Package 'norm'

December 22, 2014

Version 1.0-9.5

Date 2013/02/27

Title Analysis of multivariate normal datasets with missing values

Author Ported to R by Alvaro A. Novo <alvaro@novo-online.net>.
Original by Joseph L. Schafer <jls@stat.psu.edu>.

Maintainer John Fox <jfox@mcmaster.ca>

Description Analysis of multivariate normal datasets with missing values

License file LICENSE

URL http://www.stat.psu.edu/~jls/misoftwa.html#aut

Repository CRAN

Repository/R-Forge/Project norm

Repository/R-Forge/Revision 8

Repository/R-Forge/DateTimeStamp 2013-02-27 16:01:38

Date/Publication 2013-02-28 07:11:32

NeedsCompilation yes

License_restricts_use no

R topics documented:

.code.to.na 2
.na.to.snglcode 2
da.norm 3
em.norm 4
getparam.norm 6
imp.norm 7
loglik.norm 8
logpost.norm 9
makeparam.norm 10
mda.norm 11
mdata 12

mi.inference 12
ninvwish 13
prelim.norm 14
rngseed 15

Index **17**

`.code.to.na` *Changes missing value code to NA*

Description

Changes missing value code to NA. It's called from *'prelim.norm'*.

Usage

```
.code.to.na(x, mvcode)
```

Arguments

`x`	data object.
`mvcode`	internal input of 'prelim.norm'.

Value

Initial data object with missing values code changed to NA.

See Also

`prelim.norm`

`.na.to.snglcode` *Changes NA's to single precision missing value code*

Description

Changes NA's to single precision missing value code It's called internally by other functions in the package, e.g., *'prelim.norm'*.

Usage

```
.na.to.snglcode(x, mvcode)
```

Arguments

`x`	data object.
`mvcode`	internal input of 'prelim.norm'.

Value

Initial data object with missing values code precision changed to sinlge.

See Also

prelim.norm

---

da.norm *Data augmentation for incomplete multivariate normal data*

Description

Data augmentation under a normal-inverted Wishart prior. If no prior is specified by the user, the usual "noninformative" prior for the multivariate normal distribution is used. This function simulates one or more iterations of a single Markov chain. Each iteration consists of a random imputation of the missing data given the observed data and the current parameter value (I-step), followed by a draw from the posterior distribution of the parameter given the observed data and the imputed data (P-step).

Usage

```
da.norm(s, start, prior, steps=1, showits=FALSE, return.ymis=FALSE)
```

Arguments

s	summary list of an incomplete normal data matrix produced by the function prelim.norm.
start	starting value of the parameter. This is a parameter vector in packed storage, such as one created by the function makeparam.norm. One obvious choice for a starting value is an ML estimate or posterior mode produced by em.norm.
prior	optional prior distribution. This is a list containing the hyperparameters of a normal-inverted Wishart distribution. In order, the elements of the list are: tau (a scalar), m (a scalar), mu0 (a vector of length ncol(x), where x is the original matrix of incomplete data), and lambdainv (a matrix of dimension c(ncol(x),ncol(x))). The elements of mu0 and lambdainv apply to the data after transformation, i.e. after the columns have been centered and scaled to have mean zero and variance one. If no prior is supplied, the default is the usual noninformative prior for a multivariate normal model: tau=0, m=-1, mu0=arbitrary, and lambdainv = matrix of zeros.
steps	number of data augmentation iterations to be simulated.
showits	if TRUE, reports the iterations so the user can monitor the progress of the algorithm.
return.ymis	if TRUE, returns the output of the last I-step (imputed values of missing data) in addition to the output of the last P-step. These imputed values are useful for forming Rao-Blackwellized estimates of posterior summaries.

Value

if `return.ymis=FALSE`, returns a parameter vector, the result of the last P-step. If the value of `steps` was large enough to guarantee approximate stationarity, then this parameter can be regarded as a proper draw from the observed-data posterior, independent of `start`. If `return.ymis=TRUE`, then this function returns a list of the following two components:

`parameter`	a parameter vector, the result of the last P-step
`ymis`	a vector of missing values, the result of the last I-step. The length of this vector is `sum(is.na(x))`, where x is the original data matrix. The storage order is the same as that of `x[is.na(x)]`.

WARNING

Before this function may be used, the random number generator seed must be initialized with `rngseed` at least once in the current S session.

