Package ‘imputeYn’

October 13, 2022

Type Package

Title Imputing the Last Largest Censored Observation(s) Under Weighted Least Squares

Version 1.3

Date 2015-10-23

Author Hasinur Rahaman Khan and Ewart Shaw

Maintainer Hasinur Rahaman Khan <hasinurkhan@gmail.com>

Description Method brings less bias and more efficient estimates for AFT models.

Depends quadprog, emplik, mvtnorm, survival, boot

License GPL-2

NeedsCompilation no

Repository CRAN

Date/Publication 2015-10-23 22:08:14

R topics documented:

`imputeYn-package` *Imputing the Last Largest Censored Observation(s) Under Weighted Least Squares*

Description

Method brings less bias and more efficient estimates for AFT models.

Details

Package:	imputeYn
Type:	Package
Version:	1.3
Date:	2015-10-23
License:	GPL
Depends:	emplik, mvtnorm, quadprog, survival, boot

Author(s)

Hasinur Rahaman Khan and Ewart Shaw Maintainer: Hasinur Rahaman Khan <hasinurkhan@gmail.com>

References

Efron, B. (1967). The two sample problem with censored data. In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, Vol. 4, p. 831-853.

Jin et al. (2006). On least-squares regression with censored data. Biometrika, 93 (1), 147-161.

Khan and Shaw. (2013a). On Dealing with Censored Largest Observations under Weighted Least Squares. CRiSM working paper, Department of Statistics, University of Warwick, UK, No. 13-07. Also available in http://arxiv.org/abs/1312.2533.

Khan and Shaw (2013b). Variable Selection with The Modified Buckley-James Method and The Dantzig Selector for High-dimensional Survival Data. Proceedings 59th ISI World Statistics Congress, 25-30 August 2013, Hong Kong, p. 4239-4244.

Stute, W. (1993). Consistent estimation under random censorship when covariables are available. Journal of Multivariate Analysis, 45 , 89-103.

Examples

```
#For uncorrelated dataset
data1<-data(n=100, p=4, r=0, b1=c(2,2,3,3), sig=1, Cper=0)
imp<-imputeYn(data1$x, data1$y, data1$delta, method = "condMean", beta=NULL)
imp
```

aft.kmweight — *Computing Kaplan-Meier Weights*

Description

Compute Kaplan-Meier weights for weighted least squares method.

Usage

```
aft.kmweight(Y, delta)
```

Arguments

`Y`	survival time.
`delta`	status.

Details

Compute Kaplan-Meier weights that are used for weighted least squares to solve the AFT model under right censoring. This gives weights that are computed after implementation of Efron's (1967) tail correction.

Value

The Kaplan-Meier weights are proper in the sense that they sum one. The censoring considered here is right censoring only.

`kmwt`	The Kaplan Meier weights

Author(s)

Hasinur Rahaman Khan and Ewart Shaw

References

Stute, W. (1993). Consistent estimation under random censorship when covariables are available. Journal of Multivariate Analysis, 45 , 89-103.

Efron, B. (1967). The two sample problem with censored data. In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, Vol. 4, p. 831-853.

Examples

```
# For dataset where the last largest datum is censored and censoring level is 50 percent
data1<-data(n=100, p=2, r=0, b1=c(2,4), sig=1, Cper=0)
kmw<-aft.kmweight(data1$y,data1$delta)
kmw
```

`aft.qp` *Fiting Regularized Weighted Least Squares Method by Quadratic Programming*

Description

The AFT model is solved by quadratic programming.

Usage

```
aft.qp(X, Y, delta)
```

Arguments

`X`	the covariate matrix of size n by p.
`Y`	survival time.
`delta`	status.

Details

The AFT model is solved using weighted least squares with ridge penalty and censoring constraints (Khan and Shaw, 2013b). The optimization is based on quadratic programming that facilitates incorporating additional censoring constraints into the model. The ridge penalty used here is 0.01 sqrt(2log(p))

Value

Only gives the parameter estimates as below

`beta0`	the intercept of the AFT model
`beta`	the coefficients of the covariates

Author(s)

Hasinur Rahaman Khan and Ewart Shaw

References

Khan and Shaw. (2013a). On Dealing with Censored Largest Observations under Weighted Least Squares. CRiSM working paper, Department of Statistics, University of Warwick, UK, No. 13-07. Also available in http://arxiv.org/abs/1312.2533.

Khan and Shaw (2013b). Variable Selection with The Modified Buckley-James Method and The Dantzig Selector for High-dimensional Survival Data. Proceedings 59th ISI World Statistics Congress, 25-30 August 2013, Hong Kong, p. 4239-4244.

