including m0 mixing proportions and m0 component parameters, or a matrix with 2m0 columns, default value: inival = NULL. (if not provided, random initial values are used.)
len	number of random initial values for the EM-algorithm, default value: len = 10.
niter	number of iterations for all initial values in the EM-algorithm. The algorithm runs EM-iteration niter times from each initial value. The iteration will restart from the parameter value with the highest likelihood value at the point and run until convergence. default value: niter = 50.
tol	tolerance level for the convergence of the EM-algorithm, default value: tol = 1e-6.
rformat	form of the digital output: default of R package is used when rformat = T; If rformat = T, the digital output is rounded to the 3rd dicimal place if it is larger than 0.001, keeps 3 significant digits otherwise. The default value of rformat is F.

Value

Return the PMLE or MLE of the parameters with order = m0 (mixing proportions and component parameters), log-likelihood value at the PMLE or MLE and the penalized log-likelihood value at the PMLE.

Author(s)

Shaoting Li, Jiahua Chen and Pengfei Li

See Also

emtest.pois, pmle.pois, rmix.pois

Examples

```
#generate a random sample from a 2 component Poisson mixture,  
#compute the PMLE of the parameters under the 2 component Poisson mixture model,  
#plot the histogram of the observations and the fitted probability mass function.  
x <- rmix.pois(200,c(0.3,0.7),c(2,5))  
out <- pmle.pois(x,2)  
plotmix.pois(x,out)
```

residual1

residual diagnostics for mixture models

Description

This data set is from Lindsay and Roeder (1992). It includes the number of boys in families of size 8. The number of families is 53680. The data frame has 9 rows and 2 columns.

Usage

```
#Load this data set for subsequent analysis:  
data(residual1)
```

Format

This data frame contains the 2 columns,
count: number of boys in family.
freq: number of families with corresponding.

References

Lindsay, B. G. and Roeder, K. (1992). Residual diagnostics for mixture models. Journal of the American Statistical Association, 87(419), 785–794.

Examples

```
#fit a 2 component binomial mixture model to this data set.  
data(residual1)  
pmle.binom(residual1,8,2,1)
```

residual2	<i>residual diagnostics for mixture models</i>
-----------	--

Description

This data set is from Lindsay and Roeder (1992). It includes the number of boys in families of size 12. The number of families is 6115. The data frame has 13 rows and 2 columns.

Usage

```
#Load this data set for subsequent analysis:  
data(residual2)
```

Format

This data frame contains the 2 columns,
count: number of boys in family.
freq: number of families with corresponding.

References

Lindsay, B. G. and Roeder, K. (1992). Residual diagnostics for mixture models. Journal of the American Statistical Association, 87(419), 785–794.

Examples

```
#fit a 2 component binomial mixture model to this data set.  
data(residual2)  
pmle.binom(residual2, 12, 2, 1)
```

rmix.binom	<i>generate a random sample from a mixture of binomials</i>
------------	---

Description

Generate a random sample from a mixture of binomials.

Usage

```
rmix.binom(n, alpha, theta, size)
```

Arguments

n	sample size
alpha	vector of mixture proportions of length m, the order of the mixture.
theta	vector of probabilities of success of component distributions.
size	number of trials.

Value

Return sample of size n from an m-component binomial mixture.

Author(s)

Shaoting Li, Jiahua Chen and Pengfei Li

See Also

emtest.binom, plotmix.binom, pmle.binom

Examples

```
#generate a random sample from a 2 component binomial mixture,  
#plot the histogram of the data  
x <- rmix.binom(200,c(0.3,0.7),c(0.4,0.6),20)  
hist(x)
```

rmix.exp

generate a random sample from a mixture of exponentials

Description

Generate a random sample from a mixture of exponentials.

Usage

```
rmix.exp(n, alpha, theta)
```

Arguments

n	sample size
alpha	vector of mixture proportions, with length of m, the order of the mixture.
theta	vector of parameters of component distributions.

Value

Return a sample of size n from an m-component exponential mixture.

Author(s)

Shaoting Li, Jiahua Chen and Pengfei Li

See Also

emtest.exp, plotmix.exp, pmle.exp

Examples

```
#generate a random sample from a 2 component exponential mixture,  
#compute the sample mean and variance.  
x <- rmix.exp(200,c(0.3,0.7),c(2,8))  
mean(x)  
var(x)
```

rmix.norm	<i>generate a random sample from a mixture of normals</i>
-----------	---

Description

Generate a random sample from a mixture of normals.

Usage

```
rmix.norm(n, alpha, mu, sigma = rep(1,length(alpha)))
```

Arguments

n	sample size.
alpha	vector of mixture proportions of length m, the order of the mixture.
mu	vector of means of component distributions.
sigma	vector of standard deviations of component distributions, default value: sigma = rep(1,length(alpha)).

Value

Return a samples of size n from an m-component normal mixture.

Author(s)

Shaoting Li, Jiahua Chen and Pengfei Li

See Also

emtest.norm, emtest.norm0, plotmix.norm, plotmix.norm0, pmle.norm, pmle.norm0

Examples

```
#generate a random sample from a 2 component normal mixture,  
#plot the histogram of the data.  
x <- rmix.norm(200,c(0.3,0.7),c(-1,2),c(1,2))  
hist(x)
```

rmix.pois	<i>generate a random sample from a mixture of Poissons</i>
-----------	--

Description

Generate a random sample from a mixture of Poissons.