References

See Chapter 5 of Schafer (1996).

See Also

`rngseed`, `em.norm`, `prelim.norm`, and `getparam.norm`.

Examples

```
data(mdata)
s  <-  prelim.norm(mdata)
thetahat <- em.norm(s)   #find the MLE for a starting value
rngseed(1234567)   #set random number generator seed
theta <- da.norm(s,thetahat,steps=20,showits=TRUE)  # take 20 steps
getparam.norm(s,theta) # look at result
```

---

em.norm *EM algorithm for incomplete normal data*

Description

Performs maximum-likelihood estimation on the matrix of incomplete data using the EM algorithm. Can also be used to find a posterior mode under a normal-inverted Wishart prior supplied by the user.

Usage

```
em.norm(s, start, showits=TRUE, maxits=1000, criterion=0.0001, prior)
```

Arguments

s	summary list of an incomplete normal data matrix produced by the function prelim.norm.
start	optional starting value of the parameter. This is a parameter vector in packed storage, such as one created by the function makeparam.norm. If no starting value is supplied, em.norm chooses its own starting value.
showits	if TRUE, reports the iterations of EM so the user can monitor the progress of the algorithm.
maxits	maximum number of iterations performed. The algorithm will stop if the parameter still has not converged after this many iterations.
criterion	convergence criterion. The algorithm stops when the maximum relative difference in all of the estimated means, variances, or covariances from one iteration to the next is less than or equal to this value.
prior	optional prior distribution. This is a list containing the hyperparameters of a normal-inverted Wishart distribution. In order, the elements of the list are: tau (a scalar), m (a scalar), mu0 (a vector of length ncol(x)), and lambdainv (a matrix of dimension c(ncol(x),ncol(x))). The elements of mu0 ans lambdainv apply to the data after transformation, i.e. after the columns have been centered and scaled to have mean zero and variance one. If no prior is supplied, the default is a uniform prior, which results in maximum-likelihood estimation.

Details

The default starting value takes all means on the transformed scale to be equal to zero, and covariance matrix on the transformed scale equal to the identity. All important computations are carried out in double precision, using the sweep operator.

Value

a vector representing the maximum-likelihood estimates of the normal parameters. This vector contains means, variances, and covariances on the transformed scale in packed storage. The parameter can be transformed back to the original scale and put into a more understandable format by the function getparam.norm.

References

See Section 5.3 of Schafer (1994).

See Also

prelim.norm, getparam.norm, and makeparam.norm.

Examples

```
data(mdata)
s <- prelim.norm(mdata)   #do preliminary manipulations
thetahat <- em.norm(s)   #compute mle
getparam.norm(s,thetahat,corr=TRUE)$r #look at estimated correlations
```

getparam.norm *Extract normal parameters from packed storage*

Description

Takes a parameter vector, such as one produced by em.norm or da.norm, and returns a list of parameters on the original scale.

Usage

```
getparam.norm(s, theta, corr=FALSE)
```

Arguments

s	summary list of an incomplete normal data matrix created by the function prelim.norm.
theta	vector of normal parameters expressed on transformed scale in packed storage, such as one produced by the function em.norm.
corr	if TRUE, computes means, standard deviations, and a correlation matrix. If FALSE, computes means and a covariance matrix.

Value

if corr=FALSE, a list containing the components mu and sigma; if corr=TRUE, a list containing the components mu, sdv, and r. The components are:

mu	vector of means. Elements are in the same order and on the same scale as the columns of the original data matrix, and with names corresponding to the column names of the original data matrix.
sigma	matrix of variances and covariances.
sdv	vector of standard deviations.
r	matrix of correlations.

See Also

prelim.norm and makeparam.norm.

Examples

```
data(mdata)
s <- prelim.norm(mdata)   #do preliminary manipulations
thetahat <- em.norm(s)   #compute MLE
getparam.norm(s,thetahat,corr=TRUE)$r #look at estimated correlations
```

`imp.norm` *Impute missing multivariate normal data*

Description

Draws missing elements of a data matrix under the multivariate normal model and a user-supplied parameter

Usage

```
imp.norm(s, theta, x)
```

Arguments

`s`	summary list of an incomplete normal data matrix `x` created by the function `prelim.norm`.
`theta`	value of the normal parameter under which the missing data are to be randomly imputed. This is a parameter vector in packed storage, such as one created by `em.norm` or `da.norm`.
`x`	the original data matrix used to create the summary list `s`. If this argument is not supplied, then the data matrix returned by this function may disagree slightly with the observed values in `x` due to rounding errors.

