Package 'PHInfiniteEstimates'

March 30, 2020

Type Package

Title Tools for Inference in the Presence of a Monotone Likelihood

Version 1.3

Date 2020-03-08

Author John E. Kolassa and Juan Zhang

Maintainer John E. Kolassa <kolassa@stat.rutgers.edu>

Description Proportional hazards estimation in the presence of a partially monotone likelihood has difficulties, in that finite estimators do not exist. These difficulties are related to those arising from logistic and multinomial regression. References for methods are given in the separate function documents.

License GPL-3

Depends survival, lpSolve, mlogit, coxphf

Imports Rdpack (>= 0.7)

RdMacros Rdpack

Encoding UTF-8

RoxygenNote 7.0.1

NeedsCompilation no

Repository CRAN

Date/Publication 2020-03-29 23:00:07 UTC

R topics documented:

`bestbeta` *Newton Raphson Fitter for partial likelihood*

Description

This function implements the approximate conditional inferential approach of Kolassa and Zhang (2019) to proportional hazards regression.

Usage

```
bestbeta(fit, exclude = NULL, start = NULL, touse = NA)
```

Arguments

`fit`	Output from a Cox PH regression, with x=TRUE and y=TRUE
`exclude`	data set with stratum and patient number to exclude.
`start`	Starting value
`touse`	columns of the design matrix to use.

Value

Fitted survival analysis regression parameter of class coxph

References

Kolassa JE and Zhang J (2019). https://higherlogicdownload.s3.amazonaws.com/AMSTAT/fa4dd52c-8429-41d0-abdf-0011047bfa19/UploadedImages/NCB_Conference/Presentations/2019/kolassa_toxslides.pdf. Accessed: 2019-07-14.

`checkresults`	*Produce a graphical assessment of Monte Carlo experiment on fidelity of proportional hazards regression to the uniform ideal.*

Description

This function draws a quantile plot for Monte Carlo assessments of fit to the corrected proportional hazards fit.

Usage

```
checkresults(regnsimulation)
```

Arguments

`regnsimulation`	A matrix with six columns and as many rows as there MC samples.

Value

A list with components of consisting of simulated Wald p-values, likelihood ratio p-values, and corrected likelihood ratio p-values.

`convertbaselineltolr`	*Convert a baseline logit model data set, formatted in the long form as described in the documentation for mlogit.data from mlogit package, to a conditional logistic regression.*

Description

Convert a baseline logit model data set, formatted in the long form as described in the documentation for mlogit.data from mlogit package, to a conditional logistic regression.

Usage

```
convertbaselineltolr(dataset, choice, covs, strs = "chid", alt = "alt")
```

Arguments

`dataset`	in formatted as in the output from mlogit.data of the mlogit packages
`choice`	name of variable in dataset representing choice, a logical variable indicating whether this choice is actually chosen.
`covs`	vector of names of covariates
`strs`	name of variable in data set indicating independent subject
`alt`	name of variable in data set indicating potential choice.

Details

This function implements version of (Kolassa 2016). The multinomial regression is converted to a conditional logistic regression, and methods of (Kolassa 1997) may be applied. This function differs from `convertmtol` of this package in that `convertmtol` treats a less-rich data structure, and this function treats the richer data structure that is an output of `mlogit.data` from package `mlogit`. Data in the example is from Sanders et al. (2007).

Value

a data set on which to apply conditional logistic regression, corresponding to the baseline logit model.

References

Sanders DJ, Whiteley PF, Clarke HD, Stewart M and Winters K (2007). "The British Election Study." https://britishelectionstudy.com.

Kolassa J (1997). "Infinite Parameter Estimates in Logistic Regression." *Scandinavian Journal of Statistics*, **24**, pp. 523–530. doi: 10.1111/14679469.00078.

Kolassa J (2016). "Inference in the Presence of Likelihood Monotonicity for Polytomous and Logistic Regression." *Advances in Pure Mathematics*, **6**, pp. 331-341. doi: 10.4236/apm.2016.65024.

