Package 'MVT'

October 11, 2015

Version 0.3

Date 2015-10-09

Title Estimation and Testing for the Multivariate t-Distribution

Author Felipe Osorio

Maintainer Felipe Osorio <felipe.osorio@ucv.cl>

Description Routines to perform estimation and inference under the multivariate t-distribution. Currently, the following methodologies are implemented: multivariate mean and covariance estimation, hypothesis testing about the mean, equicorrelation and homogeneity of variances, the Wilson-Hilferty transformation, QQ-plots with envelopes and random variate generation. Some auxiliary functions are also provided.

Imports stats, utils, graphics

Suggests heavy

License GPL (>= 2)

URL http://www.ies.ucv.cl/mvt/

LazyLoad yes

NeedsCompilation yes

Repository CRAN

Date/Publication 2015-10-11 00:12:53

R topics documented:

`center.test` *One-sample location test*

Description

Performs several test for testing the mean of a multivariate-t population. Generalized Hotelling's T-squared, likelihood ratio test, score, Wald and gradient can be used as a test statistic.

Usage

```
center.test(object, center, test = "LRT")
```

Arguments

`object`	object of class `'studentFit'` representing the fitted model.
`center`	a vector indicating the hypothesized value of the mean.
`test`	test statistic to be used. One of "hotelling", "LRT" (default), "Wald", "score" or "gradient".

Value

A list of class 'center.test' with the following elements:

`statistic`	value of the statistic, i.e. the value of either Hotelling T-squared, likelihood ratio test, Wald, score or gradient test.
`parameter`	the degrees of freedom for the test statistic, which is chi-square distributed.
`p.value`	the p-value for the test.
`estimate`	the estimated mean vector.
`null.value`	the hypothesized value for the mean vector.
`method`	a character string indicating what type of test was performed.
`null.fit`	a list representing the fitted model under the null hypothesis.
`data`	name of the data used in the test.

References

Anderson, T.W. (2003). *An Introduction to Multivariate Statistical Analysis*. Wiley, New York.

Hotelling, H. (1931). The generalization of Student's ratio. *Annals of Mathematical Statistics* **2**, 360-378.

Osorio, F., and Galea, M. (2015). Statistical inference in multivariate analysis using the t-distribution. Unpublished manuscript.

Examples

```
data(cork)
x <- sweep(cork[,2:4], 1, cork[,1], "-")
colnames(x) <- c("E_N", "S_N", "W_N")
pairs(x)

fit <- studentFit(x, family = Student(eta = .25))
z <- center.test(fit, center = c(0,0,0), test = "score")
z
```

`commutation` *Commutation matrix for square matrices*

Description

This function returns a commutation matrix of order `n` which transforms, for an `n` by `n` matrix `x`, `vec(x)` into `vec(t(x))`.

Usage

```
commutation(n = 2)
```

Arguments

`n`	a non-negative integer.

Value

Returns an square commutation matrix of order n^2 .

References

Magnus, J.R., and Neudecker, H. (1979). The commutation matrix: some properties and applications. *The Annals of Statistics* **7**, 381-394.

Magnus, J.R., and Neudecker, H. (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics*, 2nd Edition. Wiley, New York.

Examples

```
a <- matrix(1:9, ncol = 3)
kmat <- commutation(nrow(a))
vec <- as.vector(a)
## vectorization of t(a):
kmat %*% vec
```

`companies` *Financial data*

Description

Data extracted from Standard & Poor's Compustat PC Plus. This dataset has been used to illustrate some influence diagnostic techniques.

Usage

```
data(companies)
```

Format

A data frame with 26 observations on the following 3 variables.

book book value in dollars per share at the end of 1992.

net net sales in millions of dollars in 1992.

ratio sales to assets ratio in 1992.

Source

Hadi, A.S., and Nyquist, H. (1999). Frechet distance as a tool for diagnosing multivariate data. *Linear Algebra and Its Applications* **289**, 183-201.