See Also

imputeYn

Examples

```
# For uncorrelated dataset
data1<-data(n=100, p=2, r=0, b1=c(2,4), sig=1, Cper=0)
fit<-aft.qp(data1$x, data1$y, data1$delta)
fit

# For correlated dataset
data2<-data(n=100, p=2, r=0.5, b1=c(2,4), sig=1, Cper=0)
fit2<-aft.qp(data2$x, data2$y, data2$delta)
fit2
```

`data` *Generating Survival Data from Log-normal AFT Model*

Description

This gives the survival data generated from log-normal AFT model.

Usage

```
data(n, p, r, b1, sig, Cper)
```

Arguments

`n`	sample size.
`p`	the number of covariates. For the AFT model each covariate is generated from Uniform(0, 1) distribution
`r`	correlation between the covariates, r is set to 0 for no correlation.
`b1`	the vector of coefficients.
`sig`	this maintains noise ratio, 1 for no noise.
`Cper`	takes specific value for generating specific censoring percentage, e.g., -0.2 for 30 censoring percentage, 0.0 for 50 censoring percentage and 0.2 for 70 percentages.

Details

Generate survival data from a log-normal AFT model (Y = alpha + X (beta) + error; Y=log(T)) where error is N(0,1). The last largest datum is generated always as censored otherwise censorship is random with censoring time generated from Uniform (c, 2c) for a suitable c.

Value

`y`	logarithmic of survival time
`x`	matrix of covariates of order n by p
`delta`	status; 1 for uncensored, o for censored
`Pper`	censoring percentage

Author(s)

Hasinur Rahaman Khan and Ewart Shaw

References

Khan and Shaw. (2013a). On Dealing with Censored Largest Observations under Weighted Least Squares. CRiSM working paper, Department of Statistics, University of Warwick, UK, No. 13-07. Also available in http://arxiv.org/abs/1312.2533.

Examples

```
#Dataset with zero correlation between the covariates and the medium censoring level
#(50 percent)
data1<-data(n=100, p=2, r=0, b1=c(2,4), sig=1, Cper=0)
data1

#Dataset with moderate correlation between the covariates and the higher censoring level
#(70 percent)
data.r<-data(n=100, p=2, r=0.5, b1=c(2,4), sig=1, Cper=0.2)
data.r
```

---

imputeYn — *Imputing the Censored Largest Datum Under Weighted Least Squares Method*

Description

The method gives imputed values for the last largest censored datum.

Usage

```
imputeYn(X, Y, delta, method = "condMean", beta = NULL)
```

Arguments

X	matrix of covariates. The order is typically n by p.
Y	response. Typically the logarithmic of the survival time.
delta	status; it includes value 1 for uncensored and value 0 for censored subject.
method	one of "condMean (conditional mean)", "condMedian" (conditional median), "RcondMean (resampling based conditional mean)", "RcondMedian (resampling based conditional median)", "PDQ (predicted difference quantity)". Default is "condMean". Here only "PDQ" method works without covariate (X).
beta	coefficients of the covariates estimated by any suitable method chosen by the user. If NULL, the coefficients are estimated using the regularized weighted least squares method with ridge penalty and the censoring constraints and then optimized by quadratic programming. Default is NULL.

Details

The method is developed only for imputing the last largest censored datum if Kaplan-Meier weights are involved in modeling. The treating/imputing methods for the last largest censored datum under weighted least squares is developed for overcoming the problem that the tail correction (Efron, 1967) results in biased and inefficient estimates. Details are discussed in Khan and Shaw (2013a). For details, see Khan and Shaw (2013a, 2013b).

Value

An "imputeYn" object is returned. It includes imputed value for the datum Y(n)+, response with imputed value for Y(n)+, status after reclassifying Y(n)+ into Y(n), coefficients of the covariates obtained with imputed value (newcoefficients) by aft.qp, and coefficients obtained without imputed value (coefficients) by aft.qp.

Author(s)

Hasinur Rahaman Khan and Ewart Shaw

References

Efron, B. (1967). The two sample problem with censored data. In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, Vol. 4, p. 831-853.

Khan and Shaw. (2013a). On Dealing with Censored Largest Observations under Weighted Least Squares. CRiSM working paper, Department of Statistics, University of Warwick, UK, No. 13-07. Also available in http://arxiv.org/abs/1312.2533.

Khan and Shaw (2013b). Variable Selection with The Modified Buckley-James Method and The Dantzig Selector for High-dimensional Survival Data. Proceedings 59th ISI World Statistics Congress, 25-30 August 2013, Hong Kong, p. 4239-4244.