Examples

```
data(voter.ml)
covs<-c("Labor","Liberal.Democrat","education")
#Fit the multinomial regression model, for comparison purposes.
## Lines beginning ## give mlogit syntax that has been made obsolete.
#Add the index attribute to the data set, giving the index of choice made and the index of the
#alternative, and a boolean variable giving choice.
##attributes(voter.ml)$index<-voter.ml[,c("chid","alt")]
##attributes(voter.ml)$choice<-"voter"
##mlogit(voter~1|Labor+Liberal.Democrat+education,data=voter.ml)
mlogit(voter~1|Labor+Liberal.Democrat+education,data=voter.ml,
   chid.var = "chid", alt.var = "alt")
#Convert to a data set allowing treatment as the equivalent conditional logistic regression.
#This result will be processed using reduceLR of this package to give an equivalent conditional
# regression model avoiding infinite estimates.
out<-convertbaselineltolr(voter.ml,"voter",c("Labor","Liberal.Democrat","education"))
#Fit the associated unconditional logistic regression for comparison purposes.
glm(out[,"y"]~out[,1:75],family=binomial)
```

`convertmtol` *Convert a polytomous regression to a conditional logistic regression.*

Description

Convert a polytomous regression to a conditional logistic regression.

Usage

```
convertmtol(xmat, str, yvec, subjects)
```

Arguments

xmat	regression matrix
str	stratum label
yvec	vector of responses
subjects	vector of subject labels passed directly to the output.

Details

Implements version of (Kolassa 2016). The multinomial regression is converted to a conditional logistic regression, and methods of (Kolassa 1997) may be applied. This function differs from `convertbaselineltolr` of this package in that the former treats the richer data structure of package `mlogit`, and this function treats a less complicated structure. Data in the example is the breast cancer data set `breast` of package `coxphf`.

Value

a data set on which to apply conditional logistic regression, corresponding to the multinomial regression model.

References

Kolassa J (1997). "Infinite Parameter Estimates in Logistic Regression." *Scandinavian Journal of Statistics*, **24**, pp. 523–530. doi: 10.1111/14679469.00078.

Kolassa J (2016). "Inference in the Presence of Likelihood Monotonicity for Polytomous and Logistic Regression." *Advances in Pure Mathematics*, **6**, pp. 331-341. doi: 10.4236/apm.2016.65024.

Examples

```
#Uses data set breast from package coxphf.
out<-convertstoml(Surv(breast$TIME,breast$CENS),breast[,c("T","N","G","CD")])
out1<-convertmtol(out[,c("T","N","G","CD")],out[,"chid"],out[,"choice"],
   out[,"patients"])
glmout<-glm.fit(out1$xmat,out1$y,family=binomial())
#In many practice examples, the following line shows which observations to retain
#in the logistic regression example.
moderate<-(fitted(glmout)<1-1.0e-8)&(fitted(glmout)>1.0e-8)
# Proportional hazards fit illustrating infinite estimates.
coxph(Surv(TIME,CENS)~ T+ N+ G+ CD,data=breast)
# Wrong analysis naively removing covariate with infinite estimate
coxph(Surv(TIME,CENS)~ T+ N+ CD,data=breast)
summary(glm((CENS>22)~T+N+G+CD,family=binomial,data=breast))

out2<-reduceLR(out1$xmat,yvec=out1$y,keep="CD")
bestcoxout<-coxph(Surv(TIME,CENS)~ T+ N+ G+ CD,data=breast,
   subset=as.numeric(unique(out1$subjects[out2$moderate])))
```

convertstoml	*Convert a proportional hazards regression to a multinomial regression.*

Description

Convert a proportional hazards regression to a multinomial regression.

Usage

```
convertstoml(survobj, covmat)
```

Arguments

survobj	A survival object, with potentially right censoring.
covmat	a matrix of covariates.

