Package ‘robustloggamma’

May 17, 2016

Version 1.0-2
Date 2016-05-16
Title Robust Estimation of the Generalized log Gamma Model
Author Claudio Agostinelli <claudio.agostinelli@unitn.it>, Al- fio Marazzi <Alfio.Marazzi@chuv.ch>, Vic- tor Yohai <vyohai@uolsinectis.com.ar> and Alex Randriamihari- soa <Alex.Randriamiharisoa@chuv.ch>
Maintainer Claudio Agostinelli <claudio.agostinelli@unitn.it>
Description Robust estimation of the generalized log gamma model is provided using Quan- tile Tau estimator, Weighted Likelihood estimator and Truncated Maximum Likelihood estima- tor. Functions for regression and censored data are also available.
Depends R (>= 2.15.1)
Imports robustbase, survival, numDeriv, RobustAFT (>= 1.4-1)
LazyData yes
License GPL (>= 2)
NeedsCompilation yes
Repository CRAN
Date/Publication 2016-05-17 00:48:06

R topics documented:

alco .................................................. 2
drg2000 ................................................. 3
loggammacenslmrob .................................... 4
loggammacensrob ..................................... 7
LogGammaDist .......................................... 8
loggammarob .......................................... 9
loggammarob.control .................................. 11
loggammarob.test ..................................... 13
loggammarob.wilks ................................... 15
sqrtloggamma ......................................... 16
summary.loggammarob ................................. 16
alcoa

Description

Three samples (with labels A, B and C) from measurement quality assurance (QA) data base of ALCOA aluminium refineries in Western Australia.

Usage

data(alcoa)

Format

A data frame with 668 observations on the following 3 variables.

dist  a numeric vector
ratio  a numeric vector
label  a factor with levels A B C

Details

Under ALCOA’s QA program, several thousand bauxite ore samples are routinely submitted to Fourier transform infrared spectroscopy (Eyer and Riley, 1999). Part of the quality assurance is the need to automatically highlight unusual spectra and this is obtained with the help of special statistical diagnostics - called representation indicators - derived from the Fourier transform.

Source

Alcoa World Alumina

References

drg2000

Examples

```r
data(alcoa)
par(mfcol=c(1,2))
boxplot(I(log(alcoa$ratio))~alcoa$label)
boxplot(I(log(alcoa$dist))~alcoa$label)
```


Description

The data refer to 70323 stays that were observed in year 2000 in a group of Swiss hospitals within a pilot study aimed at the implementation of a diagnosis-related grouping (DRG) system. DRG systems are used in modern hospital management to classify each individual stay into a group according to the patient characteristics.

Usage

```r
data(drg2000)
```

Format

A data frame with 70323 observations on the following 4 variables.

- `los` numeric. Length of the hospitalization (LOS).
- `cost` numeric. Cost of the hospitalization in Switzerland franc.
- `mdc` numeric. Major Diagnostic Category.
- `apdrg` numeric. All Patient Diagnosis Related Group.

References


Examples

```r
data(drg2000)
str(drg2000)
```
Robust estimation of an Accelerated Failure Time model with extended Log Gamma errors in presence of censored observations.

Description

Three different type of robust procedures are provided for the estimation of the parameters in an Accelerated Failure Time model with extended Log Gamma errors in presence of censored observations. Maximum Likelihood is also provided.

Usage

loggammacenslmrob(formula, delta, data, subset, weights, na.action, method = c("oneTML", "oneWL", "TWQTau", "TQTau", "ML"), model = TRUE, x = !control$compute.rd, y = FALSE, singular.ok = TRUE, contrasts = NULL, offset = NULL, control = NULL, ini = NULL, ...)