Details

This function simply performs one I-step of data augmentation.

Value

a matrix of the same form as `x`, but with all missing values filled in with simulated values drawn from their predictive distribution given the observed data and the specified parameter.

WARNING

Before this function may be used, the random number generator seed must be initialized with `rngseed` at least once in the current S session.

References

See Section 5.4.1 of Schafer (1996).

See Also

`prelim.norm`, `makeparam.norm`, and `rngseed`.

Examples

```
data(mdata)
s <- prelim.norm(mdata)   #do preliminary manipulations
thetahat <- em.norm(s)   #find the mle
rngseed(1234567)   #set random number generator seed
ximp <- imp.norm(s,thetahat,mdata)  #impute missing data under the MLE
```

---

loglik.norm *Observed-data loglikelihood for normal data*

Description

Evaluates the observed-data loglikelihood function at a user-supplied value of the parameter. This function is useful for monitoring the progress of EM and data augmentation.

Usage

```
loglik.norm(s, theta)
```

Arguments

s	summary list of an incomplete normal data matrix created by the function `prelim.norm`.
theta	vector of normal parameters expressed on transformed scale in packed storage, such as one produced by the function `em.norm`.

Value

value of the observed-data loglikelihood

References

See Section 5.3.5 of Schafer (1996)

See Also

`prelim.norm` and `logpost.norm`

Examples

```
data(mdata)
s <- prelim.norm(mdata)   #do preliminary manipulations
thetahat <- em.norm(s)   #compute MLE
loglik.norm(s,thetahat)  #loglikelihood at the MLE
```

logpost.norm *Observed-data log-posterior for normal data*

Description

Evaluates the log of the observed-data posterior density at a user-supplied value of the parameter. Assumes a normal-inverted Wishart prior. This function is useful for monitoring the progress of EM and data augmentation.

Usage

```
logpost.norm(s, theta, prior)
```

Arguments

s	summary list of an incomplete normal data matrix created by the function prelim.norm.
theta	vector of normal parameters expressed on transformed scale in packed storage, such as one produced by the function em.norm.
prior	optional prior distribution. This is a list containing the hyperparameters of a normal-inverted Wishart distribution. In order, the elements of the list are: tau (a scalar), m (a scalar), mu0 (a vector of length ncol(x), where x is the original matrix of incomplete data), and lambdainv (a matrix of dimension c(ncol(x),ncol(x))). The elements of mu0 and lambdainv apply to the data after transformation, i.e. after the columns have been centered and scaled to have mean zero and variance one. If no prior is supplied, the default is the usual noninformative prior for a multivariate normal model: tau=0, m=-1, mu0=arbitrary, and lambdainv = matrix of zeros.

Value

value of the observed-data log-posterior density

References

See Section 5.3.5 of Schafer (1996)

See Also

prelim.norm and loglik.norm

Examples

```
data(mdata)
s <- prelim.norm(mdata)   #do preliminary manipulations
prior <- list(0,.5,rep(0,ncol(mdata)),
   .5*diag(rep(1,ncol(mdata)))) #ridge prior with .5 df
thetahat <- em.norm(s,prior=prior)   #compute posterior mode
logpost.norm(s,thetahat,prior)  #log-posterior at mode
```

makeparam.norm *Convert normal parameters to packed storage*

Description

Does the opposite of getparam.norm. Converts a list of user-specified parameters to a parameter vector suitable for input to functions such as da.norm and em.norm.

Usage

```
makeparam.norm(s, thetalist)
```

Arguments

s	summary list of an incomplete normal data matrix created by the function prelim.norm.
thetalist	list of normal parameters of the same form as one produced by getparam.norm. If the list has two components, the first must be the vector of means and the second must be the covariance matrix, where means and covariances are expressed on the scale of the original data. If the list has three components, the first must be the vector of means, the second must be the vector of standard deviations, and the third must be the correlation matrix.

Value

normal parameter in packed storage, suitable for use as a starting value for em.norm, mda.norm, or mdamet.norm.

See Also

prelim.norm and getparam.norm.

