Package ‘estimability’

September 3, 2015

Type Package
Title Tools for Assessing Estimability of Linear Predictions
Version 1.1-1
Date 2015-09-02
Author Russell V. Lenth
Maintainer Russell V. Lenth <russell-lenth@uiowa.edu>
Depends stats
Description Provides tools for determining estimability of linear functions of regression coefficients, and 'epredict' methods that handle non-estimable cases correctly.
LazyData Yes
License GPL-2
NeedsCompilation no
Repository CRAN
Date/Publication 2015-09-03 10:21:26

R topics documented:

estimability-package ........................................... 1
epredict ......................................................... 2
nonest.basis ................................................... 5

Index

estimability-package  Estimability Tools for Linear Models

Description

Provides tools for determining estimability of linear functions of regression coefficients, and alternative 'epredict' methods for 'lm', 'glm', and 'mlm' objects that handle non-estimable cases correctly.

Details
When a linear model is not of full rank, the regression coefficients are not uniquely estimable. However, the predicted values are unique, as are other linear combinations where the coefficients lie in the row space of the data matrix. Thus, estimability of a linear function of regression coefficients can be determined by testing whether the coefficients lie in this row space – or equivalently, are orthogonal to the corresponding null space. This package provides functions \texttt{nonest.basis} and \texttt{is.estible} to facilitate such an estimability test. Package developers may find these useful for incorporating in their \texttt{predict} methods when new predictor settings are involved.

The package also provides \texttt{epredict} methods – alternatives to the \texttt{predict} methods in the \texttt{stats} package for "\texttt{lm}", "\texttt{glm}", and "\texttt{mlm}" objects. When the \texttt{newdata} argument is specified, estimability of each new prediction is checked and any non-estimable cases are replaced by \texttt{NA}.

\section*{Author(s)}
Russell V. Lenth <russell-lenth@uiowa.edu>

\section*{References}
Monahan, John F. (2008) \textit{A Primer on Linear Models}, CRC Press. (Chapter 3)

\section*{Description}
These functions call the corresponding S3 \texttt{predict} methods in the \texttt{stats} package, but with a check for estimability of new predictions, and with appropriate actions for non-estimable cases.

\section*{Usage}
\begin{verbatim}
## S3 method for class 'lm'
epredict(object, newdata, ...,
    type = c("response", "terms", "matrix", "estimability"),
    nonest.tol = 1e-8, nbasis = object$nonest)

## S3 method for class 'glm'
epredict(object, newdata, ...,
    type = c("link", "response", "terms", "matrix", "estimability"),
    nonest.tol = 1e-8, nbasis = object$nonest)

## S3 method for class 'mlm'
epredict(object, newdata, ...,
    type = c("response", "matrix", "estimability"),
    nonest.tol = 1e-8, nbasis = object$nonest)

eupdate(object, ...)
\end{verbatim}
### epredict

#### Arguments

- **object**: An object inheriting from `lm`
- **newdata**: A `data.frame` containing predictor combinations for new predictions
- **...**: Arguments passed to `predict` or `update`
- **nonest.tol**: Tolerance used by `is.esttbl` to check estimability of new predictions
- **type**: Character string specifying the desired result. See Details.
- **nbasis**: Basis for the null space, e.g., a result of a call to `nonest.basis`. If `nbasis` is `NULL`, a basis is constructed from `object`.

#### Details

If `newdata` is missing or `object` is not rank-deficient, this method passes its arguments directly to the same method in the `stats` library. In rank-deficient cases with `newdata` provided, each row of `newdata` is tested for estimability against the null basis provided in `nbasis`. Any non-estimable cases found are replaced with `NA`s.

The `type` argument is passed to `predict` when it is one of "response", "link", or "terms". With `newdata` present and `type = "matrix"`, the model matrix for `newdata` is returned, with an attribute "establ" that is a logical vector of length `nrow(newdata)` indicating whether each row is estimable. With `type = "estimability"`, just the logical vector is returned.

If you anticipate making several `epredict` calls with new data, it improves efficiency to either obtain the null basis and provide it in the call, or add it to `object` with the name "nonest" (perhaps via a call to `eupdate`).

`eupdate` is an S3 generic function with a method provided for "lm" objects. It updates the object according to any arguments in ..., then obtains the updated object's nonestimable basis and returns it in `object$nonest`.

#### Value

The same as the result of a call to the `predict` method in the `stats` package, except rows or elements corresponding to non-estimable predictor combinations are set to `NA`. The value for `type` is "matrix" or "estimability" is explained under details.

#### Note

The usual rank-deficiency warning from `stats::predict` is suppressed; but when non-estimable cases are found, a message is displayed explaining that these results were replaced by `NA`. If you wish that message suppressed, use `options(estimability.quiet = TRUE)`.