Arguments

formula a symbolic description of the model to be fit. See lm and formula for more details.
delta numeric, 0 or 1. 0 indicated censored observations.
data an optional data frame, list or environment (or object coercible by as.data.frame to a data frame) containing the variables in the model. If not found in data, the variables are taken from environment(formula), typically the environment from which loggammacenslmrob is called.
subset an optional vector specifying a subset of observations to be used in the fitting process.
weights an optional vector of weights to be used in the fitting process (in addition to the robustness weights computed in the fitting process).
na.action a function which indicates what should happen when the data contain NAs. The default is set by the na.action setting of options, and is na.fail if that is unset. The “factory-fresh” default is na.omit. Another possible value is NULL, no action. Value na.exclude can be useful.
method string specifying the estimator-chain. Default is TML one step.
model, x, y logicals. If TRUE the corresponding components of the fit (the model frame, the model matrix, the response) are returned.
singular.ok logical. If FALSE (the default in S but not in R) a singular fit is an error.
contrasts an optional list. See the contrasts.arg of model.matrix.default.
offset this can be used to specify an a priori known component to be included in the linear predictor during fitting. An offset term can be included in the formula instead or as well, and if both are specified their sum is used.
control a list specifying control parameters; use the function loggammarob.control and see its help page.

init an optional argument to specify or supply the initial estimate. See Details.

... additional arguments can be used to specify control parameters directly instead of (but not in addition to!) via control.

Value

An object of class lmrob; a list including the following components:

coefficients The estimate of the coefficient vector for the regression part. First element is the intercept and it would be equals to the parameter mu.

mu The estimate of the intercept parameter.

sigma The estimate of the scale parameter.

lambda The estimate of the shape parameter.

fitted.values Fitted values associated with the estimator.

residuals Residuals associated with the estimator.

cut.lower, cut.upper Cut points for the method based on TML (missing for the other methods).

iter number of iterations.

weights the specified weights (missing if none were used).

errors errors messages.

n.ret number of non zero robust weights for the method based on TML.

control control argument.

converged TRUE if the procedure converged.

method method used during the fit.

rank the numeric rank of the fitted linear model.

rweights the “robustness weights”.

df.residual the residual degrees of freedom.

degree.freedom the same as df.residual

delta as in input.

df a vector with 3 components: (number of linearly independent regressors, df.residual, number of regressors).

xlevels (only where relevant) a record of the levels of the factors used in fitting.

call the matched call.

terms the terms object used.

model if requested (the default), the model frame used.

x if requested, the model matrix used.

y if requested, the response used.

scale square root of the sigma parameter.
na.action  (where relevant) information returned by \texttt{model.frame} on the special handling of NAs.

offset  the offset used (missing if none were used).

contrasts  (only where relevant) the contrasts used.

In addition, non-null fits will have \texttt{qr} relating to the linear fit, for use by extractor functions such as \texttt{summary}.

\textbf{Author(s)}

C. Agostinelli, A. Marazzi and V.J. Yohai

\textbf{References}


\textbf{See Also}

\texttt{loggammacensrob} for the case of censored observations without covariates.

\textbf{Examples}

\footnotesize{
\begin{verbatim}
## Not run:
n <- 50
p <- 2
set.seed(1234)
X <- matrix(rnorm(p*n, sd=2), ncol=p)
mu <- 2
beta <- rep(2,p)
sigma <- 2
lambda <- 1
linear <- mu + X
y <- rloggamma(n=n, mu=linear, sigma=sigma, lambda=lambda)
cens <- rloggamma(n=n, mu=linear+3, sigma=sigma, lambda=lambda)
delta <- as.numeric(y <= cens)
y[delta==0] <- cens[delta==0]
x <- data.frame(y=as.vector(y), x1=X[,1], x2=X[,2])
res <- loggammacenslmrob(y~x1+x2, delta=delta, data=x)
summary(res)

## End(Not run)
\end{verbatim}
}
Robust estimation of the three parameters extended Log Gamma model in presence of censored observations.

Description

Five different type of robust procedures are provided for the estimation of the parameters in the three parameters extended Log Gamma model in presence of censored observations.

Usage

```
loggammacensrob(x, delta, start=NULL, weights=rep(1, length(x)),
    method=c("oneML", "oneWL", "TWQ Tau", "TQTau", "ML"), control, ...)
```

Arguments

- **x** numeric. A vector with the dataset.
- **delta** numeric, 0 or 1. 0 indicated censored observations.
- **start** NULL or numeric. A vector of length 3 to be used when method is WL, oneWL and ML, otherwise starting values are obtained from WQTau in the first two methods and QTau in the last one.
- **weights** numeric. A vector of weights used in the method TQTau.
- **method** character. The method used, see Details below. Default is oneML a one step Truncated Maximum Likelihood estimates starting from TQTau.
- **control** list. An object from function `loggammarob.control`.
- **...** further arguments can be passed directly to the function instead of using the control argument.