Examples

```
data(mdata)
s <- prelim.norm(mdata)   #do preliminary manipulations
thetahat <- em.norm(s)   #compute mle
thetahat <- getparam.norm(s,thetahat,corr=TRUE)   #extract parameters
thetahat$r   #look at mle correlations
thetahat$r[1,2] <- .5   #tweak a parameter
thetahat <- makeparam.norm(s,thetahat)   #convert to packed storage
thetahat <- em.norm(s,thetahat) #run EM again from new starting value
```

mda.norm *Monotone data augmentation for incomplete multivariate normal data*

Description

Monotone data augmentation under the usual noninformative prior, as described in Chapter 6 of Schafer (1996). This function simulates one or more iterations of a single Markov chain. One iteration consists of a random imputation of the missing data given the observed data and the current parameter value (I-step), followed by a draw from the posterior distribution of the parameter given the observed data and the imputed data (P-step). The I-step imputes only enough data to complete a monotone pattern, which typically makes this algorithm converge more quickly than da.norm, particularly when the observed data are nearly monotone. The order of the variables in the original data matrix determines the monotone pattern to be completed. For fast convergence, it helps to order the variables according to their rates of missingness, with the most observed (least missing) variable on the left and the least observed variable on the right.

Usage

```
mda.norm(s, theta, steps=1, showits=FALSE)
```

Arguments

s	summary list of an incomplete normal data matrix produced by the function prelim.norm.
theta	starting value of the parameter. This is a parameter vector in packed storage, such as one created by the function makeparam.norm. One obvious choice for a starting value is an ML estimate or posterior mode produced by em.norm.
steps	number of monotone data augmentation iterations to be simulated.
showits	if TRUE, reports the iterations so the user can monitor the progress of the algorithm.

Value

Returns a parameter vector, the result of the last P-step. If the value of steps was large enough to guarantee approximate stationarity, then this parameter can be regarded as a proper draw from the observed-data posterior, independent of start.

WARNING

Before this function may be used, the random number generator seed must be initialized with rngseed at least once in the current S session.

References

Chapter 6 of Schafer (1996).

See Also

rngseed, em.norm, prelim.norm, and getparam.norm.

Examples

```
data(mdata)
s <- prelim.norm(mdata)
thetahat <- em.norm(s)   #find the MLE for a starting value
rngseed(1234567)   #set random number generator seed
theta <- mda.norm(s,thetahat,steps=20,showits=TRUE)  # take 20 steps
getparam.norm(s,theta) # look at result
```

---

mdata *Dataset with missing values to illustrate use of package norm*

Description

Household survey with missing values. See Schafer~(1997).

References

Schafer, J.L.(1997) *Analysis of Incomplete Multivariate Data*, Chapman & Hall, London. ISBN: 0412040611

---

mi.inference *Multiple imputation inference*

Description

Combines estimates and standard errors from m complete-data analyses performed on m imputed datasets to produce a single inference. Uses the technique described by Rubin (1987) for multiple imputation inference for a scalar estimand.

Usage

```
mi.inference(est, std.err, confidence=0.95)
```

Arguments

est	a list of m (at least 2) vectors representing estimates (e.g., vectors of estimated regression coefficients) from complete-data analyses performed on m imputed datasets.
std.err	a list of m vectors containing standard errors from the complete-data analyses corresponding to the estimates in est.
confidence	desired coverage of interval estimates.

Value

a list with the following components, each of which is a vector of the same length as the components of `est` and `std.err`:

`est`	the average of the complete-data estimates.
`std.err`	standard errors incorporating both the between and the within-imputation uncertainty (the square root of the "total variance").
`df`	degrees of freedom associated with the t reference distribution used for interval estimates.
`signif`	P-values for the two-tailed hypothesis tests that the estimated quantities are equal to zero.
`lower`	lower limits of the (100*confidence)% interval estimates.
`upper`	upper limits of the (100*confidence)% interval estimates.
`r`	estimated relative increases in variance due to nonresponse.
`fminf`	estimated fractions of missing information.

METHOD

Uses the method described on pp. 76-77 of Rubin (1987) for combining the complete-data estimates from m imputed datasets for a scalar estimand. Significance levels and interval estimates are approximately valid for each one-dimensional estimand, not for all of them jointly.

References

See Rubin (1987) or Schafer (1996), Chapter 4.