Details

Implements version of (Kolassa and Zhang 2019). The proportional hazards regression is converted to a multinomial regression logistic regression, and methods of (Kolassa 2016) may be applied. This function is intended to produce intermediate results to be passed to `convertmtol`, and then to `reduceLR` of (Kolassa 1997). See examples in the `convertmtol` documentation.

Value

a data set on which to apply conditional multinomial regression, corresponding to the proportional hazards regression analysis.

References

Kolassa J (1997). "Infinite Parameter Estimates in Logistic Regression." *Scandinavian Journal of Statistics*, **24**, pp. 523–530. doi: 10.1111/14679469.00078.

Kolassa J (2016). "Inference in the Presence of Likelihood Monotonicity for Polytomous and Logistic Regression." *Advances in Pure Mathematics*, **6**, pp. 331-341. doi: 10.4236/apm.2016.65024.

Kolassa JE and Zhang J (2019). `https://higherlogicdownload.s3.amazonaws.com/AMSTAT/fa4dd52c-8429-41d0-abdf-0011047bfa19/UploadedImages/NCB_Conference/Presentations/2019/kolassa_toxslides.pdf`. Accessed: 2019-07-14.

drawdiagram — *Draw diagram for toy PH example.*

Description

Draw diagram for toy PH example.

Usage

```
drawdiagram()
```

Value

nothing.

fixcoxph — *Remove observations from a proportional hazards regression, and return the fit of the reduced model.*

Description

This function implements the approximate conditional inferential approach of Kolassa and Zhang (2019) to proportional hazards regression.

Usage

```
fixcoxph(randdat, xxx, iv, verbose = FALSE)
```

Arguments

randdat	A list with at least the component y, representing the Surv() object. I expect that this will be output from an initial non-convergent regression.
xxx	a design matrix for the gregression. I expect that this will be the $x component of the output from an initial non-convergent regression, run with x=TRUE .
iv	name of the variable of interest, as a character string
verbose	logical flag governing printing.

Value

Fitted survival analysis regression parameter of class coxph, fitted form data set with observations forcing infinite estimation removed.

References

Kolassa JE and Zhang J (2019). https://higherlogicdownload.s3.amazonaws.com/AMSTAT/fa4dd52c-8429-41d0-abdf-0011047bfa19/UploadedImages/NCB_Conference/Presentations/2019/kolassa_toxslides.pdf. Accessed: 2019-07-14.

Examples

```
data(breast) # From library coxphf
bcfit<-coxph(Surv(TIME,CENS)~ T+ N+ G+ CD,data=breast,x=TRUE)

fixcoxph(bcfit,bcfit$x,"T",Surv(TIME,CENS)~ T+ N+ G+ CD)
```

`inference` *Perform inference on conditional sample space.*

Description

This function performs classical frequentist statistical inference to a discrete multivariate canonical exponential family. It produces the maximum likelihood estimator, one- and two-sided p-values for the test that model parameters are zero, and providing confidence intervals for the parameters. The discrete probability model is given by a set of possible values of the random vectors, and null weights for these vectors. Such a discrete probability model arises in logistic regression, and this function is envisioned to be applied to the results of a network algorithm for conditional logistic regression. Examples apply this to data from Hirji et al. (1987), citing Goorin et al. (1987).

Usage

```
inference(out, alpha = 0.05, rng = c(-5, 5))
```

Arguments

`out` List of the sort provided by network.

- possible matrix with vectors of possible unconditioned values of the sufficient statistic.
- count count of entries in the conditional distribution.
- obsd Observed value of unconditioned sufficient statistics.

`alpha` Test level, or 1- confidence level.

`rng` Range of possible parameter values.

Value

List with components:

- ospv Observed one-sided p values
- tspv Observed two-sided p value.
- ci confidence interval.
- mle Maximum likelihood estimator.

References

Hirji KF, Mehta CR and Patel NR (1987). "Computing Distributions for Exact Logistic Regression." *Journal of the American Statistical Association*, **82**(400), pp. pp. 1110-1117. ISSN 01621459, doi: 10.2307/2289388.