See Also

`print.imputeYn`, `aft.qp`

Examples

```
# For uncorrelated dataset
data1<-data(n=100, p=4, r=0, b1=c(2,2,3,3), sig=1, Cper=0)
imp<-imputeYn(data1$x, data1$y, data1$delta, method = "condMean", beta=NULL)
imp

# For correlated dataset
data2<-data(n=100, p=4, r=0.5, b1=c(2,2,3,3), sig=1, Cper=0)
imp2<-imputeYn(data2$x, data2$y, data2$delta, method = "condMean", beta=NULL)
imp2
```

`imputeYn.extra` *Imputing the Last Largest tied Observations with a new method*

Description

The method provides one step ahead imputed values for tied censored observations.

Usage

```
imputeYn.extra(Y, delta, hc.Yn, method = "km.TPQ", trans.sprob=NULL,
 stime2=NULL, sprob2=NULL, trace=F)
```

Arguments

`Y`	response. Typically the logarithmic of the survival time.
`delta`	status; it includes value 1 for uncensored and value 0 for censored subject.
`hc.Yn`	set of the lifetimes for the last largest censored observations.
`method`	one of "it.PDQ" (repeated predicted difference quantity or called iterative (Khan and Shaw, 2013a)), "km.TPQ" (Kaplan-Meier trend predicted quantity or called "extrapolation" (Khan and Shaw, 2012a)). Default is "km.TPQ".
`trans.sprob`	use only for "km.TPQ". Transformation for the survival probabilities. This transformation is needed if survival probability versus survival time plot is not linear. It takes "exp" for exponential transformation, and any value for respective power transformation. Default is NULL.
`stime2`	use only for "km.TPQ". Survival times after necessary transformation if needed in order to obtain a linear relationship between the survival probability and survival time. Default is NULL.
`sprob2`	use only for "km.TPQ". Survival probability after necessary transformation if needed in order to obtain a linear relationship between the survival probability and survival time. Default is NULL.
`trace`	If TRUE then Kaplan-Meier survival plots will be printed for both data- the original and the data with imputed values. Default is FALSE.

Details

The method is developed for imputing the largest censored observations (can be seen often for heavy censored data) if Kaplan-Meier weights are involved in modeling. For example, if weighted least squares is used then the extra imputing methods will provide one step ahead of the Efron's (1967) tail correction approach. These methods satisfy the Efron's tail correction, achieve less biased estimates, and impute the largest censored observations that are also tied. Furthermore the two methods satisfy the right censoring assumption. If there is heavy censoring toward the right for right censored data, that is found as typical case for many areas like indistry, clinical trials etc, then this function provides the sensible imputations for those censored observations that are the largest and also tied observations. Details are discussed in Khan and Shaw (2013a).

Value

It includes sorted lifetimes, censoring indicators, sorted lifetimes after imputation, censoring indicators after imputation, censored lifetimes for the Y(n)+ observations, imputed lifetimes for the Y(n)+ observations, the survival times, the survival probabilities, the survival times after transformation, the survival probabilities after transformation, the transformation used, and trace.

Author(s)

Hasinur Rahaman Khan and Ewart Shaw

References

Efron, B. (1967). The two sample problem with censored data. In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, Vol. 4, p. 831-853.

Khan and Shaw. (2013a). On Dealing with Censored Largest Observations under Weighted Least Squares. CRiSM working paper, Department of Statistics, University of Warwick, UK, No. 13-07. Also available in http://arxiv.org/abs/1312.2533.

Khan and Shaw (2013b). Variable Selection with The Modified Buckley-James Method and The Dantzig Selector for High-dimensional Survival Data. Proceedings 59th ISI World Statistics Congress, 25-30 August 2013, Hong Kong, p. 4239-4244.

Examples

```
## For Channing House data (heavy censored data)##
## Not run: require(package="boot")
## Not run: time.ch<-channing[1:97,]$time #for male
## Not run: delta.ch<-channing[1:97,]$cens # for male
## Not run: hc.Yn.m<-rep(137,19) # there are 19 last largest censored male each has 137 lifetime
## Not run: imp.PDQ<-imputeYn.extra(time.ch, delta.ch,  hc.Yn=hc.Yn.m,
  method="it.PDQ", trace=T)
## End(Not run)
## Not run: imp.PDQ

## Not run: imp.TPQ<-imputeYn.extra(time.ch, delta.ch, hc.Yn=hc.Yn.m,
  method = "km.TPQ", trace=T)
## End(Not run)
## Not run: imp.TPQ
```

`lss.mod` *Modified Least Squares Pricnciple for Solving the AFT Model*

Description

This provides modified results as the package **lss** developed by Jin et al. (2006).