Hadi, A.S., and Son, M.S. (1997). Detection of unusual observations in regression and multivariate data. In: A. Ullah, D.E.A. Giles (Eds.) *Handbook of Applied Economic Statistics*. Marcel Dekker, New York. pp. 441-463.

cork *Cork borings*

Description

Measurements of the weight of cork borings taken from the north (N), east (E), south (S), and west (W) directions of 28 trees. It is of interest to compare the bark thickness (and hence weight) in the four directions.

Usage

```
data(cork)
```

Format

A data frame with 28 observations on the following 4 variables.

N north.

E east.

S south.

W west.

Source

Mardia, K.V., Kent, J.T., and Bibby, J.M. (1979). *Multivariate Analysis*. Academic Press, London.

duplication *Duplication matrix*

Description

This function returns a duplication matrix of order `n` which transforms, for a symmetric matrix `x`, `vech(x)` into `vec(x)`.

Usage

```
duplication(n = 1)
```

Arguments

`n`	order of the duplication matrix.

Value

Returns an n^2 by $n(n+1)/2$ matrix.

References

Magnus, J.R., and Neudecker, H. (1980). The elimination matrix, some lemmas and applications. *SIAM Journal on Algebraic Discrete Methods* **1**, 422-449.

Magnus, J.R., and Neudecker, H. (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics*, 2nd Edition. Wiley, New York.

Examples

```
dmat <- duplication(3)
a <- matrix(c( 1, 2, 3,
               2, 3, 4,
               3, 4, 5), nrow = 3)
vech <- a[upper.tri(a, diag = TRUE)]
vec <- dmat %*% vech
as.vector(vec)
```

`envelope` *QQ-plot with simulated envelopes*

Description

Constructs a normal QQ-plot using a Wilson-Hilferty transformation for the estimated Mahalanobis distances obtained from the fitting procedure.

Usage

```
envelope(object, reps = 50, conf = 0.95, plot.it = TRUE)
```

Arguments

`object`	an object of class `'studentFit'` representing the fitted model.
`reps`	number of simulated point patterns to be generated when computing the envelopes. The default number is 50.
`conf`	the confidence level of the envelopes required. The default is to find 95% confidence envelopes.
`plot.it`	if TRUE it will draw the corresponding plot, if FALSE it will only return the computed values.

Value

A list with the following components :

`transformed`	a vector with the z-scores obtained from the Wilson-Hilferty transformation.
`envelope`	a matrix with two columns corresponding to the values of the lower and upper pointwise confidence envelope.

References

Atkinson, A.C. (1985). *Plots, Transformations and Regression*. Oxford University Press, Oxford.

Osorio, F., and Galea, M. (2015). Statistical inference in multivariate analysis using the t-distribution. Unpublished manuscript.

See Also

`wilson.hilferty`

Examples

```
data(PFM)
fit <- studentFit(~ cuprum.D + habitat.D + planvital.D + provida.D, data = PFM,
  family = Student(eta = 0.25))
envelope(fit, reps = 500, conf = 0.95)
```

---

equicorrelation.test *Equicorrelation test*

Description

Performs several test for testing that the covariance matrix follows an equicorrelation (or compound symmetry) structure. Likelihood ratio test, score, Wald and gradient can be used as a test statistic.

Usage

```
equicorrelation.test(object, test = "LRT")
```

Arguments

`object`	object of class `'studentFit'` representing the fitted model.
`test`	test statistic to be used. One of "LRT" (default), "Wald", "score" or "gradient".

Value

A list of class 'equicorrelation.test' with the following elements:

`statistic`	value of the statistic, i.e. the value of either Likelihood ratio test, Wald, score or gradient test.
`parameter`	the degrees of freedom for the test statistic, which is chi-square distributed.
`p.value`	the p-value for the test.
`estimate`	the estimated covariance matrix.
`null.value`	the hypothesized value for the covariance matrix.
`method`	a character string indicating what type of test was performed.
`null.fit`	a list representing the fitted model under the null hypothesis.
`data`	name of the data used in the test.