Goorin A, Perez–Atayde A, Gebhardt M and Andersen J (1987). "Weekly High–Dose Methotrexate and Doxorubicin for Osteosarcoma: The Dana–Farber Cancer Institute/The Children's Hospital – Study III." *Journal of Clinical Oncology*. doi: 10.1200/JCO.1987.5.8.1178.

Examples

```
#Columns in table are:
# Lymphocytic Infiltration (1=low, 0=high)
# Sex (1=male, 0=female)
# Any Ostioid Pathology (1=yes, 0=no)
# Number in LI-Sex-AOP group
# Number in LI-Sex-AOP group with disease free interval greater than 3 y
goorin<-data.frame(LI=c(0,0,0,0,1,1,1,1),Sex=c(0,0,1,1,0,0,1,1),
   AOP=c(0,1,0,1,0,1,0,1),N=c(3,2,4,1,5,5,9,17),Y=c(3,2,4,1,5,3,5,6))

out<-network(goorin[,1:3],goorin[,4],conditionon=1:3,resp=goorin[,5])
inference(out)
```

---

network — *This function enumerates conditional sample spaces associated with logistic regression,*

Description

This function uses a network algorithm to enumerate conditional sample spaces associated with logistic regression, using a minimal version of the algorithm of Hirji et al. (1987).

Usage

```
network(
  dm,
  n = NULL,
  resp = NULL,
  conditionon = NULL,
  sst = NULL,
  addint = TRUE,
  verbose = FALSE
)
```

Arguments

dm	matrix of covariates
n	Vector of number of trials. If null, make them all ones.
resp	vector of successes. Used only to calculate the sufficient statistics, unless sufficient statistics are entered directly. Either resp or sst must be provided.
conditionon	indices of covariate matrix indicating sufficient statistics to be conditioned on.
sst	sufficient statistic vector, if input directly. Otherwise, recomputed from resp.
addint	logical, true if a column of 1s must be added to the covariate matrix.
verbose	logical; if true, print intermediate results.

Details

Examples apply this to data from Hirji et al. (1987), citing Goorin et al. (1987).

Value

For a successful run, a list with components:

- possible matrix with vectors of possible unconditioned values of the sufficient statistic.
- count count of entries in the conditional distribution.
- obsd Observed value of unconditioned sufficient statistics.

For an unsuccessful run (because of input inconsistencies) NA

References

Hirji KF, Mehta CR and Patel NR (1987). "Computing Distributions for Exact Logistic Regression." *Journal of the American Statistical Association*, **82**(400), pp. pp. 1110-1117. ISSN 01621459, doi: 10.2307/2289388.

Goorin A, Perez–Atayde A, Gebhardt M and Andersen J (1987). "Weekly High–Dose Methotrexate and Doxorubicin for Osteosarcoma: The Dana–Farber Cancer Institute/The Children's Hospital – Study III." *Journal of Clinical Oncology*. doi: 10.1200/JCO.1987.5.8.1178.

Examples

```
#Columns in table are:
# Lymphocytic Infiltration (1=low, 0=high)
# Sex (1=male, 0=female)
# Any Ostioid Pathology (1=yes, 0=no)
# Number in LI-Sex-AOP group
# Number in LI-Sex-AOP group with disease free interval greater than 3 y
goorin<-data.frame(LI=c(0,0,0,0,1,1,1,1),Sex=c(0,0,1,1,0,0,1,1),
   AOP=c(0,1,0,1,0,1,0,1),N=c(3,2,4,1,5,5,9,17),Y=c(3,2,4,1,5,3,5,6))

out<-network(goorin[,1:3],goorin[,4],conditionon=1:3,resp=goorin[,5])
inference(out)
```

newllk *Proportional hazards partial likelihood, using Breslow method for ties, excluding some observations.*

Description

This function implements the approximate conditional inferential approach of Kolassa and Zhang (2019) to proportional hazards regression.