Usage

```
rmix.pois(n, alpha, theta)
```

Arguments

n	sample size.
alpha	vector of mixture proportions of length m, the order of the mixture.
theta	vector of parameters of each component.

Value

Return a sample of size n from an m-component Poisson mixture.

Author(s)

Shaoting Li, Jiahua Chen and Pengfei Li

See Also

rmix.binom, rmix.exp, rmix.norm

Examples

```
#generate a random sample from a 2 component Poisson mixture,  
#and compute the sample mean and variance.  
x <- rmix.pois(200,c(0.3,0.7),c(2,5))  
mean(x)  
var(x)
```

 SLC

SLC (sodium-lithium countertransport)

Description

This data set contains 190 SLC measurements studied by Roeder, K. (1994). Roeder (1994) analyzed this data and concluded that a three component normal mixture with equal variance is most suitable. Chen et al. (2012) also analyzed this data and gave a slightly better fit by a two component normal mixture with unequal variance. The data frame has 190 rows and 1 column.

Usage

```
#Load this data set for subsequent analysis:
data(SLC)
```

Format

This data frame contains one column:
 SLC: 190 slc measurements.

References

Chen, J., Li, P. and Fu, Y. (2012). Inference on the order of a normal mixture. *JASA*. 107, 1096–1105.

Roeder, K. (1994), A Graphical Technique for Determining the Number of Components in a Mixture of Normals, *Journal of the American Statistical Association*, 89, 487–500.

Examples

```
#fit a 2 component normal mixture model to this data set,
#test for order = 2 using EM-test.
data(SLC)
a <- c(0.6,0.4,0.2,0.3,0.01,0.01)
pmle.norm(SLC,2,1,inival=a)
emtest.norm(SLC,2,inival=a,len=5)
```

 timesoffailure

successive failure times

Description

The data set contains the times of successive failures for the air conditioning system of each member in a fleet of 13 Boeing 720 jet aircrafts. The data frame has 213 rows and 1 column.

Usage

```
#Load this data set for subsequent analysis:  
data(timesoffailure)
```

Format

This data frame contains one column,
x: pooled failure times about 213 observations.

References

Proschan, F. (1963). Theoretical explanation of observed decreasing failure rate. *Technometrics* 5, 375–83.

Examples

```
#fit a finite exponential mixture model of order 2 to this data set.  
data(timesoffailure)  
pmle.exp(timesoffailure,2,1)
```

zscores

z-scores

Description

This data set consists of 6033 z-scores, transformed from the two-sample t-test statistics based on the prostate cancer data set of Singh et al. (2002). The prostate cancer data set consists of gene expression levels of 6033 genes of 52 prostate cancer patients and 50 normal control subjects. By Efron (2004), if the gene is not differentially expressed, the corresponding z should follow $N(0, 1)$. Normal mixture with known variance 1 is suitable to this data set. The data frame has 6033 rows and 1 columns.

Usage

```
#Load this data set for subsequent analysis:  
data(zscores)
```

Format

This data frame contains one column:
z-scores: 6033 z-scores.

References

Efron, B. (2004), Large-Scale Simultaneous Hypothesis Testing: The Choice of a Null Hypothesis, *Journal of the American Statistical Association*, 99, 96–104.

Singh, D., Febbo, P. G., Ross, K., Jackson, D. G., Manola, J., Ladd, C., Tamayo, P., Renshaw, A. A., D’Amico, A. V., Richie, J. P., Lander, E. S., Loda, M., Kantoff, P. W., Golub, T. R., and Sellers, W. R. (2002), Gene Expression Correlates of Clinical Prostate Cancer Behavior, *Cancer Cell*, 1, 203–209.

Examples

```
## Not run:  
#fit a mixture of 2 normals with equal and known variance 1 to this data set.  
data(zscores)  
pmle.norm0(zscores,var=1,2,1)  
  
## End(Not run)
```

Index

*Topic **datagen**

rmix.binom, 27
rmix.exp, 28
rmix.norm, 29
rmix.pois, 30

*Topic **datasets**

agesch, 2
earthquake, 3
grains, 12
lengthofstay, 12
pearson, 13
residual1, 26
residual2, 27
SLC, 31
timesoffailure, 31
zscores, 32

*Topic **distribution**

rmix.binom, 27
rmix.exp, 28
rmix.norm, 29
rmix.pois, 30

*Topic **dplot**

plotmix.binom, 14
plotmix.exp, 15
plotmix.norm, 16
plotmix.norm0, 17
plotmix.pois, 18

*Topic **htest**

emtest.binom, 4
emtest.exp, 5
emtest.norm, 7
emtest.norm0, 8
emtest.pois, 10

*Topic **models**

pmle.binom, 20
pmle.exp, 21
pmle.norm, 22
pmle.norm0, 23
pmle.pois, 25

agesch, 2

earthquake, 3
emtest.binom, 4
emtest.exp, 5
emtest.norm, 7
emtest.norm0, 8
emtest.pois, 10

grains, 12

lengthofstay, 12

pearson, 13
plotmix.binom, 14
plotmix.exp, 15
plotmix.norm, 16
plotmix.norm0, 17
plotmix.pois, 18
pmle.binom, 20
pmle.exp, 21
pmle.norm, 22
pmle.norm0, 23
pmle.pois, 25

residual1, 26
residual2, 27
rmix.binom, 27
rmix.exp, 28
rmix.norm, 29
rmix.pois, 30

SLC, 31

timesoffailure, 31

zscores, 32