Value

An object of class 'loggammacensrob'. A list that includes the following components:

- **mu** location parameter.
- **sigma** scale parameter.
- **lambda** shape parameter.
- **eta** estimate of E(exp(x)) parameter.
- **weights** the final weights.
- **iterations** number of iterations.
- **error** if not NULL or 0, an error occurred.
- **data** the original dataset.
- **delta** the original dataset.
- **method** the method used.
LogGammaDist

Author(s)
C. Agostinelli, A. Marazzi and V.J. Yohai

References

See Also
LogGammaDist and loggammarob for the case without censored observations.

Examples

```r
set.seed(1234)
x <- sort(rloggamma(n=80, lambda=1))
cens <- rloggamma(n=80, mu=1.5, lambda=1)
delta <- as.numeric(x <= cens)
x[delta==0] <- cens[delta==0]
res <- loggammacensrob(x, delta, method="oneTML",
    control=loggammarob.control(lower=0, upper=2, n=30))
print(res)
summary(res)
```

---

LogGammaDist

The Extended Log Gamma Distribution

Description
Density, distribution function, quantile function and random generation for the Extended Log Gamma distribution with parameters mu (location), sigma (scale) and lambda (shape).

Usage

```r
dloggamma(x, mu=0, sigma=1, lambda, log=FALSE, zero=0.0001)
ploggamma(q, mu=0, sigma=1, lambda, lower.tail=TRUE, log.p=FALSE, zero=0.0001)
qloggamma(p, mu=0, sigma=1, lambda, zero=0.0001)
rloggamma(n, mu=0, sigma=1, lambda, zero=0.0001)
```

Arguments

- `x, q` numeric. Vector of quantiles.
- `p` numeric. Vector of probabilities.
- `n` numeric. Number of observations.
- `mu` numeric. Location parameter.
- `sigma` numeric. Scale parameter.
**loggammarob**

lambda numeric. Shape parameter.

log, log.p logical. If TRUE, probabilities/densities p are returned as log(p)

lower.tail logical. If TRUE (default), probabilities are P[X <= x], otherwise, P[X > x].

zero numeric. A threshold, values of lambda smaller than zero will report the asymptotic value of the function at 0.

Value
dloggamma gives the density, ploggamma gives the distribution function, qloggamma gives the quantile function, and rloggamma generates random deviates.

See Also

 GammaDist

Examples

```r
set.seed(1234)
x <- rloggamma(10, lambda=1)
x
```

---

**Description**

Four different type of robust procedures are provided for the estimation of the parameters in the three parameters extended Log Gamma model.

**Usage**

```r
loggammarob(x, start=NULL, weights = rep(1, length(x)),
method=c("oneWL", "WQTau", "WL", "QTau", "ML"), control, ...)
```

---

**Arguments**

- **x** numeric. A vector with the dataset.
- **start** NULL or numeric. A vector of length 3 to be used when method is WL, oneWL and ML, otherwise starting values are obtained from WQTau in the first two methods and QTau in the last one.
- **weights** numeric. A vector of weights used in the method QTau.
- **method** character. The method used, see Details below. Default is oneWL a one step weighted likelihood estimates starting from WQTau.
control list. An object from function `loggammarob.control`.
digits minimal number of _significant_ digits, see `print.default`.
...

Value

An object of class ’loggammarob’. A list that includes the following components:

- mu location parameter.
- sigma scale parameter.
- lambda shape parameter.
- eta estimate of E(exp(x)) parameter.
- weights the final weights.
- iterations number of iterations.
- error used only in method `onewL`. If 1 then the Jacobian matrix is not invertible.
- data the original dataset.
- method the method used.

Author(s)

C. Agostinelli, A. Marazzi, V.J. Yohai and A. Randriamiharisoa

References


See Also

- `loggammadist`

Examples

```r
set.seed(1234)
x <- sort(rloggamma(n=80, lambda=1))
res <- loggammarob(x, control=loggammarob.control(lower=0, upper=2, n=30))
print(res)
```
Description

Tuning parameters for 'loggammarob' for all four methods implemented.