References

Osorio, F., and Galea, M. (2015). Statistical inference in multivariate analysis using the t-distribution. Unpublished manuscript.

Sutradhar, B.C. (1993). Score test for the covariance matrix of the elliptical t-distribution. *Journal of Multivariate Analysis* **46**, 1-12.

Examples

```
data(examScor)
fit <- studentFit(examScor, family = Student(eta = .25))
fit

z <- equicorrelation.test(fit, test = "LRT")
z
```

`examScor` *Open/Closed book data*

Description

Dataset from Mardia, Kent and Bibby on 88 students who took examinations in five subjects. The first two subjects were tested with closed book exams and the last three were tested with open book exams.

Usage

```
data(examScor)
```

Format

A data frame with 88 observations on the following 5 variables.

mechanics mechanics, closed book exam.

vectors vectors, closed book exam.

algebra algebra, open book exam.

analysis analysis, open book exam.

statistics statistics, open book exam.

Source

Mardia, K.V., Kent, J.T., and Bibby, J.M. (1979). *Multivariate Analysis*. Academic Press, London.

fisher.info *Fisher information matrix*

Description

Calculate the Fisher information matrix for the multivariate t-distribution.

Usage

```
fisher.info(object)
```

Arguments

object	an object of class 'studentFit' representing the fitted model.

Value

A square matrix of order $p(p+3)/2+1$ containing the Fisher information of the fitted model.

References

Osorio, F., and Galea, M. (2015). Statistical inference in multivariate analysis using the t-distribution. Unpublished manuscript.

See Also

studentFit

Examples

```
data(PFM)
fit <- studentFit(~ cuprum.D + habitat.D + planvital.D + provida.D, data = PFM,
  family = Student(eta = 0.25))
info <- fisher.info(fit)
```

homogeneity.test *Test of variance homogeneity of correlated variances*

Description

Performs several test for testing equality of $p \geq 2$ correlated variables. Likelihood ratio test, score, Wald and gradient can be used as a test statistic.

Usage

```
homogeneity.test(object, test = "LRT", type = "scale")
```

Arguments

object	object of class 'studentFit' representing the fitted model.
test	test statistic to be used. One of "LRT" (default), "Wald", "score" or "gradient".
type	one of "scale" (default) or "both" indicating the type of the hypothesis to test homogeneity of variances or variances *and* means, respectively.

Value

A list of class 'homogeneity.test' with the following elements:

statistic	value of the statistic, i.e. the value of either Likelihood ratio test, Wald, score or gradient test.
parameter	the degrees of freedom for the test statistic, which is chi-square distributed.
p.value	the p-value for the test.
estimate	the estimated covariance matrix.
null.value	the hypothesized value for the covariance matrix.
method	a character string indicating what type of test was performed.
null.fit	a list representing the fitted model under the null hypothesis.
data	name of the data used in the test.

References

Harris, P. (1985). Testing the variance homogeneity of correlated variables. *Biometrika* **72**, 103-107.

Modarres, R. (1993). Testing the equality of dependent variables. *Biometrical Journal* **7**, 785-790.

Osorio, F., and Galea, M. (2015). Statistical inference in multivariate analysis using the t-distribution. Unpublished manuscript.

Examples

```
data(examScor)
fit <- studentFit(examScor, family = Student(eta = .25))
fit

z <- homogeneity.test(fit, test = "LRT")
z
```

`kurtosis` *Mardia's multivariate kurtosis coefficient*

Description

This function computes the kurtosis of a multivariate distribution and estimates the kurtosis parameter for the t-distribution using the method of moments.

Usage

```
kurtosis(x, center, cov)
```

Arguments

`x`	vector or matrix of data with, say, p columns.
`center`	mean vector of the distribution or second data vector of length p.
`cov`	covariance matrix (p x p) of the distribution.