Usage

```
lss.mod(formula, data, subset, trace = FALSE, mcsize = 500, maxiter = 10,
tolerance = 0.001, cov = FALSE, na.action = na.exclude, residue = FALSE)
```

Arguments

`formula`	specifies a model to be fitted. The response and covariates of the model are separated by a ~ operator. The response, on the left side of ~, should be a Surv object with two columns, of which the first column is the survival time or censored time and the second column is the censoring indicator. The covariates X, on the right side of ~, should be columns with the same length as Surv object. eg: lss(Surv(time, status)~).
`data`	a data frame that contains the Surv objects and covariates.
`subset`	specifies subset of the original data frame that should be used for the model fit.
`trace`	takes logical values TRUE or FALSE. If TRUE, then the summary of every iteration will be kept. Default is FALSE.
`mcsize`	specifies the resampling number. The default is 500.
`maxiter`	specifies the maximum iteration number. The iterations will be stopped after maxiter iterations if the convergence criterion is not met. Default is 50.
`tolerance`	specifies the value of convergence criterion. Default is 0.01.
`cov`	takes logical values TRUE or FALSE. If cov=TRUE, the covariance matrices ofthe least-squares estimator will be printed. Default is FALSE.
`na.action`	takes values na.exclude or na.fail. The default is na.exclude, which deletes the observations with missing values. The other choice is na.fail, which returns an error if any missing values are found.
`residue`	takes two value TRUE or FALSE. Default is FALSE. If TRUE then it shows the residuals.

Details

This is a modified version of the package **lss** developed by Jin et al. (2006). The Least squares principle is developed for solving the AFT model with possibly right censored data. The modification is done by introducing a ridge estimator as an initial estimator that takes into account the correlation between the covariates and censoring information while optimizing using quadratic programming.

Value

The least-squares estimator, the standard error of the least-squares estimator, the Z score and the p-value for testing the hypothesis of beta=0. The covariance matrices of the least-squares estimator, if cov is set to T.

Author(s)

Hasinur Rahaman Khan and Ewart Shaw

References

Jin et al. (2006). On least-squares regression with censored data. Biometrika, 93 (1), 147-161.

Khan and Shaw. (2013a). On Dealing with Censored Largest Observations under Weighted Least Squares. CRiSM working paper, Department of Statistics, University of Warwick, UK, No. 13-07. Also available in `http://arxiv.org/abs/1312.2533`.

See Also

`imputeYn`, `print.imputeYn`

Examples

```
# For uncorrelated datset with four covariates and 50 percent censoring
data1<-data(n=100, p=4, r=0, b1=c(2,2,3,3), sig=1, Cper=0)
require(package="quadprog")
## Not run: fit.lss.mod<-lss.mod(cbind(data1$y, data1$delta) ~ data1$x,mcsize=500,trace=FALSE,maxiter=50,
 tolerance=0.01)
## End(Not run)
## Not run: fit.lss.mod

# For correlated datset with 50 percent censoring
data2<-data(n=100, p=4, r=0.5, b1=c(2,2,3,3), sig=1, Cper=0)
## Not run: fit.lss.mod2<-lss.mod(cbind(data2$y, data2$delta) ~ data2$x, mcsize=500,trace=FALSE,
 maxiter=50, tolerance=0.01)
## End(Not run)
## Not run: fit.lss.mod2
```

`print.imputeYn` *Printing the Important Components of the Class "imputeYn"*

Description

Print the imputed value for the censored largest datum.

Usage

```
## S3 method for class 'imputeYn'
print(x, digits = max(3, getOption("digits") - 3), ...)
```

Arguments

`x`	the object of class "imputeYn".
`digits`	outputs with how many digits. Default is 3.
`...`	not used.

Details

Print the imputed value for the censored largest datum, the coefficients of the covariates with or without the imputed value. The coefficients that consider the imputated value are labelled as "new coefficients."

Value

Gives three important components of the class "imputeYn"- the imputed value for the last largest censored datum, coefficients of the covariates after tail correction, and coefficients of the covariates for dataset with imputed last datum.

Author(s)

Hasinur Rahaman Khan and Ewart Shaw

References

Khan and Shaw. (2013a). On Dealing with Censored Largest Observations under Weighted Least Squares. CRiSM working paper, Department of Statistics, University of Warwick, UK, No. 13-07. Also available in http://arxiv.org/abs/1312.2533.

See Also

imputeYn

Examples

```
# For uncorrelated dataset
data1<-data(n=100, p=4, r=0, b1=c(2,2,3,3), sig=1, Cper=0)
imp<-imputeYn(data1$x, data1$y, data1$delta, method = "condMean", beta=NULL)
print(imp)

# For correlated dataset
data2<-data(n=100, p=4, r=0.5, b1=c(2,2,3,3), sig=1, Cper=0)
imp2<-imputeYn(data2$x, data2$y, data2$delta, method = "condMean", beta=NULL)
print(imp2)
```

Index