#### Author(s)

Russell V. Lenth <russell-lenth@uiowa.edu>

#### See Also

`predict.lm` in the `stats` package; `nonest.basis`.
Examples

```r
require("estimability")

# Fake data where x3 and x4 depend on x1, x2, and intercept
x1 <- -4:4
x2 <- c(-2,1,-1,2,0,2,-1,1,-2)
x3 <- 3*x1 - 2*x2
x4 <- x2 - x1 + 4
y <- 1 + x1 + x2 + x3 + x4 + c(-.5,.5,-.5,.5,-.5,.5,-.5,.5)

# Different orderings of predictors produce different solutions
mod1234 <- lm(y ~ x1 + x2 + x3 + x4)
mod4321 <- update(lm(y ~ x4 + x3 + x2 + x1))
# (Estimability checking with mod4321 will be more efficient because
# it will not need to recreate the basis)
mod4321$nonest

# test data:
testset <- data.frame(
  x1 = c(3, 6, 6, 0, 0, 1),
  x2 = c(1, 2, 2, 0, 0, 2),
  x3 = c(7, 14, 14, 0, 0, 3),
  x4 = c(2, 4, 0, 4, 0, 4))

# Look at predictions when we don't check estimability
suppressWarnings(# Disable the warning from stats::predict.lm
  rbind(p1234 = predict(mod1234, newdata = testset),
        p4321 = predict(mod4321, newdata = testset)))

# Compare with results when we do check:
  rbind(p1234 = epredict(mod1234, newdata = testset),
        p4321 = epredict(mod4321, newdata = testset))

# Note that estimable cases have the same predictions

# change mod1234 and include nonest basis
mod134 <- update(mod1234, . ~ . - x2, subset = -c(3, 7))
mod134$nonest

# When row spaces are the same, bases are interchangeable
# so long as you account for the ordering of parameters:
epredict(mod4321, newdata = testset, type = "estimability",
         nbasis = nonest.basis(mod1234)[c(1, 5:2), ])

## Not run:
### Additional illustration
example(nonest.basis) ## creates model objects warp.lm1 and warp.lm2

# The two models have different contrast specs. But the empty cell
# is correctly identified in both:
faccmb <- expand.grid(wool = c("A", "B"), tension = c("L", "M", "H"))
```
\begin{verbatim}
cbind(fac.cmb, 
    pred1 = epredict(warp.lm1, newdata = fac.cmb), 
    pred2 = epredict(warp.lm2, newdata = fac.cmb))

## End(Not run)
\end{verbatim}

---

**nonest.basis**  

*Estimability Tools*

**Description**

This documents the functions needed to test estimability of linear functions of regression coefficients.

**Usage**

\begin{verbatim}
nonest.basis(x, ...)
## S3 method for class 'qr'
nonest.basis(x, ...)
## S3 method for class 'matrix'
nonest.basis(x, ...)
## S3 method for class 'lm'
nonest.basis(x, ...)

all.estble

is.estble(x, nbasis, tol = 1e-8)
\end{verbatim}

**Arguments**

- `x` For `nonest.basis`, an object of a class in `methods("nonest.basis")`. Or, in `is.estble`, a numeric vector or matrix for assessing estimability of `sum(x * beta)`, where `beta` is the vector of regression coefficients.
- `nbasis` Matrix whose columns span the null space of the model matrix. Such a matrix is returned by `nonest.basis`.
- `tol` Numeric tolerance for assessing nonestimability. For nonzero `x`, estimability of \( \beta'x \) is assessed by whether or not \( ||N'x||^2 < \tau||x'||^2 \), where \( N \) and \( \tau \) denote `nbasis` and `tol`, respectively.
- `...` Additional arguments, currently ignored.

**Details**

Consider a linear model \( y = X\beta + E \). If \( X \) is not of full rank, it is not possible to estimate \( \beta \) uniquely. However, \( X\beta \) is uniquely estimable, and so is \( a'X\beta \) for any conformable vector \( a \). Since \( a'X \) comprises a linear combination of the rows of \( X \), it follows that we can estimate any linear
function where the coefficients lie in the row space of $X$. Equivalently, we can check to ensure that
the coefficients are orthogonal to the null space of $X$.

The constant \texttt{all.estble} is simply a 1 x 1 matrix of NA. This specifies a trivial non-estimability
basis, and using it as \texttt{nbasis} will cause everything to test as estimable.

**Value**

When $X$ is not full-rank, the methods for \texttt{nonest.basis} return a basis for the null space of $X$.
The number of rows is equal to the number of regression coefficients (including any NAs); and the
number of columns is equal to the rank deficiency of the model matrix. The columns are orthonormal. If
the model is full-rank, then \texttt{nonest.basis} returns \texttt{all.estble}. The matrix method uses $X$ itself, the \texttt{qr}
method uses the $QR$ decomposition of $X$, and the \texttt{lm} method recovers the required
information from the object.

The function \texttt{is.estble} returns a logical value (or vector, if \texttt{x} is a matrix) that is \texttt{TRUE} if the
function is estimable and \texttt{FALSE} if not.

**Author(s)**

Russell V. Lenth <russell-lenth@uiowa.edu>

**References**

Monahan, John F. (2008) \textit{A Primer on Linear Models}, CRC Press. (Chapter 3)

**Examples**

```r
require(estimability)
X <- cbind(rep(1,5), 1:5, 5:1, 2:6)
(nb <- nonest.basis(X))

# Test estimability of some linear functions for this X matrix
lfs <- rbind(c(1,4,2,5), c(2,3,9,5), c(1,2,2,1), c(0,1,-1,1))
is.estble(lfs, nb)

# Illustration on 'lm' objects:
warp.lm1 <- lm(breaks ~ wool * tension, data = warpbreaks,
              subset = -(26:38),
              contrasts = list(wool = "contr.treatment", tension = "contr.treatment"))
zapsmall(nonest.basis(warp.lm1))

warp.lm2 <- update(warp.lm1,
              contrasts = list(wool = "contr.sum", tension = "contr.helmert"))
zapsmall(nonest.basis(warp.lm2))
```
Index

*Topic models
  epredict, 2
  estimability-package, 1
  nonest.basis, 5
*Topic package
  estimability-package, 1
*Topic regression
  epredict, 2
  estimability-package, 1
  nonest.basis, 5

all.estble (nonest.basis), 5

epredict, 2, 2
estimability (estimability-package), 1
estimability-package, 1
eupdate (epredict), 2

is.estble, 2, 3
is.estble (nonest.basis), 5

nonest.basis, 2, 3, 5

predict, 2, 3
predict.lm, 3

update, 3