Value

A list with the following components :

`kurtosis`	returns the value of Mardia's multivariate kurtosis.
`kappa`	returns the excess kurtosis related to a multivariate t-distribution.
`eta`	estimated shape (kurtosis) parameter using the methods of moments, only valid if $0 \leq \eta < 1/4$.

References

Mardia, K.V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika* **57**, 519-530.

Osorio, F., and Galea, M. (2015). Statistical inference in multivariate analysis using the t-distribution. Unpublished manuscript.

Examples

```
data(companies)
S <- cov(companies)
kurtosis(companies, colMeans(companies), S)
```

mvt.control *Set control parameters*

Description

Allows users to set control parameters for the estimation routine available in MVT.

Usage

```
mvt.control(maxIter = 2000, tolerance = 1e-6, fix.shape = FALSE)
```

Arguments

maxIter	maximum number of iterations. The default is 2000.
tolerance	the relative tolerance in the iterative algorithm.
fix.shape	whether the shape parameter should be kept fixed in the fitting processes. The default is fix.shape = FALSE.

Value

A list of control arguments to be used in a call to studentFit.

A call to mvt.control can be used directly in the control argument of the call to studentFit.

Examples

```
ctrl <- mvt.control(maxIter = 500, tol = 1e-04, fix.shape = TRUE)
data(PFM)
studentFit(~ cuprum.D + habitat.D + planvital.D + provida.D, data = PFM,
  family = Student(eta = 0.25), control = ctrl)
```

PFM *Returns from the Chilean Pension Funds*

Description

Dataset from Osorio and Galea (2015) on the Chilean pension funds. The data corresponds to monthly returns of pension funds of four Pension Fund Managers (PFMs), namely Cuprum, Habitat, PlanVital and ProVida corresponding to the funds A, B, C, D and E, decreasing in terms of risk, in the period from August 2005 to December 2013.

Usage

```
data(PFM)
```

Format

A data frame with 101 observations and 21 columns including the `date` and variables `cuprum`, `habitat`, `planvital` and `provida`, each variable is related to the funds A, B, C, D, and E.

Source

Osorio, F., and Galea, M. (2015). Statistical inference in multivariate analysis using the t-distribution. Unpublished manuscript.

`rmt` *Multivariate-t random deviates*

Description

Random number generation from the multivariate-t distribution.

Usage

```
rmt(n = 1, mean = rep(0, nrow(Sigma)), Sigma = diag(length(mean)), eta = .25)
```

Arguments

`n`	the number of samples requested
`mean`	a vector giving the means of each variable
`Sigma`	a positive-definite covariance matrix
`eta`	shape parameter (must be in `[0,1/2)`). Default value is `0.25`

Details

The function `rmt` is an interface to C routines, which make calls to subroutines from LAPACK. The matrix decomposition is internally done using the Cholesky decomposition. If `Sigma` is not non-negative definite then there will be a warning message.

This parameterization of the multivariate-t includes the normal distribution as a particular case when `eta = 0`.

Value

If `n = 1` a vector of the same length as `mean`, otherwise a matrix of `n` rows of random vectors.

References

Devroye, L. (1986). *Non-Uniform Random Variate Generation*. Springer-Verlag, New York.

See Also

`rt`

Examples

```
# covariance matrix
Sigma <- matrix(c(10,3,3,2), ncol = 2)
Sigma

# generate the sample
y <- rmt(n = 1000, Sigma = Sigma)

# scatterplot of a random bivariate t sample with mean vector
# zero and covariance matrix 'Sigma'
par(pty = "s")
plot(y, xlab = "", ylab = "")
title("bivariate t sample (eta = 0.25)", font.main = 1)
```

Student *Family object for the multivariate t-distribution*

Description

Provide a convenient way to specify the details of the model used by function `studentFit`.