Usage

loggammarob.control(method="oneWL", tuning.rho=1.547647, tuning.psi=6.08, lower=-7, upper=7, n=201, max.it=750, refine.tol=1e-6, solve.tol=1e-7, nResample=100, bw=0.3, smooth=NULL, raf=c("NED","GKL","PWD","HD","SCHI2"), tau=1, subdivisions=1000, lambda.step=TRUE, sigma.step=TRUE, step=1, minw=0.04, nexp=1000, reparam=NULL, bootstrap=FALSE, bootstrap.lambda=NULL, qthreshold=0.9, nTML=2000, xmax=1e100, iter=1, pcut=0.997, compute.rd=FALSE, eps.outlier= function(nobs) 0.1 / nobs)

Arguments

- **method**: character. The method to be used. See Details below.
- **tuning.rho**: numeric. Tuning constant c1 for the Tau-estimator.
- **tuning.psi**: numeric. Tuning constant c2 for the Tau-estimator.
- **lower**: numeric. The lower limit for the search grid of the shape parameter.
- **upper**: numeric. The upper limit for the search grid of the shape parameter.
- **n**: numeric. The number of subdivisions for the search grid of the shape parameter.
- **max.it**: numeric. Maximum number of iterations.
- **refine.tol**: numeric. Relative convergence tolerance for the fully iterated best candidates.
- **solve.tol**: numeric. Relative tolerance for inversion. Hence, this corresponds to `solve.default()`'s tol.
- **nResample**: integer. Number of re-sampling candidates to be used to find the initial estimator. Currently defaults to 100 which works well in most situations.
- **bw**: numeric. Bandwidth used in the Weighted Likelihood steps.
- **smooth**: NULL or numeric. When not NULL the parameter bw is set to smooth times the square root of the starting value of the scale parameter.
- **raf**: character. Residual adjustment function used in the Weighted Likelihood steps: raf="NED": Negative Exponential Disparity RAF, raf="GKL": Generalized Kullback-Leibler Divergence Family with parameter tau (see below) RAF, raf="PWD": Power Divergence Family with parameter tau (see below) RAF, raf="HD": Hellinger Distance RAF, raf="SCHI2": Symmetric Chi-Squared Disparity RAF. Default value is "NED".
tau parameter used when raf is equal to "PWD" or "GKL".
subdivisions numeric. Number of subdivisions used in the approximation of the smoothed model density in the Weighted Likelihood steps.
lambda.step logical.
sigma.step logical.
step integer. Number of steps to be performed when method is "oneWL" (only implemented for the functions for non censored data).
minw numeric. A scalar in the interval [0,1]. When method is "oneWL" the weights smaller than minw are set to zero.
nexp integer. When method is "oneWL" number of quantile points used in the approximation of the Expected Jacobian matrix.
reparam list. When method is "oneWL" a reparametrization is possible for the "sigma" parameter. See function sqrtloggamma for an example.
bootstrap logical. To use loggammarob in boot
bootstrap.lambda numeric. An initial estimates for the shape parameter. To use loggammarob in boot
qthreshold numeric. A value in (0.5, 1] used for TQtau e TWQtau procedure. It is the quantile order to truncated the data on the right.
TML numeric. Number of elements to be considered in the grid for finding the cut points of the TML.
xmax numeric. A threshold value for the log likelihood. Used for ML.
iter numeric. Number of iterations to be performed. Only working for TML.
pcut numeric. Fraction to determined the cut points of the TML.
compute.rd logical. Indicating if robust distances (based on the MCD robust covariance estimator covMcd) are to be computed for the robust diagnostic plots. This may take some time to finish, particularly for large data sets, and can lead to singularity problems when there are factor explanatory variables (with many levels, or levels with "few" observations). Hence, is FALSE by default.
eps.outlier limit on the robustness weight below which an observation is considered to be an outlier. Either a numeric(1) or a function that takes the number of observations as an argument. Used in summary.loggammacenslmrob.

Author(s)
C. Agostinelli, A. Marazzi, V.J. Yohai and A. Randriamiharisoa

References
C. Agostinelli, A. Marazzi, V.J. Yohai and A. Randriamiharisoa (2016)

See Also

loggammarob

Examples

## Show the default settings:
str(loggammarob.control())

loggammarob.test  Robust inference for the generalized loggamma model.