Usage

```
newllk(
  beta,
  fit,
  exclude = NULL,
  minus = FALSE,
  keeponly = NULL,
  justd0 = FALSE
)
```

Arguments

`beta`	parameter vector.
`fit`	Output from a Cox PH regression, with x=TRUE and y=TRUE
`exclude`	data set with stratum and patient number to exclude.
`minus`	logical flag to change sign of partial likelyhood
`keeponly`	variables to retain. Keep all if this is null or NA.
`justd0`	logical variable, indicating whether to calculate only the function value and skip derivatives.

Value

a list with components

- d0 partial likelihood
- d1 first derivative vector
- d2 second derivative matrix

References

Kolassa JE and Zhang J (2019). https://higherlogicdownload.s3.amazonaws.com/AMSTAT/fa4dd52c-8429-41d0-abdf-0011047bfa19/UploadedImages/NCB_Conference/Presentations/2019/kolassa_toxslides.pdf. Accessed: 2019-07-14.

PHInfiniteEstimates *PHInfiniteEstimates: Tools for Proportional Hazards Estimation, and Inference on the Associate Parameters, when Other Parameters are Estimated at Infinity.*

Description

The PHInfiniteEstimates package Proportional hazards estimation in the presence of partial likelihood monitonicity has difficulties, in that finite estimators do not exist. These difficulties are related to those arising from logistic regression, addressed by (Kolassa 1997), and multinomial regression, addressed by (Kolassa 2016). Algorithms to provide conditionally similar problems in these contexts are provided.

References

Kolassa J (1997). "Infinite Parameter Estimates in Logistic Regression." *Scandinavian Journal of Statistics*, **24**, pp. 523–530. doi: 10.1111/14679469.00078.

Kolassa J (2016). "Inference in the Presence of Likelihood Monotonicity for Polytomous and Logistic Regression." *Advances in Pure Mathematics*, **6**, pp. 331-341. doi: 10.4236/apm.2016.65024.

reduceLR *Reduce a logistic regression with monotone likelihood to a conditional regression with double descending likelihood.*

Description

Reduce a logistic regression with monotone likelihood to a conditional regression with double descending likelihood.

Usage

```
reduceLR(Z, nvec = NULL, yvec = NULL, keep, sst = NULL)
```

Arguments

Z	regression matrix
nvec	vector of sample sizes
yvec	vector of responses
keep	vector of variable names to block from consideration for removal.
sst	vector of sufficient statistics

Details

This function implements version of Kolassa (1997). It is intended for use with extensions to multinomial regression as in Kolassa (1997) and to survival analysis as in Kolassa and Zhang (2019). The method involves linear optimization that is potentially repeated. Initial calculations were done using a proprietary coding of the simplex, in a way that allowed for later iterations to be restarted from earlier iterations; this computational advantage is not employed here, in favor of computational tools in the public domain and included in the R package lpSolve. Furthermore, Kolassa (1997) removed regressors that became linearly dependent using orthogonalization, but on further reflection this computation is unnecessary. Data in the examples are from Hirji et al. (1987), citing Goorin et al. (1987).

Value

a list with components

- keepme indicators of which variables are retained in the reduced data set
- moderate indications of which observations are retained in the reduced data set
- extreme indicators of which observations are removed in the reduced data set
- toosmall indicator of whether resulting data set is too small to fit the proportional hazards regression

References

Hirji KF, Mehta CR and Patel NR (1987). "Computing Distributions for Exact Logistic Regression." *Journal of the American Statistical Association*, **82**(400), pp. pp. 1110-1117. ISSN 01621459, doi: 10.2307/2289388.

Goorin A, Perez–Atayde A, Gebhardt M and Andersen J (1987). "Weekly High–Dose Methotrexate and Doxorubicin for Osteosarcoma: The Dana–Farber Cancer Institute/The Children's Hospital – Study III." *Journal of Clinical Oncology*. doi: 10.1200/JCO.1987.5.8.1178.

Kolassa J (1997). "Infinite Parameter Estimates in Logistic Regression." *Scandinavian Journal of Statistics*, **24**, pp. 523–530. doi: 10.1111/14679469.00078.