Usage

```
Student(eta = .25)
```

Arguments

`eta`	shape parameter for the multivariate t-distribution, must be confined to `[0,1/2)`.

Details

`Student` is a generic function to create info about the t-distribution which is passed to the estimation algorithm.

Examples

```
MyFmly <- Student(eta = .4)
MyFmly
```

studentFit *Estimation of mean and covariance using the multivariate t-distribution*

Description

Estimates the mean vector and covariance matrix assuming the data came from a multivariate t-distribution: this provides some degree of robustness to outlier without giving a high breakdown point.

Usage

```
studentFit(x, data, family = Student(eta = .25), subset, na.action, control)
```

Arguments

x	a formula or a numeric matrix or an object that can be coerced to a numeric matrix.
data	an optional data frame (or similar: see model.frame), used only if x is a formula. By default the variables are taken from environment(formula).
family	a description of the error distribution to be used in the model. By default the multivariate t-distribution with 0.25 as shape parameter is considered (using eta = 0 allows to tackle the multivariate normal distribution).
subset	an optional expression indicating the subset of the rows of data that should be used in the fitting process.
na.action	a function that indicates what should happen when the data contain NAs.
control	a list of control values for the estimation algorithm to replace the default values returned by the function mvt.control.

Value

A list with class 'studentFit' containing the following components:

call	a list containing an image of the studentFit call that produced the object.
family	the Student object used, with the estimated shape parameters (if requested).
center	final estimate of the location vector.
Scatter	final estimate of the scale matrix.
logLik	the log-likelihood at convergence.
numIter	the number of iterations used in the iterative algorithm.
weights	estimated weights corresponding to the assumed heavy-tailed distribution.
distances	estimated squared Mahalanobis distances.
eta	final estimate of the shape parameter, if requested.

Generic function print show the results of the fit.

References

Kent, J.T., Tyler, D.E., and Vardi, Y. (1994). A curious likelihood identity for the multivariate t-distribution. *Communications in Statistics: Simulation and Computation* **23**, 441-453.

Lange, K., Little, R.J.A., and Taylor, J.M.G. (1989). Robust statistical modeling using the t distribution. *Journal of the American Statistical Association* **84**, 881-896.

Osorio, F., and Galea, M. (2015). Statistical inference in multivariate analysis using the t-distribution. Unpublished manuscript.

See Also

cov, cov.rob, cov.trob

Examples

```
data(PFM)
fit <- studentFit(~ cuprum.D + habitat.D + planvital.D + provida.D, data = PFM,
  family = Student(eta = 0.25))
fit
```

`Weights` *Distribution of the weights from a multivariate t-distribution*

Description

Density, distribution function and quantile function for the weights distribution arising from the multivariate t-distribution with dimension `dim` and shape parameter `eta`.

Usage

```
dweights(x, eta = .25, dim, log = FALSE, scaled = TRUE)
pweights(q, eta = .25, dim, lower.tail = TRUE, log.p = FALSE, scaled = TRUE)
qweights(p, eta = .25, dim, lower.tail = TRUE, log.p = FALSE, scaled = TRUE)
```

Arguments

`x, q`	vector of quantiles.
`p`	vector of probabilities.
`eta`	shape parameter of the multivariate t-distribution, must be in the interval $[0, 1/2)$. Default value is `eta = 0.25`.
`dim`	dimension of the multivariate t-distribution.
`log, log.p`	logical; if TRUE, probabilities p are given as $\log(p)$.
`lower.tail`	logical; if TRUE (default), probabilities are $P(X \leq x)$, otherwise, $P(X > x)$.
`scaled`	logical; if TRUE, the weights are scaled to belong the interval `(0,1)`.