Description

Robust Tests and confidence intervals for the parameters of the generalized loggamma model.

Usage

loggammarob.test(x, mu = NULL, sigma = NULL, lambda = NULL, eta = NULL, type = "Wald", conf.level = 0.95, prob = 0.00001)

Arguments

x list. An object of class 'loggammarob'
mu numeric. Null value for the location parameter
sigma numeric. Null value for the scale parameter
lambda numeric. Null value for the shape parameter
eta numeric. Null value for E(exp(X)) parameter
type character. Type of inference, for now only Wald test is performed
conf.level numeric. Level of the confidence interval.
prob quantile order of the loggamma to be used in the numerical calculation of the expected Fisher Information.

Details

If no null values are provided, the function report the test for mu=0, sigma=1 and lambda=0.
Value

An object of class 'htest' containing the following components:

- **statistic**: the value of the t-statistic.
- **parameter**: the degrees of freedom for the statistic.
- **p.value**: the p-value for the test.
- **conf.int**: a confidence interval for the parameter appropriate to the specified alternative hypothesis.
- **estimate**: the estimate of the parameter(s).
- **null.value**: the specified hypothesized value of the parameter.
- **alternative**: a character string describing the alternative hypothesis.
- **method**: a character string indicating what type of test was performed.
- **data.name**: a character string giving the name(s) of the data.

Author(s)

A. Marazzi, C. Agostinelli, V.J. Yohai and A. Randriamiharisoa

References


See Also

- loggammarob

Examples

```r
set.seed(1234)
x <- sort(rloggamma(n=80, lambda=1))
res <- loggammarob(x, control=loggammarob.control(lower=0, upper=2, n=30))
loggammarob.test(res, mu=0) # only location
loggammarob.test(res, mu=0, sigma=1) # location and scale
loggammarob.test(res, eta=1) # E(exp(X))
```
Description

Performs a robust Wilks test to check equality between scale and shape parameters of a generalized log gamma model.

Usage

`loggammarob.wilks(x, thetainit = NULL, method = "L-BFGS-B", lower = c(-Inf, 1e-04), upper = c(Inf, Inf), ...)`

Arguments

- `x`: list. An object of class `loggammarob`, typically created by `loggammarob`.
- `thetainit`: NULL or numeric. A vector of length 2 to be used as initial values, first element corresponds to mu, second element to sigma/lambda under the Null Hypothesis that sigma is equal to lambda.
- `method`: parameter passed to function `optim`.
- `lower`: parameter passed to function `optim`.
- `upper`: parameter passed to function `optim`.
- `...`: further parameters passed to function `optim`.

Value

an object of class `htest`.

Author(s)

C. Agostinelli, A. Marazzi, V.J. Yohai and A. Randriamiharisoa

References


See Also

`loggammarob.test` for robust (weighted) t-test on one or more parameters of the generalized loggamma model.

Examples

```r
set.seed(1234)
x <- sort(log(rgamma(n=30, shape=2, scale=2)))
res <- loggammarob(x)
loggammarob.wilks(res)
```
sqrtloggamma

A reparametrization for the sigma parameter in loggamma model.

Description

A reparametrization for the sigma parameter in loggamma model. This is used in the function loggammarob.control.

Usage

sqrtloggamma

Format

The format is a list of 3 elements

$ \text{gam} : \text{function} (\text{sigma})$
$ \text{gaminv} : \text{function} (\text{gam})$
$ \text{delta} : \text{function} (\text{sigma})$

Details

The object must be a length of dimension 3. Each component is a function. The first component \text{gam} performs the transformation for the parameter \text{sigma}; \text{gaminv} is the inverse function and \text{delta} is the derivative of the \text{gam} function wrt \text{sigma} to the -1 power.