`ninvwish` *Random normal-inverted Wishart variate*

Description

Simulates a value from a normal-inverted Wishart distribution. This function may be useful for obtaining starting values of the parameters of a multivariate normal distribution for multiple chains of data augmentation.

Usage

```
ninvwish(s, params)
```

Arguments

s	summary list of an incomplete normal data matrix produced by the function prelim.norm.
params	list of parameters of a normal-inverted Wishart distribution. In order, the elements of the list are: tau (a scalar), m (a scalar), mu0 (a vector of length ncol(x)), and lambdainv (a matrix of dimension c(ncol(x),ncol(x))). When using this function to create starting values for data augmentation, mu0 and lambdainv should be chosen in relation to the data matrix after the columns have been centered and scaled to have mean zero and variance one.

Value

a vector in packed storage representing the simulated normal-inverted Wishart variate. This vector has the same form as parameter vectors produced by functions such as em.norm and da.norm, and may be used directly as a starting value for these functions. This vector can also be put into a more understandable format by getparam.norm.

WARNING

Before this function may be used, the random number generator seed must be initialized with rngseed at least once in the current S session.

References

See Section 5.4.2 of Schafer (1996).

See Also

rngseed, getparam.norm, em.norm and da.norm.

Examples

```
data(mdata)
s <- prelim.norm(mdata)   #do preliminary manipulations
params <- list(1,.5,rep(0,ncol(mdata)), .5*diag(rep(1,ncol(mdata)))) # gives widely dispersed values
rngseed(1234567)
start <- ninvwish(s,params)  # draw a variate
thetahat <- em.norm(s,start=start) # run EM from this starting value
```

---

prelim.norm *Preliminary manipulations for a matrix of incomplete continuous data.*

Description

Sorts rows of x by missingness patterns, and centers/scales columns of x. Calculates various bookkeeping quantities needed for input to other functions, such as em.norm and da.norm.

Usage

```
prelim.norm(x)
```

Arguments

x	data matrix containing missing values. The rows of x correspond to observational units, and the columns to variables. Missing values are denoted by NA.

Value

a list of thirteen components that summarize various features of x after the data have been centered, scaled, and sorted by missingness patterns. Components that might be of interest to the user include:

nmis	a vector of length ncol(x) containing the number of missing values for each variable in x. This vector has names that correspond to the column names of x, if any.
r	matrix of response indicators showing the missing data patterns in x. Dimension is (S,p) where S is the number of distinct missingness patterns in the rows of x, and p is the number of columns in x. Observed values are indicated by 1 and missing values by 0. The row names give the number of observations in each pattern, and the column names correspond to the column names of x.

References

See Section 5.3.1 of Schafer (1996).

Examples

```
data(mdata)
s <- prelim.norm(mdata)  #do preliminary manipulations
s$nmis[s$co] #look at nmis
s$r #look at missing data patterns
```

---

rngseed *Initialize random number generator seed*

Description

Initializes the seed value for the internal random-number generator used in missing-data programs

Usage

```
rngseed(seed)
```

Arguments

seed	a positive number > = 1, preferably a large integer.

Value

NULL.

NOTE

The random number generator seed must be set at least once by this function before the simulation or imputation functions in this package (da.norm, etc.) can be used.

Index

∗Topic **NA**
.code.to.na, 2
.na.to.snglcode, 2
∗Topic **datasets**
mdata, 12
∗Topic **distribution**
da.norm, 3
rngseed, 15
∗Topic **htest**
mi.inference, 12
∗Topic **multivariate**
loglik.norm, 8
logpost.norm, 9
mda.norm, 11
ninvwish, 13
prelim.norm, 14
∗Topic **regression**
em.norm, 4
getparam.norm, 6
imp.norm, 7
makeparam.norm, 10
∗Topic
.code.to.na, 2
.na.to.snglcode, 2
mdata, 12
mi.inference, 12
rngseed, 15
.code.to.na, 2
.na.to.snglcode, 2

da.norm, 3, *14*

em.norm, *4*, 4, *12*, *14*

getparam.norm, *4, 5*, 6, *10*, *12*, *14*

imp.norm, 7

loglik.norm, 8, *9*
logpost.norm, *8*, 9

makeparam.norm, *5–7*, 10
mda.norm, 11
mdata, 12
mi.inference, 12

ninvwish, 13

prelim.norm, *2–10*, *12*, 14

rngseed, *4*, *7*, *12*, *14*, 15