Details

The weights' distribution with parameters `eta` and `dim` $= p$ has density

$$f(x) = \frac{m^{1-(1/\eta+p)/2}}{B(1/(2\eta), p/2)} x^{1/(2\eta)-1}(m-x)^{p/2-1}$$

for $0 \leq \eta < 1/2$, $p > 0$ and $0 < x < m$, where $m = (1+p\eta)/(1-2\eta)$.

The mean is $E(X) = 1/(1-2\eta)$ and the variance is

$$Var(X) = \frac{2p\eta^2}{(1+(p+2)\eta)(1-2\eta)^2}$$

The scaled version of the weights distribution has a Beta distribution with parameters $1/(2\eta)$ and $p/2$.

Value

`dweights` gives the density, `pweights` the distribution function, and `qweights` the quantile function.

Invalid arguments will result in return value `NaN`, with a warning.

References

Abramowitz, M., and Stegun, I.A. (1972). *Handbook of Mathematical Functions*. Dover, New York. Chapter 6: Gamma and related functions.

Johnson, N.L., Kotz, S., and Balakrishnan, N. (1995). *Continuous Univariate Distributions*, volume 2, 2nd Ed. Wiley, New York. Chapter 25: Beta distributions.

Osorio, F., and Galea, M. (2015). Statistical inference in multivariate analysis using the t-distribution. Unpublished manuscript.

See Also

Distributions for other standard distributions.

`beta` for the Beta function.

Examples

```
data(companies)
fit <- studentFit(companies, family = Student(eta = .25))

# compute the 5% quantile from the estimated distribution of the weights
p <- fit$dims[2]
eta <- fit$eta
wts <- fit$weights
cutoff <- qweights(.05, eta = eta, dim = p, scaled = FALSE)

# identify observations with 'small' weights
n <- fit$dims[1]
which <- seq_len(n)[wts < cutoff]
which
```

`wilson.hilferty` *Wilson-Hilferty transformation*

Description

Returns the Wilson-Hilferty transformation of random variables with F distribution.

Usage

```
wilson.hilferty(x, center, cov, eta = 0)
```

Arguments

`x`	object of class `'studentFit'` from which is extracted the estimated Mahalanobis distances of the fitted model. Also `x` can be a vector or matrix of data with, say, p columns.
`center`	mean vector of the distribution or second data vector of length p. Not required if `x` have class `'studentFit'`.
`cov`	covariance matrix (p by p) of the distribution. Not required if `x` have class `'studentFit'`.
`eta`	shape parameter of the multivariate t-distribution. By default the multivariate normal (`eta = 0`) is considered.

Details

Let F the following random variable:

$$F = \frac{D^2/p}{1-2\eta}$$

where D^2 denotes the squared Mahalanobis distance defined as

$$D^2 = (x-\mu)^T\Sigma^{-1}(x-\mu)$$

Thus the Wilson-Hilferty transformation is given by

$$z = \frac{(1-\frac{2\eta}{9})F^{1/3} - (1-\frac{2}{9p})}{(\frac{2\eta}{9}F^{2/3} + \frac{2}{9p})^{1/2}}$$

and z is approximately distributed as a standard normal distribution. This is useful, for instance, in the construction of QQ-plots.

For `eta = 0`, we obtain

$$z = \frac{F^{1/3} - (1-\frac{2}{9p})}{(\frac{2}{9p})^{1/2}}$$

which is the Wilson-Hilferty transformation for chi-square variables.

References

Osorio, F., and Galea, M. (2015). Statistical inference in multivariate analysis using the t-distribution. Unpublished manuscript.

Wilson, E.B., and Hilferty, M.M. (1931). The distribution of chi-square. *Proceedings of the National Academy of Sciences of the United States of America* **17**, 684-688.

See Also

cov, mahalanobis, envelope

Examples

```
data(companies)
x <- companies
z <- wilson.hilferty(x, center = colMeans(x), cov = cov(x))
par(pty = "s")
qqnorm(z, main = "Transformed distances Q-Q plot")
abline(c(0,1), col = "red", lwd = 2)
```

Index