Examples

str(sqrtloggamma)
set.seed(1234)
x <- sort(rloggamma(n=50, lambda=1))
res <- loggammarob(x, control=loggammarob.control(lower=0, upper=2, n=30, reparam=sqrtloggamma))
print(res)

summary.loggammarob

Summary Method for "loggammarob" Objects

Description

Summary method for R object of class "loggammarob" and print method for the summary object.
Usage

```r
## S3 method for class 'loggammarob'
summary(object, p = NULL, conf.level = 0.95,
         prob = 1e-05, ...)
## S3 method for class 'summary.loggammarob'
print(x, digits = max(3, getOption("digits") - 3), ...)
```

Arguments

- `object`: an R object of class `loggammarob`, typically created by `loggammarob`.
- `p`: numeric. Values in the interval [0,1]. Quantile orders for which point estimation and confidence interval are required.
- `conf.level`: numeric. A scalar or a vector of length 4+length(p). This is the confidence level used to construct confidence intervals for the four parameters mu, sigma, lambda and eta and for the quantiles.
- `prob`: numeric. Value in the interval [0,1]. This is used to determine the interval of numerical integration in the evaluation of the asymptotic variance and covariance matrix. See details below.
- `x`: an R object of class `summary.loggammarob`, typically resulting from `summary(loggammarob(...))`.
- `digits`: number of digits for printing, see `digits` in `options`.
- `...`: potentially more arguments passed to methods.

Details

The `prob` argument determines the interval of the numerical integration in the evaluation of the asymptotic variance and covariance matrix with the following code:

\[
q \text{loggamma}(p = \text{prob}/2, \lambda = \hat{\lambda}) \\
q \text{loggamma}(p = 1 - \text{prob}/2, \lambda = \hat{\lambda})
\]

for the lower limit and \(q \text{loggamma}(p = \text{1-prob}/2, \lambda = \hat{\lambda})\) for the upper limit. Here \(\hat{\lambda}\) is the estimate.

Value

An object of class `summary.loggammarob`. A list that includes the following components:

- `muse`: standard error for the mu estimate.
- `sigmase`: standard error for the sigma estimate.
- `lambdase`: standard error for the lambda estimate.
- `etase`: standard error for the eta estimate.
- `muconf.int`: confidence interval for the mu parameter.
- `sigmconf.int`: confidence interval for the sigma parameter.
- `lambdacconf.int`: confidence interval for the lambda parameter.
- `etacconf.int`: confidence interval for the eta parameter.

If `p` is not NULL then

- `q`: quantiles estimate.
- `qse`: standard error for the require quantiles estimate.
- `qconf.int`: confidence interval for the require quantiles parameter.
Author(s)
C. Agostinelli, A. Marazzi, V.J. Yohai and A. Randriamiharisoa

References

See Also
loggammarob

Examples
set.seed(1234)
x <- rloggamma(n=50, lambda=1)
res <- loggammarob(x, control=loggammarob.control(lower=0, upper=2, n=30))
summary(res)
Index

*Topic **datasets**
  alcoa, 2
  drg2000, 3
  sqrtloggamma, 16

*Topic **distribution**
  LogGammaDist, 8

*Topic **models**
  loggammacenslmrob, 4
  loggammacensrob, 7
  loggamarob, 9
  loggamarob.control, 11
  loggamarob.test, 13
  loggamarob.wilks, 15
  summary.loggamarob, 16

*Topic **robust**
  loggammacenslmrob, 4
  loggammacensrob, 7
  loggamarob, 9
  loggamarob.control, 11
  loggamarob.test, 13
  loggamarob.wilks, 15
  summary.loggamarob, 16

alcoa, 2
as.data.frame, 4

boot, 12

dloggamma (LogGammaDist), 8
drg2000, 3

formula, 4

GammaDist, 9

list, 5
lm, 4
loggammacenslmrob, 4
loggammacensrob, 6, 7
LogGammaDist, 8, 8, 10
loggamarob, 8, 9, 12-15, 17, 18
loggamarob.control, 5, 7, 10, 11
loggamarob.test, 13, 15
loggamarob.wilks, 15

model.frame, 6

na.exclude, 4
na.fail, 4
na.omit, 4

offset, 4
opti, 15
options, 4, 17

ploggamma (LogGammaDist), 8
print, 16
print.default, 10
print.loggamarob (loggamarob), 9
print.summary.loggamarob
(summary.loggamarob), 16

qloggamma (LogGammaDist), 8

rloggamma (LogGammaDist), 8

sqrtloggamma, 12, 16
summary.loggamarob, 16

summaryNloggammarob
(summary.loggamarob), 16