Kolassa J (2016). "Inference in the Presence of Likelihood Monotonicity for Polytomous and Logistic Regression." *Advances in Pure Mathematics*, **6**, pp. 331-341. doi: 10.4236/apm.2016.65024.

Kolassa JE and Zhang J (2019). https://higherlogicdownload.s3.amazonaws.com/AMSTAT/fa4dd52c-8429-41d0-abdf-0011047bfa19/UploadedImages/NCB_Conference/Presentations/2019/kolassa_toxslides.pdf. Accessed: 2019-07-14.

Examples

```
#Cancer Data
Z<-cbind(rep(1,8),c(rep(0,4),rep(1,4)),rep(c(0,0,1,1),2),rep(c(0,1),4))
dimnames(Z)<-list(NULL,c("1","LI","SEX","AOP"))
nvec<-c(3,2,4,1,5,5,9,17); yvec<-c(3,2,4,1,5,3,5,6)
reduceLR(Z,nvec,yvec,c("SEX","AOP"))
#CD4, CD8 data
Z<-cbind(1,c(0,0,1,1,0,0,1,0),c(0,0,0,0,1,1,0,1),c(0,0,0,0,0,1,1,0),c(0,1,0,1,0,0,0,1))
dimnames(Z)<-list(NULL,c("1","CD41","CD42","CD81","CD82"))
```

```
nvec<-c(7,1,7,2,2,13,12,3); yvec<-c(4,1,2,2,0,0,4,1)
reduceLR(Z,nvec,yvec,"CD41")
```

---

testp *Simulate Weibull survival data from a model, perform reduction to remove infinite estimates, and calculate p values.*

Description

Operating characteristics for the approximate conditional inferential approach to proportional hazards.

Usage

```
testp(
  dataset,
  myformula,
  iv,
  ctime,
  nsamp = 10000,
  add = NULL,
  nobs = NA,
  half = FALSE,
  verbose = FALSE
)
```

Arguments

dataset	the data set to use
myformula	the formula for the Cox regression
iv	name of the variable of interest, as a character string
ctime	fixed censoring time
nsamp	number of samples.
add	preliminary results, if any.
nobs	number of observations in target models, if different from that of dataset.
half	logical flag triggering a less extreme simulation by dividing the Weibull regression parameters in half.
verbose	logical flag triggering intermediate messaging.

Details

This function is intended to verify the operating characteristics of the approximate conditional inferential approach of Kolassa and Zhang (2019) to proportional hazards regression. A Weibull regression model, corresponding to the proportional hazards regression model, is fit to the data, and new data sets are simulated from this model. P-values are calculated for these new data sets, and their empirical distribution is compared to the theoretical uniform distribution.

Value

a list with components

- out matrix with columns corresponding to p-values.
- seed random seed
- bad unused.
- srreg parametric lifetime regression

References

Kolassa JE and Zhang J (2019). https://higherlogicdownload.s3.amazonaws.com/AMSTAT/fa4dd52c-8429-41d0-abdf-0011047bfa19/UploadedImages/NCB_Conference/Presentations/2019/kolassa_toxslides.pdf. Accessed: 2019-07-14.

Examples

```
data(breast)

breattestp<-testp(breast,Surv(TIME,CENS)~ T+ N+ G+ CD,"T",72,nsamp=100,verbose=TRUE)
```

`voter.ml` *Subset of British elections data used in (Kolassa 2016). Data are from (Sanders et al. 2007).*

Description

Subset of British elections data used in (Kolassa 2016). Data are from (Sanders et al. 2007).

Usage

```
data(voter.ml)
```

References

Sanders DJ, Whiteley PF, Clarke HD, Stewart M and Winters K (2007). "The British Election Study." https://britishelectionstudy.com. Kolassa J (2016). "Inference in the Presence of Likelihood Monotonicity for Polytomous and Logistic Regression." *Advances in Pure Mathematics*, **6**, pp. 331-341. doi: 10.4236/apm.2016.65024.

Examples

```
data(voter.ml)
```

Index